Cellular diversity in the *Drosophila* midbrain revealed by single-cell transcriptomics

Vincent Croset[1], Christoph D Treiber[1,2], Scott Waddell[2]

Centre for Neural Circuits and Behaviour, The University of Oxford, Tinsley Building, Mansfield Road, Oxford, OX1 3SR, UK

[1] These authors contributed equally to this work

[2] To whom correspondence should be addressed: christoph.treiber@gmail.com, scott.waddell@cncb.ox.ac.uk

Abstract

To understand the brain, molecular details need to be overlaid onto neural wiring diagrams so that synaptic mode, neuromodulation and critical signaling operations can be considered. Single-cell transcriptomics provide a unique opportunity to collect this information. Here we present an initial analysis of thousands of individual cells from *Drosophila* midbrain, that were acquired using Drop-Seq. A number of approaches permitted the assignment of transcriptional profiles to several major brain regions and cell-types. Expression of biosynthetic enzymes and reuptake mechanisms allows all the neurons to be typed according to the neurotransmitter or neuromodulator that they produce and presumably release. Some neuropeptides are preferentially co-expressed in neurons using a particular fast-acting transmitter, or monoamine. Neuromodulatory and neurotransmitter receptor subunit expression illustrates the potential of these molecules in generating complexity in neural circuit function. This cell atlas dataset provides an important resource to link molecular operations to brain regions and complex neural processes.

Introduction

Neuroscience is typically studied at the systems, cellular, or molecular level. However, it will be necessary to bridge these traditional boundaries to fully understand how the brain operates. Such a momentous task is somewhat simplified if analyses are focused on an animal with a relatively small brain, but where systems-level processes are evident. In many respects, the vinegar fly *Drosophila melanogaster* fits the bill (Haberkern and Jayaraman, 2016). *Drosophila* have an estimated 150,000 neurons in the entire brain, of which the optic lobes, or visual neuropils, comprise two thirds of this neural mass. The remaining approximately 50,000 neurons, or midbrain, houses many key neural structures such as the mushroom bodies and central complex, which are, amongst other things, critical for memory-directed behavior (Cognigni et al., 2018) and navigation (Seelig and Jayaraman, 2015), respectively.

Recent large-scale electron-microscopy projects have generated wiring diagrams, or connectomes, of parts of the larval and adult fly nervous system (Berck et al., 2016; Eichler et al., 2017; Ohyama et al., 2015; Takemura et al., 2013, 2017a, 2017b; Tobin et al., 2017; Zheng et al., 2017). While these efforts are an essential part of the quest to decipher brain function, they are not enough. Genes determine the anatomy and mode of connectivity, the biophysical properties, and the information-processing limits of individual constituent neurons. Therefore, understanding any given wiring diagram requires a systematic view of gene expression within their functionally relevant cellular context. With this knowledge in hand, investigators can begin to examine how gene products contribute to cell- and circuit-specific functions and, ultimately, organismal behavior.

New developments in single-cell sequencing technology provide a unique means to generate such a brain-wide view of gene expression with cellular resolution. Massively parallel approaches, such as Drop-seq (Macosko et al., 2015), permit simultaneous analysis of the transcriptomes of 1000s of individual cells. In brief, each cell from a dissociated tissue is first captured with an oligonucleotide bar-coded bead in a nanoliter aqueous droplet. The bar-coding allows all of the mRNAs expressed in a cell to be assigned to the same cell, and for the identity of the cell to be remembered. Following this critical cell-specific hybridization step, all the material from 1000s of individual cells can be pooled and processed together for mRNA sequencing. Drop-seq therefore provides the means to access the transcriptomes of a representation of most cells in the fly midbrain.

A key hurdle in generating a single-cell atlas of the brain is the ability to assign individual transcriptome profiles to the correct cell, or at least cell-type. Again, using an animal whose brain has an intermediate number of neurons and presumably neural diversity simplifies the task. Moreover, years of genetic analyses in *Drosophila* have provided a considerable number of established transgenic and intrinsic markers for specific brain regions and cell-types. These identifiers often allow one to extract the relevant cell profiles from the larger dataset.

Here we report the application and an initial analysis of Drop-seq data to investigate the cellular diversity of the *Drosophila* midbrain. We demonstrate the ability to assign many single-cell profiles to identified cell-types and brain regions, and identify novel markers for these regions. Moreover, cells can be robustly classified based on their neurotransmitter (NT) profile. We find that certain neuropeptides preferentially accompany particular fast-acting transmitters, or monoamines. In addition, we detail the apparent complexity of modulatory and NT receptor subunit expression. This single-cell dataset provides an indication of the extent of neural diversity in the fly brain, and provides essential cellular context linking molecules to neural circuits and brain function.

Results

Drop-seq analysis of the *Drosophila* midbrain

We first optimized the conditions required to effectively dissociate and capture individual *Drosophila melanogaster* cells with DNA bar-coded microparticles in aqueous droplets, using a commercially available apparatus. *Drosophila* neurons are about a tenth of the size of mammalian cells. We therefore first verified the efficiency of processing insect cells and of single-cell capture by generating single-cell transcriptomes attached to microparticles (STAMPs) from a cell suspension comprised of a 1:1 mixture of *Drosophila* S2 and *Spodoptera frugiperda* (fall armyworm) Sf9 cultured cells. We then sequenced these S2/Sf9 STAMPs (Fig. 1 - figure supplement 1A). This procedure retrieved 764 barcode-associated transcriptomes, of which 368 were identifiable as *Drosophila* and 384 as *Spodoptera*. Importantly, only 12 transcriptomes contained cDNA coming from both species (Fig. 1 - figure supplement 1B), indicating that only 3.1% of all sequenced transcriptomes resulted from capturing two cells together. This analysis suggested that the Drop-seq system and our chosen parameters are suitable for barcoding single insect cells and are optimized to minimize capture of cell doublets.

We next used these same parameters to collect STAMPs from thousands of cells from the *Drosophila* midbrain, in eight independent experiments, over eight different days. Each day we isolated single-cells from 80-100 dissected brains taken from an equal number of male and female flies. Brains were removed from the head capsule, optic lobes were manually dissected away and a single-cell suspension was prepared from the remaining fly midbrains. Cells were then individually paired with DNA barcoded beads and cDNA libraries were generated from bead-bound single-cell transcriptomes, and sequenced (Fig. 1A) (Macosko et al., 2015). Pooling the data from the eight independent experiments resulted in a dataset of 19,260 cells, with each containing between 200 and 10,000 unique molecular identifiers (UMIs) and therefore, single mRNA molecules. We performed a Principle-Component Analysis (PCA) on these transcriptomes and reduced the top 50 PCs into two dimensions using t-SNE (Van Der Maaten, 2014) (Fig. 1 - figure supplement 2A). We selected the cut-off for the optimal number of UMIs per cell to be included in our analyses by generating t-SNE plots from data with a variety of quality thresholds. These analyses revealed that discarding cells with less than 800 UMIs, resulted in a data set of 10,286 high quality cells, segregated with k-means clustering into 29 cell clusters, with several corresponding to most of the known iterative cell types in the

Drosophila brain (Fig. 1 - figure supplement 2B). More stringent criteria decreased the number of cells included without further improving the clustering (Fig. 1 - figure supplement 2C). A comparison between our eight individual replicate experiments revealed that all of them contributed equally to all but one cluster (Fig. 1 - figure supplement 3). We therefore chose to use the 10,286 cells that have ≥800 UMIs from our eight pooled replicates for our subsequent analyses.

We assessed the transcript drop-out rate, by determining the number of cells that can be seen to express the male-specific long non-coding *RNA on the X 1* (*roX1*) gene (Kelley and Kuroda, 2003). The distribution of UMIs for this gene was biphasic, with one peak at 0, and another at 9 UMIs (Fig. 1 - figure supplement 4A). Since our data was prepared from an equal number of male and female brains we reasoned that these two populations must represent cells from female and male flies, respectively. We used the median between the two peaks (4.5) as cut-off to separate these two populations which revealed that 43.1% of all the cells express *roX1* and therefore can be considered male (Fig 1 - figure supplement 4B). Since this number is close to the expected 50%, this distribution suggests that transcript drop-out is low in our high-quality dataset of 10,286 cells. Interestingly, the neural Cluster J and Glial cluster 2 are almost exclusively comprised of *roX1*-negative cells, suggesting that they may contain cells that are only present in the female brain.

We manually annotated 28 clusters in the t-SNE plot of 10,286 cells, with each containing between 9 and 7167 cells (Fig. 1B). We assigned cell identity to a number of cell clusters according to the markers in Supplementary File 2 (which contains the list of genes that are significantly over-expressed in each of these clusters, compared to all others); the mushroom body (MB) Kenyon Cells (KCs), olfactory projection neurons (PNs), ellipsoid body (EB) ring neurons, monoaminergic neurons, astrocytes and other glia, and insulin producing cells (IPCs). We also identified a few cells from the ocelli, in addition to fat body tissue, some of which is present in the head capsule and therefore is also expected to be included in our dissected brain tissue. We also identified 12 additional cell clusters that we could not at this time assign to a particular neural type, and that we name with the letters A-L. Surprisingly, cluster G only contained cells obtained from a single replicate experiment (Fig. 1 - figure supplement 3). The largest cluster of all contains 7167 cells with a variety of expression profiles, that at this stage of analysis, we marked as “unannotated”, but that can nevertheless be segregated for example, based on their primary fast-acting NT (see below, and Fig. 4).

Identification of mushroom body Kenyon Cells

The easiest and most certain way to assign a Drop-seq cluster to a specific cell-type is to track the expression of a transgenically expressed marker. For this reason, our single-cell expression dataset, was generated from a genotype of flies that express an mCherry transgene specifically in the αβ subset of MB KCs (Fig. 2A). In addition, we have previously deep-sequenced the genome of this strain at high coverage (Treiber and Waddell, 2017) which enabled more precise mapping of Drop-seq reads to the genome of this particular fly strain. To our surprise, visualizing mCherry expression levels in our dataset revealed labeling of a very distinct group of cells (Fig. 2B), that allowed us to assign this cluster to αβ KCs.

The MB is a brain structure that is critical for olfactory learning and memory (de Belle and Heisenberg, 1994; Cognigni et al., 2018; Heisenberg, 2003) and it is comprised of three main classes of neurons, the αβ, α′β′ and γ neurons, that are morphologically unique and have dissociable roles in memory processing and expression (Bouzaiane et al., 2015; Perisse et al., 2013). We first identified the α′β′ and γ KC types, using the expression of the previously known general KC markers *eyeless* and *Dop1R2* (aka *Dopamine receptor in mushroom bodies*, *damb*) (Han et al., 1996; Kurusu et al., 2000). Cells expressing these two markers were contained within three distinct clusters, including the αβ cluster identified as expressing mCherry (Fig. 2C-D). The αβ and γ KCs have previously been shown to be distinguishable from the α′β′ neurons using the expression of molecular markers. The αβ and γ KCs express *short neuropeptide F precursor* (*sNPF*) (Johard et al., 2008) and *Fasciclin 2* (*Fas2*) (Cheng et al., 2001; Crittenden et al., 1998), whilst α′β′ and γ KCs express the rho guanyl-nucleotide exchange factor gene *trio* (Awasaki et al., 2000). The expression patterns of these three genes, permitted us to assign each KC cluster to one of these KC subtypes (Fig. 2E). Furthermore, we identified 26 additional genes that are differentially expressed between KC subtypes (Fig. 2F). Of these, ten have roles in gene regulation, five in signal transduction, and three in synapse function.

Identification of olfactory projection neurons

We assigned two cell clusters to PNs (Fig. 1B), based on the strong expression of two previously described markers, *cut* (*ct*) and *abnormal chemosensory jump 6* (*acj6*). The *ct* gene encodes a homeobox transcription factor involved in dendrite targeting in PNs and is known to be expressed in a subset of the antero-dorsal (ad-), lateral (l-) and ventral (v-) PNs (Komiyama and Luo, 2007). The *acj6* gene encodes a POU-domain transcription factor that is also necessary for PN development and has been described to label all adPNs and a subset of lPNs (Komiyama et al., 2003; Lai et al., 2008) (Fig. 3C). We next isolated the cells from these clusters and performed a new PCA and t-SNE analysis on the top six PCs. PNs segregated into four distinct clusters, each of which expresses a specific transcriptional signature (Fig. 3 A-B). Consistent with the expression patterns mentioned above, *ct* transcripts were found in all four PN clusters, whereas *acj6* was only identified in three of them (Clusters 1, 2 and 4; Fig. 3C). Interestingly, *ventral veins lacking* (*vvl*), another POU-domain transcription factor reported to be expressed in *acj6*-negative lPNs (Komiyama et al., 2003; Li et al., 2017) only labeled a small number of neurons, which were all part of the cluster that was negative for *acj6* (Cluster 3; Fig. 3C). Our data therefore confirm the non-overlapping expression patterns of *acj6* and *vvl*, and support the assignment of the *vvl* expressing cluster to cells including the lPNs.

To identify putative ventral PNs (vPNs), we used expression of *Lim1*, which encodes a LIM-homeodomain transcription factor reported to be expressed in most vPNs, but not in adPNs or lPNs (Komiyama and Luo, 2007; Li et al., 2017). Surprisingly, *Lim1* labeled one of the three *acj6*-positive clusters, and several neurons co-expressed both *Lim1* and *acj6* (Cluster 4; Fig. 3C). This contrasts with a previous study that indicated that *acj6* and *Lim1* expression does not overlap, as a result of these two genes being expressed in progeny derived from discrete PN neuroblasts (Komiyama and Luo, 2007). About 50% of the *acj6*-positive neurons were recently shown to express *knot* (*kn*), another transcription factor involved in dendrite morphology (Jinushi-Nakao et al., 2007; Li et al., 2017). Consistently, we found that the two $acj6^{+}/Lim1^{-}$ clusters (clusters 1 and 2) segregate according to *kn* expression (Fig. 3C).

We also identified three to eight genes in each PN cluster that were significantly over-expressed, as compared to the expression in other PN clusters (Fig. 3B). Of potential functional importance, we found that the $acj6^{+}/kn^{+}$ PNs express the *sNPF* neuropeptide gene, whereas neurons encompassing the putative lPNs express *Tachykinin* (*Tk*). These data suggest that these two classes of otherwise cholinergic

neurons might co-release different neuropeptides. Interestingly, the sNPF and Tk neuropeptides have previously been reported to have a modulatory role in the antennal lobe, although these studies concluded that the peptides were released from olfactory receptor neurons and local interneurons, respectively (Ignell et al., 2009; Nässel et al., 2008). More recently, others have also detected the expression of *Tk* in PNs (Li et al., 2017).

Assigning fast-acting neurotransmitters

We next assessed the proportion and distribution of cells in our data set that express genes that would indicate they release a particular fast-acting NT; acetylcholine (ACh), glutamate (Glu) and gamma-aminobutyric acid (GABA). We determined that cells were cholinergic, glutamatergic or GABA-ergic based on the expression of *vesicular acetylcholine transporter* (*VAChT*), *vesicular glutamate transporter* (*VGlut*) and *glutamic acid decarboxylase 1* (*Gad1*), three key proteins that are either required for the vesicular loading, or metabolism, of ACh, Glu and GABA respectively. Consistent with our expectations, this analysis labelled the cell clusters that most likely represent KCs and PNs as being cholinergic (Barnstedt et al., 2016; Tanaka et al., 2012), while the EB cluster is comprised of GABAergic cells (Fig 4A) (Kahsai et al., 2012). Reassuringly, we did not find significant NT marker expression in glia, including astrocytes.

Cells expressing these NT-specific marker genes were largely exclusive, although 6% of cells contained markers for ACh and GABA and 6% for ACh and Glu. It is therefore conceivable that some cells release excitatory and inhibitory NTs. A smaller percentage of cells expressed markers for Glu and GABA (3%), of which a third (1%) expressed all three NT markers (and therefore possibly represent multiple cell captures) (Fig. 4B).

