

NEURAL STEM CELL HETEROGENEITY AND ADULT NEURONAL DIVERSITY

JOSÉ TELES REIS

**A dissertation submitted in partial fulfillment of the requirements for the Degree of Masters in
Biomedical Research**

Dissertação para obtenção do grau de Mestre em Investigação Biomédica

at Faculdade de Ciências Médicas | NOVA Medical School of NOVA University Lisbon

August, 2020

NEURAL STEM CELL HETEROGENEITY AND ADULT NEURONAL DIVERSITY

José Teles Reis

**Supervisor: Dr. Catarina Homem, Principal Investigator at Chronic Diseases Research Center
CEDOC/ NOVA Medical School**

**A dissertation submitted in partial fulfillment of the requirements for the Degree of Masters in
Biomedical Research**

Dissertação para obtenção do grau de Mestre em Investigação Biomédica

August, 2020

Acknowledgments

I would first like to express my deep gratitude to Dr. Catarina Homem, my supervisor, for her mentorship and unwavering support throughout this year. I have admired Dr. Homem's work since my undergraduate days, and so it was an immense pleasure to work in her lab and in a research area that is so dear to me, Developmental Biology.

Dr. Homem introduced me to PhD student Graça Marques, with whom I worked very closely and learned a lot from. We tackled very interesting challenges together. My deep thanks must go out to her as well.

I wish to convey my sincere appreciation to Dr. Tiago Baptista for his support in keeping the flies well taken care off and continuing my experimental crosses during the state of emergency early this year. You were our fly hero.

All members of the *Proliferation and Fate Regulation of Stem Cells* laboratory played an important role throughout this year, as they embody the collaborative spirit on which a successful learning environment thrives. I was given important advice and participated in lively discussions.

I would also like to thank Dr. Rita Teodoro and Dr. Paulo Pereira for guidance during these two years. You have played an important role in my personal and professional development as well.

All in all, I will always cherish my Masters' years in the company of immensely talented scientists. I can only hope that the future may bring opportunities to collaborate once again.

I would like to dedicate this thesis to my loved ones, my Mother, Zoraida Teles, my Brother, Manuel Teles Reis, my Grandparents, Odete Teles and Gerardo Teles, and Ana Raquel Cruz.

Funding

This project was funded by the European Research Council
- ERC H2020-ERC-2017-STG-GA 759853-StemCellHabitat



Abstract

Vast cellular diversity can be observed in the animal nervous system. In the fruit fly, multipotent neural stem cells, neuroblasts (NB), represent the fundamental unit for the production of all neurons of the adult. These divide asymmetrically to give rise to daughter cells which then produce neurons or glia. For more than 30 years, *Drosophila* NB lineages have been a powerful model for neurodevelopmental studies, which has allowed the identification of conserved basic molecular mechanisms. Previous work has shown that various fate determinant genes are differentially expressed in between lineages, which results in the production of diverse sets of neurons by each NB in a stereotyped manner. However, we still lack a global understanding of all genetic instructions necessary in each lineage for the construction of the adult central nervous system. To try to contribute to this ultimate goal, our lab has revisited the specification of neuronal diversity through single-cell technologies, a powerful method to identify cellular heterogeneity. Specifically, in this work, we have curated a dataset of single-cell sequencing of larval NB lineages through bioinformatic methods. This allowed the identification of populations of cells that seem to match all the NB lineage cell types. By comparing the transcriptomic profiles of these cell types, we have identified multiple genes that become differentially expressed along NB lineage development. This included known NB lineage fate regulators, but also unstudied genes within this context. Furthermore, we have observed notable expression heterogeneity within NBs and neurons. Interestingly, we found that multiple genes differentially expressed between larval NBs show the same expression pattern in embryonic NBs. This supports the idea that NB lineage identity programs are conserved from embryo to larval neurogenesis. Finally, we have also validated *in vivo* the expression of a few marker genes which were found to be distinctively expressed among NBs. Overall, this work should provide future studies with a useful molecular blueprint of players potentially driving NB lineage development and the generation of neuronal diversity.

Keywords: *Drosophila*, neuronal diversity, neuroblast lineages, single-cell sequencing.

Resumo

Uma vasta diversidade celular pode ser observada no sistema nervoso animal. Na mosca da fruta são os neuroblastos (NB), células estaminais multipotentes, que geram todos os neurónios do indivíduo adulto. Estes dividem-se assimetricamente gerando células filhas que de seguida se dividem originando neurónios ou células da glia. Há mais de 30 anos que as linhagens de NBs são estudadas como modelo para entender os mecanismos moleculares básicos por detrás do desenvolvimento neuronal. Estudos mostram que diferentes genes ligados à especificação celular podem ser diferencialmente expressos entre as variadas linhagens. Como resultado, diferentes subtipos neuronais são então gerados a partir de cada NB. No entanto, ainda não temos um entendimento global de todas as instruções genéticas necessárias em cada linhagem para a construção do sistema nervoso central do adulto. De modo a contribuirmos para o seu estudo, recentemente o nosso laboratório revisitou a especificação da diversidade neuronal utilizando o método de sequenciação de célula-única, um método que permite a identificação de heterogeneidade celular. Neste trabalho, dados de sequenciação de célula-única de linhagens de NBs foram analisados com recurso a métodos bioinformáticos. Isto levou à identificação de todos os tipos celulares característicos das linhagens de NBs. Através da comparação dos perfis de expressão das várias categorias celulares, foi possível identificar múltiplos genes que se encontravam diferencialmente expressos ao longo da diferenciação das linhagens de NBs. Incluídos estavam vários genes conhecidos como estando envolvidos na especificação das linhagens de NBs, mas também alguns genes que ainda não foram estudados neste contexto. Além disto, também observámos uma notável heterogeneidade entre NBs e neurónios. Curiosamente, vários genes que foram identificados como estando diferencialmente expressos entre os NBs da larva, também apresentam este padrão no embrião. Isto apoia a ideia de que a identidade e os programas transcripcionais das várias linhagens são conservados desde a fase de neurogénese do embrião até à larva. Por último, a expressão de alguns genes marcadores que tinham sido identificados como estando heterogeneamente expressos foi também validada *in vivo*. No geral, este trabalho poderá providenciar estudos futuros um mapa dos genes potencialmente envolvidos no desenvolvimento de linhagens de NBs e na criação da diversidade neuronal.

Palavras-chave: *Drosophila*, diversidade neuronal, linhagens de neuroblastos, sequenciação de célula-única

Index

Acknowledgments	i
Abstract	iii
Index	vii
Index of figures	ix
Index of tables	xi
Abbreviation list	xiii
Introduction	1
Neuronal diversity	1
<i>Drosophila</i> neuroblast lineages, a model for neurodevelopmental studies.....	1
Molecular players behind neuroblast lineage progression.....	4
Lineage intrinsic mechanisms in <i>Drosophila</i>	5
Spatial patterning influence NB origin and lineage identity	5
Temporal patterning diversifies neuronal progeny across time.....	7
Binary fate decisions induce contrasting behaviours in sibling neurons	8
Interaction of lineage intrinsic mechanisms and neuronal diversity	8
Neuronal terminal differentiation	9
From neuroblast lineages to adult neuropils, understanding the principles of clonal organization	10
Identification of individual neuroblast lineage markers.....	12
Resolving cellular diversity with single-cell transcriptomics.....	12
Study hypothesis and aims	14
Materials and methods	15
Fly husbandry	15
Immunohistochemistry and confocal imaging	15
Bioinformatic analysis of single-cell transcriptomic data	16
Seurat object creation and filtration	16
Seurat object pre-processing	16
Dimensionality reduction, clustering analysis and major cell groupings.....	16
Identification of cluster marker genes	17
Type I NB subclustering analysis.....	18
Gene ontology set enrichment analysis.....	18
Gene annotation lists.....	18

Plots generated outside Seurat	18
Results and discussion	19
Sample quality control and filtering	19
Clustering analysis identifies NB lineage cell types	19
Analysis of gene expression changes along NB lineage differentiation and proliferation mode	23
Type I neuroblast subclustering analysis unveils transcriptional heterogeneity	28
Neuronal cluster markers reveal differences in maturation state and fate identity	33
Conclusion and future studies.....	39
References.....	43
Supplementary information	53
Supplementary figures.....	53
Supplementary files	57

Index of figures

Introduction

Fig. 1 NB origin and neurogenesis	3
Fig. 2 NB lineage proliferation modes and cell type markers	4
Fig. 3 Three lineage intrinsic mechanisms promote CNS heterogeneity in the fly	6
Fig. 4 Identification of the adult neuropil correspondents of NB lineages	11
Fig. 5 Droplet-based microfluidic scRNAseq procedures.....	14

Results and discussion

Fig. 6 Quality control and dataset filtering.....	20
Fig. 7 Clustering analysis identifies NB lineage cell types and a transition cellular state	22
Fig. 8 Quality control for the major cluster groups	25
Fig. 9 Major cell group comparison identifies gene expression changes across NB lineage development and proliferation mode.....	26
Fig. 10 Subclustering analysis identifies transcriptional heterogeneity within type I NBs ...	29
Fig. 11 Validation and localization assessment of type I NB subcluster marker genes.	32
Fig. 12 Neuronal cluster gene expression diversity reveal differences in neuronal maturation and fate identity.....	34
Fig. 13 Early born and maturation neuronal genes	35

Supplementary information

Suppl. Fig 1 Jackstraw results.....	53
Suppl. Fig.2 Expression of NB lineage cell type specific markers in all clusters.	54
Suppl. Fig.3 Cluster marker genes identified for each major cell group comparison.....	55
Suppl. Fig.4 Determining the significant PCs for type I NB subclustering	56
Suppl. Fig.5 Additionally validated type I NB subcluster markers	57

Index of tables

Materials and methods

Table 1: Material list, respective sources, and RRID identifiers 15

Supplementary information

Suppl. file 1 All type I NB subcluster markers 57

**Suppl. file 2 Transcription factor genes identified to be heterogeneously expressed in the
type I NB subclustering analysis..... 57**

Abbreviation list

abd-A	abdominal A	GMCs	Ganglion mother cells
Abd-B	Abdominal B	GO	Gene ontology
acj6	abnormal chemosensory jump 6	grh	grainyhead
Ama	Amalgam	grn	grain
Antp	Antennapedia	hb	hunchback
ase	asense	Hey	Hairy/E(spl)-related with YRPW motif
AUC	Area under the receiver operating curve	hkb	huckbein
beat	beaten path	HLH4C	Helix loop helix protein 4C
brat	brain tumor	HmgD	High mobility group protein D
CadN	Cadherin-N	hth	homothorax
cas	castor	imGMCs	immature ganglion mother cells
CB	Central brain	Imp	IGF-II mRNA-binding protein
cDNA	Complementary deoxyribonucleic acid	ind	intermediate neuroblast defective
cOPC	central OPC	INPs	intermediate neural progenitors
Cha	Choline acetyltransferase	insb	insensible
che-1	abnormal chemotaxis	isl	islet
CNS	Central nervous system	jigr1	jing interacting gene regulatory 1
crc	cryptocephal	klu	klumpfuss
croc	crocodile	Kr	Kruppel
dac	dachshund	lab	labial
Dfd	Deformed	Lac	Lachesin
dINPs	Developing intermediate neural progenitors	lncRNA	long non-coding ribonucleic acid
DII	Distal-less	lola	longitudinals lacking
DNA	Deoxyribonucleic acid	lolal	lola like
dmrt99B	doublesex-Mab related 99B	M1BP	Motif 1 Binding Protein
dpn	deadpan	mOPC	main arms of the OPC
dpr	defective proboscis extension response	mamo	maternal gene required for meiosis
Dscam	Down syndrome cell adhesion molecules	mei-P26	meiotic P26
Dsp1	Dorsal switch protein 1	mINPs	mature intermediate neural progenitors
elav	embryonic lethal abnormal vision	mira	miranda
EloC	Elongin C	mirr	mirror
emc	extra macrochaete	mRNA	Messenger ribonucleic acid
ems	empty spiracles	msh	muscle segment homeobox
en	engrailed	nAChR	nicotinic Acetylcholine Receptor
erm	earmuff	NB	Neuroblast
eve	even-skipped	NijA	Ninjurin A
exex	extra-extra	nrv3	nervana 3
Fas3	Fasciclin 3	nSyb	neuronal Synaptobrevin
fd96Cb	forkhead domain 96Cb	OL	Optic lobe
Fer2	48 related 2	OPC	Outer proliferation center
fhk	forkhead	otd	orthodenticle
gcm	glial cells missing	otp	orthopedia
GLAD	Gene List Annotation for Drosophila	pOPC	posterior arms of the OPC

