Comparative single-cell transcriptomics of complete insect nervous systems

B.T. Cocanougher[1,2†], J.D. Wittenbach[1†], X. Long[1†], A.B. Kohn[3], T.P. Norekian[3,4,5], J. Yan[1], J. Colonell[1], J.-B. Masson[7], J.W. Truman[1], A. Cardona[1,6], S. Turaga[1], R.H. Singer[1,8‡], L.L. Moroz[3,4‡*], M. Zlatic[1,2‡*]

Affiliations:

[1]Howard Hughes Medical Institute Janelia Research Campus, Ashburn, Virginia, USA.

[2]Department of Zoology, Cambridge University, Cambridge, UK.

[3]Whitney Laboratory for Marine Bioscience, University of Florida, St. Augustine, Florida, USA.

[4]Friday Harbor Laboratories, University of Washington, Friday Harbor, Washington, USA

[5]Institute of Higher Nervous Activity and Neurophysiology, Russian Academy of Sciences, Moscow, Russia

[6]Department of Physiology, Development, and Neuroscience, Cambridge University, Cambridge, UK.

[7]Decision and Bayesian Computation, CNRS UMR 3571, Neuroscience and Computational Biology Department, Institut Pasteur, Paris, France

[8]Department of Anatomy and Structural Biology, Albert Einstein College of Medicine, Bronx, New York, USA

[†]These authors contributed equally to this work.

[‡]These authors contributed equally to this work.

[*]Corresponding author. Email: moroz@whitney.ufl.edu (L.L.M.); zlaticm@janelia.hhmi.org (M.Z.)

Abstract:

Molecular profiles of neurons influence information processing, but bridging the gap between genes, circuits and behavior has been very difficult. Furthermore, the behavioral state of an animal continuously changes across development and as a result of sensory experience. How behavioral state influences molecular cell state is poorly understood. Here we present a complete atlas of the *Drosophila* larval nervous system. We develop *polyseq*, a python analysis package, and use single molecule RNA-FISH to validate our scRNAseq findings. To investigate how internal state affects cell state, we altered internal state with high-throughput behaviour protocols designed to mimic wasp sting and over activation of the memory system. We found nervous system-wide gene expression changes related to cell state. This work advances our understanding of how genes, neurons, and circuits generate behavior.

One Sentence Summary:

Single cell genomics and modulation of internal state reveal how genes, neurons, and circuits produce behavior.

Main Text:

Making sense of any complex system involves identifying constituent elements and understanding their individual functions and interactions. Neural circuits are no exception. While recent advances in connectomics (*1-9*) and live imaging techniques (*10-14*) offer unprecedented information about neural connectivity and activity, the task of identifying cell types has traditionally relied on painstaking morphological, functional, or single gene histochemical taxonomy. High-throughput single-cell RNA sequencing (scRNAseq) offers a new way forward by providing a molecular-level identity for each cell via its transcriptomic profile. Importantly, it is also scalable to populations of millions of cells without incurring exorbitant costs. These techniques have already revealed striking heterogeneity in cell populations that is lost in bulk samples. In the fruit fly, efforts are already well underway to produce connectomic (*4-7,9*), activity (*12-14*), and behavior atlases (*16-17*) of the nervous system. Much work has separately revealed the role that genes (*18-19*) and circuits (*20-21*) play in behavior; a major challenge is to combine genes, circuits, and behavior all at once. A comprehensive transcriptomic atlas is the missing piece to the connectivity, activity, and behavior maps that would create the required resource necessary to understand the complex interplay between genes, circuits, and behavior.

To this end, we developed a protocol to capture, sequence, and transcriptionally classify the molecular cell types and cell state of the entire central nervous system of the *Drosophila* larva, the first example of a full central nervous system transcriptomic atlas in an insect. Furthermore, we successfully performed cell capture experiments on single animals for the first time. Given that the *Drosophila* larva has a nervous system of approximately 10,000 neurons, our atlas of 202,107 cells has ~20X coverage of each neuron in the nervous system. Previously identified cell types were recognizable in our atlas, including motor neurons, Kenyon cells of the mushroom body, insulin producing cells, brain dopaminergic and serotonergic cells, and all glial subtypes. While scRNAseq provides nearly complete information about the transcriptional program being used by a cell at the time of collection, a drawback to the technique is a loss of spatial information. We therefore used a recently developed RNA fluorescent *in situ* hybridization (RNA-FISH) protocol to resolve the anatomical location of molecular cell types in the whole larval brain (*22*). We combined RNA-FISH with high-resolution Bessel beam structured illumination microscopy to detect and count mRNA within newly identified cells. This technique

provides ground truth for the absolute number of a particular RNA in a given molecular cell type at a particular time point. It also provides an opportunity to assess the quantitative capability of our scRNAseq approach.

Larval behavior after hatching is dominated by feeding; when a critical weight is achieved, this behavior switches to wandering in preparation for pupation (*23*). Endocrine and neuroendocrine pathways responsible for this switch have been well characterized (*24*), but the extent of molecular changes in defined cell types across the nervous system that respond to this neuroendocrine signaling are not known. To investigate such nervous system-wide changes during development, we sequenced the nervous system at four time points in development.

Previous sensory experience alters the behavioral state of an animal. Flies that are hungry form food-associated memories more easily (*25*) and flies that are intoxicated court more frequently (*26*). Male flies that lose fights are more likely to subsequently exhibit submissive behaviors and to lose second contests while male flies that win exhibit aggressive behavior and are more likely to win later fights (*27*). Are internal states controlled transcriptionally at the level of identified cell types and circuits and, if so, how? It is an open question whether memory or internal state will affect gene expression globally or only in restricted cell populations.

In order to discover the nervous-system-wide gene expression changes induced by previous experience, we examined gene expression profiles from the nervous systems of animals exposed to two experimental protocols. The first protocol involved presenting repeated pain and fear, by modeling repeated wasp sting. Fictive stings were induced using optogenetic activation of a small population of well described interneurons (*5*). Of note, no mechanical damage to the animal's surface occurred with this protocol. The second protocol involved repeated activation of higher-order central brain neurons involved in learning. Using behavioral assays before and after the stimulation we showed that each of these protocols cause a long-lasting change in the animals' behavioral state. We then analyzed the effect of fictive sting and repeated activation of the learning center to search for changes in gene expression related to cell state. We consider these "cell state" genes and find that both cell populations and individual neuron types can exhibit cell state changes.

Taken together, these results suggest the powerful role that transcriptomic atlases can play in probing the complex interplay between cell state, circuit function, and behavior.

Results:

A complete transcriptomic atlas of 202,107 single cells from the larval central nervous system was built (Fig. 1). Nervous systems were captured at four time points in development (1 hour, 24 hours, 48 hours, and 96 hours after larval hatching) and for three nervous system dissections (full CNS, brain only, and ventral nerve cord only). Four non-neuronal tissues, including ring gland cells, hemocytes, imaginal disc cells, and salivary gland cells, were also captured. The larval CNS has ~10,000 neurons, thus the 202,107 cells collected here represents ~20X coverage of the nervous system (i.e. assuming equal capture efficiency, each cell in the nervous system was collected around 20 times).

We discovered that transcriptomic cell types split into four neuronal cell classes: (1) larval functional neurons, (2) adult developing neurons, (3) neural stem cells, and (4) glia. Sixty-three transcriptomic cell types were identifiable among these four groups at the top level of the clustering (Fig. 2). These 63 cell types were further subclustered into 495 transcriptomic cell subtypes (Table S1).

Larval functional neurons participate in the circuits which control sensation and behavior. At larval hatching, embryonic neurons are born and all the neurons necessary for larval life are functional (*28*). These neurons will continue to grow and some populations, such as the Kenyon cells of the mushroom body, will add more neurons throughout development. Identifiable cells at the top level include motor neurons, Kenyon cells, and excitatory and inhibitory interneurons. These cells can be separated into brain and VNC neurons using a machine learning classifier (Fig. S1). Most larval functional neurons are present at hatching, but adult developing neurons continue to grow in number over the course of larval development. As a result, the absolute number of larval functional neurons in our atlas increases slightly over development, but the relative number of larval neurons to adult developing neurons decreases over time (Fig. 3). We

discovered multiple markers that exhibit increased expression in larval functional neurons, including the protein tyrosine phosphatase *IA-2*.

Adult developing neurons form the second class of neurons. Larvae spend much of their life feeding and growing. From initial hatching to pupation, larvae grow 1,000 times their initial length (Fig. 2) (*28*). During this growth period, the larval nervous system grows and adds developing adult neurons which remain quiescent during larval life but grow and elaborate their axonal and dendritic arbors during pupation into adult functional neurons (*29*). In our atlas, we can identify adult developing neurons through high expression neuronal markers (nSyb, elav) and a lack of synaptic and neurotransmitter genes (*VChAT, VGlut*). We also report here previously undescribed markers of this population, including *grh* and the non-coding RNA *CR31386* (Fig. 2).

Neural stem cells, including neuroblasts and ganglion mother cells (GMCs), formed the third major cell class. In the ventral nerve cord (VNC), neuroblasts divide repeatedly via asymmetric cell division to generate one renewed neuroblast and one GMC (*30*). Each GMC then divides terminally to form two neurons. This pattern forms lineages throughout the VNC which have well characterized morphology and identified transcription factor expression. A recent report shows that neurotransmitter identity is lineage-restricted (*31*). We also find specific stem cell specific expression of multiple new markers, including *jim* (Fig. 3).

Glia represent the fourth major class of cells. Surface glia provide a smooth, protective surface around the nervous system which can be appreciated in the scanning electron micrograph (Fig. 1). Glia split into 8 molecular cell types (Fig. S2). These include 6 previously described anatomical types: astrocyte-like glia, cell body glia, ensheathing glia, wrapping glia, perineural, and subperineural glia (*32*) and 2 additional molecular subtypes of these types. We also identified glial precursors in the ganglion mother cell clusters. We find *CG5955*, which codes for an L-threonine 3-dehydrogenase, is found specifically in glia.

polyseq **defines cell types**

We developed *polyseq* (github.com/jwittenbach/polyseq), an open source Python package, to perform cell clustering. *polyseq* performs functions of many popular R packages (such as Seurat) (*33*) with significantly improved runtime (Fig. S3) and the extensibility and modularity of Python. *polyseq* starts with a gene by cell matrix, which is then filtered for high quality cells, normalized, regressed, reduced, and clustered. Visualization can then be performed with t-SNE or umap for the full dataset. The software also includes inbuilt functionality for violin plots and heatmaps. The data remain in a form that is easy to integrate with the vast community of Python packages for further visualization and analysis.

We used *polyseq* to analyze a dataset from an individual animal. Each individual animal has a unique life experience. These experiences shape the epigenome through chromatin remodeling, and this change in chromatin accessibility affects gene expression (*34*). Given these within-animal differences, measuring cells from a single animal would be particularly powerful for reducing variability. All bulk RNAseq and scRNAseq work in Drosophila to date has been done in mixtures of large populations of animals (*35-37*). Here we were able to capture nearly half of all cells from the nervous system of a single animal.

The single nervous system atlas reported here consists of 3,421 cells collected in one sample from a single third instar larva (Fig. S4). After preprocessing, 3,363 cells remained for further analysis. Graph-based clustering revealed 10 clusters which were visualized with a t-distributed stochastic neighbor embedding (t-SNE) plot (Fig. 3C). The t-SNE plot is a two-dimensional visualization of high-dimensional data in which each cell is represented by a single dot and each cell is placed as close as it can be to cells with similar transcriptomic profiles. Finally, for each cluster, a linear classifier is trained on the transcriptomic profiles to classify neurons from that cluster versus all other neurons in the high dimensional space. The direction of the classifier's hyperplane in "gene space" can then be used to rank all genes in order of importance for distinguishing neurons in that cluster from all others – a simple yet surprisingly effective method of differential expression analysis.

Clusters 0, 1 and 2 contain adult developing neurons and larval functional neurons. Cluster 3 contains known markers for ganglion mother cells (GMCs) and neuroblasts (*insb*, *E(spl)m8-*

HLH) and lacks markers for neurotransmitters or receptors (Coumailleau and Schweisguth, 2014). Cluster 4 contains cells with known glial markers (*picot, fabp, santa-maria, apolpp, wrapper*) and cluster 6 contains cells with known astrocytic markers (*alrm, vir-1*). Subsetting these astrocytes and glia by known glial subtype markers revealed further cell type refinement. In addition to known markers, many putative markers have expression specific to glial and astrocytic subtypes. Cluster 10 contains markers for tracheal cells (*sog*) and lacks neuronal markers (*nSyb*) and thus represents a class of non-neuronal cells.

