Science
AAAS

Supplementary Materials for

Coordinated control of neuronal differentiation and wiring by sustained transcription factors

Mehmet Neset Özel, Claudia Skok Gibbs, Isabel Holguera, Mennah Soliman, Richard Bonneau, Claude Desplan

Correspondence to: no24@nyu.edu (M.N.Ö.), rb133@nyu.edu (R.B.), cd38@nyu.edu (C.D.)

This PDF file includes:

Other Supplementary Materials for this manuscript include the following:

Materials and Methods

Determination of candidate terminal selectors

We used Seurat v3 (*43*) to determine the sets of differentially expressed TFs that are continuously maintained in each neuron from our developmental scRNA-seq atlas of the optic lobes (*1*). First, we subset Seurat objects from each stage to retain only the neuronal clusters. Cluster markers were then calculated between these neuronal clusters only using the FindAllMarkers function (only.pos=TRUE, otherwise default parameters) which we filtered for transcription factors using a list of 629 genes from FlyBase. From these tables, we determined for each cluster, the TFs that were picked as its markers in adults as well as in at least 4 out of the 5 developmental stages (to buffer against potential errors due to technical variations in small clusters). A total of 113 TFs were found to be consistent markers of at least one cluster throughout development and in adults (*1*).

Relying on cluster markers alone can be misleading: Some genes may fail to be picked up as markers for certain clusters despite being expressed if the level of expression is not significantly higher than the average of the rest of the dataset. This could be true even for differentially expressed genes. Similarly, some pan-neuronal (or ubiquitous) genes might be picked up as markers in some clusters if the level of expression is variable. We thus turned to mixture modeling data that we previously generated to binarize the expression of every gene in each cluster and stage (*1*). This method considers, for each gene, the level of expression in all clusters in order to assign a probability of expression to each cluster (*33*). For each of the 174 neuronal clusters with mixture modeling available at all 6 stages, we determined which of the 113 marker TFs had a non-zero probably of expression for a given cluster at every stage. However, this can also be error-prone for some genes when there is not a bimodal distribution of expression levels among the clusters. Therefore, we combined the two methods: a gene was selected as a candidate terminal selector for a given cluster if i) it was found as a consistent marker of the cluster as described above, or ii) if it was continuously expressed at all stages according to mixture modeling. Lastly, we discarded all TFs that were found to be expressed in more than 150 clusters according to mixture modeling (likely pan-neuronal) in at least two of the stages. The final table (**Fig. S1A**, **Table S1**) contains 95 TFs in total, with a median of 10 TFs per cluster (min:4, max:16) whose combinations are unique to each cluster.

It is important to note that while these results represent our best efforts to binarize a complex developmental gene expression landscape, there may still be errors, especially for

genes with high variations in their levels of expression, and for smaller clusters that may be present in less than 20-30 cells in some of the stages. We thus recommend manual inspection of candidate selector expression patterns for cell types of interest as we did here (**Fig. 1B**, **Fig. 3D**) before designing experiments.

In order to calculate the distances between cluster pairs (**Fig. S1D**), we used the Seurat V4 function AggregateExpression() with the option return.seurat=TRUE to obtain averaged transcriptomes for each cluster. We then calculated the pairwise Euclidian distances between all 174 neuronal clusters (15,225 pairs), using the scaled expression values of 95 candidate terminal selectors (x-axis) or 2467 cluster markers that were ranked within the top 100 (by logFC) markers of any neuronal cluster (y-axis) at P50. Using scaled expression ensures that each gene contributes to the distance measure equally regardless of expression level. Note that the pairs involving cluster 102 appear to form a distinct regime in **Fig. S1D**. This is most likely an artefact resulting from the fact that cluster 102 is a heterogenous mixture of contaminating central brain neurons all present at very low stoichiometry, rather than an actual cell type.

To produce the cluster dendrograms in **Fig. S2**, we used the Seurat function BuildClusterTree() on the scale.data slot with features specified as all candidate terminal selectors (**S2A**), 36 homeobox candidate selectors (**S2B**) or the top 100 cluster markers as described above (**S2C**). Clusters were manually grouped into 22 branches in **Fig. S2A** and assigned specific colors accordingly, except for 6 “outlier” clusters that were not given any color. The same color scheme was then applied to the trees in **S2B** and **S2C** as well to highlight the consistency (or lack thereof) between the hierarchical organization of clusters based on different gene sets.

Animal husbandry and genetics

All experiments were performed with both female and male (mixed) *D. melanogaster* (except for sequencing experiments, where only females were used) maintained at 25 or 29°C. The precise genotypes and temperatures used for experiments in each figure panel are detailed in **Table S4**. Generally, MARCM and sequencing experiments were performed at 25°C, while RNAi and most ectopic expression experiments were performed at 29°C. Source details for all fly strains are specified in **Table S5**. Adult dissections were performed within 2 days of eclosion. For pupal dissections, white pupae (P0) were selected and then aged until P50 (50h at 25°C or ~35h at 29°C) or P30 (30h at 25°C or ~20h at 29°C). Late L3 experiments were performed using wandering larvae. Experiments involving tub-Gal80^{ts} were reared at 18°C and then

switched to 29°C two days before dissection (L3 dissections) or at late L3 stage (later dissections). Heat shocks at 37°C were applied for 15 minutes at L3 stage (MARCM) or for 7-8 minutes at early pupal (P0-10) stages (flip-out). All genotypes were analyzed in at least 5 biological replicates (brains); sample sizes are indicated in the figure legends. Animals fitting the gender and age criteria were selected randomly from the fly vials for all experiments.

Antibody generation

Polyclonal antibodies against Drgx, Pdm3, SoxN, Mef2, Vsx2, Brp and Repo were generated by Genscript (https://www.genscript.com/).

Amino acids 431-587 of Drgx protein was used to immunize rats:

SNSVAELRRKAQEHSAALLQSLHAAAAAGLAFPGLHLPPLSFAHHPALGQHVVNHNNNNTM
RMKHEAQDMTMNGLGPGSGSGSGSGSAGGGTSSAALLDLAESAVAYQQQQHATLSPPTT
PTQQSSGGVAATEGSPGSGAIAGSGSLNGNVVLTKME

Amino acids 801-1292 of Pdm3 protein was used to immunize rats:

STSAVSSTLPQISLRHPDELTAPQMDLKPLELSASTSPPAPPPRHHFGHSLRGSSTVSPKHS
PQGRMGGSGGSTTTGMNLSQHHERHDRLERLERQERHERRSHTPTATATRASVSSSSSA
GHHGGSLPSGRLSPPSSAPSNSAANSISDRGYTSPLFRTHSPQGHALSLGGSPRLERDYLG
NGPSSGTATSTSSCGAPTAAGSSATANVLSSINRLNASNGELTITKSLGAPTATATRASSASP
RDDSPGPGPSTSSVSHMQPLKLSPSSRSEPPHLSPNGNDNDNDLLMDSPNEPTINQATTNV
VDGIDLDEIKEFAKAFKLRRLSLGLTQTQVGQALSVTEGPAYSQSAICSSALAAQMYAAQLST
QQQNMFEKLDITPKSAQKIKPVLERWMKEAEESHWNRYKSGQNHLTDYIGVEPSKKRKRRT
SFTPQALELLNAHFERNTHPSGTEITGLAHQLGYEREVIRIWFCNKRQALKNTVRMMSKGMV

The following peptide from SoxN protein (*55*) was used to immunize rabbits:

LHYQTDSPDLQQQHQS**C**

The entire ORF of the Mef2 protein purified with a His-tag was used to immunize rabbits:

MGRKKIQISRITDERNRQVTFNKRKFGVMKKAYELSVLCDCEIALIIFSSSNKLYQYASTDMDR
VLLKYTEYNEPHESLTNKNIIEKENKNGVMSPDSPEAETDYTLTPRTEAKYNKIDEEFQNMMQ
RNQMAIGGAGAPRQLPNSSYTLPVSVPVPGSYGDNLLQASPQMSHTNISPRPSSSETDSGG
MSLIIYPSGSMLEMSNGYPHSHSPLVGSPSPGPSPGIAHHLSIKQQSPGSQNGRASNLRVVI
PPTIAPIPPNMSAPDDVGYADQRQSQTSLNTPVVTLQTPIPALTSYSFGAQDFSSSGVMNSA
DIMSLNTWHQGLVPHSSLSHLAVSNSTPPPATSPVSIKVKAEPQSPPRDLSASGHQQNSNG
STGSGGSSSSTSSNASGGAGGGGAVSAANVITHLNNVSVLAGGPSGQGGGGGGGGSNGN

VEQATNLSVLSHAQQHHLGMPNSRPSSTGHITPTPGAPSSDQDVRLAAVAVQQQQQQPHQ QQQLGDYDAPNHKRPRISGGWGTHHHHHH

The following epitope from Vsx2 protein was used to immunize rabbits:

TEAPTDLTTTAGATVAKERQTPTPPKTTNATMATAATSAATAATPTNAAEGNLTSVSEPQQQ PQQQQQEQQHHQPHHHQYREHHQMTMAAASRMAYFNAHAAVAAAFMPHQLAAAVHHHH QHQHQHHPHHHPHHPHGAVGGPPPPPPMQHHHPHHPHHPLLHAQGFPQLKSFAAGAGTC LPGSLAPKDFGMESLNGFGVGPNSKKKKK

Amino acids 221-401 of Brp protein was used to immunize guinea pigs:

TDVQRQQLEQQQKQLEEVRKQIDNQAKATEGERKIIDEQRKQIDAKRKDIEEKEKKMAEFDV QLRKRKEQMDQLEKSLQTQGGGAAAAGELNKKLMDTQRQLEACVKELQNTKEEHKKAATE TERLLQLVQMSQEEQNAKEKTI MDLQQALKIAQAKVKQAQTQQQQQQDAGPAGFLKSFF

The following epitope from Repo protein was used to immunize guinea pigs:

MEHDSFDDPIFGEFGGGPLNPLGAKPLMPTTTAMHPVMLGSVHELCSQQQQQQQQQRLPD CNTILPNGGGGGAGSGGAGG SPNYVTKLDFVNKMGCYSPSQKYEYISAPQKLVEHHHHHH

Drgx$^{\Delta Tm1}$ allele

CRISPR-mediated mutagenesis was performed (WellGenetics Inc) to produce a genomic deletion of the Tm1-specific enhancer element we identified in the *Drgx* locus (see below the snATAC-seq analysis section). In brief, the upstream gRNA sequence AGCAATTCGCTATCCTTCGC[TGG] and downstream gRNA sequence TTCGACTGGACAGCTTAGTC[TGG] were cloned into U6 promoter plasmid(s) separately. The cassette 3xP3-RFP, which contains a floxed 3xP3-RFP, and two homology arms were cloned into pUC57-Kan as donor template for repair. gRNAs and hs-Cas9 were supplied in DNA plasmids, together with donor plasmid for microinjection into embryos of the control strain w^{1118}. F1 flies carrying selection marker of 3xP3-RFP were further validated by genomic PCR and sequencing. CRISPR generated a 1,038 bp deletion within the 4th intron of *Drgx*. 3xP3-RFP cassette was then excised through Cre/LoxP recombination, and the excision was validated by PCR and sequencing.

Immunohistochemistry and imaging

Fly brains were dissected in ice-cold Schneider's Insect Medium (SIM) and fixed in 4% formaldehyde (in PBS) at room temperature for 30-50 minutes and washed in PBST (PBS + 0.3% Triton X-100). They were then incubated in primary antibodies (in PBST + 5% horse

serum) for 1 or 2 days at 4°C, washed three times in PBST for 10 minutes and then incubated in secondary antibodies at 4°C overnight followed by 2x 15 min washes in PBST and an additional 15 min in PBS. They were then mounted in Slowfade and imaged with a Leica SP8 confocal microscope using a 63x (NA=1.3) glycerol objective with xy resolution <200nm and z resolution = 500nm.

The following primary antibodies were used: Chicken anti-GFP (1:1000), mouse anti-V5 (1:250), rat anti-NCad (1:20), mouse anti-Brp (nc82, 1:50), guinea pig anti-Brp (1:300), mouse anti-Aop (1:100), rabbit anti-SoxN (1:250), rabbit anti-Mef2 (1:250), rat anti-Drgx (1:250), rat anti-Pdm3 (1:500), guinea pig anti-Pdm3, guinea pig anti-Repo (1:500), rabbit anti-Vmat (*56*) (1:250), guinea pig anti-Vsx1 (1:300), rabbit anti-Vsx2 (1:500), guinea pig anti-Runt (1:500), rabbit cleaved anti-Dcp-1 (1:200), rabbit P-p44/42 anti-MAPK (1:100).

The following secondary antibodies (all donkey) were used: anti-chicken Alexa 488, anti-rabbit DyLight 405, anti-rabbit Alexa 488, anti-rabbit Cy3, anti-rabbit Alexa 647, anti-mouse Alexa 488, anti-mouse Alexa 555, anti-mouse Alexa 647, anti-rat Alexa 488, anti-rat Cy3, anti-rat Alexa 647, anti-guinea pig DyLight 405, anti-guinea pig Alexa 488, anti-guinea pig Cy3, anti-guinea pig Alexa 647. All were used at 1:300 dilution except for anti-chicken Alexa 488 that was used at 1:500. The origin of all antibodies used is listed in **Table S5**.

HCR RNA-FISH

Hybridization chain reaction fluorescence in situ hybridization experiments to detect mRNAs (HCR RNA-FISH) were performed as previously described (*57*). Custom probes were designed by Molecular Instruments against the following transcripts: *5-HT7* (NM_079860.3, amplifier: B1), *Octβ1R* (NM_001170219.2, amplifier: B2), *Or63a* (NM_079171.4, amplifier: B3) and *Dh44-R1* (NM_137116.3, amplifier: B4). Hybridization, amplification buffer and wash buffers, and the fluorophore-labelled amplification hairpins (Alexa 546 for B3 and B4 and Alexa 647 for B1 and B2) were also obtained from Molecular Instruments.

Briefly, brains were dissected in ice-cold SIM and fixed in 4% formaldehyde (in PBS) at room temperature for 20 minutes and washed 3x10 minutes in PBST (PBS + 0.3% Tween-20) and for another 5 minutes in 5X SSCT. The samples were pre-hybridized in hybridization buffer for 30 minutes, and then incubated for 2 nights in hybridization buffer containing the probes at 37°C. Afterwards, they were washed 4x5 minutes in wash buffer at 37°C and for another 5 minutes in 5X SSCT. They were then pre-amplified for 30 minutes in amplification buffer (without hairpins) before being incubated overnight in amplification buffer with the hairpins at

room temperature. Finally, they were washed three times (5, 5 and 10 minutes) in 5X SSCT and another 5 minutes in PBS before mounting in Slowfade. All solutions were prepared using nuclease-free water and all incubations were performed on a nutator.

Image analysis

Images were analyzed and prepared using Imaris 7 (Bitplane) and Photoshop 2021 (Adobe). They were presented either as maximum intensity projections to highlight TF/RNA staining patterns or as 3D reconstructions using the “Blend” mode with “glow” colormap in Imaris to highlight neuronal morphologies.

Most quantifications were done manually by determining the proportion of each cell type (ignoring irrelevant cell types that may be labeled by the drivers, such as non-Tm neurons in experiments related to Tm1/2/4/6) in each brain (biological replicates) based on their stereotypical morphology, or expression of specific TFs, as specified in the figure legends. As the number of labeled neurons per brain show high variation in MARCM and flip-out experiments, all statistics were performed using the per brain proportions of indicated cell types (or their perturbed variations) and parametric, two-sided t-tests unless indicated otherwise. Bar and pie charts were generated using Prism 9 (GraphPad).

Quantification of the number of apoptotic cells for experiments shown in **Figure S4B-C** was performed semi-automatically using Imaris. Colocalization channels were created between the GFP (min. intensity=30) and Pdm3 (min. intensity=50) channels, as well as between the GFP (min. intensity=30) and cDcp1 (min. intensity=30) channels. Surface objects were then created separately for the i) GFP, ii) GFP-Pdm3 Coloc, iii) GFP-cDcp1 Coloc channels, with the same thresholds, filtered for “Quality” above 5 and “Number of Voxels” above 500. The number of cells that co-express GFP+Pdm3+cDcp1 were counted manually based on the overlap of Surface objects from (ii) and (iii). For each brain, the ratio of all dying cells were then calculated as the number of Surface objects in (iii) divided by the number of objects in (i). The ratio of dying Tm2 neurons were calculated as the ratio of triple labeled cells divided by the number of objects in (ii). The analysis was restricted to the medulla cortex as T2 and TE neurons are also labeled by GFP and Pdm3 in this experiment, but they were excluded based on their location. For experiments shown in **Figure S5F-G**, apoptotic Tm1 and Tm4 cells were counted manually since membrane GFP was used in these experiments rather than nuclear GFP. Tm4s display strong nuclear Aop staining while in Tm1s Aop can only be observed as extranuclear rings (and they are also labeled by Drgx in the control brains).

To quantify HCR-FISH experiments, somas of labeled neurons (based on CD4-tdGFP signal) were manually segmented in Imaris to create surface objects (using the “circle” tool under the “Contour” tab). Mean intensity values per cell were then exported to spreadsheets for each of FISH channel. Since the signal intensities can vary significantly across samples, for each cell, we calculated the ratio of Dh44-R1 (Mi15 marker) signal to 5-HT7 (Dm2 marker) signal, or similarly the ratio of Or63a (Mi15) signal to Octβ1R (Dm2) signal to compare across conditions (**Fig. S7F-H**).

