Fate transitions in *Drosophila* neural lineages: a single cell road map to mature neurons.

AUTHORS

Graça S. Marques[1], José Teles-Reis[1], Nikolaos Konstantinides[2], Patrícia H. Brito[3], Catarina C. F. Homem[1]

[1]CEDOC, NOVA Medical School / Faculdade de Ciências Médicas, Universidade NOVA de Lisboa, Lisboa, Portugal

[2]Department of Biology, New York University, New York, NY 10003, USA

[3]Applied Molecular Biosciences Unit-UCIBIO, Departamento de Ciências da Vida, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Caparica, Portugal

CORRESPONDENCE: catarina.homem@nms.unl.pt

ABSTRACT

Brain development requires the formation of thousands of neurons and glia and their coordinated maturation to ensure correct circuit formation. Neuronal fate is determined by several layers of gene regulatory networks during neuronal lineage progression. Once specified, neurons must still undergo a critical maturation period, involving the stepwise expression of functional feature genes such as cell surface molecules, ion channels or neurotransmitter receptors. The precise mechanisms that govern neuronal maturation remain however poorly understood. Here, we use single-cell RNA sequencing combined with a conditional genetic strategy to select and analyse neural lineages and their young progeny at a restrictive timepoint to investigate the transcriptional trajectories in young developing secondary neurons in the *Drosophila* larval brain.

Our findings reveal that neuron maturation starts very quickly after neuronal birth, and sub-divide the process of maturation into 3 distinct phases: Phase 1 is composed by immature neurons that have yet to start expressing mRNA of mature neuronal features; Phase 2 includes neurons that start transcribing but not

translating maturation markers such as neurotransmitter genes; Phase 3 neurons start translating mature neuronal features, in a coordinated fashion with the animal developmental stage.

This dataset represents a complete transcriptomic characterization of the neural lineages generated at this larval stage in the central brain and ventral nerve cord. Its analysis has also allowed for the characterization of a yet undefined transitional state, the immature ganglion mother cells, and has proven useful for the identification of known and novel fate regulators.

Keywords: Single-cell RNA sequencing, *Drosophila*, Neuron maturation, Neuroblasts, Brain, Neuronal lineage differentiation, Neural fate regulators

INTRODUCTION

Brain development is a highly complex process that requires the coordinated generation and maturation of thousands of different types of neurons and glia to ensure the formation of complex neuronal circuits. During brain development a small number of neural stem cells (NSCs) gives rise to the large neuronal diversity in the adult. Generally, NSCs divide to generate intermediate progenitors that then generate differentiated neurons and glia. Neural lineages play pivotal roles in cell fate determination, as transcriptional and molecular changes that occur at each step of lineage progression can be inherited and ultimately determine neural cell fate specification. Indeed, the type of neuron generated by each neural lineage has been shown to be determined by the combination of several layers of transcription factors and signaling pathways acting at distinct steps of lineage progression. At the level of the NSC, a series of spatial and temporal transcription factors endow NSCs a dynamic identity which ensures the formation of diverse neuron fates throughout development (Homem et al., 2015). A second series of transcriptions factors, expressed at the level of intermediate neural progenitors, further expands neuronal fate (Briscoe and Novitch, 2008; Landgraf and Thor, 2006; Wang et al., 2014). Neuronal fate can be additionally regulated by Notch signaling, which has been shown to diversify neuron type by mediating binary fate decisions at the level of the differentiated neuronal cell (Cau and Blader, 2009; Kimura et al., 2008; Shin et al., 2007; Truman et al., 2010). After the specified neurons are born, they must still

undergo a maturation period, during which they generate an axon and dendrites, form synapses and generate neuronal circuitries (Allan and Thor, 2015; Hobert, 2016). Thus, during maturation, neurons must coordinately express a combination of effector molecules such as cell surface molecules, ion channels and neurotransmitter (NT) receptors to ensure that neuronal wiring occurs only when their synaptic partners are formed (Kurmangaliyev et al., 2020). Several studies have transcriptionally analyzed mature adult neurons (Allen et al., 2020; Davie et al., 2018; Konstantinides et al., 2018; Özel et al., 2020), but very little is known about the mechanisms underlying the early stages of neuronal maturation (Bradke et al., 2020). Interestingly, it has been recently shown that the age-related transcriptomic diversity of neurons is partially lost as early as 15h after neuron birth, resulting in a transcriptomic convergence in mature neurons (Özel et al., 2020). This further indicates that the transcriptomic profiles involved in the initial phases of neuron maturation are quickly lost and can no longer be detectable in adult neurons.

Thus, the progression from NSCs to mature and functional neurons critically rely on the coordinated and precise fine-tuning of complex gene expression programs. To date, extensive efforts have been made to characterize neural lineages and the genes and mechanisms involved in neuronal specification. However, a complete view of the molecular and cellular events regulating the progression from NSCs to mature neurons is still lacking. In particular, very little is known about the mechanisms that drive neuron maturation, essential for the formation of functional neurons.

To characterize the initial transcriptional changes in neural lineages required for neuronal formation and maturation, we have analyzed the transcriptomes of neural lineages in the developing *Drosophila* larval brain by single cell sequencing. *Drosophila*'s stereotypical neuronal lineages, the existence of many cell-type-specific markers, and precise developmental time points for neuron formation and synaptogenesis make it an ideal model to understand the genetic programs that regulate neural lineage progression and neuronal maturation. *Drosophila* have approximately 150,000 neurons in the entire brain, one third in the central brain (CB) and ventral nerve cord (VNC) and the remaining in the optic lobe (OL). The adult CB and VNC neurons are in their vast majority generated during larval stages, and these neurons form many key brain structures such as the mushroom bodies and central complex, which are, including other things, critical for memory-directed behavior and

navigation (Cognigni et al., 2018; Seelig and Jayaraman, 2015). Neurons are produced from asymmetrically dividing NSCs called neuroblasts (NBs). Upon asymmetric division, NBs give rise to an intermediate progenitor that usually divides once to give rise to two neurons or glia. Neurogenesis starts in the embryo (primary neurogenesis), when the neurons that form the embryo and larval central nerve system are formed (primary neurons). The second wave of neurogenesis (secondary neurogenesis) occurs in larval and early pupal stages when the large majority of adult neurons are formed (secondary neurons) (Truman and Bate, 1988). After secondary neurons are formed, they remain immature until mid- pupal stages when synaptogenesis starts (Hartenstein, 1993; Özel et al., 2015, 2019; Muthukumar et al., 2014). The regulatory processes and gene networks involved in the formation of secondary neurons, particularly the ones from the CB and VNC, are however understudied. Previously generated bulk-seq datasets of the developing brain include both primary and secondary neurons and thus did not allow to distinguish adult neurons nor neurons with different ages. The available single-cell transcriptomic atlases did not cover the stages or the brain areas when midbrain adult neurons are generated (Allen et al., 2020; Avalos et al., 2019; Davie et al., 2018; Konstantinides et al., 2018; Kurmangaliyev et al., 2019, 2020; Özel et al., 2020); focused only on a sub-set of CB (type II lineages, not including type II NBs themselves) mixed with optic lobe lineages (Michki et al., 2021); or analyzed brains from the adult animal which contain only post-mitotic fully mature neuronal cells, (Davie et al., 2018; Konstantinides et al., 2018). Thus, to characterize secondary neurogenesis and the mechanisms that regulate secondary neuron maturation we used a conditional genetic strategy to select CB and VNC NBs and their immediate progeny, including only 0h up to 12.5h-old neurons (a time window prior to neuron transcriptomic convergence (Özel et al., 2020). Using the Chromium system (10x Genomics) we obtained ~12.6K single-cell transcriptomes of NBs and their daughter cells, ~7.2K of which were neurons.

This dataset allowed us to obtain the full transcriptome of all the described cell types in the developing midbrain, including type I and type II NBs, intermediate neural progenitors (INPs), ganglion mother cells (GMCs) and young secondary neurons in their first phases of maturation. It additionally permitted the identification, characterization and validation of an intermediate cell fate state, the immature

GMCs. It has also allowed us to identify several known as well as novel candidate regulators of lineage progression, as TFIIFα, which was experimentally validated.

Having specifically included only young neurons, of defined ages, allowed us for the first time to transcriptionally characterize the initial phases of neuron maturation. Although synaptogenesis is only described to start days later at pupal stages, (Chen et al., 2014; Hartenstein, 1993; Muthukumar et al., 2014; Özel et al., 2015, 2019) we have found that the young secondary larval neurons present in our dataset already transcribe several effector genes characteristic of mature neurons, such as ion channels and neurotransmitter (NT) receptors. This was unexpected as synaptogenesis and NT signaling genes have been shown to be transcriptionally coordinated in other systems (Kratsios et al., 2015). We further show that NT genes, such as ChaT, is transcribed but not translated in young neurons and that translation of ChAT protein does not solely depend on neuron age but is rather coordinated with the animal developmental stage only starting in pupal stages. We found that translation of ChAT initiates around stage P6-P7 (~48h after puparium formation, APF), just prior to the start of synaptogenesis in the CB (~60h APF, (Muthukumar et al., 2014)). Our results have allowed us to sub-divide the initial phases of secondary neuron maturation into 3 phases: a first phase when newly born neurons do not transcribe NT or effector genes; a second phase of maturation when neurons start transcribing, but not translating NT, ion channels and other neuron terminal feature genes; a third phase of neuron maturation when NT genes as ChAT start being translated in a coordinated manner with animal developmental stage.

RESULTS

Atlas of early fates in NB lineages in CB and VNC

To characterize the initial transcriptional changes that drive larval neural lineage progression and early maturation of secondary neurons we have specifically labeled, isolated and performed single cell transcriptome sequencing of larval CB and VNC NBs, intermediate progenitors and youngest neurons. Single cell transcriptome analysis allowed us to obtain a representation of every neural cell state, including transient or intermediate states, and different stages of neuron maturation. To ensure that only the lineages that originate adult neurons were analyzed we devised a conditional genetic strategy to temporally label neuronal lineages from *Drosophila* 3rd instar larval CB and VNC. In order to fluorescently label and isolate these neural

lineages, but not optic lobe (OL), we used the CB and VNC NB specific Vienna Tile Gal4 line#201094 (VT#201094; **Figure 1A**). With this VT-Gal4 line we drove the expression of CD8::GFP specifically in NBs. As GFP protein is stable for several hours it is inherited by the NB progeny, effectively labeling neural lineages (**Figure S1A**). To control the time window of GFP expression, we included a temperature sensitive (ts) tubGal80 in these animals, which allows GFP expression at 25ºC, but represses GFP expression at 18ºC. With this conditional genetic system, we could precisely control when NBs start expressing GFP and thus generating GFP-labeled progeny.

For our experimental set-up, precisely staged wandering 3rd instar larvae (equivalent to 105h ALH) were used: animals were raised at 18ºC and shifted to 25ºC 18h prior to dissection to initiate GFP expression in NBs. This time window allows for NBs to divide 9-12 times, depending on lineage type, and for several GFP positive neurons to be generated. Neuronal lineages can be sub-divided in two types that differ in the structure of their lineage trees. Type I lineages are composed by NBs that divide every ~1.3h to self-renew and generate ganglion mother cells (GMCs), which after ~4.2h divide to form neurons and glia (Homem et al., 2013) (**Figure 1B**). Consequently, in a 18h time window, the oldest neurons in our dataset can be at most ~12.5h-old (18h-1.3h-4.2h=12.5h), although most likely they are slightly younger as the expression of GFP protein under UAS-Gal4 takes approximately 3h (personal communication, M. Garcez). Type II NBs divide every ~1.6h to self-renew and generate an additional intermediate neural progenitor cell (INP), which requires ~6.6h to mature and divide to self-renew and generate GMCs which then generate neurons and glia (**Figure 1B**). Hence, in type I lineages, NBs, GMCs and the young neurons formed during this time window will be GFP positive. In type II lineages, due to the extra differentiation step (INPs), this 18h labeling window allows for labeling of NBs, INPs and GMCs but fewer neurons (**Figure 1B,C**). As a result, the majority of neurons represented in this dataset are from type I origin.

We sorted the labeled neural lineages by FACS, based on their size and GFP expression. Two samples were processed in parallel with the Chromium system (10x Genomics) and analyzed using the standard Seurat pipeline (Satija et al., 2015; Stuart et al., 2019). Quality control (QC) metrics attested an overall good quality of the samples, such as low percentage of mitochondrial genes (**Figure S1B**) and allowed us to set appropriate filters, resulting in a library of 12671 cells and 10250

genes. Using Seurat's graph-based clustering approach we identified 39 clusters with distinctively expressed marker genes, which were visualized in low dimensional space using the Uniform Manifold Approximation and Projection (UMAP) algorithm (McInnes et al., 2018) (**Figure 1D**). Moreover, the 2 samples seemed to be evenly distributed through the different clusters, which strengthens the good quality of our dataset (**Figure S1C**).

By checking the expression of known marker genes, we were able to identify and annotate the different cells that make up neuronal lineages (**Figure 1E,F**). NBs were globally identified by the expression of known NB markers *dpn*, *mira*, *wor* and *klu* (**Figure 1F**). Type I NBs were distinguished as they additionally co- express *ase* (cluster 0; **Figure 1F**, **S1D**), while type II NBs do not (cluster 36; Figure 1F, S1D). Type II NBs and their lineage INPs and GMCs, were additionally identified by the expression of *Sp1* and *pnt* (**Figure 1F**).

INPs from type II lineages were separated into two clusters: developing INPs (dINPs, cluster 34; **Figure 1F**, **S1D**) which include the several stages of INP maturation; and fully mature INPs (cluster 22; **Figure 1F**, **S1D**). The "dINP cluster" includes immature INPs expressing *erm* and *ase*, and the more mature INPs that express *erm*, *ase* and also *dpn*. The "INP cluster" includes the fully mature INPs, which stop expressing *pnt* and *erm* and turn on *klu* expression (Berger et al., 2012).

GMCs were identified based on the expression of *wor* and *ase*, but not *dpn* and *mira* (**Figure 1F**). Type II GMCs could be distinguished as they expressed *Sp1* (cluster 15; **Figure 1F**, **S1D**), while type I GMCs did not (clusters 1, 3, 17, 32; **Figure 1F**, **S1D**). As GMCs do not have any specific marker, usually being identified by combinatorial expression of several genes, it has been hard to specifically isolate and analyze them. Our transcriptome dataset of both type I and type II GMCs thus represents a good tool for understanding GMCs and lineages transition from asymmetric dividing self-renewing NBs to symmetric terminal dividing GMCs.

Neurons were identified based on the positive expression of *pros* or *elav*, but negative expression of *dpn*, *mira* and *wor* (clusters 2, 4-8, 11-14, 16, 18-21, 24, 26-31, 35, 37; **Figure S1D**).

Finally, we identified a group of cells (cluster 9, **Figure S1D**) as an intermediate state between type I NBs and GMCs. This annotation is supported, not only by the location of this cluster in the UMAP plot (between type I NBs and GMCs; **Figure 1E**), but also because although transcriptional close to type I NBs, they have partially lost

the expression of *dpn*, an essential marker of NBs. Namely, most of these cells (~92%), still express the NB marker *mira*, but only a third still express *dpn* (~35%) (**Figure 1F**). As these cells also differ from GMCs, where both *dpn* and *mira* are not expressed, this led us to annotate them as immature GMCs (imGMCs; **Figure 1E,F**). Notably, the presence of this transitional state indicates that the fate commitment of type I GMCs does not occur immediately after asymmetric NB division and is rather a progressive process.

In our analysis we did not identify any cluster enriched for the glial cell marker repo, indicating that no glial cells were included in our dataset (**Figure S1D**). As neural lineages also generate glial cells (Doe, 2017), this absence might have been due to the limited temporal window analyzed. We have additionally classified clusters characterized by the expression of ribosomal subunit genes or with no obvious expression of any of the previously mentioned markers as low quality (Ilicic et al., 2016) or nonannotated clusters (**Figure 1E,F**).