Two clusters with neuronal markers were separated from the central cloud of neurons. Cluster 7 contains genes known to be enriched in motor neurons (*twit, VGlut, Proc, cac*) and cluster 8 contains genes specific to Kenyon cells (*Dop1R2*) and other brain neurons. Further analysis of the expression of neurotransmitters, neuropeptides, and receptors in the motor neuron cluster confirms previous markers (*38-39*).

Monoaminergic neurons play a key role in learning in the fly (*40-41*). A single top-level cluster was identified with expression of key monoaminergic synthetic enzymes (*Trh, ple*) and transporters (*DAT, SerT*). Subclustering of this top-level cluster revealed three strong groups, corresponding to serotonergic, dopaminergic, and octopaminergic clusters, identifying previously undescribed markers of these populations of cells which separate one monoamine type from another (Fig. S5).

JhI-21, a solute carrier 7-family amino acid transporter, is a novel marker found here in 5-HT neurons. This gene encodes for a protein necessary for protein nutrition signaling, which was recently described (*42-44*). However, these reports describe the importance of the JhI-21 protein peripherally, with no mechanism for the transmitting of nutritional information to the nervous system. Here we see that the protein is expressed in the serotonin neuron itself, suggesting that serotonin neurons act directly as sensors for amino acid nutritional state.

Validating transcriptomic predictions

Insulin producing cells (IPCs) consist of fourteen neurons in the larval brain (Fig. 4A) (*45*). These participate in circuits which monitor the nutritional status of the larva and function as the

larval equivalent of the mammalian pancreas. If IPCs are ablated, larvae and adults are smaller and have a diabetic phenotype, including increased hemolymph trehalose and glucose levels (*46*). IPCs secrete insulin-like peptides which regulate hemolymph sugar levels. Graph-based clustering revealed a cluster defined by strong expression of insulin-like peptide 2 (*Ilp2*) and insulin-like peptide 5 (*Ilp5*) expression, which are canonical markers of IPCs (Fig. 4B). By subsetting the data to look only at 96 cells in the putative IPCs cluster, the neurotransmitters and receptors expressed by these could be analyzed (Fig. 4C).

Previous reports show IPCs are regulated by canonical neurotransmitters. This includes modulation by serotonin through the 5-HT1A receptor and octopamine through the Octbeta1 receptor (*47*) and by the neuropeptide allatostatin A (*48*). In our atlas we also see strong expression of receptors for dopamine (*Dop2R*), glutamate (*GluClalpha*), and Allatostatin C Receptor 2(*AstC-R2*) in IPCs (Fig. 4C).

To validate the specificity of our scRNAseq approach for identifying AstC-R2 in IPC cells, we probed AstC-R2 mRNA in a HaloTag reporter line for the IPCs. The overlap between the neurons containing the HaloTag and FISH signals confirmed the sequencing result (Fig. 4D). The colocalization of AstC-R2 with 14 IPCs suggests that all IPCs are regulated by AstC through AstC-R2. The discovery of regulation by AstC-R2 updates our model of the regulation of IPCs by adding an additional population of cells that are modulating IPC activity.

Correlating smRNA-FISH and scRNAseq

In order to determine whether scRNAseq could quantitatively capture the dynamics of expression in a single cell type, we compared scRNSeq expression to ground truth expression levels determined using single molecule RNA-FISH (smRNA-FISH). We compared the RNA levels in a small population of cells discovered in our atlas that contain expression of vesicular glutamate transporter (VGlut) and the neuropeptide Allatostatin C (AstC). We quantified the relative expression of these mRNAs using smFISH and compared the result with scRNAseq (Fig. 5).

We used smFISH to probe VGlut and AstC mRNAs simultaneously and obtain quantitative expression levels. We detected 5 groups of cells that contain AstC FISH signals, which was consistent with previous reports of AstC localization (*49*). We observed 5 pairs of cells that contained both VGlut and AstC. Among the 5 pairs, 1 pair belonged to previously reported SLP1 AstC cells (Fig. 5A). We quantified the VGlut and AstC mRNAs in these cells using a Bessel beam selective plane illumination microscope (BB-SIM) (*22*). Although quantification of VGlut and AstC mRNAs within individual cell bodies could not be obtained due to the difficulty of segmenting overlapping cell bodies, we were able to obtain an average quantification of VGlut and AstC mRNAs within these 5 pair of cells (Fig. 5B). The similarity we obtained for the VGlut and AstC expression ratio between single-molecule FISH and scRNAseq suggested that the relative quantification from scRNAseq was compatible with single molecule FISH (Fig. 5C).

Machine learning predicts spatial origin

Machine learning was used to predict the spatial origin of neurons and glia. By dissecting out and separately sequencing brain and VNC cells, ground truth labels were available for further analysis. We trained a neural network using this labeled data to recognize brain or VNC neurons with > 99% accuracy on held out data. We then used this classifier to predict the spatial origin of cells from samples of single animal origin with no brain or VNC label (Fig. S1). The ratio of brain to VNC capture was close to expected (expected ~80% VNC, observed 84% VNC). The pattern of single gene known VNC markers agreed with classifier predictions (Fig. S1). In the single animal sample, graph-based clustering revealed distinct brain and VNC clusters. Further analysis revealed that these clusters correspond strongly to motor neurons (VNC cluster) and Kenyon cells and cholinergic brain neurons (Kenyon cell cluster).

Optogenetic sting alters expression globally

To investigate if a change in internal state would alter nervous-system-wide gene expression, we examined gene expression profiles from animals exposed to repeated fictive sting. An optogenetic sting was induced by activation of the basin interneurons. The basins are first order interneurons that receive input from nociceptive (pain) and mechanosensory (vibration) sensory neurons (*5*). Such optogenetic activation of the brain evokes a rolling escape response (*5*), which

mimics the natural response to wasp sting or nociceptor activation (*50*). Of note, no mechanical damage was induced in our protocol.

The basin interneurons were activated for 15 seconds with a 45 second rest period for a total of 120 activation periods (Fig. 6A). Supervised machine learning was used to automatically detect behavior (*51*). A rolling escape response was observed at the start of the experiment; by the final activation stimulus, backing up and turning were the predominate responses (Fig. 6B).

Control animals of the same age were collected from the same food plate as experimental animals and placed on an agar plate in the dark for two hours. Immediately following the sting protocol, 2-4 animals from each group underwent the scRNAseq protocol. We were primarily interested in searching for cell state genes which could drive cell state clustering (Fig. 6D). If cells from experimental animals and controls are analyzed together, will cells cluster together based on cell type (independent of treatment group) or cell state (dependent on treatment group)? If cell type clustering is observed, it suggests that any changes in cell state induced by our protocol are minor compared to cell type specific features. But if cell state clustering is observed, it is evidence that experience driven changes are at least of comparable importance to cell type in determining genetic cell state.

The optogenetic sting protocol led to cell state clustering. Transcriptomic data from cells isolated from activated and control brains were normalized and analyzed in the same mathematical space, but the clustering that was observed was based on cell state (i.e. clustering was driven by whether the cells came from a “stung” animal or an “unstung” control). Cell state genes that differed between the stung and unstung controls were discovered (Fig. 6F).

Cell state genes were most evident in larval functional and developing neurons, including motor neurons, cholinergic, and neuropeptidergic cells. Genes that were upregulated in motor neurons following the sting protocol included non-coding RNA (*CR40469*), carbohydrate metabolic enzymes (lactate dehydrogenase, *ImpL3)*, and the ethanol-induced apoptosis effector, *Drat*.

In addition to cell state genes within the nervous system, a large group of immune cells was sequenced in the sting state, with particularly high expression of neuropeptide-like precursor 2 (*NPLP2*), a gene which has been observed in phagocytic immune cells (*52*). RNA-FISH in sting and control conditions revealed this transcript is not higher in the neurons but suggested that the lymph gland is being activated and ramping up cell numbers in the sting condition. Strong evolutionary selection pressure exists on the larva to survive predation by parasitic wasps (*53*). Larvae can survive by using their immune system to encapsulate and prevent the hatching of internalized parasitic wasp eggs. Previous work has investigated the role of signaling detected by mechanical damage of the sting and the presence of the foreign body (wasp egg) inside the larva (*54*). Our transcriptomic data suggest the immune system may also respond to a currently unknown signal generated directly by the nervous system.

Learning center overactivation alters expression locally

In a second behavior protocol, we activated all higher-order central brain neurons involved in learning and memory, called Kenyon cells (KCs). Similar to the fictive sting, we observed a change in behavioral response at the start and end of training. At the start of training, animals hunch and arrest movement at the onset of activation and crawl forward at the offset of activation (Fig. 6C). At the end of training, animals continue to hunch and stop at onset, but a larger fraction (~80%) perform a small motion before turning rather than crawling forward to offset. Also this protocol, not only altered animals response to the optogenetic activation of KCs, but also drastically altered behavior after activation. Animals greatly increased the probability of stopping and reduced the probability of crawling.

To discover potential molecular changes that could drive these behavioral changes, we analyzed the transcriptomes of animals exposed to these optogenetic training protocols and compared them to controls. Unlike the global changes in gene expression following optogenetic sting, we detected changes in the transcriptomic state of many fewer cell types following repeated activation of the higher-order brain neurons involved in learning and memory (Fig. 6G-I). We discovered a number of interesting candidate genes that were upregulated in an activity-dependent way in Kenyon cells and dopaminergic neurons, which are key cell populations in the learning and memory center.

Cell state genes with differential expression between KC activated brains and controls separated local groups of cells within clusters. Changes were observed in KCs and dopaminergic neurons (DANs) (Fig. 6H,I). Cell state genes were not limited to previously described activity related genes. They included long non-coding RNA (*noe*, *CR40469*), chromatin remodeling (*His2AV*, *mamo*), axon guidance (*trn*, *DIP1*, *fas*, *dpr14*, *fz2*), and receptor genes (*Dop2R*).

Discussion:

This work makes several contributions to the field. First, we present the first full transcriptomic atlas of an insect central nervous system. Second, we use super resolution microscopy to compare single molecule RNA-FISH with scRNAseq in the *Drosophila* larva. By combining these two techniques – the first providing information about the complete collection of RNA present in a cell and the second providing full anatomical, subcellular, and absolute quantification of a chosen RNA(s) – we provide a resource for the field of *Drosophila* neurobiology and provide an example of complementary methods for building and validating single cell transcriptomic atlases. Third, we provide an experimental paradigm for discovering a molecular signature of internal state and use this paradigm to uncover drastic gene-expression changes that accompany a state of stress evoked by repeated "optogenetic" predator attack. Our atlas is therefore a powerful resource for developmental, neuroscience, and evolutionary biology.

Separating cell type from cell state is a key challenge for transcriptomic cell atlases. In order to understand the changes in a specific cell type between health and disease, for example, it will be necessary to be able to find the same cell type among differing conditions. Here we show that such a distinction can be discovered in cases where molecular cell state is significantly altered. Furthermore, we show that even in a nervous system with drastically different (and unnatural) activity patterns in a sizable population of highly interconnected neurons (here 200 of 10,000 or 2% of the nervous system), limited changes may be observed across the entire nervous system but changes can be observed in specific cell types. We foresee such techniques being useful to investigate a wide range of internal state and cell state changes, from sleep to parasitism to circadian rhythms.

Single cell transcriptomic atlases are the missing piece required for the combined analysis of genes, circuits, and behavior. Our work here shows that transcriptomic atlases can be reliably built for multiple developmental stages of the *Drosophila* larva. Furthermore, we show that optogenetic manipulations of internal state can alter gene expression in a context dependent manner. By adding a transcriptomic atlas to the existing atlases of neuron connectivity, neuron activity, and behavior, we have set the stage for a more complete understanding of the principles of that underlie the complex interplay of genes, circuits, and behavior.

References and notes:

1. White JG, Southgate E, Thomson JN, Brenner S: The structure of the nervous system of the nematode Caenorhabditis elegans. Philos Trans R Soc Lond B Biol Sci 1986, 314:1-340.

2. Jarrell TA, Wang Y, Bloniarz AE, Brittin CA, Xu M, Thomson JN, Albertson DG, Hall DH, Emmons SW: The connectome of a decision-making neural network. Science 2012, 337:437-444.

3. Helmstaedter M, Briggman KL, Turaga SC, Jain V, Seung HS, Denk W: Connectomic reconstruction of the inner plexiform layer in the mouse retina. Nature 2013, 500:168-174.

4. Takemura SY, Bharioke A, Lu Z, Nern A, Vitaladevuni S, Rivlin PK, Katz WT, Olbris DJ, Plaza SM, Winston P et al.: A visual motion detection circuit suggested by Drosophila connectomics. Nature 2013, 500:175-181.

5. Ohyama T, Schneider-Mizell CM, Fetter RD, Aleman JV, Franconville R, Rivera-Alba M, Mensh BD, Branson KM, Simpson JH, Truman JW et al.: A multilevel multimodal circuit enhances action selection in Drosophila. Nature 2015, 520: 633-639.