Single-cell RNA sequencing

Female white pupae (P0) were selected and maintained at 25°C until P50. Brains were dissected in ice-cold SIM and incubated for 20 minutes (at 25°C) in a Collagenase/Dispase cocktail, followed by washes and mechanical dissociation of the optic lobes, using a previously described protocol (*1*). 5 brains (10 optic lobes) were dissociated together for each condition (Control and UAS-Pdm3). We then isolated the cells expressing the nuclear GFP *Stinger* from these single-cell suspensions using a BD FACSAria (see **Supplementary Information** for the FACS gating strategy). Samples were immediately processed for scRNA-seq.

Droplet-based purification, amplification and barcoding of single-cell transcriptomes were performed using Chromium™ Single Cell 3’ Reagent Kit v3.1 (10X Genomics) as described in the manufacturer’s manual. As the output concentrations from FACS were typically less than 400 cells/µl, we loaded the entire reaction volume (43.2 µl) with single-cell suspension (no extra water). The samples of each condition were processed back-to-back at each step of the entire protocol and were run on the same 10x chip to minimize batch effects. The libraries were subjected to paired-end sequencing (28x8x91) with Illumina NovaSeq 6000 (Genomics Core at NYU CGSB) to ~130,000 reads per cell (88-89% sequencing saturation, as reported by CellRanger). We mapped the sequenced libraries to the *D. melanogaster* genome assembly BDGP6.88 using CellRanger 5.0.1, which retained 3,830 cells from Control and 3,813 cells from UAS-Pdm3 library, with a median of ~2500 genes per cell. All downstream analyses were performed with Seurat v4 (*58*). The libraries were merged (without integration) and further filtered to retain the cells with 2,000 to 20,000 UMIs and less than 5% of UMIs corresponding to mitochondrial genes, to a total of 6693 cells (3303 Control, 3390 UAS-Pdm3).

scRNA-seq analysis

The data was log-normalized, followed by determination of the top 2000 variable features and scaling with default parameters in Seurat. Cell type assignments were done in a

supervised manner using a machine learning model we had previously trained on the scRNA-seq atlas of the entire optic lobe at P50 (*1*). tSNE reduction was then calculated using the top 50 principal components (**Fig. S4A**). Surprisingly, these libraries contained many cell types which we never observed using the *TmX/Tm2-Gal4* with a membrane-bound GFP (**Fig. S3D**), including (but not limited to) Mi1/4/9 and Tm3/5/9/20 types (**Fig. S4A**). It is possible that the driver has weak background expression in these other cell types, and that the very strong nuclear localization of *Stinger* allows GFP signal to be concentrated there (as opposed to a membrane GFP) allowing these cells to be picked up by FACS. Indeed, a high number of nuclei with weak GFP signal (that were never labeled by Pdm3 or Aop) could be observed in these brains (**Fig. S4B**). We thus restricted all further analysis to the cells classified as Tm1/2/4/6 or T2 neurons by the machine learning model (**Fig. 1H,J**). In addition, we realized that one of the brains dissected for the Control library was in fact a male, based on *roX1/2* expression. These cells were excluded from all the differential gene expression analyses below. They were still included in the UMAPs since those were calculated using the scaled data slot, which was corrected (regressed out) for sex differences based on *roX1* and *roX2* expression.

The top 500 variable features were re-determined after subsetting of the data, which were then re-scaled accordingly. The UMAP reduction (**Fig. S4D**, **Fig. 1H-O**) was calculated using the top 6 principal components. The unsupervised clustering was performed using the top 5 principal components and resolution 0.4 (otherwise default parameters). Differentially expressed genes (DEGs) between the indicated clusters in **Figure 1K** and **Table S2** were then determined using the FindMarkers function with default parameters, filtered to only retain the markers with adjusted p-value < 0.01. Control and UAS-Pdm3 libraries in this experiment are expected to have some batch effects. To avoid picking up DEGs which may be driven by such effects, we also calculated the list of markers (adj. p-value < 0.01) between the 4 Tm types in the *control library only*, using the FindAllMarkers function (only.pos=TRUE, otherwise default parameters). Reasoning that any gene that does not normally display differential expression between the Tm neurons should not be regulated by Pdm3 in these cell types, the number of DEGs displayed in **Figure 1K** and the full marker lists provided in **Table S2** only include such genes that were also part of this list of 1395 “Control Tm markers”.

snATAC-seq analysis

We used Seurat v4 and Signac 1.3.0 (*59*) to analyze the snATAC-seq datasets generated from whole fly brains by (*28*) at stages Adult, P48 and P24. Peak/count matrices generated by the authors, based on pre-determined regulatory regions in the fly genome, were

extracted from the CisTopic object and separate Seurat/Signac objects were created for all three stages. The following steps were then taken separately for each object/stage: The data were TF-IDF normalized and all features were set as variable by running the FindTopFeatures function with “min.cutoff = 'q0'” option. Latent semantic indexing (LSI) reduction was then calculated using the RunSVD function (n=200). We then clustered the data using the top 120 LSI dimensions (excluding the first one, which correlated strongly with sequencing depth) and with ‘algorithm’ and ‘resolution’ set to 3 in the FindClusters function. Label transfer was applied from Adult to P48 and from P48 to P24 datasets to establish (rough) linkages between the clusters of different stages. The anchors were found between the datasets with 100 CCA dimensions, then the labels were transferred with the TransferData function (weight.reduction=lsi and dims=2:80).

In order to annotate the datasets, gene activity, i.e. ‘pseudo expression’, matrices were calculated by counting the number of reads (directly from the fragment files) within 5kb upstream of TSS for every gene. These were then used to perform label transfer from the scRNA-seq datasets at corresponding stages: Adult, P50 and P30 (*1*). The anchors were found between the datasets using the variable features of the RNA dataset (CCA, 150 dims), then the labels were transferred with the TransferData function (weight.reduction=lsi and dims=1:120). The quality of label transfer was variable and particularly low at P24; we thus first focused on the Adult dataset to establish correspondence between the scRNA- and snATAC-seq clusters. Every cluster was inspected manually for its correspondence to the optic lobe clusters in the scRNA-seq dataset. If a substantial number (depending on the expected frequency of clusters) of cells belonging to any (neuronal) RNA cluster corresponded to a clearly defined ATAC cluster, those clusters were retained for further analysis (**Fig. S6A**). This process excludes all glia and the vast majority of central brain neurons, but it may also exclude optic lobe neurons that derive from central brain which are typically present in lower frequencies and may not cluster well. Using the linkages we established above between the different stages, we determined in P48 and P24 datasets the clusters that corresponded to optic lobe neuronal clusters in the Adult dataset and discarded the rest.

After isolating the optic lobe neurons in all 3 stages, we reapplied TF-IDF and SVD, and re-clustered the datasets to maximize resolution. The Adult optic lobe dataset was clustered with LSI dimensions 3 to 80 and resolution=5, while we used the dimensions 2 to 80 and resolution=3 for the P48 and P24 datasets. Multiple clusters corresponding to T4/T5 neurons were merged. We obtained 45 final clusters of optic lobe neurons at the adult stage, 37 of which

could be confidently linked to defined (though not all annotated) clusters (or groups of clusters) in the scRNA-seq atlas (**Fig. S6B**, the clusters that are still likely optic lobe neurons but could not be confidently linked to well-defined RNA clusters are grayed out). There were 59 initial clusters in the P48 dataset and 41 in P24. We took a multi-step process to harmonize the cell type assignments between different stages. First, label transfer was performed from Adult to P48 dataset, as described above for the whole datasets. We took advantage of both these transferred labels and the transferred labels from the P50 RNA dataset to manually inspect all clusters. Unlike the strategy we previously employed to harmonize the cell-type assignments in scRNA-seq datasets of different stages (*1*), label transfer between different stages was far less robust in snATAC-seq; thus we mostly respected the unsupervised assignments at each stage. If most (typically >50% but in a few cases less) of a P48 cluster was classified as a specific Adult cluster (and no other), the entire cluster was assigned the same number as in the Adult. Since the clustering resolution was higher at P48, that meant in several cases merging of clusters. However, we later subclustered the lamina neurons L1-5 at this stage, to resolve all 5 types (**Fig. S6C**, blue ellipses). In some cases where a P48 cluster was mapped to multiple Adult clusters, the P48 cluster was split based on the label transfer classifications. In other cases where the correspondence of a P48 cluster to the Adult clusters was unclear, they were assigned to new identities (not present in the Adult dataset). This was the case for TE neurons (that are not present in adult brains), but also for TmY14, Lawf and Lai neurons, which could not be mapped reliably in the Adult dataset. The remaining P48 clusters that had inconsistent mapping to Adult clusters and also could not be reliably linked to a P50 RNA cluster were kept but grayed out in tSNE (**Fig. S6C**). The process was then repeated from P48 to P24 dataset, which had the lowest clustering resolution as well as the least reliable mapping from the (P30) RNA dataset. This meant that most of the clusters had to be split manually based on the transferred labels from the P48 dataset to maintain the same resolution (**Fig. S6D**). For the same reason, a much larger proportion of the dataset remained unlinked to specific cell types and were grayed out in the tSNE.

The accessibility tracks were generated for the indicated clusters (**Fig. 2D**, **Fig. S6E**) using the CoveragePlot function with default parameters. The sequence of the genomic range chr2L:3679100-3679800 which encompasses the Tm1-specific peak in *Drgx* locus was extracted from Ensembl for motif enrichment analysis. We compiled a list of 1288 TF binding motifs, representing 411 TFs, from the CisBP database, including only those determined using direct evidence from *Drosophila*. We used the MEME suite (https://meme-

suite.org/meme/tools/ame) tool AME (*60*) to identify the enriched motifs. The only motifs that were found to be enriched with an E-value < 100 belonged to the TFs *Klu*, *org-1* and *nau*; the latter two are not expressed in the optic lobe at any stage according to our scRNA-seq datasets.

Network Inference

We inferred GRN models for two distinct groups of neurons using data from developmental stages P24-P50: i) Tm1/2/4/6 neurons that we studied in this paper (Tm network), ii) Lamina monopolar neurons L1-5 whose development have been extensively studied previously (Lamina network).

Construction of priors: Prior connectivity matrices (tables of 1s and 0s linking TFs to target genes) were determined separately for Tm and Lamina neurons using the P48 snATAC-seq data (**Fig. S6C**), Signac v1.3.0 and the Inferelator-Prior software (*40*) available on GitHub, from the release branch v0.2.3 faf5e47.

For each prior, we used GTF and fasta files from the *D. melanogaster* genome assembly BDGP6.88. MEME suite tool FIMO (*61*) was implemented with the default 1e-4 p-value threshold to scan the *Drosophila* genome for the set of TF motifs described in the previous section. Several TFs motifs, specifically those belonging to: 'ct', 'dan', 'danr', 'Eip78C', 'H2.0', 'jumu', 'Sox100B', 'Tbp', 'vis', 'Met', were never found in the genome using the default FIMO threshold (likely due to their lower information content), thus for those motifs we used a 1e-3 p-value threshold to receive scores. All TF motifs were scored and filtered based on their information content as part of the Inferelator-Prior workflow. The refined list of scored TF motifs was used to assign regulatory relationships to target genes by using a 10kb window up and downstream of each gene body. These interactions were then clustered using the sklearn DBSCAN package with parameter epsilon=1 to retain a very sparse (<0.5% density on average per TF) matrix containing only the highest confidence interactions. The command line arguments for the Inferelator-Prior pipeline can be accessed at: https://github.com/cskokgibbs/DMOLN_NetworkScripts. This pipeline requires as input a .bed file defining genomic ranges to be scanned for motifs. We tested 3 different priors (NoBed, ChromA and MergedDA) for both Tm and Lamina networks whose underlying peak sets were determined with different strategies as described below. Please note that even though **Figure S8** displays the benchmarking results for networks built with all 3 priors, the final networks used in this work were based on the MergedDA priors.

NoBed: No constraining peak file was used to build these priors. The entire 10kb up/down window for every gene was scanned for motifs. Note that this implies the NoBed priors for Tm and Lamina networks were therefore identical.

ChromA: For these priors, we extracted all fragments originating from the Tm or Lamina neurons in the P48 snATAC-seq (**Fig. S4C**, red and blue ellipses) and called peaks using ChromA (*62*) version: fcfd120 branch: Any-Genome, using the parameters: atac -spec dm6. These peak sets were then used directly as input to Inferelator-Prior. Therefore, for these priors, all regions accessible in any of the 4 Tm or 5 lamina neurons were scanned for motifs without considering differential accessibility.

MergedDA: We determined differentially accessible (DA) regions between the Tm or Lamina neurons with two different approaches. First, we used directly the peak set from (*28*). We subset the P48 Signac object for either the 4 Tm clusters or the 5 lamina clusters (**Fig. S6C**) and calculated the DA peaks using the FindAllMarkers function (logfc.threshold = 0.1, test.use = 'LR', latent.vars = 'nCount_peaks'). However, we noticed that these pre-determined regulatory regions often contained multiple peaks whose accessibility can vary independently, which could introduce artifacts. With this in mind, we re-built the peak/count matrices for Tm and Lamina clusters from the Fragments file using the FeatureMatrix function, using as features the peaks determined above by ChromA as accessible in Tm or Lamina clusters. DA peaks were then calculated for each group using the same parameters. This approach may have its own limitations, primarily due to sensitivity of peak calling since this was done separately on relatively small groups of cells here rather than the entire dataset. We thus decided to merge the two peak sets (separately for Tm and Lamina) using the GenomicRanges function reduce (min.gapwidth = 0) to obtain the final as input to Inferalator-Prior.

MergedDA networks outperformed the other priors in all metrics (**Fig. 5B**, **Fig. S8A-B**), as expected; however, ChromA and NoBed priors performed similarly, suggesting that ATAC-seq datasets are only useful for network inference when considering differential accessibility.

Inference: We used the scRNA-seq datasets from P30, P40 and P50 generated by our lab (*1*), as well as from P24, P36 and P48 generated by another group (*42*) to infer GRN models. In order to determine the sets of differentially expressed genes (DEGs) to model on, we used exclusively the “DGRP” portion of the P24-P36-P48 datasets as these are not expected to have batch effects between stages. The objects were subset for either the Tm or Lamina clusters, and then the stages were merged to keep each individual cluster from each stage as a separate class (thus, 12 classes for Tm, 15 for Lamina). DE genes (adj. p-value<0.01) were then

calculated between these classes using the FindAllMarkers function (only.pos = TRUE). The resulting sets of DE genes therefore represent those whose expression vary either across different cell types and/or different stages. Log-normalized data was used for inference; however, applying the same default scaling factor (10,000) to all stages can introduce artifacts if there are variations in the total RNA content of a given cell across different stages of development. We previously noted significant differences in the average UMI per cell values across different stages of our dataset (*1*); though it remains unclear how much of this could be attributed to technical reasons. When we normalized the cells of each stage with a value proportional to the average UMI value of that stage, we received overall very similar results compared to when we normalized with the default value. However, the benchmarks we performed with the *Hr3* RNAi dataset (described below) were significantly better (results not shown) with the stage-adjusted normalization; we thus proceeded with this method. This is consistent with the finding that *Hr3* is strongly regulated in the temporal axis during this period (*39*).

The networks were inferred using the Inferelator 3.0 (*40*), GitHub version v0.5.6 dd532f4, branch: release. Networks were learned using the multitask framework "AMuSR" (*63*) with the following parameters: regression="amusr", workflow="multitask", bootstraps=5, and cross validation seed set to 42 . Datasets were organized into two tasks representing data collected at P30-40-50 ("Desplan") and P24-36-48 ("Zipursky"). For each task, TFA was calculated by computing the dot product between the task-specific expression dataset and the pseudo-inverse of the prior. As part of the inference step, a linear model for each task is constructed that assumes each gene in the expression dataset can be modeled as a linear combination of the TFs regulating it ($\mathbf{X} = \beta\mathbf{A}$). Task specific betas represent the variance explained by each regulatory TF and are solved in the following way: We calculate the ratio between the variance of the residuals in the full model and the variance of the residuals when the model is refit, leaving one TF out at a time, as originally explained in the Inferelator (*40*). Any non-zero beta is considered evidence for a regulatory relationship between a gene and a TF. In the multi-task learning approach, after TFA is computed for each task independently, we solve the linear model for each task together resulting in an ensemble network, as well as task specific networks. Here, the model implements adaptive penalties to prefer regulatory interactions shared across the two tasks, placing a heavier penalty on interactions found to be only task-specific. Three networks were learned using the respective priors described above for both groups of cell types. In order to obtain results without sampling error, model selection was

repeated by sampling the expression data across 5 bootstraps, resampled with replacement. Tm and lamina networks were modeled on both DE genes and DE TFs, as described above. The respective priors, subset by DE genes and TFs, were used as gold standard networks for evaluating network performance. Evaluation metrics included calculating Area Under the Precision Recall (AUPR) curve to compare our predicted interactions to the gold standard network (prior). Further, we compute Matthews correlation coefficient (MCC) to obtain confusion matrices representing true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN). Calculating this score provided us with a metric to evaluate the reliability of our predicted network, as MCC obtains a high score only if the predicted interactions obtain good results for all four confusion matrices. An F1 score was additionally computed by taking the weighted average of the precision and recall. This score allows us to take into account both FP and FN, which is important to consider in the case of imbalanced classes, such as our model's assumption that a target gene can be regulated by a limited number of TFs. Finally, we compute the variance explained by each edge contained in the resulting network, allowing us to interpret which edges explain the most variance in the posterior network. Each network was benchmarked against a negative control where a network was inferred on each respective prior with the gene names shuffled. The final (multitask, MergedDA prior) Tm and Lamina networks are provided in **Table S3**. All network inference run scripts for the Inferelator are available at https://github.com/cskokgibbs/DMOLN_NetworkScripts.

