

# Neuronal cell types in the fly: single-cell anatomy meets single-cell genomics

## Authors

Alexander Shakeel Bates<sup>1</sup>, Jasper Janssens<sup>2,3</sup>, Gregory S.X.E. Jefferis<sup>1,4</sup>, Stein Aerts<sup>2,3</sup>

## Affiliations

<sup>1</sup>Neurobiology Division, MRC Laboratory of Molecular Biology, Cambridge, CB2 0QH, UK

<sup>2</sup>VIB Center for Brain & Disease Research, KU Leuven, Leuven 3000, Belgium

<sup>3</sup>Department of Human Genetics KU Leuven, Leuven 3000, Belgium

<sup>4</sup>*Drosophila* Connectomics Group, Department of Zoology, University of Cambridge, Downing Street, Cambridge, CB2 3EJ UK

## Corresponding authors

Gregory Jefferis: [jefferis@mrc-lmb.cam.ac.uk](mailto:jefferis@mrc-lmb.cam.ac.uk)

Stein Aerts: [stein.aerts@kuleuven.vib.be](mailto:stein.aerts@kuleuven.vib.be)

## Highlights

- A neuronal type is a testable hypothesis: similarity indicates functional equivalency.
- Image registration together with morphology algorithms define types across datasets.
- Single-cell transcriptomics produce cell-atlases with consistent clusters.
- Specific genetic lines bridge transcriptomic clusters to morphological types.
- Genetic driver lines, connectomics and transcriptomics make a full *Drosophila* neuronal parts list imminent.

## Abstract

At around 150,000 neurons, the adult *Drosophila melanogaster* central nervous system is one of the largest for which a complete cellular catalogue is imminent. While numerically much simpler than mammalian brains, its complexity is still difficult to parse without grouping neurons into consistent types, which can number 1-1000 cells per hemisphere. We review how neuroanatomical and gene expression data are being used to discover neuronal types at scale. The correlation between multiple co-varying neuronal properties, including lineage, gene expression, morphology, connectivity, response properties and shared behavioural significance is essential to the definition of neuronal cell type. Initial studies comparing morphological and transcriptomic definitions of neuronal type suggest that these are highly consistent, but there is

much to do to match these approaches brain-wide. Matched single-cell transcriptomic and morphological data provide an effective reference point to integrate other data types, including connectomics data. This will significantly enhance our ability to make functional predictions from brain wiring diagrams as well facilitating molecular genetic manipulation of neuronal types.

## Neuronal types as functional units in the brain

Neuronal types are as much about neuroscientists as the nervous system. They help us compare and communicate observations across different experiments and laboratories. Although this point is rarely made explicitly, it is perhaps not particularly controversial. In contrast whether and how neuronal types can be rigorously defined and whether the associated effort is worthwhile are more polarising issues. We argue that formalised approaches to defining neuronal type are essential as we are faced with increasingly complete parts lists for nervous systems. Even at the scale of *D. melanogaster*, there may be ~135,000 neurons in the adult brain, around ~20,000 in the adult ventral nervous system (VNS) (personal communication, D. Dickmann and Y. Huan) and ~10-12,000 in the larval central nervous system [\*1]. Any typing effort at this scale depends on the application of quantitative methodologies to single neuron data, which exists for *D. melanogaster* thanks to ~12,500 genetic lines [2–5], stochastic labelling [6,7], connectomics (e.g. [\*8,\*\*9]) and single-cell RNA sequencing (scRNA-seq) [\*\*10,\*11,\*12] (Figure 1). Currently, computational neuroanatomy based on 3D image data and single-cell transcriptomics are the principal approaches for discovering neuronal types at scale; we will review the key principles and technologies behind both approaches. While it is still rare to have both transcriptomic *and* morphological information at single neuron resolution, we will then cover strategies that allow these data types to be integrated and the benefits that this will bring.

But first, why is type designation useful? In a comparable field (see [13]) that also suffers from intuitive but hard to define types, species taxonomy, central principles for classification include reproductive isolation, phylogeny and phenotypic similarity. As with species, neuronal types are ultimately the reproducible product of definable genomic regulatory programs active during development. We could approximate this developmental ‘aim’ as the average of the cell phenotypes that it produces – to borrow a philosophical idea, its ‘Platonic form’. Each real neuron would be generated from this ideal form by a somewhat noisy but statistically characterisable process. Analogously, the valleys in Waddington's epigenetic landscape, representing stable cell states, have a certain width that represents the allowed variance of a cell state while retaining its unique identity. Neurons of a different type may vary significantly in some features, but still be related by others. It is very helpful to capture this in a hierarchical

scheme that posits relationships between types. These ideas might drive us to consider classes of generative models that would capture the statistical regularities of neuronal types.

Although in phylogeny, reproductive isolation is a hypothesis that can sometimes be tested, it is generally a thought experiment that calibrates definitions based on other measures e.g. genomic sequence similarity. A similar hypothesis for neuroscience might be that:

Neurons of a type can be consistently identified across individuals but not consistently subdivided within individuals.

This does not specify what neuronal features should be used for subdivision, but these could include variations in key functional properties including connectivity, morphology, gene expression and response properties. The degree of 'difference' across multiple features between two animals determines whether they can produce fertile offspring. Similarly, an assessment of 'functional equivalency' between neurons may provide an understanding of the level of feature covariance that is required to define a type. This presently requires expert input and calibration [14], but could become increasingly automated [15,16]. Although challenging, cell typing has the benefit of reducing dimensionality, by mapping many neurons onto a smaller number to units that are developmentally specified or functionally meaningful. This simplifies reasoning about brain architecture and development and facilitates scientific communication.

Neuronal typing could also use the evolutionary and developmental origin of neurons as a grouping mechanism [13,17]. Combining single-cell lineage tracing and transcriptomics [18,19], could eventually provide a completely independent cell type reference tree. Different scientists might prefer a more developmental or functional approach to cell typing based on their area of research. In the fly morphological cell types currently represent the most extensive set of annotated cell types, and also provide a convenient bridge between transcriptome and connectome. However the goal must be to integrate across these layers rather than choose a single approach to defining neuronal types: this will ultimately provide us with the genetic regulatory programmes that define morphotypes, solidify type definitions, make cross-species links, produce actionable resources and further suggest functionality.