6. Berck ME, Khandelwal A, Claus L, Hernandez-Nunez L, Si G, Tabone CJ, Li F, Truman JW, Fetter RD, Louis M et al.: The wiring diagram of a glomerular olfactory system. Elife 2016, 5.

7. Eichler, K, et al. "The complete connectome of a learning and memory centre in an insect brain." Nature 548.7666 (2017): 175.

8. Hildebrand DGC, Cicconet M, Torres RM, Choi W, Quan TM, Moon J, Wetzel AW, Scott Champion A, Graham BJ, Randlett O et al.: Whole-brain serial-section electron microscopy in larval zebrafish. Nature 2017, 545:345-349.

9. Eschbach, C., Fushiki, A., Winding, M., Schneider-Mizell, C., Shao, M., Arruda, R., ... & Zlatic, M. (2019). Multilevel feedback architecture for adaptive regulation of learning in the insect brain. *bioRxiv*, 649731.

10. Ahrens, M. B., Orger, M. B., Robson, D. N., Li, J. M., & Keller, P. J. (2013). Whole-brain functional imaging at cellular resolution using light-sheet microscopy. *Nature methods*, *10*(5), 413.

11. Prevedel, R., Yoon, Y. G., Hoffmann, M., Pak, N., Wetzstein, G., Kato, S., ... & Vaziri, A. (2014). Simultaneous whole-animal 3D imaging of neuronal activity using light-field microscopy. *Nature methods*, *11*(7), 727.

12. Lemon, W.C., et al. "Whole-central nervous system functional imaging in larval Drosophila." *Nature communications* 6 (2015).

13. Grimm, J. B., Muthusamy, A. K., Liang, Y., Brown, T. A., Lemon, W. C., Patel, R., ... & Lavis, L. D. (2017). A general method to fine-tune fluorophores for live-cell and in vivo imaging. Nature methods, 14(10), 987.

14. Vladimirov, N., Wang, C., Höckendorf, B., Pujala, A., Tanimoto, M., Mu, Y., ... & Koyama, M. (2018). Brain-wide circuit interrogation at the cellular level guided by online analysis of neuronal function. *Nature methods*, *15*(12), 1117.

15. Chhetri, R. K., Amat, F., Wan, Y., Höckendorf, B., Lemon, W. C., & Keller, P. J. (2015). Whole-animal functional and developmental imaging with isotropic spatial resolution. *Nature methods*, *12*(12), 1171.

16. Vogelstein, J. T., Park, Y., Ohyama, T., Kerr, R. A., Truman, J. W., Priebe, C. E., & Zlatic, M. (2014). Discovery of brainwide neural-behavioral maps via multiscale unsupervised structure learning. *Science*, *344*(6182), 386-392.

17. Robie, A. A., Hirokawa, J., Edwards, A. W., Umayam, L. A., Lee, A., Phillips, M. L., ... & Reiser, M. B. (2017). Mapping the neural substrates of behavior. *Cell*, *170*(2), 393-406.

18. Konopka, R. J., & Benzer, S. (1971). Clock mutants of Drosophila melanogaster. *Proceedings of the National Academy of Sciences*, *68*(9), 2112-2116.

19. Sokolowski, M. B. (2001). Drosophila: genetics meets behaviour. *Nature Reviews Genetics*, *2*(11), 879.

20. Garcia-Campmany, L., Stam, F. J., & Goulding, M. (2010). From circuits to behaviour: motor networks in vertebrates. *Current opinion in neurobiology*, *20*(1), 116-125.

21. Borst, A. (2014). Fly visual course control: behaviour, algorithms and circuits. *Nature Reviews Neuroscience*, *15*(9), 590.

22. Long, X., Colonell, J., Wong, A. M., Singer, R. H., & Lionnet, T. (2017). Quantitative mRNA imaging throughout the entire Drosophila brain. *nature methods*, *14*(7), 703.

23. Bakker, K. (1959). Feeding period, growth, and pupation in larvae of Drosophila melanogaster. *Entomologia Experimentalis et Applicata*, *2*(3), 171-186.

24. Truman, J. W. (2005). Hormonal control of insect ecdysis: endocrine cascades for coordinating behavior with physiology. *Vitamins & Hormones*, *73*, 1-30.

25. Krashes, M. J., DasGupta, S., Vreede, A., White, B., Armstrong, J. D., & Waddell, S. (2009). A neural circuit mechanism integrating motivational state with memory expression in Drosophila. *Cell*, *139*(2), 416-427.

26. Lee, H. G., Kim, Y. C., Dunning, J. S., & Han, K. A. (2008). Recurring ethanol exposure induces disinhibited courtship in Drosophila. *PLoS One*, *3*(1), e1391.

27. Trannoy, S., & Kravitz, E. A. (2017). Strategy changes in subsequent fights as consequences of winning and losing in fruit fly fights. *Fly*, *11*(2), 129-138.

28. Truman, J. W., & Bate, M. (1988). Spatial and temporal patterns of neurogenesis in the central nervous system of Drosophila melanogaster. *Developmental biology*, *125*(1), 145-157.

29. Li, H. H., Kroll, J. R., Lennox, S. M., Ogundeyi, O., Jeter, J., Depasquale, G., & Truman, J. W. (2014). A GAL4 driver resource for developmental and behavioral studies on the larval CNS of Drosophila. *Cell reports*, *8*(3), 897-908.

30. Truman, J. W., Moats, W., Altman, J., Marin, E. C., & Williams, D. W. (2010). Role of Notch signaling in establishing the hemilineages of secondary neurons in Drosophila melanogaster. *Development*, *137*(1), 53-61.

31. Lacin, H., Chen, H. M., Long, X., Singer, R. H., Lee, T., & Truman, J. W. (2019). Neurotransmitter identity is acquired in a lineage-restricted manner in the Drosophila CNS. *eLife*, *8*, e43701.

32. Freeman, M. R. (2015). Drosophila central nervous system glia. *Cold Spring Harbor perspectives in biology*, *7*(11), a020552.

33. Satija, R., Farrell, J. A., Gennert, D., Schier, A. F., & Regev, A. (2015). Spatial reconstruction of single-cell gene expression data. *Nature biotechnology*, *33*(5), 495.

34. Strahl, B. D., & Allis, C. D. (2000). The language of covalent histone modifications. *Nature*, *403*(6765), 41.

35. Crocker, A., Guan, X. J., Murphy, C. T., & Murthy, M. (2016). Cell-type-specific transcriptome analysis in the Drosophila mushroom body reveals memory-related changes in gene expression. *Cell reports*, *15*(7), 1580-1596.

36. Davie, K., Janssens, J., Koldere, D., De Waegeneer, M., Pech, U., Kreft, Ł., ... & Poovathingal, S. (2018). A single-cell transcriptome atlas of the aging Drosophila brain. *Cell*, *174*(4), 982-998.

37. Croset, V., Treiber, C. D., & Waddell, S. (2018). Cellular diversity in the Drosophila midbrain revealed by single-cell transcriptomics. *Elife*, *7*, e34550.

38. Worrell, J. W., & Levine, R. B. (2008). Characterization of voltage-dependent Ca2+ currents in identified Drosophila motoneurons in situ. *Journal of neurophysiology*, *100*(2), 868-878.

39. Sanyal, S. (2009). Genomic mapping and expression patterns of C380, OK6 and D42 enhancer trap lines in the larval nervous system of Drosophila. *Gene Expression Patterns*, *9*(5), 371-380.

40. Schwaerzel, Martin, et al. "Dopamine and octopamine differentiate between aversive and appetitive olfactory memories in Drosophila." *Journal of Neuroscience* 23.33 (2003): 10495-10502.

41. Selcho, Mareike, et al. "The role of dopamine in Drosophila larval classical olfactory conditioning." *PloS one* 4.6 (2009): e5897.

42. Ro, J., Pak, G., Malec, P. A., Lyu, Y., Allison, D. B., Kennedy, R. T., & Pletcher, S. D. (2016). Serotonin signaling mediates protein valuation and aging. *Elife*, *5*, e16843.

43. Ziegler, A. B., Augustin, H., Clark, N. L., Berthelot-Grosjean, M., Simonnet, M. M., Steinert, J. R., ... & Grosjean, Y. (2016). The amino acid transporter JhI-21 coevolves with glutamate receptors, impacts NMJ physiology, and influences locomotor activity in Drosophila larvae. *Scientific reports*, *6*, 19692.

44. Ziegler, A. B., Manière, G., & Grosjean, Y. (2018). JhI-21 plays a role in Drosophila insulin-like peptide release from larval IPCs via leucine transport. *Scientific reports*, *8*(1), 1908.

45. Schlegel, P., Texada, M. J., Miroschnikow, A., Schoofs, A., Hückesfeld, S., Peters, M., ... & Truman, J. W. (2016). Synaptic transmission parallels neuromodulation in a central food-intake circuit. Elife, 5, e16799.

46. Rulifson, E. J., Kim, S. K., & Nusse, R. (2002). Ablation of insulin-producing neurons in flies: growth and diabetic phenotypes. *Science*, *296*(5570), 1118-1120.

47. Luo, J., Becnel, J., Nichols, C. D., & Nässel, D. R. (2012). Insulin-producing cells in the brain of adult Drosophila are regulated by the serotonin 5-HT 1A receptor. *Cellular and Molecular Life Sciences*, *69*(3), 471-484.

48. Hentze, J. L., Carlsson, M. A., Kondo, S., Nässel, D. R., & Rewitz, K. F. (2015). The neuropeptide allatostatin A regulates metabolism and feeding decisions in Drosophila. *Scientific reports*, *5*, 11680.

49. Williamson, M., Lenz, C., Winther, M. E., Nässel, D. R., & Grimmelikhuijzen, C. J. (2001). Molecular cloning, genomic organization, and expression of a B-type (cricket-type) allatostatin preprohormone from Drosophila melanogaster. *Biochemical and biophysical research communications*, *281*(2), 544-550.

50. Hwang, R. Y., Zhong, L., Xu, Y., Johnson, T., Zhang, F., Deisseroth, K., & Tracey, W. D. (2007). Nociceptive neurons protect Drosophila larvae from parasitoid wasps. *Current Biology*, *17*(24), 2105-2116.

51. Jovanic, T., Masson, J. B., Truman, J. W., & Zlatic, M. (2017). Mapping neurons and brain regions underlying sensorimotor decisions and sequences. *bioRxiv*, 215236.

52. Fontana, J. R., & Crews, S. T. (2012). Transcriptome analysis of Drosophila CNS midline cells reveals diverse peptidergic properties and a role for castor in neuronal differentiation. *Developmental biology*, *372*(1), 131-142.

53. Kraaijeveld, A. R., & Godfray, H. C. J. (1997). Trade-off between parasitoid resistance and larval competitive ability in Drosophila melanogaster. *Nature*, *389*(6648), 278.

54. Sorrentino, R. P., Carton, Y., & Govind, S. (2002). Cellular immune response to parasite infection in the Drosophila lymph gland is developmentally regulated. *Developmental biology*, *243*(1), 65-80.

55. Etheredge, J. (2017). *Transcriptional profiling of Drosophila larval ventral nervous system hemilineages* (Doctoral dissertation, University of Cambridge).

56. Diaz, M.M. et al. Allatostatin-C/AstC-R2 Is a Novel Pathway to Modulate the Circadian Activity Pattern in Drosophila. Curr. Biol. 29, 13-22 (2019)

57. Meissner, G.W. et al. Mapping Neurotransmitter Identity in the Whole-Mount Drosophila Brain Using Multiplex High-Throughput Fluorescence in Situ Hybridization. GENETICS 211:437-482

58. Schindelin, J. et al. Fiji: an open-source platform for biological-image analysis. Nat Methods 9, 676-82 (2012).

59. Cardona, A. et al. TrakEM2 software for neural circuit reconstruction. PLoS One 7, e38011 (2012).

60. Lionnet, T. et al. A transgenic mouse for in vivo detection of endogenous labeled mRNA. Nat Methods 8, 165-70 (2011).

