HHS Public Access

Author manuscript

Cell Rep. Author manuscript; available in PMC 2021 May 21.

Published in final edited form as:

Cell Rep. 2021 April 27; 35(4): 109039. doi:10.1016/j.celrep.2021.109039.

The molecular landscape of neural differentiation in the developing *Drosophila* brain revealed by targeted scRNA-seq and multi-informatic analysis

Nigel S. Michki[1], **Ye Li**[2], **Kayvon Sanjasaz**[3], **Yimeng Zhao**[2], **Fred Y. Shen**[4], **Logan A. Walker**[1], **Wenjia Cao**[7], **Cheng-Yu Lee**[2,5,6,8], **Dawen Cai**[1,2,4,9,*]

[1]Biophysics LS&A, University of Michigan, Ann Arbor, MI, USA

[2]Department of Cell and Developmental Biology, University of Michigan Medical School, Ann Arbor, MI, USA

[3]Molecular, Cellular, and Developmental Biology LS&A, University of Michigan, Ann Arbor, MI, USA

[4]Neuroscience Graduate Program, University of Michigan Medical School, Ann Arbor, MI, USA

[5]Life Sciences Institute, University of Michigan, Ann Arbor, MI, USA

[6]Division of Genetic Medicine, Department of Internal Medicine, University of Michigan Medical School, Ann Arbor, MI, USA

[7]Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI, USA

[8]Comprehensive Cancer Center, University of Michigan Medical School, Ann Arbor, MI, USA

[9]Lead contact

SUMMARY

The *Drosophila* type II neuroblast lineages present an attractive model to investigate the neurogenesis and differentiation process as they adapt to a process similar to that in the human outer subventricular zone. We perform targeted single-cell mRNA sequencing in third instar larval brains to study this process of the type II NB lineage. Combining prior knowledge, *in silico* analyses, and *in situ* validation, our multi-informatic investigation describes the molecular landscape from a single developmental snapshot. 17 markers are identified to differentiate distinct maturation stages. 30 markers are identified to specify the stem cell origin and/or cell division

*Correspondence: dwcai@umich.edu.

AUTHOR CONTRIBUTIONS

N.S.M. and D.C. conceived of the project and designed experiments with critical inputs from Y.L. and C.-Y.L. N.S.M., Y.L., K.S., and Y.Z. performed experiments. N.S.M., L.A.W., and W.C. performed data analysis. N.S.M. designed and developed the MiCV software. F.Y.S. developed the HCR protocol for larval brain staining. N.S.M. and D.C. wrote the manuscript with critical insights from Y.L., L.A.W., and C.-Y.L. D.C. initiated and supervised the project.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.celrep.2021.109039.

DECLARATION OF INTERESTS

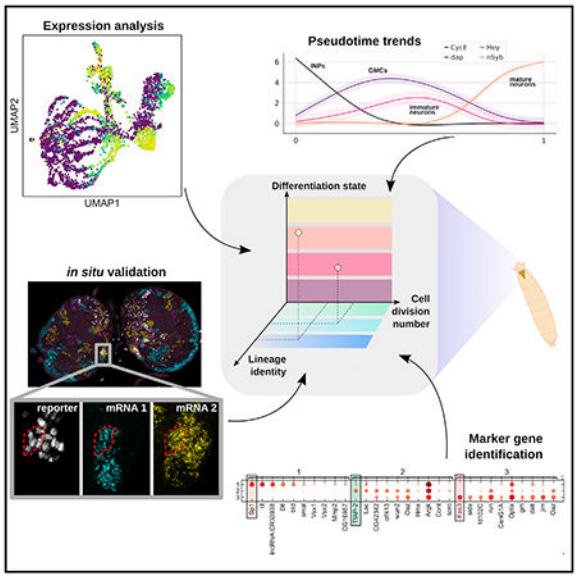
N.S.M. is the founder of MiOmics Inc. (MI, USA).

numbers of INPs, and at least 12 neuronal subtypes are identified. To foster future discoveries, we provide annotated tables of pairwise gene-gene correlation in single cells and MiCV, a web tool for interactively analyzing scRNA-seq datasets. Taken together, these resources advance our understanding of the neural differentiation process at the molecular level.

In brief

Using a combination of targeted scRNA-seq, *in situ* RNA staining, and a multi-informatic analysis paradigm, Michki et al. characterize the transcriptome landscape of thousands of type II neurons and their progenitors in the developing larval fruit fly brain.

Graphical Abstract



INTRODUCTION

The brain is generated by a set of complex fate-specification mechanisms that birth a diverse pool of neural and glial subtypes. These mechanisms rely upon some of the approximately 1,500 transcription factors found in the vertebrate genome (Zhou et al., 2017). Understanding which of these transcription factors play a role in neural fate specification remains an open area of basic research across model organisms (Bayraktar and Doe, 2013; Homem and Knoblich, 2012; Soldatov et al., 2019; Zhong et al., 2018). In particular, untangling the interplay of intrinsic (cell-specific) and extrinsic (global, spatial) fate patterning mechanisms remains particularly challenging, especially in the complex and large vertebrate brain.

Drosophila melanogaster represents a model organism that recapitulates features of vertebrate neurogenesis. Unlike the abundant type I neuroblasts (NBs, neural stem cells), the 16 type II NBs in the *Drosophila* brain adopt a neurogenesis process that is directly analogous to that observed in mammalian cortical development (Homem and Knoblich, 2012). During development, each type II NB undergoes repeated asymmetric cell divisions

to generate an NB and a sibling progeny that acquires a progenitor identity (i.e., intermediate neural progenitor, INP). Each INP undergoes limited rounds of asymmetric cell division to re-generate and to produce a ganglion mother cell (GMC), which divides once more to become two neuron(s) and/or glial cell(s). Along this NB-INP-GMC-neuron maturation process, cells express a well-defined cascade of transcription factors that mark these cell-differentiation stages (Ren et al., 2017; Syed et al., 2017). In parallel, INPs born in each division cycle may express a cascade of transcription factors unique to each NB lineage that contribute to the generation of different neural progenies (Bayraktar and Doe, 2013). It is highly plausible that the combination of these two transcription factor cascades alongside a third molecular axis, which defines unique NBs (i.e., each NB generates a distinct lineage), brings about the generation of a highly diverse neuronal pool (Figure 1A).

The advent of high throughput single-cell mRNA sequencing (scRNA-seq) technologies has enabled researchers to broadly investigate the mRNA expression landscape of hundreds of thousands of cells (Macosko et al., 2015; Ziegenhain et al., 2017). Coupled with a wide variety of analytical tools (Butler et al., 2018; Wolf et al., 2018), researchers can make hypotheses about the number of unique cellular subtypes in the brain (Cocanougher et al., 2019; Saunders et al., 2018), what the functions of these subtypes might be (Ren et al., 2019), and what subtypes might arise together along a common developmental pathway (Cao et al., 2019; Qiu et al., 2017; Soldatov et al., 2019). While such “cell atlas” style scRNA-seq datasets effectively characterize the transcriptomes of a majority of cells from a region of interest, cell populations that are classically clustered together (through *in situ* and/or functional analyses, for example) may not be identified by blind *in silico* cluster analysis (Kiselev et al., 2019). In addition, broad scRNA-seq studies often do not take advantage of the extensive collection of genetic labeling tools that can highlight classically clustered cell populations, enabling them to be studied in greater detail. For instance, a targeted approach to scRNA-seq is required if we are to confidently and efficiently describe nuanced developmental systems, such as the specification of unique neural subtypes derived from the type II NB lineages of *Drosophila*, where inclusion of non-type II-derived cells (making up the majority of the fly brain) would introduce overwhelming noise and confound our analysis.

In the type II NB lineages of *Drosophila*, we set out to broadly classify the molecular factors that define the neural progenies of dividing INPs along three key fate-patterning axes, i.e., differentiation state, division number, and progenitor lineage (Figure 1A) using targeted scRNA-seq. We created a long-living fluorescent reporter to brightly label the type II progenies at the third instar larval stage and using fluorescence-activated cell sorting (FACS) sorted them in preparation for 10× Chromium scRNA-seq (Figure 1B). We subsequently recovered transcriptomes containing 11,622 genes from 6,092 cells. Through an iterative process of cell clustering, marker gene analysis, pseudotime analysis, and *in situ* validation, we identified genes that vary in expression along all three neural fate-patterning axes mentioned above. These genes include markers that globally define the INP, GMC, and neuron differentiation stages in most NB lineages. Further *in silico* analysis suggested molecular factors that are uniquely expressed in subpopulations of INPs, GMCs, and immature and mature neurons. Subsequent *in situ* mRNA staining recovered the spatial relationship of these molecular factors, which clarified the cell division number and NB

lineage specificity. We finally identified novel markers that exclusively label distinct neural subsets. These new markers further enabled building novel neural developmental trajectories that lead to unique neuronal cell fates. Our multi-informatic approach to targeted scRNA-seq experimental design and analysis provides a roadmap for navigating the differentiation process of complex brains. Our annotated scRNA-seq data and interactive analysis tools provide valuable resources for future discoveries.

RESULTS

Type II neuroblast-derived cells are uniquely identified from the mixed optic lobe cell population using descriptive quality-control metrics and clustering

To perform *targeted* scRNA-seq, we brightly labeled the type II NB progenies with a long-lasting fluorescent reporter. We created an UAS-hH2B::2xmNG reporter fly, in which two copies of the mNeonGreen (2xmNG) fluorescent protein are fused to the C' terminus of the human histone 2B protein (hH2B). This leverages the expression of multiple copies of a bright fluorescent protein alongside the slower turnover rate of the histone protein (Tumbar et al., 2004). To validate labeling fidelity, we crossed UAS-hH2B::2xmNG to an R9D11-Gal4 driver (Weng et al., 2010). We found that mNG-labeled type II NB progenies and a small subset of medial optic lobe (OL) cells in larval brains (Figure S1A). When comparing our UAS-hH2B::2xmNG to the previously used UAS-IVS-myr::tdTomato reporter, we found that the membrane-targeted myr::tdTomato cells formed clusters that are smaller than the hH2B::2xmNG-labeled cells (Figures S1B–S1D). This indicates that the slow hH2B::2xmNG turnover preserved labeling in progeny cells in which Gal4 was no longer expressed. Finally, the bright nuclear mNG labeling enabled reliable FACS selection for targeted 10× Chromium scRNA-seq (Figure 1B, and detailed in STAR Methods).

Subsequently, we projected the scRNA-seq data onto a 2D UMAP plot and overlaid the counts of all genes, unique transcripts (UMIs), and mitochondrial genes as part of routine scRNA-seq quality control (Figure S2). When overlaying the hH2B::2xmNG reporter transcript counts, we found that mNG transcripts were expressed non-uniformly, with pockets of cells expressing the hH2B::2xmNG transcript at a significantly higher level than others in the dataset (Figure S3A). To examine whether this non-uniform expression pattern reflects true biological variance, we performed *in situ* RNA staining for *mNG* using the HCRv3 protocol (Choi et al., 2018) and imaged the native mNG fluorescence to compare the relationship of mNG transcripts and proteins (STAR Methods). We found that each of the type II clusters indeed expresses a high level of *mNG* transcripts in only a small subpopulation of cells near the tip of each lineage (Figures S3B–S3D). This spatial localization, coupled with co-expression of *mNG* transcripts with *D* in $CycE^{+}$ cells (data not shown) leads us to conclude that the R9D11 enhancer fragment's expression is tightly restricted to newly born INPs and their daughter GMCs, emphasizing the need for long-living reporters for investigation of neural subtypes derived from the type II NBs.

To further ensure the specificity of our analysis to type II cells, we performed an *in silico* filtering to exclude the optic-lobe cells that are also labeled by R9D11-Gal4 (Bayraktar et al., 2010). Based on prior literature, at least two genes are *not* expressed in the developing optic-lobe (*lncRNA:cherub* and *dati*; see *in situ* expression patterns from (Landskron et al.,

2018; Schinaman et al., 2014), respectively). In the unsupervised clustering and UMAP projection, two groups of cells can be clearly separated as *cherub*$^{+}$/*dati*$^{+}$ and *cherub*$^{-}$/*dati*$^{-}$, which we define as putative type II and OL cells, respectively (Figures 1C–1E).

To identify other potential marker genes to separate OL and type II cells, we performed a logistic regression-based marker gene analysis (Ntranos et al., 2018) comparing these two major groups against one another (Figure 1F). The transcription factors *mamo* and *bi* are upregulated in the putative OL cells when compared to the putative type II cells, among others. To confirm this, we generated HCR probes against *mamo* and *bi* as novel markers for the OL, and *lncRNA:cherub* and *dati* as markers for cells not in the OL. We subsequently stained larval *R9D11*-hH2B::2xtagBFP brains (Figure 1G) and clearly show that *bi* and *mamo* are both predominantly expressed in the OL, and *lncRNA:cherub* and *dati* are predominantly excluded from the OL. Why *mamo* is upregulated in cells in the OL is unknown. Prior work has shown that *mamo* is required for specification of α′/β′ mushroom body neurons in the developing CNS (Liu et al., 2019). Further study of its role in the OL may elucidate novel function there. On the other hand, *bi* has been shown to be indispensable for neural differentiation in the OL (Pflugfelder et al., 1990). Our finding of *bi* being excluded from the type II lineages expands our knowledge of its expression specificity.

From our *in silico* filtering process, we confidently separated the type II-derived cells from optic lobe cells that were also captured in our scRNA-seq experiment. Only these type II-derived cells were carried forward for our downstream analysis.