Analysis of Neuropeptide expression

We also analyzed the expression of neuropeptides in our Drop-seq dataset. We first investigated whether individual neuropeptide-encoding genes were preferentially expressed in neurons that co-transmit/co-release a particular fast-acting NT ACh, Glu or GABA (Fig 7A). *sNPF*, *CCHamide-2* (*CCHa2*), *Tk*, *space blanket* (*spab*), *jelly belly* (*jeb*) and *amnesiac* (*amn*) showed a strong preference for expression in cholinergic

neurons, whereas *Diuretic hormone 31* (*Dh31*) is highly biased to GABA-ergic neurons. *Neuropeptide-like precursor 1* (*Nplp1*) and *Allatostatin A* (*AstA*) were mainly expressed in glutamatergic cells.

Some other peptide-encoding genes show a strong anti-correlation with a particular transmitter. For example *neuropeptide F* (*dNPF*), *sNPF*, *Tk*, *spab*, *jeb*, *Allatostatin C* (*AstC*), *Diuretic hormone 44* (*Dh44*), *CCHa2* and *Myosuppressin* (*Ms*) were anti-correlated with GABA-ergic cells. Similarly, *Myoinhibitory peptide precursor* (*Mip*), *pigment-dispersing factor* (*PDF*) and *SIFamide* (*SIFa*) were absent from cholinergic neurons.

Ms showed an interesting bias for expression in cells that express two (Glu and ACh or Glu and GABA) or all three fast acting NTs. We also noticed that the specificity towards cells expressing only one type of fast-acting NT varied between neuropeptides, with some such as PDF, exhibiting a broad and general expression pattern, other than the anti-correlation with ACh.

The abundance and specificity of expression across the midbrain also varied between individual neuropeptides. Some neuropeptide-encoding genes are only expressed in 1-2% of cells (e.g. *CCHa2*, *amn*, *dNPF*, *Mip*, *PDF* and *SIFa*), and their release could therefore potentially represent signals of, for example, internal states. Others, such as *spab*, *sNPF* and *Nplp1*, are very broadly expressed in 20-25% of all cells (see Supplementary file 1), suggesting that these neuropeptides likely act as modulatory co-transmitters with fast-acting NTs.

Some neuropeptide expression patterns are highly specific to certain cell types. For example, *Dh31* is mainly expressed by EB neurons whereas *sNPF* is strongly expressed in KCs (Fig. 2E), in a subdivision of PNs (Fig. 3B) and in two clusters that have not yet been assigned to a specific cell-type. Furthermore, although both *spab* and *Nplp1* are very broadly expressed, their expression patterns are strongly anti-correlated, suggesting that they may have complementary functions in the *Drosophila* midbrain.

We also found transcripts for the *Drosophila* insulin-like peptides 2, 3, 5 and 6 (Fig 5B). The Ilp2, Ilp3 and Ilp5 peptides are exclusively expressed in IPCs in the brain, whilst Ilp6 is expressed in glia (Brogiolo et al., 2001; Okamoto et al., 2009). We found that Ilp2, 3 and 5 expression was weakly correlated with that of NTs, whilst Ilp6 expression is strongly correlated with cells that do not express NT markers, consistent with glial expression.

283

284 **Assignment and subdivision of monoaminergic neurons**

285 We used expression of the *vesicular monoamine transporter* (*Vmat*) gene to identify
286 monaminergic neurons in our midbrain dataset (Figure 6A). Three discrete cell
287 populations clearly expressed *Vmat*. We performed a new PCA and tSNE analysis
288 on cells from these three clusters, guided by known markers for serotonin (5-HT),
289 tyramine (Tyr), octopamine (OA) and dopamine (DA) releasing neurons. *Ddc* (Dopa
290 decarboxylase) labels 5-HT and DA neurons, *SerT* (Serotonin transporter) and *Trh*
291 (Tryptophan hydroxylase) mark 5-HT neurons, *pale* (tyrosine hydroxylase*)* and *DAT*
292 (Dopamine transporter) DA neurons, *Tdc2* (Tyrosine decarboxylase 2) Tyr and OA
293 neurons, *Tbh* (tyramine β-hydroxylase) OA neurons. These labels allowed us to
294 identify the neuronal clusters corresponding to each of these cell types (Figure 6B).

295 In addition to known markers of monoaminergic neuronal types, we found new genes
296 expressed in these populations (Fig. 6C), that are likely to have an important role for
297 their development and connectivity, such as *kekkon 1* (*kek1*) in dopaminergic
298 neurons (DANs) (Ghiglione et al., 1999), or *IGF-II mRNA-binding protein* (*Imp*) and
299 *Jim Lovell* (*lov*) in serotonergic neurons (Bjorum et al., 2013; Geng and Macdonald,
300 2006; Munro et al., 2006). High expression in Tyr neurons of *hikaru genki* (*hig*), which
301 encodes a protein generally found in the synaptic clefts of cholinergic synapses
302 (Nakayama et al., 2014, 2016), may highlight the importance of cholinergic input to
303 these neurons.

304 Many neurons release more than one NT. We therefore investigated whether the
305 different types of monoaminergic neurons co-expressed markers for fast-acting
306 transmitters. The most striking evidence in these analyses suggest that many
307 octopaminergic and tyraminergic neurons likely co-release Glu, and less of them
308 GABA, or ACh. (Fig. 6D).

309 We also tested whether monoaminergic neurons co-expressed neuropeptide genes.
310 Many mature neuropeptides are amidated at their C-terminus through the sequential
311 enzymatic action of the *Phm* and *Pal2* encoded peptidylglycine-alpha-hydroxylating
312 monooxygenase and peptidyl-alpha-hydroxyglycine alpha-amidating lyase (Han et
313 al., 2004; Jiang et al., 2000; Kolhekar et al., 1997). These genes were expressed in
314 50% and 81% of all monoaminergic neurons, respectively (Fig. 6E), suggesting that a
315 significant proportion of monoaminergic neurons likely co-release neuropeptides.
316 Indeed, we found expression of *Dh44*, *Nplp1*, *Glycoprotein hormone beta 5* (*Gpb5*)

317 and *Proctolin* (*proc*; which is not amidated) in up to 21% of DANs (Fig. 6E). 61% of
318 DANs express at least one neuropeptide and 32% express two, or more. *Dh44*,
319 *Nplp1* and *spab* were found in up to 30% of 5-HT neurons, with 90% of these
320 expressing one or two neuropeptides (Fig. 6E).

321 Perhaps surprisingly, OA and Tyr neurons contained mRNA for many neuropeptides.
322 We found that 85% of OA neurons express at least one neuropeptide, whereas 46%
323 express two, or more. Co-expression was even more evident in Tyr neurons; 83%
324 expressed one, whereas 78% express two or more. *Nplp1*, *Gpb5*, and *SIFa* were
325 detected in Tyr and OA neurons, whereas *Dh44*, *Ms* and *spab* were only identified in
326 OA neurons, and *sNPF*, *Dh31*, *Mip*, *Ilp2* and *ITP* were exclusively found in Tyr
327 neurons (Fig. 6E). *Dh44* was the most broadly expressed, being detected in 46% of
328 OA neurons. *Mip* and *SIFa* were each expressed in 44% of Tyr neurons, and were
329 co-expressed in 33% of them. Together, these results indicate that neuropeptide
330 expression, and co-expression, is a common feature of many monoaminergic
331 neurons. The obvious complexity and possible heterogeneity of neuropeptide
332 expression may reflect functional specialization of individual, or small groups of these
333 monoaminergic neurons.

334 Prior work has shown that DANs are anatomically and functionally divisible based on
335 roles in motivation, learning and memory and arousal (Huetteroth et al., 2015;
336 Krashes et al., 2009; Nall et al., 2016; Yamagata et al., 2015). Some of this DAN
337 subdivision has also been associated with the expression of particular transcription
338 factors, receptors for specific neuropeptides, or other monoamines (Dib et al., 2014;
339 Ichinose et al., 2015; Krashes et al., 2009). DANs implicated in learning and memory
340 reside in two discrete clusters, called PPL1 and PAM. PPL1 DANs mostly convey the
341 negative reinforcing effects of aversive stimuli, such as electric shock, high heat or
342 bitter taste (Aso et al., 2012; Das et al., 2014; Galili et al., 2014), whereas the
343 numerically larger PAM cluster contains DANs that appear somewhat specialized in
344 representing particular types of rewards, such as the sweet taste and nutrient value
345 of sugars, or water (Burke et al., 2012; Huetteroth et al., 2015; Lin et al., 2014; Liu et
346 al., 2012; Yamagata et al., 2015). Prior work demonstrated that PAM DANs express
347 the transcription factor *48 related 2* (*Fer2*), which is required for their development
348 and survival (Dib et al., 2014). We found that 44 neurons in the DA cluster (37%)
349 express *Fer2* (Fig. 6B). We therefore consider these *Fer2*-positive cells to represent
350 PAM DANs. 15 additional genes are significantly over-expressed in these cells, in
351 comparison to the rest of the brain (Fig. 6F). Amongst them we found *Ddc*, *ple*, *Vmat*

and *DAT*, that are essential for DA synthesis, vesicle loading and transport (Yamamoto and Seto, 2014). Potential new markers for PAM DANs include the transcription factor *scarecrow* (*scro*), the amino-acid transporter *Jhl1-21*, the Dpr-interacting protein *DIP-delta*, the PDGF- and VEGF-related growth factor *Pvf3*, the EGFR modulator *kek1,* as well as five novel genes; *CG1402*, *CG13330*, *CG17193*, *CG10384* and *CG42817*.

To corroborate the expression of these new markers in PAM neurons, we compared this data to a transcriptome profiling dataset that was acquired from sequencing mRNA extracted from populations of GFP labeled PAM DANs. We used R58E02-GAL4, a PAM-specific line (Liu et al., 2012; Pfeiffer et al., 2008) to express UAS-6xGFP (Shearin et al., 2014) specifically in PAM DANs, and purified the cells by Fluorescence Activated Cell Sorting (FACS). We prepared mRNA from GFP+ and GFP- neurons, which was subsequently reverse-transcribed and amplified using Smart-seq2, and sequenced. This analysis identified about 10 times more (143) genes that were significantly over-expressed in PAM neurons, as compared to the number retrieved with Drop-seq (Fig. 6F). This return is consistent with previous reports showing a higher recovery rate with Smart-seq2 compared to Drop-seq, but also higher levels of noise, as Smart-seq2 does not employ UMIs (Ziegenhain et al., 2017). Of the 15 genes found to be over-expressed in PAM neurons in the Drop-seq experiment, 9 (*ple*, *DAT*, *Fer2*, *Jhl-21*, *scro*, *DIP-delta*, *CG10384*, *CG17193* and *CG42817*) were also retrieved in the Smart-seq2 data (Fig. 6F). This confirms that these genes are specifically expressed in PAM neurons. Furthermore, with the exception of *ple*, *DAT* and *Fer2*, these genes have not been previously localized to PAM neurons, and therefore represent novel markers for this cell-type.

Dopamine receptors

Cells respond to DA using a variety of DA receptors in their cell membrane. Interestingly, our analysis shows that all four DA receptors are found in KCs, which form numerous synapses with DA neurons at the MB lobes (see above; Fig. 7B). Many KCs co-express several of these receptors, and 24% of them even express all four receptors (Fig. 7C). Other cell types express combinations of DA receptors, to varying degrees. In addition to KCs, *Dop1R1* (*dumb*) and *Dop1R2* (*damb*) are found in a few other clusters (Fig. 7B), in particular in several of those that we could not attribute to any cell type. This information will be helpful for further characterizing

these clusters. Consistent with evidence showing that *Dop1R1* is expressed in the Central complex where it regulates arousal (Kahsai et al., 2012; Lebestky et al., 2009), we found limited expression of *Dop1R1* in the EB, suggesting that only a subset of these neurons are involved in this process. *Dop1R1* also seems to be expressed in small numbers of monoaminergic neurons, suggesting that it may play a role in autocrine signaling. However, the main candidate receptor for DA autocrine signaling is *Dop2R*, which was found to be broadly expressed in DANs, and also in large numbers of other monoaminergic neuronal types (Fig. 7B). Interestingly, *Dop2R* expression was also detected in some PNs and IPCs as well as a few non-attributed clusters, which indicates that the activity of these neurons is also subject to dopaminergic modulation. Finally, the Dopamine/Ecdysteroid receptor (*DopEcR*) was found in several cell types, including KCs, PNs, the ocelli, and many other non-attributed clusters (Fig. 7B), suggesting a broad role for this receptor. Expression of this *DopEcR* in PNs corroborates previous data showing its involvement in pheromone sensitization in these neurons, both in flies and moths (Abrieux et al., 2014; Aranda et al., 2017).

Dopamine metabolism

DA signaling is regulated by enzymatic degradation and reuptake through transporters. Recycled metabolites can then be used to resynthesize DA. These steps can occur in different cell types, that could be DA-releasing cells, post-synaptic neurons, or glia (Yamamoto and Seto, 2014) (Fig. 7A). We therefore used our Drop-seq data to determine which cell types expressed components of the DA recycling and metabolic pathways.

As expected, the first step of DA synthesis, conversion of tyrosine into the DA precursor L-DOPA catalyzed by the *ple* encoded Tyrosine hydroxylase appears to occur exclusively in DANs (Fig. 7B). In comparison *Ddc*, which converts L-DOPA to DA, is also involved in 5-HT synthesis, and so was expressed in DA and 5-HT neurons. Interestingly, *Ddc* also labels several other neuronal populations, including α′β′ and γ KCs, one cluster of olfactory PNs, and several non-identified, alphabet labeled clusters (Fig. 7B). It is not clear if Ddc in these neurons is involved in the metabolism of DA or other aromatic L-amino acids.

Three enzymes have been described to play a role in DA degradation and recycling. The *ebony* (*e*) gene product converts DA into N-beta-alanyldopamine (NBAD)

(Hovemann et al., 1998; Suh and Jackson, 2007) and was almost exclusively expressed in astrocytes in our data (Fig. 7B). Dopamine-N-acetyltransferase, encoded by *Dat*, converts DA into N-acetyl dopamine (NADA). Interestingly, *Dat* was abundant in astrocytes, in smaller amounts in other glia, and was also detected in the EB and a few other subsets of neurons (Fig. 7B). Although these results highlight the important role of glia, and in particular astrocytes, in DA reuptake, metabolism and recycling, other cells appear to convert DA into NADA rather than into NBAD. The fate and consequence of these two metabolites in each cell type remains largely unknown. Finally, *tan* (*t*), a gene coding for a hydrolase that can convert NBAD back into DA, was not found in any cell population from the central brain itself (Fig. 7B), suggesting that this recycling pathway is not utilized there. However, several cells from the ocelli express this enzyme, consistent with the function of *t* in the histamine metabolism in photoreceptors (Borycz et al., 2002; True et al., 2005).

The vesicular monoamine transporter (encoded by *Vmat*) transports DA, 5-HT, OA and Tyr into synaptic vesicles (Martin and Krantz, 2014). As already mentioned, *Vmat* was detected in all these neuronal types. In addition, *Vmat* expression was evident in fat body cells. Although this has, to our knowledge, never been demonstrated in flies, perivascular adipose tissue in rats contains monoamines acting on the sympathetic nervous system, and is thus likely to express VMAT (Ayala-Lopez et al., 2014). The *DAT*-encoded DA transporter mediates DA reuptake of by DANs. Unlike *Vmat*, *DAT* was specifically expressed in dopaminergic, but not other monoaminergic neurons. Surprisingly, we also found *DAT* expression in α′β′ KCs, suggesting that these neurons might tightly regulate their reception duration and magnitude of DA signals.

Distribution of nicotinic neurotransmitter receptors

The response of a neuron to a particular NT is determined by the types of receptors that that cell expresses. In addition, most ionotropic NT receptors are oligomers comprised of combinations of subunits, variations of which can have very different functional characteristics (Sattelle et al., 2005). Acetylcholine is a major excitatory NT in the insect brain and is the primary fast-acting NT released from olfactory receptor neurons, olfactory PNs and MB KCs. Nicotinic acetylcholine receptors (nAChR) are heteropentamers that can be comprised of 2 or 3 alpha and the corresponding 3 or 2 beta subunits. Flies have 7 alpha subunit genes and 3 types of beta encoding genes.