This transcriptomic single cell atlas represents to our knowledge the first specific characterization of CB and VNC secondary neuronal lineages including only the first hours of secondary neuron maturation. It serves as a useful catalogue that allows for a robust analysis of wandering 3rd instar secondary neural lineages. Moreover, as it includes cells at different stages of their commitment or maturation, it allows for a thorough analysis of fate transitions as demonstrated by the identification of imGMCs; it also allows unequivocal investigation of early born adult neurons, as these were the only ones included.

Validation and characterization of imGMCs as a transition state between type I NBs and GMCs

To validate the existence of the transient imGMC population, we next carefully analysed neuronal lineages *in vivo*. Based on their UMAP location, the population we called imGMCs, is transcriptionally “between” type I NBs and GMCs (**Figure 1E**). The majority of imGMCs still express *mira*, but only one-third still express *dpn* (**Figure 1F**). The partial absence of *dpn*, an important NB marker, indicates that imGMCs are no longer fully committed to the stem cell fate and supports these cells' transitional state into a more committed cell fate. Backing up the existence of imGMCs, occasions when Dpn is retained in the nucleus of the daughter cells after NB division have been described to sometimes occur in type I lineages of the

anterior region and attributed to newborn GMCs (Boone and Doe, 2008). The presence of this intermediate cell state was also validated by us *in vivo* by immunofluorescence. Indeed, in some lineages, the daughter cells immediately next to type I NBs are Dpn+Mira+ (**Figure 2A**, arrow). We could clearly distinguish between NBs and their daughter cells, as daughter cells have a smaller size when compared to their progenitor type I NBs. The low number of these double Dpn+Mira+ imGMCs is in accordance with our transcriptomic data, where these Dpn+Mira+ imGMCs should only represent ~33% of imGMCs. This proves that the bioinformatic analysis of our dataset can efficiently discriminate transient cell states, such as imGMCs (**Figure 2B**).

The transition from NB to GMC involves dramatic changes in cell biology, as loss of self-renewal capacity and change from a cell-size wise asymmetric to symmetric division, next we wanted to identify genes that might be regulating this transition. The analysis of GMCs has been hindered by the lack of a unique cell marker for this cell fate, which has not allowed for their purification and analysis. As we have now determined that the transition from NB to GMC is a stepwise process, we compared the transcriptomes of imGMCs to their upstream type I NBs and to their downstream GMCs. We focused our analysis in differentially expressed transcription factors (TFs) and DNA binding genes (DNAB) as several TF/DNAB have been described to be master regulators of neural lineage fate transitions (e.g. Dpn, Ase, Erm, Osa; e.g. (Eroglu et al., 2014; San-Juán and Baonza, 2011; Weng et al., 2010)).

The comparison between imGMCs and type I NBs identified *E(spl)mγ*, a Notch target, as being downregulated in imGMCs (**Figure 2C**). On the other hand, imGMCs have increased expression of several co-transcriptional repressors (as *Dsp1*, *l(3)neo38*, *spen* and *sbb*) and of genes involved in neuron differentiation (as *chinmo*, *lola*, *ase*) (**Figure 2C**). A higher expression of transcriptional co-repressors and chromatin silencers is consistent with the required shutdown of the stem cell program in imGMCs to transition to a more neuronal committed fate. The comparison between imGMCs and mature GMCs has also identified several differentially expressed genes. As predicted from being a transitional state, imGMCs express higher levels of NB markers as *vfl* (or *zelda*, *zld*), *nab* and *dpn* (Reichardt et al., 2018) then their mature counterparts (**Figure 2D**). imGMCs also express higher levels of *EloC* (**Figure 2D**), involved in RNA Polymerase II (PolII) elongation, which knockdown has been shown to cause a defect in GMCs (Neumüller et al., 2011).

Interestingly, mature GMCs express higher levels of *Su(var)205* (or HP1; **Figure 2D**) which has been recently shown to be recruited by Pros to promote heterochromatin compaction to promote terminal neuronal differentiation in GMCs (Liu and Song, 2020). Supporting their increased neuronal commitment, GMCs also express higher levels of several neuron differentiation genes than imGMCs (e.g., *jigr1* and *chinmo*; **Figure 2D**).

Overall, these comparisons suggest that the transition from NBs to GMCs involves a quick downregulation of Notch, progressive transcriptional and chromatin silencing and heterochromatin formation, with simultaneous upregulation of several neuronal differentiation genes. It further identifies several differentially expressed genes between NB-imGMCs-GMCs which are candidate regulators of the NB to GMC transition.

Identification of candidate regulators of neuronal lineage progression by differential gene expression analysis

To identify cell specific markers and regulators of neuronal lineage progression in the remaining steps of neural lineage differentiation, we performed a comparative analysis of the transcriptomes for all intermediate progenitor states in both type I and type II lineages. With this goal in mind, we focused in the most differentially expressed TFs and DNAB genes between cell states, as previously done for imGMCs.

Type I NBs vs. Type I GMCs

In the previous section we have done a stepwise comparison between NBs, imGMCs and GMCs. Here we will perform a simpler comparison between NBs and GMCs (the stable cell identities), specifically from type I lineages.

Among the top TFs/DNAB genes that distinguish these two fates are known NB self-renewal factors as *dpn*, *grh*, *klu* and *E(spl)mγ* (Almeida and Bray, 2005; Berger et al., 2012; Zacharioudaki et al., 2012) and genes involved in neuron differentiation as *jigr1* and *lola* (Southall et al., 2014; Yuva-Aydemir et al., 2015) (**Figure 3A**). Notably, *grh* is identified as being more highly expressed in NBs *vs.* GMCs (**Figure 3A**) but is not a top differentially expressed gene in the comparison of NBs *vs.* imGMCs (**Figure 2C**). Consistently, *grh* expression has been previously observed *in vivo* in

both NBs and in the GMC closest to the NBs (Cenci and Gould, 2005), which positionally corresponds to imGMCs, but is absent in older, more mature GMCs.

As our transcriptional analysis revealed *E(spl)mγ* to be quickly and significantly downregulated in the transition from NB to imGMC, remaining low in mature GMCs (**Figure 2C**), we have used this marker to further validate our results *in vivo*. Consistently, *in vivo* analysis of a protein reporter revealed that *E(spl)mγ* is expressed in type I NBs, but its expression is no longer seen in the rest of the lineage, not even in the youngest GMCs closest to the NB (**Figure 3B**, arrow). Thus, *E(spl)mγ* expression pattern specific to NBs, and its quick downregulation in its immediate progeny, poses this gene as a first responder during differentiation and an important regulator of NB stemness and differentiation.

We have also identified several genes enriched in type I NBs with uncharacterized functions in NB lineages, such as *CG2051* (*Hat1*), *CG30403*, *CG4570*, potentially new fate regulators.

Type I GMCs vs. neurons

We then analysed the transition from GMCs to neurons in type I NB lineages. As expected, GMCs, being less differentiated than neurons, have more expression of genes associated with neural proliferation, such as *N*, *ase* and *wor* (**Figure 3C**). *HmgD*, a chromosomal protein involved in DNA bending and chromatin organization, was previously shown to be highly differentially expressed between NBs and neurons at the transcriptomic level (Yang et al., 2016). Our analysis now shows that *HmgD* seems to be downregulated mainly at the transition from GMCs to terminally differentiated neurons (**Figure 3C**).

Conversely, the top genes that where more expressed in neurons include genes related to neuronal development and circuitry, such as *Jim*, *Lin29* (or *dati*) and *Smox* (or *Smad2*) (Iyer et al., 2013; Schinaman et al., 2014; Zheng et al., 2006).

Type II NBs vs. dINPs

Type II lineages are characterized by the presence of transit amplifying progenitors, the INPs, which are able to divide 5-8 times to self-renew and generate GMCs, which in turn originate neurons and glia. Recently, the transcriptomic profile of type II INPs and their offspring GMCs and neurons were analyzed (Michki et al., 2021). The whole transcriptomic profile of single type II NBs remained however uncharacterized,

hindering the understanding of the genetic networks that drive these NBs to generate a transit amplifying progenitor. As previously mentioned, in our dataset, INPs are separated into dINPs (immature and maturating INPs) and INPs (fully mature INPs). The comparison between type II NBs and dINPs revealed several genes to be specifically upregulated in dINPs, as *Sp1* and *erm* (**Figure 3D**), which have been described to have a role in type II lineage differentiation (Álvarez and Díaz-Benjumea, 2018; Weng et al., 2010). Several genes that were found to be more highly expressed in GMCs *vs.* type I NBs are also found in the comparison dINPs *vs.* type II NBs (e.g., *Dsp1*, *l(3)neo38*, *Dref*; **Figure 3D**). There are also several common differentially expressed genes along the remaining intermediate progenitor states in type I and type II lineages (**Figure 3E**), reflecting a common genetic program acting during neural differentiation. Conversely, type II NBs express higher levels of *tll* (**Figure 3D**), a described marker of these cells (Hakes and Brand, 2020), nicely validating our analysis. The comparison between type II NBs and dINPs surprisingly revealed that a basal RNA PolII transcription factor, TFIIFα, is more highly expressed in NBs (Figure 3D). There have been several recent reports showing how the rate of basal transcription regulation (e.g., elongation, pausing) is important for cell fate (Bai et al., 2010). Our analysis has however revealed that there is differential expression of the basal TFIIFα between NBs and their more committed offspring, potentially indicating that differential composition of RNA PolII complex might be an important mechanism to differentially control transcription in cells with different fates in *Drosophila* neural lineages.

To test whether TfIIFα is indeed important for the fate change from type II NBs to dINPs we expressed RNAi against it specifically in type II NBs and evaluated the impact in lineages *in vivo*. Knock down of TfIIFα in type II NBs leads to defective lineages containing ectopic Dpn+Ase- NB-like cells (**Figure 3F**, pink arrowheads), a phenotype that is never seen in control brains (**Figure 3F**). In addition to this defective differentiation phenotype, knock down of TfIIFα in type II NBs results in disorganized lineages that frequently get mixed with each other at the level of (d)INPs/GMCs, thus preventing exact quantification of each differentiation state. Although TfIIFα was not on the top 20 differentially expressed gene between type I NBs and GMCs, it is still more expressed in type I NBs than in GMCs (**Figure 3G**). To test if TfIIFα has a conserved role in type I lineages, we knocked it down in type I NBs. In control type I lineages there is one type I NB (Dpn+Ase+), up to one imGMC

(Dpn+), and several GMCs (Dpn-Ase+; Figure 3H). However, RNAi of TfIIFα in type I NBs causes the formation of up to 3 cells that are Dpn+Ase+ like type I NBs, but have a smaller size (**Figure 3H**, pink arrowhead), suggesting that they are progeny that retained Dpn expression after NB division. This indicates that knock down of TfIIFα prevents proper lineage progression. In summary, TfIIFα RNAi leads to defects in type I and II NB lineage differentiation and super-numerary NB-like cells, thus showing that TfIIFα plays an important role in fate regulation.

Overall, the comparison between the transcriptional profiles of the distinct lineage differentiation states has proven to be a good strategy for the identification of several known and novel candidate regulatory genes of cell fate and neural lineage progression.

Transcriptomic differences distinguish cells with different degrees of differentiation

The differential gene expression analysis nicely revealed an increasing level of neuronal differentiation from NBs to neurons. Remarkably, the UMAP plot itself (**Figure 1E**) recapitulates the *in vivo* NB lineage progression order for both type I and type II lineages: NB→ imGMCs→ GMCs→ neurons and NB→ dINPs→ INPs→ GMCs, respectively. We have further validated the order of lineage progression in our dataset by predicting the future state of each cell using the RNA velocity method (La Manno et al., 2018) (**Figure S2A**).

As in the CB and VNC neural differentiation does not occur in perfect synchrony, some of the sequences of regulatory changes along differentiation might be overshadowed. To further identify the sequence of changes in gene expression each cell must go through from NB to a mature neuron, we have ordered cells in pseudotime using Monocle 2. This method takes advantage of individual cell asynchronous progression and uses gene expression changes to order cells along a certain trajectory such as lineage differentiation (Trapnell et al., 2014). For this analysis we assessed the clusters that were part of either type I or type II lineages. Both lineages were processed independently and, for both cases, cells were consistently ordered from the less differentiated cells (NBs) to the more differentiated ones (neurons on type I lineages and GMCs on type II lineages, **Figure 4A,B,** **S2B**). The type II lineage trajectory identified two clear GMC fate branches (**Figure 4B**). However, manual analysis of the genes differentially expressed between both

branches revealed that the same genes are expressed with only slightly different expression levels. We hypothesize that these 2 branches might be an artifact caused by the low number of cells from type II lineages available and to the lack of more differentiated cells (neurons) in the trajectory to allow a more complete comparison.

To identify genes that are dynamically expressed along differentiation, we have analysed the top 100 genes that varied the most throughout pseudotime and displayed them in a heatmap. As expected, for both lineage types, known NB fate regulators such as *mira* and *Syp* are more expressed in the cells at the beginning of the trajectory (less differentiated) in comparison to the cells at the end of the trajectory (more differentiated; **Figure 4C,D**). On the other hand, neuron regulatory genes, such as pros, are more expressed in more differentiated cells (**Figure 4C**). Interestingly the analysis of the genes that varied the most in pseudotime in type II lineages identified a small group of genes that seem to be more expressed prior to GMC generation (**Figure 4D**, 1st group) and could potentially be important for the INP to GMC fate transition, or potentially represent a transitioning state equivalent to the previously described type I imGMCs; strikingly most of these genes code for heat shock response related proteins (Hsromega, Hsp26, Hsp83, Hsp68 and DnaJ-1). Although the heat shock response was initially thought to be solely dependent on stress conditions, recent evidence suggests that it is cell specific and responds to the proliferative and metabolic needs of the cell (Li and Tennessen, 2017; Li et al., 2016; Ma et al., 2015), which could explain a specific upregulation of heat shock proteins in this major differentiation state transition.

Interestingly, both the RNA velocity method and Monocle suggest that within the neuronal population it is possible to identify different ages or degrees of differentiation. For instance, only the most differentiated neurons of this dataset (**Figure 4C**; end of trajectory) have an upregulation of genes involved in more specialized roles such as synaptic function and transmembrane transport, as is the case of *CASK*, *cpx* and *nAChRalpha2* (Buhl et al., 2013; Schulz et al., 2000; Sun et al., 2009). Since all neurons in this dataset are younger than 12.5h, these results show that in this short time window it is possible to identify age-related differences between these cells. Knowledge of how neurons develop from birth is still lacking and thus we have focused our next analysis on the early phases of neuron maturation.

Transcriptomic differences distinguish neurons by age

The initial steps of maturation required for proper neuronal development remain poorly understood. Most studies have focused on the analysis of mature neurons, however recent reports show that the transcriptomic profile of neurons converges 15h after their birth (Özel et al., 2020), meaning that fully mature neurons will no longer have a footprint of the transcriptomic alterations that occurred in the first hours of neuronal development. Our scRNA-Seq dataset includes specifically very young neurons, making it possible for us to focus on these initial steps of neuronal maturation.