Acknowledgments:

We thank J. Grimm for sharing reagents; J. Etheredge for fly stocks; S. Harrison, M. Mercer and the Janelia Fly Core for assistance with fly husbandry; A. Lemire and K. Aswath of Janelia Quantitative Genomics for assistance with sequencing. **Funding:** Supported by Janelia HHMI (M.Z.), Gates Cambridge Trust (B.T.C.), HHMI Medical Fellows Program (B.T.C.), NSF (L.L.M.). **Author contributions**: B.T.C., L.L.M., and M.Z. conceptualized the study; B.T.C., J.D.W., X.L., and J.Y. performed all data analysis and visualization; J.D.W., J.B.M., and J.Y. performed all software development, B.T.C., X.L., J.L., A.B.K., and L.L.M. performed the investigation; B.T.C., J.D.W. and X.L. curated the data; B.T.C., X.L., A.B.K., and L.L.M. developed methodology and performed validation, R.H.S., L.L.M. and M.Z. acquired funding, provided resources, and performed project administration; S.T., R.H.S., L.L.M. and M.Z. supervised the study; B.T.C., J.D.W. and X.L. wrote the original draft; B.T.C., X.L., J.D.W., R.H.S., L.L.M. and M.Z. edited the final manuscript; all authors reviewed the final manuscript. **Competing interests:** none. **Data and materials availability:** RNA sequencing data was deposited in GEO. All other data is available in the main text or supplementary materials.

Supplementary Materials:

Materials and Methods

Figures S1-S4

Tables S1-S3

Movies S1-S2

References (*56-60*)

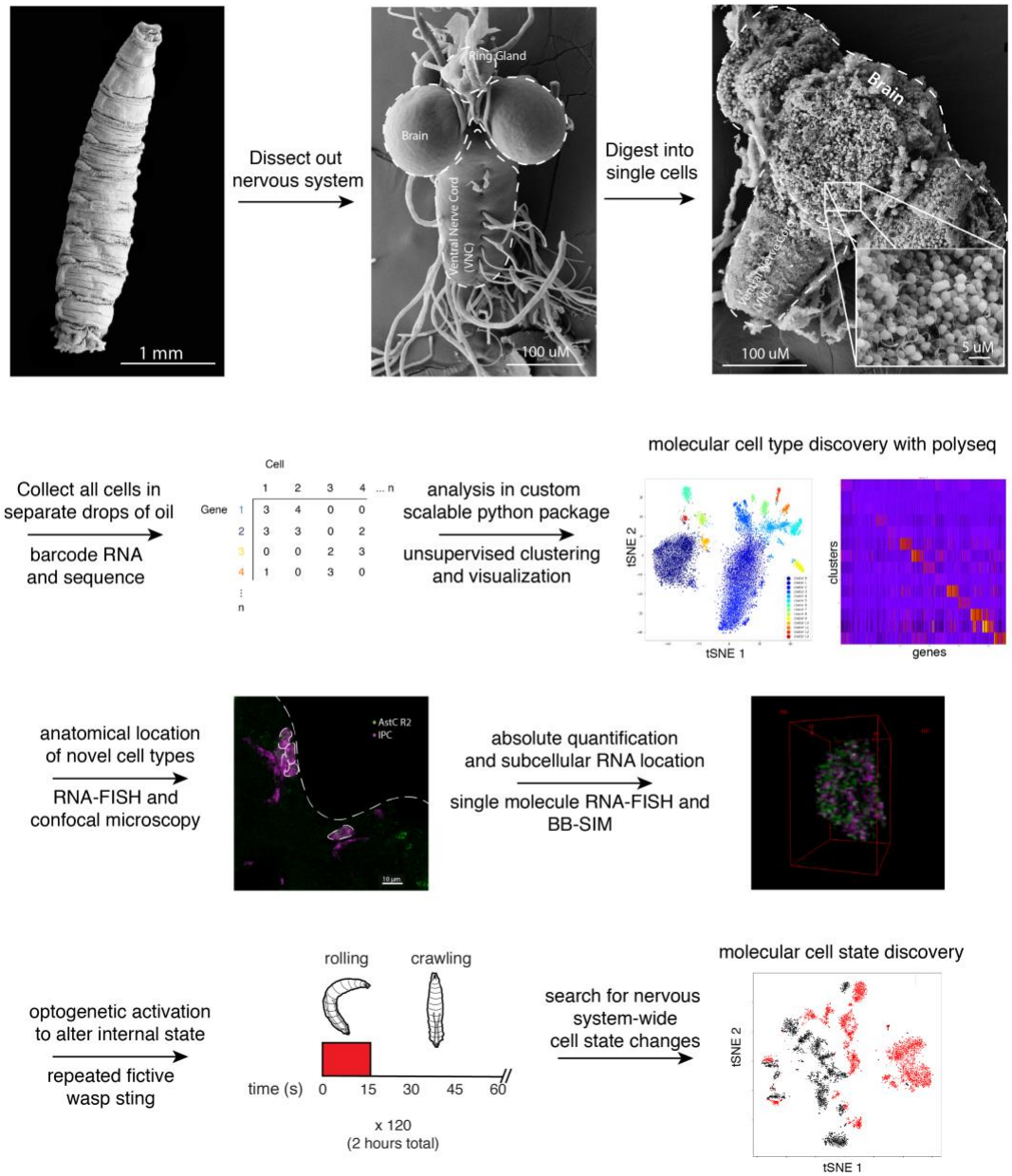

Fig. 1. Schematic of full nervous system scRNA-seq collection and analysis. In order to develop a brainwide transcriptomic atlas of the *Drosophila* nervous system with single cell resolution, we developed a protocol to digest the entire nervous system into single cells, collected the cells using a microfluidic device (10x Chromium machine, 10x Genomics, Pleasanton, CA), and sequenced the mRNA from each cell. After barcoding and sequencing, a cell by gene matrix is generated. This cell by gene matrix was then analyzed with *polyseq*, a custom python package. In total, 202,107 neurons and glia were sequenced. The anatomical location of newly defined molecular cell types were validated and identified using RNA-FISH with confocal imaging. To push the technique forward, RNA-FISH combine with Bessel beam selective plane illumination microscope (BB-SIM) was used to obtain the absolute quantification

and subcellular location of transcripts in these new cell types. Optogenetic manipulations were performed to alter the internal state of the animal, either with two hours of fictive wasp sting or two hours of overactivation of 10% of brain neurons, and scRNAseq was used once more to search for a change in molecular cell state between conditions.

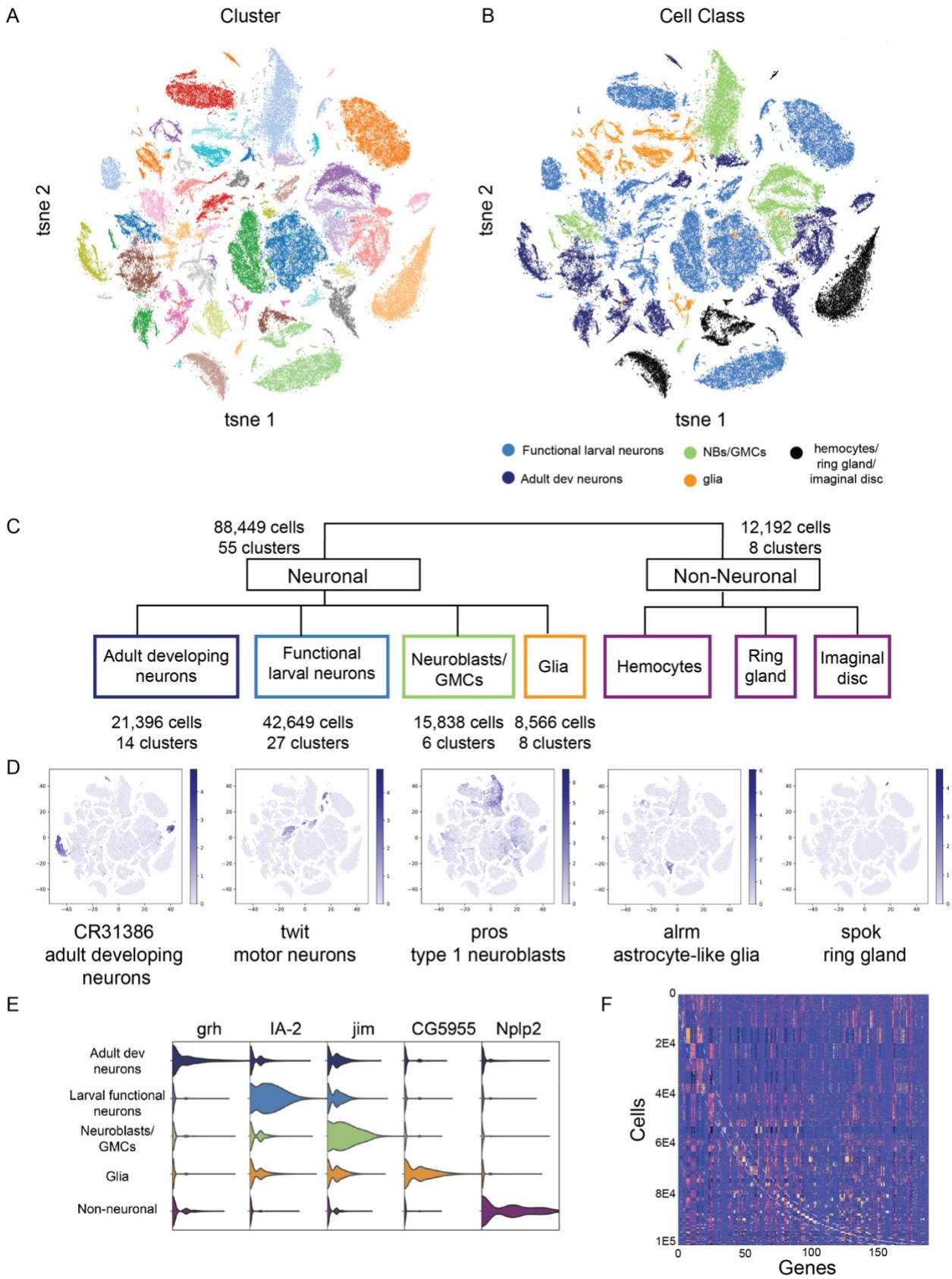
A
Cluster
tsne 2
tsne 1
B
Cell Class
tsne 2
tsne 1
Functional larval neurons
Adult dev neurons
NBs/GMCs
glia
hemocytes/
ring gland/
imaginal disc
C
88,449 cells
55 clusters
12,192 cells
8 clusters
Neuronal
Non-Neuronal
Adult developing neurons
Functional larval neurons
Neuroblasts/ GMCs
Glia
Hemocytes
Ring gland
Imaginal disc
21,396 cells
14 clusters
42,649 cells
27 clusters
15,838 cells
6 clusters
8,566 cells
8 clusters
D
CR31386
adult developing neurons
twit
motor neurons
pros
type 1 neuroblasts
alrm
astrocyte-like glia
spok
ring gland
E
grh
IA-2
jim
CG5955
Nplp2
Adult dev neurons
Larval functional neurons
Neuroblasts/ GMCs
Glia
Non-neuronal
F
Cells
Genes
0
2E4
4E4
6E4
8E4
1E5
0
50
100
150

Fig. 2. Single cell transcriptomic atlas of the larval nervous system. A. t-SNE visualization of high-quality cells colored by cluster. Genes that define each cluster and their expression levels are in the supplementary full data atlas. B, C. Cell types split into classes that are relevant to the larval nervous system, including adult developing neurons, larval functional neurons, neuroblasts and ganglion mother cells (GMCs), glia, and non-neuronal tissues. Known markers for these cell classes were used to define cell class groups. Once defined, new cell class genes were discovered. D. Expression levels for exemplary genes from each class are shown with their expression levels in the t-SNE space. E. Genes that define each cell cluster and cell class were discoverable. Violin plots show example genes from each cell class. F. Heatmap of all high-quality cells and the top 3 genes that define each cluster.