These receptors have mostly been studied at the vertebrate neuromuscular junction (Albuquerque et al., 2009) and very little is known about the composition of nAChR in neurons in a central nervous system. Although gene expression cannot explicitly inform of subunit composition, co-expression is a prerequisite that limits the potential complexity in any given neuron. We therefore analyzed the co-expression of nAChR subunits in our Drop-seq dataset. We detected the expression, at varying frequencies, of all seven known nAChR α-subunits, and two of the three known β-subunits in our samples (Fig. 8A). α1, α5, α6 and α7 are expressed in considerably more cells than α2, α3 and α4, whereas β1 is expressed in more than twice as many cells as β2. Most subunits are broadly expressed across all cell types, although some exhibit very distinct expression patterns. Most notably, α3 is broadly expressed in the midbrain, but distinctly absent in KCs. We also tested for co-expression of different combinations of receptor subunits (Figure 8B). Expression of α5 most strongly correlated with expression of α6 and β1. In contrast α3 weakly correlated with expression of α6 and β2 and α2 weakly with α4. Some of the patterns of expression are consistent with previously published pharmacological studies that tested for co-assembly of receptors by co-immunoprecipitation using α-Bungarotoxin (Chamaon et al., 2002; Schulz et al., 2000). For example, cells that express the *nAChR-α1* most frequently also express *α2*, when compared to all other nAChR subunits and these two subunits have been shown to preferentially co-assemble into the same receptor complex (Chamaon et al., 2002; Schulz et al., 2000). Similarly, *β1* is the most frequently co-expressed subunit in *β2* expressing cells, again confirming previous co-immunoprecipitation experiments. We also detected high expression levels of the secreted protein *quiver* (*qvr*), a Ly-6/neurotoxin family member, in most neurons of our sample (see Supplementary File 1). The mammalian homologue of quiver, lynx1, has been shown to bind and regulate nAChR in the mammalian nervous system (Miwa et al., 1999).

Co-expression of activity regulated genes

A recent study identified a set of genes whose expression was upregulated in response to prolonged neural activation. These activity-regulated genes (ARGs) were identified using differential bulk transcription profiling following broad neural activation, using three different artificial stimulation paradigms (Chen et al., 2016). We plotted the expression patterns of the 11 most highly upregulated genes that were identified following optogenetic neuronal activation, and found that 10 of them

were also robustly expressed in our dataset (see Supplementary file 1). Interestingly, the expression patterns of the most highly upregulated ARGs were strongly correlated (Fig 9). For example, cells that express the transcription factor *stripe* (*sr*) are more likely to also express *Hormone receptor-like in 38* (*Hr38*, p-value < 2.2×10^{-16}, Pearson's product-moment correlation) and *CG14186* (p-value < 2.2×10^{-16}). These three genes were the most highly upregulated in Chen et al. (2016), following artificial optogenetic neural stimulation. Our data therefore demonstrate that they are also likely to be co-regulated in the brain, following ordinary levels of neural activity.

We wondered whether the expression pattern of these ARGs might highlight areas of the fly midbrain that have a high intrinsic level of activity. However, no specific cluster was prominently marked with the expression of 9 out of the 10 ARGs tested. Only CG13055 strongly labeled the cluster of γ KCs (Fig. 2F and Supplementary file 2). In addition, we noticed that expression of most ARGs was slightly higher in γ KCs. Since prior work suggested that the γ neurons are the least active of all the KC subtypes (Tomchik and Davis, 2009), we speculate that ARG expression is a homeostatic neural response to reduce excitability.

Discussion

Generating an atlas of gene expression of every cell type in the human body is a goal of modern science (Regev et al., 2017). Remarkable recent advance in high-throughput single-cell RNA sequencing methods have brought this ambitious goal within reach. However, the large size of mammalian tissues means that huge numbers of cells need to be sequenced in order to capture a representative sample of the overall number. Insects, such as *Drosophila*, provide an obvious solution to the tissue size and cell number issues. Flies are complex organisms with tissues that serve analogous functions to many mammalian organs. Moreover, each one of these fly tissues is comprised of a greatly reduced number of cells, compared to their mammalian equivalent. This is perhaps most obvious when considering the brain. Whereas the mouse brain is considered to contain about 75 million neurons, the *Drosophila* brain has only 150,000. Since two thirds of these cells comprise the optic lobes, much of the computational cognitive power and behavioral orchestration is handled by about 50,000 neurons in the midbrain. In this study, we describe a global and unbiased single-cell transcriptomic analysis, using Drop-seq, that is representative of much of the *Drosophila* midbrain. This initial cell atlas of the fly brain provides a unique resource of gene expression across many cell types and regions of neuropil.

The extent of neural diversity is not known in any brain. Analysis of the fly therefore provides a useful inroad to this question. Even our initial clustering indicates a high level of neural complexity in the fly brain. Labeling the cluster plot with markers for the ACh, Glu and GABA NTs reveals that many diverse cells use each of these fast-acting NTs. For example, although the three major subsets of MB KCs are all cholinergic they each occupy a discrete cluster, and are distant to many other cholinergic neurons. The GABA-ergic ring neurons of the EB are similarly unique and distinct from other GABA-ergic neurons. At this stage, we cannot tell whether cells in the major KC subtype and EB neuron clusters are truly homogenous, or whether further iterative clustering will separate them into additional distinguishable subtypes. We might expect to find that the anatomically unique core, surface and posterior subdivisions of the αβ KCs have unique molecular profiles within the larger αβ cluster. Similarly, the EB ring neurons might be separable into layer specific subtypes. This will require additional analyses and perhaps the collection of more cells. Comparing Drop-seq profiles from the fly brain to those from larger social insects, such as ants and honeybees, and to neurons from the mammalian brain

would be useful to address the question of how a larger brain is constructed. As a brain evolves to be bigger, are there many more cell types? Or is there simply an expansion of the number of copies of each cell-type? One might imagine that just expanding the number of identical cortical units, such as pyramidal neurons or MB KCs, increases the computational power of the brain by permitting a higher degree of parallel processing and that the resulting larger networks also provide more storage space.

A key element of our analysis here is the ability to assign many single-cell molecular signatures to the relevant cell-type and brain region. We did this using a number of different approaches. Our data was collected from individual neurons taken from brains that specifically express mCherry in the αβ KCs of the mushroom body. This allowed us to unequivocally identify these neurons in our cluster plot, and demonstrates the power of sequencing cells from a brain where some specific neurons are genetically labeled. In theory, this strategy can be used to identify the profiles for any *Drosophila* cell-type in a Drop-seq dataset, providing a corresponding specific GAL4 driver line is available. This is a clear advantage of using *Drosophila* as a model for a brain cell-atlas, because thousands of GAL4 lines are available that direct expression in specific subsets of neurons in the brain. Cell-specific transgene expression therefore presents the most straightforward means to link single-cell sequencing data to neuroanatomy and will be very useful to de-orphan our currently 'unannotated' cells.

The extensive collection of cell-specific GAL4 lines were constructed by fusing potential regulatory regions from genes to GAL4 coding sequence. Their expression patterns can therefore indicate elements of the expression of the gene from which they are taken. We also used this property of the GAL4 collections to help us assign single-cell data to specific neurons. For example, we originally suspected that one of the clusters corresponded to EB ring neurons because cells in the cluster expressed the *Gad1* marker for GABA-ergic neurons and *Fas2*, an antibody for which is known to label this region of neuropil. To corroborate this assignment to EB we asked whether promoter-GAL4 lines constructed from some of the other top new markers for this cluster, such as *Dh31* and *Sox21b*, labeled EB ring neurons. Indeed, we found that the R20A02 (*Dh31*) and R73A06 (*Sox21b*) GAL4 drivers very specifically express in these neurons. Therefore, by combining the expression of known markers, and querying the specificity of new markers, it is possible to convincingly assign transcriptional profiles to cell-type.

Our initial analyses of the brain cell-atlas also immediately provided a lot of new information that is of functional importance. We focused our first investigations on neurotransmitter usage and the potential for synaptic co-release/co-transmission. These analyses clearly defined the main fast-acting transmitters used by each cell cluster. For example, the KC transmitter was fairly recently determined to be Ach (Barnstedt et al., 2016) and consequently all the KC clusters strongly labeled with the cholinergic markers *ChAT* and *VAChT*. The cell-atlas dataset therefore allows one to easily determine the neurotransmitters that a particular cell-type uses, providing the cells of interest can be identified in the cluster plot.

Important questions can also be addressed even without identifying how particular cells appear in the cluster plot. One example is our analyses of potential co-release of multiple fast acting transmitters or fast-acting transmitters with neuropeptides. Our data suggest that a small percentage of neurons might co-release ACh and Glu, or ACh and GABA. Analyzing co-expression of transmitter marker genes and neuropeptide-encoding genes revealed some very interesting and novel findings. We found that some neuropeptides, whether expressed in many or only a few cells, are exclusively detected in neurons that use a particular fast-acting transmitter. These correlations suggest a fine relationship between the fast-acting transmitter and neuropeptide-specific modulation. Our co-expression analyses also reveal extensive expression of neuropeptide-encoding and processing genes in monoaminergic neurons. It will be interesting to test whether the apparent heterogeneity of neuropeptide expression in these neurons contributes to their apparent functional specialization (Aso et al., 2012, 2014; Burke et al., 2012; Claridge-Chang et al., 2009; Huetteroth et al., 2015; Krashes et al., 2009; Lin et al., 2014; Liu et al., 2012; Yamagata et al., 2015).

The brain cell atlas is of great use to those with a gene-centered view of fly neurobiology. It is now possible to query the atlas and ask how broadly, or cell-specifically, a given gene is expressed. Our initial clustering allows one in some cases to pinpoint the expression to a defined cell-type and region of neuropil. This seems particularly valuable information if one is working with a gene, for example, one that has been implicated in neural disease, but does not know the anatomical context in which it operates. Similarly, if a constitutive mutant fly strain has pleiotropic effects, the expression pattern of the gene can indicate where the different phenotypes might manifest. Moreover, the brain-atlas dataset can provide these

answers quickly for multiple genes, and it therefore represents a terrific complement to the usual time-consuming and 'single-gene at a time' approaches, such as technically challenging *in-situ* hybridization to RNA, generating antibodies, making promoter fusions, or knocking in epitope tags to individual loci. Perhaps most importantly, querying the cell-atlas provides single-cell resolution of gene expression across all the major cell-types in the fly midbrain.

We believe that the potential uses for the cell atlas are almost endless. The data reveal a huge number of new genetic markers for known cell types, and as yet undefined cell types, in the fly brain. Many of these are likely to be functionally important and represent new entry points to guide interventionist experiments to understand how specific molecules operate within the relevant neurons and networks. Although we focused on neural cells, different classes of glia (Freeman, 2015) could also be defined in the cluster.

Our initial analysis was performed on 10,286 of the highest quality cells (≥800 UMIs) from a larger dataset of 19,260 cells. This atlas is effectively a scaffold that can now be continuously updated and expanded as additional cells are collected and sequenced. Our current dataset was derived from cells taken from unique groups of flies, processed on 8 separate days, and yet each biological replicate contributed equally to the combined data set. This robustness and reproducibility of the approach is essential to know in order to be able to add data from future experiments to the current cell cluster. Including more cells with a comparably high number of UMIs per cell should increase statistical power and permit further resolution of cell-type. Including more cells with a lower number of UMIs per cell did not improve our analysis.

The current dataset was collected from young flies that were raised under ideal conditions with ample food and water. Future experiments that aim to investigate the impact of changes to the state of the fly, such as age, bacterial infection and starvation, can use the current cell atlas as a foundation to identify changes in expression patterns that may occur in individual cells across the midbrain. Similarly, brains from flies harboring specific mutations can be molecularly characterized using the approach described here, to uncover molecular manifestations of the mutant phenotype.

643 The fly brain cell atlas described here should also be a valuable resource to
644 researchers working in other animals. Many markers for *Drosophila* cell-type are
645 likely to be conserved in other insects and arthropods, and so will be useful markers
646 for regions of the brain in these animals (Thoen et al., 2017; Wolff and Strausfeld,
647 2015). The orthologs of some of these new markers, for example those expressed in
648 subsets of dopaminergic neurons, might also extend to labeling comparable cells in
649 the mammalian brain.

Materials and methods

Fly strains

The *Drosophila* strains used were MB008B (Aso et al., 2014), R58E02 (Pfeiffer et al., 2008), *w-;+;20XUAS-6XGFP* (Shearin et al., 2014) and *w-; +; UAS-mCherry*. Flies were raised at 25°C in 12h:12h day-night cycles on standard food at 40-50% humidity.

Cell culture

S2 cells (Gibco, R69007) were grown in Schneider's medium (Gibco 21720-001) supplemented with 10% FBS (Sigma, F0804) & 1% penicillin-streptomycin (Gibco, 15070-063). Sf9 cells (Gibco, 12659-017) were grown in Sf-900™ III SFM (Gibco, 12658019). All cells were incubated at 25°C. Cells were grown in adherent cultures to confluency. Vessels were gently tapped to detach cells, and supernatants were centrifuged for 10 minutes at 100 x g. Cells were washed once with 1 x PBS and resuspended in 1 x PBS and subsequently diluted to 200 cells / ul prior to pooling and Drop-seq.

Brain dissociation and cell collection

The brain dissociation protocol was adapted from previously described methods (Harzer et al., 2013; Nagoshi et al., 2010). For each day of experiments, 80-100 central brains were individually dissected in ice-cold calcium- and magnesium-free DPBS (Gibco, 14190-086) and immediately transferred into 1 mL toxin-supplemented Schneider's medium (tSM: Gibco, 21720-001 + 50 µM d(−)-2-amino-5-phosphonovaleric acid, 20 µM 6,7-dinitroquinoxaline-2,3-dione and 0.1 µM tetrodotoxin) on ice. Brains were washed once with 1 mL tSM and incubated in tSM containing 1.11 mg/mL papain (Sigma, P4762) and 1.11 mg/mL collagenase I (Sigma, C2674). Brains were washed once more with tSM and subsequently triturated with flame-rounded 200-µL pipette tips. Dissociated brains were resuspended into 1mL PBS + 0.01% BSA and filtered through a 10 µm CellTrix strainer (Sysmex, 04-0042-2314). Cell concentration was measured using a disposable Fuchs-Rosenthal hemocytometer (VWR, 631-1096) under a Leica DMIL LED Fluo microscope, that also allowed detecting mCherry fluorescence in

dissociated KCs. Cells were diluted in PBS + 0.01% BSA up to a concentration of 200 cells/µL. Thus a typical preparation from 80 brains yielded ~2'000'000 single-cells in a volume of 10 mL.

Drop-seq procedure

Drop-seq was performed as described (Macosko et al., 2015), using a Dolomite Bio (Royston, UK) Single Cell RNA-Seq system. Cells were diluted at a concentration of 200 cells/µL into PBS + 0.01% BSA. Barcoded Beads SeqB (ChemGenes Corp., Wilmington, MA, USA) were diluted at a concentration of 200 particles/µL into 200 mM Tris pH 7.5, 6% Ficoll PM-400, 0.2% Sarkosyl, 20 mM EDTA + 50 mM DTT.

For each run, 700 µL of cells solution from dissociated brains were loaded into a microcentrifuge tube inside a reservoir connected to a Mitos P-Pump (Dolomite microfluidics, 3200016) set to provide a constant flow of 30 µL/min. The reservoir was placed on a stirring plate and agitation was provided by a stir bar placed inside the reservoir but outside the tube to maintain the cells in suspension while avoiding damaging the cells. 600 µL of beads solution were loaded into a 50 cm sample loop connected to a second Mitos P-Pump set to provide a constant flow of 30 µL/min. The sample loop was used to avoid beads sedimentation while eliminating the need for stirring, thus preventing beads damage. QX200 Droplet Generation Oil for EvaGreen (BioRad, 1864006) was loaded directly inside a third Mitos P-Pump, set to provide a constant flow of 200 µL/min. Cells, beads and oil flows were connected to a Single Cell RNA Seq Droplet Chip (Dolomite Bio) according to manufacturer's instructions, allowing pairing of single-cells with single-beads and formation of 357 pL droplets of aqueous cell/bead solution in oil. The chip was placed under a Meros High Speed Digital Microscope & Camera with a HLB M Plan Apo 5X objective in order to monitor droplet formation. Droplets were collected in 50 mL Falcon tubes. Reagents were reloaded and Falcon tubes replaced every 15 minutes.