Pseudotime analysis describes the continuous differentiation stages of type II-derived cells

Knowing that the *R9D11*-hH2B::2xmNG reporter specifically labels type II progenies from INPs to maturing neurons, we aimed to first align each cell along a continuous cellular differentiation state axis (Figure 1A). We expected this would reveal the most prominent underlying structure of our data because, in the case of type II neurogenesis, all cells will similarly transition through the INP, to GMC, to immature, to mature neuron differentiation states. Using the Markov chain-based pseudotime analysis algorithm *Palantir* was a natural choice as Markov chains describe discrete transitions that occur randomly based upon a continuous probability distribution (Setty et al., 2019). Given a properly chosen starting cell, *Palantir* aligns cells in our scRNA-seq data based upon the path of fewest transcriptomic changes propagating from the starting cell.

Cells expressing high levels of the INP markers *CycE* and *D* are good candidate starting cells for *Palantir* (Bayraktar and Doe, 2013; Yang et al., 2017). To easily identify these cells from the UMAP plot, we built a multi-informatic cellular visualization web tool (MiCV) to display the single cell co-expression pattern of multiple genes in the 2D/3D UMAP plots. Furthermore, users can conveniently select a subset of cells for specific analysis, such as picking the starting cell(s) for *Palantir*, by combining mouse-click selections from the parallel plots generated by MiCV (STAR Methods). We overlaid the pseudotime result onto the reprojected 2D UMAP plot that only included type II NB-derived cells. Based on the expression of known marker genes (Figure S4), we predicted INP, GMC, immature, and

mature neuron clusters (Figure 2A, dash lines). Interestingly, these cell maturation state clusters aligned well with the pseudotime arrangement. For example, using MiCV, we displayed the single-cell co-expression pattern of *CycE, dap*, and *nSyb* (Figure 2B), which are known to distinguish the INP, GMC/immature neuron, and mature neuron states, respectively, and found their UMAP positions matched well with their pseudotime alignments (Figure 2A).

To describe the dynamics of gene expression across pseudotime, and thus the differentiation process, we fit a gene-expression trend line to each gene detected in our scRNA-seq dataset using *PyGAM* (Servén et al., 2018). Indeed, we found that the expression peaks of four marker genes, i.e., *CycE* for INPs (Yang et al., 2017), *dap* for GMCs (Lane et al., 1996; de Nooij et al., 1996), *Hey* for a subset of the transient immature neuronal state (Monastirioti et al., 2010), and *nSyb* for maturing neurons (Deitcher et al., 1998), aligned in this exact differentiation order along the calculated pseudotimeline (Figure 2C). Hence, we can use the relative expression levels of these genes to approximate the boundaries of the continuously changing differentiation states (Figure 2A, dashed lines) in pseudotime. Subsequently, we performed gene-expression trend clustering using *phenograph* (Levine et al., 2015) to screen novel putative marker genes whose expression trend matched one of the four known marker genes' (Figures 2D–2G). Independently, we used a marker gene-based differentiation state scoring (Wolf et al., 2018) strategy to separate these differentiation stages and found similar sets of marker genes (Figure S4). Interestingly, many of the putative marker genes do not have any known function related to neural differentiation. Further pathway analysis and gene-manipulation studies will be needed to explore their exact roles in type II neurogenesis.

Nonetheless, we profiled the *in situ* expression patterns of some putative marker genes we identified in this analysis. We first synthesized HCRv3 probes against the canonical makers *CycE, dap, Hey*, and *nSyb* transcripts (STAR Methods) and used these probes to investigate their expression pattern in the type II NB-derived cells using our novel reporter fly. As predicted, these genes form largely non-overlapping expression patterns in the larval brain (Figure 2H, left panel). We found that *CycE* transcripts were expressed in large neuroblasts as indicated by the large cell bodies (Figure 2H, right panels, asterisk) and in smaller tagBFP positive cells as a marker for replicating INPs. As predicted, *dap, Hey*, and *nSyb* transcripts expressed in bands of cells that were sequentially positioned away from the neuroblast (Figure 2H, right panels, dashed lines). Next, from the gene-expression trend clustering result (Figures 2D–2G), we selected four candidate markers and performed similar HCR *in situ* mRNA profiling. The *in situ* results suggest that *ytr, tap, E(spl) m6-BFM*, and *jim* transcripts express in unique patterns (Figure 2I, right panels), and the co-expression MiCV plots indicate that these markers largely overlap the canonical makers in the respective cells (Figures 2J–2M). In particular, *E(spl)m6-BFM*, and *jim* were expressed almost exclusively in immature neurons and maturing neurons, respectively (Figures 2L–2M). However, while the putative INP marker *ytr* expressed in 96% of all the INPs, it also expressed in 37% of GMCs and 38% of maturing neurons (Figure 2J). This observation indicates that *ytr* broadly expresses in INPs and that its expression may be selectively maintained in a subset of GMCs and their progeny neurons. The putative GMC marker *tap* appears to express in subsets of INPs and approximately half of the immature neurons (Figure 2K). This suggests that tap

may be a gene that defines one daughter neuron during their mother GMC's terminal cell division.

Though many genes that trend along the differentiation state axis are potentially interesting, we highlight here the gene *E(spl)m6-BFM*, a member of the *Notch*-responsive subgroup of the "enhancer of split" family of transcription factors (Lai et al., 2000). This family of proteins is responsible for regulating a variety of developmental processes (Maier et al., 1993), and their group's function in balancing the self-renewal of differentiation in the type II neuroblasts of *Drosophila* has recently been described (Li et al., 2017). However, the specific function or restricted spatial expression of *E(spl)m6-BFM* in the developing larval brain has not been established. Based on our analysis, *E(spl)m6-BFM* marks a subset of the cells in the transient immature neuronal state, which comes about directly after the mother GMC's terminal cell division. This is similar to *Hey*, a previously identified immature neuron marker, which is upregulated in only one of the two daughter neurons of this terminal GMC division (Monastirioti et al., 2010) and activates in a *Notch*-dependent manner. Our scRNA-seq data indicate that *E(spl)m6-BFM* is expressed in both Hey^{+} cells and Hey^{-} cells that have similar pseudotime values, though $Hey^{+}/E(spl)m6\text{-}BFM^{-}$ cells are also present (Figure 2L). To validate this, we used HCR probes for both *Hey* and *E(spl)m6-BFM* and identified subsets of immature neurons that were only Hey^{+}, only $E(spl)m6\text{-}BFM^{+}$, or *Hey* and *E(spl)m6-BFM* double-positive (Figure S5). Similar to *E(spl)m6-BFM, Rbp*, a protein known to be functionally required for synaptic homeostasis and neurotransmitter release (Liu et al., 2011; Müller et al., 2015), is also upregulated only in this immature neuronal subset (data not shown). Further study will be desired to understand why either of these genes undergo a burst of expression in the *immature* neuronal state and to establish their functional roles at the protein level.

INP and GMC sub-clustering enables the identification of novel maturation pathways that are convolved with the canonical *Dichaete, grainy-head, eyeless* transitions

Having used pseudotime analysis to define the major differentiation states in the type II neurogenesis process, we next characterized the cellular heterogeneity within these states using automated scRNA-seq clustering analysis. Such analysis may or may not obviously reflect previously established models of cell type differentiation/diversity, especially when this diversity could refer to any of/all the axes of cell type differentiation (Figure 1A). Nonetheless, we performed Leiden clustering (Traag et al., 2019) with a low resolution (0.6) and overlaid the result on the reprojected UMAP (Figure 3A, left). We found that cluster 1 and 0 included 561 and 563 cells, which correspond to the INP and GMC populations in the above-mentioned pseudotime analysis, respectively. Subsequently, we took these putative INP and GMC cells and found they could be clustered into four groups of INPs and four groups of GMCs (Figure 3A, right).

To discover which genes distinguished each subcluster, we performed logistic regression-based marker gene analysis and plotted the top 10 genes that defined the INP (Figure 3B) or GMC subclusters (Figure 3C). We found that this clustering result reflects a convolution of the lineage-specific canonical *Dichaete, grainy-head, eyeless* transitions outlined in Bayraktar and Doe (2013), which have been indicated to sequentially express in young to

old INPs over the course of their division cycles (Bayraktar and Doe, 2013). *D* expression was rather specific in 74% of subcluster 4 INP cells and in 78% of subcluster 1 GMC cells, while only expressing in fewer than 28% of other subcluster cells (Figure 3D). On the contrary, *grh* and *ey* expressions are intermingled in the other subclusters (Figures 3E and 3F, respectively).

Interestingly, we found that *Sp1, TfAP-2*, and *Fas3*, among the top marker genes in this clustering analysis, not only expressed in segregated subclusters but also marked both INP and GMC subclusters (Figures 3G–3I, respectively). We suspected that the GMC subclusters specified by these genes might be the direct progenies of the INP subclusters that carry over the *Sp1, TfAP-2*, and *Fas3* transcripts. We subsequently counted the number of top 100 marker genes that were shared between each of the INP and GMC subclusters. The correlation plot strongly suggests that GMC subclusters 0, 1, and 6 are likely the progeny of INP subclusters 3, 4, and 5, respectively (Figure 3J).

The choice of clustering resolution can be somewhat arbitrary, and the 8 subclusters for INPs and for GMCs here may represent a surface level of INP patterning that can be further broken down into more subtypes. Since we saw a clear link between 6 of these 8 subclusters, we decided to pursue *in situ* validation experiments for the marker genes identified at the 8-subcluster resolution in follow-up experiments and aimed to do so in an exploratory manner, taking *Sp1, TfAP-2*, and *Fas3* (the top marker genes for the relevant GMC subclusters) as promising marker genes to investigate further.

The transcription factor *Sp1* is expressed in young INPs throughout the DM1–6 and DL1 lineages and marks a unique neural progeny

We first aimed to *in situ* profile the transcript expression of *Sp1*, a Cys2His2-type zinc finger transcription factor that is necessary for the specification of type II neuroblasts (Álvarez and Díaz-Benjumea, 2018). We reasoned that this, along with the apparent co-expression of *Sp1* with *D* in the INPs of our scRNA-seq dataset (Figures 3D and 3G, respectively), would imply that *Sp1* may be broadly expressed in young, newly matured INPs of most type II NB lineages. We synthesized HCRv3 probes against *Sp1* and *D* transcripts (STAR Methods) and validated their specificity using gene-trap reporter flies (Figure S6). When accessing their expression patterns in the type II NB-derived cells, we found that *Sp1* mRNA was expressed prominently in all type II lineages with the possible exception of DL2 (Figures 4A and 4B). On the contrary, *D* mRNA expressed prominently in DM1–3, and in much smaller subsets of cells in lineages DM4–6 (data not shown), which is consistent with previous observations (Bayraktar and Doe, 2013).

Our scRNA-seq data indicate that while *Sp1* co-expressed with *D* in more than 30% of INPs (Figure 4C), 8% and 16% of all INPs are $Sp1^{+}/D^{-}$ and $Sp1^{-}/D^{+}$, respectively. To validate the presence of these INP populations *in situ*, we used our HCR protocol to co-stain *Sp1* and *D* mRNA (Figure 4D). We show that, for instance, in the DM6 lineage, an $Sp1^{+}$, D^{+} INP progeny can be identified directly adjacent to cells where either *Sp1* or *D* is exclusively expressed (Figure 4D, enlarged box). Furthermore, we overlaid *Sp1* or *D* expressions on the UMAP plot and found that these two transcripts continue to express in maturing neurons of two exclusive subsets (detailed below). This is consistent with a previous study, which found

that the *D*-expressing young INPs specifically give rise to *D*-expressing neurons (Bayraktar and Doe, 2013). Therefore, we hypothesize that $Sp1^{+}/D^{+}$ INPs may transition to *Sp1* or *D* exclusively expressing INPs, which give rise to distinct neural subtypes. To specify whether *Sp1* protein is expressed in neurons, we labeled the type II progenies with a membrane-bound tdTomato (R9D11-CD4::tdTomato) to visualize neuron's characteristic axonal projections and coupled with an Sp1::GFP reporter line. We show as an example that the DM3 lineage generates many neurons that form a $tdTomato^{+}$ neurite bundle that are also GFP^{+}, which indicates the generation of $Sp1^{+}$ neural progeny (Figure S6B).

Next, we wondered whether *Sp1* is like *D* that expresses strictly in young INPs. We quantified our scRNA-seq data and found that *Sp1* coexpressed with the two canonical late INP markers *grh* and *ey* only in a small subset of INPs (Figure 4C). Taken together, these data support the hypothesis that *Sp1*, much like *D*, is expressed broadly in INPs with low division numbers and that these INPs are responsible for producing a neural progeny similarly marked by *Sp1* expression that is distinct from the D^{+} neural progeny.

The transcription factor *TfAP-2* and cell-adhesion molecule *Fas3* are each expressed in INPs of specific type II neuroblast lineages

We next characterized the spatial expression patterns of *TfAP-2* and *Fas3*, selected markers for the other two major putative INP subtypes identified in our low-resolution clustering (Figure 3). We generated HCR probes against mRNA of *TfAP-2* and *Fas3* in a similar manner to *Sp1* and probed their expression in reporter flies in order to identify which type II NBs generate their respective INP subsets. Unlike *Sp1*, however, *TfAP-2* and *Fas3* transcripts are expressed much more broadly across the brain and are not restricted to the type II lineages (Figures 4F and 4J).