PAT	Primary axonal tract
pb	proboscipedia
PBS	Phosphate-buffered saline
PBT	Triton X-100 and PBS
PC	Principal component
PCA	Principal component analysis
pdm	POU domain
pnt	pointed
pros	prospero
repo	reversed polarity
RNA	Ribonucleic acid
RNAseq	Ribonucleic acid sequencing
ROC	Receiver operating curve
rpr	reaper
RT	Room temperature
run	runt
Rx	Retinal homeobox
SAT	Secondary axonal tract
Scr	Sex combs reduced
scRNAseq	Single cell ribonucleic acid sequencing
sh	shaker
side	sidestep
skl	sickle
slo	slowpoke
slp	sloppy paired
stau	staufen
svp	seven-up
Syp	Syncrin
TF	Transcription factor
tll	tailless
tOPC	tips of the outer proliferation center
toy	twin of eyeless
tsh	teashirt
Ubx	Ultrabithorax
UMAP	Uniform manifold approximation and projection
UMI	Unique molecular identifier
unc-3	uncoordinated 3
VACHT	Vesicular acetylcholine transporter
VNC	Ventral nerve cord
vnd	ventral nervous system defective
Vsx1	Visual system homeobox 1
wor	worniu
wun2	wunen-2

Introduction

Neuronal diversity

The enormous heterogeneity in the adult central nervous system (CNS) was immediately made clear through the depictions of Cajal (Cajal, 1999). Critically positioned cell bodies with remarkably different arborization patterns convolute into hardwired neuronal networks. It has puzzled biologists for decades how this variability is achieved and how finely tuned circuits are assembled.

Three different models have been proposed to lie at the genesis of nervous system heterogeneity. Both from invertebrate and vertebrate neurobiology, it is accepted that all neurons derive from the differentiation of neural stem cells (Homem *et al.*, 2015). In the first model, neural stem cells originate from distinct spatial regions of the primordial tissues, which early in development acquire regional differences. This results in the expression of distinct transcription factors between neural progenitors, endowing each stem cell with its own unique transcriptional signature. This identity signature is then thought to be passed to the descendants during the neuroprogenitor division, contributing to the generation of distinct morphological and functional neuronal outputs by each progenitor. The mechanisms belonging to this model are termed lineage intrinsic or pre-mitotic (Li *et al.*, 2018) (Telley and Jabaudon, 2018). Furthermore, extrinsic signals, such as neighbouring cells or long-distance cues, have also been shown to influence neuronal topography (Kolodkin and Tessier-Lavigne, 2011). Mechanisms belonging to this second model are termed lineage extrinsic or postmitotic (Li *et al.*, 2018) (Telley and Jabaudon, 2018). Thirdly, later in the adult life, neurons have also been shown to be plastic and to adapt to life challenges. For example, circuits have been shown to be modulated in an activity-dependent manner (Guan *et al.*, 2005) or through circadian rhythm oscillations (Krzeptowski *et al.*, 2018). In this third model of neuronal diversity generation, both cell and non-cell autonomous mechanisms are considered (Harris and Littleton, 2015).

Deciphering the genetic basis of neurodevelopment has long been an objective of Biology with rewarding translational promises, such as finding new targets for regenerative medicine protocols or identifying new ways to treat neurodegenerative disorders. The work of this thesis focuses on trying to identify new molecular players which could potentially orchestrate the assembly and generation of the cellular diversity of the nervous system.

***Drosophila* neuroblast lineages, a model for neurodevelopmental studies**

One hundred and ten years of research discoveries and six Nobel prizes have established *Drosophila melanogaster* as a powerful research model (Bellen *et al.*, 2010) (Prokop, 2015). Given its fast-reproductive cycle and simplicity, the fly has allowed researchers to systematically address

the genetic mechanisms behind numerous biological phenomena. More recently, with the application of binary expression systems, such as *UAS-GAL4*, with just a fly cross it has been possible to induce the misexpression of specific genes in a temporally controlled manner (Brand and Perrimon, 1993). A number of consortiums have developed fly line collections in which *GAL4* expression is controlled by thousands of different cis-regulatory regions (Jenett et al., 2012) (Tirian and Dickson, 2017), allowing manipulations to almost all cell types within reach. This incredible genetic tractability coupled with the fact that many pathways have been shown to be conserved in mammals (Rieder and Larschan, 2014) (Wangler et al., 2015), make the fly a popular system to tackle a wide range of questions, such as to how the nervous system is assembled.

The adult *Drosophila* CNS is subset into the central brain (CB), ventral nerve cord (VNC), and the Optic lobe (OL). During development, multiple neural stem cells, Neuroblasts (NB), divide to produce the neurons that constitute the CNS (Homem and Knoblich, 2012). NBs are categorized according to their embryonic origin or according to their proliferation mode (Yaghmaeian Salmani and Thor, 2020). VNC-NBs are formed from the embryonic ventral neurogenic region (Yaghmaeian Salmani and Thor, 2020). Here, multiple clusters of cells transiently start expressing the proneural genes *achaete*, *scute*, and *lethal of scute*. However, through Notch-mediated lateral inhibition, only one cell per cluster will permanently end up expressing these genes. As a result, this cell will specify into a NB and delaminate (Fig.1A) (Skeath, 1999) (Yaghmaeian Salmani and Thor, 2020). CB-NBs arise from the procephalic neurogenic region through a variety of mechanisms. While part of CB-NBs undergoes delamination like VNC-NBs, in some areas they are generated through vertical divisions of the neuroectoderm. Furthermore, in a specific area of the procephalic region, neighbouring cells were described to both generate NBs, which suggests the non-requirement of lateral inhibition for the genesis of some progenitors (Urbach and Technau, 2003) (Yaghmaeian Salmani and Thor, 2020). Finally, OL NBs start to be generated in the embryonic stage and also continue to be produced in the larval stages. OL NBs are produced from the delamination of the OL neuroepithelium (Hakes et al., 2018) (Egger et al., 2007).

Neurogenesis in *Drosophila* occurs in two waves (Fig. 1B) (Homem and Knoblich, 2012). First, in the embryo, NBs delaminate from the neurogenic regions (Fig. 1A) and undertake multiple division cycles to give rise to larval neurons, also designated as primary neurons (Fig. 1B) (Urbach, 2003a) (Pereanu, 2006) (Homem and Knoblich, 2012). These will extend axonal projections (Primary axonal tract – PAT), which will branch and gain neuronal terminal features in the late embryonic stages (Pereanu, 2006). Here, part of the NBs suffer apoptosis, while the majority pause their cycling period and stay in an inactive state, also known as quiescence (Prokop and Technau, 1991) (Maurange and Gould, 2005). In the larval stage, the quiescent NBs are reactivated and start producing more descendants (Fig. 1B) (Homem and Knoblich, 2012). Although a small percentage of primary neurons are conserved into adulthood, it is in this second wave of neurogenesis that the great

majority of adult neurons are born (Homem and Knoblich, 2012). These, also known as secondary neurons, will similarly extend axonal projections as primary neurons (Secondary axonal tracts – SAT), but interestingly, in the great majority of lineages, only fully differentiate in the pupal stages (Truman, 1990) (Pereanu, 2006). In terms of their proliferation mode, NB lineages can be distinguished into three types (Yaghmaeian Salmani and Thor, 2020). Type I NBs divide in an asymmetric manner to create a new NB and a ganglion mother cell (GMC) that then gives rise to two neurons, two glial cells, or one neuron and one glial cell. Type II NBs produce an intermediate neural progenitor (INP) that also divides asymmetrically to self-renew and generate a GMC on each division (Fig. 2). However, the INP has a much lower mitotic potential than the type II NB (Homem and Knoblich, 2012). The generation of an INP results in a larger production of neurons by type II lineages. In 10 stem cell divisions, Type II lineages will give rise to 60 neurons as opposed to only 20 neurons by type I lineages (Yaghmaeian Salmani and Thor, 2020). Finally, type 0 NBs do not generate INPs or GMCs upon division. Instead, they divide and immediately give rise to one neuron (Yaghmaeian Salmani and Thor, 2020).

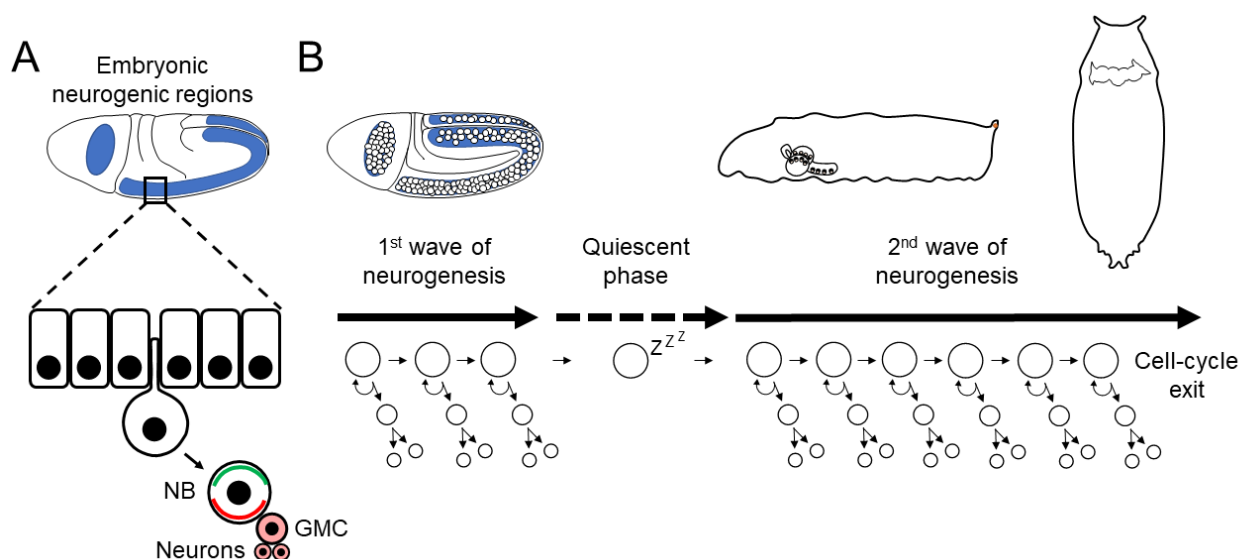


Fig. 1 NB origin and neurogenesis

A) NB delaminate from the embryonic neurogenic regions and afterwards begin cycling asymmetrically (Urbach, 2003a) (Homem and Knoblich, 2012). Proteins (inscuteable in green, miranda in red) in opposite sides of the NB help segregate differentially fate determinants into the self-renewing NB and GMC. B) Nervous system development occurs in two waves (Homem and Knoblich, 2012). Larval neurons (also termed primary neurons) are born in the embryonic and first neurogenic wave. In the late embryo, NBs stop proliferating and enter quiescence. A percentage of these NBs undergo programmed cell death. In the larva, NBs are reactivated and specify most adult neurons (also termed secondary neurons). Later in the pupa, the great majority of NBs arrest the cell cycle and stop generating neurons (Homem *et al.*, 2014). Figure was adapted from (Arefin *et al.*, 2019) (Gallaud *et al.*, 2017) (Wodarz and Huttner, 2003).