Droplets were subsequently broken and beads with captured mRNA were washed as described (Macosko et al., 2015). In brief, bead-bound mRNA was immediately reverse-transcribed using a Template Switch Oligo (5' – AAG CAG TGG TAT CAA CGC AGA GTG AAT rGrGrG – 3') and Maxima H Minus Reverse Transcriptase (Thermo Scientific, EP0753). cDNA was treated with Exonuclease I (NEB, M0293L) and amplified in multiple 50 µL PCR reactions performed on aliquots of ~2000 beads,

using a SMART PCR primer (5' – AAG CAG TGG TAT CAA CGC AGA GT – 3') and Hifi HotStart Readymix (Kapa Biosystems, KK2602) for a total of 17 cycles. 10 µL from each PCR reaction were pooled, and amplified cDNA was purified twice, with 0.6X and 1.0X volumes of Agencourt AMPure XP beads (Beckman Coulter, A63880) and quantified on a Bioanalyzer, using a High-Sensitivity DNA kit (Agilent, 5067-4626). From each sample, 2 x 600 pg of amplified cDNA were tagmented using the Nextera XT DNA Sample Preparation Kit (Illumina, FC-131-1024) with New-P5-SMART PCR hybrid (5' – AAT GAT ACG GCG ACC ACC GAG ATC TAC ACG CCT GTC CGC GGA AGC AGT GGT ATC AAC GCA GAG T*A*C – 3') and one of Nextera N701 to N706 oligos. cDNA libraries were purified twice, with 0.6X and 1.0X volumes of Agencourt AMPure XP beads (Beckman Coulter, A63880) and quantified on a Bioanalyzer, using a High-Sensitivity DNA kit (Agilent, 5067-4626). Libraries were pooled together and sequenced on an Illumina HiSeq2500 sequencer using a Custom Read1 primer (5' – GCC TGT CCG CGG AAG CAG TGG TAT CAA CGC AGA GTA C – 3') and standard Illumina Read2 primers. All oligos were synthesized by Sigma, and HPLC purified. Samples from days 1 and 2 were sequenced together, on two separate lanes. Samples from days 3-8 were sequenced together, on three separate lanes.

Data processing and alignment

Sequencing data was processed as previously described (Macosko et al., 2015; Satija et al., 2015), following the Drop-seq Computational Protocol v.1.0.1 and using the Drop-seq software tools v.1.13 from the McCarroll lab. Barcodes were extracted and reads were aligned to a combination of the *Drosophila melanogaster* genome release 6.13 (from Flybase.org) and three reference sequences for mCherry and each split-Gal4 transgenes of the flies that were used in this study. For the species mix experiments, reads were also aligned to the *Spodoptera frugiperda* genome (Kakumani et al., 2014), available at NCBI GenBank, assembly ASM221328v1. The Flybase v.FB2016_05 September gene names were used for the creation of the Digital Gene Expression (DGE) Matrix.

t-SNE analysis on whole brain data

Analysis of DGEs was performed with the Seurat 2.1.0 R package (Macosko et al., 2015; Satija et al., 2015). Cells with less than 200 genes were discarded. Several thresholds for the number of UMIs per cell were tested (see Fig. 1 - figure supplement 2). All results presented here are based on 800 and 10,000 UMIs per cell as lower and higher threshold, respectively. Data was log-normalized and scaled using default options. Variation driven by individual batches was regressed out from the normalized, scaled data. PCA analysis was performed on the data as previously described (Macosko et al., 2015). To visualize the data, spectral t-SNE dimensionality reduction was performed, using the first 50 PCAs, as instructed by a Jack Straw resampling test (Satija et al., 2015; Van Der Maaten, 2014). Clusters were identified by a shared nearest neighbor modularity optimization (Waltman and Eck, 2013), using a resolution of 2.5. Some of these clusters were subsequently manually modified (compare Fig. 1 - figure supplement 2B (unmodified) and Fig. 1B (modified)). Main markers for each identified cluster were identified as genes with Log2 FC ≥ 1 and a p-value of $p < 0.01$ (after Bonferroni correction).

Co-expression analysis

Gene co-expression was assessed by calculating the Pearson product-moment correlation of the log-normalized, scaled expression values using R. For the radar plots, the number of cells expressing each neuropeptide of interest and in addition either *VACht* (to identify cholinergic cells), *VGlut* (glutamatergic), *Gad1* (GABA-ergic) or combinations of the three were calculated and normalized to the total number of cells expressing each NT.

t-SNE analysis on PNs and monoaminergic neurons (re-clustering)

DGE columns corresponding to cells belonging to either PN or monoaminergic clusters were used for PCA analyses. For re-clustering of monoaminergic neurons, a selection of known markers (*ple*, *DAT*, *SerT*, *Trh*, *Vmat*, *Oamb*, *Ddc*, *Tdc2* and *Tbh*) was used as input for PCA analysis. In both cases, the first 6 PCAs were used for re-clustering, which was performed as above.

RNA-sequencing of PAM-DA neurons

Central brains from flies expressing a brighter, hexameric GFP (20xUAS-6xGFP; Shearin et al., 2014) specifically in PAM-DA neurons under the control of the R58E02 Gal4 line (Pfeiffer et al., 2008) were dissected and dissociated as above. Filtered cells were sorted with a MoFlo Astrios (Beckman Coulter), and both GFP+ and GFP- cells were collected. Cells were lysed, retro-transcribed and amplified (17 cycles) using the SMART-Seq v4 Ultra Low Input RNA Kit for Sequencing (Takara Clontech, 634890), according to manufacturer's instructions. Technical triplicates were made for each sample. cDNA libraries were generated with TruSeq RNA Library Prep Kit v2 (Illumina, RS-122-2001) and sequenced on an Illumina HiSeq4000 sequencer.

Acknowledgements

We thank the Bloomington Stock Center for flies. We are grateful to Michal Maj for help with FACS, Paola Cognigni for help with R and statistics, and other members of the Waddell lab for discussion. RNA-sequencing was performed at MacroGen Inc. and the Oxford Genomic Centre. CT was supported by a Wellcome Trust PhD studentship and SW is funded by a Wellcome Trust Principal Research Fellowship in the Basic Biomedical Sciences (200846/Z/16/Z) and the Bettencourt-Schueller Foundation.

Competing interests

The authors declare that no competing interests exist.

References

Abrieux, A., Duportets, L., Debernard, S., Gadenne, C., and Anton, S. (2014). The GPCR membrane receptor, DopEcR, mediates the actions of both dopamine and ecdysone to control sex pheromone perception in an insect. Front. Behav. Neurosci. *8*.

Albuquerque, E.X., Pereira, E.F.R., Alkondon, M., and Rogers, S.W. (2009). Mammalian nicotinic acetylcholine receptors: from structure to function. Physiol. Rev. *89*, 73–120.

Aranda, G.P., Hinojos, S.J., Sabandal, P.R., Evans, P.D., and Han, K.-A. (2017). Behavioral Sensitization to the Disinhibition Effect of Ethanol Requires the Dopamine/Ecdysone Receptor in Drosophila. Front. Syst. Neurosci. *11*.

812 Aso, Y., Herb, A., Ogueta, M., Siwanowicz, I., Templier, T., Friedrich, A.B., Ito, K.,
813 Scholz, H., and Tanimoto, H. (2012). Three dopamine pathways induce aversive
814 odor memories with different stability. PLoS Genet. *8*, e1002768.

815 Aso, Y., Hattori, D., Yu, Y., Johnston, R.M., Iyer, N.A., Ngo, T.-T.B., Dionne, H.,
816 Abbott, L.F., Axel, R., Tanimoto, H., et al. (2014). The neuronal architecture of the
817 mushroom body provides a logic for associative learning. ELife *3*, e04577.

818 Awasaki, T., Saito, M., Sone, M., Suzuki, E., Sakai, R., Ito, K., and Hama, C. (2000).
819 The Drosophila trio plays an essential role in patterning of axons by regulating their
820 directional extension. Neuron *26*, 119–131.

821 Ayala-Lopez, N., Martini, M., Jackson, W.F., Darios, E., Burnett, R., Seitz, B., Fink,
822 G.D., and Watts, S.W. (2014). Perivascular adipose tissue contains functional
823 catecholamines. Pharmacol. Res. Perspect. *2*.

824 Barnstedt, O., Owald, D., Felsenberg, J., Brain, R., Moszynski, J.-P., Talbot, C.B.,
825 Perrat, P.N., and Waddell, S. (2016). Memory-Relevant Mushroom Body Output
826 Synapses Are Cholinergic. Neuron *89*, 1237–1247.

827 de Belle, J.S., and Heisenberg, M. (1994). Associative odor learning in Drosophila
828 abolished by chemical ablation of mushroom bodies. Science *263*, 692–695.

829 Berck, M.E., Khandelwal, A., Claus, L., Hernandez-Nunez, L., Si, G., Tabone, C.J.,
830 Li, F., Truman, J.W., Fetter, R.D., Louis, M., et al. (2016). The wiring diagram of a
831 glomerular olfactory system. ELife *5*.

832 Bjorum, S.M., Simonette, R.A., Alanis, R., Wang, J.E., Lewis, B.M., Trejo, M.H.,
833 Hanson, K.A., and Beckingham, K.M. (2013). The Drosophila BTB domain protein
834 Jim Lovell has roles in multiple larval and adult behaviors. PloS One *8*, e61270.

835 Borycz, J., Borycz, J.A., Loubani, M., and Meinertzhagen, I.A. (2002). tan and ebony
836 Genes Regulate a Novel Pathway for Transmitter Metabolism at Fly Photoreceptor
837 Terminals. J. Neurosci. *22*, 10549–10557.

838 Bouzaiane, E., Trannoy, S., Scheunemann, L., Plaçais, P.-Y., and Preat, T. (2015).
839 Two independent mushroom body output circuits retrieve the six discrete
840 components of Drosophila aversive memory. Cell Rep. *11*, 1280–1292.

841 Brogiolo, W., Stocker, H., Ikeya, T., Rintelen, F., Fernandez, R., and Hafen, E.
842 (2001). An evolutionarily conserved function of the Drosophila insulin receptor and
843 insulin-like peptides in growth control. Curr. Biol. CB *11*, 213–221.

844 Burke, C.J., Huetteroth, W., Owald, D., Perisse, E., Krashes, M.J., Das, G., Gohl, D.,
845 Silies, M., Certel, S., and Waddell, S. (2012). Layered reward signalling through
846 octopamine and dopamine in Drosophila. Nature *492*, 433–437.

847 Chamaon, K., Smalla, K.-H., Thomas, U., and Gundelfinger, E.D. (2002). Nicotinic
848 acetylcholine receptors of Drosophila: three subunits encoded by genomically linked
849 genes can co-assemble into the same receptor complex. J. Neurochem. *80*, 149–
850 157.

851 Chen, X., Rahman, R., Guo, F., and Rosbash, M. (2016). Genome-wide identification
852 of neuronal activity-regulated genes in Drosophila. ELife *5*.

853 Cheng, Y., Endo, K., Wu, K., Rodan, A.R., Heberlein, U., and Davis, R.L. (2001).
854 Drosophila fasciclinII Is Required for the Formation of Odor Memories and for Normal
855 Sensitivity to Alcohol. Cell *105*, 757–768.

856 Claridge-Chang, A., Roorda, R.D., Vrontou, E., Sjulson, L., Li, H., Hirsh, J., and
857 Miesenböck, G. (2009). Writing memories with light-addressable reinforcement
858 circuitry. Cell *139*, 405–415.

859 Cognigni, P., Felsenberg, J., and Waddell, S. (2018). Do the right thing: neural
860 network mechanisms of memory formation, expression and update in Drosophila.
861 Curr. Opin. Neurobiol. *49*, 51–58.

862 Crittenden, J.R., Skoulakis, E.M., Han, K.A., Kalderon, D., and Davis, R.L. (1998).
863 Tripartite mushroom body architecture revealed by antigenic markers. Learn. Mem.
864 Cold Spring Harb. N *5*, 38–51.

865 Das, G., Klappenbach, M., Vrontou, E., Perisse, E., Clark, C.M., Burke, C.J., and
866 Waddell, S. (2014). Drosophila learn opposing components of a compound food
867 stimulus. Curr. Biol. CB *24*, 1723–1730.

868 Dib, P.B., Gnägi, B., Daly, F., Sabado, V., Tas, D., Glauser, D.A., Meister, P., and
869 Nagoshi, E. (2014). A Conserved Role for p48 Homologs in Protecting Dopaminergic
870 Neurons from Oxidative Stress. PLOS Genet. *10*, e1004718.

871 Eichler, K., Li, F., Litwin-Kumar, A., Park, Y., Andrade, I., Schneider-Mizell, C.M.,
872 Saumweber, T., Huser, A., Eschbach, C., Gerber, B., et al. (2017). The complete
873 connectome of a learning and memory centre in an insect brain. Nature *548*, 175–
874 182.

875 Freeman, M.R. (2015). Drosophila Central Nervous System Glia. Cold Spring Harb.
876 Perspect. Biol. *7*.

877 Galili, D.S., Dylla, K.V., Lüdke, A., Friedrich, A.B., Yamagata, N., Wong, J.Y.H., Ho,
878 C.H., Szyszka, P., and Tanimoto, H. (2014). Converging circuits mediate temperature
879 and shock aversive olfactory conditioning in Drosophila. Curr. Biol. CB *24*, 1712–
880 1722.

881 Geng, C., and Macdonald, P.M. (2006). Imp associates with squid and Hrp48 and
882 contributes to localized expression of gurken in the oocyte. Mol. Cell. Biol. *26*, 9508–
883 9516.

884 Ghiglione, C., Carraway, K.L., Amundadottir, L.T., Boswell, R.E., Perrimon, N., and
885 Duffy, J.B. (1999). The transmembrane molecule kekkon 1 acts in a feedback loop to
886 negatively regulate the activity of the Drosophila EGF receptor during oogenesis. Cell
887 *96*, 847–856.

888 Haberkern, H., and Jayaraman, V. (2016). Studying small brains to understand the
889 building blocks of cognition. Curr. Opin. Neurobiol. *37*, 59–65.

890 Han, K.-A., Millar, N.S., Grotewiel, M.S., and Davis, R.L. (1996). DAMB, a Novel
891 Dopamine Receptor Expressed Specifically in Drosophila Mushroom Bodies. Neuron
892 *16*, 1127–1135.

893 Han, M., Park, D., Vanderzalm, P.J., Mains, R.E., Eipper, B.A., and Taghert, P.H.
894 (2004). Drosophila uses two distinct neuropeptide amidating enzymes, dPAL1 and
895 dPAL2. J. Neurochem. *90*, 129–141.

896 Harzer, H., Berger, C., Conder, R., Schmauss, G., and Knoblich, J.A. (2013). FACS
897 purification of Drosophila larval neuroblasts for next-generation sequencing. Nat.
898 Protoc. *8*, 1088–1099.

899 Heisenberg, M. (2003). Mushroom body memoir: from maps to models. Nat. Rev.
900 Neurosci. *4*, 266–275.

901 Hovemann, B.T., Ryseck, R.P., Walldorf, U., Störtkuhl, K.F., Dietzel, I.D., and
902 Dessen, E. (1998). The Drosophila ebony gene is closely related to microbial peptide
903 synthetases and shows specific cuticle and nervous system expression. Gene *221*,
904 1–9.

905 Huetteroth, W., Perisse, E., Lin, S., Klappenbach, M., Burke, C., and Waddell, S.
906 (2015). Sweet taste and nutrient value subdivide rewarding dopaminergic neurons in
907 Drosophila. Curr. Biol. CB *25*, 751–758.

908 Ichinose, T., Aso, Y., Yamagata, N., Abe, A., Rubin, G.M., and Tanimoto, H. (2015).
909 Reward signal in a recurrent circuit drives appetitive long-term memory formation.
910 ELife *4*, e10719.

