

# Transcriptional Programs of Circuit Assembly in the *Drosophila* Visual System

## Highlights

- Multiplex single-cell sequencing of many developing brain samples in parallel
- mRNA expression of >150 *Drosophila* visual system neurons throughout development
- A synchronous pan-neuronal program regulates gene expression in different neurons
- Dynamic cell-type-specific expression of cell surface proteins during wiring

## Authors

Yerbol Z. Kurmangaliyev, Juyoun Yoo, Javier Valdes-Aleman, Piero Sanfilippo, S. Lawrence Zipursky

## Correspondence

lzipursky@mednet.ucla.edu

## In Brief

Kurmangaliyev et al. use a genetic multiplexing strategy to generate a comprehensive atlas of the developing *Drosophila* visual system. This atlas includes transcriptional programs for more than 150 distinct neuronal populations. It covers multiple stages of neuronal development, from the early steps of circuit assembly to the adult functional brain.



## NeuroResource

# Transcriptional Programs of Circuit Assembly in the *Drosophila* Visual System

Yerbol Z. Kurmangaliyev,<sup>1</sup> Juyoun Yoo,<sup>1,2</sup> Javier Valdes-Aleman,<sup>1</sup> Piero Sanfilippo,<sup>1</sup> and S. Lawrence Zipursky<sup>1,3,\*</sup><sup>1</sup>Department of Biological Chemistry, Howard Hughes Medical Institute, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA 90095, USA<sup>2</sup>Neuroscience Interdepartmental Program, University of California, Los Angeles, Los Angeles, CA 90095, USA<sup>3</sup>Lead Contact\*Correspondence: [lzipursky@mednet.ucla.edu](mailto:lzipursky@mednet.ucla.edu)<https://doi.org/10.1016/j.neuron.2020.10.006>

## SUMMARY

Precise patterns of synaptic connections between neurons are encoded in their genetic programs. Here, we use single-cell RNA sequencing to profile neuronal transcriptomes at multiple stages in the developing *Drosophila* visual system. We devise an efficient strategy for profiling neurons at multiple time points in a single pool, thereby minimizing batch effects and maximizing the reliability of time-course data. A transcriptional atlas spanning multiple stages is generated, including more than 150 distinct neuronal populations; of these, 88 are followed through synaptogenesis. This analysis reveals a common (pan-neuronal) program unfolding in highly coordinated fashion in all neurons, including genes encoding proteins comprising the core synaptic machinery and membrane excitability. This program is overlaid by cell-type-specific programs with diverse cell recognition molecules expressed in different combinations and at different times. We propose that a pan-neuronal program endows neurons with the competence to form synapses and that cell-type-specific programs control synaptic specificity.

## INTRODUCTION

Neural circuits in brains from worms to mammals are largely hard wired. Genetic and molecular studies have provided mechanistic insights into how circuits develop (Sanes and Zipursky, 2020). A small, largely evolutionarily conserved set of intercellular signaling molecules and a more diverse set of transcription factors (TFs) act early in development to pattern neural tissue and regulate neuronal cell fate (Holguera and Desplan, 2018). TFs called terminal selectors have been identified that regulate selective features of neurons, such as the specific neurotransmitter a neuron uses (Hobert and Kratsios, 2019). Combinatorial transcriptional programs have been identified in *Drosophila* that specify unique dendritic and axonal morphologies (Enriquez et al., 2018). At later stages of development, transcriptional programs also specify patterns of synaptic connectivity, with examples of specific TFs controlling the expression of genes encoding proteins that regulate wiring (Morey et al., 2008; Liu et al., 2018). Recent single-cell sequencing studies in the *Drosophila* olfactory and visual systems suggest that a relatively small number of TFs drive neuron diversity and that a much larger set of cell surface proteins regulated by them specify patterns of synaptic connectivity (Li et al., 2017a; Kurmangaliyev et al., 2019).

The formation of connections between neurons unfolds during development in a stepwise fashion. During axon guidance, growth cones at the leading edge of axons extend along stereotyped paths to their target regions. Here, other neurons elaborate

dendrites. Within the target region, axons intermingle with dendritic processes of tens to hundreds of different neuron types. It is during this period in development that neurites form synaptic connections and acquire their mature signaling properties. Whereas genetic and biochemical studies have led to a detailed molecular description of axon guidance (Dickson, 2002; Tessier-Lavigne and Goodman, 1996), our understanding of how neurites discriminate between one another to select their appropriate synaptic partners remains fragmentary (Sanes and Zipursky, 2020). Several recent advances have brightened the prospects of uncovering molecules, mechanisms, and principles underlying synaptic specificity. Among these, single-cell RNA sequencing (scRNA-seq) provides a unique opportunity to characterize expression of cell recognition molecules expressed in synaptic partners during synapse formation.

The *Drosophila* visual system is well suited to studying the genetic programs of synaptic specificity due to the availability of synapse-level connectivity maps of more than a hundred well-defined cell types (Fischbach and Dittrich, 1989; Takemura et al., 2013, 2015) and the availability of genetic and molecular tools to manipulate discrete classes of neurons. A number of recent studies used single-cell sequencing to characterize neurons in the adult fly brain, including the visual system (Davie et al., 2018; Konstantinides et al., 2018; Croset et al., 2018; Allen et al., 2020). However, these datasets are of limited use for studying developmental processes, as transcriptomes during development are dynamic and quite different from their adult



counterparts (Li et al., 2017a; Sarin et al., 2018; Kurmangaliyev et al., 2019).

Here, we set out to generate a transcriptional atlas of the developing *Drosophila* visual system during circuit assembly. We devised a strategy for single-cell sequencing of multiple time points in a pooled fashion to minimize the confound of batch effects. We identified 162 distinct neuronal populations and followed 88 of them at seven time points just prior to, during, and following synapse formation. This enabled us to uncover a common (pan-neuronal) transcriptional program in all neuron types that proceeds in a synchronous fashion. This program includes genes encoding core protein components of synapses and membrane excitability. The diversity of cell-type-specific programs was driven by dynamically expressed sets of cell recognition molecules and other genes involved in intercellular interactions. We propose that the pan-neuronal program endows all neurons with a competence to form synapses, whereas synaptic specificity is determined by cell-type-specific repertoires of cell surface proteins with distinct temporal expression patterns and recognition specificities.

## RESULTS AND DISCUSSION

### Profiling of Different Stages of Neuronal Development in Parallel in a Single Experiment

In this study, we generated gene expression maps of postmitotic neurons in the fly visual system throughout pupal development (~100 h) using high-throughput scRNA-seq (Zheng et al., 2017). During this time, neurons complete axon guidance, elaborate distinct axonal and dendritic morphologies, select appropriate synaptic partners, form synapses, and acquire distinct biophysical properties.

Two different experimental designs were used. First, we profiled optic lobes at 24, 48, 72, and 96 h after puparium formation (APF) (Figures 1A and S1). Each time point was profiled as a separate experiment, and for the 48-h time point, we profiled two independent samples. We refer to this dataset as W1118 based on the strain used for these experiments. This provided a high-resolution dataset for unsupervised identification of cell types at different stages of development. The largest number of cells (51,000) was profiled at 48 h APF, when the transcriptional differences between neuronal cell types were maximal (Li et al., 2017a).

A second set of experiments was undertaken to follow changes in gene expression at 12-h intervals to more precisely define the developmental trajectory of each cell type (Figures 1B and S1). We developed a strategy to profile neurons from multiple time points in a single experiment. This strategy exploits natural genetic variation in wild-type strains from the *Drosophila* Genetic Reference Panel (DGRP) to mark cells isolated at different developmental stages (Kang et al., 2018; Mackay et al., 2012; Huang et al., 2014). Males from individual DGRP strains were crossed to females from a common reference strain (W<sup>1118</sup>) to reduce differences in genetic background (Fear et al., 2016). Pupae from the resulting progeny were collected every 12 h from 0 to 96 h APF. For every time point, we staged three individual animals, each tagged by a unique wild-type genotype. Optic lobes for all time points were dissected, pooled, and processed as a single sample from tissue dissociation to

sequencing. The resulting single-cell transcriptomes for each time point and individual animals were separated based on SNPs (single-nucleotide polymorphisms) captured in mRNA sequences and matched with the genotypes of parental strains (Figure 1C). In this way, we profiled cells at nine stages of development with three biological replicates for each time point (total 27 samples) in a single experiment. We performed two rounds of this experiment with reshuffled wild-type parental genotypes (Figures 1B and S1).

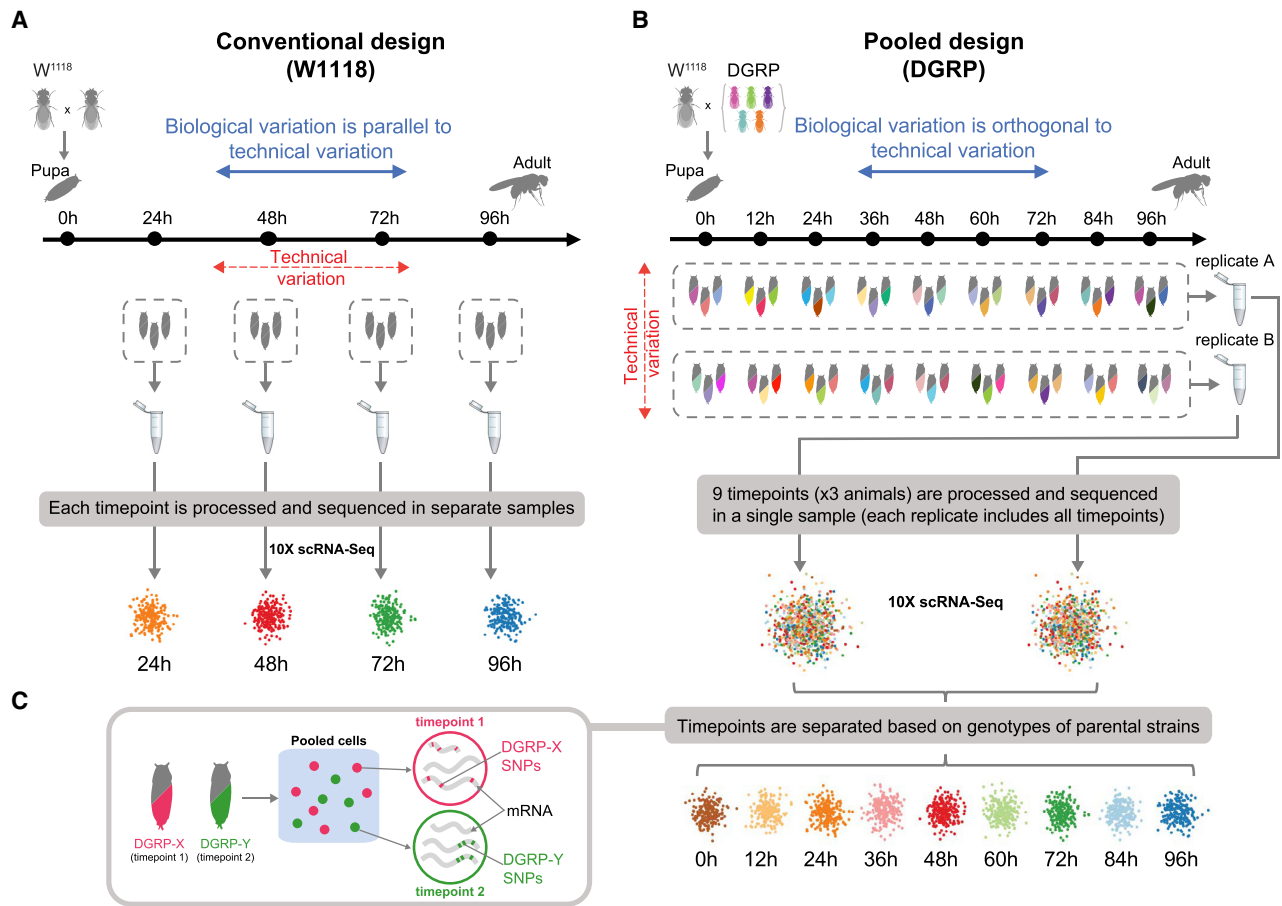
Profiling pools of cells together from all time points in a single experiment has several advantages. First, it is highly efficient and cost effective. The entire experiment, from genetic crosses to single-cell capture, took ~2 weeks. Second, it largely eliminates batch effects. An inherent limitation in the conventional design of scRNA-seq is the technical variation (batch effects) between different samples (Figure S2). In droplet-based scRNA-seq, for instance, mRNA released from tissue dissociation prior to cell capture is a source of contaminants confounding comparison between samples. In the conventional design, for instance, a contaminant (e.g., a transcript expressed at high levels at a specific time in a subset of cells) may appear as ubiquitously upregulated at this time in all neurons. In the pooled design, however, background contaminants that change with time are distributed among all cells and times and thus are not erroneously attributed to biological differences. Similarly, in pooled samples, other sources of technical variability of unknown etiology arising in the conventional design are eliminated. Finally, genetic multiplexing enables identification of cell doublets, further improving the quality of datasets and reducing cost through maximizing single-cell capture rates in scRNA-seq platforms (Kang et al., 2018).