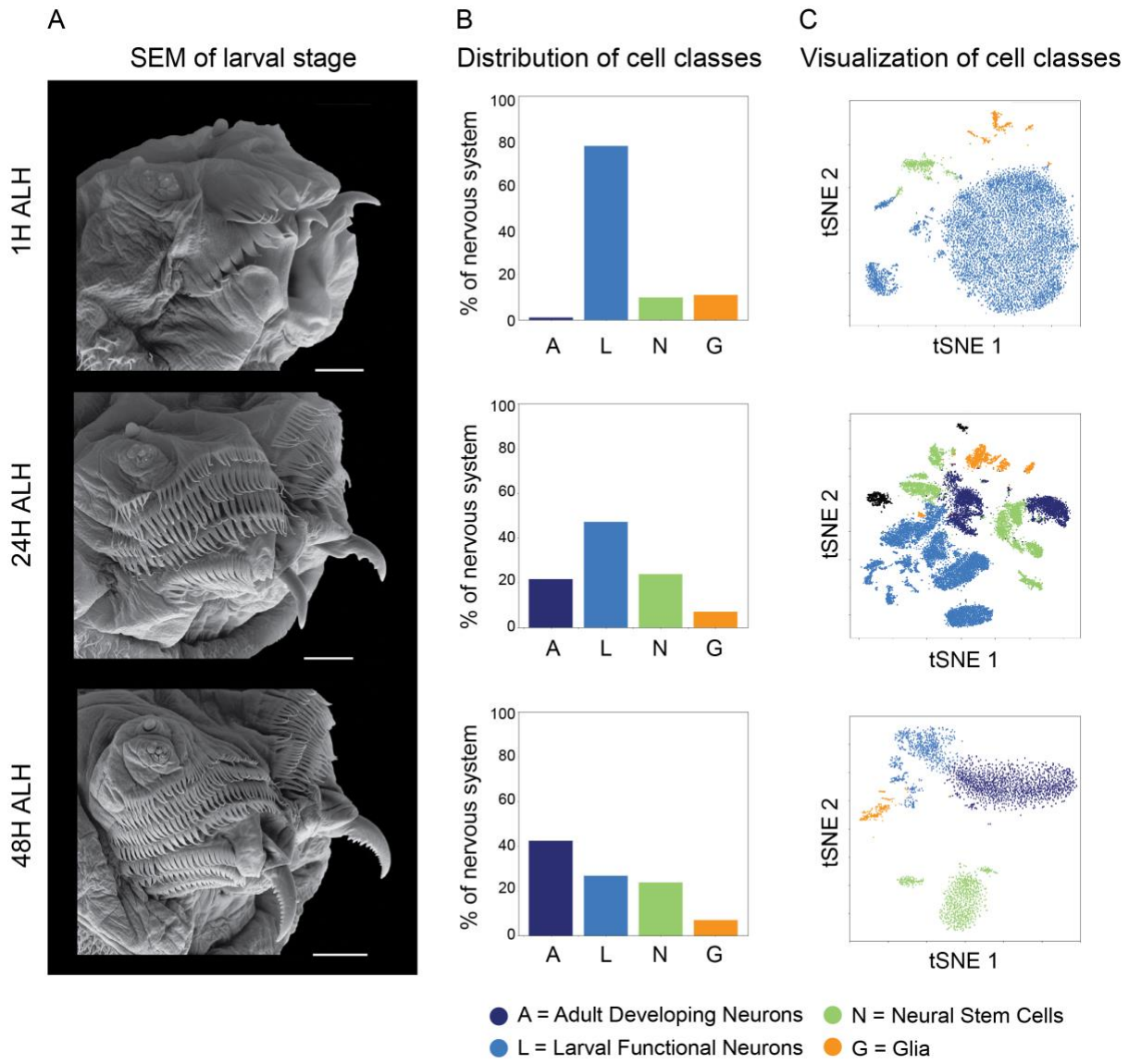

Fig. 3. Single cell transcriptomic atlas of the larval nervous system across developmental stages. A. Scanning electron microscopy (SEM) image of larval anatomy during development. Three developmental stages were sequenced separately to build age specific atlases (Scale bar = 10 μm for 1H, 20 μm for 24H and 40 μm for 48 H). Larval size and elaboration of mouth hooks continues during this developmental period. B. Distribution of cell classes. C. t-SNE clustering of stage specific samples. Just after hatching (1 hour after larval hatching (ALH)), there is one primary central cluster of neurons. Neuron classes are recognizable, including cholinergic neurons, motor neurons, astrocyte-like glia, and neuroblasts. At 24 hours ALH multiple main groups of neurons are recognizable with markers consistent with an increase in neuroblasts and ganglion mother cells (GMCs). In addition, there are large groups of neurons, here labeled "developing adult neurons" which have neuron markers but few or no genes expressed for synaptic transmitters and receptors. This is consistent with the large burst of neurons born at this point in development which lay dormant until adult life. At 48 hours ALH the functional larval and developing adult populations are still identifiable.

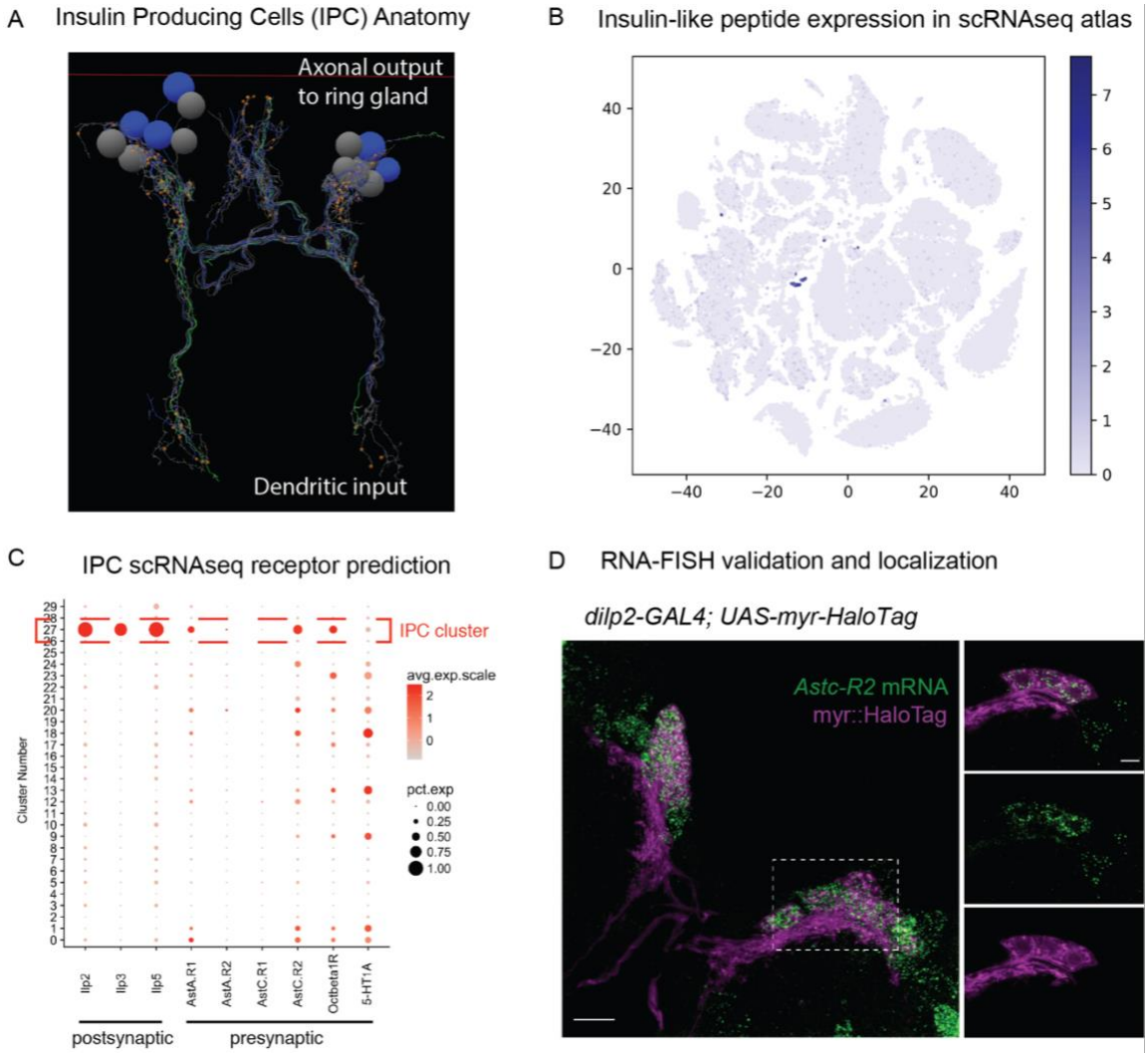

Fig. 4. Transcriptomic Atlas predicts previously unknown neuropeptide phenotype for insulin producing cells and is verified by RNA-FISH. A. Anatomy of insulin-producing cells (IPCs). The IPCs are a group of 7 bilaterally symmetrical neurons which receive input through their dendrites (purple) about the nutritional state of the animal and release insulin like peptides (ILP2, ILP3, ILP5) through their axons (green), which synapse on the ring gland, to control carbohydrate balance. They are homologous in function to the vertebrate pancreatic beta islet cells. B,C. The RNAseq atlas built in this study discovered the IPCs as a separate cluster (cluster 27) with expression in the IPCs of receptors for octopamine, serotonin, and allatostatin A, which matched previous literature. Surprisingly, the atlas also suggested the presence of a previously unrecognized receptor in the IPCs for allatostatin C (AstC-R2). D. Detection of AstC-R2 mRNA in IPCs. Maximum-intensity projection of the confocal stack of a brain in which the IPCs are labeled with a fluorescent HaloTag ligand (Magenta) and AstC-R2 mRNA is detected by FISH (green), bar 10µm. Dashed lines outline area where the single z plane is shown on the left panels, bar 5µm. Movie of D is in supplement (Movie S1).

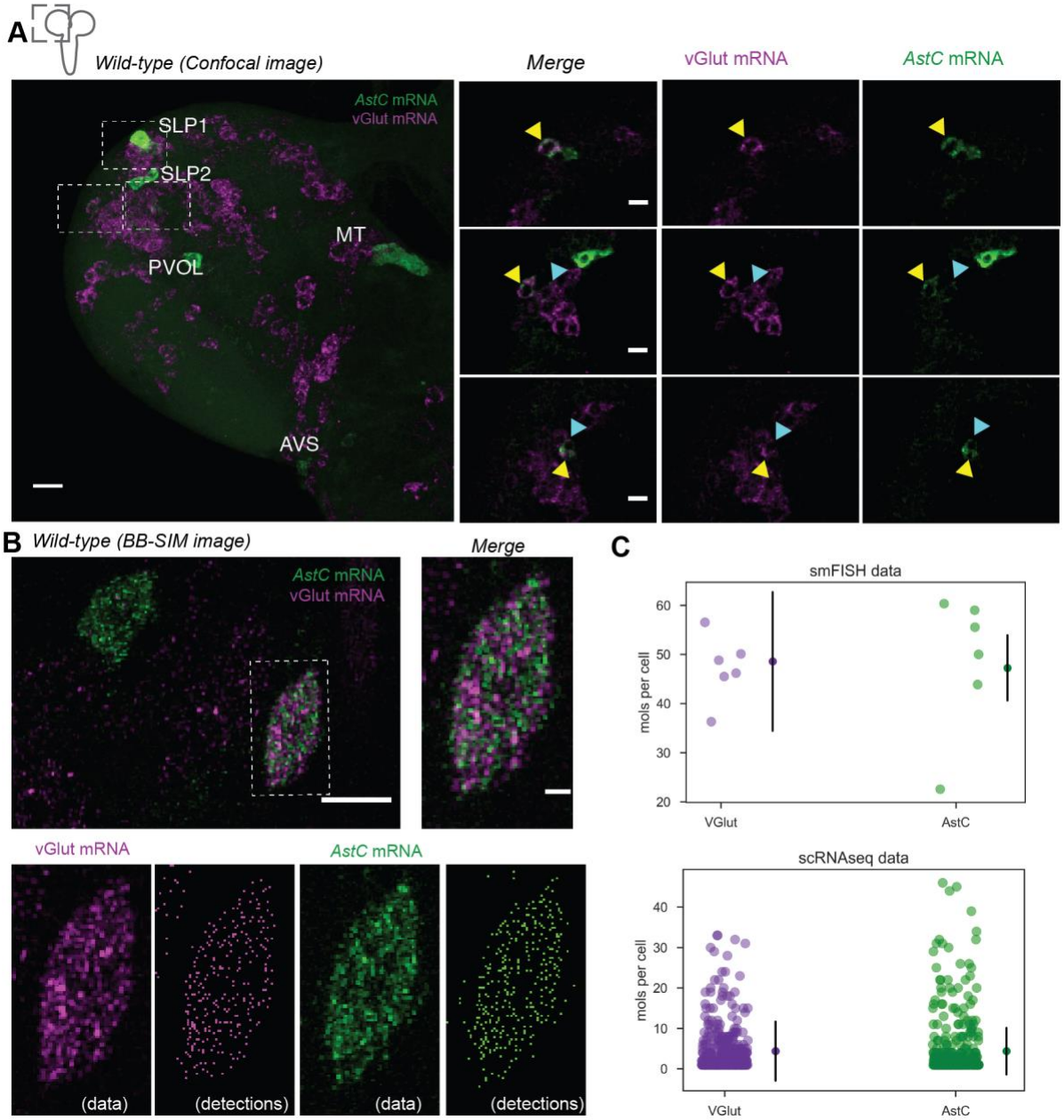

Fig. 5. Correlation between single molecule FISH and scRNAseq. A. Identifying cells that have coexpression of AstC (green) and vGlut (magenta) in the whole brain B. Maximum-intensity projections of BB-SIM stack of the AstC and vGlut mRNA FISH channels, bar 10µm. Dashed lines outline 2 cells that co-express AstC and vGlut mRNAs are shown on the right panels, bar 1µm. Lower panels show individual FISH channel and the reconstructions obtained using spot-counting algorithm. C. Comparison of the quantification of AstC and vGlut mRNAs between smFISH and scRNAseq.