911 Ignell, R., Root, C.M., Birse, R.T., Wang, J.W., Nässel, D.R., and Winther, Å.M.E.
912 (2009). Presynaptic peptidergic modulation of olfactory receptor neurons in
913 Drosophila. Proc. Natl. Acad. Sci. U. S. A. *106*, 13070–13075.

914 Jiang, N., Kolhekar, A.S., Jacobs, P.S., Mains, R.E., Eipper, B.A., and Taghert, P.H.
915 (2000). PHM is required for normal developmental transitions and for biosynthesis of
916 secretory peptides in Drosophila. Dev. Biol. *226*, 118–136.

917 Jinushi-Nakao, S., Arvind, R., Amikura, R., Kinameri, E., Liu, A.W., and Moore, A.W.
918 (2007). Knot/Collier and Cut Control Different Aspects of Dendrite Cytoskeleton and
919 Synergize to Define Final Arbor Shape. Neuron *56*, 963–978.

920 Johard, H.A.D., Enell, L.E., Gustafsson, E., Trifilieff, P., Veenstra, J.A., and Nässel,
921 D.R. (2008). Intrinsic neurons of Drosophila mushroom bodies express short
922 neuropeptide F: relations to extrinsic neurons expressing different neurotransmitters.
923 J. Comp. Neurol. *507*, 1479–1496.

924 Kahsai, L., Carlsson, M.A., Winther, A.M.E., and Nässel, D.R. (2012). Distribution of
925 metabotropic receptors of serotonin, dopamine, GABA, glutamate, and short
926 neuropeptide F in the central complex of Drosophila. Neuroscience *208*, 11–26.

927 Kakumani, P.K., Malhotra, P., Mukherjee, S.K., and Bhatnagar, R.K. (2014). A draft
928 genome assembly of the army worm, Spodoptera frugiperda. Genomics *104*, 134–
929 143.

930 Kelley, R.L., and Kuroda, M.I. (2003). The Drosophila roX1 RNA gene can overcome
931 silent chromatin by recruiting the male-specific lethal dosage compensation complex.
932 Genetics *164*, 565–574.

933 Kolhekar, A.S., Roberts, M.S., Jiang, N., Johnson, R.C., Mains, R.E., Eipper, B.A.,
934 and Taghert, P.H. (1997). Neuropeptide amidation in Drosophila: separate genes

encode the two enzymes catalyzing amidation. J. Neurosci. Off. J. Soc. Neurosci. *17*, 1363–1376.

Komiyama, T., and Luo, L. (2007). Intrinsic control of precise dendritic targeting by an ensemble of transcription factors. Curr. Biol. CB *17*, 278–285.

Komiyama, T., Johnson, W.A., Luo, L., and Jefferis, G.S.X.E. (2003). From lineage to wiring specificity. POU domain transcription factors control precise connections of Drosophila olfactory projection neurons. Cell *112*, 157–167.

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

Kurusu, M., Nagao, T., Walldorf, U., Flister, S., Gehring, W.J., and Furukubo-Tokunaga, K. (2000). Genetic control of development of the mushroom bodies, the associative learning centers in the Drosophila brain, by the eyeless, twin of eyeless, and Dachshund genes. Proc. Natl. Acad. Sci. U. S. A. *97*, 2140–2144.

Lai, S.-L., Awasaki, T., Ito, K., and Lee, T. (2008). Clonal analysis of Drosophila antennal lobe neurons: diverse neuronal architectures in the lateral neuroblast lineage. Development *135*, 2883–2893.

Lebestky, T., Chang, J.-S.C., Dankert, H., Zelnik, L., Kim, Y.-C., Han, K.-A., Wolf, F.W., Perona, P., and Anderson, D.J. (2009). Two different forms of arousal in Drosophila are oppositely regulated by the dopamine D1 receptor ortholog DopR via distinct neural circuits. Neuron *64*, 522–536.

Li, H., Horns, F., Wu, B., Xie, Q., Li, J., Li, T., Luginbuhl, D.J., Quake, S.R., and Luo, L. (2017). Classifying Drosophila Olfactory Projection Neuron Subtypes by Single-Cell RNA Sequencing. Cell *171*, 1206–1220.e22.

Lin, S., Owald, D., Chandra, V., Talbot, C., Huetteroth, W., and Waddell, S. (2014). Neural correlates of water reward in thirsty Drosophila. Nat. Neurosci. *17*, 1536–1542.

Liu, C., Plaçais, P.-Y., Yamagata, N., Pfeiffer, B.D., Aso, Y., Friedrich, A.B., Siwanowicz, I., Rubin, G.M., Preat, T., and Tanimoto, H. (2012). A subset of dopamine neurons signals reward for odour memory in Drosophila. Nature *488*, 512–516.

Macosko, E.Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, A.R., Kamitaki, N., Martersteck, E.M., et al. (2015). Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. Cell *161*, 1202–1214.

Martin, C.A., and Krantz, D.E. (2014). Drosophila melanogaster as a genetic model system to study neurotransmitter transporters. Neurochem. Int. *73*, 71–88.

Miwa, J.M., Ibanez-Tallon, I., Crabtree, G.W., Sánchez, R., Sali, A., Role, L.W., and Heintz, N. (1999). lynx1, an endogenous toxin-like modulator of nicotinic acetylcholine receptors in the mammalian CNS. Neuron *23*, 105–114.

975 Munro, T.P., Kwon, S., Schnapp, B.J., and St Johnston, D. (2006). A repeated IMP-
976 binding motif controls oskar mRNA translation and anchoring independently of
977 Drosophila melanogaster IMP. J. Cell Biol. *172*, 577–588.

978 Nagoshi, E., Sugino, K., Kula, E., Okazaki, E., Tachibana, T., Nelson, S., and
979 Rosbash, M. (2010). Dissecting differential gene expression within the circadian
980 neuronal circuit of Drosophila. Nat. Neurosci. *13*, 60–68.

981 Nakayama, M., Matsushita, F., and Hama, C. (2014). The Matrix Protein Hikaru genki
982 Localizes to Cholinergic Synaptic Clefts and Regulates Postsynaptic Organization in
983 the Drosophila Brain. J. Neurosci. *34*, 13872–13877.

984 Nakayama, M., Suzuki, E., Tsunoda, S., and Hama, C. (2016). The Matrix Proteins
985 Hasp and Hig Exhibit Segregated Distribution within Synaptic Clefts and Play Distinct
986 Roles in Synaptogenesis. J. Neurosci. *36*, 590–606.

987 Nall, A.H., Shakhmantsir, I., Cichewicz, K., Birman, S., Hirsh, J., and Sehgal, A.
988 (2016). Caffeine promotes wakefulness via dopamine signaling in Drosophila. Sci.
989 Rep. *6*, 20938.

990 Nässel, D.R., Enell, L.E., Santos, J.G., Wegener, C., and Johard, H.A. (2008). A
991 large population of diverse neurons in the Drosophila central nervous system
992 expresses short neuropeptide F, suggesting multiple distributed peptide functions.
993 BMC Neurosci. *9*, 90.

994 Ohyama, T., Schneider-Mizell, C.M., Fetter, R.D., Aleman, J.V., Franconville, R.,
995 Rivera-Alba, M., Mensh, B.D., Branson, K.M., Simpson, J.H., Truman, J.W., et al.
996 (2015). A multilevel multimodal circuit enhances action selection in Drosophila.
997 Nature *520*, 633–639.

998 Okamoto, N., Yamanaka, N., Yagi, Y., Nishida, Y., Kataoka, H., O'Connor, M.B., and
999 Mizoguchi, A. (2009). A fat body-derived IGF-like peptide regulates postfeeding
1000 growth in Drosophila. Dev. Cell *17*, 885–891.

1001 Perisse, E., Yin, Y., Lin, A.C., Lin, S., Huetteroth, W., and Waddell, S. (2013).
1002 Different kenyon cell populations drive learned approach and avoidance in
1003 Drosophila. Neuron *79*, 945–956.

1004 Pfeiffer, B.D., Jenett, A., Hammonds, A.S., Ngo, T.-T.B., Misra, S., Murphy, C.,
1005 Scully, A., Carlson, J.W., Wan, K.H., Laverty, T.R., et al. (2008). Tools for
1006 neuroanatomy and neurogenetics in Drosophila. Proc. Natl. Acad. Sci. U. S. A. *105*,
1007 9715–9720.

1008 Regev, A., Teichmann, S.A., Lander, E.S., Amit, I., Benoist, C., Birney, E.,
1009 Bodenmiller, B., Campbell, P.J., Carninci, P., Clatworthy, M., et al. (2017). Science
1010 Forum: The Human Cell Atlas. ELife *6*, e27041.

1011 Satija, R., Farrell, J.A., Gennert, D., Schier, A.F., and Regev, A. (2015). Spatial
1012 reconstruction of single-cell gene expression data. Nat. Biotechnol. *33*, 495–502.

1013 Sattelle, D.B., Jones, A.K., Sattelle, B.M., Matsuda, K., Reenan, R., and Biggin, P.C.
1014 (2005). Edit, cut and paste in the nicotinic acetylcholine receptor gene family of
1015 Drosophila melanogaster. BioEssays News Rev. Mol. Cell. Dev. Biol. *27*, 366–376.

1016 Schulz, R., Bertrand, S., Chamaon, K., Smalla, K.-H., Gundelfinger, E.D., and
1017 Bertrand, D. (2000). Neuronal Nicotinic Acetylcholine Receptors from Drosophila. J.
1018 Neurochem. *74*, 2537–2546.

1019 Seelig, J.D., and Jayaraman, V. (2015). Neural dynamics for landmark orientation
1020 and angular path integration. Nature *521*, 186–191.

1021 Shearin, H.K., Macdonald, I.S., Spector, L.P., and Stowers, R.S. (2014). Hexameric
1022 GFP and mCherry reporters for the Drosophila GAL4, Q, and LexA transcription
1023 systems. Genetics *196*, 951–960.

1024 Suh, J., and Jackson, F.R. (2007). Drosophila ebony activity is required in glia for the
1025 circadian regulation of locomotor activity. Neuron *55*, 435–447.

1026 Takemura, S., Bharioke, A., Lu, Z., Nern, A., Vitaladevuni, S., Rivlin, P.K., Katz,
1027 W.T., Olbris, D.J., Plaza, S.M., Winston, P., et al. (2013). A visual motion detection
1028 circuit suggested by Drosophila connectomics. Nature *500*, 175–181.

1029 Takemura, S.-Y., Aso, Y., Hige, T., Wong, A., Lu, Z., Xu, C.S., Rivlin, P.K., Hess, H.,
1030 Zhao, T., Parag, T., et al. (2017a). A connectome of a learning and memory center in
1031 the adult Drosophila brain. ELife *6*.

1032 Takemura, S.-Y., Nern, A., Chklovskii, D.B., Scheffer, L.K., Rubin, G.M., and
1033 Meinertzhagen, I.A. (2017b). The comprehensive connectome of a neural substrate
1034 for “ON” motion detection in Drosophila. ELife *6*.

1035 Tanaka, N.K., Endo, K., and Ito, K. (2012). Organization of antennal lobe-associated
1036 neurons in adult Drosophila melanogaster brain. J. Comp. Neurol. *520*, 4067–4130.

1037 Thoen, H.H., Marshall, J., Wolff, G.H., and Strausfeld, N.J. (2017). Insect-Like
1038 Organization of the Stomatopod Central Complex: Functional and Phylogenetic
1039 Implications. Front. Behav. Neurosci. *11*, 12.

1040 Tobin, W.F., Wilson, R.I., and Lee, W.-C.A. (2017). Wiring variations that enable and
1041 constrain neural computation in a sensory microcircuit. ELife *6*.

1042 Tomchik, S.M., and Davis, R.L. (2009). Dynamics of learning-related cAMP signaling
1043 and stimulus integration in the Drosophila olfactory pathway. Neuron *64*, 510–521.

1044 Treiber, C.D., and Waddell, S. (2017). Resolving the prevalence of somatic
1045 transposition in Drosophila. ELife *6*, e28297.

1046 True, J.R., Yeh, S.-D., Hovemann, B.T., Kemme, T., Meinertzhagen, I.A., Edwards,
1047 T.N., Liou, S.-R., Han, Q., and Li, J. (2005). Drosophila tan Encodes a Novel
1048 Hydrolase Required in Pigmentation and Vision. PLOS Genet. *1*, e63.

1049 Van Der Maaten, L. (2014). Accelerating t-SNE Using Tree-based Algorithms. J
1050 Mach Learn Res *15*, 3221–3245.

1051 Waltman, L., and Eck, N.J. van (2013). A smart local moving algorithm for large-scale
1052 modularity-based community detection. Eur. Phys. J. B *86*, 471.

1053 Wolff, G.H., and Strausfeld, N.J. (2015). Genealogical correspondence of mushroom
1054 bodies across invertebrate phyla. Curr. Biol. CB *25*, 38–44.

Yamagata, N., Ichinose, T., Aso, Y., Plaçais, P.-Y., Friedrich, A.B., Sima, R.J., Preat, T., Rubin, G.M., and Tanimoto, H. (2015). Distinct dopamine neurons mediate reward signals for short- and long-term memories. Proc. Natl. Acad. Sci. U. S. A. *112*, 578–583.

Yamamoto, S., and Seto, E.S. (2014). Dopamine dynamics and signaling in Drosophila: an overview of genes, drugs and behavioral paradigms. Exp. Anim. *63*, 107–119.

Zheng, Z., Lauritzen, J.S., Perlman, E., Robinson, C.G., Nichols, M., Milkie, D., Torrens, O., Price, J., Fisher, C.B., Sharifi, N., et al. (2017). A Complete Electron Microscopy Volume Of The Brain Of Adult Drosophila melanogaster. BioRxiv 140905.

Ziegenhain, C., Vieth, B., Parekh, S., Reinius, B., Guillaumet-Adkins, A., Smets, M., Leonhardt, H., Heyn, H., Hellmann, I., and Enard, W. (2017). Comparative Analysis of Single-Cell RNA Sequencing Methods. Mol. Cell *65*, 631–643.e4.

Figure legends

Figure 1 – Drop-seq reveals neuronal clusters in the *Drosophila* brain. (A) Schematic of the experimental procedure. *Drosophila* brains were dissected and dissociated prior to Drop-seq. After sequencing and alignment, a digital expression matrix containing information about the number of UMIs found for each gene, in each cell, was generated and used for PCA and subsequent analyses. See Methods section for details. **(B)** Two-dimensional representation (t-SNE) of 10,286 *Drosophila* brain cells, manually classified into 28 clusters. Based on the recovery of cell-types of known abundance in the brain, we estimate that there are 45,000 cells in the fly midbrain.

Figure 2 – Identification of Kenyon Cells and mushroom body-specific genes. (A) mCherry labeling of MB008B neurons in the *Drosophila* brain. Neuropils, labeled by nc82 anti-Brp antibody, are shown in grey. **(B)** Expression of mCherry in the t-SNE-clustered brain cells shown in Figure 1B. mCherry-positive cells are labeled in red and represent αβ Kenyon Cells (KCs). Intensity of red (or other colors in the panels below) is proportional to the normalized expression level. **(C)** and **(D)** Expression of *eyeless* and *Dop1R2* (*damb*), in t-SNE-clustered brain cells. The three numbered clusters containing blue-labeled cells are KCs. **(E)** Expression of *sNPF*, *Fas2* and *trio* in the three t-SNE clusters numbered in (C) and (D). Cells in light blue, orange and purple express each of these genes in αβ, γ, and α′β′ KCs, respectively. *sNPF* and *Fas2* are mostly expressed in αβ and γ KCs, while *trio* is mostly expressed in γ and α′β′ KCs. **(F)** Violin plots showing the main markers that distinguish KC subtypes of KCs from each other (pairwise comparisons for genes expressed in > 50% of cells in either cluster; Log2 FC > 1.5, Wilcoxon rank-sum test with Bonferroni-corrected p-value < 0.01). The right hand column indicates expression of these genes in non-MB neurons.