## Shared Morphology

Electron microscopy (EM) methods will soon generate a synaptic-resolution 'parts lists' for a small business of flies (perhaps 1-2 larval, 2-3 adult brains and 2 VNS in < 5 years), which can

be compared to extant, annotated light-level morphologies [20] and leveraged to improve our understanding of circuit computations (e.g. [21]). EM morphologies can also be matched to those revealed by genetic labelling [22,23]. Genetic driver lines, especially the intersectional split-GAL4 system, give stable genetic access to small constellations of neuronal types, each of a shared morphology [24–26]. In the best cases they can selectively label just one morphological neuronal type. We therefore expect strong similarities between subdividing neurons by morphology and by gene expression.

Neuronal morphologies are most often extracted from confocal microscopy stacks or EM data as ‘skeletons’, i.e. points in 3D space connected by 1D line segments. This is typically easier than acquiring accurate volumetric reconstructions. Insect neurons typically have highly reproducible spatial locations (e.g.[27]). For co-visualisation and quantitative comparison, skeletons from different brains can be ‘registered’ into a common template space (Figure 1a). Most confocal imaging pipelines image the whole fly brain or VNS with a neuropil counterstain, allowing high accuracy registration; several algorithms can then be used to assess neuronal similarity in this space. *NBLAST* [25] is a fast and sensitive comparison tool that has been validated for expert-annotated neuronal types across datasets. It segments skeletons/image data into vectors and calculates neuron-neuron similarity scores between these vector clouds. Bridging registrations between templates mean that differently registered datasets can be compared, including between sexes and species [28]. The appropriate template might be the whole-brain or a repeated neural structure; there are ~750 visual cartridges and columns that subserve each of the fly’s ommatidia, and repeated circuits within 26 VNS hemi-neuromeres. Indeed, more success has been had in classifying optic column neurons [21] registered into a standard column space [29], than not [30]. However, when working with incomplete dye-fills and registration-incurred spatial-offsets, some manual intervention is needed to best classify types after an unsupervised clustering [14].

Shared morphology alone is highly indicative of shared ontogeny, connectivity, response properties and physiologically-relevant gene expression. *D. melanogaster*’s ~200 central brain [31], ~1000 optic lobe [32], ~800 VNS neuromere [33] neuroblasts consistently create the same neuronal structures and the same, diverse types [34], with some lineages identifiable across taxa (e.g.[35]). Each neuron is unipolar, its single cell body fiber is sent into the neuropil, fasciculating with other neurons of similar developmental origin before bifurcation, in a stereotyped manner [36]. It has been suggested that cell body fascicles form a natural grouping system amenable to exhaustive identification by quantitative methods [36], not least because as a lineage-proxy they may indicate common gene expression [37]. For the same reasons,

internal tracts have been extremely helpful in delineating olfactory projection neuron types across insects [38].

Neurite apposition is suggestive of synaptic connectivity [39], which in turn is indicative of circuit role. Therefore, the overlap of both axon and dendrite in 3D space is often used for cell typing (e.g. [14,40,41,\*\*42]), i.e. a high NBLAST score, as arbors are typically stereotyped within 2-3  $\mu\text{m}$  (adult brain  $\sim 4 \times 10^6 \mu\text{m}^3$ ) [20]. With EM reconstructions from the same single brain, investigators have found mid-to-high axon-dendrite overlap highly predictive of connectivity [43,44], even if they do not always correlate [45,46] (different rules may apply for dendro-dendritic and axo-axonic connections). Clustering neurons by connectivity and NBLAST can produce similar results [\*\*15], but morphology may be a more consistent indicator of neuronal type. Weak connections can account for a large fraction of a connectome [21,47] and can be manifold and variable between identified neurons across hemispheres [\*8], repeated structures [45] or animals [46], perhaps as a product of small-scale activity-dependent structural changes or intrinsic developmental variability.

In terms of response properties, it might be unsurprising that isomorphic neurons close to the sensory periphery respond similarly to stimuli, especially when their dendrites are clearly compartmentalised, for example into glomeruli (e.g. [48]). But crucially, it seems that even in integrative neuropils, neurons of high morphological similarity have stereotyped responses to stimuli and consistent basic electrical properties [\*14].

## Shared transcriptome

Neuronal types are defined by specific genetic programs active in particular developmental settings, which generate unique sets of mRNAs. These can be determined by whole transcriptome profiling. Un-targeted bulk microarray and RNA-seq methods average gene expression over diverse neurons, and have contributed relatively little to the identification of neuronal types in insects [49]. Rather, using expression data to type neurons relies on having a large amounts of single neuronal type, or single-cell expression data (Figure 1b).

*Drosophilid* neurons have a small volume ( $\sim 5$ -500 femtolitre somata), but lack of myelination means insect brain neurons can typically be dissociated without drastic cellular damage. Methods based on droplet microfluidics parallelise scRNA-seq for thousands of cells [50,51] (Figure 1b). These pipelines isolate a cell with a barcoded bead in a droplet, lyse it within the droplet, and then capture mRNA content via polyT-oligos on the bead's surface. This has enabled three cell-atlases, built from dissociating hundreds of mixed-sex *D. melanogaster*

central brains [\*12], optic lobes [\*11], or whole brains [\*\*10] (<http://scope.aertslab.org>) to achieve tens of thousands of high-quality single-cell transcriptomes.

These single-cell profiles cluster into coherent groups, many of which can be linked to known morphological classes by marker gene expression [\*\*10,\*11,\*12] (Figure 1b). Across the cell-atlases, annotated types seemed robust to variations in cell state, though differences in metabolic and activity-regulated expression between neuronal types can be seen [\*\*10,\*12]. Some similar neuronal types seem to converge in their expression profiles after circuit assembly [\*\*52], however a majority of neuronal clusters retain their transcriptomic identity through adult life [\*\*10].