In summary, we generated a comprehensive dataset of gene expression in visual system neurons covering every 12 h of pupal development. After quality control, the dataset comprises 208,976 single-cell transcriptomes with a median of 1,650 genes and 5,148 transcripts detected per cell (minimum 2,000 transcripts). Each of the W1118 and DGRP approaches represented approximately half of the dataset. The complete experimental design and distribution of cells across time points are shown in Figure S1.

### A Transcriptional Atlas of the Developing *Drosophila* Visual System

Transcriptional differences between time points for a single neuronal cell type can be comparable to the differences between cell types (Li et al., 2017a; Sarin et al., 2018). Therefore, to follow individual cell types over time, we first integrated single-cell datasets across different stages of development. Both W1118 and DGRP samples were analyzed together to achieve maximal resolution in cell type diversity and to generate a unified atlas of transcriptional cell types. Samples from 0 and 12 h APF differed substantially from other time points both in tissue composition and coverage and were analyzed separately (the early dataset). The main analysis was focused on samples from 24 to 96 h APF (the main dataset).

The integrative analysis of the main dataset was performed using CCA (canonical correlation analysis)- and MNN (mutual nearest neighbors)-based workflow implemented in Seurat 3 (Stuart et al., 2019). The integration was performed on the levels of



**Figure 1. Transcriptional Profiling of Multiple Stages of Neuronal Development in a Single Experiment**

Schematics of experimental designs.

(A) Conventional design (W1118 dataset). Four stages of pupal development were profiled in separate experiments.

(B) Pooled design (*Drosophila* Genetic Reference Panel [DGRP] dataset). Nine stages of pupal development were pooled and profiled in the same experiment. Cells from each time point were tagged using three unique wild-type strains from the DGRP. Wild-type strains were reshuffled between time points in two independent replicates.

(C) Single-cell transcriptomes for each time point were separated based on SNPs (single-nucleotide polymorphisms) captured in mRNA molecules and matching with unique genotypes of pupae.

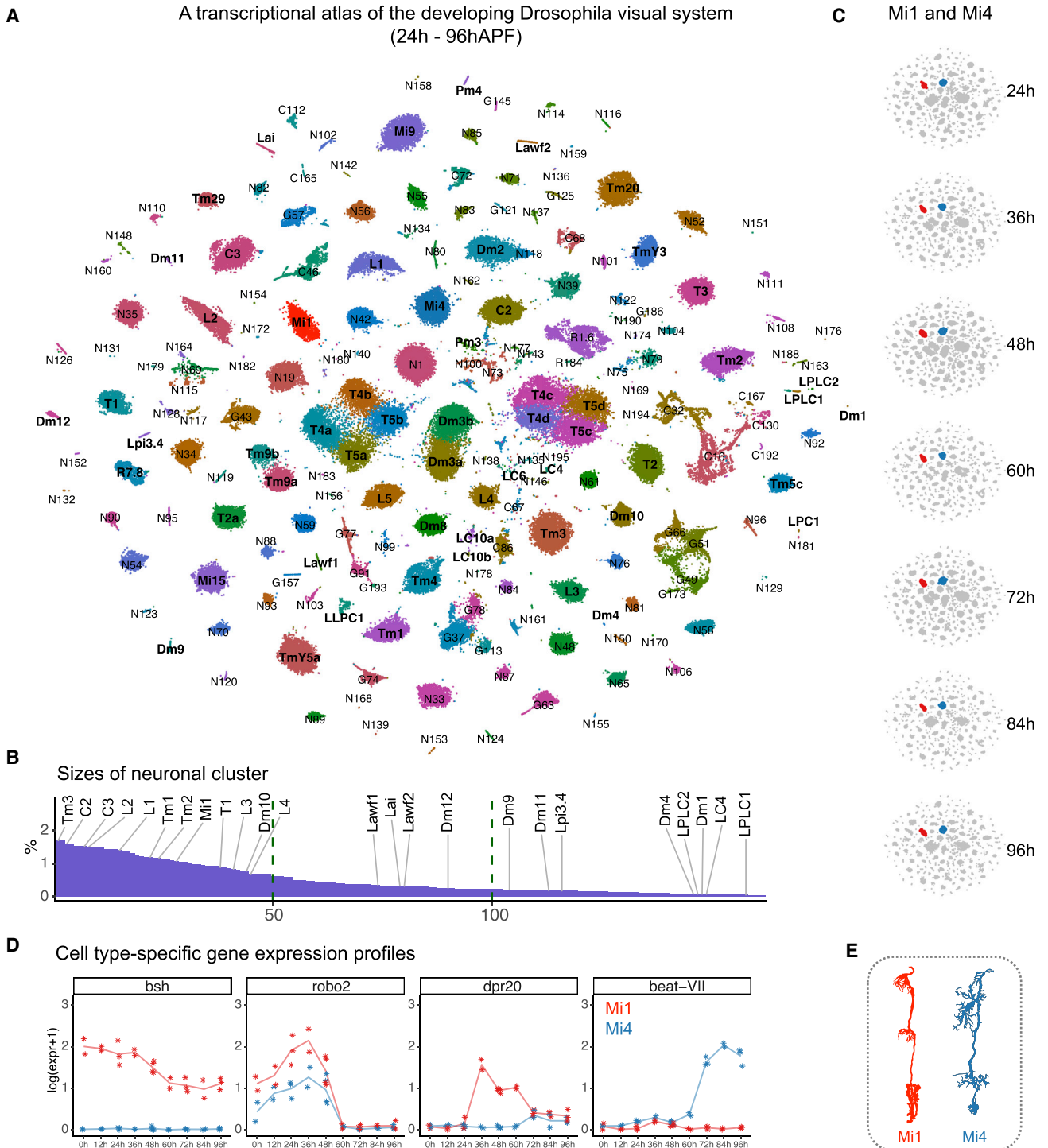
In the conventional design, biological differences between time points cannot be separated from technical variation between different samples. In the pooled design, differences between time points are independent of technical variation between samples. This minimizes the confound of batch effects on constructing an accurate time series of gene expression between time points (see effects on background contributed by ambient RNA in Figure S2). See also Figure S1 for details on experimental design and workflow of the analysis.

individual time points and replicates. An integrated dataset was used for dimensionality reduction followed by a graph-based clustering method that revealed 196 transcriptionally distinct cell populations (see STAR Methods for details). The results of dataset integration and clustering analysis were visualized using t-distributed stochastic neighbor embedding (tSNE) plots (Figures 2 and S3). Importantly, the integrative analysis was used strictly for joint clustering of cell types across different samples (i.e., time points and replicates). Once we identified clusters, however, we returned to the original (pre-integration) normalized gene expression values for each time point and sample (Figure 2D).

Most of the clusters were detectable in all samples and comprise a comparable fraction of cells across different time

points, while some were distributed less evenly (Figure S3). The variation in cluster proportions could be due to technical reasons (e.g., sampling variation or differences in tissue dissections) or biological differences in cell composition at different stages. Unsupervised analysis grouped clusters into four major classes (Figure S4), which were annotated based on expression of known marker genes (see STAR Methods for details): 162 clusters of neurons (*elav*<sup>+</sup>), 3 clusters of photoreceptor cells (*chp*<sup>+</sup>), 19 clusters of glia (*repo*<sup>+</sup>), and 11 clusters of non-neuronal cells of unknown origin (*elav*<sup>-</sup>/*repo*<sup>-</sup>).

Next, we matched neuronal clusters to previously identified cell types. As a reference, we used two recent datasets from bulk sequencing of several dozen morphologically distinct cell



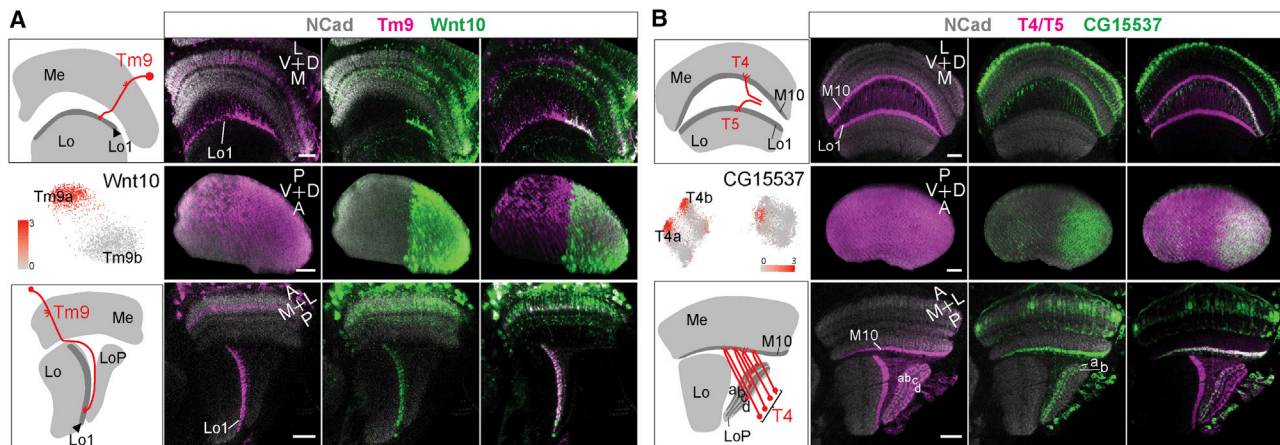
**Figure 2. A Transcriptional Atlas of the Developing *Drosophila* Visual System**

(A) tSNE plot of the main dataset. This includes samples from 24 to 96 h APF (DGRP and W1118). All samples were integrated and clustered together. Distinct transcriptional clusters are color coded and labeled by identities of clusters (i.e., names of neurons for known morphological cell types).

(B) The proportion of each neuronal cluster relative to the total number of neurons (median across all samples). Vertical bars correspond to the neuronal clusters ordered by average proportion (selected clusters of known morphological cell types are highlighted). All of the identified unicolumnar neurons were among the largest 50 clusters.

(legend continued on next page)





**Figure 3. Regional Specialization of Neurons**

(A and B) Transcriptionally distinct variants of Tm9 (A) and T4/T5 (B) neurons. The arrangement of cell types in the optic lobe and tSNE plots with expression patterns of marker genes for newly identified subtypes are shown on the left in each panel. All Tm9 (A) and T4/T5 (B) neurons are labeled in magenta. Expression patterns of marker genes in adult are shown in green. (A) There are two transcriptionally distinct populations of Tm9 neurons. They are defined by mutually exclusive expression of Wnt10 and Wnt4 (not shown) genes. Wnt10 is restricted to Tm9 neurons targeting the dorsal half of the lobula. The middle row shows the cropped lobula 1 layer (Lo1). (B) A subpopulation of T4 neurons is defined by expression of CG15537 restricted to T4 neurons targeting the dorsal third of the medulla and lobula plate layers a/b. The middle row shows the cropped medulla 10 layer (M10). Scale bar, 20  $\mu$ m.

types in the adult (Davis et al., 2020; Konstantinides et al., 2018). We identified a set of cluster-specific marker genes and used them for correlation analysis (Figures S5A and S5B). In addition, we used expression patterns of TFs to manually curate matched clusters (Figure S5C). Most of the reference cell types had a one-to-one match with a single cluster. In a few cases, a reference dataset had a high correlation with multiple clusters representing further heterogeneity. For example, T4/T5 neurons matched eight clusters with similar correlation values. Expression of previously identified marker genes allowed us to match these clusters to the eight known morphological subtypes of T4/T5 neurons (Kurmagaliyev et al., 2019). Moreover, we identified new transcriptional subtypes (e.g., two Dm3 and two Tm9 neurons; see below). Overall, cell-type-specific expression profiles of known morphological cell types were in good agreement with previously validated marker genes (e.g., Tan et al., 2015; Cosmanescu et al., 2018). For example, the TF bsh has been shown to be expressed in three cell types (Hasegawa et al., 2013), and it was detected only in the corresponding clusters (Figure S5C). The transcriptional similarities between clusters were also consistent with known anatomical and functional relationships between cell types (Figure S6). In total, we were able to match identities of 58 neuronal clusters to known morphological cell types (Figure 2A).