Within the type II progenies, *TfAP-2* mRNA appeared to be expressed prominently in INPs of the DM4–6 lineages as well as a subset of their downstream progeny (Figures 4E and 4F, green outline). However, we did not observe strong *TfAP-2* expression in any other lineage, implying that expression of this marker is primarily lineage restricted (Figures 4E and 4F, arrowheads). Interestingly, *TfAP-2* co-expressed in fewer D^{+} but many more grh/ey^{+} INPs than *Sp1* does in our scRNA-seq data, which indicates that $TfAP\text{-}2^{+}$ INPs have likely undergone some cell divisions before expressing this marker gene (Figures 4C versus 4G). Although *TfAP-2* expresses in fewer lineages than *Sp1*, our scRNA-seq data (data not shown) and *in situ* profiling (Figure 4H) showed that these two genes do indeed co-express in cells belonging to those few lineages. *TfAP-2* plays broad roles in development (Monge et al., 2001), but in the context of the central brain it has been shown to play a role in developing and maintaining the neural circuitry required for night-sleep in adult flies (Kucherenko et al., 2016). Consistently, we found in our scRNA-seq data that *TfAP-2* expressed in a subset of neurons that are distinct from the $Sp1^{+}$ or D^{+} population (data not shown). *TfAP-2*'s expression in neurons is distinct from the previously identified late INP progeny genes *grh* and *ey*; the latter two were not found in neurons in our scRNA-seq data (data not shown). *TfAP-2* (*ap-2*) is significantly orthologous to the human transcription factors TFAP2A/B (Flybase, 2019), and its role in sleep can be traced back to *C. elegans* (Turek et al., 2013). Taken together, this would imply that at least this particular role for

TfAP-2 in the central brain may be evolutionarily conserved and that the neurons generated by *TfAP-2*$^{+}$ INPs in the DM4–6 lineages may play a role in night-sleep circuit maintenance.

Based on our *in situ* RNA staining, *Fas3* mRNA was found to express most prominently in the INPs of DM1–3 (Figures 4I and 4J). Similar to *TfAP-2*, our scRNA-seq data suggest that *Fas3* co-expressed in fewer *D*$^{+}$ but many more *grh/ey*$^{+}$ INPs than *Sp1* does, which indicates that *Fas3* INPs have likely undergone some cell divisions before expressing this marker gene (Figures 4C versus 4K). Again, our scRNA-seq data (data not shown) and *in situ* profiling (Figure 4L) showed that *Fas3* and *Sp1* co-express in a significant fraction of cells. *Fas3* is interesting as a marker gene for INPs as it is *not* a transcription factor but rather a membrane-bound, homophilic cell-adhesion molecule that plays a strong role in synaptic targeting and axonal guidance in a subset of neurons in the central and peripheral nervous systems (Kose et al., 1997; Snow et al., 1989), along with cell-adhesion-mediated morphological development throughout the entirety of the fly (Wells et al., 2013). Why *Fas3* would be expressed so strongly in a subset of INPs is unknown.

A unique combination of transcription factors and surface molecules define putative neural sub-progenies of young INPs

With low resolution (0.6) global clustering, our scRNA-seq data already showed a much greater subtype diversity in neurons (12 clusters) than in GMCs or INPs (1 cluster each) (Figure 5A). We performed logistic-regression based marker gene analysis on these specific clusters to identify the top 100 marker genes for each cluster that are most uniquely expressed with the top 10 marker genes of clusters 4, 6, and 8 are plotted in Figure 5B (full plot for all clusters are shown in Figure S7). Subsequently, we analyzed the top 100 marker genes using the DAVID Functional Annotation Tool (Huang et al., 2009a, 2009b) in order to identify sets of genes that form functionally associated groups based on associated gene ontology (GO) terms. We identified the first GO term from the top three highly enriched functional groups and find that these terms indicate that transcription factors and surface molecules are predominant markers for these three (Figure 5C), as well as all other neural subsets (data not shown).

In the Bayraktar and Doe (2013) study, *bsh* was found to express in a non-overlapping subset of neurons that do not express *D* in the young INP progeny. The same study also specified that there are other young INP-derived neurons are *Bsh*$^{-}$ and *D*$^{-}$, whose markers were not identified using the available method and antibody probes. Interestingly, our automatic analysis reveals that neuron clusters 4, 6, and 8 differentially express *Sp1*$^{+}$, *D*$^{+}$, and *bsh*$^{+}$, respectively (Figure 5B). As we show that *Sp1* expresses in young INPs, it is plausible that *Sp1* is a marker gene that labels the previously unspecified young INP derived neurons. Indeed, our scRNA-seq data show that *Sp1, D*, and *bsh* were expressed in three distinct maturing neuron populations (Figures 5D, 5F, and 5E, respectively). This *in silico* analysis permits rapid identification of transcription factors that potentially belong to the same regulatory pathways to specify neuronal fate. For example, selected from the specific marker gene list, the transcription factors *Ets65A, dac*, and *Awh* are highly co-expressed with neurons expression *Sp1, D*, and *bsh*, respectively (Figures 5G, 5I, 5H, respectively).

Distinct surface molecules are also differentially expressed in different subsets of neurons, which may indicate their roles in forming functionally distinct circuits (Figures 5J–5L). Among them, *Fas3* appears to co-express in a large proportion of *Sp1* neurons, regardless of their low degree of co-expression in the INP and GMC stages (cf. Figures 5D versus 5J). To validate that Fas3 protein is translated in neurons of this developmental stage, we used a Fas3 antibody to stain our novel type II lineage reporter fly and found that it labels neurons in the DM1–3 lineages that form neurite bundles across the commissure (Figures 5M and 5N). It is plausible that the expression of *Fas3* in INP may play a role in enabling some of the neural progenies of DM1–3 to either form these axonal bundles or for them to find their final targets across the commissure early on in the neural maturation process. We further calculated the pairwise gene-gene correlation scores across the whole transcriptome, filtered the highly correlated and anticorrelated genes, and provided functional annotations (Tables S1 and S2; STAR Methods). This will aid others to rapidly discover their own candidate genes of interest.

DISCUSSION

Drosophila type II neural lineages as a model system to study complex neurogenesis processes

To enable the brain's complex functions, vastly diverse neuronal types need to be rapidly generated at a very large scale during development. To reveal how neural stem cells populate the developing brain, efforts have been made to identify cell types and their lineage relationships. For instance, focuses on neuro-development in mouse (Habib et al., 2017; Han et al., 2018; Ponti et al., 2013; Saunders et al., 2018; Soldatov et al., 2019), human brain tissues (Habib et al., 2017), and the developing human prefrontal cortex (Zhong et al., 2018) revealed intermediate stem cells (and critical genes involved) as an important mechanism for rapid cortical expansion. Underlying this rapid and diverse differentiation process is the constant change of gene-expression profiles in all cells. However, the molecular mechanisms that lead to functionally distinct neurons in the mammalian brain remain challenging to describe in detail. This is because, on the one hand, neuronal fate determination involves many genes, and, on the other hand, neural progeny cells originating from distinct lineages undergo rapid migration, which leads their intermingling nature in space.

Although they are the minority (8 stem cells per hemisphere) in the *Drosophila* central brain, the *Drosophila* type II neural lineage has a neurogenesis process analogous to the mammal's rapid cortical expansion (Homem and Knoblich, 2012). Compared to their mammalian counterparts, the *Drosophila* type II neural lineage has the advantage of being non-migrating in the larval stage. With proper labeling, type II progeny cells of the same lineage can be identified as a segregated cell cluster. Importantly, the cells' spatial relationship within a cluster serves as a considering factor when determining the age and maturation stage of these cells (Boone and Doe, 2008; Homem and Knoblich, 2012). The small stem cell pool and mammal-like lineage composition make the *Drosophila* type II neural lineage an attractive model to study the complex brain development process. In addition, many important genes and signaling pathways are conserved throughout evolution (Homem and Knoblich, 2012; Mariano et al., 2020; Ogawa and Vallender, 2014), which makes revealing

the molecular mechanisms of *Drosophila* type II neuron differentiation a meaningful primer to study the human analogs in the outer subventricular zone.

Summary of this work

In this work, use targeted single-cell transcriptome analysis to advance our understanding of the *Drosophila* type II neuron differentiation process. After initially separating the transcriptomes of the type II neuroblast-derived cells from those labeled in the optic lobes, we show that pseudotime analysis techniques can be used to define a maturation axis and extract putative marker genes that specify the INP, GMC, immature neuron, and mature neuron differentiation stages. Broadly expressed, not limited to the type II NB progenies, these marker genes of different maturation stages indeed form intersectional patterns that represent the spatial organization of the neurogenesis progress in the larval brain. Compared to previous antibody-based and gene-manipulation-based screening strategies, scRNA-seq data permit a high-throughput assessment of the whole gene-expression profile to rapidly identify candidate genes for functional study. For instance, in the past, *Hey* has been shown to mark one of the two immature neurons derived from the final cell division, and its role is exclusive as an inhibitor of *Notch* signaling in this immature neuron (Monastirioti et al., 2010). From our scRNA-seq analysis, *E(spl)m6-BFM*, a member of the enhancer-of-split family of transcription factors (Lai et al., 2000), and *Rbp*, a rim-binding protein responsible for synaptic homeostasis and neurotransmitter release (Liu et al., 2011; Müller et al., 2015) are exclusively upregulated in *only* the transient immature neuronal differentiation state directly after GMC division. These two marker genes can be used to guide the exploration of Hey^{-} immature neurons in future studies. Functional knockouts of these two genes will be critical to understanding their function in newly born neurons as it pertains to their maturation and any early functional role they may play in the developing brain.

Further higher-resolution clustering of the INP and GMC cells identified transcriptomically correlated subclusters between these two stages, which supports the idea that parallel maturation transitions happen at the same developmental time point. However, scRNA-seq data alone cannot distinguish whether these parallel transitions are due to the co-existence of earlier and newly born INPs in all NB lineages or due to the intrinsic differences among NB lineages. We therefore *in situ* profiled the marker genes selected from the scRNA-seq selected candidates and restored their missing spatial information that indicates the maturation stage as well as the NB lineage identity. In addition, combined with prior knowledge, whether a marker gene is expressed in younger or earlier born INPs can also be speculated. Our findings conclude that *Sp1* is expressed in the young INPs of nearly all NB lineages, whereas *TfAP-2* and *Fas3* express in older INPs belonging to specific NB lineages. Interestingly, we found that *Sp1* and *TfAP-2* expressed not only in neural progenitors but also in maturing neurons. These transcription factors seem to intermingle with the NB lineage-specific *D/grh/ey* cascades in the INP stage but eventually differentiate into completely exclusive neuron populations. Finally, higher-resolution clustering of neurons in our scRNA-seq dataset revealed that transcription factors and surface molecules are predominant markers for distinct neuronal subtypes at the third instar larval stage. This implies that most neurons of the type II NB progenies have not started to gain their differentiated functions at this stage of development.

Combining *in silico* scRNA-seq analysis and *in situ* mRNA imaging, we discovered many transcription factors and surface molecules that potentially play important roles in generating neuronal subtypes in an NB-specific, INP-specific, or function-specific manner. These discoveries helped us to gain a comprehensive understanding of the molecular landscape along all three major neural developmental axes that define a cell's progenitor lineage identity, progenitor cell division number, and differentiation state (Figure 6). This model provides a general guidance for biologists to disentangle the differentiation process in complex systems beyond the *Drosophila* brain.

Challenges and opportunities

We sequenced approximately 4,000 cells that were neurons originating from 8 *Drosophila* type II neuroblast lineages (16, if we assume no symmetry across the two central brain lobes). With low-resolution clustering, we identified 13 molecularly distinct neural subtypes. Increasing the clustering resolution just a bit higher we could identify more than 20 that are still distinct (data not shown). Similarly, as we show with the INPs/GMCs in our dataset, a low-resolution clustering can often mask the cellular diversity that is present in the system. As we know that each type II neuroblast generates approximately 38 INPs throughout their developmental lifespan (Bayraktar et al., 2010; Bello et al., 2008), the presented clustering in this paper only captures *part* of the INP diversity. One straightforward thought is to increase the number of sequenced single cells so that higher clustering resolution may eventually reveal even the most subtle differences between each of the hundreds of INPs in the type II system. However, as transcription factor cascades involved in INP division/maturation intertwine with those involved in NB specification and differentiation, we expect that the INP heterogeneity can be untangled somewhat using a higher clustering resolution but still fails to provide us with a coherent view of the complex lineage, maturation, and differentiation landscape we are attempting to characterize. These issues highlight the challenge of deconvoluting the INP maturation, NB lineage, and differentiation state axes and the need for a holistic, integrated approach to experimental design and subsequent bioinformatic analysis.

The data we have presented here were collected at a single developmental time point (late third instar), but we know that type II neurogenesis precedes and continues after this stage. Repeating these scRNA-seq experiments at more developmental time points will reveal more in what order molecularly defined neural subsets are generated. Using recently developed analytical techniques to "stitch" these multi-time-point datasets together (Lin et al., 2019; Tran and Bader, 2019) will be advantageous to align all the cells along a unified developmental time axis. To overcome the limitation of the R9D11-Gal4 driver, which does not label neuroblasts nor the fully mature neurons, a *permanent* labeling strategy, similar to the one used in Bayraktar et al. (2010) but covering all lineages more reliably for FACS, is required. More critically, such permanent labeling needs to be paired with technologies that provide single-lineage specification resolution, such as the introduction of single-neuroblast lineage barcoding techniques. Genetic constructs based around CRISPR-Cas9 (Raj et al., 2017; Spanjaard et al., 2018) and the Cre/Lox system (Kalhor et al., 2018; Pei et al., 2017; Weber et al., 2016) have been developed for this purpose, although which exact lineage was labeled by a particular barcode was still unknown. The introduction of a spectrally unique

barcode for each neuroblast lineage, in a similar vein to the recently developed *Bitbow* lineage tracking strategy (Li et al., 2020; Veling et al., 2019), would be advantageous as they can provide direct *in situ* evidence for neuroblast lineage identity.