The Inferelator does not report the sign (activating or repressing) of inferred interactions. Thus, for visualization purposes, we calculated Pearson correlations of expression levels using the stage-adjusted, log-normalized data with all the stages merged for the respective cell types. The network visualizations were generated with the Python package jp_gene_viz version 2.30.1 on branch 'Master' using the filters described in figure legends.

Benchmarking the networks with TF perturbations: As the Inferelator attempts to model all biologically relevant (between cell types and stages) differential gene expression among the input cells, this provides an opportunity to test if the learned models are generalizable to novel regulatory states never seen by the algorithm. We took advantage of 3 independent RNA-seq datasets acquired in the background of specific TF perturbations (described below) for this purpose. For each dataset, TFA was calculated for each cell using the respective priors (Lamina for *Hr3* and *Erm*, Tm for *pdm3* datasets) and the log-normalized expression values. We incorporated the betas collected from the relevant networks outlined above, describing the relative contribution of TFs to the model of each gene in our sample, to provide this benchmark

with unseen, independent data. Taking the dot product of the estimated TFA and independent betas for each experiment provided us with reconstructed (predicted) expression matrices. The command line arguments for these steps are provided in the same Github repository referenced above. Note that since the models were built on DEGs only, the predicted expression matrices also contain only these genes (1394 genes for Tm, 1684 for lamina datasets). We evaluated the results produced by all three prior generation methods described above for these benchmarks (**Fig. S8H-J**) but the UMAPs shown are only for the reconstructions generated with the "MergedDA" priors (**Fig. 5D-F**, **Fig. S8E-G**).

Hr3 RNAi: The Seurat object provided by the authors (*39*) was converted to match the gene names to our reference annotation, and subset to retain only the P48 stage. The predicted expression matrices were generated using the Lamina priors to calculate TFA, and the betas generated specifically from the "Zipursky" task described above, since these datasets contain the matching time point (P48) and were produced by the same lab using the same sequencing technology (10x v3.1). A merged Seurat object was then created with both the real (including only the DEGs present in the predicted expression) and predicted expression matrices, identification of top 1000 variable features, scaling and PCA were performed with the default parameters. UMAP visualization (**Fig. S8E**) was created using 30 principal components. The object was then split based on real vs. predicted expression, and integration was performed using the FindIntegrationAnchors (reduction="cca", anchor.features=1000, dims=1:30) and the IntegrateData (dims=1:30) functions. The data were scaled again and UMAP was generated using the same parameters as above (**Fig. 5D**). Then, separately for each of the 5 cell types contained in this dataset (L1-5), we determined the DEGs (adj. p-value<0.01) between control (w RNAi) and Hr3 RNAi conditions using both the real and predicted clusters (non-integrated data). For each cell type, precision is defined as the ratio of correctly determined DEGs between the conditions in predicted clusters (i.e., those that were also DEGs between the real clusters) (TP/(TP+FP)), and the recall is defined as the ratio of these correctly predicted DEGs to all DEGs between the real clusters (TP/(TP+FN)) (**Fig. S8H**).

Erm mutant: We obtained the Transcripts Per Million (TPM) results from a bulk RNA sequencing experiment of FACSed L3 neurons mutant for Erm (dFezf1) at stage P40 (*21*) and converted the gene names to match our reference annotation. We then simulated single-cell transcriptomes, separately for each of the 5 replicates of control (FRT40) and mutant experiments (**Fig. S8D**). The expression values were re-normalized to a total of 1 by dividing each value with the sum of all values for that sample. A probability distribution table was then

determined based on cumulative sums: starting from 0 and increasing gradually to 1 for the last (8905^{th}) gene in the dataset, with the amount of increase for each gene from the previous one being proportional to its level of expression. Thus, for all expressed genes, their value is higher than the previous gene in the list, while the value remains constant (from the previous one) if the gene is not expressed. For each simulated single-cell, the number of UMIs it will contain was randomly chosen from a normal distribution with mean=2500, s.d.=2000 and a minimum of 1000 UMIs. Each UMI was then assigned to a gene by picking a random number between 0 and 1, and determining the first occurrence of a number smaller than that in the probability distribution. Lastly, 100 simulated cells with at least 500 genes expressed were selected per replicate. Overall, the process aims to recapitulate the noisy sampling of mRNAs with high drop-out rates as characteristic of scRNA-seq experiments. Log-normalization, variable feature selection and scaling was applied using Seurat and the UMAP visualization (**Fig. S8D**) was created using the top 3 principal components. As this data contains only a single cell-type (with and without perturbation of only one TF), we observed that 3 PCs were enough capture all biological variation, as the inclusion of more PCs only resulted in increasing distances between the replicate clusters of the same condition.

The predicted expression matrices were then generated using the simulated single-cells, as described above for Hr3, but the betas were used from the "Desplan" task in this case since this task contains cells sequenced at the same stage (P40) as the Erm dataset. Integration between the real and predicted cells was performed using 200 anchors and 10 reduced dimensions (otherwise same as above) and the UMAPs were again calculated using the top 3 PCs (**Fig. 5E**, **Fig. S8F**). As there is only one cell type in this case, a single precision and recall value was determined for each tested prior (**Fig. S8I**). Interestingly, when we used the non-specific "NoBed" prior instead of the MergedDA prior, the predicted clusters in both Hr3 and Erm reconstructions tended to display many more DEGs between the conditions (higher recall), but with lower precision (**Fig. S8H-I**).

UAS-Pdm3: As both benchmarks above were on lamina neurons, we also aimed to similarly benchmark the Tm networks using the scRNA-seq dataset generated in this paper (**Fig. 1J**). TFAs calculated using the Tm prior, as well as the betas from the wild-type Tm networks (Zipursky task) were used to generate the predicted expression matrices. Normalization, scaling, dimensionality reduction and integration was performed using the same parameters as described above for the Hr3 dataset (**Fig. 5F**, **Fig. S8G**). Precision and recall metrics were

calculated for the DEGs between indicated subclusters in **Figure S8J** (referring to the same cluster numbers as in **Fig. 1K**).

Relatively higher metrics for these reconstructions are likely a result of changes in this dataset representing complete transformations of one cell type into another, unlike the lamina datasets where “unnatural” regulatory states were created. The one exception to this was Tm1 subcluster 6, where DEG analysis against the control Tm1 cells revealed extremely low recall (**Fig. S8J**). As we have discussed in the first section, this cluster consists of cells that express ectopic *pdm3* (**Fig. 1N**), but still retain Tm1 fate and appear morphologically normal (**Fig. S3F**). It is therefore likely that the changes observed in these cells do not represent a significant shift in their regulatory state, potentially explaining the low performance of the model in reconstructing them.

GO Enrichment: We filtered the inferred GRN model of Tm neurons to only include the interactions with combined confidence greater than 80% and variance explained greater than 1% (the same filtering as in Fig. 5C). Within this filtered network, we determined the predicted targets of Hr3 (167 genes), Pdm3 (120 genes), Mef2 (112 genes) and subjected them (separately) to GO enrichment analysis for ‘Molecular Function’ terms using The Gene Ontology Resource (http://geneontology.org/). We filtered the terms to include only those with fold enrichment greater than 2. We used REVIGO (http://revigo.irb.hr/) to remove redundant terms and group the related ones (*64*) with a similarity index of 0.5. The TreeMap tables were then exported from REVIGO and inputted to the Python package CirGO (*65*) to create the summary graphs in **Figure S9E**.

Figure S1

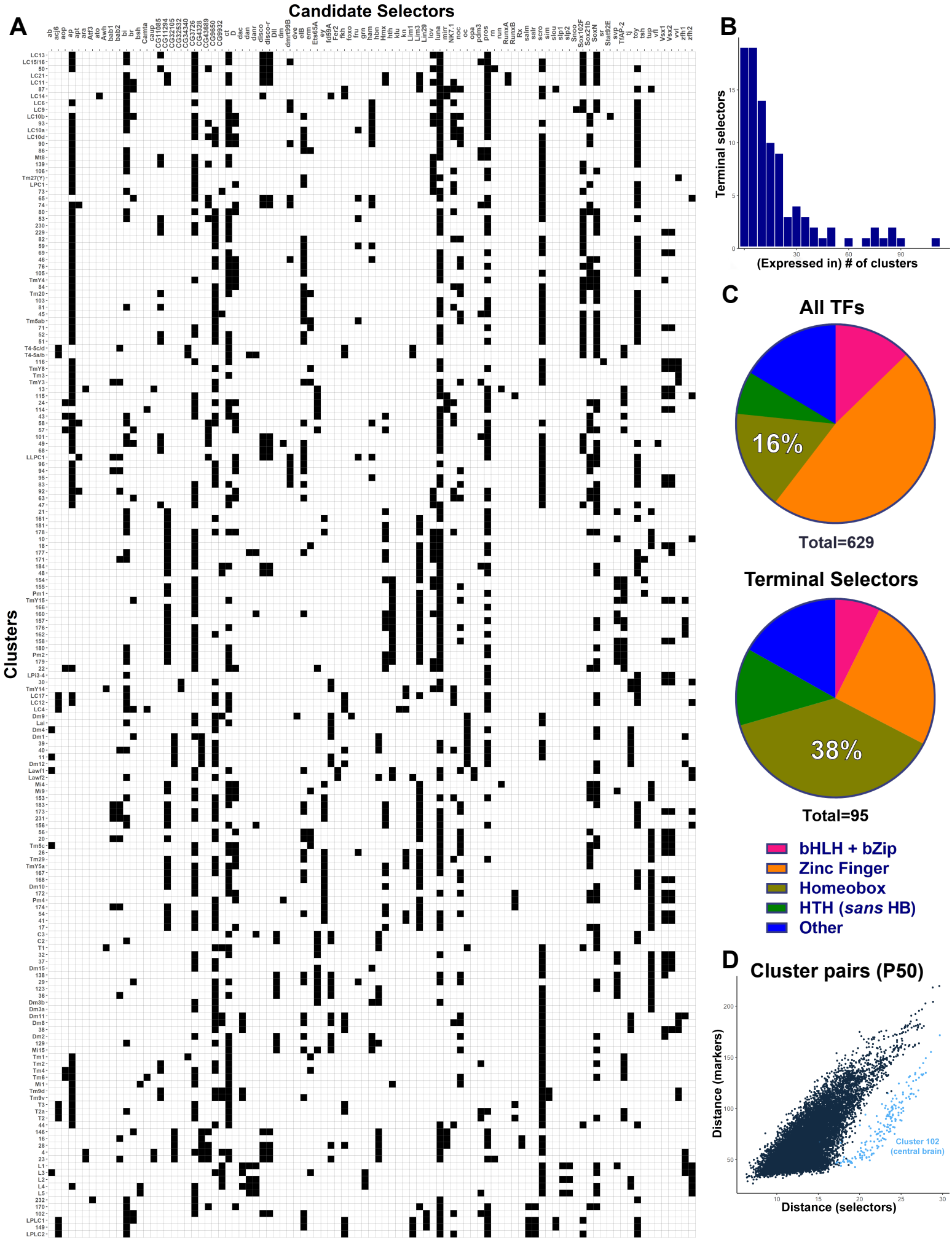

A
Candidate Selectors
Clusters
B
Terminal selectors
(Expressed in) # of clusters
C
All TFs
16%
Total=629
Terminal Selectors
38%
Total=95
bHLH + bZip
Zinc Finger
Homeobox
HTH (sans HB)
Other
D
Cluster pairs (P50)
Distance (markers)
Distance (selectors)
Cluster 102
(central brain)

Figure S1: Candidate terminal selectors of optic lobe neurons

A, Differentially expressed TFs (columns) found to be continuously expressed in each optic lobe neuronal cluster (rows) throughout its development according to (*1*), see Methods and Table S1. **B,** Histogram depicting the distribution of candidate selectors in (A) according to the number of clusters they were found to be continuously expressed (x-axis bin size: 5). **C,** Pie charts displaying the proportions of major TF classes within all TFs in the *Drosophila* genome (top), and the TFs found to be a candidate selector for at least one cell type in A. Basic domain, zinc finger, homeobox, helix-turn-helix (excluding homeobox) categories are represented, with the other TF classes merged in “Other”. **D,** Euclidian distances between all possible pairs (see Methods) of the 175 neuronal clusters shown in (A), calculated using the scaled expression values of 95 terminal selectors (x-axis) or 2467 cluster markers (y-axis). Note that the pairs involving cluster 102 (light blue) form a distinct regime because this cluster does not represent a genuine cell type but is rather a heterogenous mixture of contaminating central brain neurons.

Figure S2

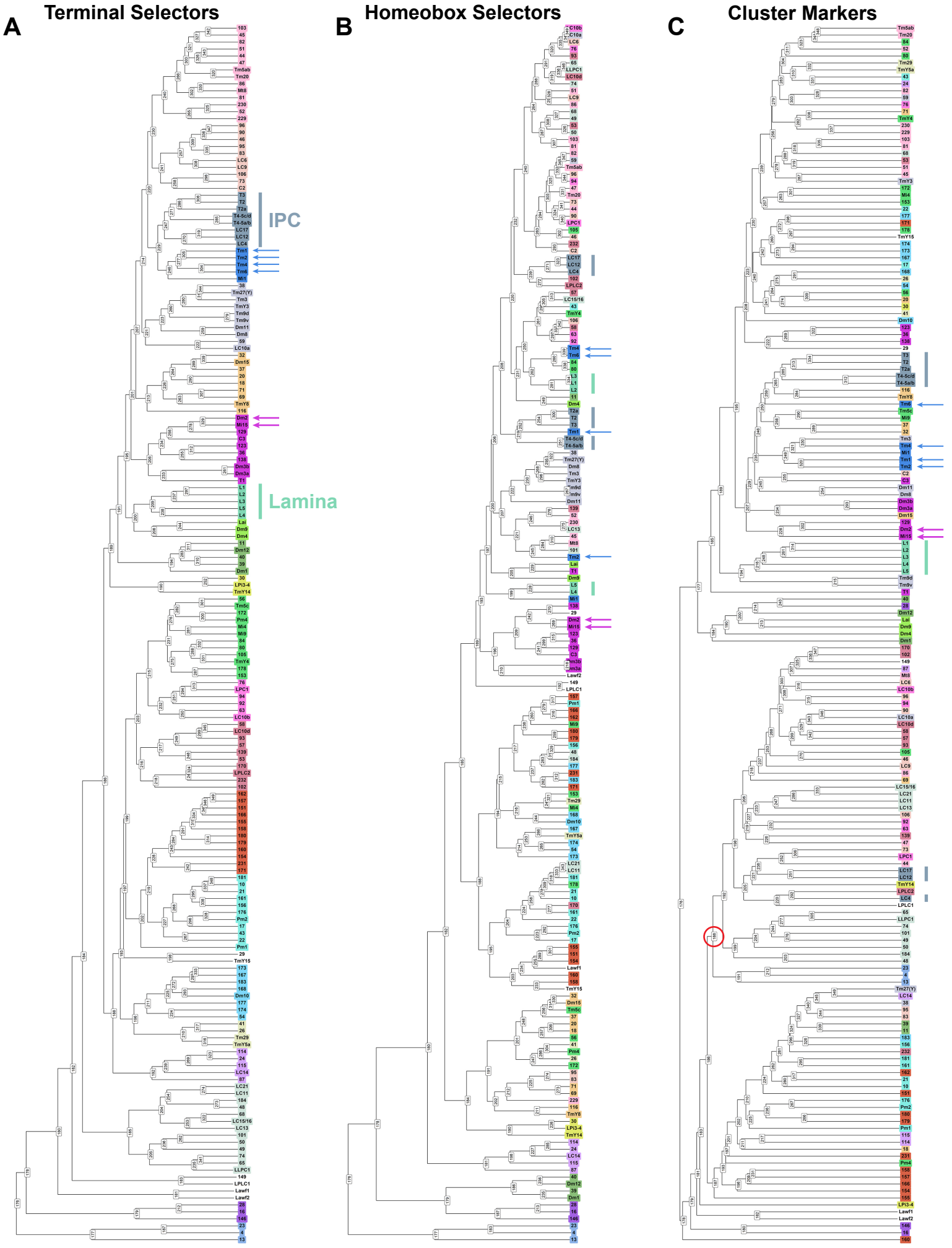
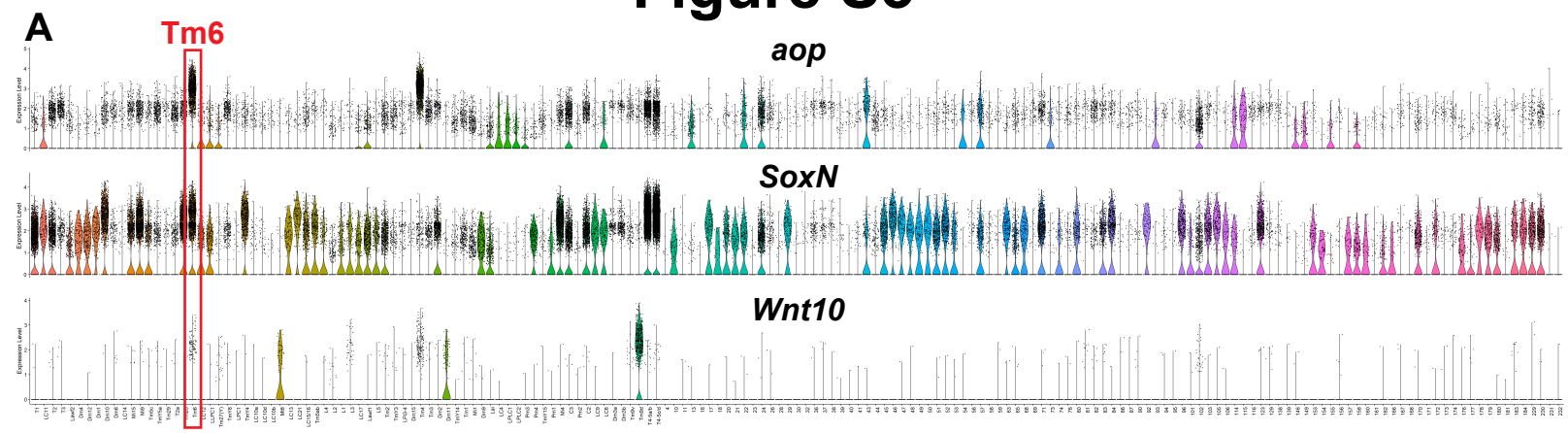
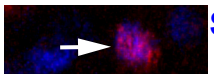

A
Terminal Selectors
IPC
Lamina
B
Homeobox Selectors
C
Cluster Markers

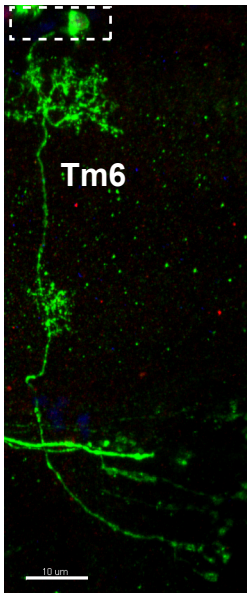
Figure S2: Hierarchical cluster trees of the optic lobe neurons

Dendrograms of neuronal clusters in the P50 scRNA-seq dataset (*1*), based on the scaled expression values of 95 candidate terminal selectors (**A**), 36 homeobox terminal selectors (**B**), or 2467 clusters markers (**C**) calculated at this stage. Clusters were colored in all panels according to their branch membership in (A), except for 6 outlier clusters that were not assigned to any color. IPC: neurons deriving from the inner proliferation center. Blue arrows: Tm1/2/4/6 neurons, purple arrows: Mi15 and Dm2 neurons. Red circle in (C) marks a branch in which all annotated neurons are known to send axonal projections to the central brain.