The age of the neurons in our dataset ranges from 0 to 12.5h old. In spite of the tight age-window of the neurons in our sample, both Monocle and Velocyto analyses indicate that these young neurons can be grouped based on their transcriptional differences throughout pseudotime, with very young neurons closer to GMCs and "older" neurons further away (**Figure 4A,C,** S2A). A simple analysis to identify the expression pattern of known neuronal markers characteristic of very young neurons, as Hey, showed that indeed only the neurons closest to GMCs indeed express this marker (**Figure 5A**). Hey is a target of Notch, previously shown to be expressed transiently only in Notch^{ON} early born neurons (Monastirioti et al., 2010), and consistently only approximately half of these very young neurons express this gene (**Figure 5A**). In neuron clusters farther away from GMCs in the UMAP, several genes related to more mature neurons are expressed, consistent with these clusters representing older neurons. This includes several genes coding for neuron effector proteins as ion transporters and neurotransmitter pathway members (e.g. *nrv3* and *nSyb*, respectively; **Figure 5B**). The neuronal clusters furthest from GMCs, as clusters 5 and 35, predicted to be the older neurons in this population, indeed express the highest number of ion channels (**Figure 5C,** S3A) and adhesion molecules (**Figure 5C'**), important regulators of axonal development and neuronal circuits (Ranscht, 2000; Tan et al., 2015). Furthermore, genes involved in neurotransmitter activity and biosynthesis, such as the fast-acting neurotransmitters VGlut, VAChT, Gad1 and Vmat are only expressed in clusters farther away from GMCs (**Figure 5D**). From our total neuron population 10,47% already express VGlut, 8.72% express VAChT, 3.15% express Gad1, and 0.87% express Vmat (Figure S3B). Interestingly these young neurons express only one NT identity, as only a very

low percentage of the neurons in our data set presented counts for simultaneous expression of NT marker genes. The assessment of multiple combinations of the 4 fast acting neurotransmitters revealed that the highest percentage amounted to 0.65% for the simultaneous presence of VGlut and VAChT (Figure S3B).

Overall, the analysis of this neuron population seems to separate them into two sub-groups which for simplicity we have classified into: “Phase 1” of maturation, which includes the very young immature neurons not transcribing neuronal feature genes as NT and ion channels; “Phase 2” of maturation, which includes the older neurons transcribing neuronal feature genes.

To verify if there is indeed a (clear) transcriptional difference between genes expressed in younger/less mature neurons and older/more mature ones, we have subset the neuronal population and analyzed it independently with Monocle. We identified 2 major tendencies of transcriptional profiles that vary throughout pseudotime, which seem to recapitulate the 2 previously identified maturation phases (**Figure 5D**). Group 1 represents a transcriptional profile composed by genes that are more highly expressed in cells at the beginning of the trajectory, meaning in younger, immature, phase 1 neurons. A GO analysis showed an enrichment of terms associated with cell fate determination, regulation of neurogenesis and neuron fate commitment (Table S1). Group 2 refers to genes that are more expressed in cells at the end of the trajectory, meaning the oldest and more mature, phase 2 neurons. Group 2 is enriched for terms related with synaptic transmission and neurotransmitter regulation (Table S2).

Overall, the transcriptional analysis of young secondary neurons revealed that unexpectedly quickly after birth neurons initiate their maturation process. Indeed ~12.5h after neuron birth is enough for these cells to initiate transcription of NT identity and several other functional neuronal features associated with more mature functional neurons. Additionally, this analysis has permitted to sub-divide the process of neuron maturation in a first phase, corresponding to very young neurons prior to expression of mature neuronal features, and a second phase, where neurons initiate transcription of mature features of neuronal genes (**Figure 5E**).

Neurons in phase 2 of maturation transcribe the neurotransmitter gene ChAT, but do not translate it into protein

The expression of NT genes, ion channels and other neuronal functional genes in secondary neurons in larval stages was surprising as these neurons will only initiate synaptogenesis days later in pupal stages (Muthukumar et al., 2014). So, next we wanted to determine if the presence of mRNA in these phase 2 neurons is already accompanied by the expression of the respective protein. We used the fast-acting neurotransmitter choline acetyltransferase (ChAT) as a case-study, as its mRNA is rarely present in phase 1 neurons and is expressed in phase 2 neurons (**Figure 6A**), in a sub-set that likely corresponds to acetylcholinergic neurons as ChAT is a well-established marker of cholinergic neurons (Hamid et al., 2019; Salvaterra and McCaman, 1985). In addition, there is a commercially available antibody for this protein, allowing for the study of its localization *in vivo*.

We labeled cells using a similar strategy as to generate this atlas, raising the animals at 18ºC and shifting them to 25ºC 18h before dissection at the equivalent to 105h ALH (**Figure 6B**). However, at this point we used a permanent labeling strategy that also labeled neurons with GFP in a manner that was equivalent to the ones present in our dataset. Interestingly, none of the GFP+ neurons are also stained with the ChAT antibody, indicating that the protein is not expressed (**Figure 6B'**). As a positive control for the antibody, ChAT positive cells can be seen in the middle of the central brain, none of which labeled with GFP, likely corresponding to old primary neurons (**Figure 6B', arrowhead**). These results indicate that even though the oldest neurons in our dataset (phase 2) already have ChAT mRNA molecules, these are not yet being translated into protein.

Initiation of ChAT translation is coordinated with animal developmental stage rather than neuron age

The presence of ChAT mRNA but the absence of protein suggests that there is a delay in the initiation of its translation in maturing neurons. To understand what determines translation initiation of ChAT in maturing neurons we tested the following two hypotheses: 1) translation of ChAT is dependent of neuron age and requires neurons to be older/of a certain age; 2) translation of ChAT is coordinated with animal developmental stage starting only when synaptogenesis begins in pupal stages. In order to test these hypotheses, we generated permanent labeling clones which allowed us to fluorescently label neuronal lineages, and thus neurons from birth, and follow them until a certain age and/or a specific developmental stage. In a

18h clone (starting at 87h ALH) with analysis at 105h ALH (wandering 3rd instar larva) there is no expression of ChAT protein in GFP labeled neurons (**Figure 6B**). So, next, we allowed these neurons to age for a longer time by developing a 72h-clone, allowing the older neurons to be up to 66.5h old, and analyzing these clones at 159h ALH (~48h APF; **Figure 6C**). In these clones, co-expression of GFP and ChAT proteins are detected in some neurons (**Figure 6C'**). However, at this point, these results still fit both hypotheses, as neurons may be expressing ChAT because they are older than in the previous clones, or simply because the animal itself is older and closer to the onset of synaptogenesis {Muthukumar et al., 2014, #17453}. In order to uncouple neuronal age from the animal age, we generated a clone for the same duration of 72h but at an earlier time during animal development. We have thus induced clone formation at 50h ALH (**Figure 6D**). This new timeline still allows neurons to age up to 66.5h-old, but the animal itself, though still a pupa, would be only ~12h APF (122h ALH). Interestingly, in these clones there is no co-expression of GFP and ChAT (**Figure 6D'**). These results allowed us to conclude that ChAT protein translation does not start depending on the neuronal age itself, being rather dependent on the age of the animal. As we have designed this experiment to allow the neurons in these clones to undergo the larva to pupa transition, which is driven by a pulse of the steroid hormone ecdysone, these results further allow us to conclude that neither the larva to pupa transition, nor the ecdysone pupariation peak, are sufficient to initiate ChAT translation. To further narrow down when ChAT translation initiates, we have generated a 48h clone, analysed at ~24h APF, which showed that at this earlier pupal development time point there is very little co-expression of GFP and ChAT (**Figure S4**).

Based on these data, we propose that within neuronal maturation, there is a "phase 1" where newly-born neurons do not transcribe NTs and other terminal effector molecules, followed by a "phase 2" when neurons start transcribing NT, ion channels and other terminal feature genes. In "phase 2", and using ChAT as an example, we show that its mRNA is however kept untranslated. Afterwards, approximately at 48h APF, a "phase 3" of maturation starts, when the neuron begins translation of ChAT in a coordinated manner with animal and brain development (**Figure 7A**). At this point, we can only suggest this model for neuroactive molecules, more specifically for ChAT, but we propose this to be a conserved mechanism for NT genes and possibly other molecules associated with more differentiated neuronal characteristics, such

as ion channels. Interestingly several RNA binding proteins and translation regulators such as Syncrip (Syp), musashi (msi), pumilio (pum), brain tumor (Brat), or polyA-binding protein interacting protein 2 (Paip2), are expressed in phase 1 and phase 2 neurons (**Figure 7B**), which might provide a molecular mechanism for translation inhibition of NT genes during larval and young pupal stages. Although the pupariation pulse of Ecdysone does not seem to be the trigger of the molecular mechanisms that will ultimately drive NT translation, other important pulses of Ecdysone occur during pupal development which might provide for the overarching temporal cue for translation initiation and phase 3 of neuron maturation.

These findings reveal that neuron maturation starts very quickly after neuron formation, and sub-divide this process into 3 distinct phases. It raises the possibility that neuron maturation is coordinated with the stage of brain development and synaptogenesis by temporal regulation of translation of NT genes and other neuronal terminal features. Such mechanism would ensure that neurons are only fully mature, when their synaptic partners are as well. Transcription without translation may provide an efficient responsive mechanism to ensure that protein synthesis can quickly initiate for the final phase of neuron maturation and synapse formation.

DISCUSSION

In this study we generated a comprehensive atlas of CB and VNC NB lineages in the developing *Drosophila* wandering larval brain. This allowed us to analyze the transcriptional changes throughout neural lineage progression, from the less differentiated NB to young neurons. Moreover, the fact that our temporally regulated labeling strategy allowed us to collect and analyze the NB lineages generated within a short and defined temporal window (18h maximum) means that we were able to greatly remove the contribution of temporal variation from our dataset and analyse precisely aged cells.

Although previous single cell studies of specific aspects of the *Drosophila* brain have been made (Avalos et al., 2019; Allen et al., 2020; Croset et al., 2018; Davie et al., 2018), our work is the first atlas of the CB and VNC NB lineages of the developing brain specifically including secondary neurogenesis. This atlas represents a valuable resource for studying the regulatory genes and networks involved in NSC proliferation, lineage differentiation, neural cellular diversity and neuron maturation.

Notably, we were able to identify not only all previously described cell states within NB lineages, but also a transient cell state, the imGMCs, proving that our dataset has enough resolution to resolve brief states of differentiation. Cases of newborn GMCs retaining Dpn after NB division have already been reported by (Boone and Doe, 2008) in the CB. Somewhat similar occurrences of a NB diving and originating Dpn+ cells, not necessarily NBs, have also been described in the OL (Mora et al., 2018; Pinto-Teixeira et al., 2018). Although this transition state between type I NBs and GMCs has already been seen *in vivo*, this is the first study where such population is characterized regarding its transcriptome. Moreover, the presence of this transitional state shows that commitment to a GMC fate does not occur immediately after NB asymmetric division; thus, a time lag after the asymmetric inheritance of basal polarity proteins by the GMC is required for the full establishment of GMCs. Although we did not identify imGMCs in type II lineages, we do not exclude that they may also exist and could have been missed due to the overall lower cell numbers for type II lineages. Since in type II lineages, INPs go through a maturation period themselves, it would not be surprising that type II GMCs would also go through such maturating steps, and it would be interesting to address this in future studies.

Interestingly, we did not identify any cluster that presented a clear glial identity (that would be identified for example by the expression of repo). A likely explanation for this is the fact that glial cells are described to be generated from CB and VNC NBs during limited periods of time and those might not be featured within the labeling temporal window used for our dataset (Baek et al., 2013; Bayraktar and Doe, 2013; Enriquez et al., 2018; Truman et al., 2004).

By comparing the transcriptional profile of different cell states throughout differentiation we were able to identify possible candidates to regulate lineage differentiation. One example is TFIIFα, a member of the RNA PolII machinery, that was identified in our dataset as being differentially expressed in type I NBs *vs.* GMCs, which we have shown to be important to maintain proper NB larval lineage development *in vivo*. TFIIFα has important roles during transcription initiation, elongation and even pausing (Gong et al., 1993; Landick, 2006; Price et al., 1989; Rossignol et al., 1999). Based on its role in different steps of transcription, TFIIFα, either alone or associated with other TFs, might be influencing transcription start site selection (Freire-Picos et al., 2005; Khaperskyy et al., 2008) or preventing RNA PolII

pausing (Price et al., 1989; Rossignol et al., 1999). Alterations in the usual rates by which these processes happen, could affect the transcription of genes responsible for NB differentiation, thus justifying the important role of TFIIFα in regulating lineage progression. In the future it might be interesting to study the mechanisms that directly or indirectly link TFIIFα to NB lineage regulation. We have additionally identified several other potential regulators of state progression within NB lineages and it would be interesting to validated them, as well as to understand their mechanisms of action. Ultimately, such studies would contribute to better understand not only NB proliferation, but also progression towards differentiation.

One of the most interesting findings in this study was determining that young neurons (<12.5h-old) have transcriptional profiles that reflect their different ages and consequently different degrees of maturation. Our analysis has led us to propose their sub-division into 3 phases of neuron maturation: A “phase 1”, composed by immature neurons that have yet to start expressing mRNA of more mature neuronal features; Followed by a “phase 2” of maturation, composed by the oldest neurons in our dataset, which start expressing mRNA of maturation markers such as genes involved in axonal development, neuronal circuit organization, neurotransmitter activity and ion channels; And finally, a “phase 3” of neuronal maturation where not only mRNA, but also the protein of more mature neuronal features is expressed. Using the NT gene ChAT as a case study, we found that although this gene is expressed in “phase 2” neurons in our transcriptome dataset, its protein is not. We have shown that regardless of neuronal age, it is not until approximately 48h APF that ChAT protein starts being expressed (“phase 3”). This indicates that ChAT translation depends on the age of the animal, rather than on the age of the neuron itself. Ultimately, such timeline seems to match the beginning of synaptogenesis, which is described to start around 60h APF in the CB (Muthukumar et al., 2014). The early expression of mRNA of neuron maturation markers was unexpected since the neurons formed in larval stages, meant for the adult brain, only terminally differentiate in the pupal period (Dumstrei et al.,l 2003; Pereanu and Hartenstein, 2006; Truman, 1992). So, if ChAT protein is only expressed days after, why does its mRNA starts being expressed so much sooner? And how is its translation being inhibited, and how is it then initiated? One hypothesis is that translation inhibitors might be acting during phases 1 and 2 of neuronal maturation, specifically keeping terminal neuronal genes untranslated and consequently keeping the neuron in a

state that is not yet fully mature, until the time comes when their synaptic partners are formed and ready to connect. This hypothesis is supported by the fact that we can find several RNA biding molecules and translation inhibitors such as Syp, pum and brat expressed in the neurons in our dataset. Moreover, according to this hypothesis, in order to allow the translation of ChAT and other NT-associated molecules to be possible in “phase 3”, these translation inhibitors would then need to be downregulated or inactivated at the appropriate pupal stage, a model which would need to be confirmed *in vivo*. As upstream coordinators of this switch in post-transcriptional regulation and neuron maturation, we can envision hormones, which as systemic signals are ideally placed to coordinate animal and organ development. Although we have shown that the pupariation pulse of the hormone ecdysone is not sufficient to initiate ChAT translation, we do not exclude that other pupal pulses of ecdysone may provide such a function. Overall, a mechanism of temporal regulation of translation may provide an efficient responsive mechanism to ensure that protein synthesis can quickly initiate for the final phase of neuron maturation and synapse formation. Ultimately, the identification of these 3 different phases of neuronal maturation represents an important foundation for further studies to understand the mechanism and timelines that regulate neuronal maturation.

REFERENCES

Allan, D.W., and Thor, S. (2015). Transcriptional selectors, masters, and combinatorial codes: Regulatory principles of neural subtype specification. Wiley Interdiscip. Rev. Dev. Biol. *4*, 505–528.

Allen, A.M., Neville, M.C., Birtles, S., Croset, V., Treiber, C.D., Waddell, S., Goodwin, S.F., and Mann, R.S. (2020). A single-cell transcriptomic atlas of the adult Drosophila ventral nerve cord. Elife *9*, 1–32.