Finally, our work identifies several transcription factors that are specifically expressed in subsets of cells of the type II neuroblast progenies. Our *in silico* and *in situ* results showed that their expressions are either constrained to particular developmental stages or in subsets of cells that are born in different orders. It would be desired to perform follow-up experiments to reveal whether these transcription factors play important roles in specifying the terminal fates of type II neuronal subtypes.

STAR★METHODS

RESOURCE AVAILABILITY

Lead contact—Further information and requests for resources should be directed to the lead contact: Dawen Cai (dwcai@umich.edu).

Material availability—Fly lines generated in this study include the [;;UAS-hH2B::2xmNG] and [;;UAS-hH2B::2xtagBFP] lines which will be deposited to the Indiana Bloomington *Drosophila* Stock Center.

Plasmids generated in this study include the pMUH-20xUAS-hH2B::2xmNG/2xtagBFP-p10pA plasmids used to generate the aforementioned fly lines and will be deposited to the Addgene plasmid repository.

HCR probes used in this study were designed and synthesized by Molecular Instruments (Los Angeles, CA, USA) and their exact sequences are the intellectual property of the aforementioned company. The lot numbers of the HCR probes are provided in the Key Resources Table and can be requested from Molecular Instruments.

Data and code availability—Sequencing data generated in this study is available from the Gene Expression Omnibus (https://www.ncbi.nlm.nih.gov/geo/) under accession ID GSE153723.

Jupyter notebooks used for scRNA-seq analysis are available upon request. The source code for the MiCV web tool is available at https://github.com/Cai-Lab-at-University-of-Michigan/MiCV. A web server with preloaded datasets including the one reported here is available at https://micv.works.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Flies were reared at 25°C on standard CT medium with a 12h/12h light/dark cycle. For FACS selection of type-II derived cells for scRNA-seq, [;;R9D11-Gal4] (BD40731) virgin female flies were crossed to male [;;UAS-hH2B::2xmNG] (this study) flies in vials prepared with fresh yeast paste to promote mating. F1 progeny were collected at approximately the late 3rd instar stage, as larvae are crawling up the vial walls to prepare for pupation. No selection was made based on larval sex.

For IHC and HCR experiments, [;;R9D11-CD4::tdTom] (BD35847) virgin female flies were crossed to male flies of the following genotypes: [;;Sp1::EGFP] (BD38669), in vials prepared with fresh yeast paste to promote mating. Alternatively [;;R9D11-Gal4] (BD35847) virgin female flies were crossed to male [;;UAS-hH2B::2xtagBFP] (this study) flies in a similar manner. F1 progeny were collected at approximately the late 3rd instar stage, as larvae are crawling up the vial walls to prepare for pupation. No selection was made based on larval sex.

METHOD DETAILS

Dissociation and FACS selection of type-II derived cells—[;;R9D11-Gal4/UAS-hH2B::2xmNG] larvae (n = 20) were rinsed and their brains dissected using dissection scissors and forceps at the late L3 stage (wandering larvae) in ice cold Rinaldini's solution. These brains were subsequently transferred to a poly-L-lysine coated coverslip that was immersed in Rinaldini's solution, attaching only the VNC to coverslip and leaving the central brain lobes unattached. These brain lobes were then further dissected using a tungsten needle by inserting the needle into each brain lobe at approximately the midpoint of the lobe and moving the needle laterally. This process removed a lot but not all the cells on the lateral portions of each brain lobe, which includes the developing optic lobe. The remaining OL cells were later excluded from our final scRNA-seq dataset using known marker genes (detailed above).

Dissected brains were transferred to a DNA low-binding 1.5mL tube in 30μL of dissection liquid (Rinaldini's solution) using a p200 pipette equipped with a siliconized p200 tip that was cut and flame-smoothed approximately 1/4 of the way up the tip. The siliconized tips are lower-binding and make it less likely for brains to stick to them. Cutting the tip and smoothing the opening makes it easier for the brains to move into the tip. The 1.5mL tube was pre-filled with 50μL of fresh, cold Rinaldini's solution, and upon transfer of the brains, 10μL of 20mg/mL papain, 10μL of 20mg/mL type-I collagenase, and 1μL of 15μM ZnCl were added to the tube, bringing the total reaction volume to 100μL. The tube was closed and mixed gently by flicking, then incubated on a heat block at 37°C for 1hr. During this incubation, the tube was flicked for mixing at 10min intervals, flicking the tube until the brains are visibly disturbed into the solution.

After the 1hr incubation, 2μL of 100μM E-64 solution was added to the mixture to stop the papain digestion. To break down the apparent intact brains, the mixture was triturated at a ~1 Hz frequency for 30 times using a p100 pipette set to 70μL and equipped with an uncut p200 siliconized tip. After the first 5 triturations, the brains should be seen largely dissociated to the naked eye. Further triturations break down the brain completely into single-cell suspensions including the VNC, which is quite resilient to dissociation.

After trituration, the cell suspension was diluted with 400μL Schneider's media + 10% FBS which further quenches the enzymatic digestion and stabilizes the cells. 1 μL of DRAQ5 DNA stain (Thermo Fisher Scientific Inc.) was added to label cells apart from debris generated in the dissociation process.

The sorting-ready cell suspension was transferred to a 5mL plastic FACS snap-cap tube on ice. Cells from non-Gal4 driver brains were dissociated in a similar manner and were sorted first on a Sony MA900 FACS machine to set the gate for using DRAQ5 to separate DNA containing cells from debris and set the gate for non-mNG expressing cells.

Sorted cells were captured in a DNA low-binding 1.5mL tube pre-filled with 100μL of Schneider's media + 10% FBS. Cells were spun down at 400x g for 4 minutes and the solution volume was reduced to 40μL before resuspending by gentle pipetting with a p200 siliconized pipette tip. 5μL of this suspension was removed to count cells using an epifluorescence microscope by plating them in a single well of a 96 well plate, pre-filled with 45μL of Schneider's media + 10% FBS. The rest of the cells were transported on ice to the University of Michigan Advanced Genomics Core and approximately 10,000 cells were loaded for 10X Chromium V3 sequencing following the manufacturer's instruction.

HCR *in situ* mRNA staining of L3 larval brains—We adapted with only minor changes from protocols described in the original third generation HCR paper (Choi et al., 2018). In brief, late-stage third instar larvae were dissected in room-temperature (RT) PBS as previously described and transferred to a 500μL tube containing PBS on ice. Brains were washed once in PBS for 1min standing, then washed in 4% RNase-free PFA at RT with 0.5% Tween-20, nutating for 20min. Brains were then washed twice with RNase-free 0.5% PBSTween for 20min each, nutating. Brains were then washed with 200μL of HCR amplification buffer at 37°C, nutating for 1hr. HCR probes (Molecular Instruments) were added to a final concentration of 5nM, and the sample was incubated at 37°C overnight, nutating. After this incubation, brains were washed 2x in HCR washing buffer at RT for 30min each, nutating. Brains were then incubated in 200μL amplification buffer at RT for 30min, nutating. 2.5uL of each imager hairpin (with attached dyes) was independently raised to a temperature of 95°C for 90sec in a thermocycler then snap-cooled to 4C immediately. 2μL of each hairpin was then added to the brains and incubated overnight at RT, nutating. Finally, brains were washed 2x with 2X SSCT at RT for 30min each, nutating, then once again with 2X SSC at RT for at least 10 minutes, nutating. Brains were subsequently mounted on a coverslip coated with poly-L-lysine that is submerged with Prolong Diamond mounting media (Thermo Fisher Scientific Inc) for imaging.

HCR probe design—Sequences provided to Molecular Instruments for HCR probe design were constructed by identifying the largest contiguous sequence present across all unique transcripts for each of our mRNAs of interest. As the information on the relative expression of individual isoforms of each transcript is in general not readily available, this provided for the highest possible detection probability at the expense of transcript-isoform specificity.

IHC staining of L3 larval brains—Brains were dissected in PBS and fixed in 4%PFA + 0.5% Triton X-100 (Triton) at RT for 20min, nutating. Brains were rinsed 2x with PBST (0.5% Triton X-100) at RT, then washed 1x with 0.5% PBSTriton at RT for 30min, nutating. For primary antibody staining, brains were incubated in Starting Block + 0.5% Triton at RT for 30min. Antibodies were then added and the brains were incubated at 4C overnight, nutating. Brains were then washed as described above followed by incubating in Starting Block + 0.5% Triton at RT for 30min. Secondary antibodies were added and brains were

incubated at RT for 2hr, protected from light. Brains were finally rinsed 2X in PBST (0.5% Triton X-100) at RT for 1min each, then washed 2X in PBS for 30min. Brains were subsequently mounted as described in the HCR section above.

Antibodies were diluted as the following: Mouse-anti-Fas3 1:50 (DHSB), Rat-anti-dpn (1:1000) (C-YL lab), Donkey-anti-Mouse (AF488) 1:500 (Jackson ImmunoResearch Laboratories, Inc.), Donkey-anti-Rat (AF647) 1:500 (Jackson ImmunoResearch Laboratories, Inc.).

scRNA-sequencing—Two replicate experiments were performed: one on a 10X Chromium v2 chip, and one on a 10X Chromium v3 chip. Input cell counts were approximately equal across replicates.

Approximately 10,000 type-II derived cells were used as input to a single channel of a 10X Chromium chip. The mRNA was subsequently reverse transcribed, amplified, and prepared for sequencing on an Illumina NovaSeq-6000 chip (University of Michigan Advanced Genomics Core). The library was sequenced for a total of 385M paired-end reads with 28bp for the cell barcode and UMI and 110bp for cDNA inserts.

QUANTIFICATION AND STATISTICAL ANALYSIS

scRNA-seq mapping and downstream analysis—Reads were mapped using both Cell Ranger (for initial analysis) and STAR-solo (for our final analysis, with mNG added to the genome) (Dobin et al., 2013) to the *Drosophila* genome assembly provided by ENSEMBL, build BDGP6 (2014-07).

The downstream scRNA-seq analysis was performed using scanpy (Wolf et al., 2018), and our analysis was formalized into the MiCV web tool generated in this work (https://micv.works). In brief, cells were filtered by requiring between 200-4100 unique genes/cell (to exclude debris and some doublets) and genes were filtered by requiring at least 2 cells to express it at greater than 1 UMI/cell. UMI counts were normalized to a total sum of 1e6 counts/cell (conversion to counts-per-million/CPM) and subsequently log-transformed by calculating ln(1+CPM) for each gene for each cell. The top 2000 highly variable genes were identified using the cell-ranger method (Zheng et al., 2017) and these genes were used to perform a principal component analysis (PCA, n = 50pcs). As two replicate experiments (batches) needed to be integrated across different sequencing chemistries (10X v2 and v3), the *harmony* (Korsunsky et al., 2019) data integration algorithm was used to batch-correct this PCA representation of the data before proceeding to neighborhood identification (k = 20), and finally a UMAP projection (2D). Clusters were identified using the Leiden algorithm (Traag et al., 2019), an optimized version of the Louvain algorithm (Blondel et al., 2008), with varying clustering resolutions. For most of the type-II only UMAP projections displayed in this work, the clustering resolution was 0.6, with 1 being a standard default (and higher numbers leading to more granular clustering of cells). Marker genes were identified using logistic regression analysis, implemented in scanpy.

scRNA-seq pseudotime trajectory inference—Pseudotime trajectories were generated using palantir (Setty et al., 2019). A starting cell for the trajectory (ID:

TCATGTTGTTCTGACA) was identified using a high expression of CycE and D, and two terminal branch cells (IDs: AATCGACGTAATCAGA and AAGCGTTTCCTATTGT) were identified by choosing cells at the maximal points of the two major neural branches in the UMAP projections of the type-II cells. The choice of terminal cells was not necessary for the automatic identification of these 2 branches by palantir. They are provided here for data reproducibility purposes. Default parameters were used throughout the rest of the pseudotime trajectory inference.

scRNA-seq pseudotime gene expression trend fitting—Pseudotime gene expression trends were generated using gene expression data after imputation using MAGIC (van Dijket al., 2018) as recommended by the palantir documentation, with a step size of 1 (meaning data was imputed only using very similar/nearby cells). Imputed data were clipped so that 0 was the minimum value for imputed gene expression. PyGAM's PoissonGAM class was used to generate trends for each gene, with trends being built up by 5 splines of order 3.

scRNA-seq co-expression quantification in INPs—INPs were considered to express a specific gene if the following criterion was satisfied: ln(CPM+1) > 4.5. INPs co-expressed two genes if both genes simultaneously met the above criterion.

scRNA-seq global pairwise gene-gene correlation calculation—All genes across our entire type-II scRNA-seq dataset were paired together and their Spearman rank-order correlation coefficients were calculated using scipy, with the log-normalized expression values for each cell being passed in as 1D vectors for each gene. These gene pairs were then filtered to only include pairs with a correlation coefficient above 0.6 (correlated) or below −0.6 (anti-correlated) and their respective gene-gene pairs were output to a table. Gene group membership information from Flybase for each gene in the pair was added to these tables if it was available and was omitted otherwise.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

We thank all members of the Cai lab who contributed to the discussion and revision of the manuscript. All authors acknowledge support from the University of Michigan Flow Cytometry Core and its National Institutes of Health (NIH) support, NIH 5P30CA046592-31, as well as the University of Michigan Advanced Genomics Core. N.S.M. acknowledges support from NIH 1T32EB005582. F.Y.S. acknowledges the support from NIH 1F31NS11184701. D.C. acknowledges support from the University of Michigan (CDB IDEA Awards in Stem Cell Biology, MCubed2.0) and Michigan Economic Development Corporation (Mi-TRAC). C.-Y.L. acknowledges support from the NIH 1R01NS107496.