Figure 3 – Sub-populations of olfactory projection neurons. (A) Re-clustering of the two Projection Neuron (PN) clusters from Figure 1B. Clusters are color-coded. **(B)** Dot plots showing the main markers distinguishing PN populations from each other (pairwise comparisons for genes expressed in > 50% of cells in either cluster; Log2 FC > 1.5, Wilcoxon rank-sum test with Bonferroni-corrected p-value < 0.01). Dot diameter represents the fraction of cells expressing each gene in each cluster, as shown in scale. Color intensity represents the average normalized expression level. **(C)** t-SNE-plots for some known markers of PNs. *ct* labels all PN clusters, *acj6* and

vvl are mutually exclusive in Clusters 1, 2 and 4, versus Cluster 3, while *Lim1* is specifically expressed in Cluster 4.

Figure 4 – Distribution of fast-acting neurotransmitters. (A) t-SNE plot showing the distribution of cells expressing *vesicular acetylcholine transporter* (*VAChT*, cholinergic neurons), *vesicular glutamate transporter* (*VGlut*, glutamatergic neurons) and *glutamic acid decarboxylase 1* (*Gad1*, GABA-ergic neurons). For graphical reasons only cells expressing each marker above a log normalized value of 2 are shown. **(B)** Quantification of cells expressing markers displayed in A. The difference to 100% are cells that did not express any of the three markers.

Figure 5 – Co-expression of neuropeptides in fast-acting neurotransmitter neurons. (A) Radar plots showing the co-expression of 16 neuropeptides with the three fast-acting NTs. Displayed is the deviation of NT co-expression from the average distribution of each gene and combination of genes. **(B)** Distribution of four *insulin-like peptides*, including the fat-body specific *Ilp6*, based on the deviation from the random distribution of fast-acting neurotransmitter.

Figure 6 – Genetic markers and co-transmission in monoaminergic neurons. (A) Expression of the Vesicular Monoamine Transporter (*Vmat*) labels three clusters from Figure 1B (cells in blue, highlighted with arrowheads and circles. **(B)** Re-clustering of the three cell populations labeled in (A). Four sub-clusters are identified, representing dopaminergic, tyraminergic, octopaminergic, and serotonergic neurons. **(C)** Dot plots showing the main markers distinguishing monoamine populations from each other (pairwise comparisons for genes expressed in > 50% of cells in either cluster; Log2 FC > 1, Wilcoxon rank sum test with Bonferroni-corrected p-value < 0.01). Dot diameter represents the fraction of cells expressing each gene in each cluster, as shown in scale. Color intensity represents the average normalized expression level. **(D)** Percentage of cells in each monoaminergic cluster that are co-expressing markers for fast neurotransmitter-releasing neurons (*VGlut*, *Gad1*, and *VAChT*). **(E)** Dot plots showing expression of genes encoding neuropeptides and neuropeptide amidating enzymes across monoaminergic populations. Dot diameter represents the fraction of cells expressing each gene in each cluster, as shown in scale. Color intensity represents the average normalized expression level. **(F)** Comparison of genes overexpressed in PAM dopaminergic neurons compared to the rest of the brain, measured with Drop-seq or with FACS and Smart-seq2. 9 of the 15 genes identified with Drop-seq (labeled in bold) were also found in the Smart-seq2 dataset.

Figure 6 – Localization of genes involved in dopamine metabolism and signaling. (A) Schematics of a dopaminergic synapse, representing the major proteins involved in dopamine signaling and metabolism. Drawing inspired by Yamamoto and Seto, 2014 **(B)** Dot plots showing the expression of these genes across all cell populations identified in the *Drosophila* brain. Dot diameter represents the fraction of cells expressing each gene in each cluster, as shown in scale. Color intensity represents the average normalized expression level.

Figure 8 – Expression patterns of nicotinic acetylcholine receptors. (A) Prevalence of nicotinic receptor subunits. **(B)** Heatmap showing Pearson correlation coefficient Z-scores for each receptor subunit pair.

Figure 9 – Co-expression of neuronal activity markers. Heatmap showing Pearson correlation coefficient Z-scores of activity-regulated genes (ARGs), as reported by Chen et al., 2016 (histogram on top). Ten most highly upregulated genes following ChR2-XXL-induced activation of all neurons that are expressed in the brain, ranked by their correlation.

Supplementary Figure Legends

Figure 1 – Figure supplement 1– Preliminary validation of Drop-seq on insect cells. (A) Schematic of a Drop-seq experiment that was run to analyze a mixture of *Drosophila* (S2; blue) and *Spodoptera* (Sf9; red) cells. **(B)** Scatter plot (Barnyard) indicating the number of transcripts from *Drosophila* or *Spodoptera* associated with each STAMP. Of 768 STAMPs analyzed, only 12 (1.56%) carried transcripts from both species (purple dots).

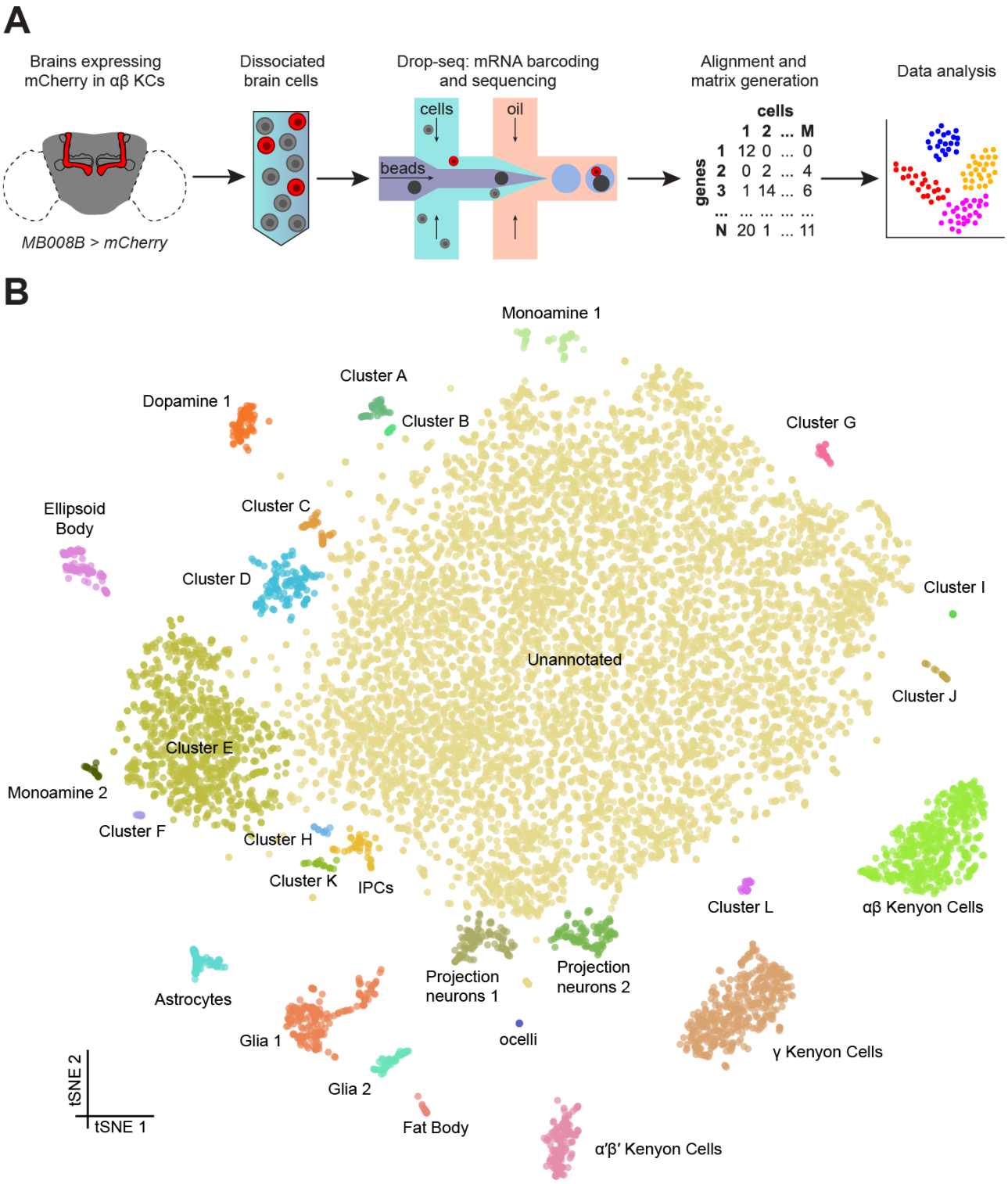
Figure 1 – Figure supplement 2 – Comparison of different quality filters. (A-C) Left: histograms representing the distribution of numbers of UMIs per cell selected with cutoffs of 200 (A), 800 (B) and 1800 (C) UMIs per cell. Right: t-SNE plots of the same data set as in Fig. 1B. obtained with these cutoffs to compare the impact of number of cells and UMI coverage on t-SNE clusters.

Figure 1 – Figure supplement 3 – t-SNE plot showing all eight replicates. t-SNE plot as in Fig. 1B, with cells from each replicate in a different color.

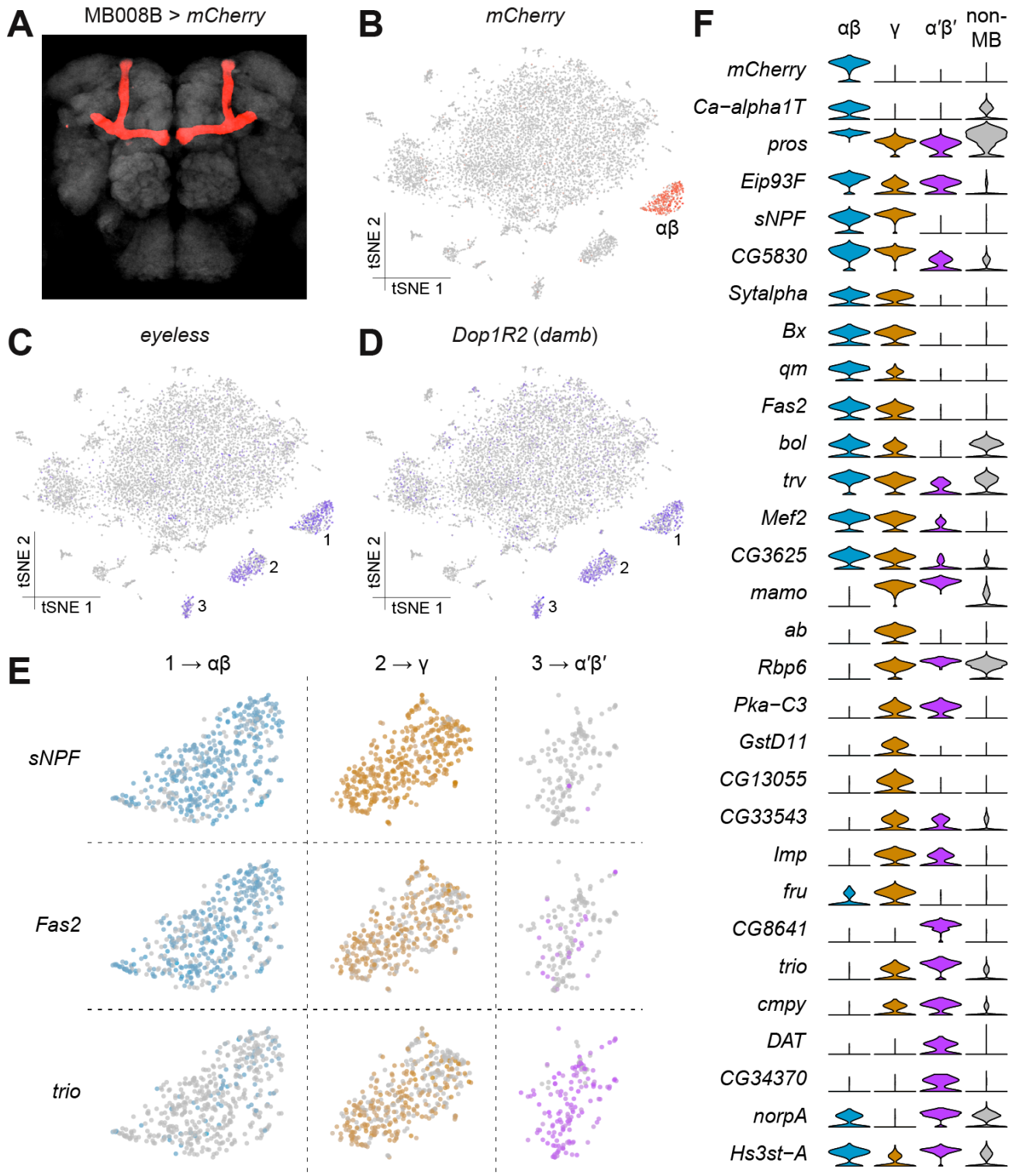
Figure 1 – Figure supplement 4 – Sex-determination of individual brain cells. (A) Histogram showing the number of UMIs per cell. The dashed line indicates the threshold that was chosen to separate *roX1*-positive and *roX1*-negative cells. **(B)** t-

1172 SNE plot showing putative male and female cells, based on *roX1* expression. Pie-
1173 chart shows the frequency of “male” and “female” cells.