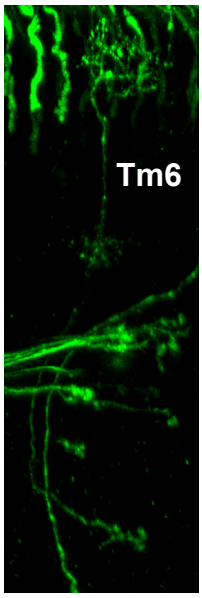
Figure S3



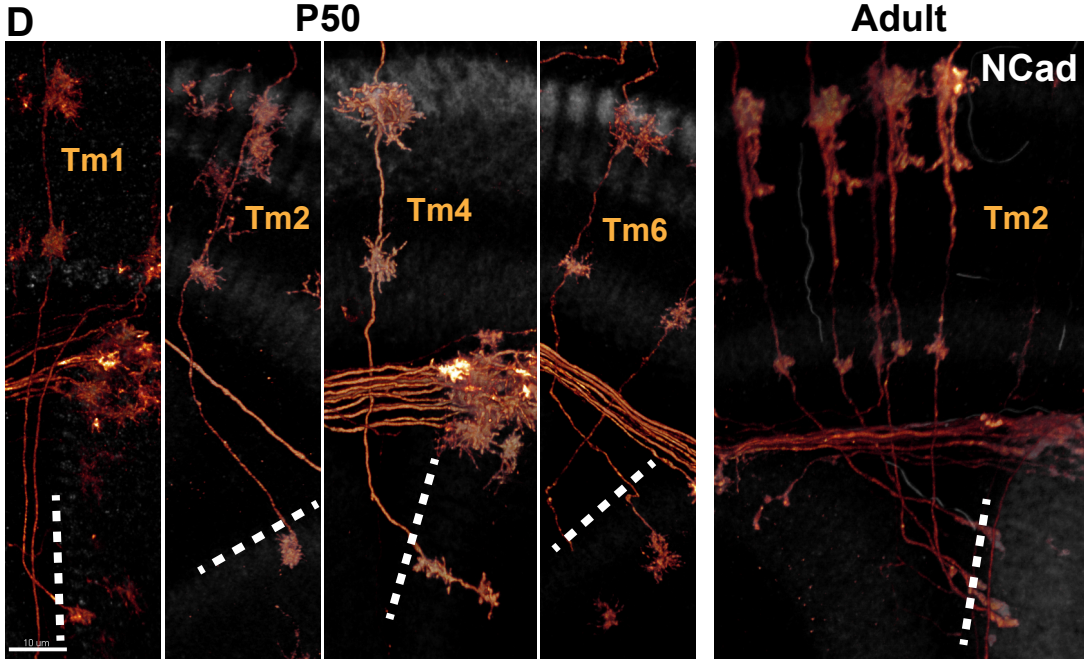




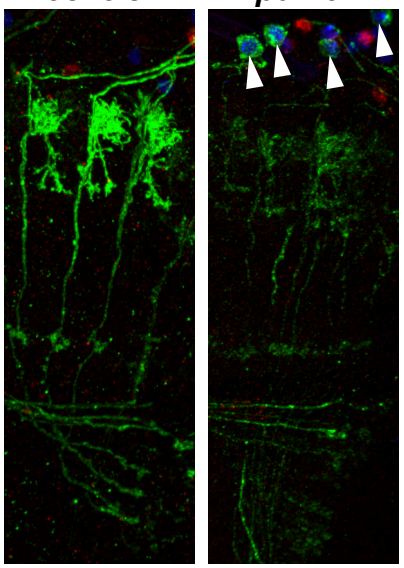
TmX/Tm4,6-Gal4



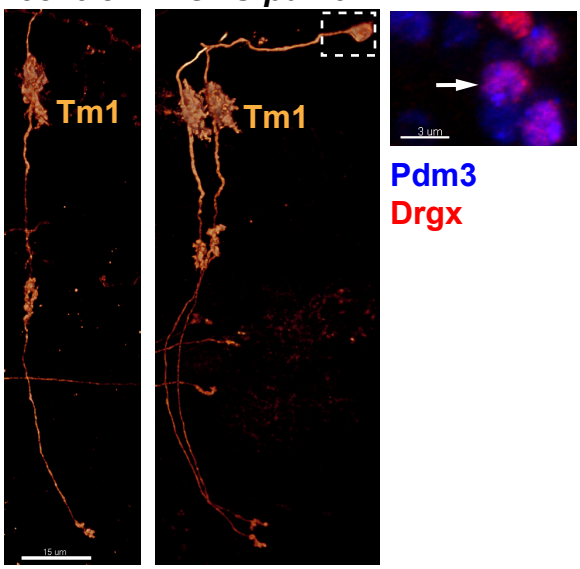
Wnt10-Gal4



TmX/Tm2-Gal4



TmX/Tm2 Pdm3 Aop



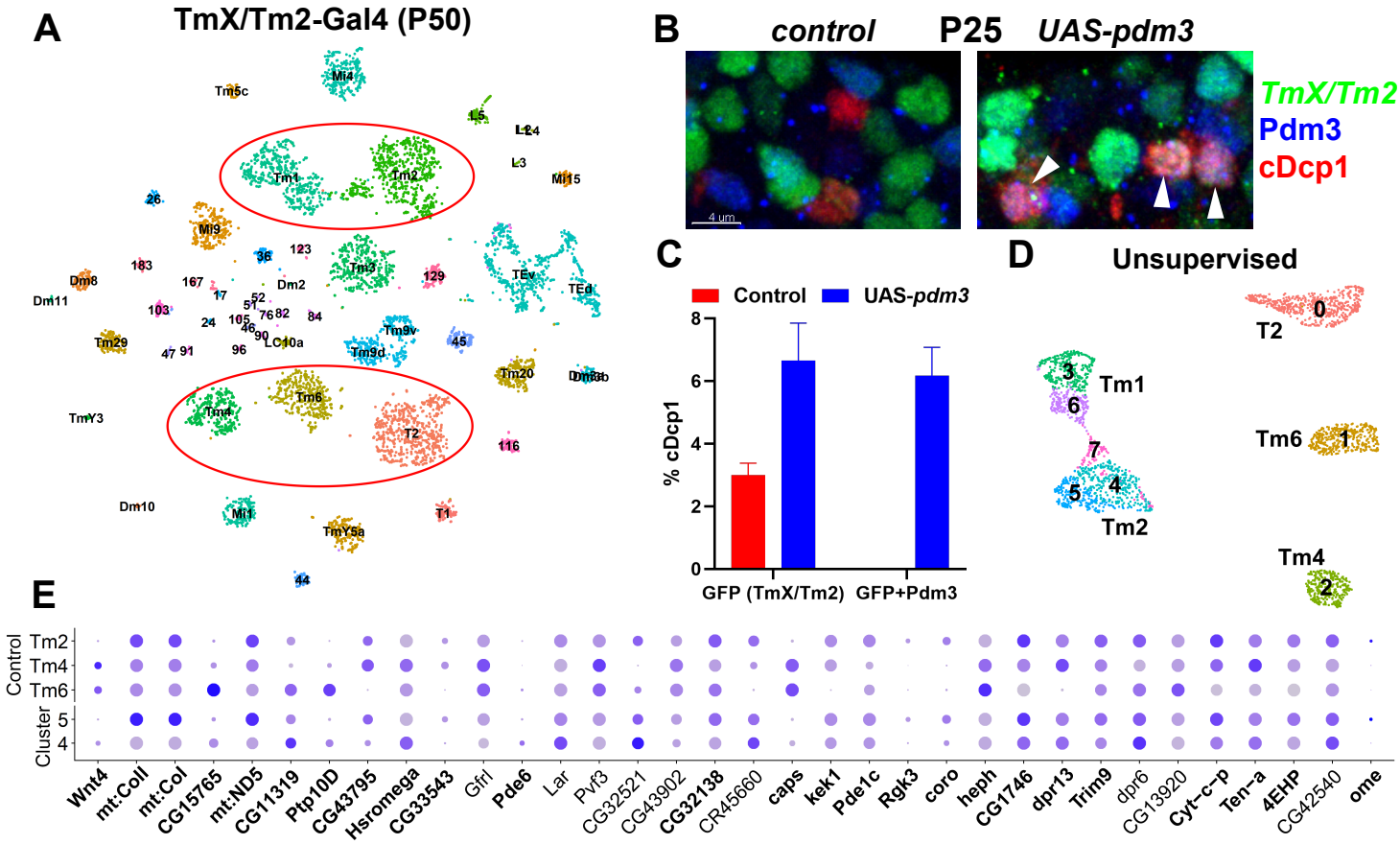
Pdm3
Drgx

Tm1,4-Gal4

Figure S3: Tm6 annotation and *pdm3* as the selector of Tm2 neurons (Supplement to Fig. 1)

A, Violin plots displaying the log-normalized expression of *aop*, *SoxN* and *Wnt10* in all adult optic lobe neurons (*1*). Cluster #62 (red rectangle) exclusively expresses all 3 genes. **B,** Max. projections of *TmX/Tm4,6-Gal4* (left) or *Wnt10-Gal4* (right) driving CD4-tdGFP (flip-out), displaying an adult Tm6 neuron (see Fig. 1A) with anti-SoxN (blue) and anti-Aop (red) for the inset in B. Arrow points to the nucleus of the neuron displayed below. **D,** 3D reconstructions of *TmX/Tm2-Gal4* driving CD4-tdGFP (flip-out) at P50 and adult optic lobes with anti-NCad (white), displaying representative Tm neurons. Dashed lines mark the border of lobula neuropil. **E,** FRT40A (control) and $pdm3^{1}$ MARCM clones labeled with *TmX/Tm2-Gal4* and CD4-tdGFP in adult optic lobes (maximum projection), with anti-Pdm3 (blue) and anti-Aop (red). Only Tm neurons observed in the mutants still expressed Pdm3 (arrowheads) and were labeled very weakly with GFP, suggesting this is leaky expression from MARCM (incomplete Gal80 suppression) rather than mutant clones (n=10 brains). **F,** 3D reconstructions of *Tm1,4-Gal4* driving CD4-tdGFP (flip-out) and UAS-pdm3.short (n=6 brains), displaying representative adult Tm1 neurons. Inset shows a max. projection of the indicated soma (arrow) with anti-Pdm3 (blue) and anti-Drgx (red). Scale bars: 10 μm (B-E), 15 μm (F) and 3 μm (F, inset).

Figure S4


A
TmX/Tm2-Gal4 (P50)
Mi4
Tm5c
L5
L24
L3
Mi15
Tm1
Tm2
26
Mi9
36
123
183
Dm8
167
17
Dm2
Tm3
129
TEv
TEd
Dm11
103
24
105
46
51
52
76
82
84
90
Tm29
47
91
96
LC10a
Tm9v
Tm9d
45
Tm20
Dm3ab
TmY3
Tm4
Tm6
T2
116
Dm10
Mi1
T1
TmY5a
44
B
control
P25
UAS-pdm3
TmX/Tm2
Pdm3
cDcp1
4 um
C
Control
UAS-pdm3
% cDcp1
8
6
4
2
0
GFP (TmX/Tm2)
GFP+Pdm3
D
Unsupervised
0
T2
3
Tm1
6
7
5
4
Tm2
Tm6
1
Tm4
2

E
Control
Tm2
Tm4
Tm6
Cluster
5
4
Wnt4
mt:Coll
mt:Col
CG15765
mt:ND5
CG11319
Ptp10D
CG43795
Hsromega
CG33543
Gfrl
Pde6
Lar
Pvf3
CG32521
CG43902
CG32138
CR45660
caps
kek1
Pde1c
Rgk3
coro
heph
CG1746
dpr13
Trim9
dpr6
CG13920
Cyt-c-p
Ten-a
4EHP
CG42540
ome

Figure S4: *pdm3* instructs transcriptionally complete neuronal fate conversions (Supplement to Fig. 1)

A, scRNA-seq of FACSed neurons from P50 optic lobes, *TmX/Tm2-Gal4* driving nuclear GFP (Stinger) and UAS-pdm3.short. tSNE was calculated using 50 principal components. Cells are colored and labeled according to neural network (*1*) classifications. Red ellipses mark the cells subset for further analysis (Fig. 1H-O). **B,** Max. projections of P25 medulla cortex, with *TmX/Tm2-Gal4* driving Stinger (green) and UAS-pdm3.short (right), with anti-Pdm3 (blue) and anti-cleaved Dcp1 (red) marking apoptotic cells. Arrowheads indicate the nuclei with triple labeling. Scale bar: 4 μm. **C,** Quantification of (B), displaying the percentage of $cDcp1^+$ neurons in each indicated group. n= 2313 (control, GFP), 561 (control, GFP+Pdm3), 2640 (UAS-pdm3, GFP), 2157 (UAS-pdm3, GFP+Pdm3) neurons in 5 brains per condition. p=0.006 (GFP) and p<0.0001 (GFP+Pdm3) for change in the proportion of $cDcp1^+$ neurons. Error bars denote SEM. **D,** Same as in Fig. 1J, with cells colored and labeled according to unsupervised clustering (as in 1K). **E**, Dot plot displaying the log-normalized expression of the 34 differentially expressed genes between clusters 4 and 5 in (D), in cells classified as Tm2/4/6 in the control library only (top) and in clusters 4 and 5 (bottom). Diameters of the dots indicate the proportion of cells in each cluster expressing the genes, while color intensity is proportional to the level of expression. Markers highlighted in bold are consistent with cluster 4 cells having been converted from the Tm4/6 fate.