Almeida, M.S., and Bray, S.J. (2005). Regulation of post-embryonic neuroblasts by Drosophila Grainyhead. Mech. Dev. *122*, 1282–1293.

Álvarez, J.A., and Díaz-Benjumea, F.J. (2018). Origin and specification of type II neuroblasts in the Drosophila embryo. Dev. *145*, 1–10.

Avalos, C.B., Brugmann, R., and Sprecher, S.G. (2019). Single cell transcriptome atlas of the drosophila larval brain. Elife *8*, 1–25.

Baek, M., Enriquez, J., and Mann, R.S. (2013). Dual role for Hox genes and Hox co-factors in conferring leg motoneuron survival and identity in Drosophila. Dev. *140*,

2027–2038.

Bai, X., Kim, J., Yang, Z., Jurynec, M.J., Akie, T.E., Lee, J., LeBlanc, J., Sessa, A., Jiang, H., DiBiase, A., et al. (2010). TIF1γ Controls Erythroid Cell Fate by Regulating Transcription Elongation. Cell *142*, 133–143.

Bayraktar, O.A., and Doe, C.Q. (2013). Combinatorial temporal patterning in progenitors expands neural diversity. Nature *498*, 449–455.

Berger, C., Harzer, H., Burkard, T.R., Steinmann, J., van der Horst, S., Laurenson, A.S., Novatchkova, M., Reichert, H., and Knoblich, J.A. (2012). FACS Purification and Transcriptome Analysis of Drosophila Neural Stem Cells Reveals a Role for Klumpfuss in Self-Renewal. Cell Rep. *2*, 407–418.

Boone, J.Q., and Doe, C.Q. (2008). Identification of Drosophila type II neuroblast lineages containing transit amplifying ganglion mother cells. Dev. Neurobiol. *68*, 1185–1195.

Bradke, F., Di Giovanni, S., and Fawcett, J. (2020). Neuronal Maturation: Challenges and Opportunities in a Nascent Field. Trends Neurosci. *43*, 360–362.

Briscoe, J., and Novitch, B.G. (2008). Regulatory pathways linking progenitor patterning, cell fates and neurogenesis in the ventral neural tube. Philos. Trans. R. Soc. B Biol. Sci. *363*, 57–70.

Buhl, L.K., Jorquera, R.A., Akbergenova, Y., Huntwork-Rodriguez, S., Volfson, D., and Littleton, J.T. (2013). Differential regulation of evoked and spontaneous neurotransmitter release by C-terminal modifications of complexin. Mol. Cell. Neurosci. *52*, 161–172.

Cau, E., and Blader, P. (2009). Notch activity in the nervous system: To switch or not switch? Neural Dev. *4*, 36.

Cenci, C., and Gould, A.P. (2005). Drosophila Grainyhead specifies late programmes of neural proliferation by regulating the mitotic activity and Hox-dependent apoptosis of neuroblasts. Development *132*, 3835–3845.

Chen, Y., Akin, O., Nern, A., Tsui, C.Y.K., Pecot, M.Y., and Zipursky, S.L. (2014). Cell-type-specific labeling of synapses in vivo through synaptic tagging with recombination. Neuron *81*, 280–293.

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

Croset, V., Treiber, C.D., and Waddell, S. (2018). Cellular diversity in the Drosophila

midbrain revealed by single-cell transcriptomics. Elife *7*, 1–31.

Davie, K., Janssens, J., Koldere, D., De Waegeneer, M., Pech, U., Kreft, Ł., Aibar, S., Makhzami, S., Christiaens, V., Bravo González-Blas, C., et al. (2018). A Single-Cell Transcriptome Atlas of the Aging Drosophila Brain. Cell *174*, 982-998.e20.

Doe, C.Q. (2017). Temporal patterning in the drosophila CNS. Annu. Rev. Cell Dev. Biol. *33*, 219–240.

Dumstrei, K., Wang, F., Nassif, C., and Hartenstein, V. (2003). Early development of the Drosophila brain: V. Pattern of postembryonic neuronal lineages expressing DE-cadherin. J. Comp. Neurol. *455*, 451–462.

Enriquez, J., Rio, L.Q., Blazeski, R., Bellemin, S., Godement, P., Mason, C., and Mann, R.S. (2018). Differing Strategies Despite Shared Lineages of Motor Neurons and Glia to Achieve Robust Development of an Adult Neuropil in Drosophila. Neuron *97*, 538-554.e5.

Eroglu, E., Burkard, T.R., Jiang, Y., Saini, N., Homem, C.C.F., Reichert, H., and Knoblich, J.A. (2014). SWI/SNF complex prevents lineage reversion and induces temporal patterning in neural stem cells. Cell *156*, 1259–1273.

Freire-Picos, M.A., Krishnamurthy, S., Sun, Z.W., and Hampsey, M. (2005). Evidence that the Tfg1/Tfg2 dimer interface of TFIIF lies near the active center of the RNA polymerase II initiation complex. Nucleic Acids Res. *33*, 5045–5052.

Gong, D. wei, Horikoshi, M., and Nakatani, Y. (1993). Analysis of cDNA encoding drosophila transcription initiation factor TFIIFα (RAP74). Nucleic Acids Res. *21*, 1492.

Hakes, A.E., and Brand, A.H. (2020). Tailless/TLX reverts intermediate neural progenitors to stem cells driving tumourigenesis via repression of asense/ASCL1. Elife *9*, 1–21.

Hamid, R., Hajirnis, N., Kushwaha, S., Saleem, S., Kumar, V., and Mishra, R.K. (2019). Drosophila Choline transporter non-canonically regulates pupal eclosion and NMJ integrity through a neuronal subset of mushroom body. Dev. Biol. *446*, 80–93.

Hartenstein, V. (1993). Atlas of drosophila development.

Hobert, O. (2016). Terminal Selectors of Neuronal Identity. In Current Topics in Developmental Biology, (Academic Press Inc.), pp. 455–475.

Homem, C.C.F., Reichardt, I., Berger, C., Lendl, T., and Knoblich, J.A. (2013). Long-term live cell imaging and automated 4D analysis of Drosophila neuroblast lineages. PLoS One *8*, 1–10.

Homem, C.C.F., Repic, M., and Knoblich, J.A. (2015). Proliferation control in neural stem and progenitor cells. Nat. Rev. Neurosci. *16*, 647–659.

Iyer, E.P.R., Iyer, S.C., Sullivan, L., Wang, D., Meduri, R., Graybeal, L.L., and Cox, D.N. (2013). Functional Genomic Analyses of Two Morphologically Distinct Classes of Drosophila Sensory Neurons: Post-Mitotic Roles of Transcription Factors in Dendritic Patterning. PLoS One *8*.

Khaperskyy, D.A., Ammerman, M.L., Majovski, R.C., and Ponticelli, A.S. (2008). Functions of Saccharomyces cerevisiae TFIIF during Transcription Start Site Utilization. Mol. Cell. Biol. *28*, 3757–3766.

Kimura, Y., Satou, C., and Higashijima, S.I. (2008). V2a and V2b neurons are generated by the final divisions of pair-producing progenitors in the zebrafish spinal cord. Development *135*, 3001–3005.

Konstantinides, N., Kapuralin, K., Fadil, C., Barboza, L., Satija, R., and Desplan, C. (2018). Phenotypic Convergence: Distinct Transcription Factors Regulate Common Terminal Features. Cell *174*, 622-635.e13.

Kratsios, P., Pinan-Lucarré, B., Kerk, S.Y., Weinreb, A., Bessereau, J.L., and Hobert, O. (2015). Transcriptional coordination of synaptogenesis and neurotransmitter signaling. Curr. Biol. *25*, 1282–1295.

Kurmangaliyev, Y.Z., Yoo, J., Locascio, S.A., and Lawrence Zipursky, S. (2019). Modular transcriptional programs separately define axon and dendrite connectivity. Elife *8*, 1–18.

Kurmangaliyev, Y.Z., Yoo, J., Valdes-Aleman, J., Sanfilippo, P., and Zipursky, S.L. (2020). Transcriptional Programs of Circuit Assembly in the Drosophila Visual System. Neuron *108*, 1045-1057.e6.

Landgraf, M., and Thor, S. (2006). Development of Drosophila motoneurons: Specification and morphology. Semin. Cell Dev. Biol. *17*, 3–11.

Landick, R. (2006). The regulatory roles and mechanism of transcriptional pausing. Biochem. Soc. Trans. *34*, 1062–1066.

Li, H., and Tennessen, J.M. (2017). Methods for studying the metabolic basis of Drosophila development. Wiley Interdiscip. Rev. Dev. Biol. *6*.

Li, J., Chauve, L., Phelps, G., Brielmann, R.M., and Morimoto, R.I. (2016). E2F coregulates an essential HSF developmental program that is distinct from the heat-shock response. Genes Dev. *30*, 2062–2075.

Ma, X., Xu, L., Alberobello, A.T., Gavrilova, O., Bagattin, A., Skarulis, M., Liu, J.,

Finkel, T., and Mueller, E. (2015). Celastrol protects against obesity and metabolic dysfunction through activation of a HSF1-PGC1α transcriptional axis. Cell Metab. *22*, 695–708.

La Manno, G., Soldatov, R., Zeisel, A., Braun, E., Hochgerner, H., Petukhov, V., Lidschreiber, K., Kastriti, M.E., Lönnerberg, P., Furlan, A., et al. (2018). RNA velocity of single cells. Nature *560*, 494–498.

McInnes, L., Healy, J., and Melville, J. (2018). UMAP: Uniform manifold approximation and projection for dimension reduction. ArXiv.

Michki, N.S., Li, Y., Sanjasaz, K., Zhao, Y., Shen, F.Y., Walker, L.A., Cao, W., Lee, C.-Y., and Cai, D. (2021). The molecular landscape of neural differentiation in the developing Drosophila brain revealed by targeted scRNA-seq and multi-informatic analysis. Cell Rep. *35*, 109039.

Monastirioti, M., Giagtzoglou, N., Koumbanakis, K.A., Zacharioudaki, E., Deligiannaki, M., Wech, I., Almeida, M., Preiss, A., Bray, S., and Delidakis, C. (2010). Drosophila Hey is a target of Notch in asymmetric divisions during embryonic and larval neurogenesis. Development *137*, 191–201.

Mora, N., Oliva, C., Fiers, M., Ejsmont, R., Soldano, A., Zhang, T.T., Yan, J., Claeys, A., De Geest, N., and Hassan, B.A. (2018). A Temporal Transcriptional Switch Governs Stem Cell Division, Neuronal Numbers, and Maintenance of Differentiation. Dev. Cell *45*, 53-66.e5.

Muthukumar, A.K., Stork, T., and Freeman, M.R. (2014). Activity-dependent regulation of astrocyte GAT levels during synaptogenesis. Nat. Neurosci. *17*, 1340–1350.

Neumüller, R.A., Richter, C., Fischer, A., Novatchkova, M., Neumüller, K.G., and Knoblich, J.A. (2011). Genome-wide analysis of self-renewal in Drosophila neural stem cells by transgenic RNAi. Cell Stem Cell *8*, 580–593.

Özel, M.N., Langen, M., Hassan, B.A., and Hiesinger, P.R. (2015). Filopodial dynamics and growth cone stabilization in Drosophila visual circuit development. Elife *4*.

Özel, M.N., Kulkarni, A., Hasan, A., Brummer, J., Moldenhauer, M., Daumann, I.M., Wolfenberg, H., Dercksen, V.J., Kiral, F.R., Weiser, M., et al. (2019). Serial Synapse Formation through Filopodial Competition for Synaptic Seeding Factors. Dev. Cell *50*, 447-461.e8.

Özel, M.N., Simon, F., Jafari, S., Holguera, I., Chen, Y.C., Benhra, N., El-Danaf,

R.N., Kapuralin, K., Malin, J.A., Konstantinides, N., et al. (2020). Neuronal diversity and convergence in a visual system developmental atlas. Nature *589*, 88–95.

Pereanu, W., and Hartenstein, V. (2006). Neural lineages of the Drosophila brain: A three-dimensional digital atlas of the pattern of lineage location and projection at the late larval stage. J. Neurosci. *26*, 5534–5553.

Pinto-Teixeira, F., Koo, C., Rossi, A.M., Neriec, N., Bertet, C., Li, X., Del-Valle-Rodriguez, A., and Desplan, C. (2018). Development of Concurrent Retinotopic Maps in the Fly Motion Detection Circuit. Cell *173*, 485-498.e11.

Price, D.H., Sluder, A.E., and Greenleaf, A.L. (1989). Dynamic interaction between a Drosophila transcription factor and RNA polymerase II. Mol. Cell. Biol. *9*, 1465–1475.

Ranscht, B. (2000). Cadherins: molecular codes for axon guidance and synapse formation. Int. J. Dev. Neurosci. Off. J. Int. Soc. Dev. Neurosci. *18*, 643–651.

Reichardt, I., Bonnay, F., Steinmann, V., Loedige, I., Burkard, T.R., Meister, G., and Knoblich, J.A. (2018). The tumor suppressor Brat controls neuronal stem cell lineages by inhibiting Deadpan and Zelda. EMBO Rep. *19*, 102–117.

Rossignol, M., Keriel, A., Staub, A., and Egly, J.M. (1999). Kinase activity and phosphorylation of the largest subunit of TFIIF transcription factor. J. Biol. Chem. *274*, 22387–22392.

Salvaterra, P.M., and McCaman, R.E. (1985). Choline acetyltransferase and acetylcholine levels in Drosophila melanogaster: A study using two temperature-sensitive mutants. J. Neurosci. *5*, 903–910.

San-Juán, B.P., and Baonza, A. (2011). The bHLH factor deadpan is a direct target of Notch signaling and regulates neuroblast self-renewal in Drosophila. Dev. Biol. *352*, 70–82.

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

Schinaman, J.M., Giesey, R.L., Mizutani, C.M., Lukacsovich, T., and Sousa-Neves, R. (2014). The KRÜPPEL-Like Transcription Factor DATILÓGRAFO Is Required in Specific Cholinergic Neurons for Sexual Receptivity in Drosophila Females. PLoS Biol. *12*.

Schulz, R., Bertrand, S., Chamaon, K., Smalla, K.H., Gundelfinger, E.D., and Bertrand, D. (2000). Neuronal nicotinic acetylcholine receptors from drosophila: Two different types of α subunits coassemble within the same receptor complex. J. Neurochem. *74*, 2537–2546.

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

Shin, J., Poling, J., Park, H.C., and Appel, B. (2007). Notch signaling regulates neural precursor allocation and binary neuronal fate decisions in zebrafish. Development *134*, 1911–1920.

Southall, T.D., Davidson, C.M., Miller, C., Carr, A., and Brand, A.H. (2014). Dedifferentiation of neurons precedes tumor formation in lola mutants. Dev. Cell *28*, 685–696.

Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., Hao, Y., Stoeckius, M., Smibert, P., and Satija, R. (2019). Comprehensive Integration of Single-Cell Data. Cell *177*, 1888-1902.e21.

Sun, M., Liu, L., Zeng, X., Xu, M., Liu, L., Fang, M., and Xie, W. (2009). Genetic interaction between Neurexin and CAKI/CMG is important for synaptic function in Drosophila neuromuscular junction. Neurosci. Res. *64*, 362–371.

Tan, L., Zhang, K.X., Pecot, M.Y., Nagarkar-Jaiswal, S., Lee, P.-T., Takemura, S.-Y., McEwen, J.M., Nern, A., Xu, S., Tadros, W., et al. (2015). Ig Superfamily Ligand and Receptor Pairs Expressed in Synaptic Partners in Drosophila. Cell *163*, 1756–1769.

Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., Lennon, N.J., Livak, K.J., Mikkelsen, T.S., and Rinn, J.L. (2014). The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. Nat. Biotechnol. *32*, 381–386.