There are around 105 NBs in the CB per lobe, of which 8 lineages have type II NBs, and the remaining are thought to be type I (Doe, 2017). In the VNC, 30 type I NBs can be found per

hemisegment (Doe, 2017) and specific lineages have been shown to switch into type 0 mode in late embryogenesis (Baumgardt et al., 2014). Finally, in the OL, 700 NBs are present per lobe (Doe, 2017), of which the large majority are thought to have a type I division pattern. However, in the tOPC, early born NBs have been shown to divide in a type 0 pattern, and switch to a type I behaviour when they age (Bertet et al., 2014).

Molecular players behind neuroblast lineage progression

A growing body of literature has examined the genetic basis behind fate decision and NB lineage progression. A number of transcription factors have been found to be critical regulators. The TF deadpan (dpn), has been shown to be important for stem cell self-renewal and maintenance (Doe, 2008) (Zhu *et al.*, 2012). Accordingly, dpn expression is restricted to NBs, and additionally to INPs in type II lineages (Fig.2). The TFs *worniu* (*wor*) and *asense* (*ase*) are expressed in NBs and GMCs (Brand *et al.*, 1993) (Ashraf and Ip, 2001), except for Type II NBs which are *ase* negative (Bowman *et al.*, 2008) (Fig.2). This absence of *ase* has been found to be critical for the differences between type I and type II lineage proliferation modes (Bowman *et al.*, 2008), which are further regulated by pointed (*pnt*) and earmuff (*erm*) TFs (Zhu *et al.*, 2011) (Weng *et al.*, 2010).

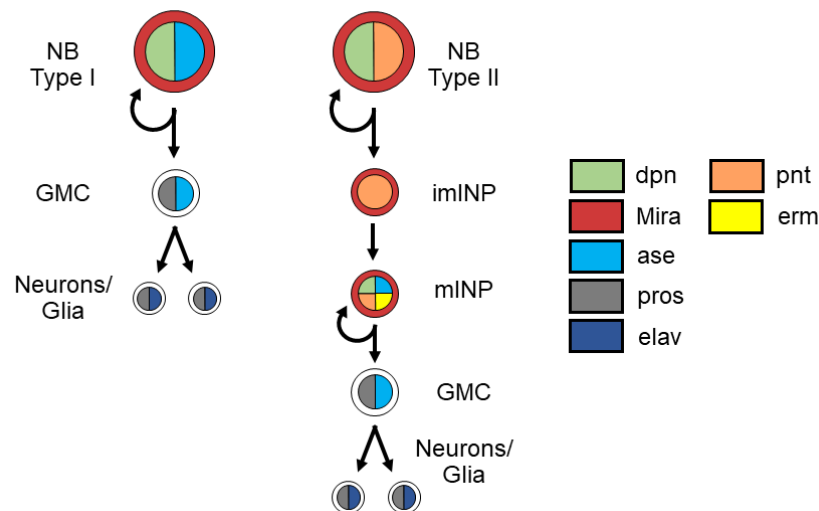


Fig. 2 NB lineage proliferation modes and cell type markers

NB lineages can be distinguished in terms of their proliferation mode and molecular markers (deadpan – green; miranda – red; asense – turquoise; prospero – gray; elav – dark blue; pointed – orange; earmuff – yellow). Type I lineages are comprised by type I Neuroblasts (NB type I), Ganglion mother cell (GMC) and Neurons and/or Glia. Type II lineages are constituted by type II Neuroblasts (NB type II), Immature intermediate neural progenitors (imINP), Mature intermediate neural progenitors (mINP), GMCs, and Neurons and/or Glia.

On the other hand, genes like the TF *prospero* (*pros*) and *brain tumor* (*brat*), a RNA binding protein which can repress translation, have been shown to limit neural stem cell self-renewal and to promote neuronal differentiation (Doe, 2008) (Lee *et al.*, 2006) (Loedige et al., 2014). Lastly, a critical step in

lineage progression is the asymmetric segregation of the latter factors upon stem cell division into the differentiating daughter cell (Ikeshima-Kataoka *et al.*, 1997) (Lee *et al.*, 2006). This is regulated by scaffolding proteins like miranda (mira) (Ikeshima-Kataoka *et al.*, 1997).

Lineage intrinsic mechanisms in *Drosophila*

Three lineage intrinsic mechanisms have been found to be conserved and critical for the specification of neuronal diversity. These three mechanisms can be termed as spatial patterning, temporal patterning, and binary cell fate decisions (Lin and Lee, 2012). For the sake of simplicity, below I only develop on studies in *Drosophila*, but examples for these mechanisms have been shown in the worm (Jukam and Desplan, 2010) and mammals (Kohwi and Doe, 2013) (Holguera and Desplan, 2018).

Spatial patterning influence NB origin and lineage identity

Early events during *Drosophila* oogenesis induce the expression of a group of genes which confer spatial identity to the cells of the future embryo (Gilbert and Barresi, 2016). In the anterior to posterior axis, the first spatial patterning gene group to become activated are the gap genes, which define wide regions of the embryo (French, 2001). These then activate a cascade of three other groups of genes which divide the embryo into even more restricted spatial domains. These are the pair-rule, segment polarity and homeotic genes (French, 2001) (Gilbert and Barresi, 2016). In a parallel and similar fashion, maternal factors also determine the embryo dorsal-ventral axis through the expression of a set of genes in a columnar restricted pattern (Stein and Stevens, 2014).

Spatial patterning genes have been found to have a critical role in neurogenesis. These groups of genes are widely expressed in the embryonic neurogenic regions and in NB lineages (Fig.3A). Early studies have shown that mutations in these genes result in the disappearance of NBs from specific areas of the embryo. Interestingly, these areas are tightly correlated with the spatial expression pattern of these genes, suggesting that these play a role in NB origin (Younossi-Hartenstein *et al.*, 1997) (Weiss *et al.*, 1998). Moreover, loss-of-function experiments have shown that gap genes such as *tailless* (*tll*), *orthodenticle* (*otd*), *empty spiracles* (*ems*) are necessary for the expression of lethal of scute, which is necessary for NB specification (Younossi-Hartenstein *et al.*, 1997).

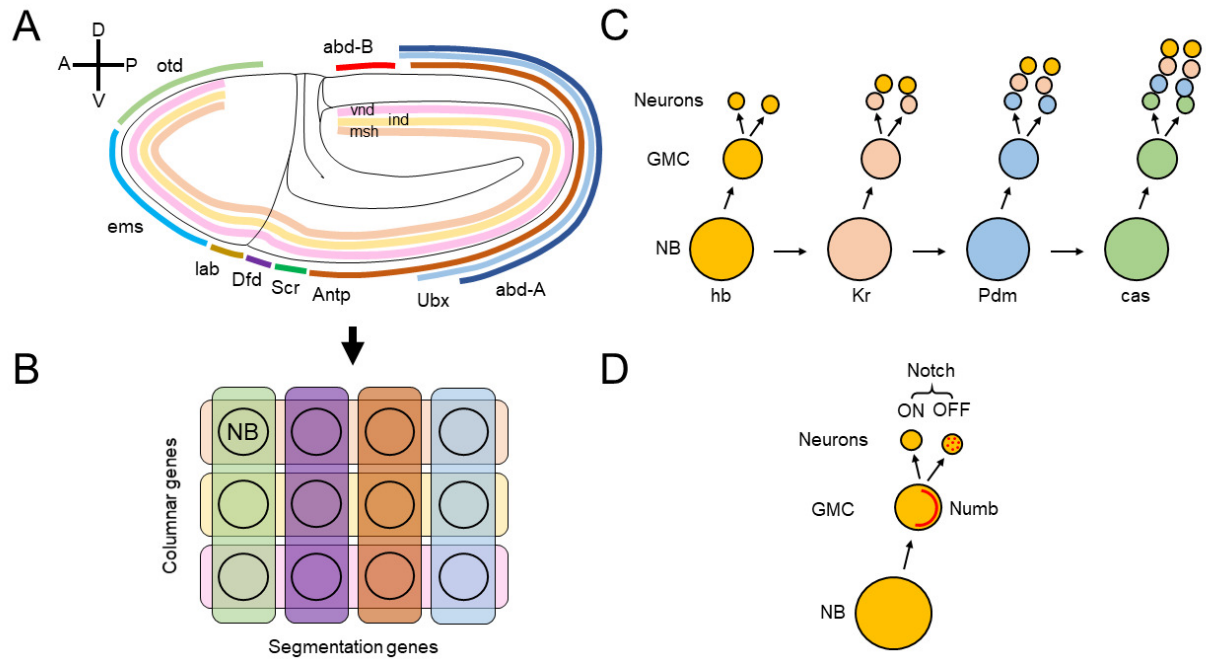


Fig. 3 Three lineage intrinsic mechanisms promote CNS heterogeneity in the fly

A) NBs born from distinct neurogenic regions express different spatial patterning genes. Here are represented a few spatial patterning genes and their NB expression domains. In the anterior to posterior orientation, NBs express the gap genes *otd* and *ems*, homeotic genes *labial* (*lab*), *Deformed* (*Dfd*), *Sex combs reduced* (*Scr*), *Antennapedia* (*Antp*), *Ultrabithorax* (*Ubx*), *abdominal-A* (*abd-A*) and *Abdominal-B* (*AbdB*) (Allan and Thor, 2015). In the dorsal ventral orientation, NB express *muscle segment homeobox* (*msh*), *intermediate neuroblast defective* (*ind*), and *ventral nervous system defective* (*vnd*) in a columnar fashion (Allan and Thor, 2015). **B)** The overlay of spatial patterning genes “creates a Cartesian coordinate map”, where all NBs inherit a unique expression profile (Lin and Lee, 2012). **C)** NB express a TF cascade (termed temporal factors) generating different neuronal progeny as time progresses. Here is shown an example from the VNC lineages in the embryo, which express hunchback (*hb*)→ Kruppel (*Kr*)→ POU domain (*pdm*)→ castor (*cas*) (Lin and Lee, 2012). **D)** Numb, a Notch inhibitor, is asymmetrically localized before the GMC division and is only segregated to one daughter cell. This results in contrasting binary behaviours in newly born neurons due to different levels of Notch activity (Lin and Lee, 2012). This figure was adapted from (Allan and Thor, 2015) and (Lin and Lee, 2012).