REFERENCES

Álvarez J-A, and Díaz-Benjumea FJ (2018). Origin and specification of type II neuroblasts in the *Drosophila* embryo. Development 145, dev158394. [PubMed: 29567672]

Bayraktar OA, and Doe CQ (2013). Combinatorial temporal patterning in progenitors expands neural diversity. Nature 498, 449–455. [PubMed: 23783519]

Bayraktar OA, Boone JQ, Drummond ML, and Doe CQ (2010). Drosophila type II neuroblast lineages keep Prospero levels low to generate large clones that contribute to the adult brain central complex. Neural Dev. 5, 26. [PubMed: 20920301]

Bello BC, Izergina N, Caussinus E, and Reichert H (2008). Amplification of neural stem cell proliferation by intermediate progenitor cells in Drosophila brain development. Neural Dev. 3, 5. [PubMed: 18284664]

Blondel VD, Guillaume J-L, Lambiotte R, and Lefebvre E (2008). Fast unfolding of communities in large networks. J. Stat. Mech 2008, P10008.

Boone JQ, and Doe CQ (2008). Identification of Drosophila type II neuroblast lineages containing transit amplifying ganglion mother cells. Dev. Neurobiol 68, 1185–1195. [PubMed: 18548484]

Butler A, Hoffman P, Smibert P, Papalexi E, and Satija R (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat. Biotechnol 36, 411–420. [PubMed: 29608179]

Cao J, Spielmann M, Qiu X, Huang X, Ibrahim DM, Hill AJ, Zhang F, Mundlos S, Christiansen L, Steemers FJ, et al. (2019). The single-cell transcriptional landscape of mammalian organogenesis. Nature 566, 496–502. [PubMed: 30787437]

Choi HMT, Schwarzkopf M, Fornace ME, Acharya A, Artavanis G, Stegmaier J, Cunha A, and Pierce NA (2018). Third-generation *in situ* hybridization chain reaction: multiplexed, quantitative, sensitive, versatile, robust. Development 145, dev165753. [PubMed: 29945988]

Cocanougher BT, Wittenbach JD, Long X, Kohn AB, Norekian TP, Yan J, Colonell J, Masson J-B, Truman JW, Cardona A, et al. (2019). Comparative single-cell transcriptomics of complete insect nervous systems. bioRxiv. 10.1101/785931.

de Nooij JC, Letendre MA, and Hariharan IK (1996). A cyclin-dependent kinase inhibitor, Dacapo, is necessary for timely exit from the cell cycle during Drosophila embryogenesis. Cell 87, 1237–1247. [PubMed: 8980230]

Deitcher DL, Ueda A, Stewart BA, Burgess RW, Kidokoro Y, and Schwarz TL (1998). Distinct requirements for evoked and spontaneous release of neurotransmitter are revealed by mutations in the Drosophila gene neuronal-synaptobrevin. J. Neurosci 18, 2028–2039. [PubMed: 9482790]

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, and Gingeras TR (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29, 15–21. [PubMed: 23104886]

Flybase (2019). Computation of D. melanogaster genes relevant to disease based on their orthology to human “disease genes. https://flybase.org/reports/FBrf0241599.html.

Habib N, Basu A, Avraham-Davidi I, Burks T, Choudhury SR, Aguet F, Gelfand E, Ardlie K, Weitz DA, Rozenblatt-Rosen O, et al. (2017). DroNc-Seq: Deciphering cell types in human archived brain tissues by massively-parallel single nucleus RNA-seq. bioRxiv. 10.1101/115196.

Han X, Wang R, Zhou Y, Fei L, Sun H, Lai S, Saadatpour A, Zhou Z, Chen H, Ye F, et al. (2018). Mapping the Mouse Cell Atlas by Microwell-Seq. Cell 172, 1091–1107.e17. [PubMed: 29474909]

Homem CCF, and Knoblich JA (2012). Drosophila neuroblasts: a model for stem cell biology. Development 139, 4297–4310. [PubMed: 23132240]

Huang W, Sherman BT, and Lempicki RA (2009a). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res. 37, 1–13. [PubMed: 19033363]

Huang W, Sherman BT, and Lempicki RA (2009b). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc 4, 44–57. [PubMed: 19131956]

Kalhor R, Kalhor K, Mejia L, Leeper K, Graveline A, Mali P, and Church GM (2018). Developmental barcoding of whole mouse via homing CRISPR. Science 361, eaat9804. [PubMed: 30093604]

Kiselev VY, Andrews TS, and Hemberg M (2019). Challenges in unsupervised clustering of single-cell RNA-seq data. Nat. Rev. Genet 20, 273–282. [PubMed: 30617341]

Korsunsky I, Millard N, Fan J, Slowikowski K, Zhang F, Wei K, Baglaenko Y, Brenner M, Loh P-R, and Raychaudhuri S (2019). Fast, sensitive and accurate integration of single-cell data with Harmony. Nat. Methods 16, 1289–1296. [PubMed: 31740819]

Kose H, Rose D, Zhu X, and Chiba A (1997). Homophilic synaptic target recognition mediated by immunoglobulin-like cell adhesion molecule Fasciclin III. Development 124, 4143–4152. [PubMed: 9374410]

Kucherenko MM, Ilangovan V, Herzig B, Shcherbata HR, and Bringmann H (2016). TfAP-2 is required for night sleep in Drosophila. BMC Neurosci. 17, 72. [PubMed: 27829368]

Lai EC, Bodner R, and Posakony JW (2000). The enhancer of split complex of Drosophila includes four Notch-regulated members of the bearded gene family. Development 127, 3441–3455. [PubMed: 10903170]

Landskron L, Steinmann V, Bonnay F, Burkard TR, Steinmann J, Reichardt I, Harzer H, Laurenson A-S, Reichert H, and Knoblich JA (2018). The asymmetrically segregating lncRNA cherub is required for transforming stem cells into malignant cells. eLife 7, e31347. [PubMed: 29580384]

Lane ME, Sauer K, Wallace K, Jan YN, Lehner CF, and Vaessin H (1996). Dacapo, a cyclin-dependent kinase inhibitor, stops cell proliferation during Drosophila development. Cell 87, 1225–1235. [PubMed: 8980229]

Lee C-Y, Robinson KJ, and Doe CQ (2006). Lgl, Pins and aPKC regulate neuroblast self-renewal versus differentiation. Nature 439, 594–598. [PubMed: 16357871]

Levine JH, Simonds EF, Bendall SC, Davis KL, Amir AD, Tadmor MD, Litvin O, Fienberg HG, Jager A, Zunder ER, et al. (2015). Data-Driven Phenotypic Dissection of AML Reveals Progenitor-like Cells that Correlate with Prognosis. Cell 162, 184–197. [PubMed: 26095251]

Li X, Chen R, and Zhu S (2017). bHLH-O proteins balance the self-renewal and differentiation of Drosophila neural stem cells by regulating Earmuff expression. Dev. Biol 431, 239–251. [PubMed: 28899667]

Li Y, Walker LA, Zhao Y, Edwards EM, Michki NS, Cheng HPJ, Ghazzi M, Chen TY, Chen M, Roossien DH, et al. (2020). Bitbow: a digital format of Brainbow enables highly efficient neuronal lineage tracing and morphology reconstruction in single brains. bioRxiv. 10.1101/2020.04.07.030593.

Lin Y, Ghazanfar S, Wang KYX, Gagnon-Bartsch JA, Lo KK, Su X, Han Z-G, Ormerod JT, Speed TP, Yang P, and Yang JYH (2019). scMerge leverages factor analysis, stable expression, and pseudoreplication to merge multiple single-cell RNA-seq datasets. Proc. Natl. Acad. Sci. USA 116, 9775–9784. [PubMed: 31028141]

Liu KSY, Siebert M, Mertel S, Knoche E, Wegener S, Wichmann C, Matkovic T, Muhammad K, Depner H, Mettke C, et al. (2011). RIM-binding protein, a central part of the active zone, is essential for neurotransmitter release. Science 334, 1565–1569. [PubMed: 22174254]

Liu L-Y, Long X, Yang C-P, Miyares RL, Sugino K, Singer RH, and Lee T (2019). Mamo decodes hierarchical temporal gradients into terminal neuronal fate. eLife 8, e48056. [PubMed: 31545163]

Macosko EZ, Basu A, Satija R, Nemesh J, Shekhar K, Goldman M, Tirosh I, Bialas AR, Kamitaki N, Martersteck EM, et al. (2015). Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. Cell 161, 1202–1214. [PubMed: 26000488]

Maier D, Marte BM, Schäfer W, Yu Y, and Preiss A (1993). Drosophila evolution challenges postulated redundancy in the E(spl) gene complex. Proc. Natl. Acad. Sci. USA 90, 5464–5468. [PubMed: 8516287]

Mariano V, Achsel T, Bagni C, and Kanellopoulos AK (2020). Modelling learning and memory in Drosophila to understand Intellectual Disabilities. Neuroscience 445, 12–30. [PubMed: 32730949]

Monastirioti M, Giagtzoglou N, Koumbanakis KA, 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. [PubMed: 20040486]

Monge I, Krishnamurthy R, Sims D, Hirth F, Spengler M, Kammermeier L, Reichert H, and Mitchell PJ (2001). Drosophila transcription factor AP-2 in proboscis, leg and brain central complex development. Development 128, 1239–1252. [PubMed: 11262226]

Müller M, Genç Ö, and Davis GW (2015). RIM-binding protein links synaptic homeostasis to the stabilization and replenishment of high release probability vesicles. Neuron 85, 1056–1069. [PubMed: 25704950]

Ntranos V, Yi L, Melsted P, and Pachter L (2018). Identification of transcriptional signatures for cell types from single-cell RNA-Seq. bioRxiv. 10.1101/258566.

Ogawa LM, and Vallender EJ (2014). Evolutionary conservation in genes underlying human psychiatric disorders. Front. Hum. Neurosci 8, 283. [PubMed: 24834046]

Pei W, Feyerabend TB, Rössler J, Wang X, Postrach D, Busch K, Rode I, Klapproth K, Dietlein N, Quedenau C, et al. (2017). Polylox barcoding reveals haematopoietic stem cell fates realized in vivo. Nature 548, 456–460. [PubMed: 28813413]

Pflugfelder GO, Schwarz H, Roth H, Poeck B, Sigl A, Kerscher S, Jonschker B, Pak WL, and Heisenberg M (1990). Genetic and molecular characterization of the optomotor-blind gene locus in Drosophila melanogaster. Genetics 126, 91–104. [PubMed: 2121596]

Ponti G, Obernier K, Guinto C, Jose L, Bonfanti L, and Alvarez-Buylla A (2013). Cell cycle and lineage progression of neural progenitors in the ventricular-subventricular zones of adult mice. Proc. Natl. Acad. Sci. USA 110, E1045–E1054. [PubMed: 23431204]

Qiu X, Mao Q, Tang Y, Wang L, Chawla R, Pliner HA, and Trapnell C (2017). Reversed graph embedding resolves complex single-cell trajectories. Nat. Methods 14, 979–982. [PubMed: 28825705]

Raj B, Wagner DE, McKenna A, Pandey S, Klein AM, Shendure J, Gagnon JA, and Schier AF (2017). Simultaneous single-cell profiling of lineages and cell types in the vertebrate brain by scGESTALT. bioRxiv. 10.1101/205534.

Ren Q, Yang C-P, Liu Z, Sugino K, Mok K, He Y, Ito M, Nern A, Otsuna H, and Lee T (2017). Stem Cell-Intrinsic, Seven-up-Triggered Temporal Factor Gradients Diversify Intermediate Neural Progenitors. Curr. Biol 27, 1303–1313. [PubMed: 28434858]

Ren J, Isakova A, Friedmann D, Zeng J, Grutzner SM, Pun A, Zhao GQ, Kolluru SS, Wang R, Lin R, et al. (2019). Single-cell transcriptomes and whole-brain projections of serotonin neurons in the mouse dorsal and median raphe nuclei. eLife 8, e49424. [PubMed: 31647409]

Saunders A, Macosko E, Wysoker A, Goldman M, Krienen F, Bien E, Baum M, Wang S, Goeva A, Nemesh J, et al. (2018). A Single-Cell Atlas of Cell Types, States, and Other Transcriptional Patterns from Nine Regions of the Adult Mouse Brain. bioRxiv. 10.1101/299081.

Schinaman JM, Giesey RL, Mizutani CM, 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, e1001964. [PubMed: 25291190]

Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B, et al. (2012). Fiji: an open-source platform for biological-image analysis. Nat. Methods 9, 676–682. [PubMed: 22743772]

Servén D, Brummitt C, and Abedi H (2018). pyGAM: Generalized Additive Models in Python. Zenodo.