Truman, J.W. (1992). The eclosion hormone system of insects. Prog. Brain Res. *92*, 361–374.

Truman, J.W., and Bate, M. (1988). Spatial and temporal patterns of neurogenesis in the central nervous system of Drosophila melanogaster. Dev. Biol. *125*, 145–157.

Truman, J.W., Schuppe, H., Shepherd, D., and Williams, D.W. (2004). Developmental architecture of adult-specific lineages in the ventral CNS of Drosophila. Development *131*, 5167–5184.

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

Wang, Y.C., Yang, J.S., Johnston, R., Ren, Q., Lee, Y.J., Luan, H., Brody, T., Odenwald, W.F., and Lee, T. (2014). Drosophila intermediate neural progenitors produce lineagedependent related series of diverse neurons. Dev. *141*, 253–258.

Weng, M., Golden, K.L., and Lee, C.-Y. (2010). dFezf/Earmuff maintains the restricted developmental potential of intermediate neural progenitors in Drosophila. Dev. Cell *18*, 126—135.

Yang, C.P., Fu, C.C., Sugino, K., Liu, Z., Ren, Q., Liu, L.Y., Yao, X., Lee, L.P., and Lee, T. (2016). Transcriptomes of lineage-specific Drosophila neuroblasts profiled by genetic targeting and robotic sorting. Dev. *143*, 411–421.

Yuva-Aydemir, Y., Xu, X.L., Aydemir, O., Gascon, E., Sayin, S., Zhou, W., Hong, Y., and Gao, F.B. (2015). Downregulation of the Host Gene jigr1 by miR-92 Is Essential for Neuroblast Self-Renewal in Drosophila. PLoS Genet. *11*.

Zacharioudaki, E., Magadi, S.S., and Delidakis, C. (2012). bHLH-O proteins are crucial for Drosophila neuroblast self-renewal and mediate notch-induced overproliferation. Development *139*, 1258–1269.

Zheng, X., Zugates, C.T., Lu, Z., Shi, L., Bai, J.M., and Lee, T. (2006). Baboon/dSmad2 TGF-β signaling is required during late larval stage for development of adult-specific neurons. EMBO J. *25*, 615–627.

Figure legends

Figure 1. Type I and type II NB lineages identified by scRNA-Seq in wandering larvae brain.

(A) Pattern of VT201094-Gal4 driver expression in the whole brain, showing GFP expression in NBs from CB/VNC, but not OL. Views from anterior and posterior sides; scale bar, 50 µm.

(B) Schematic representation of type I and type II lineages; cells colored by expression of dpn (grey), ase (red), pnt (black), klu (purple), erm (orange), pros (yellow) and elav (blue); green outline indicates the cells in which VT201094 expression is driven.

(C) Close up of type I and type II NB lineages in the anterior side of CB; dpn (white), ase (red), GFP (green), elav (blue); scale bar, 10 µm.

(D) UMAP plot of 12.7K cells composed of NB lineages from CB and VNC.

(E) Annotated cell types on the UMAP plot.

(F) Dot plot showing the main genes that identify each cell type.

Figure 2. Characterization of NB-imGMCs-GMCs transition in type I lineages.

(A) Close up of a type I NB lineage with an imGMC. type I NBs are $Dpn^{+}Mira^{+}$, and upon division, transient imGMCs (white arrow) retain Dpn expression, losing its expression once they became GMCs; mira (red), dpn (green); scale bar, 10 µm

(B) Schematic representation of type I lineages including imGMCs; cells colored by expression of dpn (grey/gradient), ase (red), klu (purple), mira (orange), pros (yellow) and elav (blue).

(C) Dotplot showing the top TFs/DNAB differentiating type I NBs *vs.* imGMCs.

(D) Dotplot showing the top TFs/DNAB differentiating imGMCs *vs.* type I GMCs.

Figure 3. Differentially expressed markers throughout the different differentiation states in NB lineages.

(A) Dotplot showing the top TFs/DNAB distinguishing type I NBs *vs.* type I GMCs.

(B) E(spl)mγ-HLH is expressed in type I NBs, but not in imGMCs (blue arrow); mira (red), E(spl)mγ-HLH reporter (green); scale bar, 10 µm.

(C) Dotplot showing the top TFs/DNAB distinguishing type I GMCs *vs.* neurons.

(D) Dotplot showing the top TFs/DNAB distinguishing type II NBs *vs.* dINPs.

(E) Venn diagram comparing TFs/DNAB genes overlapping between Type I NBs vs. Type I GMCs and Type I NBs *vs*. dINPs.

(F) Type II lineages of 3rd instar larva; control (upper panels) and TFIIFα RNAi (lower panels); schematic representations (left) and immunofluorescence images (right); cells colored by expression of Dpn (red), Ase (green); blue arrowheads indicate primary type II NBs; pink arrowheads indicate examples of extra NB-like cells; scale bar, 10 µm.

(G) Dot plot showing the expression of *TfIIFα* throughout the differentiation stages of type I and type II lineages. Dpn expression shown as reference.

(H) Type I lineages of 3rd instar larva; control (upper panels) and TFIIFα RNAi (lower panels); schematic representations (left) and immunofluorescence images (right); cells colored by expression of Dpn (red), Ase (green); blue arrowheads indicate primary type II NBs; pink arrowheads indicate examples of extra NB-like cells; scale bar, 10 µm.

Figure 4. Transcriptomic dynamics of differentiation in type I and type II NB lineages.

(A and B) Cell trajectory for type I NB lineages (A) and type II NB lineages (B); cell ordering and trajectory defined based on pseudotime analysis using monocle; colored by cell type.

(C and D) Heatmap with the top 100 genes most differentially expressed throughout pseudotime for type I NB lineages (C) and type II NB lineages (D); differential expression analysis performed with monocle.

Figure 5. Transcriptomic differences in neurons identify 2 phases of maturation.

(A) Feature plot for the young neuronal markers (*Hey*).

(B) Feature plots for maturation markers (*nrv3*, *nSyb*)

(C,C') Chord diagrams showing the correspondence between Ion channels (C) and Immunoglobulin and Cadherin super families (C') and the clusters in which they are differentially expressed.

(D) Feature plots for genes necessary for neurotransmitter biosynthesis (*VGlut*, *VAChT*, *Gad1*, *Vmat*).

(E) Heatmap with the top 100 genes most differentially expressed throughout pseudotime for neurons; differential expression analysis performed with monocle.

(F) Schematic representation of a UMAP plot showing the division of neurons according to the 2 phases of neuronal maturation proposed by us.

Figure 6. ChAT expression in neurons with different ages.

(A) Feature plot for *ChAT* (gene necessary for neurotransmitter biosynthesis).

(B-D) Clones induced in NB lineages: (B) 18h-clone induced at 87h and analysed at 105h ALH (wandering 3rd instar larvae); (C) 72h-clone induced at 87h and analysed at 159h ALH (approximately 48h after pupa formation-APF); (D) 72h-clone induced at 50h and analysed at 122h ALH (approximately 12h APF); schematic representations (B,C,D) and immunofluorescence images (B',C',D'); cells colored by expression of ChAT (red) and GFP (green); blue arrowhead indicate examples of labeled cells expressing ChAT protein; scale bar, 10 µm.

Figure 7. Neuronal maturation occurs in 3 phases possibly regulated the translation level.

(A) Model for neuronal maturation.

(B) Feature plot for translation inhibitors (*brat*, *pum*, *msi*, *Paip2*, Syp).

METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Guinea pig anti-dpn (dilution 1:1000)	gift from Juergen Knoblich (Homem et al., 2013)	RRID:AB_2568007
rat polyclonal anti-ase (dilution 1:200)	gift from Juergen Knoblich (Eroglu et al., 2014)	RRID:AB_2567568
mouse monoclonal anti-elav (dilution 1:100)	DSHB	RRID:AB_528217
rabbit monoclonal anti-mira (dilution 1:1000)	Eurogentec (non comercial)	N/A
mouse monoclonal anti-ChAT4B1 (dilution 1:500)	DSHB	RRID:AB_528122
goat anti-guinea pig Alexa 647 (dilution 1:1000)	Thermo Fisher Scientific	RRID:AB_2735091
goat anti-rat Alexa 647 (dilution 1:1000) goat anti-mouse Alexa 647 (dilution 1:1000)	Invitrogen Invitrogen	RRID:AB_141778 RRID:AB_2535804
goat anti-rat Alexa 568 (dilution 1:1000)	Thermo Fisher Scientific	RRID:AB_2534121
goat anti-rabbit Alexa 568 (dilution 1:1000)	Invitogen	RRID:AB_143157
goat anti-mouse Alexa 405 (dilution 1:1000)	Invitrogen	RRID:AB_221604
Chemicals, Peptides, and Recombinant Proteins		
aqua/polymount	Polysciences, Inc	N/A
Critical Commercial Assays		
DynaBeads® MyOne™ Silane Beads	Thermo Fisher Scientific	Cat. No. 37002D
Chromium™ Single Cell 3' Library and Gel Bead Kit v2, 4 rxns	10x Genomics	Cat. No. 120267
Chromium™ Single Cell A Chip Kit, 16 rxns	10x Genomics	Cat. No. 1000009
Deposited Data		
Single-cell RNA sequencing of *Drosophila melanogaster* Central Brain and Ventral Nerve Cord (Raw and analyzed data)	this study	GEO:
Experimental Models: Organisms/Strains		
Fly: *D. melanogaster:* w^{1118}	gift from António Jacinto	N/A
Fly: *D. melanogaster:* VT012709-Gal4	gift from Juergen Knoblich	N/A
Fly: *D. melanogaster:* UAS-CD8::GFP; *tub*-Gal80[ts]	this study	N/A

Fly: *D. melanogaster:* UAS-*dicer2*; *wor*-Gal4*ase*-Gal80; UAS-*CD8::GFP*	gift from Juergen Knoblich (Homem et al., 2014)	N/A
Fly: *D. melanogaster:* UAS-*dicer2*; *ase*-Gal4 UAS-*CD8::GFP*	gift from Juergen Knoblich (Harzer et al., 2014)	N/A
Fly: *D. melanogaster:* TFIIFα RNAi	Vienna Drosophila Resource Center	RRID:FlyBase_FBst0481806
Fly: *D. melanogaster:* RNAi-TK (empty)	Vienna Drosophila Resource Center	VDRC ID#60100
Fly: *D. melanogaster:* UAS-*3xFLAG-Cas9-VPR; tub*-Gal4	Bloomington Drosophila Stock Center	RRID:BDSC_67048
Fly: *D. melanogaster: pnt*-Gal4	gift from Juergen Knoblich	N/A
Fly: *D. melanogaster:* E(spl)mγ-HLH-GFP	Bloomington Drosophila Stock Center	RRID:BDSC_66401
Fly: *D. melanogaster:* UAS-*FLP*, *Ubi-p63E*-STOP-*Stinger*	Bloomington Drosophila Stock Center	RRID:BDSC_28282
Software and Algorithms		
R Statistical Computing Software version 4.0.0	N/A	https://www.r-project.org/
Fiji	(Schindelin et al., 2012)	https://fiji.sc/; RRID:SCR_002285
Adobe Photoshop	Adobe	free license
Adobe Illustrator	Adobe	free license
Cell Ranger	10x Genomics	https://support.10xgenomics.com/single-cell-dna/software/release-notes/1-1#header; RRID:SCR_017344
Seurat version 3.1.5	(Satija et al., 2015)	https://satijalab.org/seurat/; RRID:SCR_007322
Velocyto v.0.17.17	(La Manno et al., 2018)	http://velocyto.org/; RRID:SCR_018167
Monocle version 2.16.0	(Trapnell et al., 2014)	http://cole-trapnell-lab.github.io/monocle-release/
PANTHER molbiotools	(Chen et al., 2013; Kuleshov et al., 2016) Vladimir Cermák	http://pantherdb.org/; RRID:SCR_004869 http://molbiotools.com/
Circlize version 0.4.9	(Gu et al., 2014)	http://cran.r-project.org/web/packages/circlize/; RRID:SCR_002141

RESOURCE AVAILABILITY

Lead Contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Catarina Homem (catarina.homem@nms.unl.pt).

Materials Availability

This study did not generate new unique reagents.

Data and Code Availability

The datasets generated during this study are available in GEO.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Fly strains

For the scRNA-Seq experiments, temporal labeling of NBs and their lineage was achieved by crossing VT012709-Gal4 (enhancer of Koi and CG15236; CB/VNC NB driver) with UAS-CD8::GFP; tub-Gal80ts males. Fly crosses were set up at 25ºC, allowing ~12h-egg lays. First instar larvae (L1) were synchronized upon hatching and transferred to 18ºC to inactivate CD8 expression. CD8 expression was activated 18h previously to the dissection timepoint at 105h.

For the permanent labeling experiments, males from ;;UAS-*FLP*, *Ubi-p63E*-STOP-*Stinger* driver were crossed with ;tub-Gal80ts; VT201094 females and raised as previously described. CD8 expression was activated at 50h or 87h ALH and clones were allowed to form for 18h, 48h or 72h at 25ºC.

w^{1118} was used as control for RNAi experiments. UAS-*dicer2* ; *ase*-Gal4 UAS-*CD8::GFP* was used as a type I NBs driver (Harzer et al., 2014) and UAS-*dicer2* ; *wor*-Gal4 *ase*-Gal80 ; UAS-*CD8::GFP* was used as a type II NB driver (Homem et al., 2014). The RNAi lines were crossed to either the control or type I and type II drivers; fly crosses for RNAi experiments were set up at 29ºC to increase UAS-Gal4 expression and fluorescence intensity.

For other experiments the lines RNAi-TK (empty) and ;UAS-*3xFLAG-Cas9-VPR; pnt*-Gal4 were crossed and used as control.

Please check the Key Resources Table for more complete information about the *Drosophila* lines used in this study.

METHOD DETAILS

Brain Dissociation and Cell Sorting

One hundred and twenty-four third instar larvae (105h ALH) were collected and dissected in supplemented Schneider's medium (10% fetal bovine serum (Sigma), 20 mM Glutamine (Sigma), 0.04 mg/mL L-Glutathione (Sigma), 0.02 mg/mL Insulin (Sigma) Schneider's medium (Sigma)). After dissection, brains were transferred to Chan & Gehring solution (Chan et al., 1971) 2 % FBS, and washed once. After this, they were enzymatically dissociated in Chan & Gehring solution 2 % FBS with 1 mg/mL Papain (Sigma) and 1 mg/mL Collagenase I (Sigma) for 1hr at 30ºC. Afterwards, brains were washed once with Chan & Gehring 2 % FBS solution and once more with supplemented Schneider's medium. After these washing steps, brains were resuspended in PBS (phosphate buffered saline) 0.1% BSA (Sigma) and mechanically disrupted using a pipette tip. The cell suspension was filtered through a 30 µL mesh into a 5 mL FACS tube (BD Falcon) and immediately sorted by fluorescence activated cell sorting (FACS) (FACS Aria II, BD). GFP positive NBs and their lineage were collected in a drop of PBS 0.1% BSA. Since NBs represent a lower percentage of sorted cells when compared to neurons, they were sorted separately in order to assure an enrichment of less differentiated cells in the final pool. Cells were resuspended in 0.1% BSA at a final concentration of approximately 400 cell/µL and immediately processed according to the 10x Genomics protocol.