Besides this role, spatial patterning genes have also been found to be important in defining neuronal progeny identity and diversity. Lundell *et al.* showed that the gap gene *huckbein* (*hkb*) and the segment polarity gene *engrailed* (*en*) were specifically co-expressed in an embryonic lineage NB7-3 and that this lineage generated primary serotonin neurons (Lundell *et al.*, 1996). Upon mutation of *hkb* or *en* there was a strong decrease in the number of serotonin neurons. Interestingly, in *hkb* mutant embryos, NB7-3 was still observed, suggesting that this spatial patterning gene does not have a role in NB delamination but in neuronal progeny identity (Lundell *et al.*, 1996). Another example has been identified in the second wave of neurogenesis. The gap gene *otd* is expressed in a few NB lineages of the CB, which include the LALv1 that give rise to interneurons which target the central complex (Sen *et al.*, 2014). Larval-induced LALv1 *otd* mutant clones generated neurons which lost the expression of LALv1 characteristic markers and morphology. Surprisingly, both at the

molecular level and projection targets, Sen *et al.* showed that these neurons shifted into an Antennal lobe fate akin to the ALad1 lineage (Sen *et al.*, 2014). In accordance, ALad1 lineages are *otd* negative (Sen *et al.*, 2014). More recently, the influence of spatial patterning has been further shown by the work of Sen *et al.* Here, the authors studied the chromatin accessibility profile of specific embryonic NBs which have been patterned by different spatial genes (Sen *et al.*, 2019). These were the NB5-6, which expresses *gooseberry* (*gsb*), and the NB7-4 which is positive for *en* (Sen *et al.*, 2019). Interestingly, these lineages were found to have distinct open chromatin. This suggests a mechanism where lineage specific spatial factors could generate neural diversity by differentially restricting the accessible genome to common fate determinants (Sen *et al.*, 2019). Indeed, after studying the genome occupancy of the TF hunchback (*hb*), which acts in both NBs during a specific time window, this was found to correlate with lineage-specific chromatin profiles (Sen *et al.*, 2019). Overall, these experiments showcase the critical role of spatial patterning in defining neuronal identity and diversity.

In conclusion, spatial patterning genes have roles in early NB formation but also continue to play an important role throughout neurogenesis. The overlay of spatial patterning genes “creates a Cartesian coordinate map”, where all NBs inherit a unique expression profile which will determine the future of the neuronal progeny (Fig.3B) (Lin and Lee, 2012).

Temporal patterning diversifies neuronal progeny across time

Early lineage tracing studies revealed that a single NB could generate various subtypes of neurons over time (Schmid *et al.*, 1999) (Lin and Lee, 2012). This behaviour was explained by the expression of different transcription factors (TFs) in neural progenitors in a sequence-dependent manner (Fig.3C). These TFs, termed temporal factors, are often expressed at the stem cell level for a few divisions and passed to their progeny, diversifying neurons in a time series (Lin and Lee, 2012). In the embryonic VNC most NB lineages studied express a four TF cascade: *hb*→ *Kruppel* (*Kr*)→ *POU* domain (*pdm*)→ *castor* (*cas*) (Fig.3C) (Doe, 2017). In post-embryonic neurogenesis, temporal patterning mechanisms are less well characterized. Recently, there have been great advances in the understanding of temporal patterning mechanisms in the OL. Outer proliferation center (OPC) NBs undergo the TF cascade *homothorax* (*hth*)→ *klu*→ *eyeless* (*ey*)→ *sloppy paired* (*slp*)→ *Dichaete* (*D*)→ *tll* (Li *et al.*, 2013) (Rossi *et al.*, 2017). Additionally, a subset of these progenitors in the tips of the OPC (tOPC), express an alternative but similar TF cascade: *Distal-less* (*Dll*)→ *ey*→ *slp*→ *D* (Bertet *et al.*, 2014) (Rossi *et al.*, 2017). As for the CB and VNC larval lineages, Liu *et al.* have shown that there is an opposing gradient of two RNA binding proteins: IGF-II mRNA-binding protein (*Imp*) → *Syncrin* (*Syp*) (Liu *et al.*, 2015). The latter has been shown to be critical for the generation of the three main types of the mushroom body neurons (Liu *et al.*, 2015) (Rossi and Desplan, 2020). It has also been shown that five TF genes, *chinmo*, *cas*, *seven-up* (*svp*), *mamo* and

Broad (Br), have a temporally regulated expression in NBs and/or newly born neurons (Maurange *et al.*, 2008) (Ren *et al.*, 2017) (Liu *et al.*, 2019) (Syed *et al.*, 2017). Interestingly, *chinmo* and *mamo* temporal protein gradients have been suggested to be post-transcriptionally regulated by *Imp* and *Syp* (Zhu *et al.*, 2006) (Liu *et al.*, 2019). Furthermore, this RNA binding protein gradient has also been shown to be present in antennal lobe and type II lineages, suggesting that it is a common feature to all larval lineages (Lee *et al.*, 2020) (Liu *et al.*, 2015) (Ren *et al.*, 2017). But overall, in the CB and VNC these players seem insufficient to justify the enormous neuronal diversity observed in the adult. A given type I NB lineage could generate around 30 morphologically different temporal fates during the second wave of neurogenesis (Lee *et al.*, 2020). However, the mechanisms and players behind this are unknown. Furthermore, recent work by Lee *et al.*, where a large number of central brain lineages were genetically traced, showed that some lineages produce similar neuronal fates in a cyclical manner (Lee *et al.*, 2020). This suggests that there could exist additional temporal patterning mechanisms in CB and VNC lineages besides the opposing *Imp/Syp* gradient and temporally expressed TFs.

Binary fate decisions induce contrasting behaviours in sibling neurons

Another mechanism of diversification of the CNS is known to occur at the level of the GMC division. In several lineages, neurons that derive from the same GMC have been shown to acquire distinct fates (Fig.3D) (Lin and Lee, 2012). This is the result of differential Notch activity in the newly born neurons, which is caused by the asymmetric segregation of the Notch inhibitor *Numb* in GMCs (Fig.3D) (Lin and Lee, 2012). Hence, these sibling neurons are usually termed NotchON and NotchOFF hemilineages (Truman *et al.*, 2010).

A commonly observed behaviour is for one of the hemilineages to die as a consequence of Notch activity (Lin and Lee, 2012). In the tOPC, NotchON neurons, which are derived from the ey window, undergo programmed cell death (Bertet *et al.*, 2014). However, other examples have shown that the NotchON state does not always predict apoptosis. Again in the lineages of the tOPC, but in the *slp* and *D* windows, NotchOFF neurons undergo apoptosis (Bertet *et al.*, 2014). Furthermore, sister hemilineages can both survive but convert into different neuronal identities. In the lateral antennal lobe lineages, Lin *et al.* have shown that NotchON neurons develop into interneurons, while the NotchOFF sibling neuron transform into projection neurons (Lin *et al.*, 2010) (Lin and Lee, 2012). These differences between binary fates point out to the importance of lineage identity and how it could influence the outcome of the Notch decision.

Interaction of lineage intrinsic mechanisms and neuronal diversity

In the end, the great variety of neuronal subtypes produced are the consequence of the three-layered regulation by the intrinsic mechanisms (Lin and Lee, 2012). The spatial identity to which each NB

lineage was endowed, and the different temporal windows will specify along with Notch the diversity of all hemilineages (Lin and Lee, 2012). This multi-layered regulation has recently been very well demonstrated in the OPC lineages by the work of Erclik *et al.* The OPC is spatially partitioned into three different main areas, the main arms of the OPC (mOPC), the central OPC (cOPC), and the posterior arms of the OPC (pOPC) (Erclik *et al.*, 2017). This partition is the result of the expression of three genes: *Optix* in the mOPC, *Visual system homeobox 1* (*Vsx1*) in the cOPC, and *Retinal homeobox* (*Rx*) in the pOPC (Gold and Brand, 2014) (Erclik *et al.*, 2017). Interestingly, Erclik *et al.* showed that as a result of spatial patterning, lineages in the different areas of the OPC have different binary fate behaviours. In the mOPC NotchOFF hemilineages die, while in the cOPC and pOPC survive (Erclik *et al.*, 2017). Moreover, different sets of TF were shown to characterize the hemilineages of the different OPC areas. NotchOFF hemilineages in the cOPC were positive for *Vsx1*, *svp* and *Lim3*, but in the pOPC hemilineages no longer had *Vsx1* and some expressed *teashirt* (*tsh*) (Erclik *et al.*, 2017). As a consequence of this, different neuronal terminal fates are generated. cOPC NotchOFF hemilineages differentiate into medulla Pm3 neurons, and NotchOFF hemilineages from the pOPC specify into Pm1 and Pm2 neurons (Erclik *et al.*, 2017). Lineage intrinsic mechanisms thus set the trail for the journey of neuronal maturation.

Neuronal terminal differentiation

Early born neurons must then undergo a process of maturation to integrate into a functional circuit which is capable of executing complicated behavioural tasks. There are at least two milestones that one neuron must achieve in order to perform its function. First, neurons must acquire the proper morphology that matches its identity established by the intrinsic and extrinsic developmental programs. This step includes not only an increased membrane growth but also establishing the right partner connections. A group of genes that play a major role in this are cell adhesion molecules. Widely studied examples include the Down syndrome cell adhesion molecules (*Dscam*), which promote neuronal tiling and self-avoidance recognition (Soba *et al.*, 2007), the beaten path-Sidestep (beat-side) and the Defective proboscis extension response-Dpr interacting proteins (Dpr-DIP) families, which are involved in synaptic target specificity (Siebert *et al.*, 2009) (Carrillo *et al.*, 2015) (Tan *et al.*, 2015). Furthermore, mature neurons must also possess all the molecular machinery necessary for neuronal transmission impulse. This consists on all the necessary proteins for neurotransmitter synthesis, degradation and transport, ion channels for membrane depolarization, and proteins required for vesicle transport and exocytosis. But how exactly do lineage intrinsic mechanisms regulate neuronal terminal differentiation? Although multiple examples have taught us that they are necessary, it is still poorly characterized in *Drosophila* how spatiotemporal patterning genes and binary fate decisions control neuronal maturation genes. This has been better described in *C. elegans* where specific TFs have been found to directly regulate neuronal maturation genes (Allan and Thor, 2015). These genes are designated as terminal selectors, and importantly, are

expressed through out development and maintained in the adult, ensuring neuronal terminal programs stay active (Hobert and Kratsios, 2019). Examples in *C. elegans* include *uncoordinated 3* (*unc-3*) which is expressed in the nerve cord motor neurons or abnormal *chemotaxis 1* (*Che-1*) in ASE sensory neurons (Kratsios *et al.*, 2012) (Uchida *et al.*, 2003) (Hobert, 2016). In the fly, TFs *islet* (*isl*) and *even-skipped* (*eve*) have been found to be expressed and modulate electrophysiologically different sets of primary motor neurons (Wolfram *et al.*, 2012) (Pym *et al.*, 2006) (Allan and Thor, 2015). This was proposed to be the result of the negative regulation of ion channels genes *shaker* (*sh*) and *slowpoke* (*slo*) by *isl* and *eve*, respectively (Wolfram *et al.*, 2012) (Pym *et al.*, 2006). Moreover, they have also been shown to regulate the expression of cell adhesion molecules necessary for the establishment of proper neuronal connections (Allan and Thor, 2015), suggesting that these genes are terminal factors.