Neuronal type specificity in GAL4 lines is generated by DNA elements that control GAL4 expression, therefore epigenomics could reveal regulatory programs that define type. ChIP-seq has been performed for specific histone modifications on immuno-purified nuclei from octopaminergic and Kenyon cells [53], more recently TaDa on neural stem cells [54]. Chromatin accessibility profiling at single-cell level [55] has been applied to the *D. melanogaster* embryo, in which developing cell types were revealed based on unsupervised clustering of single-cell chromatin accessibility profiles [56].

Neurons from the cell-atlases can already be grouped by proxies for neurotransmitter expression. It might be possible that some types express multiple fast-acting transmitters, and it seems that some neuropeptidergic or aminergic expression correlates/anti-correlates with certain fast-transmitters [\*12]. In depth analysis of receptor expression may reveal the sign of neuronal connections, while expression of e.g. voltage-gated ion channels may reveal aspects of their electrical properties. This provides powerful information about circuit function when linked to morphologies (e.g. [57]), and especially, to connectomes [23].

## **Bridging morphology and transcriptome**

A recent attempt at cell body counting indicates that there are ~135,000 cells in the adult *D. melanogaster* brain, of which ~5,000 are glia (personal communication, Dion Dickmann and Yifu Huan). Independent estimates have suggested ~36-45,000 neurons in the adult's central brain [\*12,31], ~120,000 in the optic lobes [2] and ~80,000 in the VNS. EM indicates the first instar larva has ~10-12,000 neurons [1]. Currently, Virtual Fly Brain (<http://virtualflybrain.org>) records 837 adult and 398 larval neuronal types from the literature (additionally, 158 adult and 443 larval peripheral nervous system types). Golgi staining yielded good type estimates for the repeated cartridges of the adult lamina (~12) and columns of the medulla (~59) [58], repeated ~750-fold. Using ~20,000 skeletons with neurites in the central brain from the *FlyCircuit* dataset

(<http://flycircuit.tw>) and *NBLAST*, followed by affinity propagation [\*\*15], we estimate the number of cell body fascicles that supply neurons to the adult *D. melanogaster* central brain at ~340 and neuronal types at ~3,800 (Figure 2a), an underestimation as *FlyCircuit* is not exhaustive. A central brain type on average may constitute only ~5 neurons per hemisphere, with many types constituting single cells and a few, like Kenyon cells, hundreds.

To understand brain circuits, molecular detail must be overlaid onto these neuronal types (e.g. [23,57]). The largest extant scRNA-seq dataset [\*\*10] is estimated to represent ~1 x coverage of the central brain. Already at such shallow coverage more than one thousand putative transcriptomic neuronal cell clusters can be identified (Figure 2b). Given that most neuronal types contain only a few neurons we predict that at least 10-100 x cell coverage will be needed to identify all morphological types. The whole brain atlas estimates that 45% of brain neurons are cholinergic, 10% GABAergic, 15% glutamatergic [\*\*10] in close agreement with recent whole brain immunostaining (personal communication, D. Dickmann and Y. Huan). It will be hugely beneficial to neural circuits work to identify fast and modulatory neurotransmitters expressed by different neuronal types defined through connectomic work.

In the future, we may be able to directly link morphology and transcriptome by determining transcriptomes in tissue. For example, recent techniques allow capturing mRNA in sectioned tissue and identifying it with gene-specific probes using an imaging-based approach [59] or through sequencing [60]. Currently, more common ways combine bulk or single-cell RNA-seq with the GAL4 system driving expression of a fluorescent label. One can select targeted neurons by manual picking [e.g. 61], bead purification [23,53,57,62], fluorescence activated cell (FAC) sorting [63], which can be followed by plate-based techniques that isolate single neurons [\*\*52], or ribosome-pulldown (with tagged ribosomes used to isolate actively transcribed mRNA [64]). This has shown that many neuronal classes with different morphologies indeed exhibit distinct transcriptomes, revealing marker genes that can identify them across datasets. For example, different mushroom body output neurons (admittedly from different lineages), picked from sparse driver lines and profiled at near single-cell resolution (~4-14 neurons were pooled) can be differentiated by their expression profiles [61]. Interestingly, cell surface signalling genes were sufficient to separate these types; certain gene classes may be preferred for defining type. Split-GAL4 lines can target just a subset of a putative neuronal type as well as off-target neurons; transcriptomic and anatomical work may therefore need to test multiple lines per type as is commonly done in behavioural studies.

More impressively, a scRNA-seq study [\*\*52] retrieved ~1,000 pupal olfactory projection neurons from a driver line labelling ~40 types from just three lineages, and resolved 30 clusters. Interestingly, these clusters were recapitulated by full brain droplet-based scRNA-seq (Figure 1b). It is not clear whether the discrepancy between 30 transcriptomic clusters versus 40

morphological clusters is a product of sampling bias, if the scRNA-seq lacked sufficient sensitivity or whether these clusters actually do map almost exactly onto types defined morphologically. Using sparser drivers and lineage mapping to target specific types, the investigators verified that four glomerulus' projection neurons map one-to-one with a cluster. Two mapped one-to-two, likely because the constituent neurons of both types are produced in two separate waves of neurogenesis [52]. It remains to be seen whether this expression difference is functionally meaningful.

The second main bridge to morphology is using reported expression profiles and marker genes to identify neurons from large-scale datasets without using transgenic labelling of neuronal populations. The authors of [10,11,12] identified many of the best studied neuronal types in their atlases. For example, Kenyon cells can be easily separated into three classes ( $\gamma$ ,  $\alpha'/\beta'$  and  $\alpha/\beta$  lobes) [10,12], but not the finer resolution of seven types ( $\gamma_d$ ,  $\gamma_m$ ,  $\alpha'/\beta'_{ap}$ ,  $\alpha'/\beta'_m$ ,  $\alpha/\beta_p$ ,  $\alpha/\beta_s$  and  $\alpha/\beta_c$ ) described by neuroanatomists [42]. However, these seven types likely have slightly different transcriptomes, as suggested recently by RNA-seq on sorted nuclei [62]. Higher coverage scRNA-seq (e.g. with plate-based techniques) or larger cell numbers, may indeed reveal seven clusters.