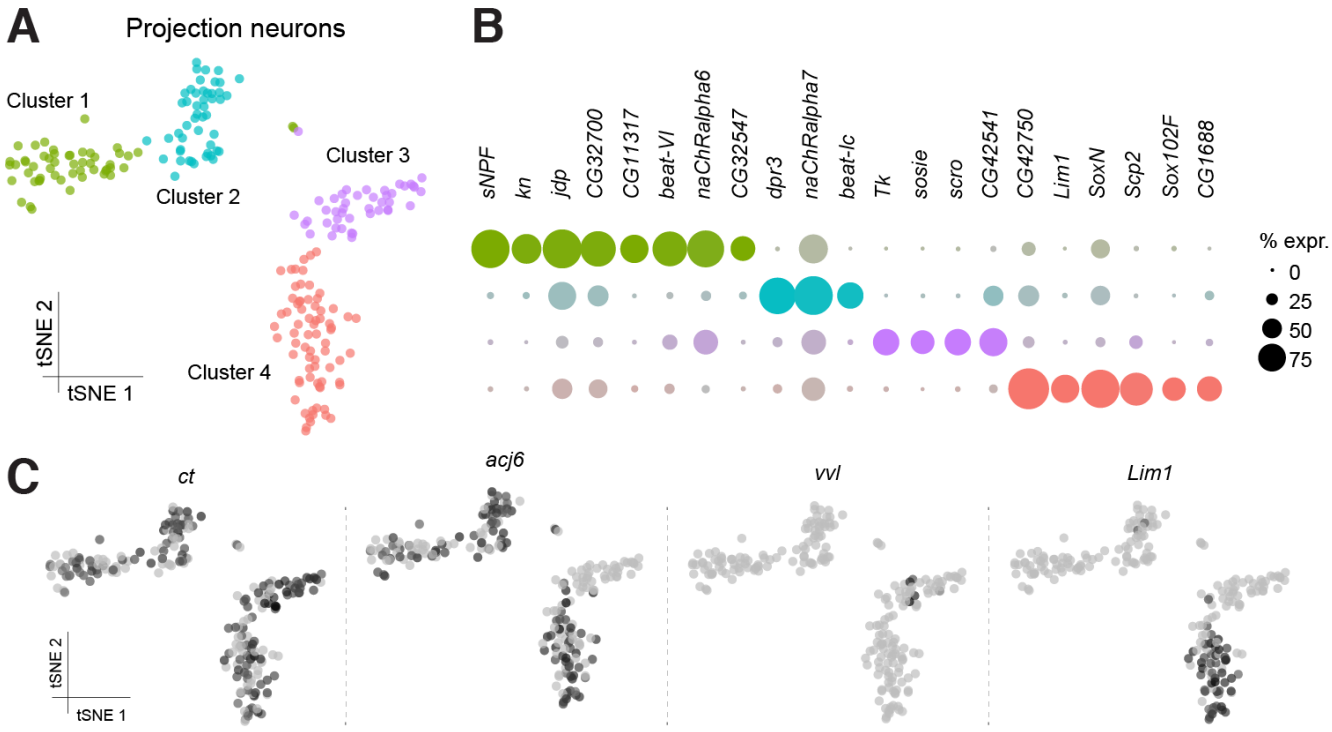
1174 **Figure 1**



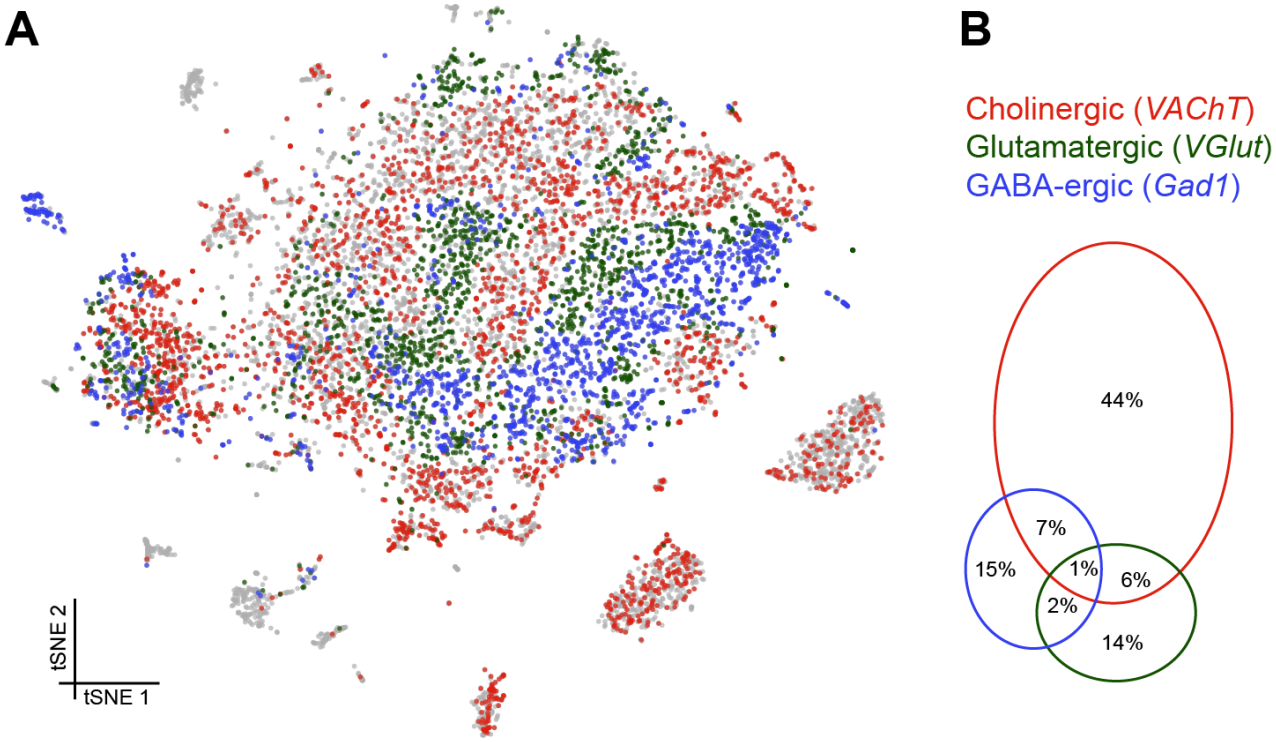
1175 **Figure 2**



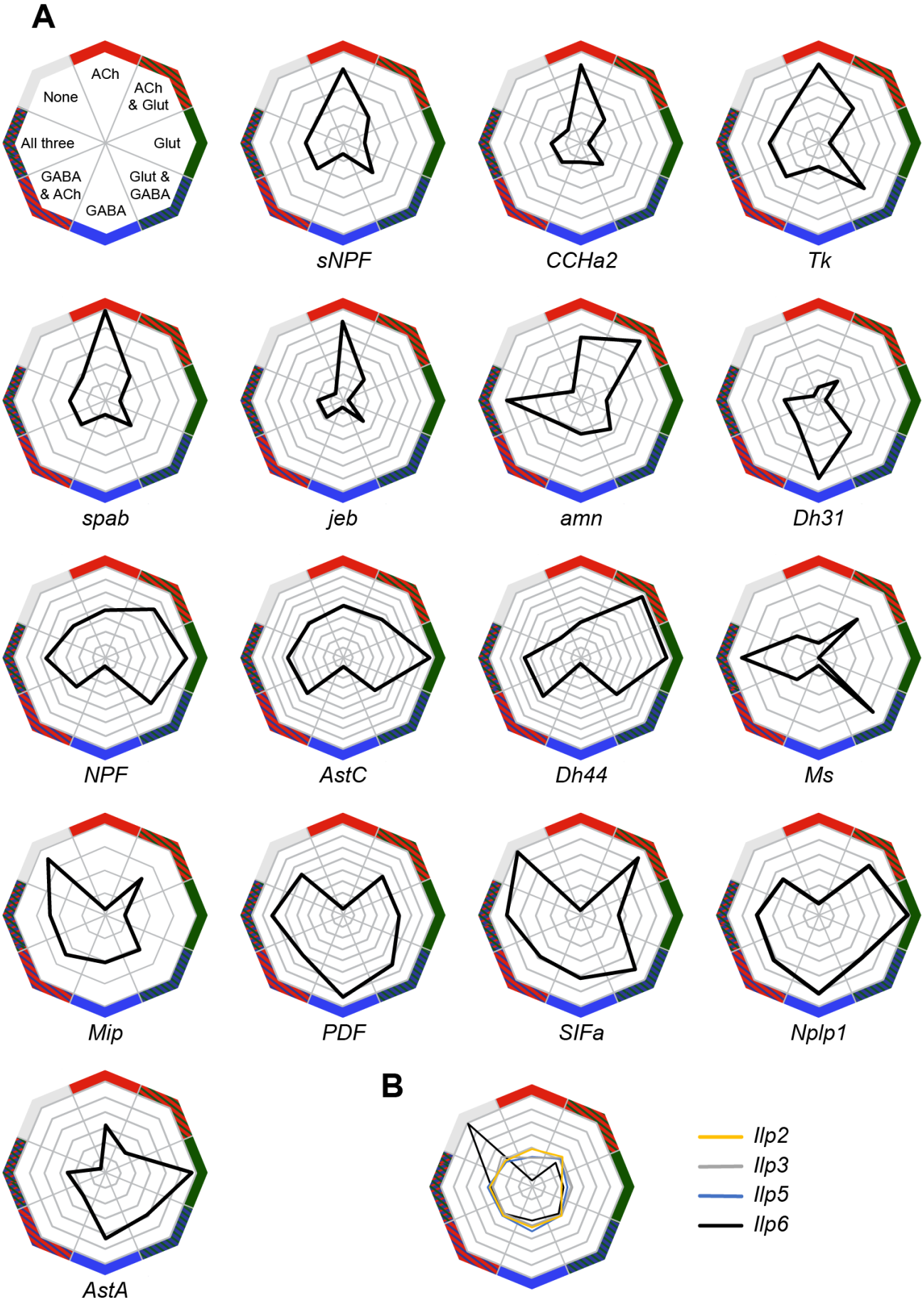
1176 **Figure 3**



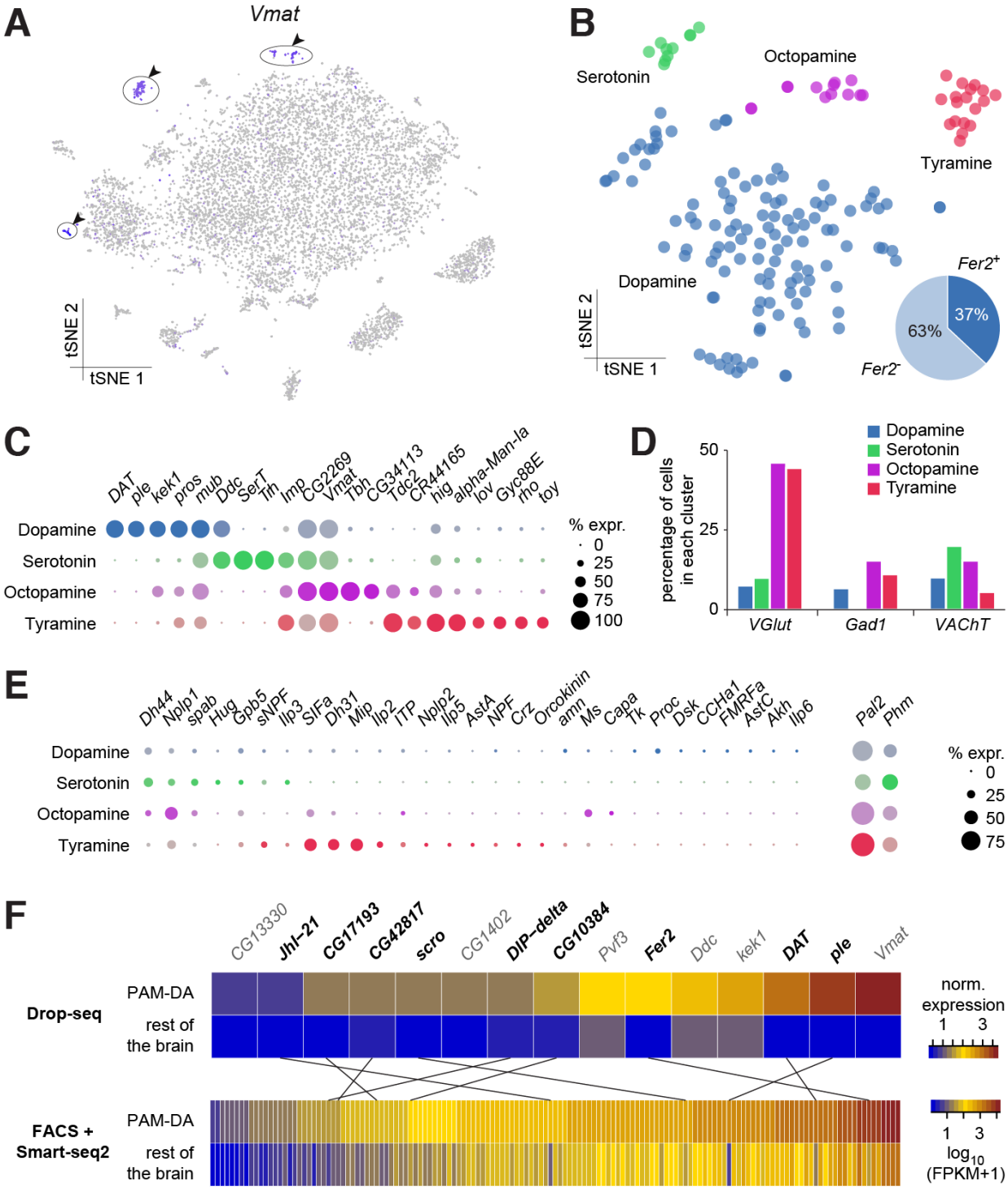
1177 **Figure 4**



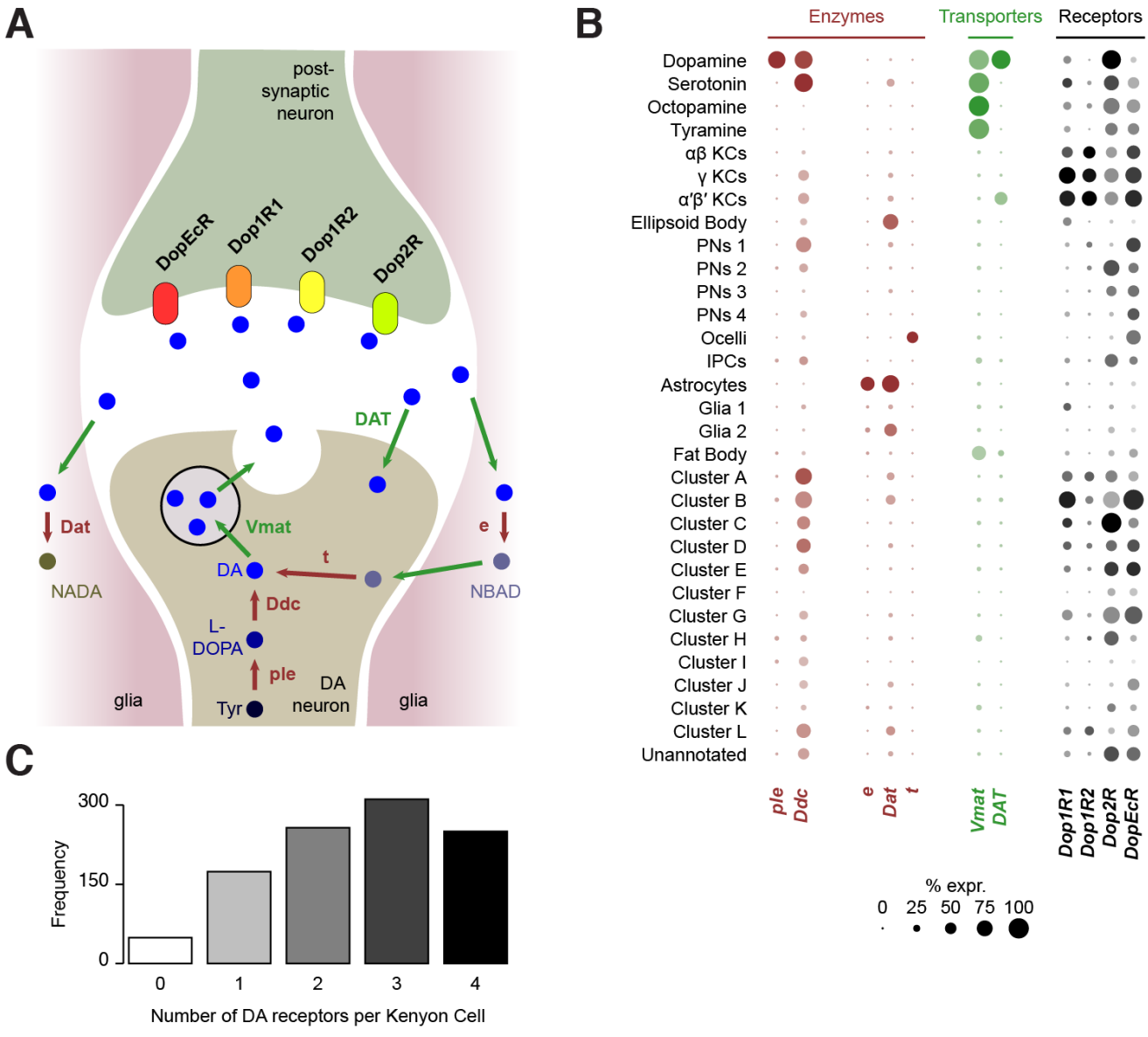
1178 **Figure 5**



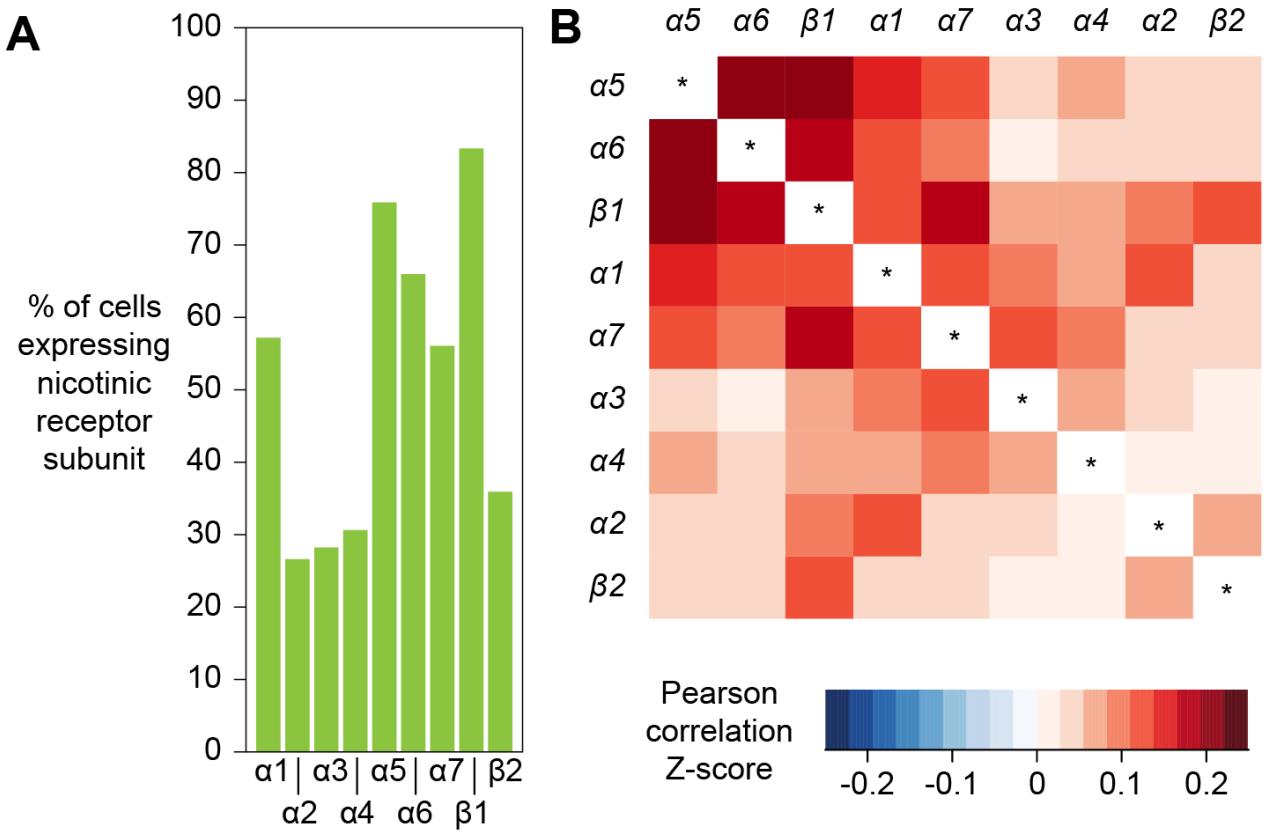
1179 **Figure 6**