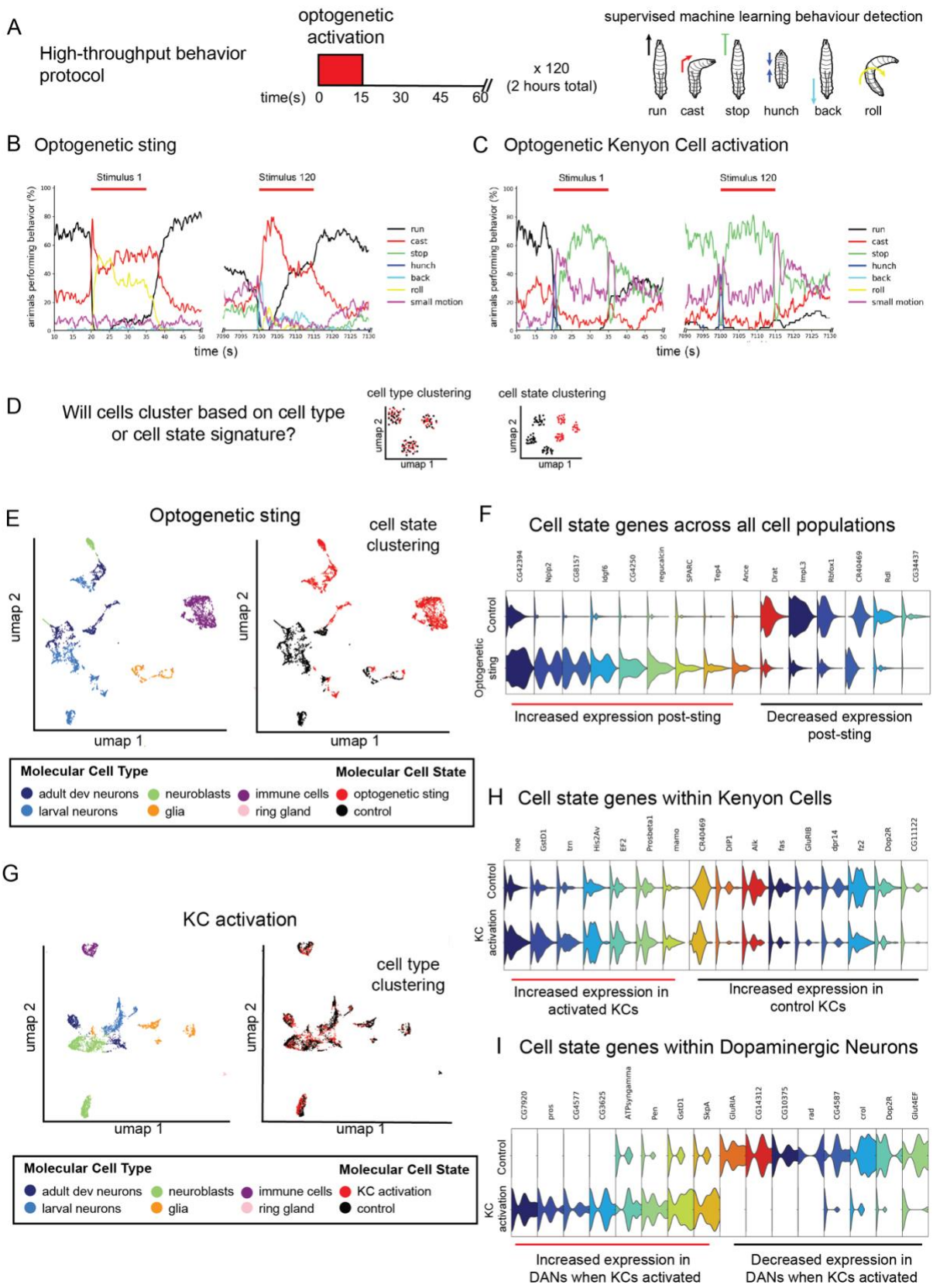
A
High-throughput behavior protocol
optogenetic activation
time(s) 0 15 30 45 60
x 120
(2 hours total)
supervised machine learning behaviour detection
run cast stop hunch back roll
B Optogenetic sting
Stimulus 1
Stimulus 120
animals performing behavior (%)
time (s)
run
cast
stop
hunch
back
roll
small motion
C Optogenetic Kenyon Cell activation
D
Will cells cluster based on cell type or cell state signature?
cell type clustering
cell state clustering
umap 2
umap 1
E
Optogenetic sting
cell state clustering
Molecular Cell Type
adult dev neurons
larval neurons
neuroblasts
glia
immune cells
ring gland
Molecular Cell State
optogenetic sting
control
F
Cell state genes across all cell populations
CG42394 Nplp2 CG8157 Idgf6 CG4250 regucalcin SPARC Tep4 Ance Drat Impl3 Rbfox1 CR40469 Rdl CG34437
Control
Optogenetic sting
Increased expression post-sting
Decreased expression post-sting
G
KC activation
cell type clustering
Molecular Cell Type
adult dev neurons
larval neurons
neuroblasts
glia
immune cells
ring gland
Molecular Cell State
KC activation
control
H Cell state genes within Kenyon Cells
noe GstD1 trn Hisc2Av EF2 Prosbeta1 mamo CR40469 Dlp1 Alk fas GluRIB dpr14 fz2 Dop2R CG11122
Control
KC activation
Increased expression in activated KCs
Increased expression in control KCs
I Cell state genes within Dopaminergic Neurons
CG7920 pros CG4577 CG3625 ATPsyngamma Pen GstD1 SkpA GluRIA CG14312 CG10375 rad CG4587 crol Dop2R Glut4EF
Control
KC activation
Increased expression in DANs when KCs activated
Decreased expression in DANs when KCs activated

Fig. 6. Optogenetic sting and KC activation induce a change in internal state which is observable as cell state changes in gene expression. A. High-throughput behavior assays were performed using previously described equipment (*55*). Fifteen seconds of optogenetic activation were delivered every 60 seconds over a two-hour period. Larval behavior was recorded, and supervised machine learning techniques were used to classify behavior as one of six behaviors. B. For the optogenetic sting protocol, animals first respond to basin activation by performing an escape response, including a fast turning response and rolling as previously described (*5*). Over the course of the experiment, a switch in behavior is observed from rolling to backing up. C. Kenyon cell activation led to a hunch followed by freezing. At the start of training, animals crawled forward to offset. After training, the hunch and freezing response remained at onset, but the offset response switched to a turn rather than a forward crawl. D. Transcriptomic atlases were produced from animals that underwent each behavior protocol and matched controls. We were primarily interested in whether cell state or cell type clustering would be observed after altering internal state (sting) or overactivating the memory center (KC activation). Cell type clustering would result in cells from activated animals and controls mixing in the same clusters while cell state clustering would lead to separation of cells from activated and control animals. E,F. The optogenetic sting protocol led to cell state clustering. All cell classes (as defined in Fig. 2) were identifiable for both conditions. A large immune cell population was specific to the sting condition, suggesting an immune response to fictive wasp sting. F. Cell state genes were discovered with upregulated and downregulated expression that separated all cell populations. G-I. KC activation led to cell type clustering and local cell state changes. Even though a larger total number of cells were activated (~200 KCs versus 64 basins), a less dramatic switch was observed in behavior and in the transcriptome. There were differentially expressed genes in two key populations of cells in the learning and memory center: KCs and dopaminergic neurons.

Supplementary Materials for

Comparative single-cell transcriptomics of complete insect nervous systems

B.T. Cocanougher†, J.D. Wittenbach†, X. Long†, A.B. Kohn, T.P. Norekian, J. Yan, J. Colonell, J.B. Masson, J.W. Truman, A. Cardona, S. Turaga, R.H. Singer‡, L.L. Moroz‡*, M. Zlatic‡*

†These authors contributed equally to this work.

‡These authors contributed equally to this work.

Correspondence to: moroz@whitney.ufl.edu (L.L.M.); zlaticm@janelia.hhmi.org (M.Z.)

This PDF file includes:

Materials and Methods
Figs. S1 to S5
Captions for Table S1 to S3
Captions for Movies S1 to S2

Other Supplementary Materials for this manuscript include the following:

Tables S1 to S3
Movies S1 to S2

Materials and Methods

Cell isolation and collection for scRNA-seq:

Drosophila larvae were grown on standard fly food at 25°C. In order to label most of the neurons in the nervous system, the GMR_57C10-GAL4 driver line was crossed to pJFRC29-10XUAS-IVS-myr::GFP-p10 in attP40; pJFRC105-10XUAS-IVS-nlstdTomato in VK00040 (hereafter UAS-J29;J105) virgins (*55*).

Larvae were dissected at 1 hour, 24 hours, 48 hours, or 96 hours after larval hatching (ALH). All dissections were performed in cold adult hemolymph solution (AHS) with no calcium or magnesium at pH 7.4. Quality of single cell isolation was investigated by visual inspection with compound and confocal microscopy. Samples were placed on ice during waiting periods. Samples were isolated and run on the 10x Chromium Single Cell 3' immediately after cell dissociation.

First the complete central nervous system (CNS) was dissected from every animal. The dissected nervous systems were kept in cold AHS on ice. For those samples where the brain and the ventral nerve cord (VNC) were sequenced separately, the separation of the brain from the VNC was performed using fine tipped forceps and MicroTools (Cat #: 50-905-3315, Electron Microscopy Sciences). The time from digestion (the part of the protocol most likely to induce cell stress) to on the 10x Genomic instrument was never longer than 30 minutes.

After separation of the brain from the VNC, the desired tissue was placed in 18 µL of AHS on ice. Once all samples were prepared, 2 µL of 10x neutral protease (Cat #: LS02100, Worthington Biochemical Corp, Lakewood, NJ, USA) was added to a final volume of 20 µL. The intact brain tissue was digested for 5 minutes. The tissue was then transferred to a fresh drop of 20 µL of AHS.

Each sample was triturated with a clean, thinly pulled glass electrode until no tissue was visible under a dissection scope. All debris (pieces of nerve and undigested neuropile) was removed. Samples with fluorescent markers were observed under a fluorescence microscope to approximate cell density. The samples were then loaded onto the 10x Chromium chip.

Single cell capture and library construction was performed using the 10x Chromium machine and the Chromium Single Cell 3' v2 Library and Gel Bead Kit (10x Genomics, Pleasanton, CA). Manufacturer's recommendations were followed for cell collection and library construction. Libraries were sequenced with an Illumina HiSeq.

scRNA-seq analysis

Bioinformatic analysis was performed using Cell Ranger software (Version 1.3.1, 10x Genomics, Pleasanton, CA, USA), the Seurat R package (*33*) and custom software in R and Python, including the *polyseq* Python package developed here. Software to train classifiers using neural networks was built with TensorFlow. The *polyseq* packages as well as all scripts used for analysis in the study are available on GitHub.

Briefly, Cell Ranger was used to perform demultiplexing, alignment, filtering, and counting of barcodes and UMIs, with the output being a cell-by-genes matrix of counts. To further ensure that only high-quality cells were retained, any cell that registered counts in a unique number of genes below a baseline threshold was removed. To reduce the dimensionality of the data for computational tractability, any gene that was not expressed in a baseline number of cells was also dropped.

To account for the fact that raw counts tended to span many orders of magnitude (~ 10_0-10_5), counts were transformed via log(counts + 1). To control for cell size and sequencing depth, the sum of the (log-transformed) counts within each cell used as a regressor for a linear regression model to predict the (log-transformed) counts for each gene (one linear regression model per gene, with each cell being a sample). "Gene expression levels" were then quantified as the z-scored residuals from the fitted models (i.e. standard deviations above/below the predicted log-transformed counts for a particular gene across all cells).

Next, to further reduce the dimensionality of the data in preparation for downstream clustering and embedding operations (both of which have computational costs that scale poorly with the dimensionality of the feature space), principal component analysis was performed with cells as samples and gene expression levels as features. The top K principal components (PCs) were retained as features for downstream analyses. For the lager cell atlas dataset, K was chosen to retain a desired percentage of the total variance. For the smaller cell state datasets, K was chosen

automatically via a shuffle test – on each shuffle, gene expression levels for each gene were randomly permuted across all cells and the percent variance explained by the top PC was recorded; the 95th percentile of this value across all shuffles was then used as a threshold to determine the cutoff point for keeping PCs with respect to percent variance explained by a particular PC.

Based on these top PCs, cells were clustered using the Louvain-Jaccard graph-based clustering approach. Briefly, the k-nearest neighbor graph between cells was is computed. Edge weights are then determined using the Jaccard index, which measures the fraction of shared neighbors between any two nodes. Finally, the Louvain community detection algorithm is applied to this graph to partition the nodes into clusters; this algorithm seeks to optimize weight of connections with each cluster relative to those between clusters.