These approaches are delivering the first examples of transcriptome-morphology bridges. To achieve higher resolution, statistical and machine-learning approaches like the ones used in the *Drosophila* embryo, may help to extrapolate spatial transcriptome mapping to the entire brain, although it is currently not clear how effective these techniques will be in the brain [65,66]. Another type of additional information to resolve the neuronal type tree relates to cell lineage information. Recent work in *C. elegans* and *D. rerio* has shown that 'genome scarring' using CRISPR-Cas9 during cell divisions can reveal single-cell lineage information, in conjunction with single-cell transcriptomes [18,19,67]. This approach may enable single-cell transcriptomes to be linked to their developmental origin, while also predicting coarse morphological features.

Transcriptomic cell-atlases contain very large clusters of neurons that do not form tight clusters. For example, in [10] the six largest clusters contain 50% of all cells; these are of unknown cell classes although their transmitters differ. There are several possible explanations. First, these large clusters may contain types present in small copy numbers (minimally one cell per hemisphere). Increasing the number of cells would therefore reveal more clusters; indeed downsampling cells from the current atlases shows that the number of *bona fide* clusters decrease [10]. Secondly, neurons of different developmental origins and with different circuit roles may converge onto equivalent adult transcriptomes if their mature function does not depend upon differential gene expression [52].



Finally, rather than only bridging morphology back to the transcriptome, we anticipate bi-directionality in that it will be possible to use single-cell genomics to design genetic driver lines that target groups of neurons defined by cell-atlas clusters. We also envision that scRNA-seq will suggest differentially expressed factors whose 'misexpression' through experimentation might convert one neuronal type, or its key traits, to another's (e.g. [68,69]), or encourage new branching patterns and circuit integration.

## Outlook

*D. melanogaster* is post-genomic. In the next few years it will be post-connectomic and post-transcriptomic at cellular resolution. It is therefore one of very few systems in which comprehensive brain-wide single-cell transcriptomics and morphological analyses can meet. Minimally, we might expect all of its neuronal types to become as well defined as those of the mammalian retina (~100 types) [13] or *C. intestinalis* (52) [70], and eventually *C. elegans* (118) [71]. For these species, morphological classification [70,72,73] matches gene expression and/or physiology well [71,74,75]. Gene expression programs [17], and even certain morphologies can be recognised across related species [76,77].

Of course defining neuronal types is only the beginning: ultimately we hope to make functional predictions based on a synthesis of this information. We can imagine a number of benefits: transcriptomic information could identify factors that specify individual neuronal types, but also the logic by which similar neuronal types – including those originating from a common lineage – are diversified. Likewise candidate wiring specificity factors could be identified by linked expression and connectomics data through genetic driver lines / morphology. Molecular information could also yield circuit hypotheses based on the distribution of transmitter and receptor distributions within a circuit [23].

scRNA-seq has identified up to 200 transcriptomic types within mammalian cortex [13] (*M. musculus*, cortex ~170 mm<sup>3</sup>). Recent work examined two diverse cortical subvolumes, finding 133 transcriptomic types, of which 79 were shared [78]. Perhaps most comparably to the cortex, the mushroom body lobes (volume 1.4x10<sup>-4</sup> mm<sup>3</sup>) only contain 9 intrinsic and 42 extrinsic known types [42,79], however, estimates for regions like the lateral horn (1x10<sup>-4</sup> mm<sup>3</sup>) are already >200 [14]. While it is early days for this field, these initial mammalian estimates suggest that the mean number of cells per type may be much higher than in fly. Of course it is possible – indeed quite likely – that larger studies will identify more cortical neuronal types and also that subcortical brain areas may have a significantly higher density of neuronal types. Smaller, arthropod brains may be unable to expend time and energy developing and maintaining large, highly-plastic neuronal ensembles outside memory systems. They may instead rely more

heavily on information accumulated on an evolutionary timescale and encoded in the genome to determine the wiring of a large number of pre-specified neuronal types even in highly integrative brain regions. Numerous studies have shown that manipulation of types present as single neurons per hemisphere, in larva and in adults, can produce strong behavioural phenotypes. Notably, this is true even of many types that represent neither first order sensory input nor muscle innervation, e.g. in phenotypes for learning and memory [80,81], courtship [82], locomotor control [8,83], etc. Figuring out why there are so many types, and to what degree their roles are unique, necessary or redundant, remains an important future challenge.

The deep conservation revealed by genome sequencing has both underscored and accelerated the impact of *D. melanogaster* work across species. We do not yet know the extent to which its neuronal types and their gene expression programs are conserved, but we are confident that linking molecular and circuit neuroscience approaches to neuronal types will underscore the utility of neuronal type as the organisational unit for neural circuits, as well as the power of fly neurobiology.