Figure S5

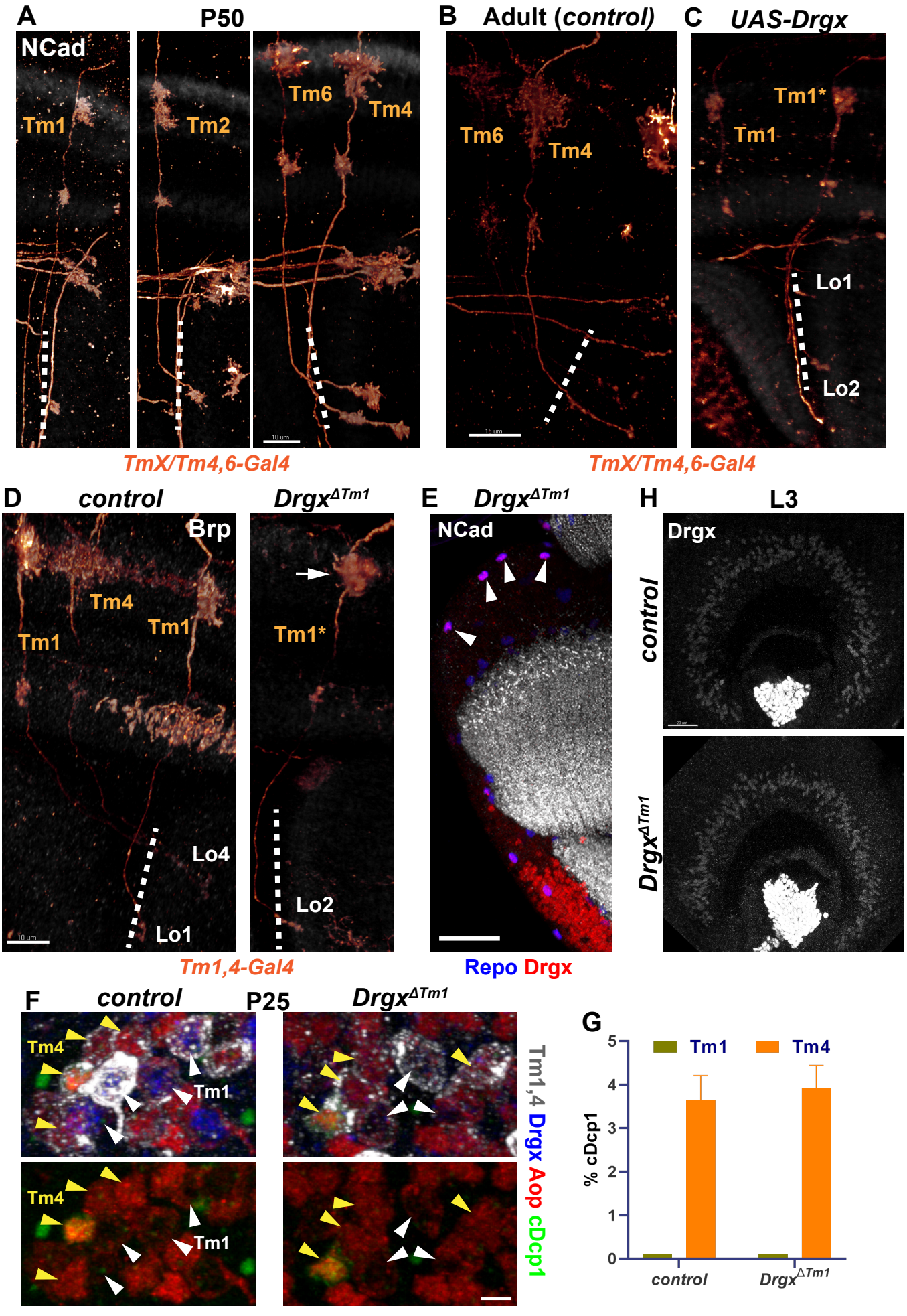

A
P50
NCad
Tm1
Tm2
Tm6
Tm4
B
Adult (control)
Tm6
Tm4
C
UAS-Drgx
Tm1*
Tm1
Lo1
Lo2
TmX/Tm4,6-Gal4
TmX/Tm4,6-Gal4
D
control
Drgx^ΔTm1
Brp
Tm4
Tm1
Lo4
Lo1
Tm1*
Lo2
Tm1,4-Gal4
E
Drgx^ΔTm1
NCad
Repo Drgx
H
L3
Drgx
control
Drgx^ΔTm1
F
control
P25
Drgx^ΔTm1
Tm4
Tm1
Tm4
Tm1
Tm1,4 Drgx Aop cDcp1
G
Tm1
Tm4
% cDcp1
0
1
2
3
4
5
control
Drgx^ΔTm1

Figure S5: Drgx as the terminal selector of Tm1 neurons (Supplement to Fig. 2)

A-C, 3D reconstructions of *TmX/Tm4,6-Gal4* driving CD4-tdGFP (flip-out) and UAS-Drgx (**C** only) with anti-NCad (white), displaying representative Tm neurons in P50 (**A**) and adult (**B-C**) optic lobes. The Tm1 labeled with an asterisk abnormally targets to Lo2 layer instead of Lo1. Quantified in Fig. 2B. **D,** 3D reconstructions of adult neurons labeled with *Tm1,4-Gal4* driving CD4-tdGFP (flip-out) in the background of heterozygous (control) or homozygous $Drgx^{\Delta Tm1}$ allele (same as in Fig. 2G), with anti-Brp (white). Quantified in Fig. 2H. The Tm1 labeled with an asterisk targets to Lo2 and has an abnormal dendritic arbor (arrow). **E,** Same as in Fig. 2G with anti-NCad (white), anti-Repo (blue) and anti-Drgx (red), showing that the only remaining $Drgx^{+}$ cells in the medulla cortex (arrowheads) of $Drgx^{\Delta Tm1}$ mutants are glia (n=3 brains). **F,** Max. projections of P25 neurons labeled with *Tm1,4-Gal4* driving CD4-tdGFP in control or $Drgx^{\Delta Tm1}$ brains, with anti-Drgx (blue), anti-Aop (red) and anti-cleaved Dcp1 (green). White arrowheads: Tm1, yellow arrowheads: Tm4 neurons based on the pattern of Aop staining. **G,** Quantification of (F). No Tm1 neurons were stained with cDcp1 in either condition. n= 902 (control, Tm1), 804 (control, Tm4), 620 ($Drgx^{\Delta Tm1}$, Tm1), 734 ($Drgx^{\Delta Tm1}$, Tm4) neurons in 6 brains per condition. p=0.72 (Tm4), error bars denote SEM. **H,** Max. projections of L3 optic lobes showing anti-Drgx staining in w^{1118} (control, n=5 brains) and homozygous $Drgx^{\Delta Tm1}$ mutants (n=7 brains). Scale bars: 10 μm (A, D-E), 15 μm (B-C), 3 μm (F) and 20 μm (H). Dashed lines mark the border of lobula neuropil.

Figure S6

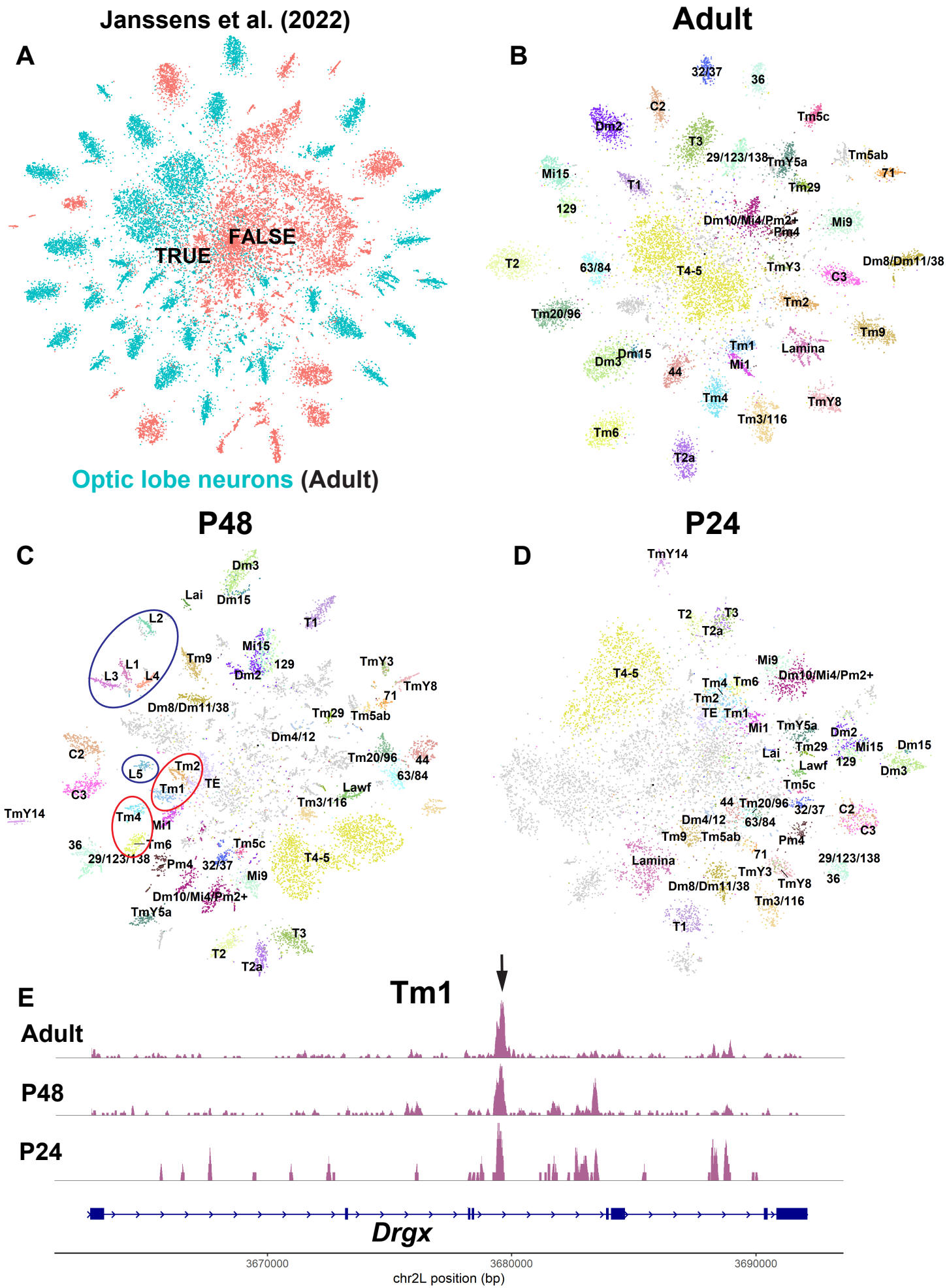

Janssens et al. (2022)
Adult
A
B
TRUE
FALSE
Optic lobe neurons (Adult)
P48
P24
C
D
E
Tm1
Adult
P48
P24
Drgx
3670000
3680000
3690000
chr2L position (bp)

Figure S6: Annotation of optic lobe neurons in snATAC-seq datasets

A-D, Clusters corresponding to the optic lobe neurons in Adult (A), P48 and P24 snATAC-seq experiments (*28*) were isolated, clustered and annotated (**B-D**) separately at the indicated stages (see Methods for details). tSNE visualizations were calculated using the top 120 reduced dimensions (LSI), excluding the first dimension. Clusters labeled by numbers are unannotated neurons in the reference scRNA-seq atlas (*1*). Clusters that could not be linked to defined (groups of) clusters in the reference atlas were grayed out. Ellipses in (C) indicate the clusters used for building the Inferelator priors for lamina (blue) and Tm (red) neurons for GRN inference. **E,** Aggregated accessibility tracks of *Drgx* locus for the Tm1 cluster, using the TF-IDF normalized snATAC-seq data at the indicated stages (B-D). Arrow points to the Tm1-specific enhancer deleted in (Fig. 2G-I). See also Fig. 2D.

Figure S7

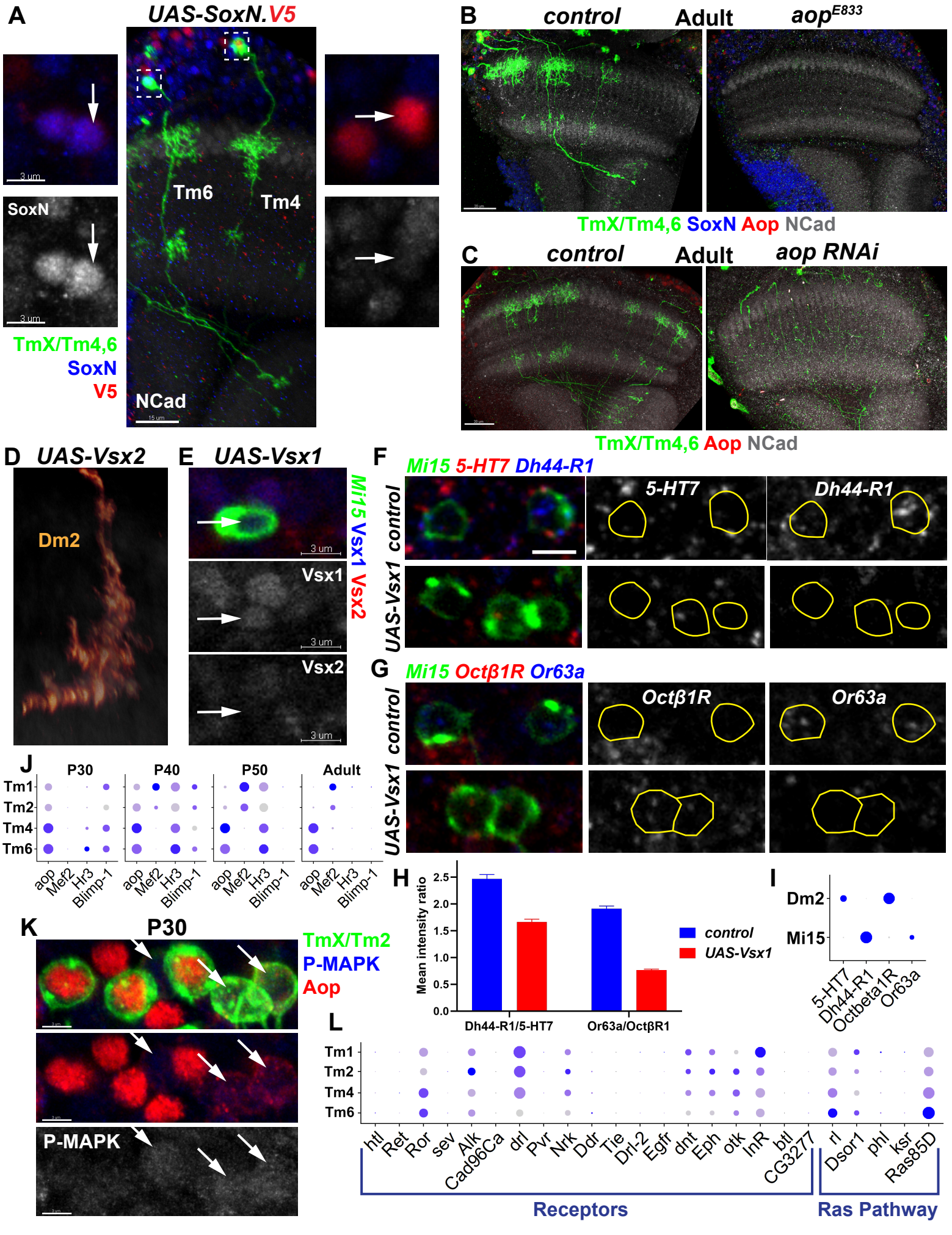
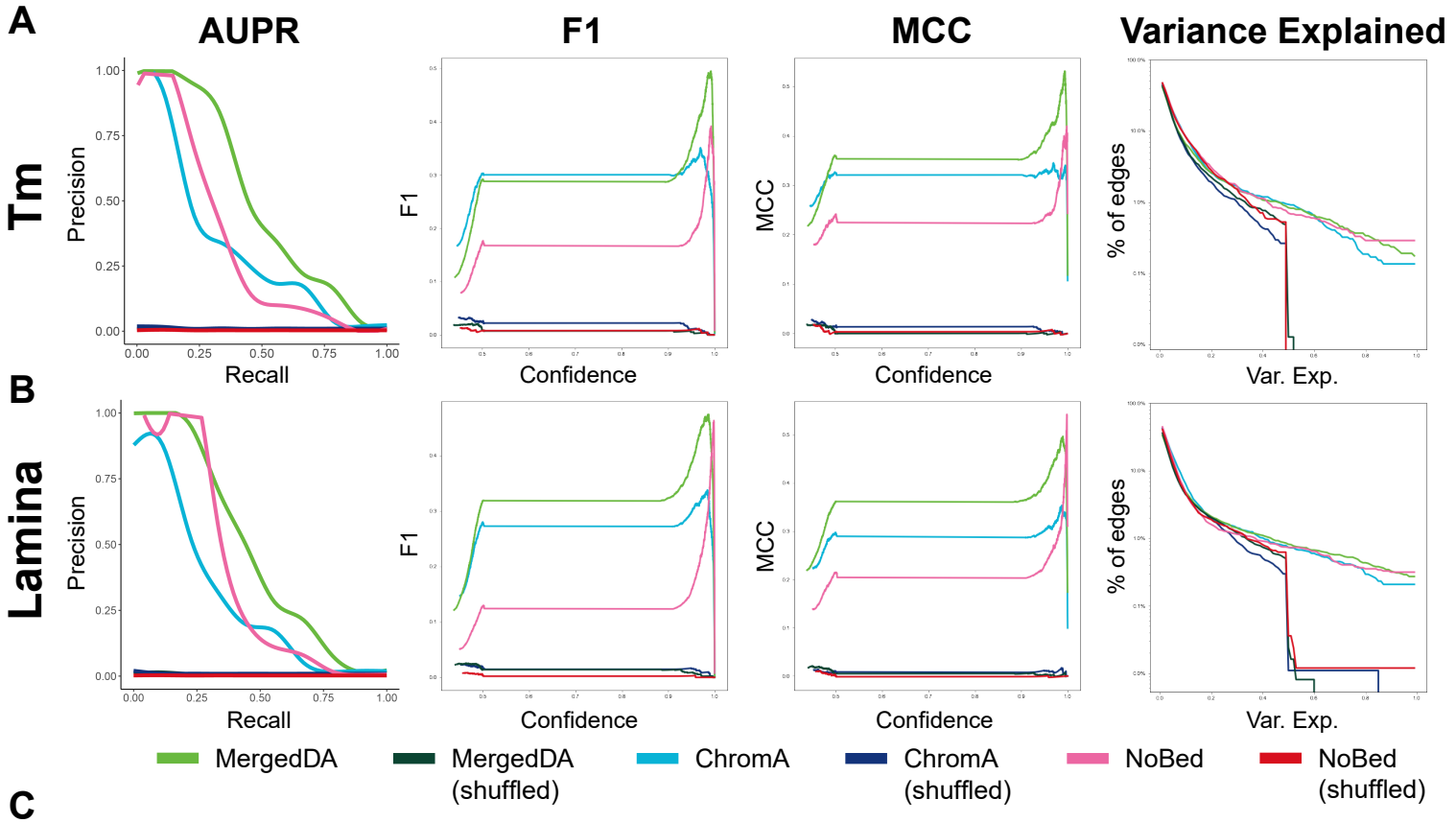
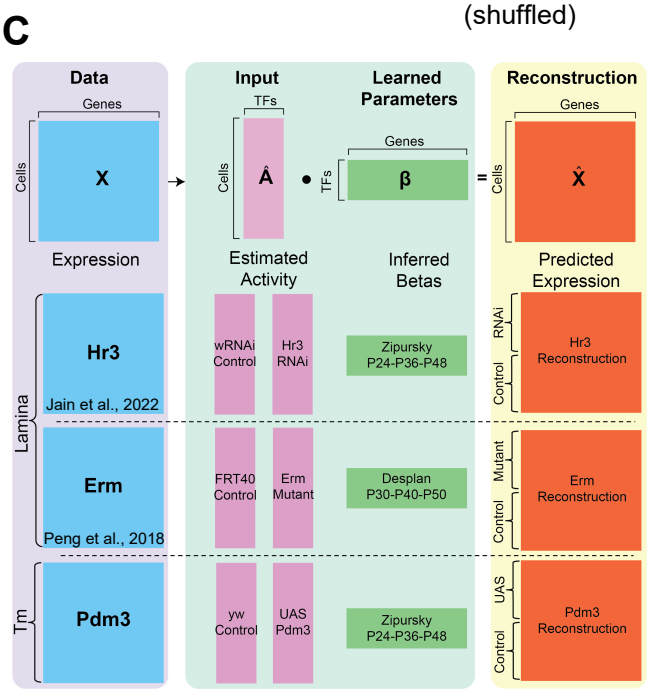