The optic lobe neuropils consist of repeated columns with stereotyped cell-type composition and connectivity (Takemura

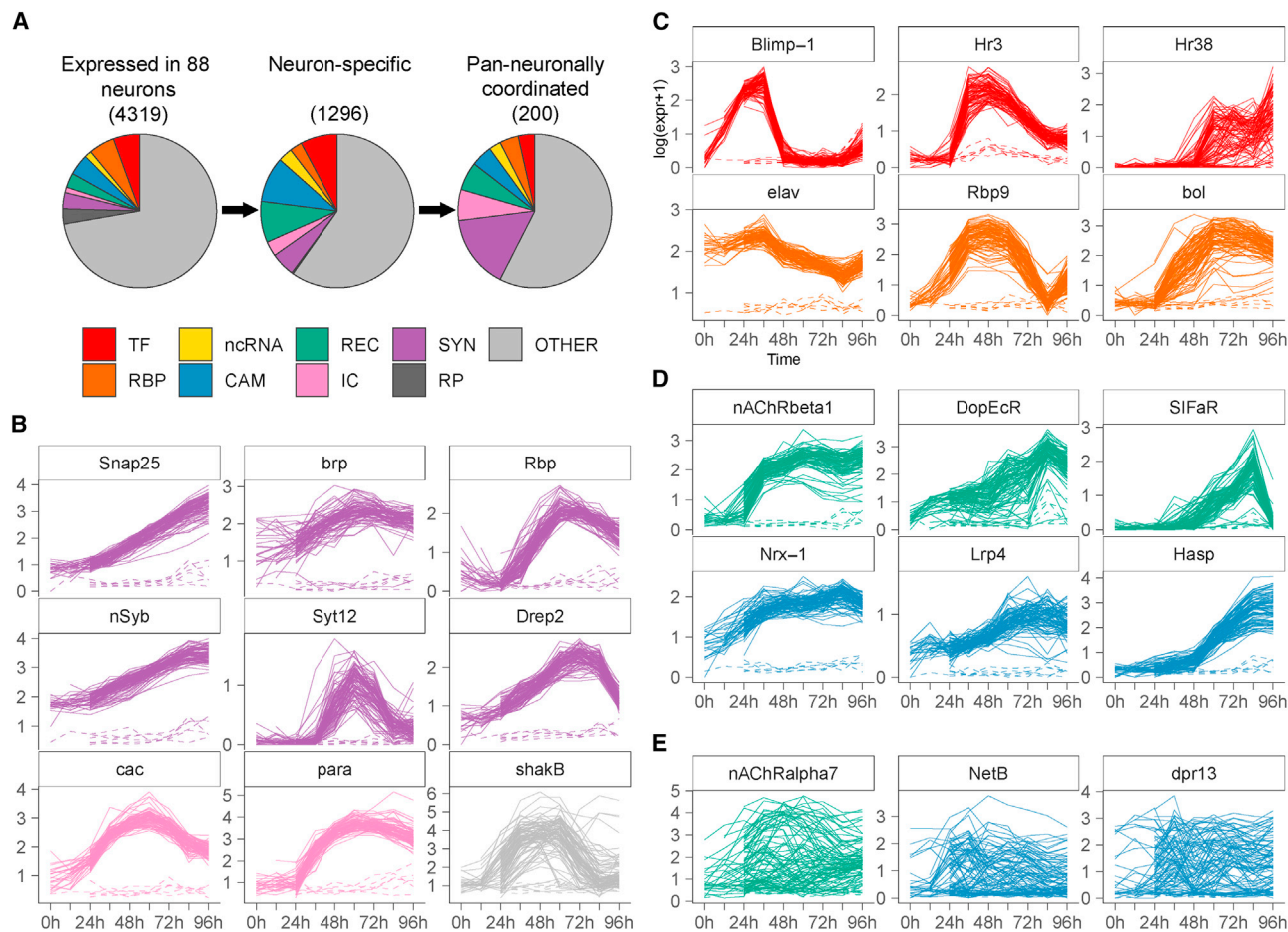
et al., 2015). Many cell types are present in a single copy in each column (unicolumnar neurons). In agreement with this, all matched uncolumnar neurons were among the largest 40–50 transcriptional clusters in our data, and each represented on average 0.7%–1.7% of the cells in each sample. We estimate that these 40–50 largest clusters represent cell types with roughly one copy per column (Figure 2B).

In parallel to the main atlas, we performed similar integrative analysis for early time points (i.e., 0 and 12 h APF; Figure S7). These samples were integrated together with samples from 24 h APF to generate a separate atlas covering the first day of pupal development (from 0 to 24 h APF). Since cells from 24 h APF were present in both parts of the atlas, we were able to match clusters between the two datasets. The differences between cell types were less prominent at earlier stages, and fewer cells were profiled. As a result, many related cell types were not resolved into separate clusters. For example, Tm1, Tm2, and Tm4 neurons and one unidentified cluster (N19) grouped together, suggesting a close developmental relationship between these cell types.

Transcriptional heterogeneity can reflect neuronal diversity beyond morphological cell types. For example, Dm8 neurons comprise two transcriptionally distinct subtypes with the same morphology but different synaptic specificities (Menon et al., 2019; Courgeon and Desplan, 2019). We found new subtypes for Dm3, Tm9, and T4 neurons. Tm9 neurons, for instance,

(C–E) Examples of cell-type-specific transcriptional data for Mi1 (red) and Mi4 neurons. Importantly, once we identified clusters through the integrative analysis, we returned to the original (pre-integration) normalized expression values for each time point and sample. (C) tSNE plots of the main dataset split by time points with Mi1 and Mi4 clusters highlighted. (D) Examples of cell-type-specific gene expression patterns. Stars indicate average expression levels in individual samples (i.e., time points and replicates; for instance, there are four individual replicates at 48 h). Lines indicate average expression for a given cell type. Expression values are shown for both the early (0–12 h) and main (24–96 h) datasets. Values are shown in log-scaled normalized expression levels. (E) Morphologies of Mi1 and Mi4 neurons (adapted from Fischbach and Dittrich, 1989).

See also Figures S3–S7.



**Figure 4. A Common (Pan-neuronal) Program of Neuronal Development**

(A) Distribution of genes in different categories expressed in neurons (left), enriched in neurons compared to glia (neuron-specific) (middle), and coordinated in a pan-neuronal fashion (right). Functional categories are color coded, including transcription factors (TFs), RNA-binding proteins (RBPs), non-coding RNA (ncRNA), cell adhesion molecules (CAMs), receptors and ligands (RECs), ion channels (ICs), synaptic proteins (SYNs), and ribosomal genes (RP).

(B–D) Expression patterns of selected pan-neuronally coordinated genes for 88 neurons (solid) and 6 glial (dashed) cell types from 24 to 96 h APF. For a subset of cell types, the expression is also shown at 0 and 12 h APF (i.e., clusters from the early dataset). Colors indicate functional categories of genes as in (A). Plots shown in log-scaled normalized expression levels. (B) Synaptic proteins and ICs. (C) Transcription factors (TFs) and RNA-binding proteins (RBPs). (D) Receptors and CAMs.

(E) For comparison, examples of broadly expressed genes with cell-type-specific expression patterns that are not coordinated are shown. Expression patterns of all 200 pan-neuronally coordinated genes ordered by average correlation values are shown in [Figures S8 and S9](#).

formed two clusters defined by mutually exclusive expression of *Wnt4* and *Wnt10* genes. *Wnt10* expression *in vivo* was restricted to dorsal Tm9 neurons ([Figure 3A](#)). We conclude that identified clusters represent ventral (*Wnt4*<sup>+</sup>) and dorsal (*Wnt10*<sup>+</sup>) subtypes of Tm9 neurons. We also noticed a stable subpopulation of T4 neurons expressing an uncharacterized gene (CG15537) throughout development that was enriched in T4 a/b subtypes. This group of cells did not form a separate cluster at the resolution of the analysis. The expression of CG15537, however, was restricted to T4 neurons targeting the dorsal third of the medulla and lobula plate layers a/b ([Figure 3B](#)). The product of CG15537 is predicted to be a cell surface protein with a hormone binding domain. Among the other genes specific to the same subpopulation of T4 neurons are a neuropeptide receptor (*TrissinR*) and *Tbh*, a key component of the octopamine biosynthesis pathway.

We hypothesize that these neurons are a regional T4 subtype with distinct neuromodulatory characteristics. In this way, we found and validated two cases of transcriptionally defined regional subtypes of neurons with unique spatial distributions. The discovery of new subtypes, even for such well-studied cell types as T4/T5 neurons, underscores the power of single-cell sequencing to reveal subtle aspects of cellular diversity.

Taken together, we generated a comprehensive transcriptional atlas of the developing *Drosophila* visual system. This atlas covers more than 150 transcriptionally distinct neuronal populations, including most of the known abundant cell types. The one-to-one correspondence between transcriptional clusters and known morphological cell types, the expression of known markers, and the match with the expected proportions of cells indicate that this is a highly reliable resource for following

individual cell types at multiple stages of development (Figure 2C). We tracked 88 neuronal cell types every 12 h of pupal development from 24 to 96 h APF (minimum of 10 cells in either DGRP replicate). This largely encompasses developmental times preceding synapse formation through the formation of the mature connectome. We also tracked 22 of these cell types in the early dataset (both at 0 and 12 h APF) covering all 100 h of pupal development. The cell-type-specific gene expression profiles were highly reproducible across replicates (Figure 2D). For an additional transcriptional analysis of the developing *Drosophila* visual system, see the contemporaneous study from the Desplan lab (Özel et al., 2020).

### Common and Cell-type-Specific Components of Neuronal Genetic Programs

#### A Common (Pan-neuronal) Program of Neuronal Development

All neurons share similar properties (e.g., synaptic transmission and membrane excitability). We sought to assess whether there was a common neuronal differentiation pathway giving rise to these features or whether they emerge in different ways in different neurons.

To do so, we followed 88 neuronal and 6 glial cell types from 24 to 96 h APF (see STAR Methods for details). First, we defined neuron-specific genes that were highly enriched in neurons compared to glia. Genes known to be involved in neuronal development and function were enriched among neuron-specific genes (Figure 4A).

Neuron-specific genes showed a wide range of temporal dynamics during development. We calculated Pearson's correlation coefficients between expression patterns of each gene across different cell types and averaged them using Fisher's Z transformation (Corey et al., 1998). Many neuron-specific genes were expressed in a highly coordinated fashion across different cell types. We hypothesize that they are part of a common genetic program unfolding in all neurons in parallel (a "pan-neuronal program"). Pan-neuronal genes were defined using the following criteria: (1) a gene was detected in at least half the neuronal cell types, and (2) the average Pearson's correlation coefficient was  $>0.75$  (see STAR Methods for details). In total, 200 genes met these criteria (Figure 4A). Expression patterns of these genes are shown in Figures 4, S8, and S9. The list of reported genes represents a subset of the top pan-neuronally genes with different levels of coordination. For instance, increasing the requirement for number of cell types in which gene was detected to 80% results in a slightly more conservative list of the genes (182 of 200 reported genes still pass this criterion).

The core components of synapses and membrane excitability were among the most tightly coordinated groups of genes (Figures 4, S8, and S9; Harris and Littleton, 2015). These include components of the synaptic release machinery (e.g., Snap25 and nSyb), the presynaptic active zone (e.g., Brp and Rbp), and many ion channels (e.g., Cac and Para). Some of the neurotransmitter receptor subunits were also expressed in a highly coordinated fashion in most neurons (e.g., nAChRbeta1 and Rdl), and others were cell-type specific, with diverse temporal dynamics (e.g., nAChRalpha7). A number of genes implicated in synaptic function in more specific contexts were also regulated

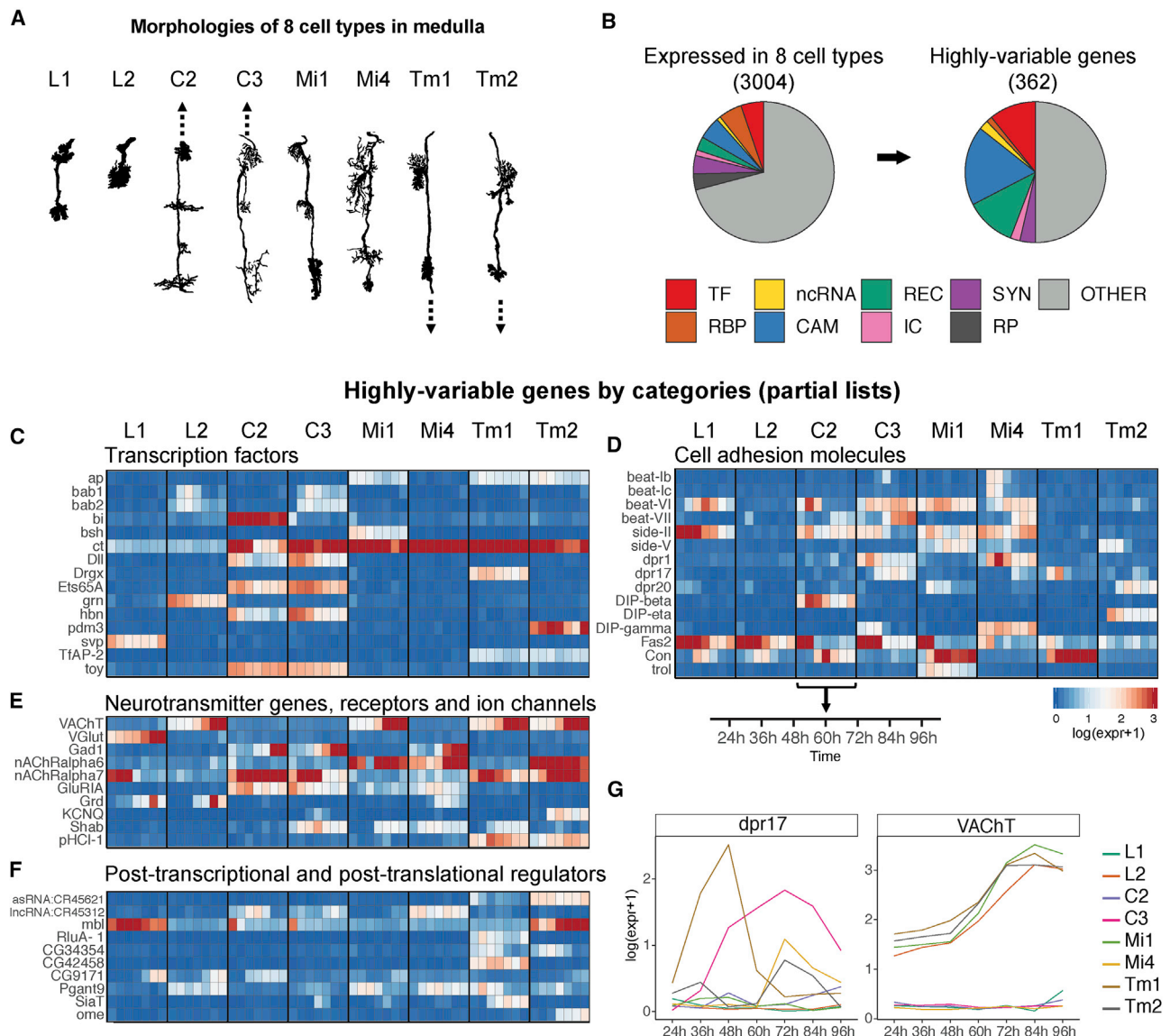
in a pan-neuronal fashion (e.g., Drep2; Andlauer et al., 2014). These included cell adhesion molecules (CAMs) that could contribute to synaptic adhesion complexes (Nrx-1, Hasp, and Lrp4; Li et al., 2007, Nakayama et al., 2016, Mosca et al., 2017). Other cell recognition molecules that contribute to connectivity in many contexts (e.g., Dscam1 and CadN) were also expressed pan-neuronally but were not specific to neurons or did not change appreciably during development. Other genes with similar expression patterns may represent uncharacterized constitutive components of synapses.