10x Genomics experimental procedure

Approximately 25k of the sorted cells (NB lineages) were used to construct single-cell libraries; libraries were obtained using Chromium Single Cell 3' reagent Kits v2 (10x Genomics) standard protocol. Cells were equally divided into 2 samples (duplicates) and loaded in 2 wells of a Single Cell A Chip, aiming for an estimated target cell recovery of ~7k cells. Cells were then partitioned into nanoliter-scale Gel Bead-In-EMulsions (GEMs) and reverse-transcribed using an Eppendorf Mastercycler pro Thermal Cycler (Eppendorf), set for 53 ºC during 45 min, 85 ºC for 5 minutes and hold at 4 ºC. Post reverse transcription incubation GEMs were then broken and the cDNA was recovered and cleaned using Silane DynaBeads (Thermo Fisher Scientific). The next step consisted in amplifying the cDNA, by incubating the

samples in a Thermal Cycler programmed for 98 ºC during 3 minutes, 10 cycles of 98 ºC for 15 sec, 67 ºC for 20 sec and 72 ºC for 1 min, followed by 72 ºC for 1 min and hold at 4 ºC. The amplified cDNA was then cleaned using SPRIselect and quantified using a TapeStation (Agilent Technologies). The amplified cDNA was fragmented, end-repared and A-tailed by incubating in a Thermal Cycler at 32 ºC for 5 min, 65 ºC for 30 min and hold at 4 ºC; next, the cDNA went through a double-sided size selection using SPRIselect. Subsequently, the samples went through adaptor ligation, by incubating in a Thermal Cycler at 20 ºC for 15 minutes, after which there was a new SPRIselect cleanup step. Afterwards, samples were attributed independent indexes and amplified by PCR using a Thermal Cycler set for 98 ºC for 45 sec, 14 cycles at 98 º for 20 sec, 54 ºC for 30 sec and 72 ºC for 20 sec, followed by 72 ºC for 1 min and hold at 4 ºC. The generated library went through a new double-sided size selection by SPRIselect and run on a TapeStation for quality control and quantification.

Both samples were subjected to paired-end sequencing using the NovaSeq 6000 system (Genome Technology Center at NYU Langone Health).

scRNA-Seq Raw Datasets

Each sequenced sample was processed with Cell Ranger Version 3.0.1 for alignment, barcode assignment and UMI counting. Samples were mapped to BDGP6 reference genome from Ensembl.

scRNA-Seq Dataset Pre-processing

The filtered gene matrices obtained after Cell Ranger processing were analyzed with R package Seurat 3.1.5. Only cells that had at least 200 unique feature counts were included in the analysis. Moreover, we only kept cells with a percentage of mitochondrial genes inferior to 20%. Higher percentages of mitochondrial genes are usually indicative of cell damage/rupture, and consequently, of altered overall transcriptional content (Ilicic et al., 2016). These initial quality control steps resulted in a dataset with 12671 cells and 10250 genes.

Samples were normalized using the NormalizeData function, and the top 2000 most variable features were then identified using the FindVariableFeatures function. Next, ScaleData was used to scale all genes; within this step, the percentage of

mitochondrial genes was also regressed out, in order to avoid artefacts in subsequent analysis.

scRNA-Seq Dataset Clustering

We performed a Principal Component Analysis (PCA), using the previously calculated top 2000 most variable genes. Next, we used an elbow plot and a jackstraw approach to identify significant PCs. For the complete dataset analysis, we used the first 51 PCs, as not to include PC that was not significant ($p > 0.05$). Within the FindClusters function, the resolution parameter was set to 1.55 as it resulted in a granularity that allowed the identification of smaller cell populations such as type II NBs and INPs. The combination of these parameters originated 39 clusters.

Moreover, clusters were annotated into major groups corresponding to the different cell types identified in our dataset, resulting in 10 major clusters. This annotation was performed based on well described markers (dpn, ase, pros, pnt, etc), as well as based on relative cell localization within the UMAP plot.

RNA Velocity Dynamics

RNA velocity analysis was performed with the python version of Velocyto v.0.17.17 package (La Manno et al., 2018). We used the subcommand 'velocity run' to create a loom file for the cells that survived the filtering steps of Seurat pipeline using the *Drosophila melanogaster* genome annotation file (Drosophila_melanogaster.BDGP6.88.gtf) and the bam file with sorted reads that was estimated using the default parameters of the Cellranger software (10x Genomics). We masked repetitive regions using the genome expressed repetitive annotation file downloaded from UCSC genome browser. The loom file created separates molecule counts into 'spliced', 'unspliced' or 'ambiguous'. To estimate RNA velocity parameters, we adapted the pipeline used in the analysis of the mouse hippocampus dataset from La Manno et al., 2018. We started by removing cells with extremely low unspliced detection requiring the sum of unspliced molecules per cell to be above the 0.4 threshold. We also selected genes that are expressed above a threshold of total number of molecules in any of the clusters requiring 40 minimum expressed (spliced) counts in at least 30 cells, after which we kept the top 3000 highly expressed and variant genes on the basis of a coefficient of variation CV vs mean fit that uses a nonparametric fit (Support Vector Regression). We applied a

final filter to the dataset by selecting genes on the basis of their detection levels and cluster-wise expression threshold. This filter kept genes with unspliced molecule counts above a detection threshold of 25 minimum expressed counts detected over 20 cells, and with average counts of unspliced and spliced expression bigger than 0.01 and 0.08 respectively in at least one of the clusters. Finally, both spliced and unspliced counts were normalize for the cell size by dividing by the total number of molecules in each cell, and multiplying the mean number of molecules across all cells to. All filtering steps resulted in a dataset of 12604 cells and 1086 genes to be used in the RNA velocyto analysis. For the preparation of the gamma fit we smooth the data using a kNN neighbors pooling approach (velocyto subcommand knn_imputation) and k=500 with calculations performed in the reduced PCA space defined by the top 99 principal components. Velocity calculation and extrapolation to future states of the cells was performed under the assumption of constant velocity. Analysis pipeline can be obtained from the corresponding author.

Single-cell Trajectories

Pseudotime analysis was performed using R package Monocle v2.16.0. The Seurat object, containing all filtering and clustering information was imported to Monocle and subset accordingly. For the analysis of type I lineage the subset included clusters annotated as “Type I NBs”, “imGMCs”, “Type I GMCs” and all clusters identified as “Neurons”. For the analysis of type II lineage, the subset included “Type II NBs”, “dINPs”, “INPs” and “Type II GMCs”; no neuronal clusters were included as, the duration of our labeling combined with the longer cell division timings within type II lineage (Homem et al., 2013), likely did not allow any neurons resulting from type II lineages to be labeled. For the analysis of the neuronal population, the subset included all clusters annotated as “Neurons”, and within those, only cells with 0 counts for *wor* and *repo* were processed.

The 3 subsets (type I lineage, type II lineage and neurons) were then processed independently, but using the same pipeline. Differences in mRNA across cells were normalized and “dispersion” values were calculated using the functions estimateSizeFactors and estimateDispersions, respectively.

To construct single cell trajectories, we started by using the differentialGeneTest function to extract the genes distinguishing different clusters; these genes were then marked to be used for clustering in subsequent calls by using setOrderFilter.

Afterwards, dimensions were reduced by using the Discriminative Dimension Reduction Tree (DDRTree) method, and finally, ordered using the orderCells function.

In order to perform differential expression analysis, we used differentialGeneTest again, however, this time to obtain the differentially expressed genes as a function of pseudotime. These genes were then ordered by qvalue, and the top 100 hits were presented for each type of lineage.

Gene annotation lists

TFs/DNAB genes were selected from the Gene List Annotation for *Drosophila* (GLAD) (Hu et al., 2015), obtained from //www.flyrnai.org/tools/glad/web/. The list for ion channels was also obtained from GLAD.

For the cadherin super family analysis, we used a gene list from FlyBase, obtained under the group “CADHERINS” – FBgg0000105.

For the immunoglobulin super family analysis, genes were selected according to FlyBase FBrf0167517 (Vogel et al., 2003).

In all three cases, gene symbols were updated according to our dataset. Moreover, we used these gene lists to show the correspondence between cluster and cluster marker genes. To do this, we created chord diagrams using the R package “circlize”, adapted according to Allen et al., 2020.

Immunofluorescence

Larval brains were dissected and fixed in 4% paraformaldehyde for 30 min at room temperature; afterwards were washed 3 times with PBS with 0.1% Triton X-100 (PBT). Fixed brains were incubated for 20 min in PBS with 0.5% Triton X-100 and 1% Normal Goat Serum (blocking solution) and incubated with the primary antibodies over-night at 4ºC. Next day, brains were washed 3 times, blocked 20 min and incubated with the secondary antibodies for 2h at room-temperature. Afterwards, brains were washed 3 times, incubated 10 min in PBS, mounted in aqua/polymount (Polysciences, Inc) and imaged.

Immunofluorescent images were acquired on a LSM880 (Carl Zeiss GmbH). Adobe Photoshop (Adobe) and/or Fiji were used to adjust brightness and contrast; figure panels were prepared in Illustrator (Adobe).

QUANTIFICATION AND STATISTICAL ANALYSIS

TapeStation

Quantification and quality of cDNA and respective libraries generated for 10x Genomics Data was assessed with TapeStation following the standard protocol available at https://www.agilent.com/cs/library/usermanuals/public/ScreenTape_HSD5000_QG.pdf (Agilent Technologies).

Differential Expression between Clusters

Differential expression analysis was performed within TFs/DNAB only. We used Seurat to identify the specific markers for each cluster. For that, we used the Receiver Operating Curve (ROC) test to find the differentially expressed genes between clusters. Within that analysis, we selected an AUC (Area Under the ROC Curve) > 0.5, to assure that the only hits were from genes with predictive values to classify that cluster. Moreover, to assure that none of the hits is a scarcely expressed gene, we only considered genes expressed in at least 25% of the cells of either one of the groups compared. Furthermore, we established that the average log fold change between the two populations being compared should be higher than 0.4 (**Figures 2C,D** and **3,A,C,D**), except for the case presented next. For neuronal cluster comparison (**Figure S3A**), the average log fold change was altered to 0.25; moreover, the dot plots for the top cluster markers of these analysis only showed genes with a pct.1 > 0.5 and pct.2 < 0.2.

GO term enrichment analysis was performed in PANTHER with the statistical overrepresentation test for biological process.

The Venn Diagram (**Figure 3E**) was generated using the online tool molbiotools. By selecting the Multiple List Comparator we were able to compare the overlapping TFs/DNAB genes between the lists comparing type I NBs *vs.* type I GMCs and type II NBs *vs*. dINPs.

ACKNOWLEDGEMENTS

We are in debt with C. Desplan for the encouragement and the opportunity to perform the scRNA-Seq experiments. We thank F. Simon for his help during the scRNA-Seq protocol. We also thank C. Desplan and his group for experimental help and discussion regarding this study. We thank J. Knoblich and F. Bonnais for fly

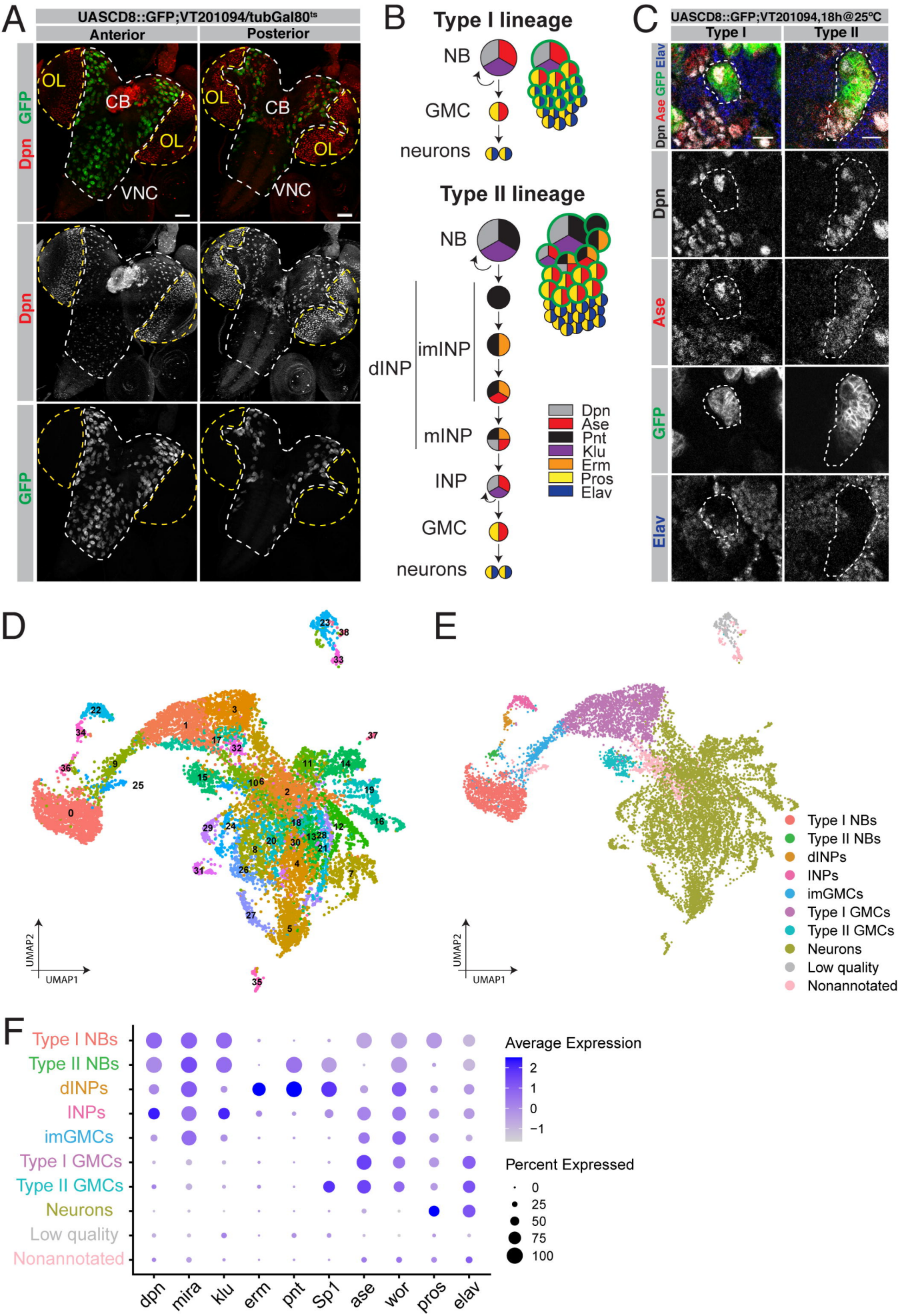
lines and antibodies. We thank A.M. Venda for providing IFs images for E(spl)-mγ-HLH expression in imGMCs. We thank all members of the C.C. Homem Lab and F. Pinto-Teixeira for helpful discussion. We also thank R. Teodoro for revising the manuscript. We also thank CEDOC's Microscopy and Fly facilities for technical support.