A
UAS-SoxN.V5
SoxN
TmX/Tm4,6
SoxN
V5
Tm6
Tm4
NCad
B
control
Adult
aopE833
TmX/Tm4,6 SoxN Aop NCad
C
control
Adult
aop RNAi
TmX/Tm4,6 Aop NCad
D UAS-Vsx2
Dm2
E UAS-Vsx1
Mi15 Vsx1 Vsx2
Vsx1
Vsx2
F
Mi15 5-HT7 Dh44-R1
control
UAS-Vsx1
5-HT7
Dh44-R1
G
Mi15 Octβ1R Or63a
control
UAS-Vsx1
Octβ1R
Or63a
J
P30
P40
P50
Adult
Tm1
Tm2
Tm4
Tm6
aop
Mef2
Hr3
Blimp-1
H
Mean intensity ratio
control
UAS-Vsx1
Dh44-R1/5-HT7
Or63a/OctβR1
I
Dm2
Mi15
5-HT7
Dh44-R1
Octbeta1R
Or63a
K
P30
TmX/Tm2
P-MAPK
Aop
P-MAPK
L
Tm1
Tm2
Tm4
Tm6
htl
Ret
Ror
sev
Alk
Cad96Ca
drl
Pvr
Nrk
Ddr
Tie
Drl-2
Egfr
dnt
Eph
otk
InR
btl
CG3277
rl
Dsor1
phl
ksr
Ras85D
Receptors
Ras Pathway

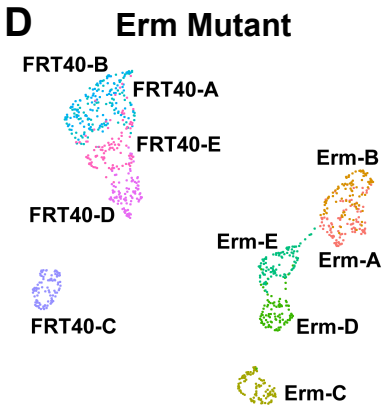
Figure S7: Additional terminal selectors of optic lobe neurons and regulation of the Tm selector network by RTK signaling (Supplement to Figs. 3 and 4)

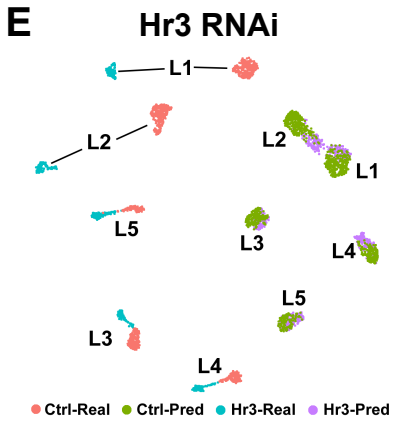
A, *TmX/Tm4,6-Gal4* driving UAS-SoxN.V5 and CD4-tdGFP (flip-out, n=3 brains). Max. projection of an adult optic lobe with anti-NCad (white), anti-SoxN (blue) and anti-V5 (red). Insets show the indicated somas (arrows) with anti-SoxN also shown in single channel (bottom). **B,** Max. projections of adult optic lobes in FRT40A (control) and aop^{E833} MARCM clones (n=3 brains) labeled with *TmX/Tm4,6-Gal4* and CD4tdGFP, with anti-NCad (white), anti-SoxN (blue) and anti-Aop (red). **C,** Max. projections of adult optic lobes with *TmX/Tm4,6-Gal4* driving CD4-tdGFP (flip-out) and *aop* RNAi (n=8 brains), with anti-NCad (white) and anti-Aop (red). **D,** Same as in Fig. 3F, with *Mi15-Gal4* driving UAS-Vsx2 instead of Vsx1. Quantified in Fig. 3E. **E,** Same as in Fig. 3F, max. projection of somas with anti-Vsx1 (blue) and anti-Vsx2 (red) that are also shown in single channel (n=3 brains). Note that Vsx1/2 are almost always co-expressed in the optic lobe neurons except for the Mi15 neurons in this experiment that express Vsx1 ectopically (arrow). **F-G,** *Mi15-Gal4* driving UAS-Vsx1 and CD4-tdGFP in adult brains, displaying single z-slices. Fluorescence *in situ* hybridization signal against (**F**) *5-HT7* (red), *Dh44-R1* (blue), and (**G**) *Octβ1R* (red), *Or63a* (blue). Also shown in single channel, yellow lines demarcate the somas labeled by GFP as shown in the left panels. **H,** Quantification of (F-G). The ratio of mean signal intensities for the indicated channels in Mi15 somas. N=90 (control) and 134 (Vsx1) neurons for *Dh44-R1*/*5-HT7*, $p<0.0001$. N=151 (control) and 180 (Vsx1) neurons for *Or63a*/*Octβ1R*, $p<0.0001$. Error bars denote SEM. **I,** Dot plot displaying the log-normalized expression of indicated genes in Dm2 and Mi15 clusters in the P96 scRNA-seq dataset (*42*). **J,L** Dot plots displaying the log-normalized expression in Tm1/2/4/6 clusters of indicated TFs in the P30, P40, P50 and adult scRNA-seq datasets (*1*) (**J**), or of all RTK receptors and Ras signaling pathway components expressed in the optic lobe at P30 (**L**). **K**, *TmX/Tm2-Gal4* driving CD4-tdGFP (flip-out) in P30 medulla cortex (max projection), with anti-P-MAPK (blue) and anti-Aop (red). Arrows indicate $GFP^{+}Aop^{-}$ Tm1/2 neurons in which PMAPK (single-channel, bottom) can be detected. Note that Aop^{+} neurons do not display PMAPK signal (n=5 brains). Scale bars: 15 μm (A), 20 μm (B-C) and 3 μm (A-insets, E-J).

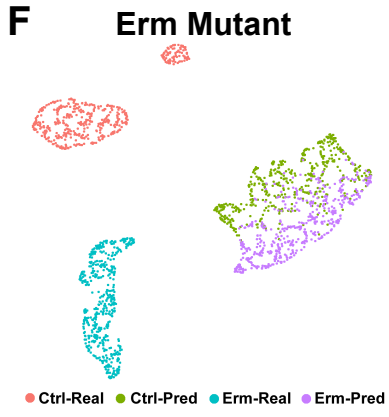
Figure S8

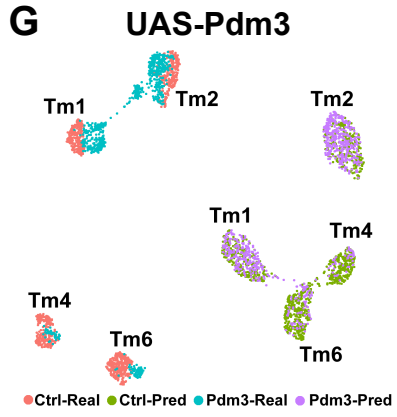

A
AUPR
F1
MCC
Variance Explained
Tm
Precision
Recall
F1
Confidence
MCC
Confidence
% of edges
Var. Exp.
B
Lamina
Precision
Recall
F1
Confidence
MCC
Confidence
% of edges
Var. Exp.
MergedDA
MergedDA (shuffled)
ChromA
ChromA (shuffled)
NoBed
NoBed (shuffled)

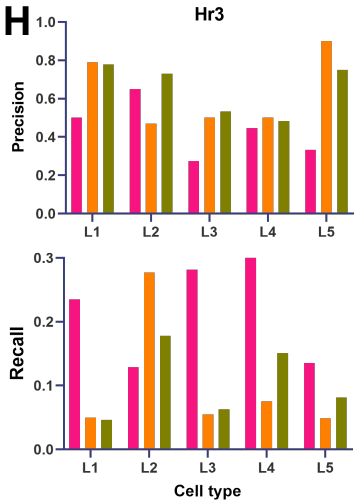

C
Data
Input
Learned Parameters
Reconstruction
Genes
Cells
X
Expression
TFs
Â
Estimated Activity
β
Inferred Betas
X̂
Predicted Expression
Lamina
Hr3
Jain et al., 2022
wRNAi Control
Hr3 RNAi
Zipursky P24-P36-P48
Hr3 Reconstruction
RNAi
Control
Erm
Peng et al., 2018
FRT40 Control
Erm Mutant
Desplan P30-P40-P50
Erm Reconstruction
Mutant
Control
Tm
Pdm3
yw Control
UAS Pdm3
Zipursky P24-P36-P48
Pdm3 Reconstruction
UAS
Control

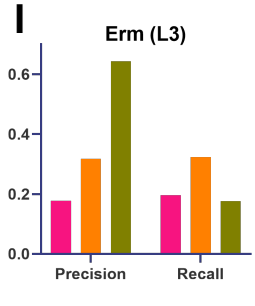

D
Erm Mutant
FRT40-B
FRT40-A
FRT40-E
FRT40-D
FRT40-C
Erm-B
Erm-E
Erm-A
Erm-D
Erm-C

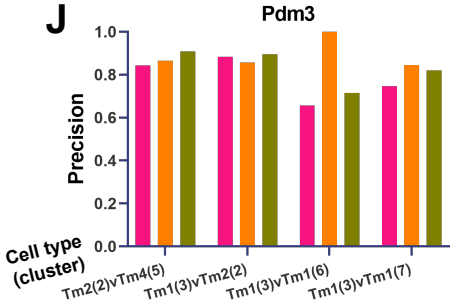

E
Hr3 RNAi
L1
L2
L2
L1
L5
L3
L4
L5
L3
L4
Ctrl-Real
Ctrl-Pred
Hr3-Real
Hr3-Pred


F
Erm Mutant
Ctrl-Real
Ctrl-Pred
Erm-Real
Erm-Pred


G
UAS-Pdm3
Tm1
Tm2
Tm2
Tm1
Tm4
Tm4
Tm6
Tm6
Ctrl-Real
Ctrl-Pred
Pdm3-Real
Pdm3-Pred


H
Hr3
Precision
Recall
L1
L2
L3
L4
L5
Cell type
NoBed
ChromA
MergedDA


I
Erm (L3)
Precision
Recall


J
Pdm3
Precision
Cell type (cluster)
Tm2(2)vTm4(5)
Tm1(3)vTm2(2)
Tm1(3)vTm1(6)
Tm1(3)vTm1(7)

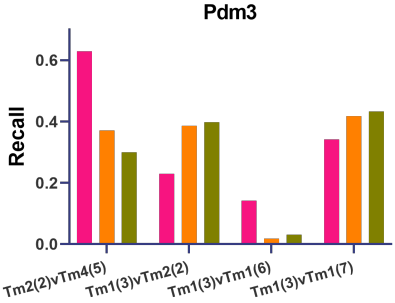
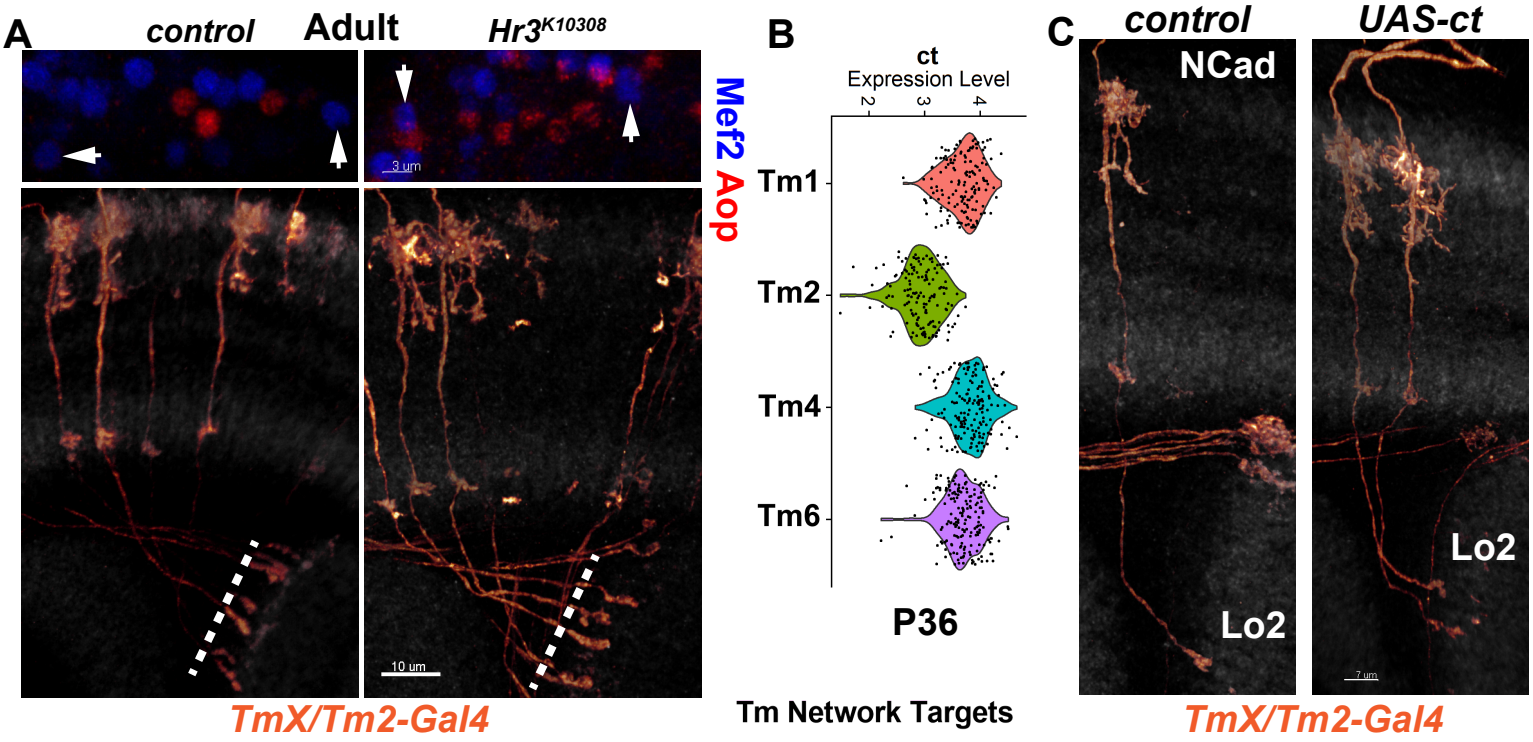
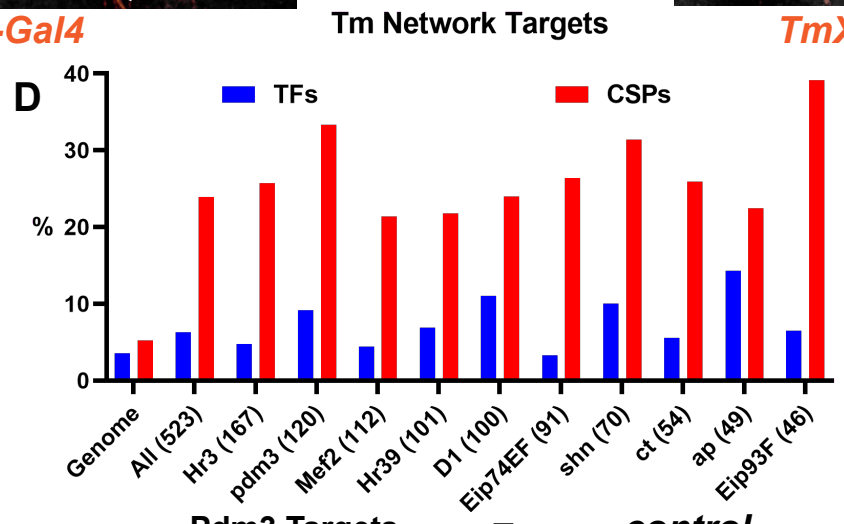

Pdm3
Recall
Tm2(2)vTm4(5)
Tm1(3)vTm2(2)
Tm1(3)vTm1(6)
Tm1(3)vTm1(7)