Most of the genes encoding presynaptic proteins were gradually upregulated during mid-pupal stages (36–48 h APF) and peaked  $\sim 60$  h APF. The timing and synchrony of this expression parallels the synchronous onset of patterned stimulus-independent neuronal activity (PSINA) in all neurons at  $\sim 48$  h APF (Akin et al., 2019). The gene encoding the gap junction protein ShabB is expressed pan-neuronally and peaks at mid-pupal stages (36–60 h APF). This would be consistent with a role for gap junctions at early stages of PSINA; indeed, attenuating ShabB activity depresses PSINA throughout the developing visual system (B. Bajar, S.L.Z., and O. Akin, unpublished data). Many other pan-neuronal genes were expressed transiently at particular stages of development. These patterns may pinpoint undescribed discrete steps in neuronal differentiation in all neurons (Figures 4, S8, and S9).

The synchrony of pan-neuronal genes suggests a common regulatory mechanism underlying their developmental timing. Several TFs exhibited temporally specific and pan-neuronally coordinated expression patterns (Figures 4, S8, and S9). Among these, Blimp-1 and Hr3 are controlled by EcR as part of the ecdysone regulated transcriptional cascade. Other genes in this cascade were also expressed in a tightly coordinated fashion in all neuron types, but they were not specific to neurons (data not shown). Another steroid hormone receptor, Hr38, was also expressed largely in a pan-neuronal fashion and was restricted to the second half of pupal development. Hr38 is an activity-regulated immediate early gene (Chen et al., 2016). It has also been shown to be indirectly activated by ecdysteroids, but independently of the EcR pathway (Baker et al., 2003). Other genes expressed in a pan-neuronal fashion also suggest a role for intercellular signaling in regulating the pan-neuronal program. These include two G-protein coupled receptors: DopEcR, a receptor for both ecdysone and dopamine (Srivastava et al., 2005), and SIFaR, a receptor for the conserved neuropeptide SIFamide (Jørgensen et al., 2006).

Pan-neuronal genes can also be regulated at post-transcriptional levels. A number of RNA-binding proteins (RBPs) and long non-coding RNA (lncRNA) genes were pan-neuronally coordinated (Figures 4, S8, and S9). For instance, Elav has been shown to function in neuron-specific splicing and lengthening of 3' untranslated regions (3' UTRs) of transcripts (Lisbin et al., 2001; Hilgers et al., 2012). Changes in 3' UTR length have been implicated in the regulation of RNA localization and translation, two prominent features of gene regulation in neurons (Tian and Manley, 2017). Two other Elav paralogs, Fne and Rbp9, implicated in synaptic growth at the larval neuromuscular junction (Zaharieva et al., 2015), are also expressed pan-neuronally, peaking at mid-pupal stages. RBPs involved in





**Figure 5. Cell-type-Specific Programs of Neuronal Development**

(A) Morphologies of eight neuronal cell types in the medulla (adapted from Fischbach and Dittrich, 1989). Arrows indicate direction of further extension of axons away from the cell body.

(B) Distribution of functional categories of 362 highly variable genes (HVGs; see STAR Methods) expressed in any one of these eight cell types. Functional categories are color coded as in Figure 4.

(C–F) Heatmaps with expression patterns of selected HVGs from the indicated functional categories. Expression patterns of other HVGs are shown in Figure S10.

(G) Examples of temporal patterns of cell-type-specific genes. dpr17 is expressed with different dynamics and expression levels in different cell types. VACHT is also expressed in cell-type-specific fashion, but temporal patterns and expression levels are coordinated across cell types. Expression patterns from seven time points at 12-h intervals between 24 and 96 h APF are shown. Values are shown in log-scaled normalized expression levels.

translational regulation, such as Bol, are also expressed pan-neuronally. Bol is an inhibitor of axon pruning in the mushroom body (Hooper et al., 2008) but may act more broadly to modulate expression of a core set of neuronal genes.

Taken together, these data suggest that the expression of pan-neuronal genes may be regulated at both the transcriptional and post-transcriptional levels and may, in addition, be coordinated by intercellular signaling.

**Cell-type-Specific Programs of Neuronal Development**

Neurons exhibit great diversity in their terminal morphologies, connectivity patterns, and physiological properties. We sought to characterize differences in transcriptomes that could underlie the development of these cell-type-specific features of neurons.

To explore these differences, we focused on eight cell types representing different morphological and functional classes of

neurons (Figure 5A). Some of them represent pairs of anatomically related neurons with different connectivity patterns (e.g., C2/C3 and Tm1/Tm2). We defined a set of 362 highly variable genes that were differentially expressed among these cell types (see STAR Methods for details). They represented 12% of all expressed genes. Three classes, TFs, CAMs, and receptors with ligands, accounted for 41% of highly variable genes (Figure 5B). TFs and CAMs were previously shown to be primary drivers of transcriptional diversity of neurons in different parts of the nervous system, both during development and in adults (Li et al., 2017a; Allen et al., 2020).

The expression patterns of TFs and CAMs were qualitatively different (Figures 5C and 5D). Most TFs were expressed in a binary fashion across cell types (i.e., on/off manner) and were expressed throughout development. Closely related cell types often expressed similar combinations of TFs (e.g., C2/C3 cells). Each neuron expresses discrete combinations of TFs that define their cell type. CAMs were more volatile in their expression patterns (Figure 5D). Many were expressed transiently during development, and their specificity could be switched between cell types at different time points. Moreover, quantitative differences in expression levels between neurons or in the same neuron at different developmental times were common. For example, *dpr17* was expressed at higher levels but with different temporal patterns by Tm1 (36–48 h APF) and C3 cells (48–96 h APF) and at lower levels and with similar temporal patterns by Tm2 and Mi4 cells (72–84 h APF). Thus, neurons express repertoires of CAMs with continuous variation in their expression levels. These diverse and rapidly changing cell recognition landscapes mirror the changes in morphology and patterns of synaptic connections elaborated by different neurons.

Genes encoding neurotransmitter-specific functions exhibited interesting temporal patterns of expression. For example, the vesicular acetylcholine transporter (VACHT) is expressed exclusively by cholinergic neurons. However, the temporal dynamics and expression levels of VACHT are tightly coordinated between different types of cholinergic neurons (Figure 5G). Thus, the choice of neurotransmitter is strictly cell-type specific, but the temporal patterns of expression are shared between different cell types.

Neurons also expressed genes involved in neuronal circuit functions in cell-type specific ways. These included genes encoding neurotransmitter biosynthetic enzymes, transporters, receptors, and ion channel subunits (Figure 5E). These genes can define functional diversity of neurons in adult circuits. For example, different combinations of neurotransmitter receptor subunits could result in distinct properties of synaptic connections. Interestingly, some of the genes that control synaptic communication and membrane excitability in the adult were also expressed during development. They may regulate intercellular communication underlying circuit assembly, most notably during the second half of pupal development in which each neuron type exhibits unique patterns of neuronal activity (i.e., PSINA).

Neuronal diversity is further shaped by differential expression of genes involved in post-transcriptional and post-translational mechanisms of gene regulation. We identified a number of cell-type specific RBPs and lncRNAs (Figure 5F). These genes

could contribute to cell identity at post-transcriptional levels. For example, expression of the cell-type-specific RBP muscleblind (Mbl) regulates alternative splicing of pan-neuronally expressed *Dscam2* to generate isoforms with different recognition specificities (Li and Millard, 2019). We also identified several genes involved in post-translational modification of proteins, including glycosylation and proteolytic enzymes, which are expressed in a cell-type-specific fashion (Figure 5F).

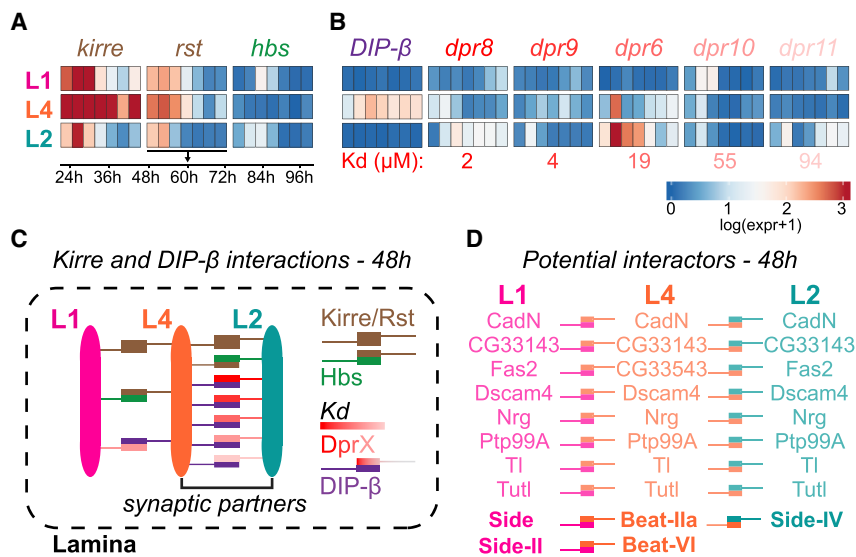
In summary, genetic programs of neuronal development comprise a common program of neuronal maturation regulating synapse formation and membrane excitability properties, overlaid by cell-type-specific programs encoding dramatically changing repertoires of cell surface proteins. These proteins regulate interactions between different neuron types controlling the assembly and specificity of connections between neurons.

### Approaching Synaptic Specificity through Transcriptomics of Synaptic Partners

Each neuron type in the fly visual system has a unique set of synaptic inputs and outputs and each of the different neuron types profiled expresses cell recognition proteins with different combinations, levels, and temporal profiles. This transcriptional atlas, in combination with the connectome, provides a resource to explore synaptic specificity mechanisms by combining it with developmental, biochemical, and genetic studies (Figure S11). Here, as an example of how this approach may be applied, we consider a simpler part of the connectome in the lamina and use a retrospective analysis to correlate expression data with previously described biochemical and genetic studies (Figure 6).

During early pupal development, L4 neurons extend dendrites that contact L1 and L2 axons. Later in pupal development, L4 selectively forms synapses with L2 (Rivera-Alba et al., 2011). Previous genetic studies have shown that three different families of cell recognition molecules of the immunoglobulin superfamily regulate this process. *Dscam2* and *Dscam4* are required for targeting of L4 dendrites to a bundled pair of L1 and L2 axons (Tadros et al., 2016). Kirre, a synaptic adhesion protein, is required for synapses between L4 and L2 (Lüthy et al., 2014), and DIP- $\beta$  promotes L2 specificity (Xu et al., 2019). Previous biochemical studies have reported the binding specificities of *Dscam2* and *Dscam4* (Millard et al., 2007; Özkan et al., 2013), the affinities of different Dprs for binding to DIP- $\beta$  (Cosmanescu et al., 2018), and the interaction of Kirre with its ligands (Özkan et al., 2013).