In order to visualize the results of the analysis, the PC features were also used to perform a nonlinear embedding into two dimensions. This was performed via either the t-SNE or the UMAP algorithm.

Once cluster identities were determined, the original gene expression level data was to determine important genes for defining each cluster. For each cluster, gene expression levels were used as features and a binary indicator of whether or not a cell came from the cluster in question was used as a target. This data was then used to fit a linear classifier (viz, a support vector classifier) to separate in-cluster cells from the rest of the population. The unit normal vector from the linear classifier was then extracted and the components used to rank order genes in terms of importance for defining that cluster. This same technique was also used to find important genes for groups defined by methods other than clustering.

mRNA in situ hybridization

FISH probes were designed based on transcript sequences using the online Stellaris Designer and purchased from Biosearch Technologies. Probe sequences for AstC and vGlut were previously report in here (*22, 56*) and probe sequences for AstC-R2, Hug, NPNL2 are in supplementary table1. Each probe is 18-22nt long with a 3' end amine-modified nucleotide that allows directly couple to an NHS-ester dye according to the manufacturer's instructions (Life Technologies). Dye-labeled probes were separated from the excess free dyes using the Qiagen Nucleotide Removal Columns. FISH protocol was described previously (*22, 57*). The brains of 3rd instar larvae were dissected in 1xPBS and fixed in 2% paraformaldehyde diluted PBS at room temperature for 55 min. Brain tissues were washed in 0.5% PBT, dehydrated, and stored in 100% ethanol at 4°C. After exposure to 5% acetic acid at 4 °C for 5 minutes, the tissues were fixed in 2% paraformaldehyde in 1xPBS for 55 min at 25 °C. The tissues were then washed in 1× PBS with 1% of NaBH4 at 4 °C for 30 min. Following a 2 hour incubation in prehybridization buffer (15% formamide, 2× SSC, 0.1% Triton X-100) at 50 °C, the brains were introduced to hybridization buffer (10% formamide, 2× SSC, 5× Denhardt's solution, 1 mg/ml yeast tRNA, 100 µg/ml, salmon sperm DNA, 0.1% SDS) containing FISH probes at 50 °C for 10 h and then at 37 °C for an additional 10 h. After a series of wash steps, the brains were dehydrated and cleared in xylenes.

Imaging

For confocal imaging, the tissues were mounted in DPX. Image Z-stacks were collected using an LSM880 confocal microscope fitted with an LD LCI Plan-Apochromat 25x/0.8 oil or Plan-Apochromat 63x/1.4 oil objective after the tissue cured for 24 hours. For single molecule imaging, we use a previous described Bessel beam selective plane illumination microscope (BB-SIM). Detail construction of microscope and the imaging procedure is described previously (*22*). Briefly, this BB-SIM is engineered to image in medium matched to the measured refractive index (RI) of xylene-cleared *Drosophila* tissue with axial resolution of 0.3 µm and lateral resolution of 0.2 µm. For BB-SIM imaging, the tissues were mounted on a 1.5x3mm poly-lysine coated coverslip attached to a 30mm glass rod. The imaging process requires the objectives and tissues immersed in the imaging medium consist with 90% 1,2-dichlorobenzene, 10% 1,2,4-trichlorobenzene with refractive index = 1.5525. Two orthogonally mounted excitation objectives are used to form Bessel beams, which are stepped to create an illumination sheet periodically striped along x or y, while a third objective (optical axis along the z direction) detects fluorescence. To employ structured illumination analysis, we collect multiple images with the illumination stripe pattern shifted to tile the plane in x, and repeat the process orthogonally to tile the plane in y. The sample is then moved in z, and the imaging repeated, and so on to image the 3D volume.

Imaging analysis

To quantify the number of vGlut and AstC mRNAs in cells contain both vGlut and AstC, we first manually segmented cells from BB-SIM z-stacks that have both vGlut and AstC FISH signals using the Fiji plugin TrakEM2 (*58*, *59*). After identifying the individual fluorescent spots in segmented cells used a previously described Matlab algorithm (*60*), we calculate the number of mRNAs per cell. Reconstructed images were generated using Matlab code that draws spots centered on each of the detected spots positions (*60*).

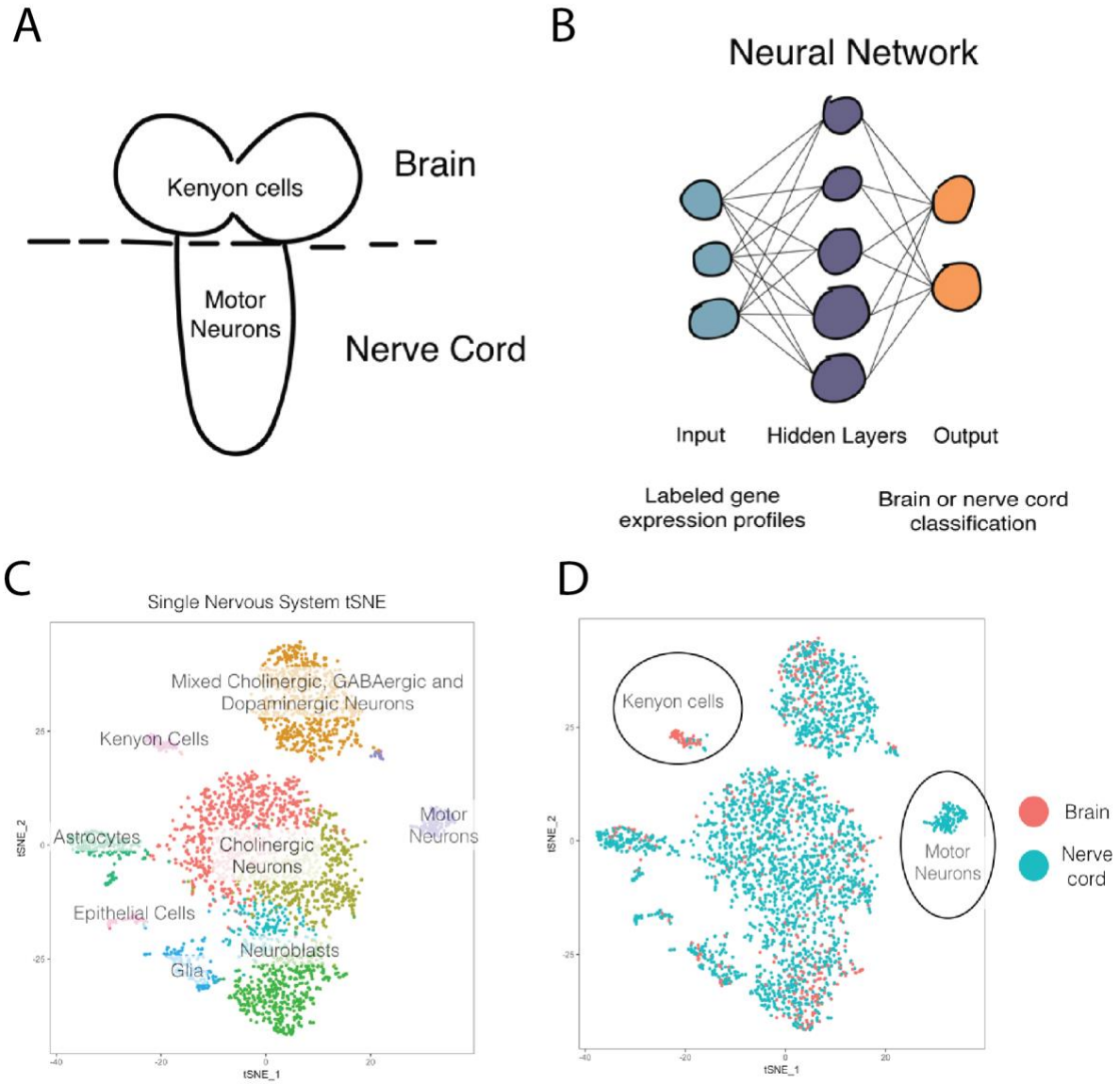

Fig. S1. Machine learning separates brain and VNC neurons.

A. We dissected the brain from the nerve cord and sequenced the RNA from each population separately. This provided ground truth labels which we could then feed into (B) a neural network to train a classifier to predict spatial origin from the brain of the VNC. C. We then used our classifer (built on separate data) to predict the brain or VNC origin of cells from a nervous system that was sequenced in one sample (no brian or VNC dissection). D. The classifier correctly separated brain from VNC cells in an intact sample. Most clusters were appropriately mixed brain and VNC cells, however one pure VNC population (motor neurons) and one pure brain population (Kenyon cells) were appropriately labeled by our machine learning classifier.

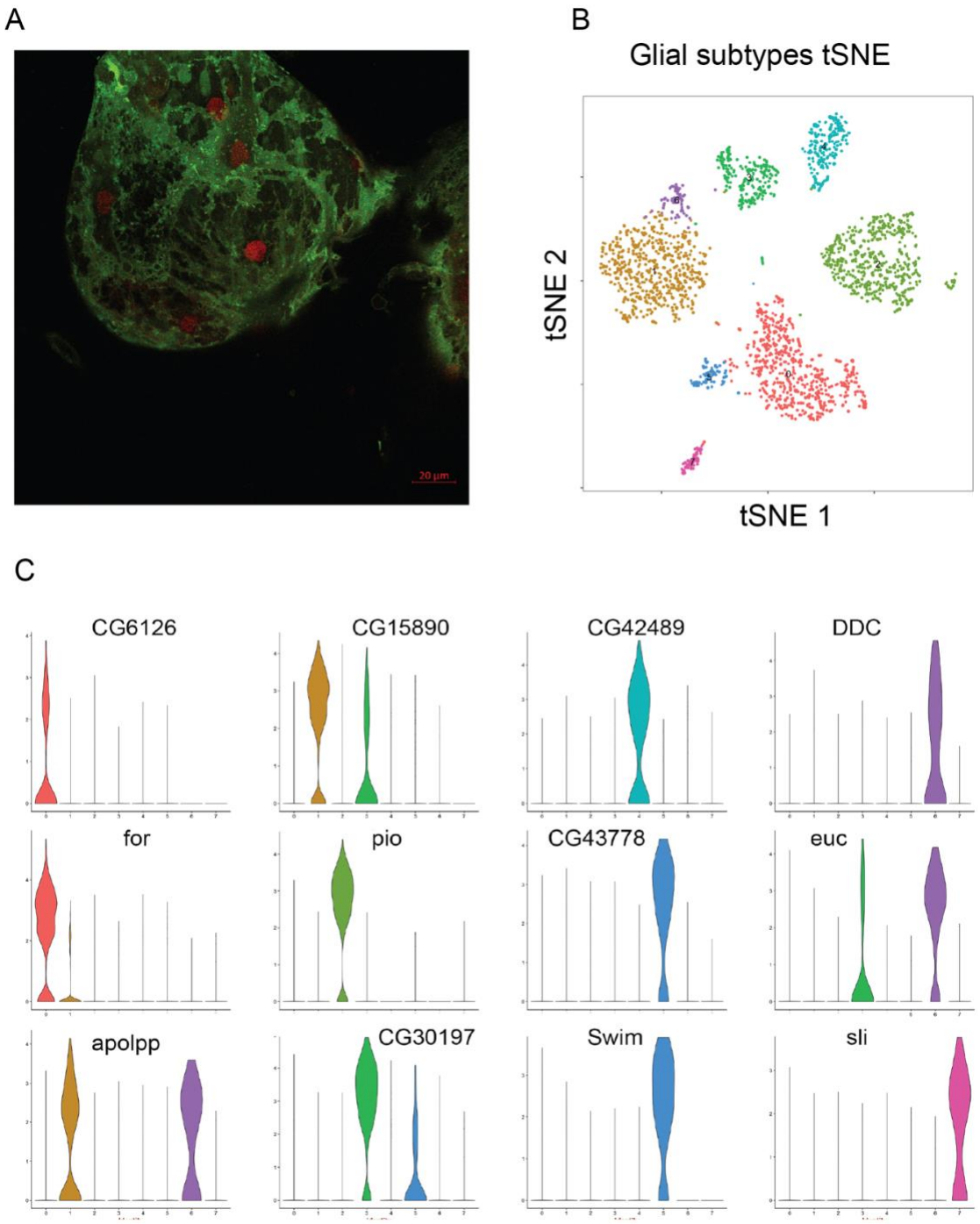

Fig. S2. Glial subtypes were recognizable in the atlas.

Glial subtypes were recognizable by extracting glial cells and analyzing them separately. A) Confocal image of surface glia using the repo-GAL4 driver line (scale bar = 20µm.). B) Glial subtypes break into 6 primary groups with 2 small subgroups. C) Genes which separate each type of glia.