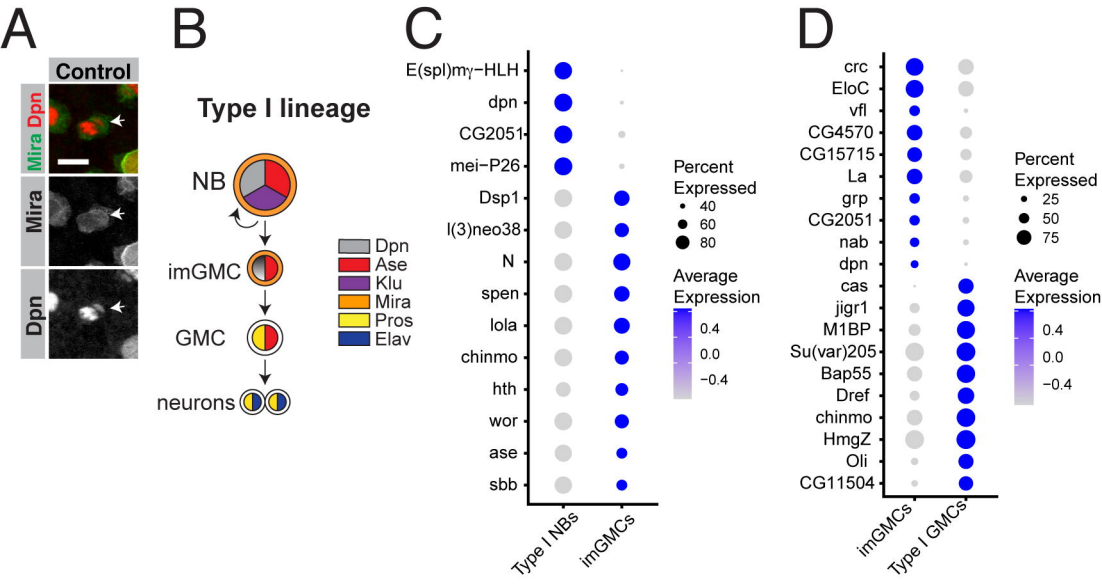
Stocks obtained from the Bloomington *Drosophila* Stock Center (NIH P40OD018537) were used in this study. The monoclonal antibodies were obtained from the Developmental Studies Hybridoma Bank, created by the NICHD of the NIH and maintained at The University of Iowa, Department of Biology, Iowa City, IA 52242. CEDOC's Fly facility was funded by CONGENTO LISBOA-01-0145-FEDER-022170.

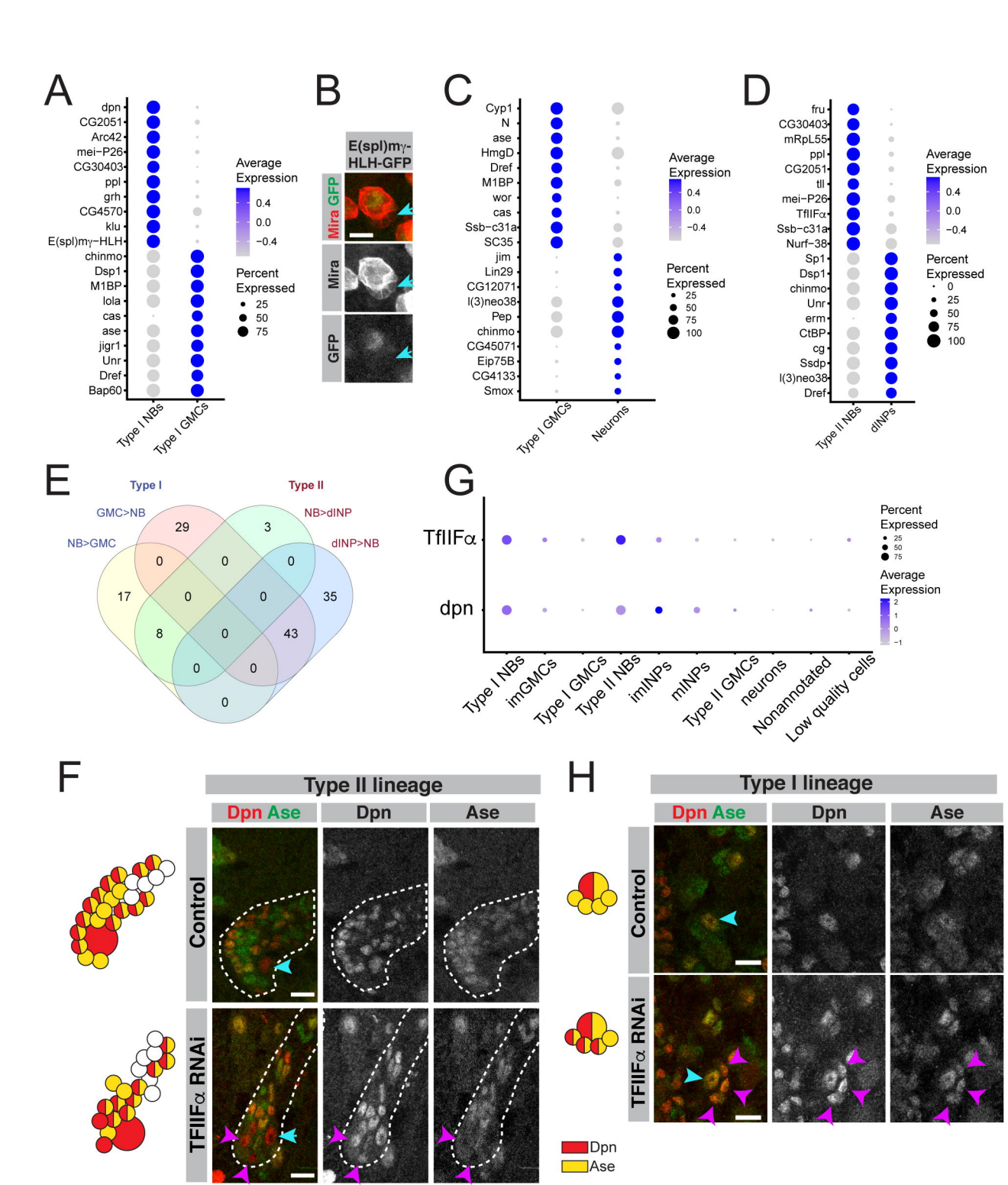
This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (H2020-ERC-2017-STG-GA 759853-StemCellHabitat); by Wellcome Trust and Howard Hughes Medical Institute (HHMI-208581/Z/17/Z-Metabolic Reg SC fate); EMBO Installation grant (H2020-EMBO-3311/2017/G2017); by Fundação para a Ciência e Tecnologia (IF/01265/2014/CP1252/CT0004 and PD/BD/114253/2016 to GSM). N.K. was supported by the National Eye Institute (K99 EY029356-01). This work was also supported by an NYU Abu Dhabi Research Institute grant to the NYUAD Center for Genomics and Systems Biology (ADHPG-CGSB).

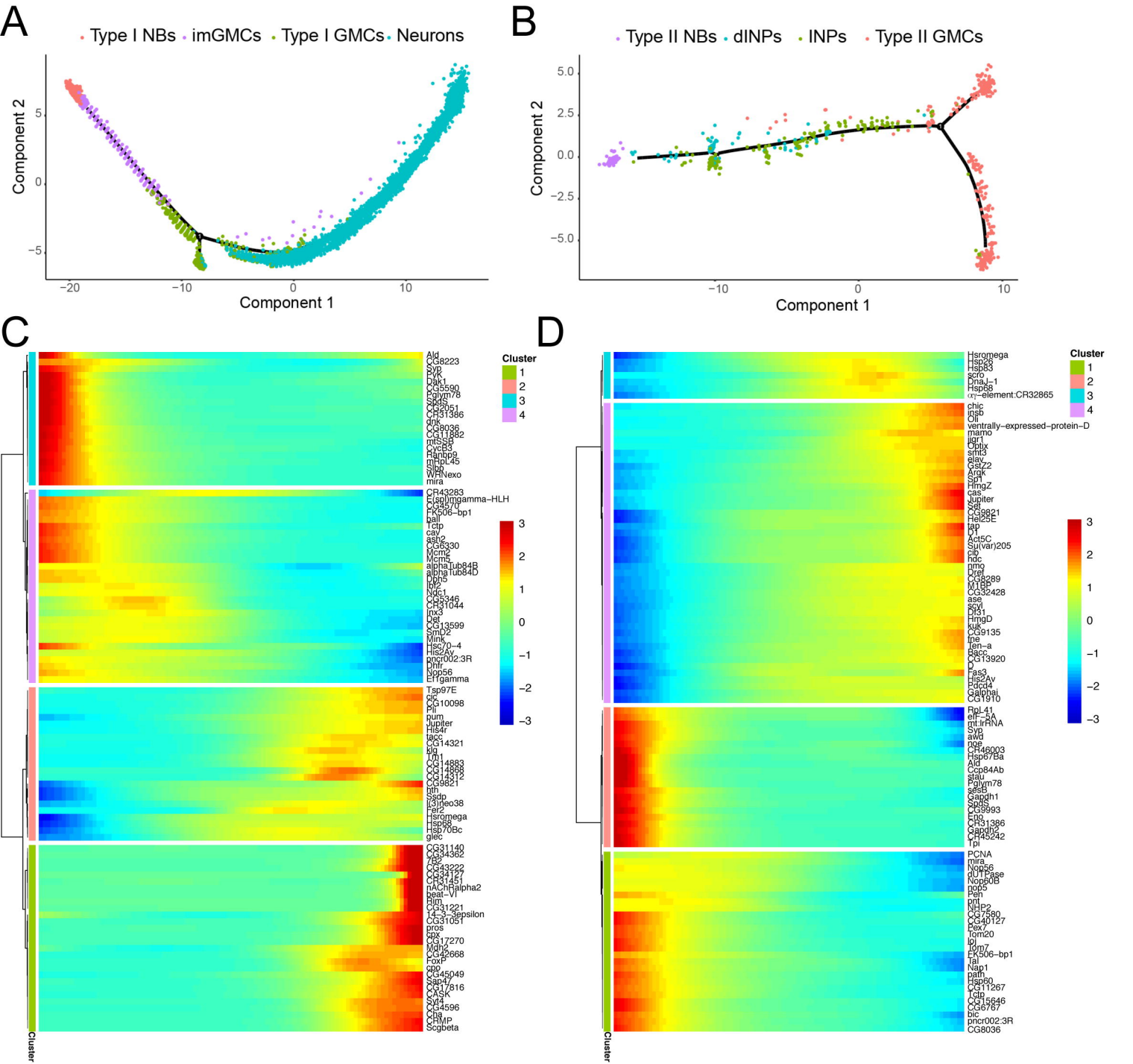
AUTHOR CONTRIBUTIONS

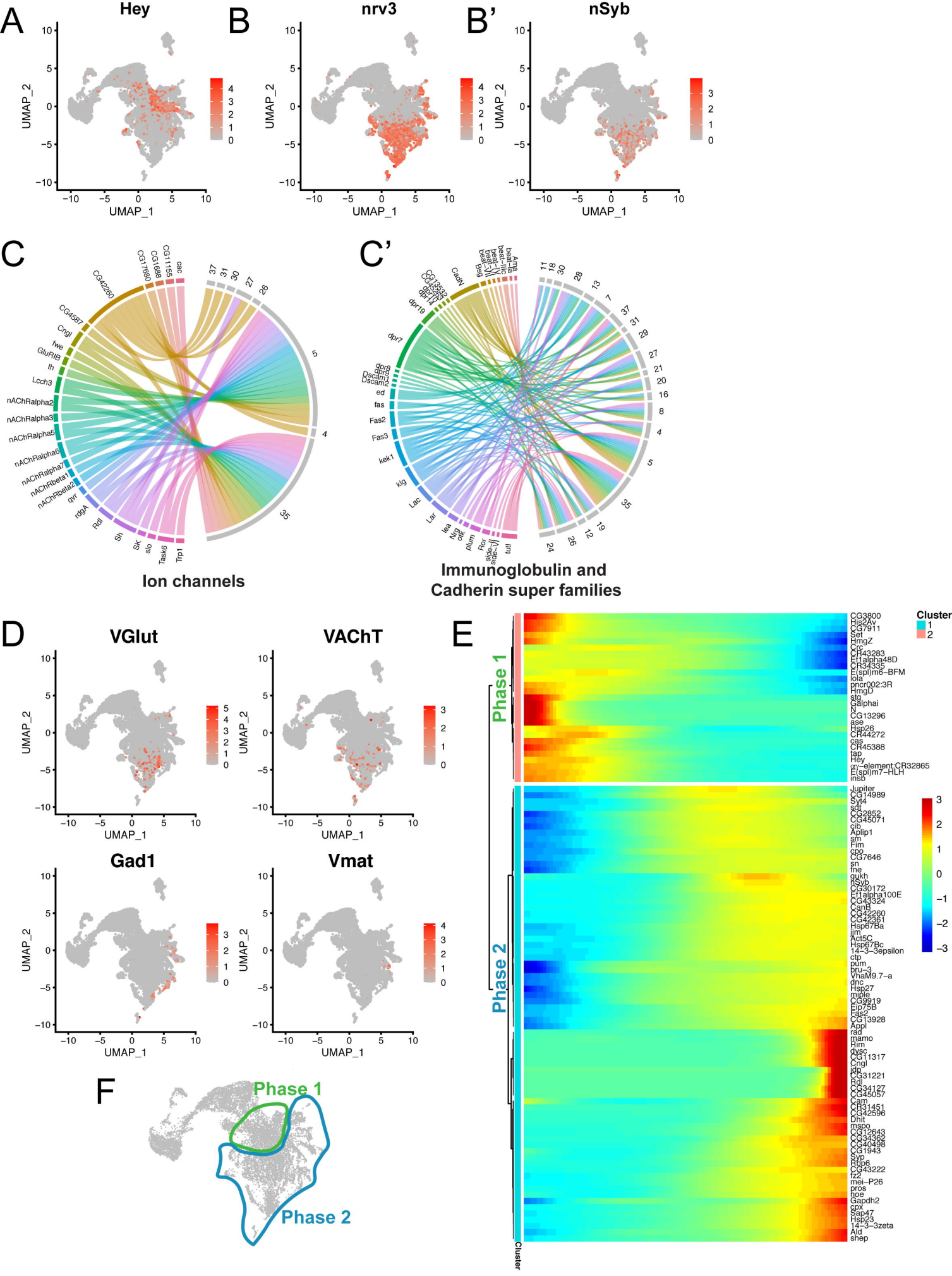
G.S.M. and C.C.H designed the project. G.S.M. performed the genetic experiments and G.S.M. and C.C.H. analyzed the data. G.S.M and N.K. acquired the scRNA-Seq data and G.S.M. and J.T.R. performed bioinformatic analysis using Seurat, G.S.M. performed the analysis using Monocle and P.B. performed the analysis using Velocyto. G.S.M. and C.C.H. wrote the manuscript.