Due to the incredible diversity and size of adult CNS, neuronal maturation is bound to be a complex biological phenomenon. Further studies of the neuronal transcriptional networks of the fly could potentially provide clues on additional players which could be categorized as terminal selectors.

From neuroblast lineages to adult neuropils, understanding the principles of clonal organization

Immunostaining of the CNS of *Drosophila* for synaptic markers reveals areas of high synaptic density. These, which can also be easily morphologically discerned, are termed neuropils, and are the result of an enrichment for axonal tracts, complex terminal arborizations and dendrites (Yu *et al.*, 2013). Part of the neuropils in the fly nervous system have already been very well studied, not only in terms of their neuronal subtype composition, but also regarding its behavioural function and connectivity with other neuropils. In the central brain, this includes the antennal lobes, which are the first center for odor processing and discrimination (Grabe and Sachse, 2018), the mushroom body, which is required for associative olfactory memory (Heisenberg, 2003), and a group of neuropils which collectively are termed the central complex (Yu *et al.*, 2013). The latter is composed by the ellipsoid body, the fan-shaped body, the noduli, and the protocerebral bridge (Yu *et al.*, 2013), and play important roles in sleep and spatial navigation (Donlea *et al.*, 2018) (Giraldo *et al.*, 2018). There are in total 33 neuropils in the CB of the fly (Yu *et al.*, 2013).

A question that arises is how each NB lineage, the fundamental units for the development of the CNS, contribute individually to each neuropil and the formation of the adult connectome? Understanding what the final innervation path of each NB clone is not only valuable for the study of neuronal circuitry and behaviour, but also ultimately critical for identifying what are the set of mechanistic rules guiding the development of each lineage. Thus, identifying the adult counterparts of each NB lineage is fundamental to solve the mystery of what are the players driving the specification of each neuronal identity.

In a major advance in 2013, two parallel works by the labs of Tzumin Lee (Yu et al., 2013) and Kei Ito (Ito et al., 2013), have undergone exhaustive lineage tracing for the large majority of post-embryonic lineages in the CB. The researchers have been able to show, for the neuronal progeny of each CB NB lineage, how complex their arborizations are and to map to which neuropils they wire to.

In order to achieve this, the authors used similar strategies. Focusing on the tool used Yu et al., this was possible through a technique called Flip-Out MARCM (Fig. 4), which allows the generation of GFP-labelled clones. If the event of FLP-induced mitotic recombination happens at the level of the NB, the lineage may become permanently labelled (Fig. 4).

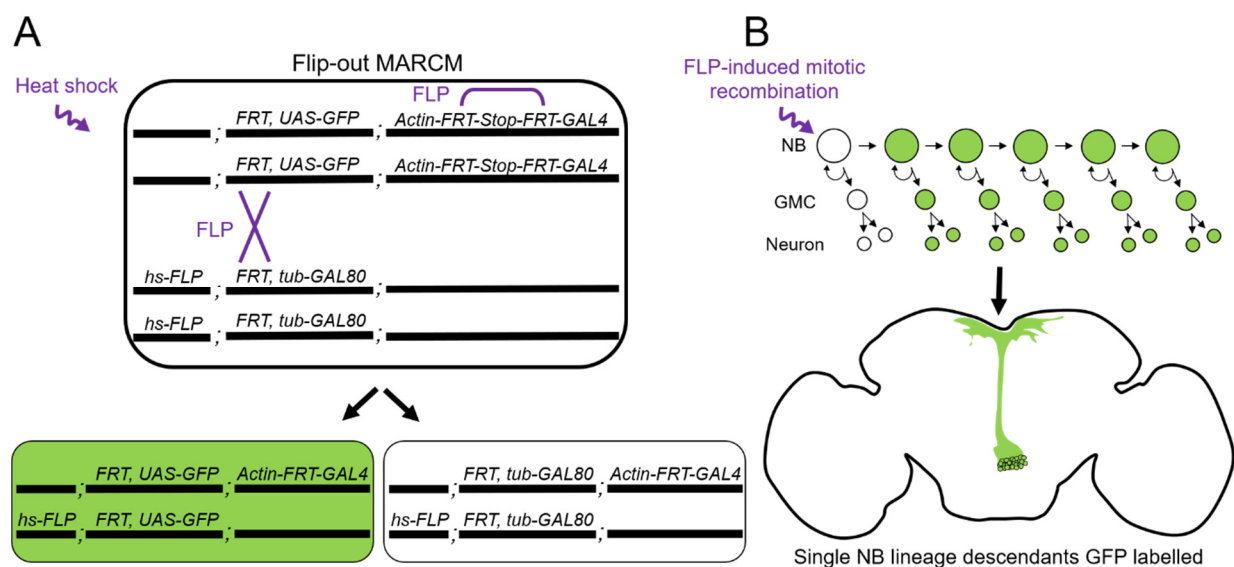


Fig. 4 Identification of the adult neuropil correspondents of NB lineages

A) Schematic shows the Flip-out MARCM tool used by Yu et al. 2013, to map the adult progeny of most CB lineages (Yu et al., 2013). Briefly, a heat-shock induces the expression of a Flipase recombinase which can remove a stop cassette allowing GAL4 to be expressed. Additionally, Flipase induced mitotic recombination allows one of the descendants to lose the tub-GAL80 insert, activating GAL4 which induces GFP expression. B) If the event of FLP-induced mitotic recombination occurs at the level of the NB, the whole lineage can become labelled, except for the first two neurons. Upon dissection of the adult brain, a cluster of neurons which are clones of a single NB lineage (or more if the system was activated in multiple NBs, which is rare) may be labelled. This allows the study of cell body positioning, tract projection site and arborization complexity.

Given that in the fruit fly neurons do not migrate upon being generated, the cell bodies derived from the same progenitor are found to be grouped together in the cortex of brain (Yu et al., 2013). MARCM labelling allowed the counting of the number of neurons produce per lineage. The great majority of lineages had between 30-150 neurons, with two lineages generating just 9 neurons (Yu et al., 2013). This variance in progeny number spotlights how different NB lineages are and that they must have different rules governing their development. Expectedly, the mushroom body lineages and one antennal lobe lineage, that do not undergo quiescence (Yu et al., 2013) (Lee, 2017), and the type II

lineages, which have a different proliferation mode, produced a higher number of neurons which was around 200 (Yu et al., 2013).

Furthermore, in the adult brain, neurons from the same lineage were found to extend their projections in a bundle as during development (Yu et al., 2013) (Ito et al., 2013). In the majority of lineages, one or two fiber bundles were identified, which is reminiscent of the binary fate decision diversification mechanism (Ito et al., 2013).

After extending together in tracts, neurons from the same clone were found to heterogeneously arborize into multiple neuropils (Ito et al., 2013). Interestingly, more than 40 lineages were found to arborize into 15 or more neuropils, and just 7 were identified to extend in 5 or less (Yu et al., 2013). This showcases the complexity of the clonal organization into functional networks, and abolished the idea that NBs were fully modular units and contributed to a single neuropil (Yu et al., 2013).

Identification of individual neuroblast lineage markers

Understanding how a small set of neural progenitors generates the adult CNS and the complex underlying behaviour will require a deep knowledge on all sets of genes driving the development of each lineage. Previous research efforts have allowed the creation of gene marker atlas for embryonic NBs (Urbach, 2003b) (Broadus *et al.*, 1995) (Doe, 1992) (Urbach *et al.*, 2016). Furthermore, by providing genetic tools to identify restricted groups of lineages, this also yielded the opportunity to track NBs all the way into the second wave of neurogenesis (Lacin and Truman, 2016). Interestingly, Lacin *et al.* showed that multiple TFs found to be heterogeneous between embryonic NB lineages were also found to be expressed in larval lineages (Lacin and Truman, 2016). However, until recently, the analysis of NB lineage heterogeneity had not been done at the whole-genome level. In 2016, Yang *et al.* created a robotic picking device to sort NBs of three specific larval lineages and subsequently performed Bulk RNAseq (Yang *et al.*, 2016). These were the Mushroom body, Antennal lobe and Type II NB lineages. Interestingly, a number of TFs were identified to be heterogeneously expressed in each lineage (Yang *et al.*, 2016). Additionally, the authors identified genes which were in lower levels in these three lineages, but enriched in Bulk RNAseq data of all NBs together (Yang *et al.*, 2016). This prompts the use other methods which could further disentangle the heterogeneity of all NB lineages.

Resolving cellular diversity with single-cell transcriptomics

The high-throughput measurement of RNA levels has been broadly applied as a methodology to identify novel potential players regulating a multitude of biological phenomena (Tan et al., 2015) (Atkins et al., 2016) (Alyagor et al., 2018). This was first achieved through the use of bulk RNA sequencing strategies, where the output is an average expression value for all genes in a given group of cells. However, a major drawback to this is that the information on genes which are

heterogeneously expressed is lost (Chen *et al.*, 2019). On the other hand, single-cell RNA sequencing (scRNAseq) allows this data to be retrieved by sequencing the RNA of each cell. In the past scRNAseq experiments have been largely limited by technical obstacles. Each sequencing library had to be prepared from a single cell isolated per well. This created expensive costs in consumables and laborious protocols. In recent years there have been a number of significant improvements which resulted in an increased use of this strategy. One of these improvements has been the introduction of barcoding technologies, which allows to pool all cells together, and run the reactions for sequencing library preparation in just one tube (Svensson *et al.*, 2018). This has contributed to the rise of the limit of cells sequenced per sample, which is currently in the thousands (Svensson *et al.*, 2018) (Salomon *et al.*, 2019).

A popular methodology for scRNAseq are droplet-based microfluidic procedures (Salomon *et al.*, 2019). Here, single cells are individually distributed into water-in-oil droplets through the use of microfluidic systems (Fig. 5A) (Salomon *et al.*, 2019). Inside each droplet it is also inserted a barcoded bead which has thousands of oligonucleotides attached (Fig. 5A). Each of these contain three main sequences which are critical for scRNAseq library preparation (Fig. 5B). The first sequence is the cell barcode which identifies the cell introduced into the droplet. This sequenced must be different between barcoded beads, but shared between the oligonucleotides attached to the same bead (Salomon *et al.*, 2019). Moreover, every oligonucleotide will also contain a sequenced termed the unique molecular identifier (UMI), which as the name suggests should be different for every oligo. This later makes it possible to identify the number of raw RNA counts for the multiple genes (Salomon *et al.*, 2019). Finally, all oligos have a poly-d(T) sequence to hybridize the mRNA released from cells upon membrane lysis. Inside each droplet all necessary reagents for reverse transcription are also packaged. After the induction of this reaction, all mRNAs should be converted into cDNA and include the cell barcode and UMI sequences (Salomon *et al.*, 2019).