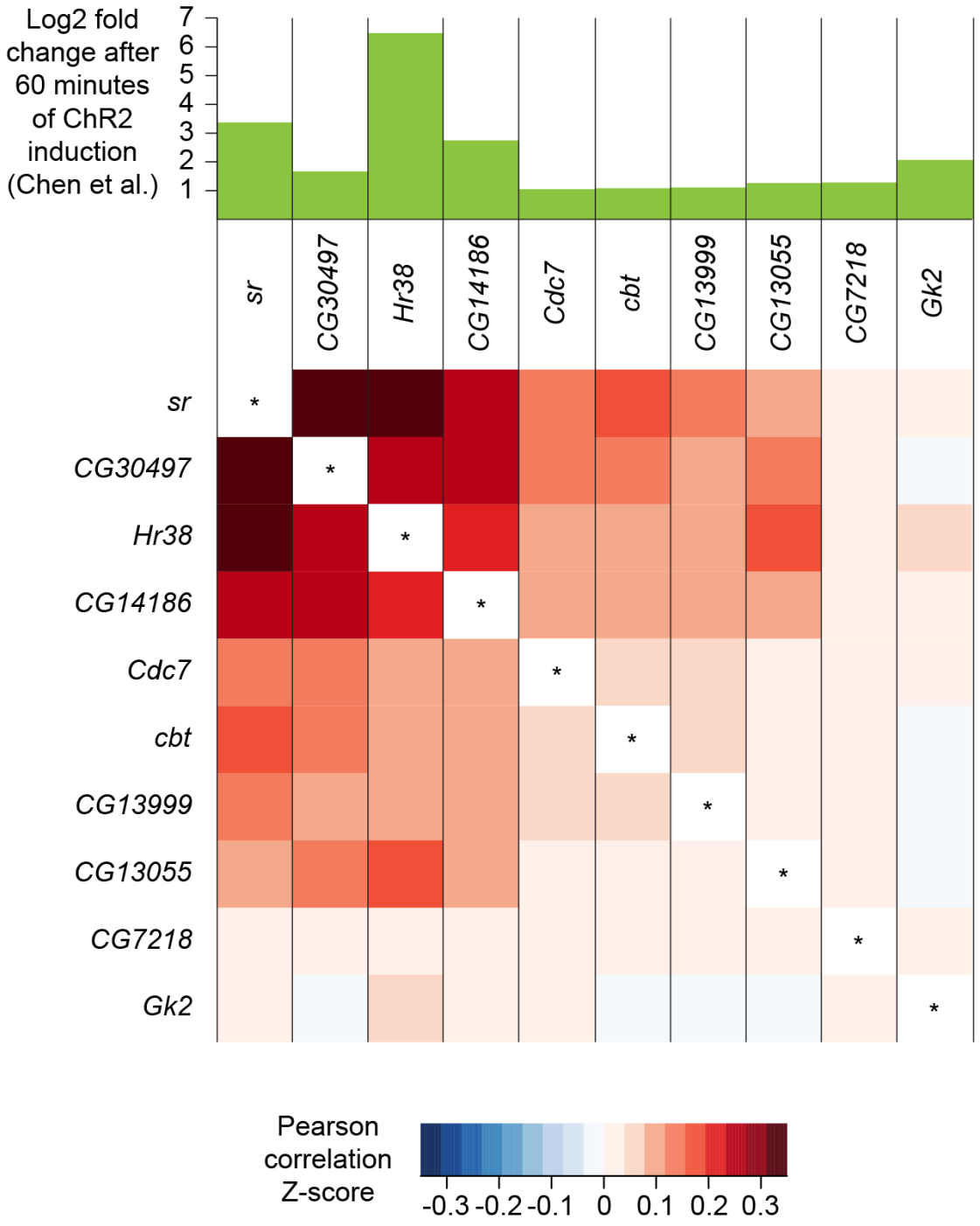
1180 **Figure 7**



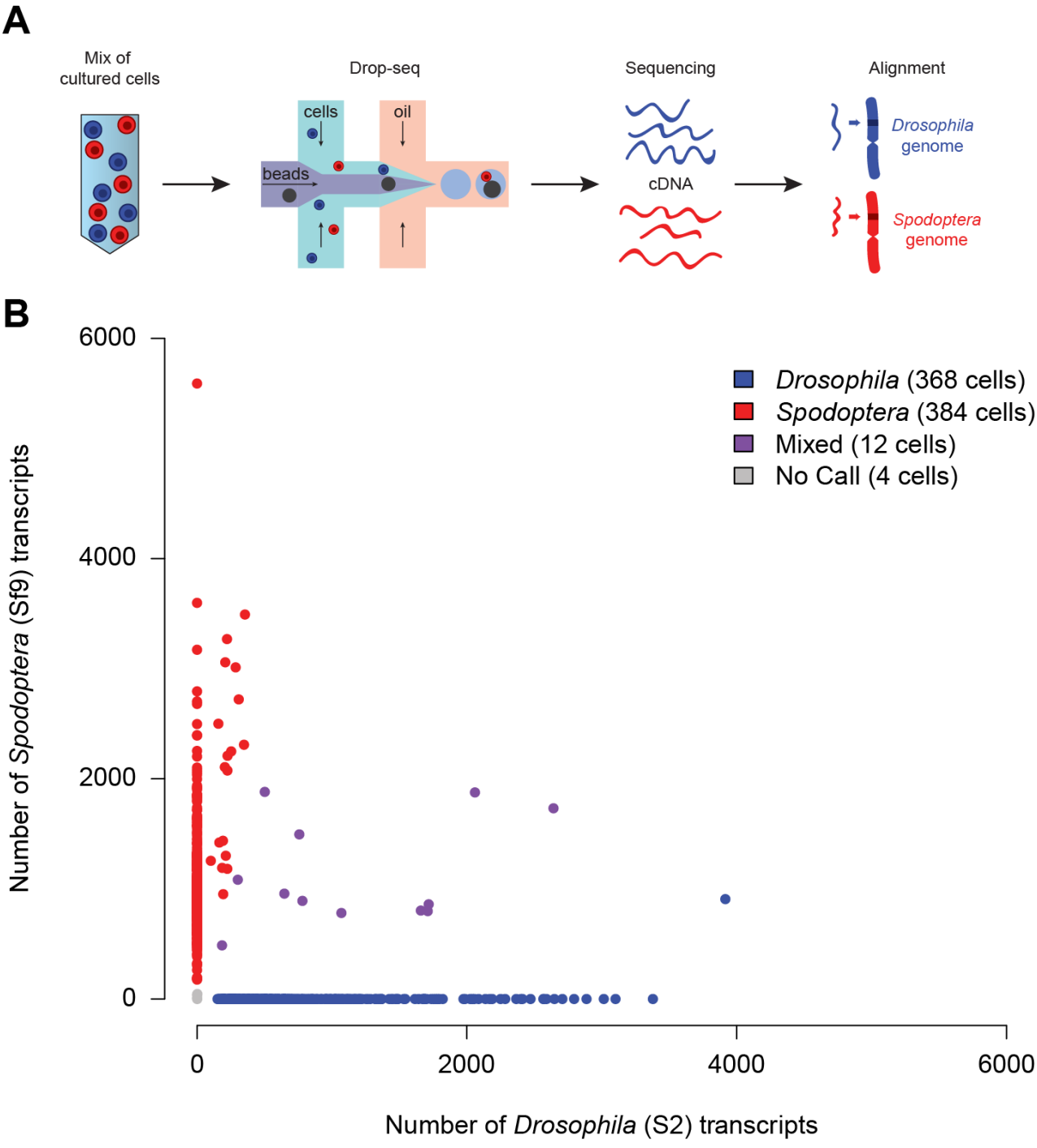
1181 **Figure 8**



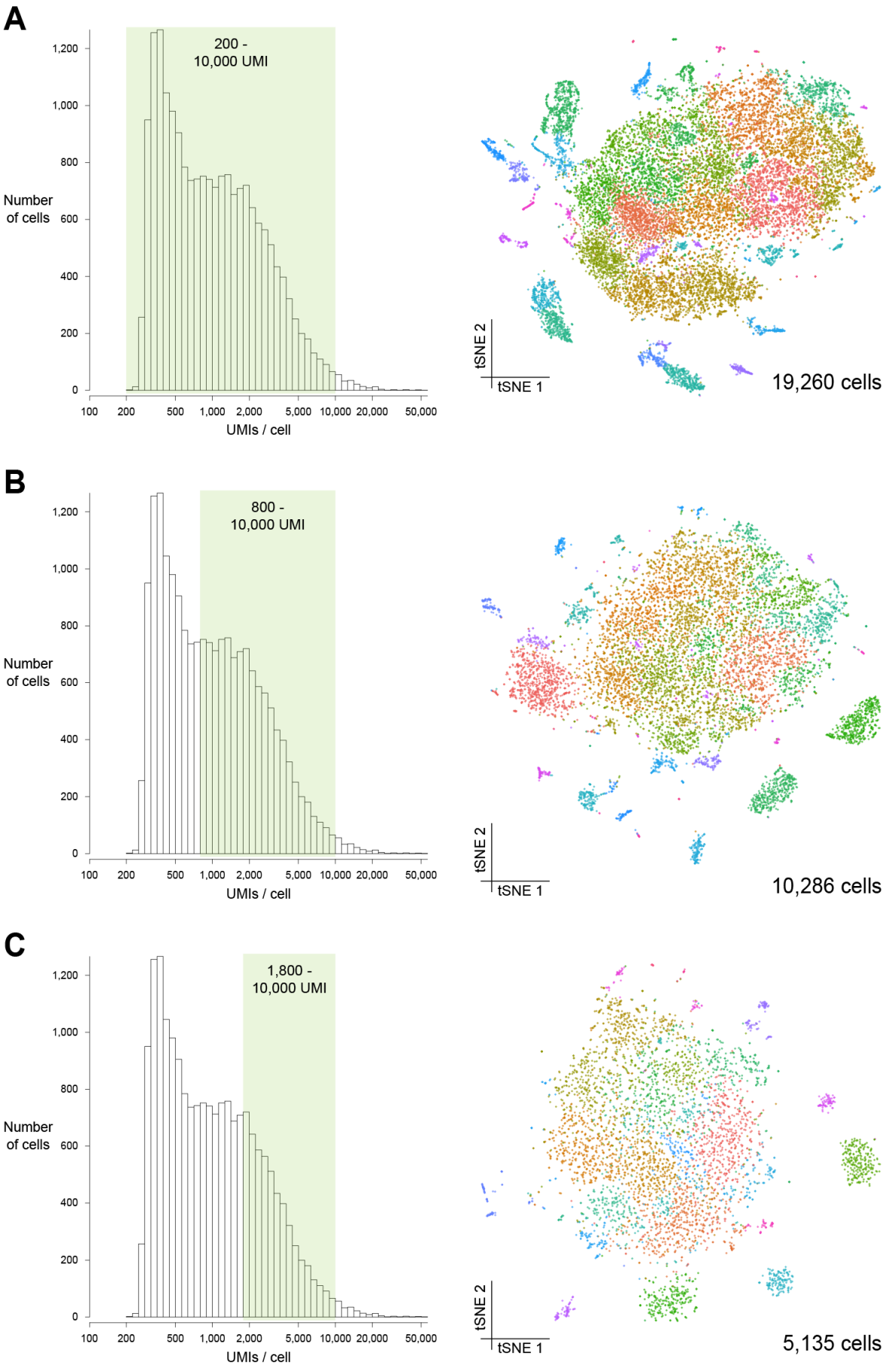
1182 **Figure 9**



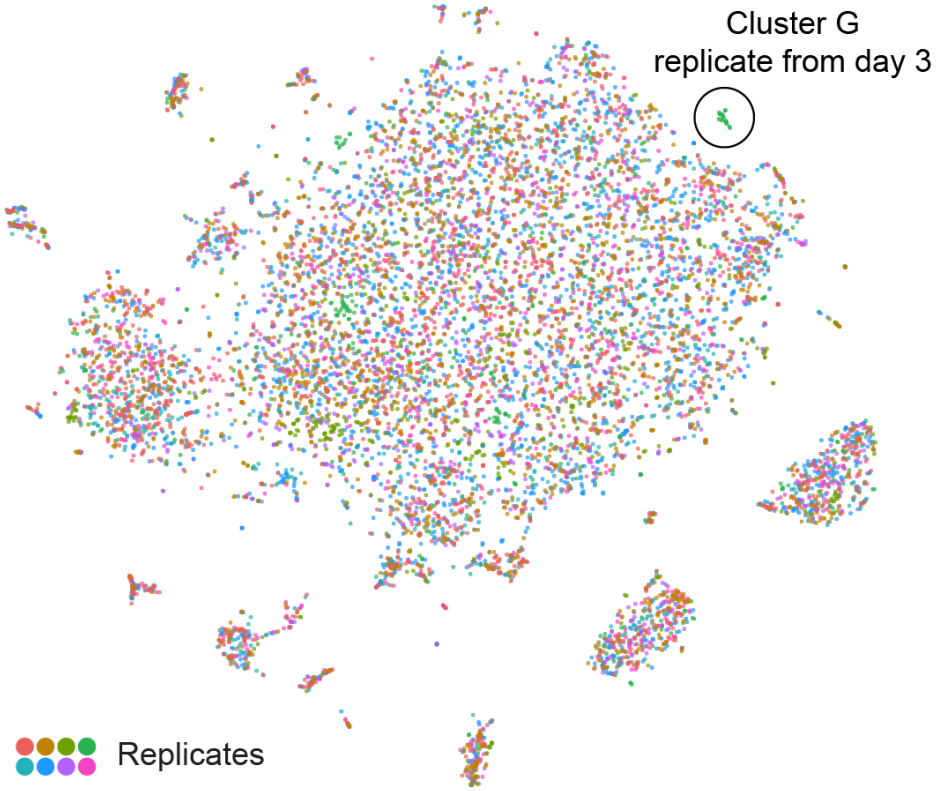
1183 # Figure 1 - Figure supplement 1



1184 # Figure 1 - Figure supplement 2



1185 # Figure 1 - Figure supplement 3



1186 # Figure 1 - Figure supplement 4