The temporal expression patterns of these genes in L4 are predictive of their function (Figures 6A–6C). *Dscam2* and *Dscam4* are expressed during targeting (see Figure S11). As their levels decrease, there is a concomitant increase in the expression of Kirre and DIP- $\beta$  just prior to the onset of synaptogenesis (Figures 6A and 6B). As the Kirre ligands (Kirre itself, Rst, and Hbs) (Özkan et al., 2013) are expressed in a similar way in L1 and L2, it is unlikely that Kirre alone determines specificity (Figure 6A). By contrast, the expression patterns of DIP- $\beta$  ligands (i.e., Dprs) and their affinities (Cosmanescu et al., 2018), however, would favor interactions between L4 and L2, consistent with the decrease in the fidelity of synaptic specificity seen in DIP- $\beta$  mutants (Xu et al., 2019) (Figures 6B and 6C). As both the Kirre and DIP- $\beta$  synaptic phenotypes are partial, other recognition



**Figure 6. Cell Surface Interactions between L4 and L1 or L4 and L2**

(A) Expression patterns of *kirre*, *rst*, and *hbs*.  
 (B) Expression patterns of DIP- $\beta$  and its Dpr binding partners from scRNA-seq data in this paper. Different Dpr proteins bind to DIP- $\beta$  with different affinities, as shown in red (Cosmanescu et al., 2018).  
 (C) Protein interactions regulating synapses between L4 and L2 are summarized. Kirre is required for synapses, but its ligands are not differentially expressed between L1 and L2. DIP- $\beta$  biases synapses to L2. This may be the result of differential expression of Dpr paralogs with different affinities between L1 and L2 (Xu et al., 2019). Interactions shown are based on genes expressed at 48 h (i.e., onset of synapse formation).  
 (D) Many other cell adhesion proteins of the immunoglobulin, leucine-rich repeat, and cadherin superfamilies are also expressed in L1, L2, and L4 at 48 h. Based on biochemical studies (Özkan et al., 2013), these proteins may bind to each other on the surface of opposing membranes as indicated. Values are shown in log-scaled normalized expression levels. For expression patterns of these proteins, see Figure S11. See text for further discussion.

molecules must act in a redundant fashion during synapse formation (Lüthy et al., 2014; Xu et al., 2019).

We sought to address whether the atlas provides candidates for these redundant functions. Our dataset is a rich resource for doing so, allowing us to follow the expression of cell surface proteins expressed by these neurons during synapse formation or, of course, in an analogous way for many other synaptic partners in the visual system (Figure S11). At the onset of synapse formation, eight pairs of recognition proteins (both heterophilic and homophilic) are expressed in both the synaptic partner L2/L4 pair and in the L1/L4 pair, which do not form synapses (Figure 6D). Several of these proteins are associated with synapse formation in flies and mice (e.g., N-cadherin; Schwabe et al., 2014), the L1 family protein Nrg (Ango et al., 2004; Tai et al., 2019), and the mammalian IgSF9b ortholog turtle (Woo et al., 2013), and these may act in a redundant way with Kirre to regulate synapse formation. As different Beat/Side protein family pairs (Siebert et al., 2009; Özkan et al., 2013; Li et al., 2017b) are differentially expressed in L1 and L2, they may work in parallel to DIP/Dpr interactions to determine specificity. These interactions may promote synapses between L2/L4, inhibit inappropriate L1/L4 synapses, or both. These data support the idea that multiple protein pairs on opposing membranes may contribute to synapse formation between these neurons, and it is the differences in the cell-type-specific expression of highly diversified protein families (i.e., DIPs/Dprs and Beats/Sides) that bias the specificity of synapse formation toward L2 (Xu et al., 2019).

We have observed a similar level of complexity in the potential interactions between cell surface proteins expressed by other neurons. For example, there are over a hundred neuron types that extend axons and dendrites into the medulla neuropil and they must discriminate between one another to select appropriate synaptic partners. Each neuron expresses many dozens of cell recognition molecules in a cell-type-specific and temporally dynamic fashion (Figure S11). These findings provide a mo-

lecular correlate for the modest phenotypes seen in many wiring mutants, alluding to marked redundancy. It raises the possibility that specificity may largely reflect biases in connectivity rather than all or none selectivity. That is, neurons may select synaptic partners not by a mechanism demanding an “exact match” between cell surface molecules but rather by selecting the “best match” among potential partners. This may confer robustness to the establishment of a complex stereotyped connectome. Unravelling the precise mechanisms by which this complexity in cellular recognition landscapes translates into highly specific patterns of synaptic connectivity will rely on future studies incorporating biochemical, genetic, and computational analyses.

### Concluding Remarks

Understanding how patterns of synaptic connections are determined during development remains a central question in neural development. In this study, we followed the development of 88 different neuron types at many time points spanning synaptogenesis. In so doing, we uncovered a pan-neuronal program running synchronously in all neurons regardless of cell type. This program was enriched, in particular, for synaptic proteins and proteins regulating electrically excitable membranes. The neuron-specific expression of pan-neuronal genes may be controlled by a common gene regulatory program. Alternatively, each cell type may regulate expression of these genes independently through a myriad of cell-type specific regulatory programs. The coordinated expression of these genes in all neurons may be controlled by a separate timing mechanism. Whatever the underlying mechanisms are for regulating this process in a similar way in all neurons, we speculate that this program endows neurons with a broad potential to form synapses.

By contrast to the pan-neuronal program of development, the expression of cell recognition molecules is profoundly asynchronous and highly cell-type specific. This diversity mirrors cell-type-specific morphological changes of axons and dendrites in

the target region and the complexity and specificity of synaptic connectivity each neuron elaborates. These rapid changes may be necessary to choreograph the specific interactions between the processes of many different neuron types as they seek to discriminate between appropriate and inappropriate synaptic partners and determine the number and distribution of synapses between them. Whether the dynamic expression pattern we observe is largely cell autonomous in nature or whether it reflects a dialog between neurons during circuit assembly is unknown. Many signaling pathway components, including those regulated by neural activity, growth factors, neuropeptides, and ecdysteroids, all of which can influence gene expression, are expressed in these neurons.

Traditionally, circuit assembly has been studied genetically one gene at a time. Molecules to study were selected based upon their patterns of expression or phenotypes of mutants. In these approaches, investigators were largely ignorant of other recognition molecules expressed in these neurons. Phenotypes in almost all cases were incomplete and often embarrassingly mild, supporting a general consensus in the field that this reflects substantial redundancy, which is advantageous for constructing a robust circuit but a bane to researchers. Transcriptome profiling revealed that each neuron expresses a unique combination of a vast number of cell recognition molecules, many of which have the potential to promote interactions between them. These findings and many genetic studies underscore the confound of redundancy in studying the logic and molecular mechanisms underlying specificity. In a broad sense we speculate that these proteins act together through common mechanisms and variations on them to sculpt circuitry. Uncovering the molecular logic of synaptic specificity will rely on combining transcriptomics with detailed studies of specific synapses using an increasingly sophisticated tool kit of genetic, biochemical, physiological, and imaging approaches.

The development of high-throughput single-cell technologies has led to an exponential increase in transcriptional studies in many different systems (Zheng et al., 2017). Profiling of multiple experimental conditions with internal controls for technical variation in a single experiment contributes to these advances (Kang et al., 2018). The pooled strategy, as we describe here, can be easily adapted for a variety of experimental designs, including comparative studies in different mutant backgrounds or in altered environments. In principle, similar approaches could be used in many other systems, where the genotype information could be obtained in parallel to transcriptional profiling. The rapid and economically feasible generation of temporally resolved transcriptional datasets can be used to tackle a variety of long-standing questions in developmental biology.

## STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- **KEY RESOURCES TABLE**
- **RESOURCE AVAILABILITY**
  - Lead Contact
  - Materials Availability

- Data and Code Availability
- **EXPERIMENTAL MODEL AND SUBJECT DETAILS**
  - Genetic design
  - *Drosophila* transgenic lines
- **METHOD DETAILS**
  - Droplet-based single-cell RNA-Seq
  - Tissue dissociation and single-cell suspensions
  - Immunohistochemistry / Immunofluorescence
  - Confocal microscopy and image analysis
- **QUANTIFICATION AND STATISTICAL ANALYSIS**
  - Raw data processing
  - Demultiplexing of DGRP samples
  - Quality control and filtering of single cells
  - Integrative analysis of the main dataset (24-96h APF)
  - Integrative analysis of the early dataset (0h-24h APF)
  - Matching transcriptional clusters to major classes of cell types (main dataset)
  - Matching transcriptional clusters to cell types (main dataset)
  - Matching transcriptional clusters to cell types (early dataset)
  - Cell-type-specific transcriptional profiles
  - Pan-neuronally coordinated genes
  - Highly variable genes
  - Functional categories of genes

## SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at <https://doi.org/10.1016/j.neuron.2020.10.006>.

## ACKNOWLEDGMENTS

We thank members of the Zipursky lab for critical reading and discussion of the manuscript. We thank the University of California, Los Angeles (UCLA) BSCRC Flow Cytometry core for assistance with fluorescence-activated cell sorting (FACS) purification of cells. We thank Michael Mashock and Xinmin Li from UCLA Technology Center for Genomics and Bioinformatics for assistance with scRNA-seq. We thank Hugo J. Bellen for transgenic fly lines. S.L.Z. is an Investigator of the Howard Hughes Medical Institute.

## AUTHOR CONTRIBUTIONS

Y.Z.K., J.Y., and S.L.Z. conceived the project and designed the experiments. Y.Z.K., J.Y., J.V.-A., and P.S. performed experiments. Y.Z.K. analyzed data. Y.Z.K., J.Y., J.V.-A., P.S., and S.L.Z. interpreted results and wrote the manuscript.

## DECLARATION OF INTERESTS

S.L.Z. is a member of the Neuron Advisory Board.

Received: July 10, 2020  
Revised: August 31, 2020  
Accepted: October 3, 2020  
Published: October 29, 2020

## REFERENCES

Akin, O., Bajar, B.T., Keles, M.F., Frye, M.A., and Zipursky, S.L. (2019). Cell-type-specific patterned stimulus-independent neuronal activity in the *Drosophila* visual system during synapse formation. *Neuron* 101, 894–904.e5.