Setty M, Kiseliovas V, Levine J, Gayoso A, Mazutis L, and Pe'er D (2019). Characterization of cell fate probabilities in single-cell data with Palantir. Nat. Biotechnol 37, 451–460. [PubMed: 30899105]

Snow PM, Bieber AJ, and Goodman CS (1989). Fasciclin III: a novel homophilic adhesion molecule in Drosophila. Cell 59, 313–323. [PubMed: 2509076]

Soldatov R, Kaucka M, Kastriti ME, Petersen J, Chontorotzea T, Englmaier L, Akkuratova N, Yang Y, Häring M, Dyachuk V, et al. (2019). Spatiotemporal structure of cell fate decisions in murine neural crest. Science 364, eaas9536. [PubMed: 31171666]

Spanjaard B, Hu B, Mitic N, Olivares-Chauvet P, Janjuha S, Ninov N, and Junker JP (2018). Simultaneous lineage tracing and cell-type identification using CRISPR-Cas9-induced genetic scars. Nat. Biotechnol 36, 469–473. [PubMed: 29644996]

Syed MH, Mark B, and Doe CQ (2017). Steroid hormone induction of temporal gene expression in *Drosophila* brain neuroblasts generates neuronal and glial diversity. eLife 6, e26287. [PubMed: 28394252]

Traag VA, Waltman L, and van Eck NJ (2019). From Louvain to Leiden: guaranteeing well-connected communities. Sci. Rep 9, 5233. [PubMed: 30914743]

Tran TN, and Bader G(2019). Tempora: cell trajectory inference using time-series single-cell RNA sequencing data. bioRxiv. 10.1101/846907.

Tumbar T, Guasch G, Greco V, Blanpain C, Lowry WE, Rendl M, and Fuchs E (2004). Defining the epithelial stem cell niche in skin. Science 303, 359–363. [PubMed: 14671312]

Turek M, Lewandrowski I, and Bringmann H (2013). An AP2 transcription factor is required for a sleep-active neuron to induce sleep-like quiescence in C. elegans. Curr. Biol 23, 2215–2223. [PubMed: 24184105]

van Dijk D, Sharma R, Nainys J, Yim K, Kathail P, Carr AJ, Burdziak C, Moon KR, Chaffer CL, Pattabiraman D, et al. (2018). Recovering Gene Interactions from Single-Cell Data Using Data Diffusion. Cell 174, 716–729.e27. [PubMed: 29961576]

Veling MW, Li Y, Veling MT, Litts C, Michki N, Liu H, Ye B, and Cai D (2019). Identification of Neuronal Lineages in the Drosophila Peripheral Nervous System with a “Digital” Multi-spectral Lineage Tracing System. Cell Rep. 29, 3303–3312. [PubMed: 31801091]

Weber TS, Dukes M, Miles DC, Glaser SP, Naik SH, and Duffy KR (2016). Site-specific recombinatorics: in situ cellular barcoding with the Cre Lox system. BMC Syst. Biol 10, 43. [PubMed: 27363727]

Wells RE, Barry JD, Warrington SJ, Cuhlmann S, Evans P, Huber W, Strutt D, and Zeidler MP (2013). Control of tissue morphology by Fasciclin III-mediated intercellular adhesion. Development 140, 3858–3868. [PubMed: 23946443]

Weng M, Golden KL, and Lee C-Y (2010). dFezf/Earmuff maintains the restricted developmental potential of intermediate neural progenitors in Drosophila. Dev. Cell 18, 126–135. [PubMed: 20152183]

Wolf FA, Angerer P, and Theis FJ (2018). SCANPY: large-scale single-cell gene expression data analysis. Genome Biol. 19, 15. [PubMed: 29409532]

Yang L, Titlow J, Ennis D, Smith C, Mitchell J, Young FL, Waddell S, Ish-Horowicz D, and Davis I (2017). Single molecule fluorescence in situ hybridisation for quantitating post-transcriptional regulation in Drosophila brains. Methods 126, 166–176. [PubMed: 28651965]

Zheng GXY, Terry JM, Belgrader P, Ryvkin P, Bent ZW, Wilson R, Ziraldo SB, Wheeler TD, McDermott GP, Zhu J, et al. (2017). Massively parallel digital transcriptional profiling of single cells. Nat. Commun 8, 14049. [PubMed: 28091601]

Zhong S, Zhang S, Fan X, Wu Q, Yan L, Dong J, Zhang H, Li L, Sun L, Pan N, et al. (2018). A single-cell RNA-seq survey of the developmental landscape of the human prefrontal cortex. Nature 555, 524–528. [PubMed: 29539641]

Zhou Q, Liu M, Xia X, Gong T, Feng J, Liu W, Liu Y, Zhen B, Wang Y, Ding C, and Qin J (2017). A mouse tissue transcription factor atlas. Nat. Commun 8, 15089. [PubMed: 28429721]

Ziegenhain C, Vieth B, Parekh S, Reinius B, Guillaumet-Adkins A, Smets M, Leonhardt H, Heyn H, Hellmann I, and Enard W (2017). Comparative Analysis of Single-Cell RNA Sequencing Methods. Mol. Cell 65, 631–643. [PubMed: 28212749]

Highlights

- 4,035 type II cells from developing *Drosophila* brains are profiled using scRNA-seq
- Pseudotime analysis reveals genes that vary during neural differentiation
- Identified marker genes specify INPs in age- and lineage-restricted fashions
- Transcription factors and surface molecules mark developing neural progenies

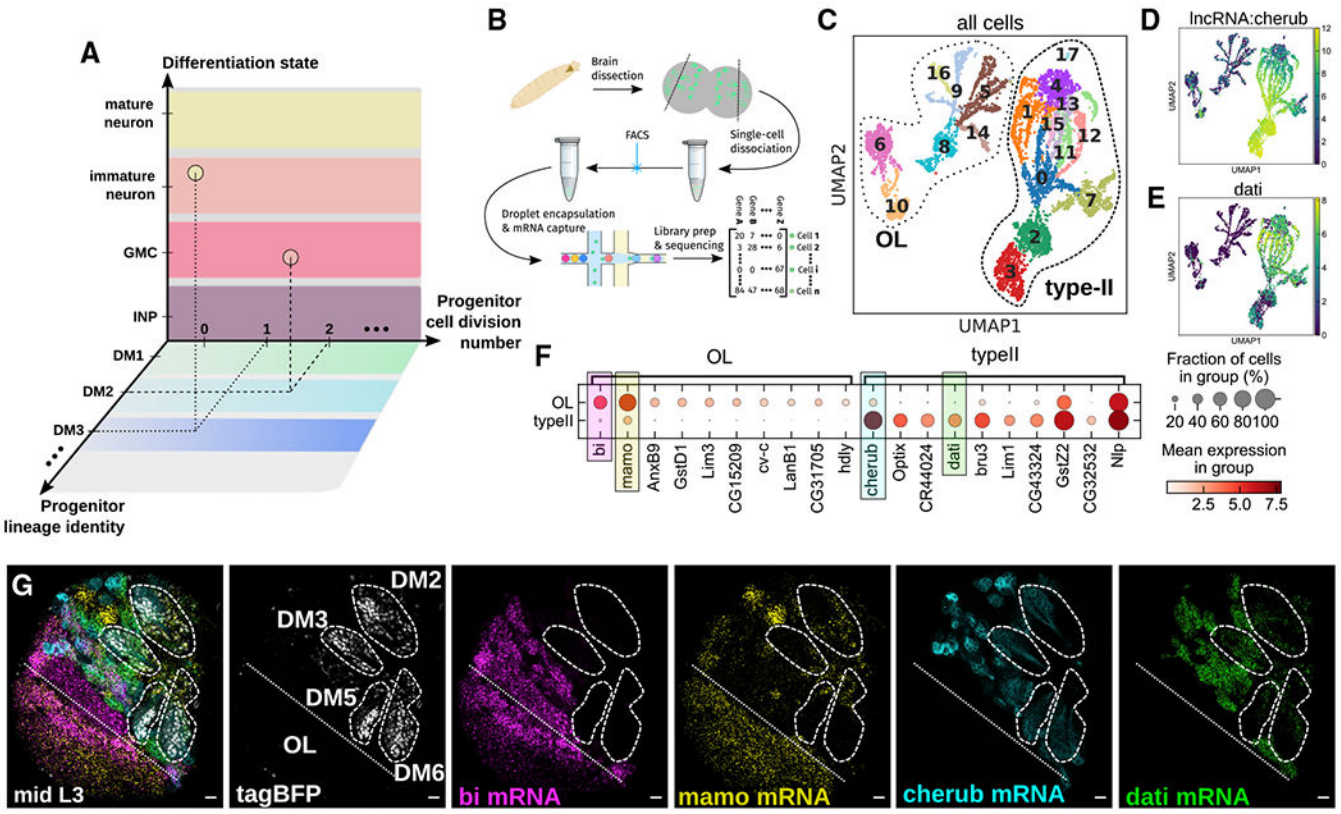


Figure 1. *Drosophila* type II neuronal fate-specification model, experiment overview, and *in silico* dissection of the optic lobe and type II-derived cells

(A) A diagram of the major axes that determine cell "state" in this work. Each sequenced cell is defined in part by factors that are specific to the lineage identity, intermediate progenitor cell division number, and differentiation state.

(B) Overview of our targeted scRNA-seq experimental strategy.

(C) Cells plotted in the first 2 dimensions of a UMAP projection. Color represents an automatic cluster assignment by the Leiden algorithm (resolution = 0.5).

(D and E) Expression of the long non-coding RNA *cherub* and the transcription factor *dati* are known to be exclusive of the optic lobe in third instar larvae. Groups of cells that lack expression of these genes are likely optic lobe cells that also express *Gal4* under the control of the R9D11 fragment of the *erm* promoter.

(F) Separating the putative type II/optic lobe cells into two groups and performing logistic regression analysis reveals genes that are upregulated between the two.

(G) A single z-slice of one brain lobe from the developing (mid L3 stage) larval brain. UAS-hH2B::2xtagBFP is driven under the control of R9D11-Gal4 and marks the type II lineages, only four of which are visible in this z-slice. *lncRNA:cherub* and *dati* mRNA are largely expressed by type II cells, while *bi* and *mamo* mRNA are largely expressed in the developing optic lobe (boundary marked by the diagonal line).

Scale bars: 10 μm in all images.

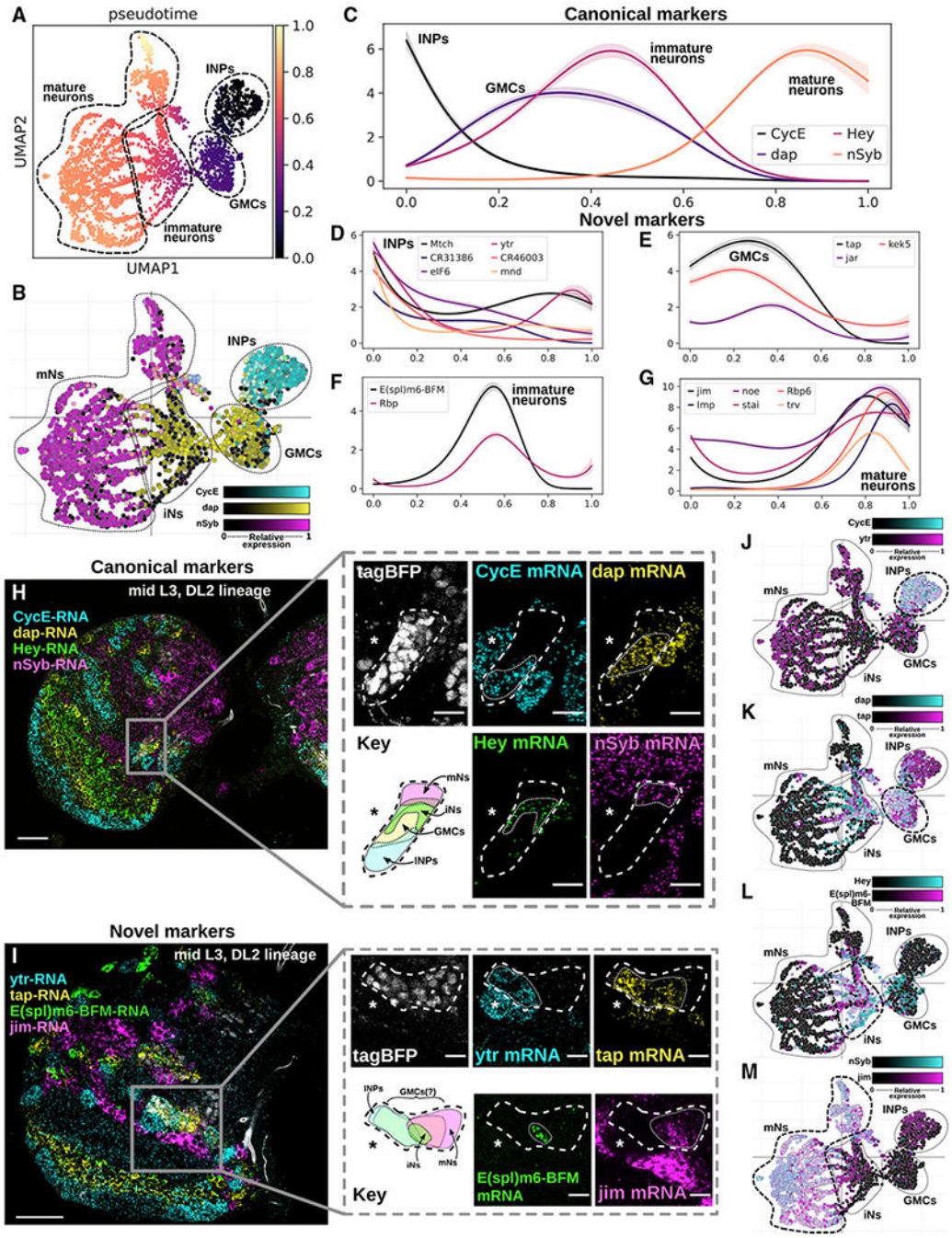


Figure 2. Pseudotime analysis reveals signature genes that vary along the cell-differentiation axis

(A) Pseudotime analysis establishes a global ordering of cells along the differentiation state axis.