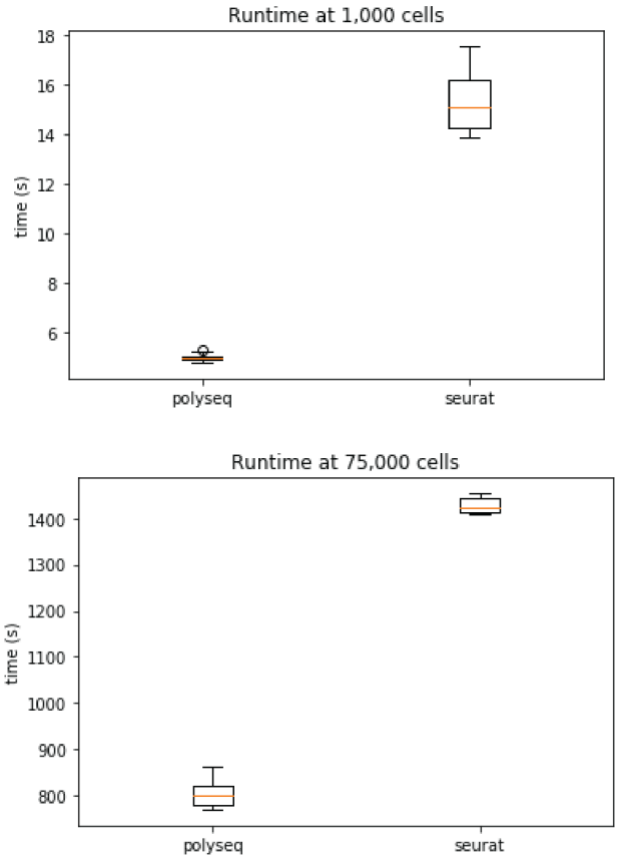

Fig. S3. Runtime for a typical pipeline to analyze single cell data in *Seurat* (R) versus *polyseq* (python).

Equivalent analysis pipelines were built in *Seurat* and *polyseq* to analyze single cell RNAseq data for datasets of 1,000 cells and 75,000 cells on a single laptop running Mac OS with 16GB RAM and a 2.6 GHz Intel Core i7. Polyseq outperformed Seurat for both small and medium datasets. Jupyter notebooks used for testing that contain code for Seurat and polyseq is available upon request.

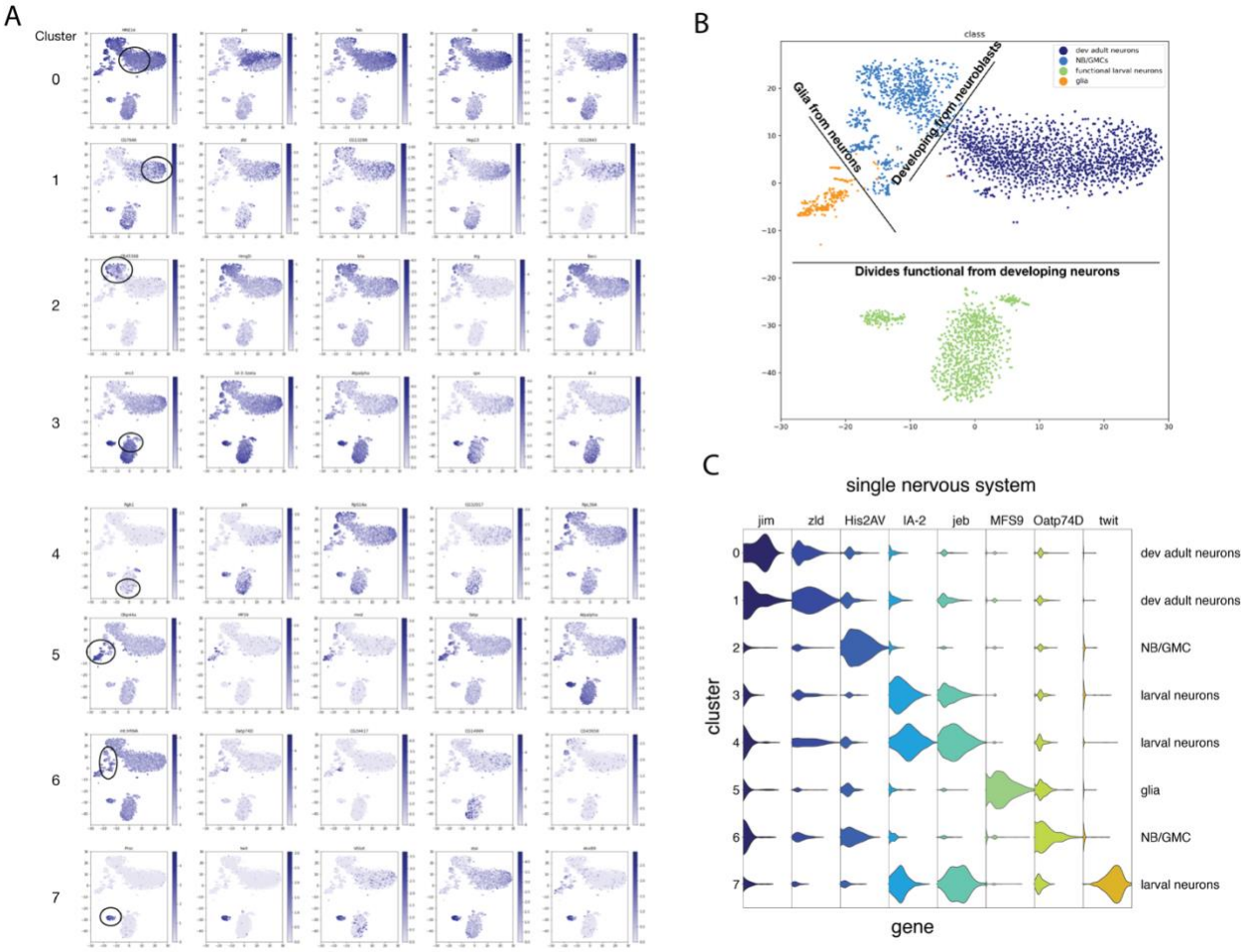

Fig. S4. Complete nervous system atlas from an individual animal.

A. t-SNE of the complete single nervous system with cell clusters colored by gene expression of the top genes in each cluster. For each cluster in A there is a combination of genes which separate the clusters into recognizable molecular cell types and cell classes. B. Lines can be drawn in the t-SNE space that separate each of the cell classes we define here (adult developing neurons, functional larval neurons, neural stem cells, and glia). C. Violin plot of characteristic genes which separate each of the 8 top level clusters.

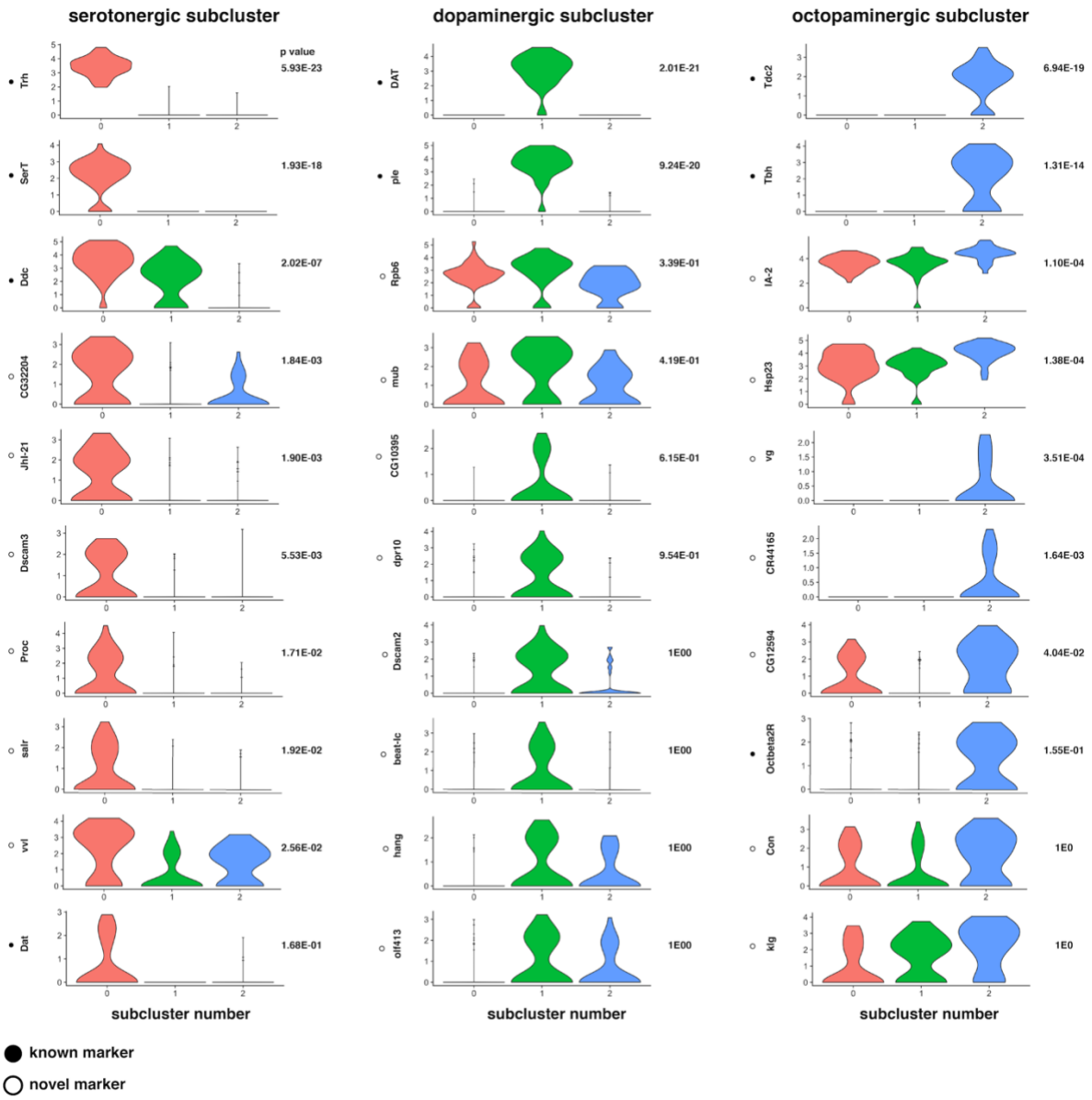

Fig. S5. Unsupervised clustering separates recognizable serotonergic, dopaminergic, octopaminergic neurons into subclusters with known validated markers and novel markers. A cluster of cells was discovered using unsupervised machine learning techniques with markers indicative of dopaminergic cells. This cluster was separated and clustered once more, revealing three separate clusters with gene markers indicating a serotonergic subcluster, a dopaminergic subcluster and an octopaminergic subcluster. The top genes that separate these clusters from one another were then computed and violin plots were generated. Known markers were the top genes for each subcluster and gave them recognizable identity – for example, tryptophan hydroxylase (Trh), which is used to synthesize serotonin from tryptophan, and serotonin transporter (SerT) were the top genes for the first subcluster, which could then be appropriately labeled containing serotonergic cells. In addition to known gene markers, new makers were also discovered for each group of cells.

Table S1. Atlas of cell types in the complete larval nervous system.

The top genes for the 63 cell type clusters and 495 cell type subclusters are provided with the mean expression inside the cluster, mean expression outside the cluster, and p-value. These clusters correspond to the Fig. 2.

Table S2. Fictive sting full nervous system atlas.

The top genes for the 14 cell type clusters obtained from animals that were fictively stung and controls. Tables include the top genes for each cluster, the mean expression inside the cluster, the mean expression outside the cluster, and the p-value. These clusters correspond to Fig. 6E-F.

Table S3. KC overactivation full nervous system atlas.

The top genes for the 14 cell type clusters obtained from animals that had KCs repeatedly activated and controls. Tables include the top genes for each cluster, the mean expression inside the cluster, the mean expression outside the cluster, and the p-value. These clusters correspond to Fig. 6G-I.

Movie S1. Z-stacks through RNA-FISH of AstC in insulin producing cells.

Z stacks are shown through the confocal stack of a brain in which the IPCs are labeled with a fluorescent HaloTag ligand (Magenta) and AstC-R2 mRNA is detected by FISH (green), bar 10µm. This movie corresponds to figure 4 in the main text.

Movie S2. Single molecule imaging to correlate between single molecule FISH and scRNAseq.

Cells were identified with coexpression of AstC (green) and vGlut (magenta) in the whole brain. This movie corresponds to figure 5 in the main text and shows a view through z-stacks of BB-SIM images of AstC and vGlut mRNA FISH channels.