- Allen, A.M., Neville, M.C., Birtles, S., Croset, V., Treiber, C.D., Waddell, S., and Goodwin, S.F. (2020). A single-cell transcriptomic atlas of the adult *Drosophila* ventral nerve cord. *eLife* 9, e54074.
- Andlauer, T.F.M., Scholz-Kornehl, S., Tian, R., Kirchner, M., Babikir, H.A., Depner, H., Loll, B., Quentin, C., Gupta, V.K., Holt, M.G., et al. (2014). Drep-2 is a novel synaptic protein important for learning and memory. *eLife* 3, e03895.
- Ango, F., di Cristo, G., Higashiyama, H., Bennett, V., Wu, P., and Huang, Z.J. (2004). Ankyrin-based subcellular gradient of neurofascin, an immunoglobulin family protein, directs GABAergic innervation at purkinje axon initial segment. *Cell* 119, 257–272.
- Baker, K.D., Shewchuk, L.M., Kozlova, T., Makishima, M., Hassell, A., Wisely, B., Caravella, J.A., Lambert, M.H., Reinking, J.L., Krause, H., et al. (2003). The *Drosophila* orphan nuclear receptor DHR38 mediates an atypical ecdysteroid signaling pathway. *Cell* 113, 731–742.
- Butler, A., Hoffman, P., Smibert, P., Papalexi, E., and Satija, R. (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* 36, 411–420.
- Chen, X., Rahman, R., Guo, F., and Rosbash, M. (2016). Genome-wide identification of neuronal activity-regulated genes in *Drosophila*. *eLife* 5, e19942.
- Corey, D.M., Dunlap, W.P., and Burke, M.J. (1998). Averaging correlations: expected values and bias in combined Pearson *r* s and Fisher's *z* transformations. *J. Gen. Psychol.* 125, 245–261.
- Cosmanescu, F., Katsamba, P.S., Sergeeva, A.P., Ahlsen, G., Patel, S.D., Brewer, J.J., Tan, L., Xu, S., Xiao, Q., Nagarkar-Jaiswal, S., et al. (2018). Neuron-subtype-specific expression, interaction affinities, and specificity determinants of DIP/Dpr cell recognition proteins. *Neuron* 100, 1385–1400.e6.
- Courgeon, M., and Desplan, C. (2019). Coordination between stochastic and deterministic specification in the *Drosophila* visual system. *Science* 366, eaay6727.
- Croset, V., Treiber, C.D., and Waddell, S. (2018). Cellular diversity in the *Drosophila* midbrain revealed by single-cell transcriptomics. *eLife* 7, e34550.
- Davie, K., Janssens, J., Koldere, D., De Waegeneer, M., Pech, U., Kreft, L., Aibar, S., Makhzami, S., Christiaens, V., Bravo González-Bias, C., et al. (2018). A single-cell transcriptome atlas of the aging *Drosophila* brain. *Cell* 174, 982–998.e20.
- Davis, F.P., Nern, A., Picard, S., Reiser, M.B., Rubin, G.M., Eddy, S.R., and Henry, G.L. (2020). A genetic, genomic, and computational resource for exploring neural circuit function. *eLife* 9, e50901.
- Dickson, B.J. (2002). Molecular mechanisms of axon guidance. *Science* 298, 1959–1964.
- Enriquez, J., Rio, L.Q., Blazeski, R., Bellemin, S., Godement, P., Mason, C., and Mann, R.S. (2018). Differing strategies despite shared lineages of motor neurons and glia to achieve robust development of an adult neuropil in *Drosophila*. *Neuron* 97, 538–554.e5.
- Fear, J.M., León-Novelo, L.G., Morse, A.M., Gerken, A.R., Van Lehmann, K., Tower, J., Nuzhdin, S.V., and McIntyre, L.M. (2016). Buffering of Genetic Regulatory Networks in *Drosophila melanogaster*. *Genetics* 203, 1177–1190.
- Fischbach, K.-F., and Dittrich, A.P.M. (1989). The optic lobe of *Drosophila melanogaster*. I. A Golgi analysis of wild-type structure. *Cell Tissue Res.* 258, 441–475.
- Harris, K.P., and Littleton, J.T. (2015). Transmission, development, and plasticity of synapses. *Genetics* 201, 345–375.
- Hasegawa, E., Kaido, M., Takayama, R., and Sato, M. (2013). Brain-specific homeobox is required for the specification of neuronal types in the *Drosophila* optic lobe. *Dev. Biol.* 377, 90–99.
- Hilgers, V., Lemke, S.B., and Levine, M. (2012). ELAV mediates 3' UTR extension in the *Drosophila* nervous system. *Genes Dev.* 26, 2259–2264.
- Hobert, O., and Kratsios, P. (2019). Neuronal identity control by terminal selectors in worms, flies, and chordates. *Curr. Opin. Neurobiol.* 56, 97–105.
- Holguera, I., and Desplan, C. (2018). Neuronal specification in space and time. *Science* 362, 176–180.
- Hoopfer, E.D., Penton, A., Watts, R.J., and Luo, L. (2008). Genomic analysis of *Drosophila* neuronal remodeling: a role for the RNA-binding protein Boule as a negative regulator of axon pruning. *J. Neurosci.* 28, 6092–6103.
- Huang, W., Massouras, A., Inoue, Y., Peiffer, J., Ràmia, M., Tarone, A.M., Turlapati, L., Zichner, T., Zhu, D., Lyman, R.F., et al. (2014). Natural variation in genome architecture among 205 *Drosophila melanogaster* Genetic Reference Panel lines. *Genome Res.* 24, 1193–1208.
- Jørgensen, L.M., Hauser, F., Cazzamali, G., Williamson, M., and Grimmelikhuijzen, C.J.P. (2006). Molecular identification of the first SIFamide receptor. *Biochem. Biophys. Res. Commun.* 340, 696–701.
- Kang, H.M., Subramaniam, M., Targ, S., Nguyen, M., Maliskova, L., McCarthy, E., Wan, E., Wong, S., Byrnes, L., Lanata, C.M., et al. (2018). Multiplexed droplet single-cell RNA-sequencing using natural genetic variation. *Nat. Biotechnol.* 36, 89–94.
- Konstantinides, N., Kapuralin, K., Fadil, C., Barboza, L., Satija, R., and Desplan, C. (2018). Phenotypic convergence: distinct transcription factors regulate common terminal features. *Cell* 174, 622–635.e13.
- Kurmangaliyev, Y.Z., Yoo, J., LoCascio, S.A., and Zipursky, S.L. (2019). Modular transcriptional programs separately define axon and dendrite connectivity. *eLife* 8, e50822.
- Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* 27, 2987–2993.
- Li, J.S.S., and Millard, S.S. (2019). Deterministic splicing of *Dscam2* is regulated by *Muscleblind*. *Sci. Adv.* 5, eaav1678.
- Li, J., Ashley, J., Budnik, V., and Bhat, M.A. (2007). Crucial role of *Drosophila* neurexin in proper active zone apposition to postsynaptic densities, synaptic growth, and synaptic transmission. *Neuron* 55, 741–755.
- Li, H., Horns, F., Wu, B., Xie, Q., Li, J., Li, T., Luginbuhl, D.J., Quake, S.R., and Luo, L. (2017a). Classifying *Drosophila* olfactory projection neuron subtypes by single-cell RNA sequencing. *Cell* 171, 1206–1220.e22.
- Li, H., Watson, A., Olechwiec, A., Anaya, M., Sorooshiyari, S.K., Harnett, D.P., Lee, H.P., Vielmetter, J., Fares, M.A., Garcia, K.C., et al. (2017b). Deconstruction of the beaten Path-Sidestep interaction network provides insights into neuromuscular system development. *eLife* 6, e28111.
- Lisbin, M.J., Qiu, J., and White, K. (2001). The neuron-specific RNA-binding protein ELAV regulates neuroglial alternative splicing in neurons and binds directly to its pre-mRNA. *Genes Dev.* 15, 2546–2561.
- Liu, J., Reggiani, J.D.S., Laboulaye, M.A., Pandey, S., Chen, B., Rubenstein, J.L.R., Krishnaswamy, A., and Sanes, J.R. (2018). *Tbr1* instructs laminar patterning of retinal ganglion cell dendrites. *Nat. Neurosci.* 21, 659–670.
- Lüthy, K., Ahrens, B., Rawal, S., Lu, Z., Tarnogorska, D., Meinertzhagen, I.A., and Fischbach, K.-F. (2014). The irre cell recognition module (IRM) protein Kirre is required to form the reciprocal synaptic network of L4 neurons in the *Drosophila* lamina. *J. Neurogenet.* 28, 291–301.
- Mackay, T.F.C., Richards, S., Stone, E.A., Barbadilla, A., Ayroles, J.F., Zhu, D., Casillas, S., Han, Y., Magwire, M.M., Cridland, J.M., et al. (2012). The *Drosophila melanogaster* Genetic Reference Panel. *Nature* 482, 173–178.
- Menon, K.P., Kulkarni, V., Takemura, S.-Y., Anaya, M., and Zinn, K. (2019). Interactions between Dpr11 and DIP- $\gamma$  control selection of amacrine neurons in *Drosophila* color vision circuits. *eLife* 8, e48935.
- Millard, S.S., Flanagan, J.J., Pappu, K.S., Wu, W., and Zipursky, S.L. (2007). *Dscam2* mediates axonal tiling in the *Drosophila* visual system. *Nature* 447, 720–724.
- Morey, M., Yee, S.K., Herman, T., Nern, A., Blanco, E., and Zipursky, S.L. (2008). Coordinate control of synaptic-layer specificity and rhodopsins in photoreceptor neurons. *Nature* 456, 795–799.
- Mosca, T.J., Luginbuhl, D.J., Wang, I.E., and Luo, L. (2017). Presynaptic LRP4 promotes synapse number and function of excitatory CNS neurons. *eLife* 6, e27347.
- Nakayama, M., Suzuki, E., Tsunoda, S., and Hama, C. (2016). The matrix proteins Hasp and Hig exhibit segregated distribution within synaptic clefts and play distinct roles in synaptogenesis. *J. Neurosci.* 36, 590–606.

- Özel, M.N., Simon, F., Jafari, S., Holguera, I., Chen, Y.C., Benhra, N., El-Danaf, R.N., Kapuralin, K., Malin, J.A., Konstantinides, N., and Desplan, C. (2020). Neuronal diversity and convergence in a visual system developmental atlas. *Nature*. In press. <https://doi.org/10.1038/s41586-020-2879-3>.
- Özkan, E., Carrillo, R.A., Eastman, C.L., Weiszmann, R., Waghray, D., Johnson, K.G., Zinn, K., Celniker, S.E., and Garcia, K.C. (2013). An extracellular interactome of immunoglobulin and LRR proteins reveals receptor-ligand networks. *Cell* 154, 228–239.
- Pei, J., Kinch, L.N., and Grishin, N.V. (2018). FlyXCDB: a resource for *Drosophila* cell surface and secreted proteins and their extracellular domains. *J. Mol. Biol.* 430 (18 Pt B), 3353–3411.
- Rivera-Alba, M., Vitaladevuni, S.N., Mishchenko, Y., Lu, Z., Takemura, S.Y., Scheffer, L., Meinertzhagen, I.A., Chklovskii, D.B., de Polavieja, G.G., and de Polavieja, G.G. (2011). Wiring economy and volume exclusion determine neuronal placement in the *Drosophila* brain. *Curr. Biol.* 21, 2000–2005.
- Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139–140.
- Sanes, J.R., and Zipursky, S.L. (2020). Synaptic specificity, recognition molecules, and assembly of neural circuits. *Cell* 181, 1434–1435.
- Sarin, S., Zuniga-Sanchez, E., Kurmangaliyev, Y.Z., Cousins, H., Patel, M., Hernandez, J., Zhang, K.X., Samuel, M.A., Morey, M., Sanes, J.R., and Zipursky, S.L. (2018). Role for Wnt signaling in retinal neuropil development: analysis via RNA-Seq and in vivo somatic CRISPR mutagenesis. *Neuron* 98, 109–126.e8.
- Schwabe, T., Borycz, J.A., Meinertzhagen, I.A., and Clandinin, T.R. (2014). Differential adhesion determines the organization of synaptic fascicles in the *Drosophila* visual system. *Curr. Biol.* 24, 1304–1313.
- Siebert, M., Banovic, D., Goellner, B., and Aberle, H. (2009). *Drosophila* motor axons recognize and follow a Sidestep-labeled substrate pathway to reach their target fields. *Genes Dev.* 23, 1052–1062.
- Srivastava, D.P., Yu, E.J., Kennedy, K., Chatwin, H., Reale, V., Hamon, M., Smith, T., and Evans, P.D. (2005). Rapid, nongenomic responses to ecdysteroids and catecholamines mediated by a novel *Drosophila* G-protein-coupled receptor. *J. Neurosci.* 25, 6145–6155.
- Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., 3rd, Hao, Y., Stoeckius, M., Smibert, P., and Satija, R. (2019). Comprehensive integration of single-cell data. *Cell* 177, 1888–1902.e21.
- Tadros, W., Xu, S., Akin, O., Yi, C.H., Shin, G.J., Millard, S.S., and Zipursky, S.L. (2016). Dscam proteins direct dendritic targeting through adhesion. *Neuron* 89, 480–493.
- Tai, Y., Gallo, N.B., Wang, M., Yu, J.-R., and Van Aelst, L. (2019). Axo-axonic innervation of neocortical pyramidal neurons by GABAergic chandelier cells requires AnkyrinG-associated L1CAM. *Neuron* 102, 358–372.e9.
- Takemura, S.Y., Bharioke, A., Lu, Z., Nern, A., Vitaladevuni, S., Rivlin, P.K., Katz, W.T., Olbris, D.J., Plaza, S.M., Winston, P., et al. (2013). A visual motion detection circuit suggested by *Drosophila* connectomics. *Nature* 500, 175–181.
- Takemura, S.Y., Xu, C.S., Lu, Z., Rivlin, P.K., Parag, T., Olbris, D.J., Plaza, S., Zhao, T., Katz, W.T., Umayam, L., et al. (2015). Synaptic circuits and their variations within different columns in the visual system of *Drosophila*. *Proc. Natl. Acad. Sci. USA* 112, 13711–13716.
- Tan, L., Zhang, K.X., Pecot, M.Y., Nagarkar-Jaiswal, S., Lee, P.T., Takemura, S.Y., McEwen, J.M., Nern, A., Xu, S., Tadros, W., et al. (2015). Ig superfamily ligand and receptor pairs expressed in synaptic partners in *Drosophila*. *Cell* 163, 1756–1769.
- Tessier-Lavigne, M., and Goodman, C.S. (1996). The molecular biology of axon guidance. *Science* 274, 1123–1133.
- Thurmond, J., Goodman, J.L., Strelets, V.B., Attrill, H., Gramates, L.S., Marygold, S.J., Matthews, B.B., Millburn, G., Antonazzo, G., Trovisco, V., et al.; FlyBase Consortium (2019). FlyBase 2.0: the next generation. *Nucleic Acids Res.* 47 (D1), D759–D765.
- Tian, B., and Manley, J.L. (2017). Alternative polyadenylation of mRNA precursors. *Nat. Rev. Mol. Cell Biol.* 18, 18–30.
- Woo, J., Kwon, S.-K., Nam, J., Choi, S., Takahashi, H., Krueger, D., Park, J., Lee, Y., Bae, J.Y., Lee, D., et al. (2013). The adhesion protein IgSF9b is coupled to neuroigin 2 via S-SCAM to promote inhibitory synapse development. *J. Cell Biol.* 201, 929–944.
- Xu, C., Theisen, E., Maloney, R., Peng, J., Santiago, I., Yapp, C., Werkhoven, Z., Rumbaut, E., Shum, B., Tarnogorska, D., et al. (2019). Control of synaptic specificity by establishing a relative preference for synaptic partners. *Neuron* 103, 865–877.e7.
- Zaharieva, E., Haussmann, I.U., Bräuer, U., and Soller, M. (2015). Concentration and localization of coexpressed ELAV/Hu proteins control specificity of mRNA processing. *Mol. Cell. Biol.* 35, 3104–3115.
- Zhao, H., Sun, Z., Wang, J., Huang, H., Kocher, J.-P., and Wang, L. (2014). CrossMap: a versatile tool for coordinate conversion between genome assemblies. *Bioinformatics* 30, 1006–1007.
- Zheng, G.X.Y., Terry, J.M., Belgrader, P., Ryvkin, P., Bent, Z.W., Wilson, R., Ziraldo, S.B., Wheeler, T.D., McDermott, G.P., Zhu, J., et al. (2017). Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* 8, 14049.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
<b>Antibodies</b>		
chicken anti-GFP (1:1000)	Abcam	Cat#13970; RRID:AB_300798
rabbit anti-dsRed (1:200)	Clontech	Cat#632496; RRID:AB_2737298
rat anti-NCad (1:40)	Developmental Studies Hybridoma Bank	Cat#DN-Ex#8; RRID:AB_528121
goat anti-chicken Alexa Fluor 488 (1:200)	Invitrogen	Cat#A11039; RRID:AB_142924
goat anti-rabbit Alexa Fluor 568 (1:500)	Invitrogen	Cat#A11011; RRID:AB_143157
goat anti-rat Alexa Fluor 647(1:500)	Invitrogen	Cat#A21247; RRID:AB_141778
<b>Chemicals, Peptides, and Recombinant Proteins</b>		
Schneider's Drosophila Medium	GIBCO	Cat#21720-024
PBS	Bioland Scientific LLC	Cat#PBS01-03
paraformaldehyde	Electron Microscopy Sciences	Cat#15710
Triton X-100	Sigma-Aldrich	Cat#T9284
Normal Goat Serum	Sigma-Aldrich	Cat#G6767
Papain	Worthington	Cat#LK003178
Liberase <sup>TM</sup> protease	Sigma-Aldrich	Cat# 5401119001
EverBrite mounting medium	Biotium	Cat#23001
<b>Deposited Data</b>		
Raw sequencing data and processed dataset	This paper	NCBI GEO: GSE156455
<i>Drosophila melanogaster</i> reference genome (dm6) and gene annotations (FlyBase, release 6.29)	Thurmond et al., 2019	<a href="https://flybase.org/">https://flybase.org/</a>
<i>Drosophila</i> Genetic Reference Panel genotypes (DGRP Freeze 2.0)	Mackay et al., 2012	<a href="http://dgrp2.gnets.ncsu.edu/">http://dgrp2.gnets.ncsu.edu/</a>
Reference bulk RNA-Seq dataset 1	Davis et al., 2020	NCBI GEO: GSE116969
Reference bulk RNA-Seq dataset 2	Konstantinides et al., 2018	NCBI GEO: GSE103772
<i>Drosophila melanogaster</i> extracellular domain database (FlyXCDB)	Pei et al., 2018	<a href="http://prodata.swmed.edu/FlyXCDB">http://prodata.swmed.edu/FlyXCDB</a>
<b>Experimental Models: Organisms/Strains</b>		
<i>D. melanogaster</i> : w[1118]	Bloomington Drosophila Stock Center	BDSC:5909; RRID:BDSC_5909
<i>D. melanogaster</i> : 23G12-LexA (T4/T5)	Bloomington Drosophila Stock Center	BDSC:65044; RRID:BDSC_65044
<i>D. melanogaster</i> :24C08-LexA (Tm9)	Bloomington Drosophila Stock Center	BDSC:62012; RRID:BDSC_62012
<i>D. melanogaster</i> : Wnt10-TG4	Gift from Hugo Bellen	CR02107
<i>D. melanogaster</i> : CG15537-TG4	Gift from Hugo Bellen	CR02107
<i>D. melanogaster</i> : LexAop-myr::tdTomato	Zipursky lab	N/A
<i>D. melanogaster</i> : 10XUAS-myr::GFP	Bloomington Drosophila Stock Center	BDSC:32197; RRID:BDSC_32197
<i>D. melanogaster</i> : DGRP-21	Bloomington Drosophila Stock Center	BDSC:28122; RRID:BDSC_28122
<i>D. melanogaster</i> : DGRP-28	Bloomington Drosophila Stock Center	BDSC:28124; RRID:BDSC_28124
<i>D. melanogaster</i> : DGRP-40	Bloomington Drosophila Stock Center	BDSC:29651; RRID:BDSC_29651
<i>D. melanogaster</i> : DGRP-129	Bloomington Drosophila Stock Center	BDSC:28141; RRID:BDSC_28141
<i>D. melanogaster</i> : DGRP-177	Bloomington Drosophila Stock Center	BDSC:28150; RRID:BDSC_28150
<i>D. melanogaster</i> : DGRP-181	Bloomington Drosophila Stock Center	BDSC:28151; RRID:BDSC_28151
<i>D. melanogaster</i> : DGRP-189	Bloomington Drosophila Stock Center	BDSC:28152; RRID:BDSC_28152
<i>D. melanogaster</i> : DGRP-235	Bloomington Drosophila Stock Center	BDSC:28275; RRID:BDSC_28275