## Figure 1 - Anatomical and transcriptomic pipelines for neuronal typing

Pipelines used to cell type *Drosophilid* neurons by **(a)** morphology and **(b)** transcriptomics to create and co-cluster different datasets. For morphology **(a)**, synaptic-scale, nanometer resolution data can be acquired using large scale volume EM, with manual or increasingly semi-automated neuron reconstruction. The volume collected in [9] is estimated to contain 100 m of cable with estimates for total manual reconstruction reaching 500-2000 person-years [20]. Light-level data can be acquired by micrometer confocal images of genetic driver lines registered to a common template brain. For typing purposes, stochastic labelling of neurons from broad GAL4 lines can reveal individual neuron morphologies (e.g. [7]), but sparse lines, usually based on the split GAL4 technique are required for reproducible access [24]. Identifying two parental lines to intersect for the split GAL4 technique can be a significant bottleneck in obtaining selective genetic access. Neuronal morphologies from any image source can be represented as a digital 'skeleton' for further analysis. Datasets registered to different templates, can be quantitatively compared and co-visualised, even between sexes and species, by bridging registrations [28]. NBLAST, or other algorithms, can be used to compare query neurons to annotated libraries to assign identities, or cluster neurons to discover groups and build annotations (see Fig.3). Here, we co-cluster olfactory projection neurons reconstructed from EM data [9] with annotated light-level *FlyCircuit* neurons [15] in order to assign the EM neurons a neuronal type identity, that comes with a wealth of associated information from the literature (a sub-set shown). Dashed lines represent the image registration process. Pink text names different template brains. In **(b)** two different methodologies are shown to generate transcriptomic data. The first technique is based on sorting out known cell populations from the fly brain through the use of a genetic marker, followed by either bulk RNA-seq or well-based single-cell RNA-seq, e.g. SMART-Seq2. This leads to the generation of transcriptomes of well defined populations and neuronal types. The second technique allows for large-scale unbiased single-cell RNA-seq through droplet microfluidics, e.g. Drop-seq, InDrop or 10x Chromium. Through droplet-based techniques atlases of whole brains can be generated. Both techniques are complementary since

transcriptomes from sorted populations can be used to annotate atlases acquired by large-scale single-cell RNA-seq [84]. Clusters of similar cells from the atlas can also be co-clustered with sorted populations, through specific algorithms [85], finding similar clusters in both data sets (data shown from [52] and [10] for olfactory projection neurons).

## Figure 2 - Defining neuronal types for entire brains

We examine pan-brain datasets for *D. melanogaster* neuronal morphologies from *FlyCircuit* and gene expression and [10]. **(a)** NBLAST clustering of 19941 registered neurons with arbor in the central brain, from *FlyCircuit*. We used affinity propagation on just the cell body fibers [15], and hierarchically clustered (by Ward's criterion) 1709 exemplar cell body fibers into 343 clusters (whole neurons and cell body fibers are shown for six clusters), cutting the dendrogram at  $h = 2.23$  to yield cell body fascicles matching those described for the lateral horn [14]. Implementing our proposal to define neuronal types by clustering within-lineage, we give an example of a hierarchical clustering of all neurons of a single estimated cell body fascicle. We cut the dendrogram at  $h = 0.65$ , a cut height that best replicates neuronal types of the lateral horn that have previously been semi-manually annotated in the same dataset [14]. Doing this with each estimated cell body fascicle, yields a neuronal type estimate of 3798 (clusters of >1 skeleton). In **(b)** the t-SNE with annotated clusters from [10] is shown (i), as well as some newly annotated clusters (red dashed boxes) using data from [23]. (ii) Dendrogram showing the hierarchical clustering of mean cells from each cluster based on cosine distance. In total 104 clusters are shown, which can be sub-clustered to a total of 230-1190 clusters. (iii) The dopaminergic neuron cluster from [10] could be further divided in multiple subclusters, of which some express *Fer2*, a known marker for PAM neurons. (iv) Main neurotransmitter expression of the three single-cell atlases generated of the adult fly brain, visualised through SCoPe.

## Declaration of interests

Declarations of interest: none.

## Acknowledgements

We thank D. Osumi-Sutherland for cell type queries of <http://virtualflybrain.org>, M-J. Dolan for providing confocal microscopy exemplar images, P. Schlegel and M. Costa for providing olfactory projection neuron morphological clustering and D. Dickmann, Y. Huan and J. Tuthill for neuron number estimates. We thank S. Frechter for interesting discussions, and M-J. Dolan, S. Liu and D. Pappasokrati for comments on this manuscript. This work was supported by the Wellcome Trust (203261/Z/16/Z), an ERC Consolidator grant (649111), the Boehringer Ingelheim Fonds and Herchel Smith Scholarship (A.S.B.), a PhD fellowship of The Research Foundation – Flanders (J.J., 1199518N) and core support from the MRC (MC-U105188491) to G.S.X.E.J. We apologise to those authors whose instrumental primary work has been cited through later articles or reviews, due to citation limitations.

## References

1. 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.
2. Jenett A, Rubin GM, Ngo T-TB, Shepherd D, Murphy C, Dionne H, Pfeiffer BD, Cavallaro A, Hall D, Jeter J, et al.: **A GAL4-driver line resource for *Drosophila* neurobiology**. *Cell Rep* 2012, **2**:991–1001.
3. Manning L, Heckscher ES, Purice MD, Roberts J, Bennett AL, Kroll JR, Pollard JL, Strader ME, Lupton JR, Dyukareva AV, et al.: **A resource for manipulating gene expression and analyzing cis-regulatory modules in the *Drosophila* CNS**. *Cell Rep* 2012, **2**:1002–1013.
4. Li HH, Kroll JR, Lennox SM, Ogundeyi O, Jeter J, Depasquale G, Truman JW: **A GAL4 driver resource for developmental and behavioral studies on the larval CNS of *Drosophila***. *Cell Rep* 2014, **8**:897–908.
5. Tirian L, Dickson B: **The VT GAL4, LexA, and split-GAL4 driver line collections for targeted expression in the *Drosophila* nervous system**. *bioRxiv* 2017, doi:10.1101/198648.
6. Chiang A-S, Lin C-Y, Chuang C-C, Chang H-M, Hsieh C-H, Yeh C-W, Shih C-T, Wu J-J, Wang G-T, Chen Y-C, et al.: **Three-dimensional reconstruction of brain-wide wiring networks in *Drosophila* at single-cell resolution**. *Curr Biol* 2011, **21**:1–11.
7. Nern A, Pfeiffer BD, Rubin GM: **Optimized tools for multicolor stochastic labeling reveal diverse stereotyped cell arrangements in the fly visual system**. *Proc Natl Acad Sci U S A* 2015, **112**:E2967–76.
8. 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.
9. Zheng Z, Lauritzen JS, Perlman E, Saalfeld S, Fetter RD, Bock Correspondence DD: **A Complete Electron Microscopy Volume of the Brain of Adult *Drosophila melanogaster***. *Cell* 2018, **174**:1–14.
10. Davie K, Janssens J, Koldere D, De Waegeneer M, Pech U, Kreft Ł, Aibar S, Makhzami S, Christiaens V, González-Blas CB, et al.: **A Single-Cell Transcriptome Atlas of the Aging *Drosophila* Brain**. *Cell* 2018, **174**:982–998.e20.
11. Konstantinides N, Kapuralin K, Fadil C, Barboza L, Satija R, Desplan C: **Phenotypic Convergence: Distinct Transcription Factors Regulate Common Terminal Features**. *Cell* 2018, **174**:622–635.e13.
12. Croset V, Treiber C, Waddell S: **Cellular diversity in the *Drosophila* midbrain revealed by single-cell transcriptomics**. *Elife* 2018, **7**:e34550.
13. Zeng H, Sanes JR: **Neuronal cell-type classification: Challenges, opportunities and the path forward**. *Nat Rev Neurosci* 2017, **18**:530–546.
14. Frechter S, Bates AS, Tootoonian S, Dolan M-J, Manton JD, Jamasb A, Kohl J, Bock D, Gregory S X: **Functional and Anatomical Specificity in a Higher Olfactory Centre**. *bioRxiv* 2018, doi:10.1101/336982.