(B) A multi-color UMAP expression plot generated by the MiCV web tool shows the expression of 3 canonical marker genes for the INP, GMC, and mature neuron states.

(C) The pseudo-temporal expression pattern of 4 genes that are known markers for the 4 major differentiation states.

(D–G) Pseudo-temporal expression patterns of groups of marker genes that do not have known functions associated with cellular differentiation state. These gene-expression trends are similar to the known marker genes plotted in (C).

(H and I) HCRv3 *in situ* mRNA staining images for both known (H) and novel (I) differentiation state marker genes in single z-slices of the DL2 lineage of mid third instar larval brains. UAS-hH2B::2xtagBFP is driven under the control of R9D11-Gal4 and marks the type II lineages. Asterisks denote the location of the putative type II NB. Thick dashed lines denote the boundaries of the tagBFP-labeled type II NB progenies. Thin dotted lines denote the boundaries of type II progeny cells expressing indicated mRNAs.

(J–M) Multi-color UMAP expression plots illustrate the expression pattern of the canonical and novel marker genes from (H) and (I), respectively.

Scale bars: 30 μm in overviews of (H) and (I) and 10 μm in insets of (H) and (I).

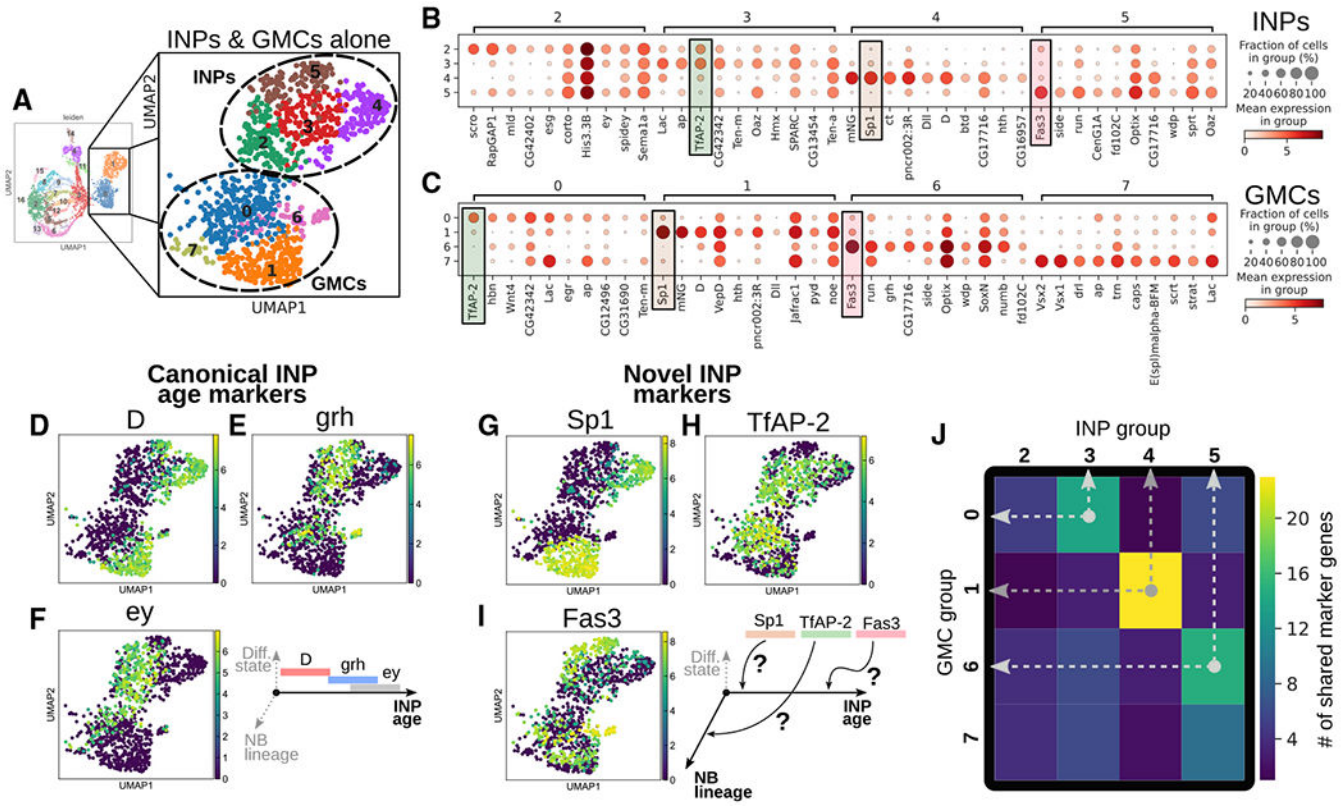


Figure 3. Sub-clustering of INPs and GMCs reveals transcription factors beyond the canonical *D-grh-ey* transition that vary along a combination of the NB lineage and INP division number patterning axes

(A) Left: Leiden clustering reveals INPs and GMCs to be in cluster 1 and 0, respectively. Right: higher resolution clustering on separated INPs and GMCs further divides them into 4 subclusters each.

(B and C) Marker gene analysis revealed that mostly transcription factors specific INP and GMC subclusters, respectively.

(D–F) Expression UMAP plots of the well-established temporally varying INP genes *D, grh*, and *ey*, respectively. *D* appears to separate cleanly at the mRNA level in the INPs of our dataset; however, *grh* and *ey* are broadly co-expressed.

(G–I) Expression UMAP plots of the INP/GMC cluster-specific genes *Sp1, TfAP-2*, and *Fas3*, which are found to correlate INP subclusters 3, 4, and 5 to GMC subclusters 0, 1, and 6, respectively.

(J) A correlation plot shows the number of top 100 marker genes that are shared between each INP and GMC subcluster. This simple similarity metric indicates a hypothesis that cells in GMC subclusters 0, 1, and 6 are the direct progenies of cells in INP subcluster 3, 4, and 5, respectively. INP group 2 and GMC group 7 are both clearly distinct from the other INP and GMC subtypes but share very few marker genes and so are unlikely to be related.

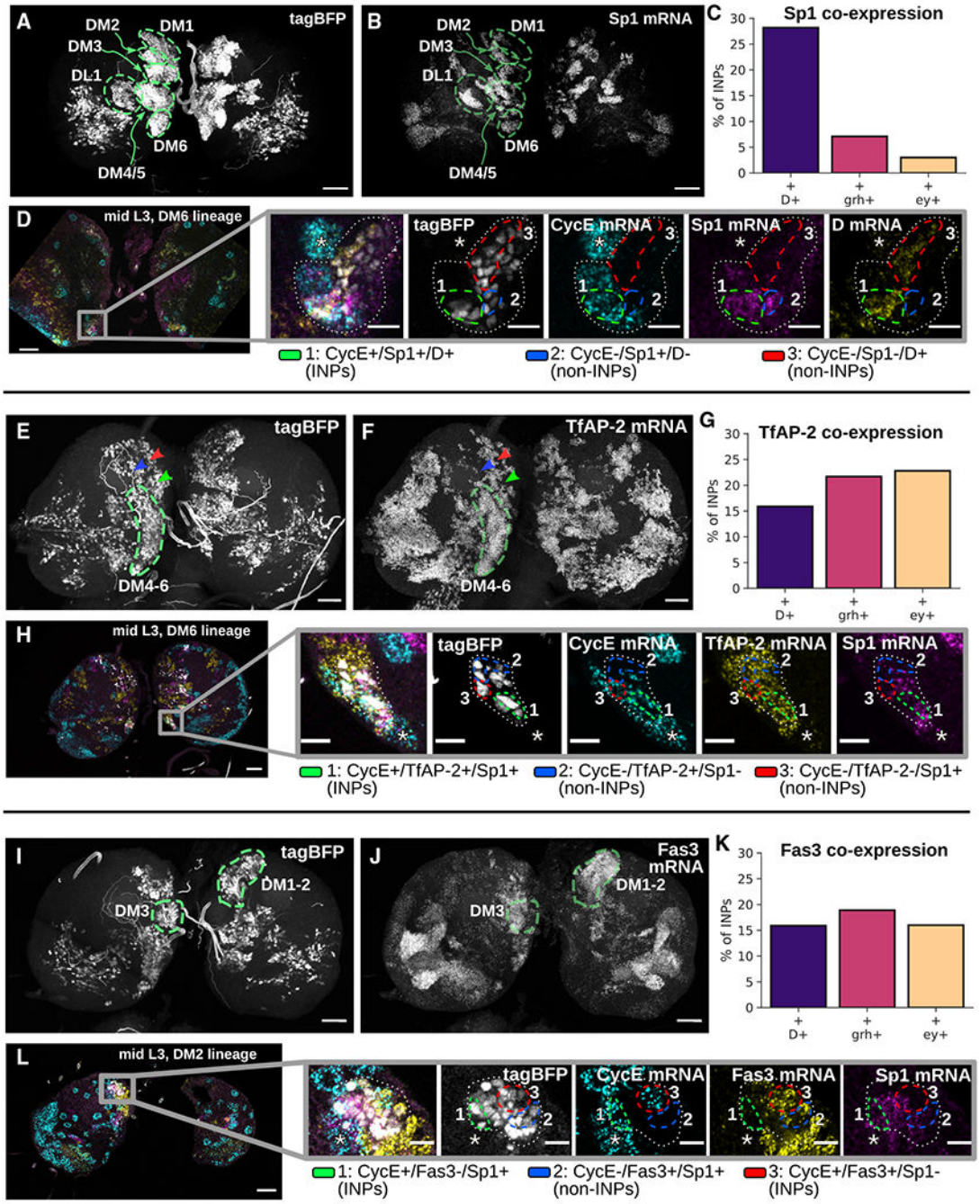


Figure 4. Sp1, TfAP-2, and Fas3 are each expressed by INPs of specific NB lineages

(A and B) Maximum Z-projections (45 μm thick) show tagBFP fluorescence and *Sp1* mRNA HCR staining in an L3 larval ;;R9D11-Gal4/UAS-H2B::tagBFP fly brain, respectively. Green dashed lines indicate the expression of *Sp1* mRNA in all type II NB-derived lineages except for DL2.

(C) *Co*-expression quantification of *Sp1* with *D, grh*, and *ey* in all INPs (n = 561).

(D) HCR staining showcases the expression patterns of *Sp1* and *D* mRNAs in lineage DM6. Dashed lines highlight region 1 of INPs that co-express *Sp1* and *D* mRNA, region 2 of non-

INP cells where *Sp1* mRNA alone is detected, and region 3 of non-INP cells where D mRNA alone is detected. White dotted lines denote the DM6 lineage boundary. Asterisks denote the position of the DM6 neuroblast.

(E and F) Maximum Z-projections (45 μm thick) as in (B) and (C), with *TfAP-2* mRNA HCR staining. Within the type II NB lineages, *TfAP-2* mRNA is highly expressed in cells belonging to DM 4–6 (dashed lines) and possibly DL1. Though some expression is seen nearby to DM1–3, *TfAP-2* is not expressed in tagBFP^{+} cells belonging to those lineages (arrowheads).

(G) Co-expression quantification of *TfAP-2* as in (C).

(H) HCR staining showcases the expression patterns of *CycE, Sp1*, and *TfAP-2* mRNAs in lineage DM6, where we can find *TfAP-2* expressed in *CycE*$^{+}$ INPs that have *Sp1* expression (green dashed lines) or not (red dashed lines), as well as in *CycE*$^{-}$ progeny cells (blue dashed lines).

(I and J) Maximum Z-projections (45 μm thick) as in (B) and (C), with *Fas3* mRNA HCR staining. Within the type II NB lineages, *Fas3* mRNA is highly expressed in cells belonging to DM1–3 (dashed lines).

(K) Co-expression quantification of *Fas3* as in (C).

(L) HCR staining showcases the expression patterns of *CycE, Fas3*, and *Sp1* mRNAs in lineage DM2. We find a clear expression of *Fas3* in both INPs and their progeny in NB lineage DM2, where we can find *Fas3* expressed in *CycE*$^{+}$ INPs that have *Sp1* expression (green dashed lines) or not (red dashed lines), as well as in *CycE*$^{-}$ progeny cells (blue dashed lines).

Scale bars: 30 μm in (A), (B), (E), (F), (I), and (J) and in the overviews of (D), (H), and (L); 10 μm in insets of (D), (H), and (L). Minimum expression threshold: ln(CPM^{+}1) >4.5 in (C), (G), and (K).