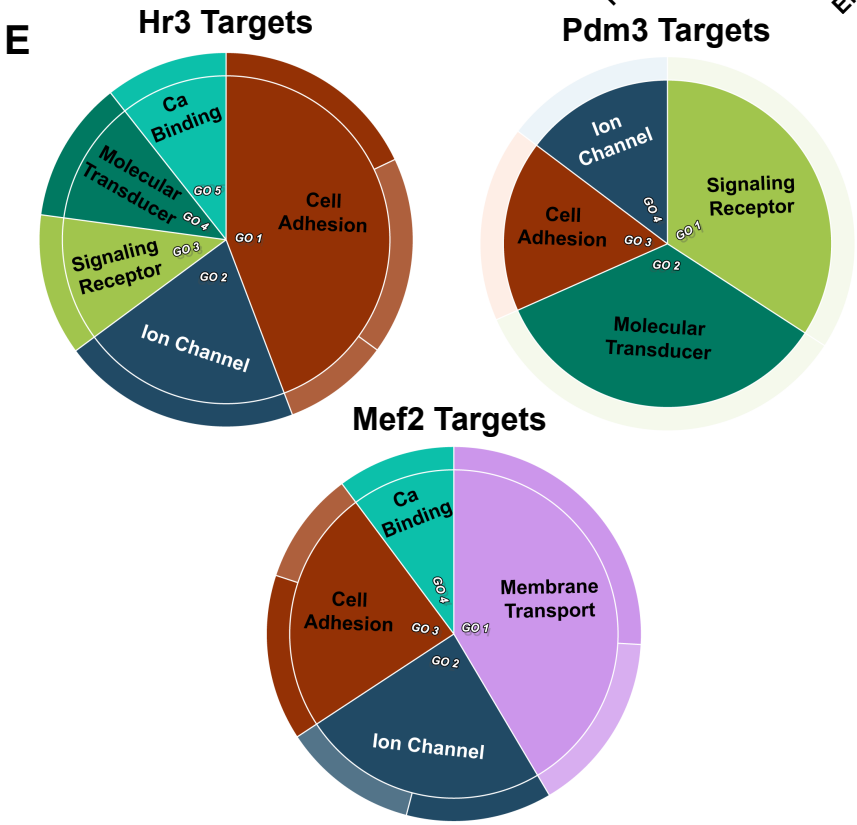
Figure S8: Computational inference and benchmarking of GRNs in optic lobe neurons (Supplement to Fig. 5)

A-B, Curves of Area under Precision and Recall (AUPR), F1 scores, Mathews Correlation Coefficient (MCC) and variance explained by the edges inferred (multitask) from (**A**) Tm1/2/4/6 and (**B**) L1-5 cell types using three different priors for each network, and their shuffled negative controls. See Methods for detailed description of the priors and displayed metrics. **C,** Reconstruction method for predicting gene expression profiles of the RNA-seq experiments shown. TFA is calculated for each cell as described in Fig. 5A using the same priors for the respective lamina (for Hr3 and Erm) and Tm (for Pdm3) network inference. Dot products of these TFA matrices with the learned betas from the indicated inference tasks, i.e., the weights between TFs and their targets as determined by the Inferelator using only the wild-type data, generate a predicted expression matrix matching the original data. **D,** Simulated single-cell transcriptomes (100 per replicate, see Methods) from the Erm bulk RNA-seq experiment at P40 (*21*). UMAP visualization was calculated using 3 principal components (PCs). **E-G,** scRNA-seq datasets of (**E**) L1-5 neurons at P48 expressing Hr3 RNAi (*39*), (**F**) simulated-single Erm mutant L3 neurons from (D), (**G**) Tm1/2/4/6 neurons at P50 overexpressing UAS-pdm3.short (Fig. 1). Both the original ("Real") and predicted ("Pred") (based on networks built with the MergedDA prior) transcriptomes are shown, and the cells are colored according to this status, as well as their condition of origin. UMAPs were calculated using 30 PCs (E,G) or 3 PCs (F) on the non-integrated gene expression. See Fig. 5D-F for UMAPs of the same data after Seurat integration was performed between the real and predicted clusters. **H-J,** Comparative analysis of real and predicted differentially expressed genes (DEGs) between the experimental conditions for cell types or clusters shown in E-G (see Fig. 1K for the cluster numbers in J), calculated separately using the networks generated with three different priors. Precision is defined as the ratio of correctly determined DEGs between the conditions in predicted clusters (i.e., those that were also DEGs between the real clusters), and the recall is defined as the ratio of these correctly predicted DEGs to all DEGs between the real clusters.

Figure S9


A
control
Adult
Hr3K10308
Mef2 Aop
3 um
10 um
TmX/Tm2-Gal4
B
ct
Expression Level
2
3
4
Tm1
Tm2
Tm4
Tm6
P36
C
control
UAS-ct
NCad
Lo2
Lo2
7 um
TmX/Tm2-Gal4


Tm Network Targets
D
TFs
CSPs
40
30
% 20
10
0
Genome
All (523)
Hr3 (167)
pdm3 (120)
Mef2 (112)
Hr39 (101)
D1 (100)
Eip74EF (91)
shn (70)
ct (54)
ap (49)
Eip93F (46)


E
Hr3 Targets
Ca Binding
Molecular Transducer
Signaling Receptor
Ion Channel
Cell Adhesion
GO 1
GO 2
GO 3
GO 4
GO 5
Pdm3 Targets
Ion Channel
Cell Adhesion
Signaling Receptor
Molecular Transducer
Mef2 Targets
Ca Binding
Cell Adhesion
Membrane Transport
Ion Channel

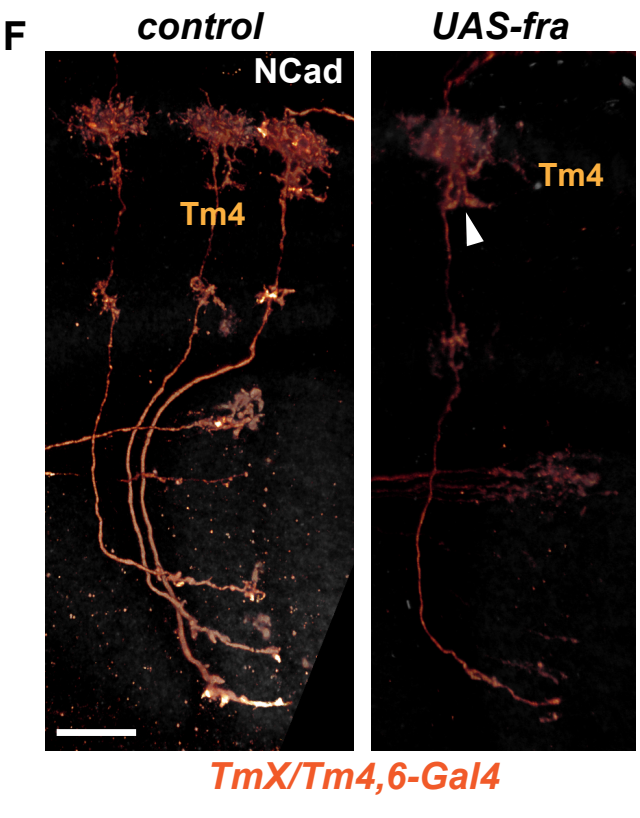

F
control
UAS-fra
NCad
Tm4
Tm4
TmX/Tm4,6-Gal4

Figure S9: Additional validation experiments for the GRN models of Tm neurons (Supplement to Fig. 5)

A, FRT42D (control) and *Hr3*^{K10308} MARCM clones labeled with *TmX/Tm2-Gal4* and CD4-tdGFP showing (bottom) 3D reconstructions of representative adult neurons with anti-NCad (white), and (top) the max. projections of somas with anti-Mef2 (blue) and anti-Aop (red). Arrows point to the nuclei of Tm2 neurons shown in bottom panels (n=5 brains). Dashed lines mark the border of lobula neuropil. **B,** Violin plot displaying the log-normalized expression of *ct* in Tm1/2/4/6 clusters in the P36 scRNA-seq dataset (*42*). **C,** *TmX/Tm2-Gal4* driving CD4-tdGFP (flip-out) and UAS-ct (n=6 brains). 3D reconstructions of GFP expression for the representative neurons in each condition with anti-NCad (white). **D,** Percentages of genes encoding TFs and cell-surface proteins (CSPs) in the entire genome (17,551 genes), among all target genes included in the filtered GRN model of Tm neurons (the same as in Fig. 5C), and among the predicted targets (in the same model) of each TF shown. The total number of genes considered for each group are indicated in parentheses. **E,** GO enrichment analysis of the predicted targets of Hr3, Pdm3 and Mef2 in the filtered Tm GRN model (same as in D). Molecular Function terms with greater than two-fold enrichment were summarized by REVIGO to eliminate redundant terms and group related ones together. Areas within the graphs were determined by the p-values of the terms. **F,** *TmX/Tm4,6-Gal4* driving CD4tdGFP (flip-out) and UAS-fra (N=37 Tm4 neurons in 9 brains). 3D reconstructions of GFP expression for the representative adult neurons in each condition with anti-NCad (white). Arrowhead points to the dendritic fork normally characteristic of Tm2s (compare to control in C). Scale bars: 10 μm (A-bottom), 3 μm (A-top), 7 μm (C) and 15 μm (F).

Table S1 (separate file): Candidate terminal selectors of the optic lobe neurons: Source data for Figure S1A.

Table S2 (separate file): Differential gene expression analysis of the Tm subclusters in scRNA-seq

Table S3 (separate file): Inferelator models learned from the Tm and Lamina neurons: Produced using the respective MergedDA priors, multitask from the P30-40-50 and P24-36-48 datasets.

Table S4: Genotypes of animals used for all experiments in figure panels and the temperature at which each experiment was performed

Table S5: Origins of all reagents, fly strains, antibodies and software used in the study

Table S6: Full names of all gene symbols referenced in the study

Table S4

Ctrl: Control, Exp: Experiment, FSF: FRT-stop-FRT

Panels	Genotype	Temp
S3B, S5A	hsFLP2:PEST/+; UAS-FSF-CD4tdGFP/+; R35H01-Gal4/+	25
S3C	hsFLP2:PEST/+; TI(CRIMIC.TG4.2)^{Wnt10CR01661}/+ ; UAS-FSF-CD4tdGFP/+	25
S3D, S7K	hsFLP2:PEST/+; UAS-FSF-CD4tdGFP/+; R71F05-Gal4/+	25
1C, S3E	Ctrl: hsFLP122, UAS-CD8GFP/+; FRT42D, tub-Gal80/FRT42D; R71F05-Gal4/UAS-CD4tdGFP Exp: hsFLP122, UAS-CD8GFP/+; FRT42D, tub-Gal80/FRT42D, pdm3^{1}; R71F05-Gal4/UAS-CD4tdGFP	25
1D-F	Ctrl: hsFLP2:PEST/+; UAS-FSF-CD4tdGFP/+; R71F05-Gal4/+ Exp: hsFLP2:PEST/+; UAS-FSF-CD4tdGFP/ P(TRiP.HMJ21205); R71F05-Gal4/+	29
1G	Ctrl: Same as above Exp: hsFLP2:PEST/+; UAS-FSF-CD4tdGFP/ UAS-pdm3.short; R71F05-Gal4/+	25
1H-O, S4B	Ctrl: ;UAS-Stinger/+; R71F05-Gal4/+ Exp: ;UAS-Stinger/UAS-pdm3.short; R71F05-Gal4/+	25
S3F	Ctrl: hsFLP2:PEST/+; UAS-FSF-CD4tdGFP/+; 27b-Gal4/+ Exp: hsFLP2:PEST/+; UAS-FSF-CD4tdGFP/UAS-pdm3.short; 27b-Gal4/+	25
2A-C, S5B-C	Ctrl: hsFLP2:PEST/+; UAS-FSF-CD4tdGFP/+; R35H01-Gal4/+ Exp: hsFLP2:PEST/+; UAS-FSF-CD4tdGFP/ Mi(Hto-WP)Drgx^{GLO}; R35H01-Gal4/+	29
2E	; tub-Gal80^{ts}/+; pxb-Gal4, UAS-CD8GFP/+	29
2F	; tub-Gal80^{ts}/+; pxb-Gal4, UAS-CD8GFP/UAS-Klu.HA	29
2G-I, S5D-F	Ctrl: hsFLP2:PEST/+; Drgx$^{\Delta Tm1}$/+ ; 27b-Gal4/UAS-FSF-CD4tdGFP Exp: hsFLP2:PEST/+; Drgx$^{\Delta Tm1}$/Drgx$^{\Delta Tm1}$; 27b-Gal4/UAS-FSF-CD4tdGFP	25
S5H	Ctrl: w^{1118} ;; Exp: w^{1118} ; Drgx$^{\Delta Tm1}$/Drgx$^{\Delta Tm1}$;	25
3A-B	Ctrl: hsFLP122, UAS-CD8GFP/+; FRT40A, tub-Gal80/FRT40A; R35H01-Gal4/UAS-CD4tdGFP Exp: hsFLP122, UAS-CD8GFP/+; FRT40A, tub-Gal80/FRT40A, SoxN^{NC14}; R35H01-Gal4/UAS-CD4tdGFP	25
S7A	hsFLP2:PEST/+; tub-Gal80^{ts}/UAS-SoxN.V5; R35H01-Gal4/UAS-FSF-CD4tdGFP	29
S7B	Ctrl: hsFLP122, UAS-CD8GFP/+; FRT40A, tub-Gal80/FRT40A; R35H01-Gal4/UAS-CD4tdGFP Exp: hsFLP122, UAS-CD8GFP/+; FRT40A, tub-Gal80/FRT40A, aop^{E833}; R35H01-Gal4/UAS-CD4tdGFP	25
S7C	Ctrl: hsFLP2:PEST/+; UAS-FSF-CD4tdGFP/+; R35H01-Gal4/+ Exp: hsFLP2:PEST/+; UAS-FSF-CD4tdGFP/+; R35H01-Gal4/ P(TRiP.HMS01256)	29
3C	Ctrl: hsFLP2:PEST/+; UAS-FSF-CD4tdGFP/+; R71F05-Gal4/+ Exp: hsFLP2:PEST/+; UAS-FSF-CD4tdGFP/+; R71F05-Gal4/ P(TRiP.HMS01256)	29
3F-G, S7E-G	Ctrl: hsFLP2:PEST/+; UAS-FSF-CD4tdGFP/+; R76F01-Gal4/+ Exp: hsFLP2:PEST/+; UAS-FSF-CD4tdGFP/UAS-Vsx1; R76F01-Gal4/+	29

S7D	hsFLP2:PEST/+; UAS-FSF-CD4tdGFP/UAS-Vsx2.CC; R76F01-Gal4/+	29
4A,E	Ctrl: hsFLP2:PEST/+; UAS-FSF-CD4tdGFP/+; R71F05-Gal4/+ Exp: hsFLP2:PEST/+; UAS-FSF-CD4tdGFP/P(TRiP.HMS01691); R71F05-Gal4/+	29
4B,F	hsFLP2:PEST/+; UAS-FSF-CD4tdGFP/UAS-aop.WT; R71F05-Gal4/+	29
4C,G	hsFLP2:PEST/+; UAS-FSF-CD4tdGFP/UAS-aop.ACT; R71F05-Gal4/+	29
5H, S9A	Ctrl: hsFLP122, UAS-CD8GFP/+; FRT42D, tub-Gal80/FRT42D; R71F05-Gal4/UAS-CD4tdGFP Exp: hsFLP122, UAS-CD8GFP/+; FRT42D, tub-Gal80/FRT42D, $Hr3^{K10308}$; R71F05-Gal4/UAS-CD4tdGFP	25
5J	Ctrl: hsFLP2:PEST/+; UAS-FSF-CD4tdGFP/+; R71F05-Gal4/+ Exp: hsFLP2:PEST/+; UAS-FSF-CD4tdGFP/$tub\text{-}Gal80^{ts}$; R71F05-Gal4/UAS-Blimp1	29
5K	Ctrl: hsFLP2:PEST/+; UAS-FSF-CD4tdGFP/+; 27b-Gal4/+ Exp: hsFLP2:PEST/+; UAS-FSF-CD4tdGFP/+; 27b-Gal4/P(TRiP.HMS00924)	29
S9C	Ctrl: hsFLP2:PEST/+; UAS-FSF-CD4tdGFP/+; R71F05-Gal4/+ Exp: hsFLP2:PEST/Mi(Hto-WP)ct^{BRO}; UAS-FSF-CD4tdGFP/+; R71F05-Gal4/+	29
S9F	Ctrl: hsFLP2:PEST/+; UAS-FSF-CD4tdGFP/+; R35H01-Gal4/+ Exp: hsFLP2:PEST/+; UAS-FSF-CD4tdGFP/+; R35H01-Gal4/ UAS-fra	29