(Continued on next page)

**Continued**

REAGENT or RESOURCE	SOURCE	IDENTIFIER
D. melanogaster: DGRP-304	Bloomington Drosophila Stock Center	BDSC:25177; RRID:BDSC_25177
D. melanogaster: DGRP-307	Bloomington Drosophila Stock Center	BDSC:25179; RRID:BDSC_25179
D. melanogaster: DGRP-320	Bloomington Drosophila Stock Center	BDSC:29654; RRID:BDSC_29654
D. melanogaster: DGRP-324	Bloomington Drosophila Stock Center	BDSC:25182; RRID:BDSC_25182
D. melanogaster: DGRP-348	Bloomington Drosophila Stock Center	BDSC:55019; RRID:BDSC_55019
D. melanogaster: DGRP-354	Bloomington Drosophila Stock Center	BDSC:55020; RRID:BDSC_55020
D. melanogaster: DGRP-355	Bloomington Drosophila Stock Center	BDSC:55038; RRID:BDSC_55038
D. melanogaster: DGRP-374	Bloomington Drosophila Stock Center	BDSC:28185; RRID:BDSC_28185
D. melanogaster: DGRP-382	Bloomington Drosophila Stock Center	BDSC:28189; RRID:BDSC_28189
D. melanogaster: DGRP-383	Bloomington Drosophila Stock Center	BDSC:28190; RRID:BDSC_28190
D. melanogaster: DGRP-391	Bloomington Drosophila Stock Center	BDSC:25191; RRID:BDSC_25191
D. melanogaster: DGRP-395	Bloomington Drosophila Stock Center	BDSC:55022; RRID:BDSC_55022
D. melanogaster: DGRP-406	Bloomington Drosophila Stock Center	BDSC:29657; RRID:BDSC_29657
D. melanogaster: DGRP-437	Bloomington Drosophila Stock Center	BDSC:25194; RRID:BDSC_25194
D. melanogaster: DGRP-441	Bloomington Drosophila Stock Center	BDSC:28198; RRID:BDSC_28198
D. melanogaster: DGRP-461	Bloomington Drosophila Stock Center	BDSC:28200; RRID:BDSC_28200
D. melanogaster: DGRP-492	Bloomington Drosophila Stock Center	BDSC:28203; RRID:BDSC_28203
D. melanogaster: DGRP-505	Bloomington Drosophila Stock Center	BDSC:55024; RRID:BDSC_55024
D. melanogaster: DGRP-508	Bloomington Drosophila Stock Center	BDSC:28205; RRID:BDSC_28205
D. melanogaster: DGRP-589	Bloomington Drosophila Stock Center	BDSC:28213; RRID:BDSC_28213
D. melanogaster: DGRP-748	Bloomington Drosophila Stock Center	BDSC:28224; RRID:BDSC_28224
D. melanogaster: DGRP-790	Bloomington Drosophila Stock Center	BDSC:28232; RRID:BDSC_28232
D. melanogaster: DGRP-805	Bloomington Drosophila Stock Center	BDSC:28237; RRID:BDSC_28237
D. melanogaster: DGRP-810	Bloomington Drosophila Stock Center	BDSC:28239; RRID:BDSC_28239
D. melanogaster: DGRP-819	Bloomington Drosophila Stock Center	BDSC:28242; RRID:BDSC_28242
D. melanogaster: DGRP-850	Bloomington Drosophila Stock Center	BDSC:28249; RRID:BDSC_28249
D. melanogaster: DGRP-897	Bloomington Drosophila Stock Center	BDSC:28260; RRID:BDSC_28260

**Software and Algorithms**

Cell Ranger (3.1.0)	10X Genomics	<a href="https://www.10xgenomics.com/">https://www.10xgenomics.com/</a> ; RRID:SCR_017344
CrossMap (v0.3.6)	<a href="#">Zhao et al., 2014</a>	<a href="http://crossmap.sourceforge.net/">http://crossmap.sourceforge.net/</a> ; RRID:SCR_001173
samtools/bcftools (1.8)	<a href="#">Li, 2011</a>	<a href="http://samtools.github.io/">http://samtools.github.io/</a> ; RRID:SCR_005227
demuxlet (version 2)	<a href="#">Kang et al., 2018</a>	<a href="https://github.com/statgen/popscl">https://github.com/statgen/popscl</a>
Seurat (3.1.2)	<a href="#">Stuart et al., 2019</a>	<a href="https://github.com/satijalab/seurat/">https://github.com/satijalab/seurat/</a> ; RRID:SCR_016341
edgeR (3.26.8)	<a href="#">Robinson et al., 2010</a>	<a href="https://bioconductor.org/">https://bioconductor.org/</a> ; RRID:SCR_012802
Fiji (ImageJ)	ImageJ	<a href="https://imagej.nih.gov/ij/">https://imagej.nih.gov/ij/</a> ; RRID:SCR_002285

**RESOURCE AVAILABILITY**

**Lead Contact**

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, S. Lawrence Zipursky ([LZipursky@mednet.ucla.edu](mailto:LZipursky@mednet.ucla.edu)).

**Materials Availability**

This study did not generate new unique reagents.



### Data and Code Availability

The accession number for the raw sequencing data and the final processed dataset (the atlas) reported in this paper is NCBI GEO: GSE156455. The atlas includes metadata for individual single cells, single-cell gene expression matrix and tSNE embeddings. Metadata for single cells includes sample identities (dataset, time point, replicate, genotype) and cluster identities (i.e., clusters, classes, types, subtypes).

## EXPERIMENTAL MODEL AND SUBJECT DETAILS

### Genetic design

An overview of the general workflow of single-cell sequencing experiments is shown in [Figure S1](#). Flies were reared at 25°C on standard medium. For experiments with conventional design (W1118), w[1118] (BDSC #5905) female white pre-pupae (0h APF) were collected and reared until dissection at target time points (5-6 animals per sample). For experiments with pooled design (DGRP), virgin females w[1118] (BDSC #5905) were crossed to males of 35 isogenic wild-type strains from the *Drosophila* Reference Genetic Panel ([Mackay et al., 2012](#); [Huang et al., 2014](#)). F1-generation female white pre-pupae were collected and reared until dissection at target time points. Animals for pooled samples with mixtures of target time points were staged with 12h intervals and were dissected and processed simultaneously. For every target time point, we used three individual animals with unique genotypes (total 27 animals per sample). For two replicates of DGRP experiments, genotypes used for each time point were reshuffled. All genotypes used in the study are shown in [Figure S1](#).

### *Drosophila* transgenic lines

The following transgenic lines were used for validation of regional subtypes of Tm9 and T4/T5 neurons: 24C08-LexA (Tm9, BDSC #62012), 23G12-LexA (T4/T5, BDSC #65044), Wnt10-TG4 (gift from Hugo Bellen, CR01661), CG15537-TG4 (gift from Hugo Bellen, CR02107), LexAop-myr::tdTomato (Zipursky laboratory), 10XUAS-myr::GFP (BDSC #32197).

## METHOD DETAILS

### Droplet-based single-cell RNA-Seq

Single-cell suspensions were processed using the 10X Genomics Chromium 3' v3 platform. For W1118 experiments, 4 lanes of a Chromium Chip were loaded per each sample; for DGRP experiments, 8 lanes of a Chromium Chip were loaded per each sample (i.e., replicates A and B). Loading volumes were estimated based on cell concentrations to capture around 8,000 single cells per one lane. Single-cell RNA-Seq libraries were generated using the manufacturer's protocol, with 12 cycles of PCR for cDNA amplification. All RNA-Seq libraries were sequenced using 4 lanes of Illumina NovaSeq 6000 S4 platform (28bp + 150bp).

### Tissue dissociation and single-cell suspensions

Optic lobes were collected in a single Eppendorf tube per sample. Brain tissue was incubated in papain (Worthington #LK003178) and Liberase protease (Sigma-Aldrich #5401119001) cocktail at 25°C for 15 min. Tissue was gently washed twice with PBS, then washed with 0.04% BSA in PBS and dissociated mechanically by pipetting. Cell suspension was filtered through a 20 µm cell-strainer (Corning #352235), stained with DRAQ5 (abcam #ab108410), and sorted by FACS (BD FACSAria II) to isolate single-cells and measure cell concentrations.

### Immunohistochemistry / Immunofluorescence

Brain dissections and immunostaining were performed as described in ([Kurmangaliyev et al., 2019](#)). Brains were dissected in ice-cold Schneider's *Drosophila* Medium (GIBCO #21720-024), and fixed in PBS (Bioland Scientific LLC #PBS01-03) containing 4% paraformaldehyde (Electron Microscopy Sciences, Cat#15710) for 25 min at room temperature (RT). Brains were rinsed repeatedly with PBST (PBS containing 0.5% Triton X-100 (Sigma #T9284)), and incubated in blocking solution (PBST containing 10% Normal Goat Serum (Sigma #G6767)) for at least 1 hr at RT prior to incubation with anti-body. Brains were incubated sequentially with primary and secondary antibodies diluted in blocking solution overnight at 4°C, with at least 2 PBST rinses followed by 2 hr incubations at RT in between and afterward. Brains were transferred to 50% (for 30 min), then 100% EverBrite mounting medium (Biotium #23001) and mounted on slides for confocal microscopy.