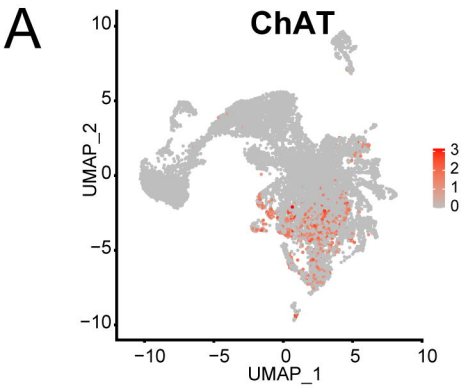

A
UASCD8::GFP;VT201094/tubGal80ts
Anterior
Posterior
Dpn GFP
Dpn
GFP
OL
CB
OL
VNC
B
Type I lineage
NB
GMC
neurons
Type II lineage
NB
dINP
imINP
mINP
INP
GMC
neurons
Dpn
Ase
Pnt
Klu
Erm
Pros
Elav
C
UASCD8::GFP;VT201094,18h@25ºC
Type I
Type II
Dpn Ase GFP Elav
Dpn
Ase
GFP
Elav
D
UMAP2
UMAP1
E
UMAP2
UMAP1
Type I NBs
Type II NBs
dINPs
INPs
imGMCs
Type I GMCs
Type II GMCs
Neurons
Low quality
Nonannotated
F
Type I NBs
Type II NBs
dINPs
INPs
imGMCs
Type I GMCs
Type II GMCs
Neurons
Low quality
Nonannotated
dpn
mira
klu
erm
pnt
Sp1
ase
wor
pros
elav
Average Expression
2
1
0
−1
Percent Expressed
0
25
50
75
100

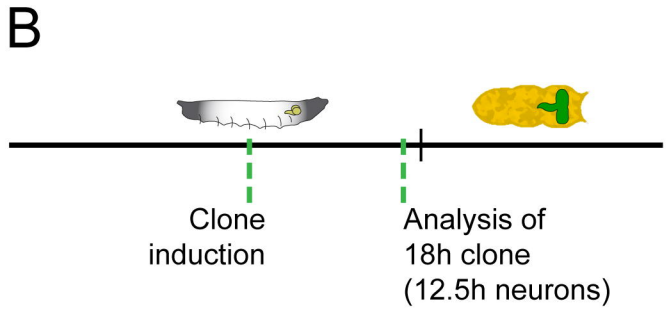

A
Control
Mira Dpn
Mira
Dpn
B
Type I lineage
NB
imGMC
GMC
neurons
Dpn
Ase
Klu
Mira
Pros
Elav
C
E(spl)mγ−HLH
dpn
CG2051
mei−P26
Dsp1
l(3)neo38
N
spen
lola
chinmo
hth
wor
ase
sbb
Type I NBs
imGMCs
Percent Expressed
40
60
80
Average Expression
0.4
0.0
−0.4
D
crc
EloC
vfl
CG4570
CG15715
La
grp
CG2051
nab
dpn
cas
jigr1
M1BP
Su(var)205
Bap55
Dref
chinmo
HmgZ
Oli
CG11504
imGMCs
Type I GMCs
Percent Expressed
25
50
75
Average Expression
0.4
0.0
−0.4

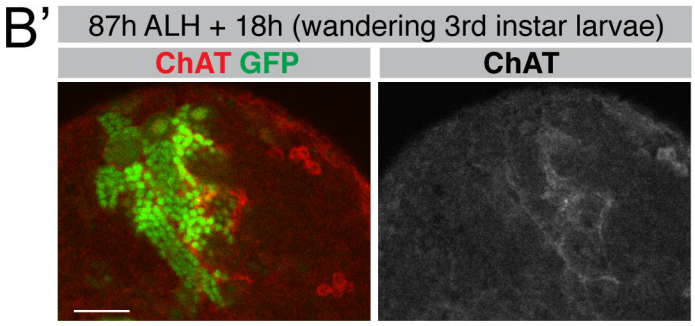

A
dpn
CG2051
Arc42
mei−P26
CG30403
ppl
grh
CG4570
klu
E(spl)mγ−HLH
chinmo
Dsp1
M1BP
lola
cas
ase
jigr1
Unr
Dref
Bap60
Type I NBs
Type I GMCs
Average Expression
0.4
0.0
−0.4
Percent Expressed
25
50
75
B
E(spl)mγ-HLH-GFP
Mira GFP
Mira
GFP
C
Cyp1
N
ase
HmgD
Dref
M1BP
wor
cas
Ssb−c31a
SC35
jim
Lin29
CG12071
l(3)neo38
Pep
chinmo
CG45071
Eip75B
CG4133
Smox
Type I GMCs
Neurons
Average Expression
0.4
0.0
−0.4
Percent Expressed
25
50
75
100
D
fru
CG30403
mRpL55
ppl
CG2051
tll
mei−P26
TfIIFα
Ssb−c31a
Nurf−38
Sp1
Dsp1
chinmo
Unr
erm
CtBP
cg
Ssdp
l(3)neo38
Dref
Type II NBs
dINPs
Average Expression
0.4
0.0
−0.4
Percent Expressed
0
25
50
75
100
E
Type I
Type II
GMC>NB
NB>GMC
NB>dINP
dINP>NB
29
3
0
0
0
17
0
0
35
8
0
43
0
0
0
G
TfIIFα
dpn
Type I NBs
imGMCs
Type I GMCs
Type II NBs
imINPs
mINPs
Type II GMCs
neurons
Nonannotated
Low quality cells
Percent Expressed
25
50
75
Average Expression
2
1
0
−1
F
Type II lineage
Dpn Ase
Dpn
Ase
Control
TFIIFα RNAi
H
Type I lineage
Dpn Ase
Dpn
Ase
Control
TFIIFα RNAi
Dpn
Ase

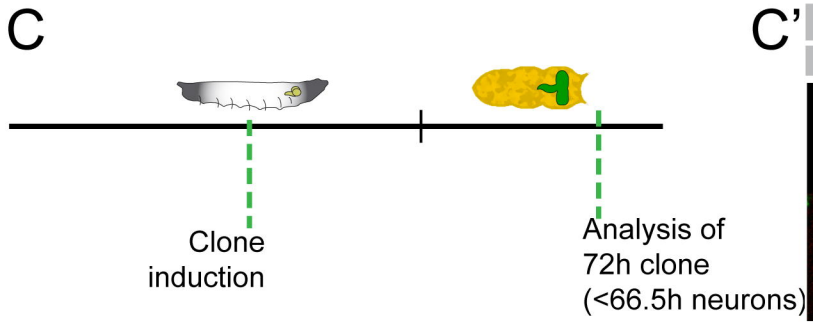

A
Type I NBs
imGMCs
Type I GMCs
Neurons
Component 2
Component 1
B
Type II NBs
dINPs
INPs
Type II GMCs
Component 2
Component 1
C
Cluster
1
2
3
4
D
Cluster
1
2
3
4

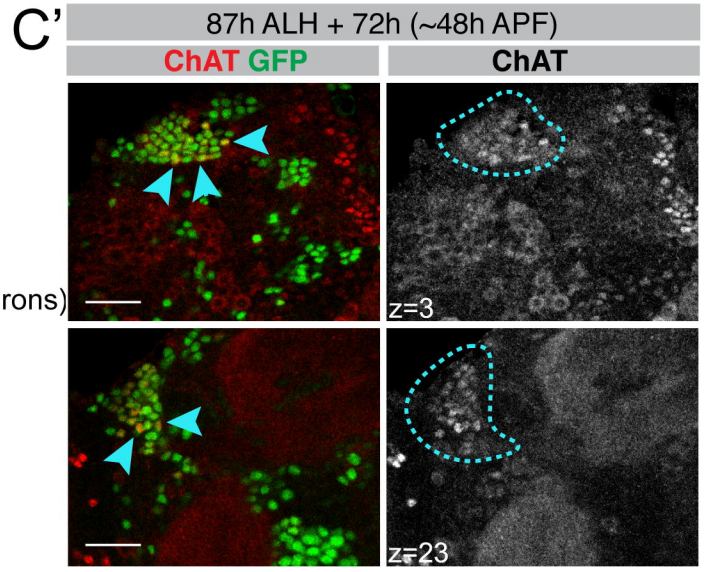

A
Hey
B
nrv3
B'
nSyb
UMAP_1
UMAP_2
C
Ion channels
C'
Immunoglobulin and
Cadherin super families
D
VGlut
VAChT
Gad1
Vmat
E
Phase 1
Phase 2
Cluster
F
Phase 1
Phase 2

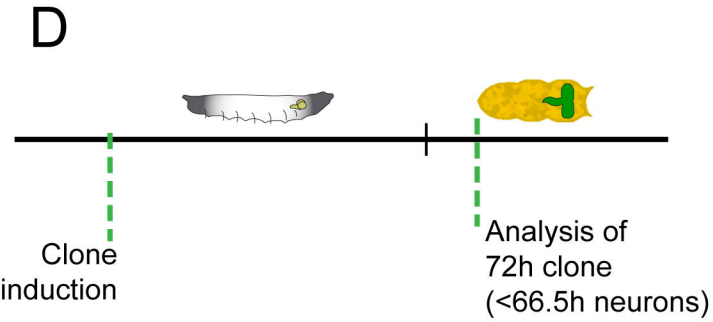

A
ChAT
UMAP_2
UMAP_1
10
5
0
−5
−10
3
2
1
0

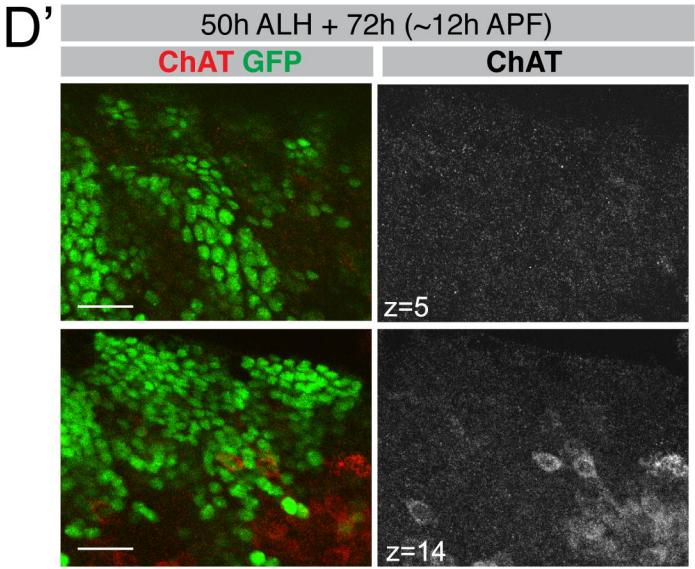

B
Clone induction
Analysis of 18h clone (12.5h neurons)


B'
87h ALH + 18h (wandering 3rd instar larvae)
ChAT GFP
ChAT

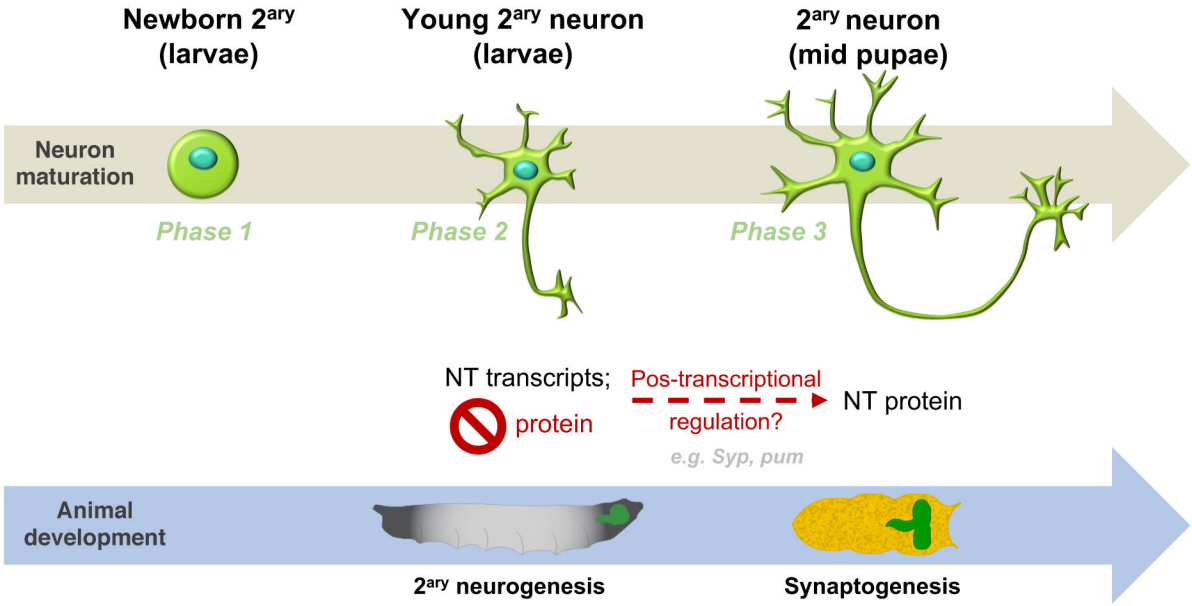

C
Clone induction
Analysis of 72h clone (<66.5h neurons)


C'
87h ALH + 72h (~48h APF)
ChAT GFP
ChAT
z=3
z=23

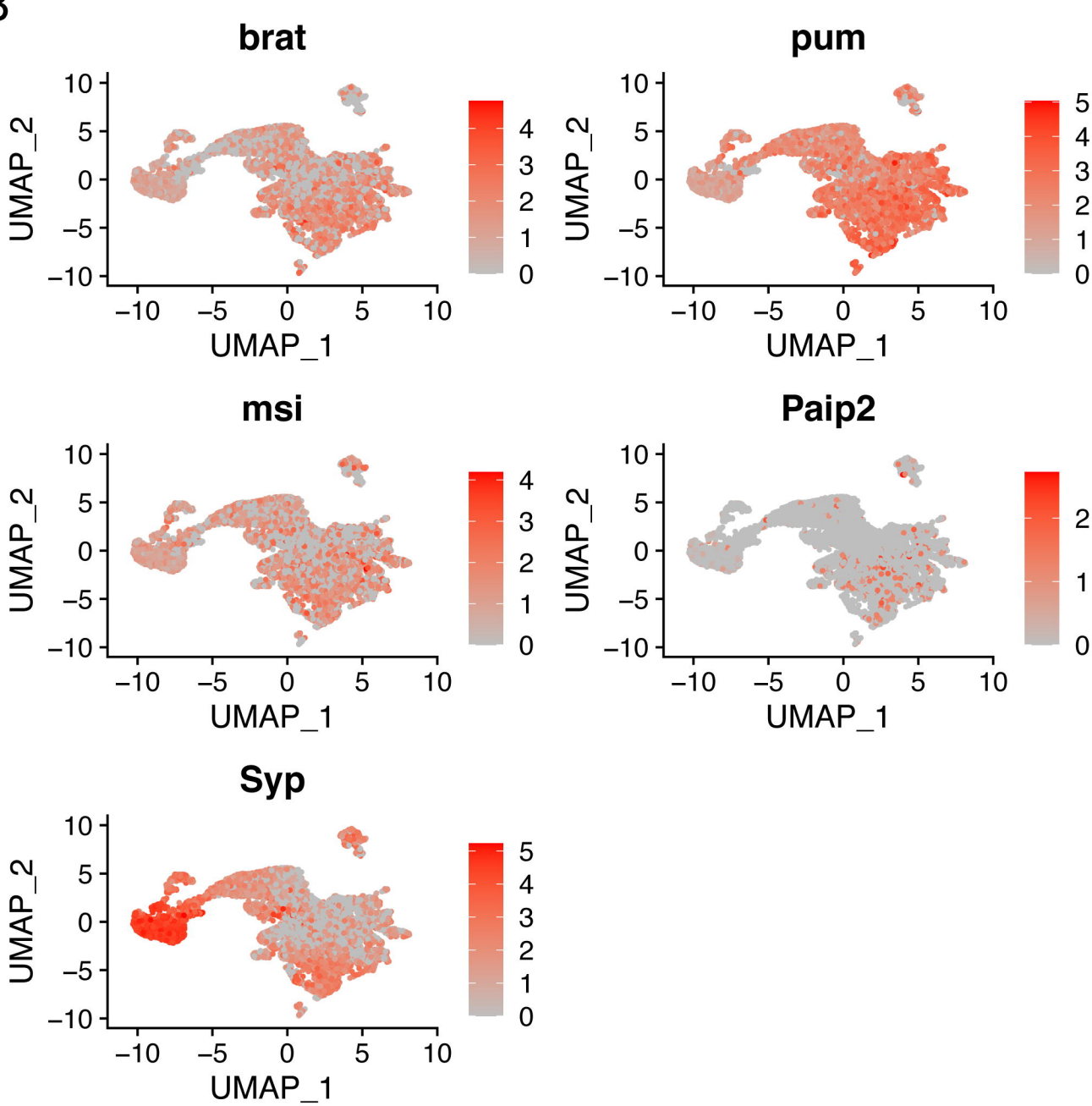

D
Clone induction
Analysis of 72h clone (<66.5h neurons)


D'
50h ALH + 72h (~12h APF)
ChAT GFP
ChAT
z=5
z=14

A


Newborn 2ary (larvae)
Young 2ary neuron (larvae)
2ary neuron (mid pupae)
Neuron maturation
Phase 1
Phase 2
Phase 3
NT transcripts;
protein
Pos-transcriptional regulation?
NT protein
e.g. Syp, pum
Animal development
2ary neurogenesis
Synaptogenesis

B


brat
pum
msi
Paip2
Syp
UMAP_1
UMAP_2