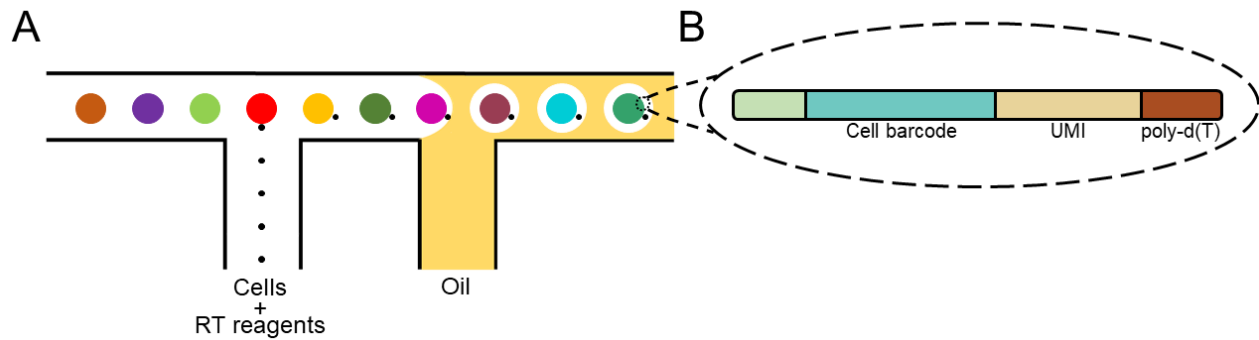


Fig. 5 Droplet-based microfluidic scRNAseq procedures

A) Each cell is put inside a water-in-oil droplet along with a barcoded bead and all reagents for reverse transcription reaction. B) Schematic of the critical sequences for scRNAseq library preparation in the oligonucleotides of barcoded beads. The first sequence identifies the cell (cell barcode), while the unique molecular identifier (UMI) will identify a single transcript. The poly-d(T) sequence is necessary for hybridizing to the cell's mRNA and initiating reverse transcription (Salomon *et al.*, 2019). This figure was adapted from the 10XGenomics user guide (10XGenomics, 2020).

Study hypothesis and aims

The identification of NB heterogenous genes by the cornerstone works of Desplan, Doe, Urbach, Technau, *et al.* have been pivotal for the understanding of the genesis of neuronal diversity. However, due to the nature of the methodology of these studies, based in expression screens, the number of genes detected is limited to the number of existing reagents. It is thus possible that additional undiscovered genes, such as transcription factors, can distinguish NB lineages and further explain the origin of CNS cellular diversity. In *Drosophila*, both the adult brain and VNC have been single-cell sequenced and it has been possible to retrieve previously known cell types and to identify new marker genes (Konstantinides *et al.*, 2018) (Croset *et al.*, 2018) (Davie *et al.*, 2018) (Allen *et al.*, 2020). Revisiting neuronal development through the eye of single-cell technologies is thus very tempting. Recently, our group has performed scRNAseq of 3rd instar larval NB lineages (Marques *et al.*, unpublished). The aims of this thesis were to characterize NB lineages at the single-cell level and to search for additional transcriptional heterogeneity between larval NBs and neurons. For this, we have optimized a bioinformatic pipeline for the analysis of this dataset and validated the expression of a few identified marker genes. With this work we establish the framework for future studies aiming at exploring the assembly of the complex and diverse nervous system.

Materials and methods

Table 1: Material list, respective sources, and RRID identifiers.

Materials	Source	RRID identifier
Reagents		
Aqua-Poly/Mount coverslipping medium	Polysciences	N/A
Paraformaldehyde	Sigma-Aldrich	N/A
Phosphate-buffered saline	Invitrogen	N/A
Normal goat serum	Jackson Immunoresearch	RRID: AB_2336990
Triton X-100	Sigma-Aldrich	N/A
Vienna food		
Barley Malt Syrup (80g)	Próvida	N/A
Beetroot Syrup (22g)	Grafschafter	N/A
Agar (8g)	NZYTEch	N/A
Biological Corn flour (80g)	Próvida	N/A
Soya flour (10g)	A. Centazzi	N/A
Instant Yeast (18g)	Saf-instant	N/A
Propionic acid (8ml)	Argos	N/A
15% Niapagin in 96% EtOH (12ml)	Tegosept, Dutscher UK	N/A
Antibodies		
Chicken anti-GFP (1:100)	Abcam	RRID:AB_300798
Rabbit anti-Miranda (1:1000)	Eurogentec (non comercial)	N/A
Mouse anti-Elav (1:100)	DSHB	RRID:AB_528217
Rat anti-Elav (1:100)	DSHB	RRID:AB_528218
Goat Alexa fluor 488 anti-chicken (1:1000)	Invitrogen	RRID:AB_142924
Goat Alexa fluor 568 anti-rabbit (1:1000)	Invitrogen	RRID:AB_143157
Goat Alexa fluor 647 anti-rat (1:1000)	Invitrogen	RRID:AB_141778
Goat Alexa fluor 647 anti-mouse (1:1000)	Invitrogen	RRID:AB_2535804
Drosophila lines		
y[1] w[*]; P{y[+t7.7] w[+mC]=croc-GFP.FPTB}attP40/CyO	BDSC	RRID:BDSC_83419
w[1118]; PBac{y[+mDint2] w[+mC]=fkh-GFP.FPTB}VK00037/SM5	BDSC	RRID:BDSC_43951
w[1118]; PBac{y[+mDint2] w[+mC]=grn-GFP.FPTB}VK00037	BDSC	RRID:BDSC_58483
y[1] w[*]; P{y[+t7.7] w[+mC]=mirr-GFP.FPTB}attP40	BDSC	RRID:BDSC_68183
y[1] w[*]; PBac{y[+mDint2] w[+mC]=Optix-GFP.FPTB}VK00031	BDSC	RRID:BDSC_83422
w[1118]; PBac{y[+mDint2] w[+mC]=run-GFP.FPTB}VK00033/TM6C, Sb[1]	BDSC	RRID:BDSC_64827
y[1] w[*]; P{y[+t7.7] w[+mC]=Six4-GFP.FPTB}attP40/CyO	BDSC	RRID:BDSC_67733

Fly husbandry

All protein traps used in this work were created by the ModERN project (Kudron *et al.*, 2018) and ordered from the Bloomington *Drosophila* Stock Center. Flies were cultured in wide vials with Vienna recipe food, inside humidity and temperature-controlled (25°C) incubators.

Immunohistochemistry and confocal imaging

Each genotype was allowed to lay eggs in Vienna recipe food, progeny hatched and developed until the 3rd instar wandering stage. Subsequently, larvae were collected into small petri dishes with ice-cold 1x Phosphate-buffered saline (PBS). Dissection was performed by first removing just the posterior side of the larvae which included the gut, trachea and fat body. Afterwards, larvae were

fixed in paraformaldehyde 4% for 30 mins, which was followed by a 0.1% PBT (Triton X-100, PBS 1x) wash for permeabilization. Next, blocking (0,5% normal goat serum, 0,5% Triton X-100, PBS 1x) was performed for 1 hour at room temperature (RT) and then larvae were left overnight at 4°C with primary antibodies and blocking. In the next day, specimens were washed, and blocking was performed again for 1 hour. Secondary antibodies and blocking were then added, and larvae were left in the dark at RT for 4 hours. This was followed by PBT and PBS washings. The CNS was then fully dissected and mounted with Aqua-Poly/Mount. Slides were left to dry and kept at 4°C until imaging. Confocal imaging was performed in a Carl Zeiss LSM880; Z-Stacks were acquired of all imaged brains. 5 brains were analysed per genotype.

Bioinformatic analysis of single-cell transcriptomic data

Seurat object creation and filtration

For data quality control and filtration, dimensionality reduction, clustering, and visualization we used the R Seurat package 3.1.5 (Macosko *et al.*, 2015) (Satija *et al.*, 2015) (Satija lab, 2020). Output filtered gene count matrices from the 10XGenomics processing pipeline Cellranger were used as input for '*Read10x*' function to introduce data into R. Subsequently, the Seurat object was created through '*CreateSeuratObject*'. The two samples of this study were loaded independently and afterwards merged through '*MergeSeurat*'. Cells with < 200 expressed genes or which had ≥ 20 percent of transcripts from mitochondrial genes were filtered out to remove potentially low-quality cells. Additionally, genes which were barely expressed, < 3 cells, were also removed.

Seurat object pre-processing

To neutralize sequencing depth biases, for each cell, the counts of each gene were normalized by all counts of the cell, multiplied by a scale factor 10 000, and afterwards log-transformed (Satija lab, 2020). This was applied through '*NormalizeData*'. The top 2000 highly variable genes were then determined through '*FindVariableFeatures*'. The expression of all genes was then centred, scaled, and mitochondrial gene expression heterogeneity was regressed out as we considered this "uninteresting variability" (Satija lab, 2020). This was achieved through the function '*ScaleData*'.

Dimensionality reduction, clustering analysis and major cell groupings

The following steps were performed only on the top 2000 highly variable genes identified above, as it has been suggested to "highlight" the important variability (Satija lab, 2020) (Luecken and Theis, 2019). To understand which gene sets explain the majority of the observed heterogeneity between cells, we ran a principal component analysis (PCA) and the Jackstraw permutation test (Satija lab, 2020). With this approach, it was possible to "summarize" our dataset into 51 dimensions (Suppl. Fig. 1) (Luecken and Theis, 2019). In the Jackstraw, we discarded the first PC to not be significant

(p-value < 0.05) and all subsequent PCs. Next, using PCA Euclidean distances, a K-Nearest Neighbour graph where each node is a cell was constructed. This was followed by the generation of a Shared Nearest Neighbour graph. This was achieved through the function '*FindNeighbors*'. From this graph, we used the Louvain community detection algorithm for clustering. The latter was employed by the function '*FindClusters*'. Here, we had to specify a value for a parameter called 'resolution', which controls the subdivision of clusters. We decided to use a 'resolution' of 1.55, given that it allowed the proper separation of the expected major cell types in our analysis. Next, we utilized the Uniform manifold approximation and projection (UMAP) algorithm to visualize the clusters in two-dimensional space (McInnes *et al.*, 2018) (Satija lab, 2020). Finally, we grouped clusters into 8 major cell types: Type I NBs (cluster 0), Type II NBs (cluster 36), dINPs (cluster 34), INPs (cluster 22), imGMCs (cluster 9), Type I GMCs (clusters 1, 3, 17 and 32), Type II GMCs (cluster 15), and neurons (clusters 2, 4, 5, 6, 7, 8, 11, 12, 13, 14, 16, 18, 19, 20, 21, 24, 26, 27, 28, 29, 30, 31, 35, 37). The identity of each cluster was identified by the observation of the expression of cell type specific markers (*dpn*, *mira*, *klu*, *erm*, *pnt*, *Sp1*, *ase*, *pros*, *elav*) (Suppl. Fig. 2). Additionally, there were also a few clusters which showed indicators of low quality cells (clusters 10, 23 and 25). Finally, a few clusters were not annotated as they did not show an obvious expression of the cell type-specific markers above (clusters 33 and 38) (Suppl. Fig. 2).

Identification of cluster marker genes

For the identification of marker genes both the functions '*FindMarkers*' and '*FindAllMarkers*' were used. Both are similar, however '*FindMarkers*' allows the identification of marker genes between specific clusters, and '*FindAllMarkers*' instead compares a given cluster versus all other cells. Here, Seurat allows the use of tests to identify differentially expressed genes. We used the ROC (Receiver operating curve) test which provides a "classifier" AUC (area under the ROC curve) for each gene (Satija lab, 2020). As explained by the Seurat documentation, this value informs about the potential of a specific gene as a marker. If this value is 1, this gene is a perfect positive marker for a given cluster. On the other hand, if this value is 0 this gene is considered to be the perfect negative marker. In other words, if this value is 1, all cells in the cluster evaluated express the gene at higher levels than the other counterpart. The opposite for negative markers. If a marker has an AUC value of 0.5 it cannot distinguish the populations compared (Satija lab, 2020). We only kept genes which have a AUC > 0.5 and an average log fold change > 0.25. Additionally, genes which were not expressed within 25% of the cells in either of the populations compared were also discarded. Finally, in the top marker plots for the NB subclustering and neuronal cluster analysis, only are shown genes which are expressed in more than 50% of the cells of the cluster analysed (pct.1) and expressed in less than 20% of the rest of the cells (pct.2).