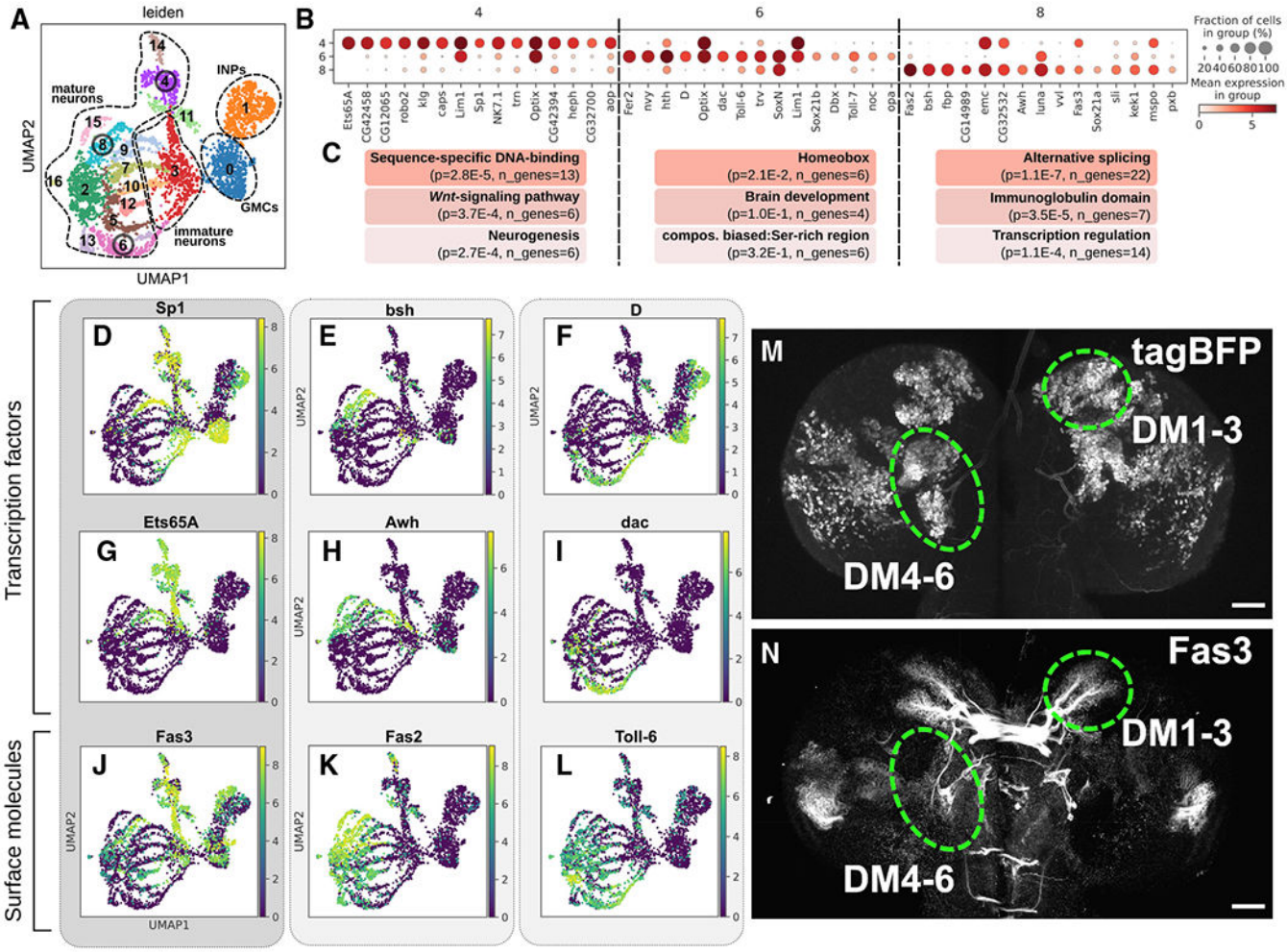


Figure 5. A unique combination of transcription factors and surface molecules define putative neural sub-progenies of young INPs

(A) Automatic Leiden clustering (resolution = 0.6) of the type II scRNA-seq data, with putative neural subtypes 4, 6, and 8 outlined, representing the *Sp1, bsh*, and *D*$^{+}$ neural progenies, respectively.

(B) Marker gene detection for the three selected neural subtypes showing the top 15 marker genes as identified using the t test_overestim_var function in scanpy.

(C) Top Gene Ontology (GO) functional annotations for the top 100 marker genes for cells in each of clusters 4, 6, and 8, respectively (p values are Benjamini corrected; n_genes refers to the number of marker genes annotated with the respective GO term).

(D–F) Log-fold expression values of *Sp1, bsh*, and *D*, respectively, showing three unique neural lineages are marked by these three transcription factors.

(G and J) Log-fold expression values of the transcription factor *Ets65A* and the cell-surface molecule *Fas3* that mark the *Sp1*$^{+}$ neural progeny.

(H and K) Log-fold expression of the transcription factor *Awh* and surface molecule *Fas2* that mark the *bsh*$^{+}$ neural progeny.

(I and L) Log-fold expression of the transcription factor *dac* and surface molecule *Toll-6* that mark the *D*$^{+}$ neural progeny.

(M and N) Maximum Z-projections show tagBFP fluorescence and *Fas3* antibody staining in an L3 larval ;;R9D11-Gal4/UAS-H2B::tagBFP reporter fly brain, respectively. It appears that neurons from DM1–3 that produce commissure-crossing axons are prominently labeled by *Fas3*, whereas neurons from DM4–6 are largely unstained. Scale bars: 30 μm in (M) and (N).

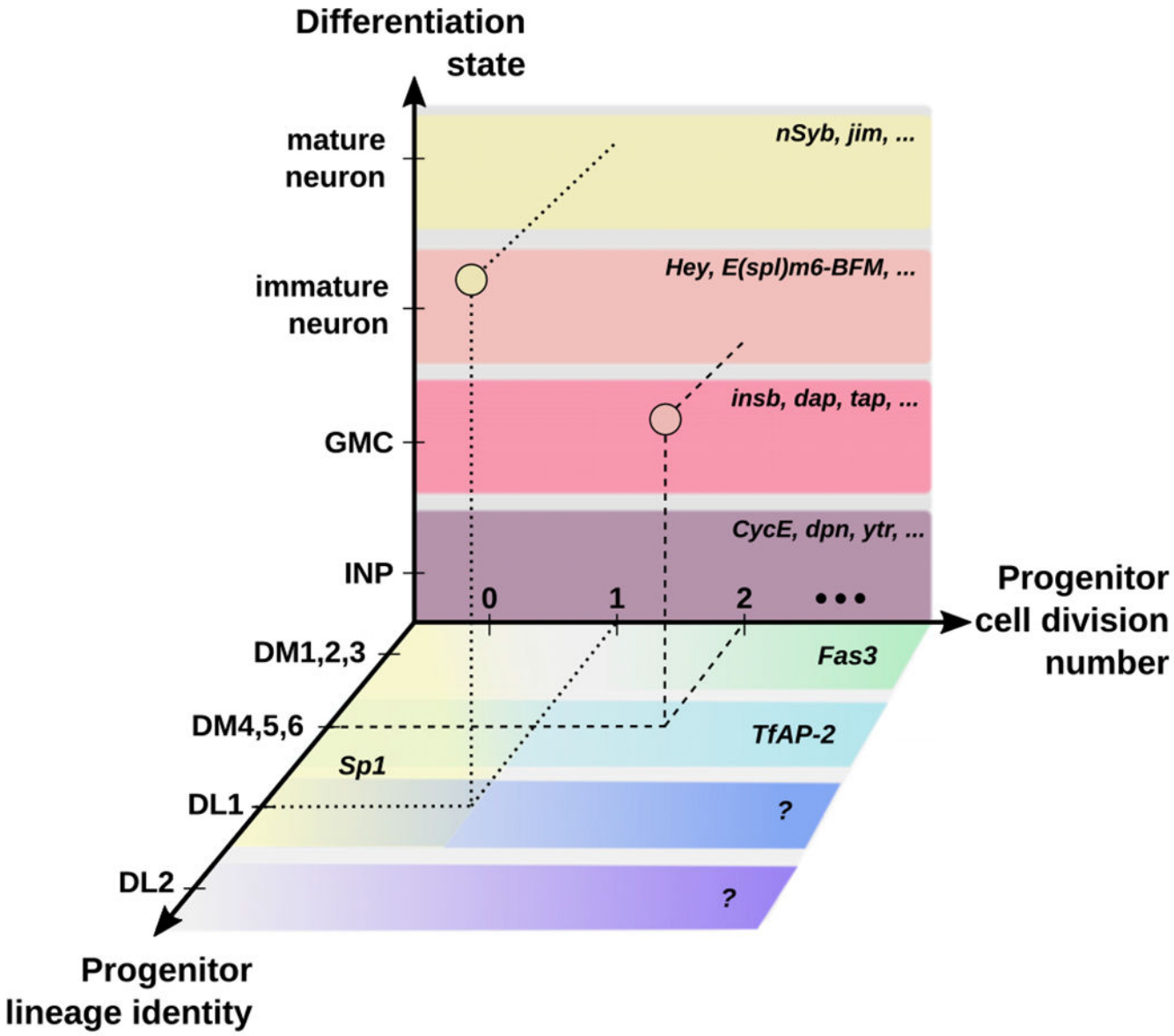


Figure 6. A *Drosophila* type II neuronal fate-specification model illustrates the complex molecular network that determines the neural differentiation process

Despite its small scale and apparent simplicity, the complex interplay of molecular factors that vary along the differentiation state, lineage identity, and progenitor cell division number axes are responsible for determining the fate of each cell derived from the type II neuroblasts of *Drosophila*. In this diagram, some of the most prominent molecular factors from the literature or identified and validated in this work are shown to occupy different domains along these three axes. Multi-time-point analysis and *in situ* validation will enable us to continue to fill in the blanks and develop a more complete roadmap of the type II neurogenesis process across development.

KEY RESOURCES TABLE

REAGENT OR RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Mouse monoclonal anti-Fas3	DHSB	Cat# 7G10, RRID:AB_528238
Rat monoclonal anti-Dpn	Lee et al. (2006)	NA
Donkey-anti-Ms (AF488)	Jackson ImmunoResearch Laboratories, Inc.	Cat# 715-545-151
Donkey-anti-Rt (AF647)	Jackson ImmunoResearch Laboratories, Inc.	Cat# 712-605-150
Chemicals, peptides, and recombinant proteins		
Papain	Millipore Sigma	Cat# P4762-25MG
Collagenase type I	Millipore Sigma	Cat# SCR103
E-64	Millipore Sigma	Cat# E3132-1MG
Fetal Bovine Serum	Millipore Sigma	Cat# F0926-50ML
Schneider's Media	Millipore Sigma	Cat# S0146-500ML
DRAQ5	abcam	Cat# ab108410
Dextran sulfate, 50% solution	Millipore Sigma	Cat# S4031
Critical commercial assays		
10X chromium v3 single-cell gene expression kit	10X Genomics	Cat# 1000154
10X chromium v2 single-cell gene expression kit	10X Genomics	Cat# 120234
Deposited data		
Raw reads and analyzed counts matrices	This study	GEO: [ID here]
Experimental models: organisms/strains		
D. melanogaster, R9D11-Gal4 driver line: w[1118]; P{y[+t7.7] w[+mC] = GMR9D11-GAL4}attP2	BDSC	RRID:BDSC_40731
D. melanogaster, R9D11-CD4::tdTomato membrane reporter line: w[1118]; P{y[+t7.7] w[+mC] = R9D11-CD4-tdTom}attP2/TM6B, Tb[1]	BDSC	RRID:BDSC_40731
D. melanogaster: yw;;UAS-hH2B::2xmNG	This study	NA
D. melanogaster: yw;;UAS-hH2B::2xTagBFP2	This study	NA
D. melanogaster, Sp1::EGFP protein fusion reporter line: w[1118]; PBac{y[+mDint2] w[+mC] = Sp1-EGFP.S}VK00033	BDSC	RRID:BDSC_38669
D. melanogaster, UAS-IVS-myr::tdTomato membrane reporter line: w[*]; P{y[+t7.7] w[+mC] = 10XUAS-IVS-myr::tdTomato}attP40	BDSC	RRID:BDSC_32222
Oligonucleotides		
mNeonGreen HCR probe set	Molecular Instruments	PRC014
CycE HCR probe set	Molecular Instruments	PRD167
D HCR probe set	Molecular Instruments	PRC881
Sp1 HCR probe set	Molecular Instruments	PRC883
TfAP-2 HCR probe set	Molecular Instruments	PRD168
Fas3 HCR probe set	Molecular Instruments	PRC900
ytr HCR probe set	Molecular Instruments	PRE680

REAGENT OR RESOURCE	SOURCE	IDENTIFIER
E(spl)m6-BFM HCR probe set	Molecular Instruments	PRE684
tap HCR probe set	Molecular Instruments	PRE682
jim HCR probe set	Molecular Instruments	PRE686
lncRNA:cherub HCR probe set	Molecular Instruments	PRG382
dati HCR probe set	Molecular Instruments	PRG385
mamo HCR probe set	Molecular Instruments	PRG383
bi HCR probe set	Molecular Instruments	PRG384
Software and algorithms		
Fiji/ImageJ	Schindelin et al., 2012	RRID:SCR_002285 https://fiji.sc/
scanpy scRNA-seq analysis software	Wolf et al., 2018	RRID:SCR_018139
STAR RNA-seq aligner	Dobin et al., 2013	RRID:SCR_015899 https://github.com/alexdobin/STAR
Palantir pseudotime trajectory fitting software	Setty et al., 2019	https://github.com/dpeerlab/Palantir
PyGAM model fitting software	Servén et al., 2018	https://github.com/dswah/pyGAM
MiCV web tool	This study	https://micv.works https://github.com/Cai-Lab-at-University-of-Michigan/MiCV