Primary antibodies and dilutions used in this study: chicken anti-GFP (Abcam #13970, 1:1000), rabbit anti-dsRed (Clontech #632496, 1:200), rat anti-NCad (Developmental Studies Hybridoma Bank (DSHB) DN-Ex#8, 1:40). Secondary antibodies and dilutions used in this study were as follows: goat anti-chicken Alexa Fluor 488 (AF488) (Invitrogen #A11039, 1:200), goat anti-rabbit AF568 (Invitrogen #A11011, 1:200), goat anti-rat AF647 (Invitrogen #A21247, 1:500).

Immunofluorescence images were acquired using a Zeiss LSM 880 confocal microscope with Zen digital imaging software. Optical sections or maximum intensity projections were level-adjusted, cropped and exported for presentation using ImageJ software (Fiji). Reported expression patterns were reproducible across three or more biological samples.

### Confocal microscopy and image analysis

Immunofluorescence images were acquired using a Zeiss LSM 880 confocal microscope with Zen digital imaging software. Optical sections or maximum intensity projections were level-adjusted, cropped and exported for presentation using ImageJ software (Fiji). Reported expression patterns were reproducible across three or more biological samples.

## QUANTIFICATION AND STATISTICAL ANALYSIS

### Raw data processing

Fastq files with raw reads were processed using Cell Ranger (3.1.0) with default parameters. Reference genome and transcriptome were based on FlyBase (release 6.29, [Thurmond et al., 2019](#)).

### Demultiplexing of DGRP samples

Single-cell transcriptomes from DGRP samples were demultiplexed based on parental genotypes using demuxlet with default parameters (version 2, <https://github.com/statgen/popsicle>; [Kang et al., 2018](#)). The genotypes of wild-type strains were downloaded from the DGRP web site (<http://dgrp2.gnets.ncsu.edu/>; [Mackay et al., 2012](#); [Huang et al., 2014](#)). Demultiplexing was based on genotypes of 35 DGRP strains that were used in experiments (each sample included 27 of them and the remaining genotypes were used as negative controls). The coordinates of genomic variants for DGRP strains were updated from dm3 to dm6 version of the *Drosophila melanogaster* reference genome using CrossMap ([Zhao et al., 2014](#)). The exonic variants for w[1118] (BDSC #5905) were called based on BAM files produced by Cell Ranger for W1118 samples. Variants were called using bcftools pipeline with default parameters (version 1.8; [Li, 2011](#)). The genotypes of F1 heterozygotes were assembled based on parental genotypes. The genomic variants used for demultiplexing were filtered using following criteria: (1) only biallelic single-nucleotide polymorphisms (SNPs) with maximum minor allele count of 8 among analyzed DGRP strains; (2) SNPs had to be represented by a reference genome allele in w[1118] strain (min.depth > 10). In total, 176,636 SNPs were used for demultiplexing. Only 37 of 246,308 cell barcodes (before filtering) were erroneously assigned to the genotypes that were not used in the given samples (negative controls), indicating high accuracy of recovered sample identities of single cells.

### Quality control and filtering of single cells

The initial set of cell barcodes called by Cell Ranger were filtered based on following criteria: (1) number of transcripts from 2,000 to 20,000; (2) maximum 10% of mitochondrial transcripts; (3) for DGRP samples, we removed predicted doublets and cells assigned to wrong genotypes; (4) we also removed cells with more than 3 transcripts aligned to male-specific lncRNA:roX1 and lncRNA:roX2 genes. Only one of the samples had a considerable number of potential male cells suggesting contamination by a male pupa (W1118, 48h APF, replicate A, 15% of cells). In total, 208,976 cells passed all criteria and were used for further analysis ([Figure S1](#)).

### Integrative analysis of the main dataset (24-96h APF)

All steps of single-cell data analysis were performed using methods implemented in Seurat V3 (version 3.1.2; [Butler et al., 2018](#); [Stuart et al., 2019](#)). Samples from 24h to 96h APF were analyzed together. Raw gene counts were normalized by total number of transcripts per cell and log-transformed (function `NormalizeData`). Integration was performed on 3000 highly variable genes selected across all samples (function `SelectIntegrationFeatures`). We excluded from this set mitochondrial genes, ribosomal protein genes, genes encoding heat-shock proteins and genes from oxidative phosphorylation complexes (FlyBase gene groups). We performed integration of samples using “reference-based” implementation of CCA (canonical correlation analysis) and MNN (mutual nearest neighbors) based workflow. The integration was performed on the levels of individual time points and replicates (e.g., DGRP, replicate A, 96h). W1118 samples from 48h (replicate A) and 72h APF were used as reference datasets, and dimensionality of the dataset was set to 200 (functions `FindIntegrationAnchors` and `IntegrateData`). Integrated dataset was scaled and used for principal component analysis (PCA). The first 200 PCs were used for graph-based clustering (functions `FindNeighbors` and `FindClusters`, `k.param = 50`, `resolution = 10`). The same set of PCs was used to generate t-distributed stochastic neighbor embedding (tSNE) plots, and to generate a dendrogram of clusters (function `BuildClusterTree`).

### Integrative analysis of the early dataset (0h-24h APF)

DGRP samples from early time points were integrated separately. The analysis was performed similar to the main dataset. We used DGRP 24h samples as reference datasets. All parameters of the analysis were the same as those used for the main dataset.

### Matching transcriptional clusters to major classes of cell types (main dataset)

The first three PCs of integrated dataset separated the major classes of cell types: PC1 separated neurons (*elav+*) from other populations; PC2 separated two groups of non-neuronal clusters from each other (*repo+* glia from *repo-* cells); and PC3 separated photoreceptor cells (*chp+*) from the remaining cells. We computed centroids of each cluster in the space of these 3 PCs and applied K-means clustering. This allowed us to group transcriptional clusters into four major classes of cell types ([Figure S4](#)).

### Matching transcriptional clusters to cell types (main dataset)

Neuronal clusters were matched to known morphological cell types using two bulk RNA-Seq datasets from adult animals (Davis et al., 2020; Konstantinides et al., 2018). For Davis et al. (2020) normalized expression profiles per cell type were downloaded from NCBI GEO: GSE116969 (dataTable4). For Konstantinides et al. (2018), expression matrix with raw gene counts was downloaded from NCBI GEO: GSE103772. Flybase gene IDs were matched to gene symbols, and gene counts were normalized using edgeR (Robinson et al., 2010). For both datasets we focused on reference profiles for individual optic lobe neurons; Figures S5A and S5B).

For single-cell transcriptomes, we used cells from the W1118/96h sample. The analysis was performed on original expression values (i.e., uncorrected expression matrix). We computed the average expression profile for each cell type. Next, we identified genes enriched in each neuronal cluster in the same sample. Cluster-enriched genes were identified using Wilcoxon rank-sum test (function FindMarkers, fold-change > 2, adjusted  $p < 0.01$ ). Each cluster was compared to the background set of neurons with equal proportions of cells from each neuronal cluster (max. 30 cells from each cluster). The union of top 20 enriched genes for each cluster was used for correlation analysis (top marker genes, total 826 genes).

Pearson's correlation coefficients were computed between log-transformed reference expression profiles and the average expression profile of each transcriptional cluster for top marker genes (Figures S5A and S5B). Almost all reference datasets had best mutual match with a single transcriptional cluster (only exclusion was LC16 cells from Davis et al., 2020). Matching of cell types that were present in both reference datasets were also concordant. In a few cases multiple clusters matched with a single reference profile with similar correlation values. In these cases, reference cell types included multiple distinct transcriptional subtypes (both known and newly identified, see main text for details). The final set of matched cell types was manually curated based on expression of known marker genes (Figure S5C). In particular, we curated known subtypes of T4/T5 neurons, newly identified subtypes for Dm3 and Tm9 neurons, and we also inferred identity of T2a cluster based on similarity to the related and matched T2 and T3 clusters.

We used two levels of cluster identities for few cell types: cell types and cell subtypes. Subtypes were defined for T4/T5 neurons, Tm9 and Dm3 neurons. All further analyses were performed on the level of cell types (e.g., all T4/T5 neurons are considered as a single cell type). We also annotated two large clusters of photoreceptors that corresponded to R1-6 and R7/R8 cell types. The remaining cluster of R cells represented only 0.3% of photoreceptors and likely represents cells undergoing apoptosis (defined by expression of *grim* and *rpr*). In total, 58 neuronal and 2 photoreceptor clusters were matched to known cell types and subtypes in the main dataset.

### Matching transcriptional clusters to cell types (early dataset)

The early dataset included cells from DGRP/24h samples. These cells were also analyzed in the main dataset and have assigned cluster and cell-type identities. We used these shared cells to directly transfer cluster identities between two datasets (on the level of cell types, see above). Clusters from the early and the main datasets were considered matching if the majority of 24h cells in both clusters were the same (minimum 90% cells in each direction). In this way, we were able to match and transfer identities for 58 of 196 transcriptional clusters from the main to the early dataset (Figure S7). In few cases, matching of closely related clusters were curated manually (i.e., T4/T5 neurons and photoreceptors). Also note that some of the matched clusters in the early datasets largely consisted of cells from 24h APF (i.e., was not detected at 0h and 12h APF). 22 neuronal clusters that were matched to the main dataset were represented by at least 10 cells at both 0h and 12h APF. Some other clusters included cells only from 0h and 12h APF. These clusters may represent cells from other parts of the brain captured due to differences in tissue dissections at different pupal stages, or represent transient progenitor populations.

### Cell-type-specific transcriptional profiles

The average transcriptional profiles were computed for each cluster at given time point in given sample/replicate (e.g., Mi1 in DGRP, 24h APF, replicate A, see Figure 2D). Averaging was performed in non-log space for the original normalized expression values (i.e., uncorrected expression matrix). The downstream analyses of cell-type-specific profiles were focused on DGRP samples (i.e., analyses of pan-neuronally coordinated and highly variable genes). In particular, we focused on 88 neuronal and 6 glial clusters that could be followed at every time point from 24h to 96h APF (minimum 10 cells in one of the replicates). Expression values for two DGRP replicates were further averaged to obtain summarized cell-type-specific transcriptional profiles. For visualizations of cell-type-specific expression patterns we used log<sub>1p</sub>-transformed expression values (i.e.,  $\log(x+1)$ ). In heatmaps, we capped the maximum expression values to 20.

### Pan-neuronally coordinated genes

The analysis was based on 88 neuronal and 6 glial cell types in DGRP samples from 24h to 96h APF (see above). First, we defined neuron-specific genes by comparing average expression in neurons and glia (fold-change > 8). For each gene, enrichment was estimated at a time point in which it was expressed in the largest number of neurons, or (in case of ties) with highest average expression. Next, we computed the average correlation between expression patterns of neuron-specific genes across different neurons. Pearson's correlation coefficients were averaged using Fisher's Z transformation (Corey et al., 1998). Genes expressed in at least half of the neurons (min. expression 1,  $n > 44$ ) and with average Pearson's correlation coefficient higher than 0.75 were defined as pan-neuronally coordinated genes. The expression levels at earlier time points (for cell types that could be followed in the early dataset) were used only for visualizations.

### Highly variable genes

The analysis was focused on 8 neuronal cell types from 24h to 96h APF (Figure 5). We determined genes with highly variable expression patterns across these cell types. Coefficients of variation (CV) were calculated for all genes that were expressed in any of these cell types (min. expression 1). For each gene, CV was calculated at a time point in which it was expressed in the largest number of neurons, or (in case of ties) with highest average expression. To take into account the relationship between variability and average expression, CV values were grouped into 10 bins based on similarity of mean expression and transformed into Z-scores within each of these bins (scaled CV). We used a cut-off of 1.5 for scaled CV values to define the set of top highly variable genes.

### Functional categories of genes

We defined several functional categories of genes. Some genes could be classified into more than one category. Thus, classification was performed sequentially, and every gene was assigned to only one category (e.g., gene classified as “ion channel” cannot be classified as “synaptic genes,” see below). Gene classifications were based on gene groups (GG) and gene ontology (GO) annotations obtained from FlyBase (release 6.29). Annotations for cell adhesion molecules were obtained from FlyXCDB (<http://prodata.swmed.edu/FlyXCDB>; Pei et al., 2018). Functional categories were defined in following order: (1) ribosomal proteins (from GG); (2) transcription factors (from GG); (3) RNA binding proteins (from GO, term GO:0003723; excluding ribosomal proteins (GG), translation factors (GG) and tRNA genes (GG)); (4) non-coding RNA genes (based on gene annotations); (5) cell adhesion molecules (from FlyXCDB, protein domains: Ig, EGF, LRR, fn3, Cadherin); (6) receptor and ligands (from GG, groups: “transmembrane receptors,” “receptor ligands”). (7) ion channels (from GG); (8) synaptic genes (from GO, GO:0007268).