Type I NB subclustering analysis

Cluster 0 (NB type I) was subsetted and cells with ≤ 3 counts of *dpn* were removed to avoid cluster contamination. Highly variable features were identified again, followed also by scaling and PCA. To select the number of PCs, we determined the significant PCs through the ‘*JackStraw*’ function and inspected the top correlated genes in each PC through ‘*DimHeatMap*’ (Satija lab, 2020). PC1, 2 and 4 had multiple genes which could potentially account for unimportant variation for this study, such as heat shock, ribosomal, mitochondrial genes, and also genes related to somatic sex differences (Suppl. Fig 4A). We therefore decided to use the top 30 statistically significant PC as chosen above, and additionally excluded PC1, 2 and 4 (Suppl. Fig 4B). K-Nearest Neighbour and Shared-Nearest Neighbor graph was constructed as previously mentioned. As for the resolution parameter in ‘*FindClusters*’ we experimented from 0.1 to 5. Although results did not vary much, at higher resolutions we were able to identify subtle differences between NBs which we later verified by immunostaining. This parameter was fixed at 2.9 since it allowed the identification of a marker *croc*, which was also confirmed by immunostaining to be heterogeneous in type I NBs.

Gene ontology set enrichment analysis

Gene ontology set enrichment analysis was performed on the marker genes of all neuronal clusters and on the marker genes of all Type I NB subclusters. Marker genes which had an AUC value > 0.5 and an average log fold change > 0.25 were loaded into the FlyEnrichr web tool (Kuleshov *et al.*, 2019). Results were shown in bar graphs which indicate the combined score for a given gene ontology (GO) term (Kuleshov *et al.*, 2019).

Gene annotation lists

Gene lists for TF/Deoxyribonucleic acid (DNA) binding and ion channel genes were retrieved from the Gene List Annotation for *Drosophila* (GLAD). Gene list for the cadherin super family was taken from Flybase under the gene group “CADHERINS” - FBgg0000105. The immunoglobulin super family genes were retrieved from the publication (Vogel, 2003) through Flybase - FBrf0167517. Gene symbols from all lists were updated to match the ones in our analysis.

Plots generated outside Seurat

The Venn diagram was produced through an online tool (molbiotools) and shows the overlap of TF/DNA binding marker genes identified in the type I NBs vs. type I GMCs and type II NBs vs. dINPs comparisons. The chord diagrams have been generated through the package ‘*circlize*’ (Gu *et al.*, 2014) and were adapted and used as described in (Allen *et al.*, 2020) to show the correspondence between clusters and cluster marker genes.

Results and discussion

To gain a deeper insight into neural stem cell development and the generation of neuronal diversity, our group has recently done scRNAseq of 3rd instar larval NB lineages (Marques *et al.*, unpublished). In order to isolate these cells, a Gal4 fly line with restricted expression to the CB and VNC NB lineages (Vienna tiles line 201094) was crossed to *UAS-CD8::GFP; UAS-Gal80ts* line. The descendants were then put at 18°C until 87 hours after larval hatch and subsequently shifted to 25°C for 18 hours to activate GFP expression. This conditional genetic system allowed GFP to label only type I lineage NBs, GMCs and Neurons, and Type II lineage NBs, INPs and GMCs. GFP was expressed in over half of CB and VNC NB lineages in the larval CNS. Cells were then fluorescence-activated cell sorted and single cells libraries were prepared using 10X Chromium technologies and subsequently sequenced (Marques *et al.*, unpublished). Two independent samples were obtained through this strategy. In this work, a bioinformatic pipeline was optimized for the analysis of this dataset and for the construction of an atlas for larval NB lineages.

Sample quality control and filtering

The structures of both samples in this study were first evaluated through the use of Seurat's quality control metrics. These were the number of expressed genes (nFeature_RNA), the number of RNA molecules (nCount_RNA), and the percentage of mitochondrial reads (percent.mt) per cell (Satija lab, 2020). These metrics were shown to be similar between both samples (Fig.6A). Interestingly, it was possible to distinguish two cell populations in terms of the number of genes expressed per cell (Fig.6A and B) (discussed ahead). Next, to discard potentially dying or under stress cells, two filters were applied. Cells which expressed a low number of genes and had high mitochondrial content were removed. The idea behind this assumption is that cell membrane rupture could lead to cytoplasmatic RNA depletion and enrichment for mitochondrial RNA (Illicic *et al.*, 2016). Given this, 94 cells were eliminated. Additionally, genes which were barely expressed, in less than 3 cells, were also removed. In the end, a dataset of 12671 cells, with a mean of 1635 genes expressed was achieved after the use of these filters (Fig. 6B).

Clustering analysis identifies NB lineage cell types

To identify cells with a similar transcriptomic profile a clustering analysis was performed. Using PCA and Seurat's graph-based clustering approach (Satija lab, 2020) it was possible to identify 39 clusters with distinctively expressed marker genes. These results were then visualized in low dimensional space using the UMAP algorithm (Fig.7A) (McInnes *et al.*, 2018) (Satija lab, 2020)

It has been previously shown that a sample-bias upon clustering could exist. This phenomenon termed batch effects is the consequence of technical variability introduced for example by mRNA capture efficiency by barcoded beads' oligonucleotides (Tung *et al.*, 2017). To understand if batch

effects could be influencing the clustering results, cells were labelled by sample in the UMAP plot. Cells appeared to distribute evenly through the different clusters independently of their sample origin, suggesting the absence of batch effects (Fig. 6C).

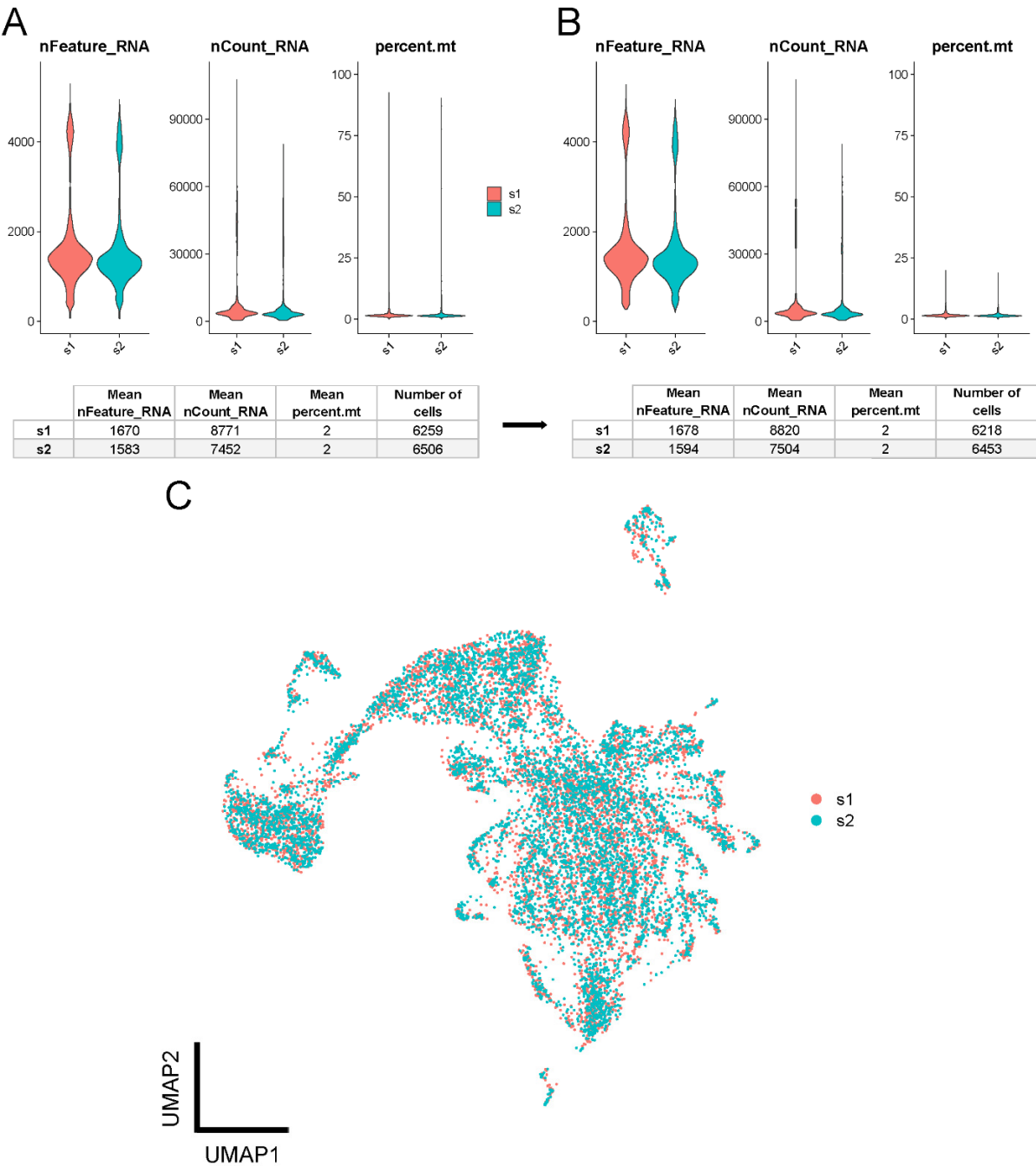


Fig. 6 Quality control and dataset filtering

A) Seurat's quality control metrics for both samples (s1 and s2) in the unfiltered dataset and **B)** after filtration. Top: Violin plots show the cells' distribution according to the number of genes expressed (nFeature_RNA), the number of RNA molecules (nCount_RNA) and the percentage of mitochondrial reads per cell (percent.mt). Bottom: table summarizes the mean for each of these metrics and the number of cells per sample. **C)** UMAP graph shows the distribution of cells from the two different samples.

Through the analysis of the cluster expression of known cell-type specific markers it was possible to group clusters into 8 major cell types, where all characteristic NB lineage cell types were identified (Fig. 7B, 7C).

Type I NBs expressed the markers *dpn*, *mira*, *klu*, *ase* and *wor* (Fig. 7C). Type II NBs were identified by similarly expressing these genes in the great majority of cells, with the exception of *ase* (Fig. 7C), a characteristic feature which has been shown to be critical for the maintenance of type II proliferation mode (Bowman *et al.*, 2008). Additionally, these cells also expressed *pnt* (Fig. 7C) which was previously described to be necessary for *ase* negative regulation in type II NBs (Zhu *et al.*, 2011).