Table S5

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Critical Commercial Resources		
Chromium Next GEM Single Cell 3' GEM, Library & Gel Bead Kit v3.1	10X Genomics	1000128
Chromium Next GEM Chip G Kit	10X Genomics	1000127
Single Index Kit T Set A	10X Genomics	1000213
10x™ Magnetic Separator	10X Genomics	230003
PCR Tubes 0.2 ml 8-tube strips	Eppendorf	0030124286
DNA LoBind Tubes, 1.5 ml	Eppendorf	022431021
SPRIselect Reagent Kit	Beckman Coulter	B23318
High Sensitivity DNA Kit	Agilent Technologies	5067-4626
Hemacytometer C-Chip 0.02 mm	Fisher	22-600-109
Qubit® dsDNA HS Assay Kit	Thermo Fisher Scientific	Q32854
pluriStrainer Mini 20 µm	pluriSelect	43-10020-40
Chemicals, Peptides, and Recombinant Proteins		
Ethanol, Pure (200 Proof, anhydrous)	Sigma	459836
10% Tween 20	Bio-Rad	1610781
Formaldehyde	Merck KGaA	1.03999.1000
PBS 20X	Growcells.com	MRGF-6396
Triton X-100	Sigma	T8787
Schneider's Drosophila Medium [+] L-Glutamine	Gibco	21720-024
Collagenase (C. histolyticum)	Sigma	C0130
Dispase II	Sigma	D4693-1G
BSA	Sigma	A7906
DPBS	Corning/Fisher	21-031-CV
SSC 20X	Quality Biological	351-003-131
Experimental Model: Organisms/Strains		
Drosophila, w1118	WellGenetics	N/A
Drosophila, hsFLP2:PEST (X)	Ref (*66*)	N/A
Drosophila, 27b-Gal4 (III)	Ref (*30*)	N/A
Drosophila, pxb-Gal4 (III)	Ref (*29*)	N/A
Drosophila, UAS-pdm3.short (II)	Ref (*25*)	N/A
Drosophila, UAS-Vsx1 (II)	Ref (*5*)	N/A
Drosophila, FRT40A, aop^{E833}	Ref (*32*)	N/A

Drosophila, FRT42D, pdm3^{1}	Ref (*24*)	N/A
Drosophila, R35H01-Gal4	Bloomington Drosophila Stock Center (BDSC)	49922
Drosophila, R71F05-Gal4	BDSC	48303
Drosophila, R76F01-Gal4	BDSC	39934
Drosophila, Wnt10-Gal4	BDSC	86484
Drosophila, hsFLP122 (X)	BDSC	8862
Drosophila, UAS-CD4:tdGFP (III)	BDSC	35836
Drosophila, UAS-Stinger (II)	BDSC	1956
Drosophila, UAS-pdm3.RNAi (II)	BDSC	53887
Drosophila, UAS-Mef2.RNAi (II)	BDSC	38247
Drosophila, UAS-aop.RNAi (III)	BDSC	34909
Drosophila, UAS-ct.RNAi (III)	BDSC	33967
Drosophila, UAS-aop.WT (II)	BDSC	5790
Drosophila, UAS-aop.ACT (II)	BDSC	5789
Drosophila, UAS-Drgx (II)	BDSC	56546
Drosophila, UAS-ct (X)	BDSC	56538
Drosophila, UAS-SoxN.V5 (II)	BDSC	83300
Drosophila, tub-Gal80^{ts} (II)	BDSC	7108
Drosophila, FRT40A	BDSC	8212
Drosophila, FRT42D	BDSC	1802
Drosophila, FRT40A, tub-Gal80	BDSC	5192
Drosophila, FRT42D, tub-Gal80	BDSC	9917
Drosophila, FRT40A, SoxN^{NC14}	BDSC	91794
Drosophila, UAS-fra (III)	BDSC	8814
Drosophila, UAS-Klu.HA (III)	FlyORF	F000578
Drosophila, UAS-Vsx2.CC (II)	FlyORF	F003265
Drosophila, FRT42D, Hr3^{K10308}	Kyoto Stock Center	111288
Drosophila, UAS-FSF-CD4:tdGFP (II)	Gift from Robin Hiesinger	N/A
Drosophila, UAS-FSF-CD4:tdGFP (III)	Gift from Robin Hiesinger	N/A
Drosophila, UAS-Blimp-1 (III)	Gift from Jens Rister and Gerald Call	N/A
Drosophila, Drgx$^{\Delta Tm1}$ (II)	This paper (WellGenetics)	N/A
Deposited Data		
Single-cell RNA sequencing (Raw and analyzed data)	This paper	GEO: GSE199734

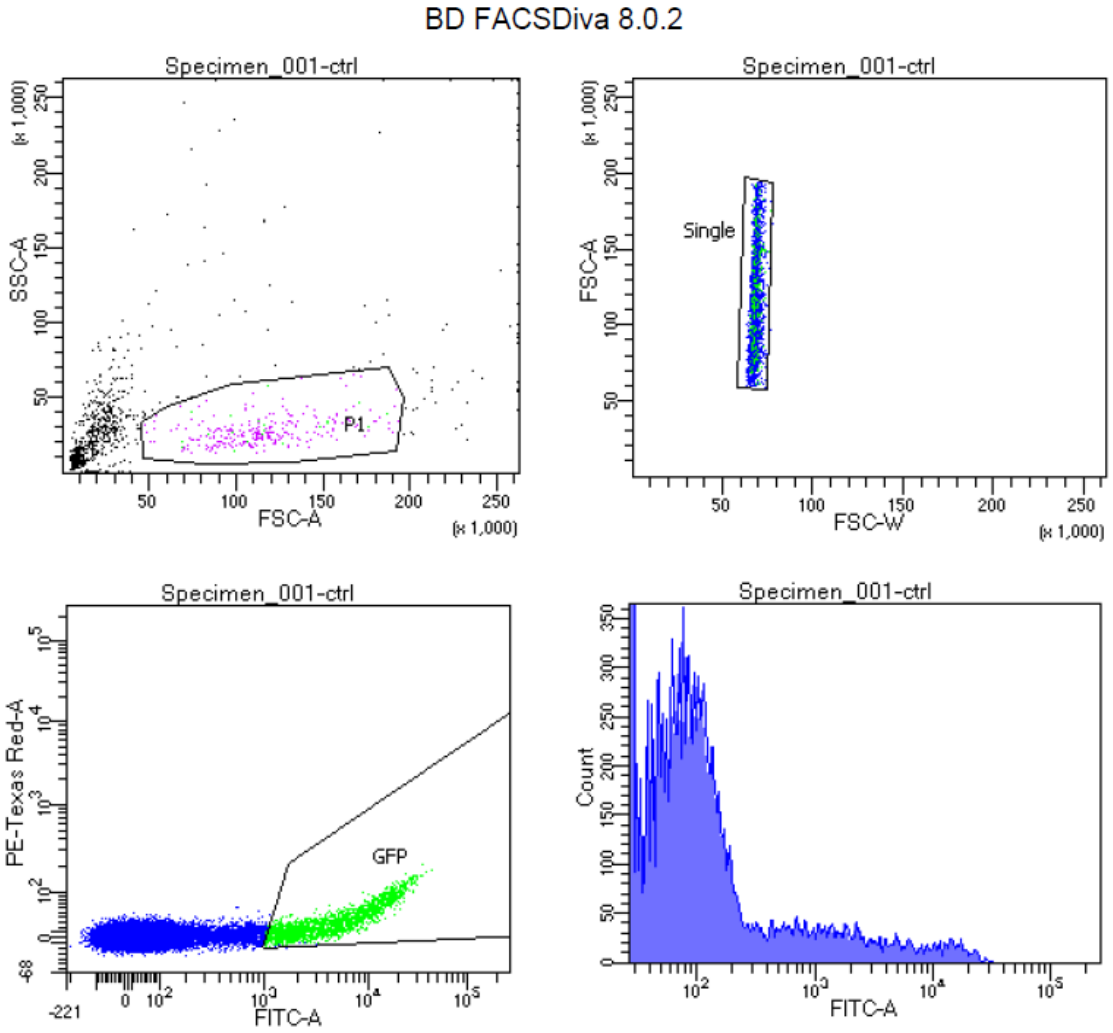
Antibodies		
Mouse anti-Brp	Developmental Studies Hybridoma Bank (DSHB)	nc82
Mouse anti-Aop	DSHB	8B12H9
Rat anti-NCad	DSHB	DN-ex8
Mouse anti-V5:DyLight550	BioRad	MCA1360D550GA
Chicken anti-GFP	Millipore Sigma	06-896
Rabbit anti-(cleaved)Dcp1	Cell Signaling Technology	9578
Rabbit P-p44/42 anti-MAPK	Cell Signaling Technology	9101S
Guinea Pig anti-Pdm3	Ref (*24*)	N/A
Rabbit anti-Vmat (N-term)	Ref (*56*)	N/A
Guinea Pig anti-Vsx1	Ref (*5*)	N/A
Guinea Pig anti-Runt	Ref (*6*)	N/A
Guinea Pig anti-Brp	This paper	N/A
Rabbit anti-SoxN	This paper	N/A
Rabbit anti-Mef2	This paper	N/A
Rat anti-Drgx	This paper	N/A
Rat anti-Pdm3	This paper	N/A
Guinea Pig anti-Repo	This paper	N/A
Rabbit anti-Vsx2	This paper	N/A
Donkey anti-chicken Alexa 488	Jackson ImmunoResearch	703-545-155
Donkey anti-rabbit DyLight 405	Jackson ImmunoResearch	711-475-152
Donkey anti-rabbit Alexa 488	ThermoFisher Scientific	A31570
Donkey anti-rabbit Cy3	Jackson ImmunoResearch	711-165-152
Donkey anti-rabbit Alexa 647	Jackson ImmunoResearch	711-605-152
Donkey anti-mouse Alexa 488	Jackson ImmunoResearch	715-545-150
Donkey anti-mouse Alexa 555	ThermoFisher Scientific	A31570
Donkey anti-mouse Alexa 647	Jackson ImmunoResearch	715-605-151
Donkey anti-rat Alexa 647	Jackson ImmunoResearch	712-605-153
Donkey anti-rat Cy3	Jackson ImmunoResearch	712-165-153
Donkey anti-rat Alexa 488	Jackson ImmunoResearch	712-545-150
Donkey anti-guinea pig DyLight 405	Jackson ImmunoResearch	706-475-148
Donkey anti-guinea pig Alexa 488	Jackson ImmunoResearch	706-545-148
Donkey anti-guinea pig Cy3	Jackson ImmunoResearch	706-165-148
Donkey anti-guinea pig Alexa 647	Jackson ImmunoResearch	706-605-148
Software and Algorithms		
Prism 9	GraphPad	N/A
Imaris	Bitplane, Switzerland	N/A
CellRanger	10X Genomics	N/A

R Statistical Computing Software version 4	R Project	N/A
RStudio	RStudio, Inc	N/A
Python 2.7	Anaconda, Inc	N/A
Python 3.6	Anaconda, Inc	N/A

Table S6

Symbol	Gene Name		Symbol	Gene Name
bsh	brain-specific homeobox		eIB	elbow B
hth	homothorax		Ets65A	E26 transformation specific at 65A
vvl	ventral veins lacking		Vsx1	Visual system homeobox 1
Lim1	LIM homeobox 1		Vsx2	Visual system homeobox 2
erm	earmuff		5-HT7	5-hydroxytryptamine (serotonin) receptor 7
SoxN	SoxNeuro		Octβ1R	Octopamine β1 receptor
Drgx	Dorsal root ganglia homeobox		Or63a	Odorant receptor 63a
TfAP-2	Transcription factor AP-2		Dh44-R1	Diuretic hormone 44 receptor 1
aop	Anterior open		MAPK	mitogen activated protein kinase
Wnt10	Wnt oncogene analog 10		InR	Insulin-like receptor
ap	apterous		Alk	Anaplastic lymphoma kinase
scro	scarecrow		Hr3	Hormone receptor 3
ct	cut		Hr39	Hormone receptor-like in 39
Camta	Calmodulin-binding transcription activator		Eip74EF	Ecdysone-induced protein 74EF
pdm3	pou domain motif 3		Eip93F	Ecdysone-induced protein 93F
Mef2	Myocyte enhancer factor 2		Fezf2	FEZ Family Zinc Finger 2
Dcp-1	Death caspase-1		Dscam4	Down syndrome cell adhesion molecule 4
Wnt4	Wnt oncogene analog 4		fra	frazzled
Klu	Klumpfuss		Pet1	PC12 ETS Domain-Containing Transcription Factor 1
repo	reversed polarity		Lmx1a	LIM Homeobox Transcription Factor 1 Alpha
Vmat	Vesicular monoamine transporter		Lmx1b	LIM Homeobox Transcription Factor 1 Beta
Dll	Distal-less		Nurr1	Nuclear receptor related 1 protein
fd59A	forkhead domain 59A		Imp	IGF-II mRNA-binding protein
ham	hamlet		Syp	Syncrip
noc	no ocelli			

FACS Gating Strategy



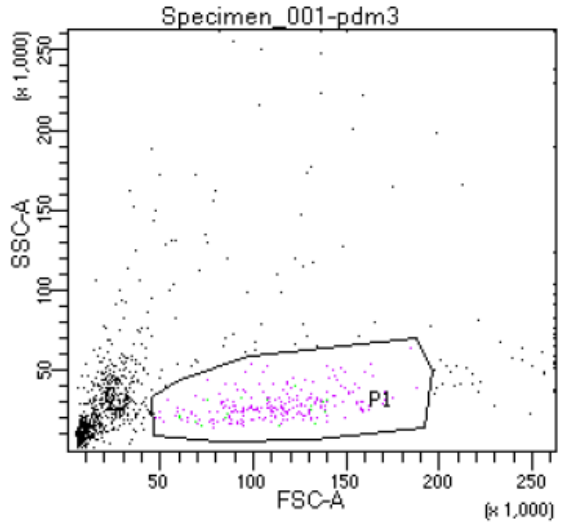
Experiment Name:	Neset_006
Specimen Name:	Specimen_001
Tube Name:	ctrl
Record Date:	Sep 21, 2021 2:41:41 PM
$OP:	Administrator
GUID:	f6614b59-4cf6-4096-9a90-8c92...

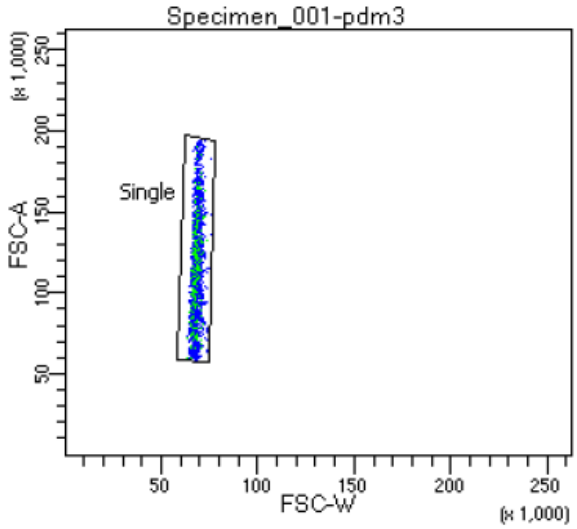
Population	#Events	%Parent	DsRed-A Mean	DsRed-A Mean
All Events	100,000	####	####	4
P1	27,843	27.8	####	2
Single	26,749	96.1	####	2
GFP	2,059	7.7	####	2

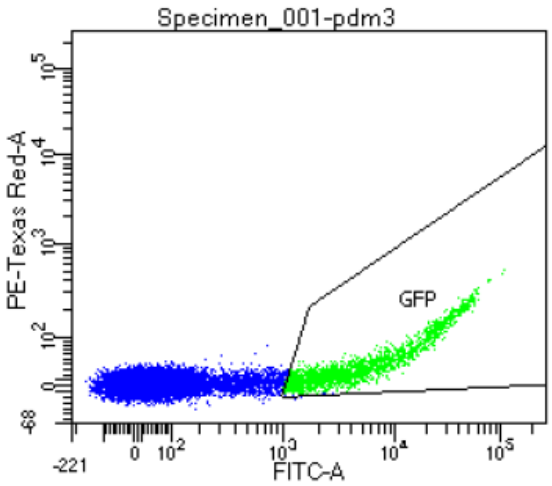
Tube: ctrl

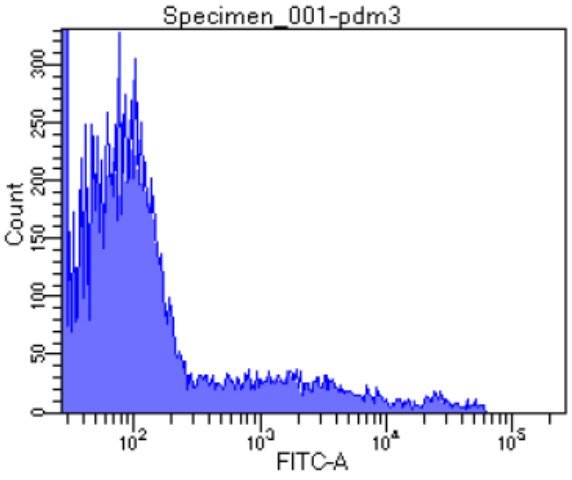
Population	#Events	%Parent	%Total
All Events	100,000	####	100.0
P1	27,843	27.8	27.8
Single	26,749	96.1	26.7
GFP	2,059	7.7	2.1

BD FACSDiva 8.0.2









Experiment Name:	Neset_006
Specimen Name:	Specimen_001
Tube Name:	pdm3
Record Date:	Sep 21, 2021 2:53:50 PM
$OP:	Administrator
GUID:	39f6ec45-bf38-40fd-ada3-8a3c...

Population	#Events	%Parent	DsRed-A Mean	DsRed-A Mean
All Events	100,000	####	####	4
P1	24,382	24.4	####	3
Single	23,338	95.7	####	3
GFP	2,054	8.8	####	4

Tube: pdm3

Population	#Events	%Parent	%Total
All Events	100,000	####	100.0
P1	24,382	24.4	24.4
Single	23,338	95.7	23.3
GFP	2,054	8.8	2.1