15. Costa M, Manton JD, Ostrovsky AD, Prohaska S, Jefferis GSXE: **NBLAST: Rapid, Sensitive Comparison of Neuronal Structure and Construction of Neuron Family Databases.** *Neuron* 2016, **91**:293–311.
16. Armañanzas R, Ascoli GA: **Towards the automatic classification of neurons.** *Trends Neurosci* 2015, **38**:307–318.
17. Arendt D, Musser JM, Baker CVH, Bergman A, Cepko C, Erwin DH, Pavlicev M, Schlosser G, Widder S, Laubichler MD, et al.: **The origin and evolution of cell types.** *Nat Rev Genet* 2016, **17**:744–757.
18. Spanjaard B, Hu B, Mitic N, Olivares-Chauvet P, Janjuha S, Ninov N, Junker JP: **Simultaneous lineage tracing and cell-type identification using CRISPR-Cas9-induced genetic scars.** *Nat Biotechnol* 2018, **36**:469–473.
19. Raj B, Wagner DE, McKenna A, Pandey S, Klein AM, Shendure J, Gagnon JA, Schier AF: **Simultaneous single-cell profiling of lineages and cell types in the vertebrate brain.** *Nat Biotechnol* 2018, **36**:442–450.
20. Schlegel P, Costa M, Jefferis GS: **Learning from connectomics on the fly.** *Curr Opin Insect Sci* 2017, **24**:96–105.
21. Takemura S-Y, 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.
22. Dolan M-J, Frechter S, Bates AS, Dan C, Huoviala P, Roberts RJV, Schlegel P, Dhawan S, Tabano R, Dionne H, et al.: **Neurogenetic dissection of the Drosophila innate olfactory processing center.** *bioRxiv* 2018, doi:10.1101/404277.
23. Davis FP, Nern A, Picard S, Reiser MB, Rubin GM, Eddy SR, Henry GL: **A genetic, genomic, and computational resource for exploring neural circuit function.** *bioRxiv* 2018, doi:10.1101/385476.
24. Luan H, Peabody NC, Vinson CR, White BH: **Refined spatial manipulation of neuronal function by combinatorial restriction of transgene expression.** *Neuron* 2006, **52**:425–436.
25. Dionne H, Hibbard KL, Cavallaro A, Kao J-C, Rubin GM: **Genetic Reagents for Making Split-GAL4 Lines in Drosophila.** *Genetics* 2018, **209**:31–35.
26. Pfeiffer BD, Ngo T-TB, Hibbard KL, Murphy C, Jenett A, Truman JW, Rubin GM: **Refinement of tools for targeted gene expression in Drosophila.** *Genetics* 2010, **186**:735–755.
27. Jefferis GSXE, Potter CJ, Chan AM, Marin EC, Rohlifing T, Maurer CR Jr, Luo L: **Comprehensive maps of Drosophila higher olfactory centers: spatially segregated fruit and pheromone representation.** *Cell* 2007, **128**:1187–1203.
28. Manton JD, Ostrovsky AD, Goetz L, Costa M, Rohlifing T, Gregory S X: **Combining genome-scale Drosophila 3D neuroanatomical data by bridging template brains.** *bioRxiv* 2014, doi:10.1101/006353.
29. Zhao T, Plaza SM: **Automatic Neuron Type Identification by Neurite Localization in the Drosophila Medulla.** *arXiv [q-bioNC]* 2014,
30. Li Y, Wang D, Ascoli GA, Mitra P, Wang Y: **Metrics for comparing neuronal tree shapes based on persistent homology.** *PLoS One* 2017, **12**:1–24.
31. Ito K, Hotta Y: **Proliferation pattern of postembryonic neuroblasts in the brain of Drosophila**

**melanogaster**. *Dev Biol* 1992, **149**:134–148.

32. Hofbauer A, Campos-Ortega JA: **Proliferation pattern and early differentiation of the optic lobes in *Drosophila melanogaster***. *Roux Arch Dev Biol* 1990, **198**:264–274.
33. Birkholz O, Rickert C, Berger C, Urbach R, Technau GM: **Neuroblast pattern and identity in the *Drosophila* tail region and role of doublesex in the survival of sex-specific precursors**. *Development* 2013, **140**:1830–1842.
34. Yu H-H, Awasaki T, Schroeder MD, Long F, Yang JS, He Y, Ding P, Kao J-C, Wu GY-Y, Peng H, et al.: **Clonal Development and Organization of the Adult *Drosophila* Central Brain**. *Curr Biol* 2013, **23**:633–643.
35. Boyan G, Liu Y, Khalsa SK, Hartenstein V: **A conserved plan for wiring up the fan-shaped body in the grasshopper and *Drosophila***. *Dev Genes Evol* 2017, **227**:253–269.
36. Cardona A, Saalfeld S, Arganda I, Pereanu W, Schindelin J, Hartenstein V: **Identifying neuronal lineages of *Drosophila* by sequence analysis of axon tracts**. *J Neurosci* 2010, **30**:7538–7553.
37. Yang C-P, Fu C-C, Sugino K, Liu Z, Ren Q, Liu L-Y, Yao X, Lee LP, Lee T: **Transcriptomes of lineage-specific *Drosophila* neuroblasts profiled by genetic targeting and robotic sorting**. *Development* 2016, **143**:411–421.
38. Ito K, Shinomiya K, Ito M, Armstrong JD, Boyan G, Hartenstein V, Harzsch S, Heisenberg M, Homberg U, Jenett A, et al.: **A systematic nomenclature for the insect brain**. *Neuron* 2014, **81**.
39. Rees CL, Moradi K, Ascoli GA: **Weighing the Evidence in Peters' Rule: Does Neuronal Morphology Predict Connectivity?** *Trends Neurosci* 2017, **40**:63–71.
40. Jeanne JM, Fişek M, Wilson RI: **The Organization of Projections from Olfactory Glomeruli onto Higher-Order Neurons**. *Neuron* 2018, **98**:1198–1213.e6.
41. Lin C-Y, Chuang C-C, Hua T-E, Chen C-C, Dickson BJ, Greenspan RJ, Chiang A-S: **A comprehensive wiring diagram of the protocerebral bridge for visual information processing in the *Drosophila* brain**. *Cell Rep* 2013, **3**:1739–1753.
42. Aso Y, Hattori D, Yu Y, Johnston RM, Nirmla A, Ngo T-T, Dionne H, Abbott LF, Axel R, Tanimoto H, et al.: **The neuronal architecture of the mushroom body provides a logic for associative learning**. *Elife* 2014,
43. Gerhard S, Andrade I, Fetter RD, Cardona A, Schneider-Mizell CM: **Conserved neural circuit structure across *Drosophila* larval development revealed by comparative connectomics**. *Elife* 2017, **6**:e29089.
44. Tobin WF, Wilson RI, Lee W-CA: **Wiring variations that enable and constrain neural computation in a sensory microcircuit**. *Elife* 2017, **6**:e24838.
45. Takemura S-Y, Xu CS, Lu Z, Rivlin PK, Parag T, Olbris DJ, Plaza S, Zhao T, Katz WT, Umayam L, et al.: **Synaptic circuits and their variations within different columns in the visual system of *Drosophila***. *Proc Natl Acad Sci U S A* 2015, **112**:13711–13716.
46. Couton L, Mauss AS, Yunusov T, Diegelmann S, Evers JF, Landgraf M: **Development of connectivity in a motoneuronal network in *Drosophila* larvae**. *Curr Biol* 2015, **25**:568–576.
47. Varshney LR, Chen BL, Paniagua E, Hall DH, Chklovskii DB: **Structural Properties of the *Caenorhabditis elegans* Neuronal Network**. *PLoS Comput Biol* 2011, **7**.
48. Badel L, Ohta K, Tsuchimoto Y, Kazama H: **Decoding of Context-Dependent Olfactory Behavior in *Drosophila***. *Neuron* 2016, **91**:155–167.

49. Trapnell C: **Defining cell types and states with single-cell genomics.** *Genome Res* 2015, **25**:1491–1498.
50. Macosko EZ, Basu A, Satija R, Nemesh J, Shekhar K, Goldman M, Tirosh I, Bialas AR, Kamitaki N, Martersteck EM, et al.: **Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets.** *Cell* 2015, **161**:1202–1214.
51. Klein AM, Mazutis L, Akartuna I, Tallapragada N, Veres A, Li V, Peshkin L, Weitz DA, Kirschner MW: **Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells.** *Cell* 2015, **161**:1187–1201.
52. Li H, Horns F, Wu B, Xie Q, Li J, Li T, Luginbuhl DJ, Quake SR, Luo L: **Classifying Drosophila Olfactory Projection Neuron Subtypes by Single-Cell RNA Sequencing.** *Cell* 2017, **171**:1206–1220.e22.
53. Henry GL, Davis FP, Picard S, Eddy SR: **Cell type-specific genomics of Drosophila neurons.** *Nucleic Acids Res* 2012, **40**:9691–9704.
54. Marshall OJ, Brand AH: **Chromatin state changes during neural development revealed by in vivo cell-type specific profiling.** *Nat Commun* 2017, **8**:2271.
55. Buenrostro JD, Wu B, Litzenburger UM, Ruff D, Gonzales ML, Snyder MP, Chang HY, Greenleaf WJ: **Single-cell chromatin accessibility reveals principles of regulatory variation.** *Nature* 2015, **523**:486–490.
56. Cusanovich DA, Reddington JP, Garfield DA, Daza RM, Aghamirzaie D, Marco-Ferrerres R, Pliner HA, Christiansen L, Qiu X, Steemers FJ, et al.: **The cis-regulatory dynamics of embryonic development at single-cell resolution.** *Nature* 2018, **555**:538–542.
57. Pankova K, Borst A: **RNA-seq transcriptome analysis of direction-selective T4/T5 neurons in Drosophila.** *PLoS One* 2016, **11**:1–13.
58. Fischbach K-F, Dittrich APM: **The optic lobe of Drosophila melanogaster. I. A Golgi analysis of wild-type structure.** *Cell Tissue Res* 1989, **258**.
59. Wang X, Allen WE, Wright MA, Sylwestrak EL, Samusik N, Vesuna S, Evans K, Liu C, Ramakrishnan C, Liu J, et al.: **Three-dimensional intact-tissue sequencing of single-cell transcriptional states.** *Science* 2018, **361**.
60. Ståhl PL, Salmén F, Vickovic S, Lundmark A, Navarro JF, Magnusson J, Giacomello S, Asp M, Westholm JO, Huss M, et al.: **Visualization and analysis of gene expression in tissue sections by spatial transcriptomics.** *Science* 2016, **353**:78–82.
61. Crocker A, Guan XJ, Murphy CT, Murthy M: **Cell-Type-Specific Transcriptome Analysis in the Drosophila Mushroom Body Reveals Memory-Related Changes in Gene Expression.** *Cell Rep* 2016, **15**:1580–1596.
62. Shih M-FM, Davis FP, Henry GL: **Nuclear transcriptomes of the seven neuronal cell types that constitute the Drosophila mushroom bodies.** *bioRxiv* 2018, doi:10.1101/412569.
63. Tan L, Zhang KX, Pecot MY, Nagarkar-Jaiswal S, Lee PT, Takemura SY, McEwen JM, Nern A, Xu S, Tadros W, et al.: **Ig Superfamily Ligand and Receptor Pairs Expressed in Synaptic Partners in Drosophila.** *Cell* 2015, **163**:1756–1769.
64. Thomas A, Lee P-J, Dalton JE, Nomie KJ, Stoica L, Costa-Mattioli M, Chang P, Nuzhdin S, Arbeitman MN, Dierick HA: **A versatile method for cell-specific profiling of translated mRNAs in Drosophila.** *PLoS One* 2012, **7**:e40276.
65. Karaiskos N, Wahle P, Alles J, Boltengagen A, Ayoub S, Kipar C, Kocks C, Rajewsky N, Zinzen RP:

**The *Drosophila* embryo at single-cell transcriptome resolution.** *Science* 2017, **3235**:eaan3235.

66. Nitzan M, Karaïskos N, Friedman N, Rajewsky N: **Charting a tissue from single-cell transcriptomes.** *bioRxiv* 2018, doi:10.1101/456350.
67. Alemany A, Florescu M, Baron CS, Peterson-Maduro J, van Oudenaarden A: **Whole-organism clone tracing using single-cell sequencing.** *Nature* 2018, **556**:108–112.
68. Tea JS, Chihara T, Luo L: **Histone deacetylase Rpd3 regulates olfactory projection neuron dendrite targeting via the transcription factor Prospero.** *J Neurosci* 2010, **30**:9939–9946.
69. Kohl J, Ostrovsky AD, Frechter S, Jefferis GSXE: **A Bidirectional Circuit Switch Reroutes Pheromone Signals in Male and Female Brains.** *Cell* 2013, **155**:1610–1623.
70. Ryan K, Lu Z, Meinertzhagen IA: **The CNS connectome of a tadpole larva of *Ciona intestinalis* (L.) highlights sidedness in the brain of a chordate sibling.** *Elife* 2016, **5**:e16962.
71. Hobert O, Glenwinkel L, White J: **Revisiting Neuronal Cell Type Classification in *Caenorhabditis elegans*.** *Curr Biol* 2016, **26**:R1197–R1203.
72. 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.
73. Albertson DG, Thomson JN: **The pharynx of *Caenorhabditis elegans*.** *Philos Trans R Soc Lond B Biol Sci* 1976, **275**:299–325.
74. Seung HS, Sümbül U: **Neuronal cell types and connectivity: lessons from the retina.** *Neuron* 2014, **83**:1262–1272.
75. Sharma S, Wang W, Stolfi A: **Single-cell transcriptome profiling of the *Ciona* larval brain.** *bioRxiv* 2018, doi:10.1101/319327.
76. Nelson SB, Sugino K, Hempel CM: **The problem of neuronal cell types: a physiological genomics approach.** *Trends Neurosci* 2006, **29**:339–345.
77. Bumbarger DJ, Riebesell M, Rödelsperger C, Sommer RJ: **System-wide rewiring underlies behavioral differences in predatory and bacterial-feeding nematodes.** *Cell* 2013, **152**:109–119.
78. Tasic B, Yao Z, Graybuck LT, Smith KA, Nguyen TN, Bertagnolli D, Goldy J, Garren E, Economo MN, Viswanathan S, et al.: **Shared and distinct transcriptomic cell types across neocortical areas.** *Nature* 2018, **563**:72–78.
79. Takemura S-Y, Aso Y, Hige T, Wong A, Lu Z, Xu CS, Rivlin PK, Hess H, Zhao T, Parag T, et al.: **A connectome of a learning and memory center in the adult *Drosophila* brain.** *Elife* 2017, **6**:e26975.
80. Saumweber T, Rohwedder A, Schleyer M, Eichler K, Chen Y-C, Aso Y, Cardona A, Eschbach C, Kobler O, Voigt A, et al.: **Functional architecture of reward learning in mushroom body extrinsic neurons of larval *Drosophila*.** *Nat Commun* 2018, **9**:1104.
81. Aso Y, Sitaraman D, Ichinose T, Kaun KR, Vogt K, Belliard-Guérin G, Plaçais P-Y, Robie AA, Yamagata N, Schnaitmann C, et al.: **Mushroom body output neurons encode valence and guide memory-based action selection in *Drosophila*.** *Elife* 2014, **3**:e04580.
82. von Philipsborn AC, Liu T, Yu JY, Masser C, Bidaye SS, Dickson BJ: **Neuronal control of *Drosophila* courtship song.** *Neuron* 2011, **69**:509–522.
83. Bidaye SS, Machacek C, Wu Y, Dickson BJ: **Neuronal control of *Drosophila* walking direction.** *Science* 2014, **344**:97–101.



84. Kiselev VY, Yiu A, Hemberg M: **scmap: projection of single-cell RNA-seq data across data sets.** *Nat Methods* 2018, **15**:359–362.
85. Butler A, Hoffman P, Smibert P, Papalexi E, Satija R: **Integrating single-cell transcriptomic data across different conditions, technologies, and species.** *Nat Biotechnol* 2018, **36**:411–420.