Furthermore, we identified a group of cells which correspond to a mixture of immature and young mature INPs, which we termed developing INPs (dINPs). This cluster expressed *pnt* and *erm* (Fig. 7C) which has been shown *in vivo* to be expressed in both these cell populations (Zhu *et al.*, 2011) (Janssens *et al.*, 2014). Additionally, there was a cluster which localized close to dINPs in the UMAP plot, which expressed *dpn*, *mira*, *klu*, and *ase*, but only a low percentage of these cells expressed *erm* and *pnt* (~30% express *erm* and ~16% express *pnt*) (Fig. 7C). Indeed, *pnt* and *erm* have been shown to be downregulated as INPs age (Zhu *et al.*, 2011) (Janssens *et al.*, 2014), and thus we believe that these cells are older INPs, after the entire maturation period, which we have named INPs.

Moreover, we identified clusters which correspond to GMCs. These clusters showed depletion of *dpn* and *mira* expression, but expressed *ase* and *wor* (Fig. 7C). Interestingly, one of the GMC clusters showed expression of *Sp1* (Fig. 7C) which has been recently described to be restricted to type II lineages (Rives-Quinto *et al.*, 2020), and thus most likely are type II lineage GMCs. Multiple other clusters were annotated as neurons and were characterized by the depletion of stemness related genes *dpn*, *mira*, and *wor*, and expression of pro-neural genes *pros* or *elav* (Fig. 7C). Expectedly, cluster UMAP disposition greatly recapitulated NB lineage progression order *in vivo* (NB→ GMCs→ Neurons) (Fig. 7B).

It was also possible to identify a group of cells which seem to be in a transition state between type I NBs and GMCs, which we named immature GMCs (imGMCs). Roughly two-thirds of these cells no longer express the TF *dpn* critical for neural stem cell self-renewal, but nearly all still express the NB marker *Mira* (Fig. 7C).

Surprisingly, we did not identify any cluster enriched for the glial cell marker *reversed polarity* (*repo*). Previous work has shown that gliogenesis can be associated with specific temporal windows (Doe, 2017). It is thus possible that at 105 ALH there is no glia being formed.

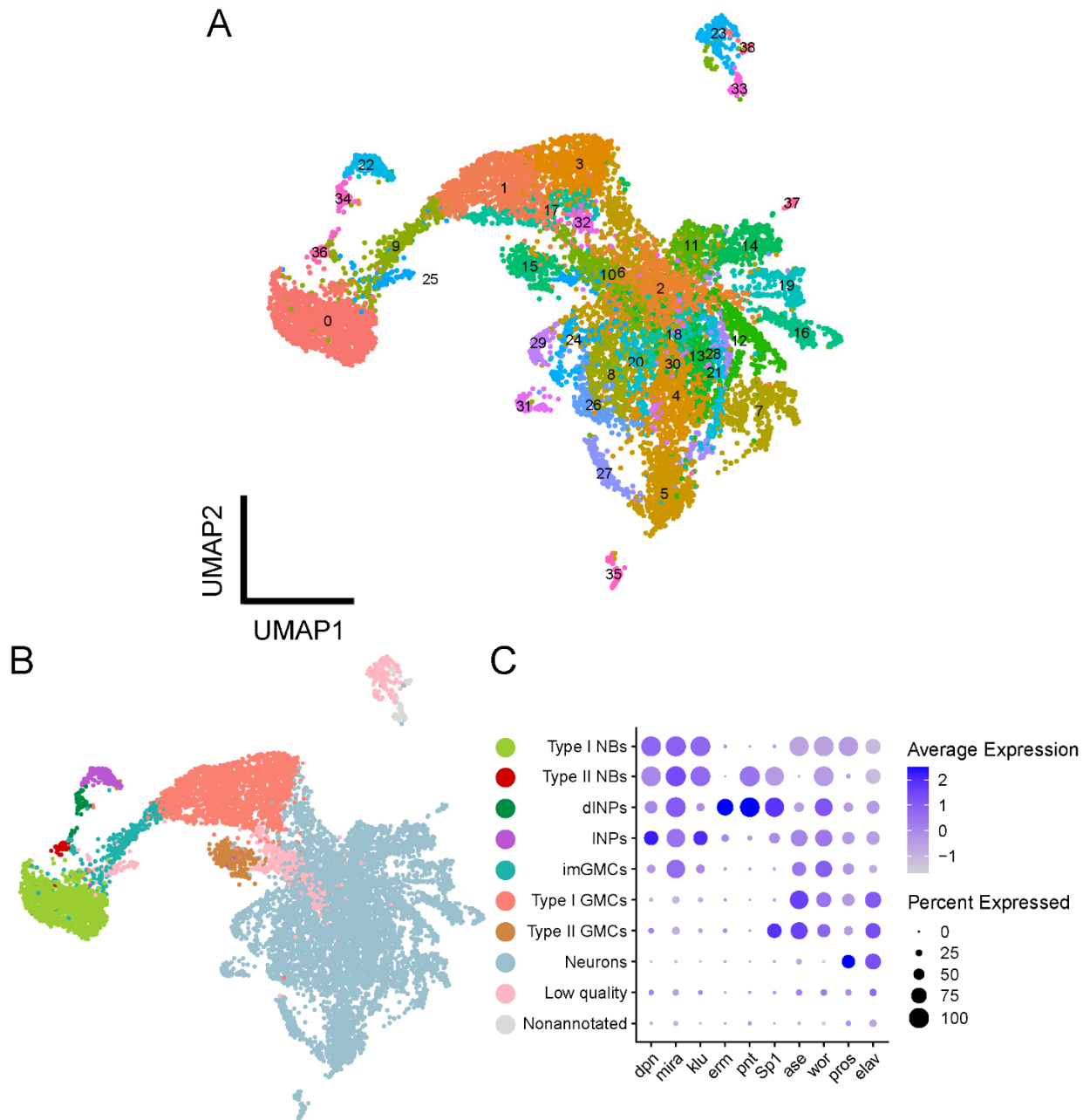


Fig. 7 Clustering analysis identifies NB lineage cell types and a transition cellular state
A) UMAP plot shows the 39 clusters identified **B)** UMAP plot with cells colored by major cluster grouping. Type I NBs – yellow green; Type II NBs – red; dINPs – spring green; INPs – violet; imGMCs – sea green; Type I GMCs – salmon; Type II GMCs – brown; Neurons – light blue; Low quality – pink; Nonannotated – gray. **C)** Expression of NB lineage cell type specific markers (*dpn*, *mira*, *klu*, *erm*, *pnt*, *Sp1*, *ase*, *wor*, *pros*, *elav*) across the different major cluster groupings. Major cell group average scaled expression level is indicated by the colour scale. Dot dimension informs about the proportion of cells per major group that express the NB lineage cell type specific markers.

Moreover, clusters which did not show an obvious expression of the previously mentioned cell type-specific markers were not annotated. Finally, there were also a few clusters which seem to correspond to low quality cells. Cluster 10 and 23 had several mitochondrial and ribosomal proteins in the top cluster markers, respectively (data not shown), which could potentially indicate cell stress. Part of cells in cluster 23 also expressed *Ninjurin A* (NijA) (data not shown), which has been shown by Broderick *et al.* to be found in higher levels after injury and to cause non-apoptotic cell death (Broderick *et al.*, 2012). Additionally, some cells in cluster 25 expressed *mira* and *dpn*, but this cluster together with cluster 10 were the ones with less RNA molecules and number of genes expressed (data not shown). It is thus possible that these cells could represent cells with high-dropout levels, and thus they were also categorized as low quality cells.

When analysing the quality control metrics for the different major cell types, neural stem cells (NBs type I and II) were shown to express a higher number of genes (nFeature_RNA) than more differentiated cell types (Fig. 8A). This is most likely the result of a higher mitotic activity. Indeed, this higher number of genes in NBs was responsible for the second cell population observed above (Fig. 6A). Furthermore, neural stem cells also had more RNA molecules (Fig. 8B), which possibly is a direct consequence of the expression of a broader set of genes. Interestingly, similar results were also observed in a single-cell study of the *Drosophila* adult ovary. Germ stem cells also expressed a wider range of genes and had more RNA molecules than their somatic descendants (Slaidina *et al.*, 2020).

Analysis of gene expression changes along NB lineage differentiation and proliferation mode

To better characterize the molecular players driving NB lineage development, we analysed the differentially expressed genes between the major groups previously identified. Type I NBs and its descendants GMCs were compared first. Analysis of the TFs/DNA binding (TFs/DNAB) proteins marker genes that differentially characterize these two clusters identified multiple factors that had already been described to play a role in NB lineage development, confirming the quality of our approach. Among the identified genes by our analysis are *grainyhead* (grh), *klu* and *E(spl)mgamma-HLH*, all been shown to be important for NB self-renewal (Almeida and Bray, 2005) (Berger *et al.*, 2012) (Zacharioudaki *et al.*, 2012). In accordance, all these players were found to be enriched in type I NBs, but depleted in type I GMCs (Fig. 9A). On the other hand, genes which have been demonstrated to promote lineage progression and neuronal differentiation were shown to be upregulated in type I GMCs. These included genes such as *jing interacting gene regulatory 1* (jigr1) and *longitudinals lacking* (lola) (Yuva-Aydemir *et al.*, 2015) (Southall *et al.*, 2014) (Fig. 9A). In this major cell type comparison, we also identified novel genes, with uncharacterized functions in NB lineages, such as *CG2051*, *CG30403*, *CG4570* and *Motif 1 Binding Protein* (M1BP) (Fig. 9A). The

post-translational modification of trimethylation on histone 3 lysine 9 (H3K9me3) is widely associated with heterochromatinization (Ninova et al., 2019). This histone mark is known to recruit Heterochromatin protein 1 (HP1), which plays an important role in heterochromatin compaction and spreading (Ninova et al., 2019) (Ninova et al., 2020), and thus is important for epigenetic gene silencing. Interestingly, *CG30403*, which is part of the family of MADF-BESS transcriptional regulators (Shukla et al., 2014), has been shown to physically interact with the Heterochromatin protein 1 (HP1) (Alekseyenko et al., 2014) (Ninova et al., 2020). Recently, Marshall and Brand have profiled the genomic occupancy of known epigenetic key players in specific NB lineages to get a better appreciation of how chromatin states vary during fate acquisition (Marshall and Brand, 2017). Interestingly, they have found that in mature neurons, the genes responsible for stem cell programs, such as *dpn*, *Mira*, *E(spl)mgamma*, where in the great majority occupied by HP1 (Marshall and Brand, 2017). This suggests that neural stem cell programs are mainly epigenetically silenced through H3K9me3 and HP1. Given that in our dataset we found *CG30403* to be enriched in NBs, and that the product of this gene interacts with HP1, it would be interesting to test if this gene could be important for the maintenance of stem cell programs through a negative interaction with HP1.

Genes responsible for asymmetric division, like *mira* and *staufer* (*stau*) (Li et al., 1997), were found to be upregulated in type I NBs (Suppl. Fig. 3A). Conversely, *insensible* (*insb*), which has recently been suggested to post-translationally reduce the activity of *dpn* to ensure termination of the stem cell program and differentiation commitment (Komori et al., 2018), was shown to be upregulated in type I GMCs (Suppl. Fig. 3A). We also identified two long non-coding RNAs (lncRNA) in the top 30 markers in the comparison of type I NBs and type I GMCs. One was the lncRNA *cherub* (CR43283), which was present in higher levels in type I GMCs (Suppl. Fig. 3A). Recently, Landskron et al. showed that *cherub* is asymmetrically split in the NBs to the descendants and regulates type II lineage tumours (Landskron et al., 2018). The other was an unstudied lncRNA CR31386, which we found to be enriched in type I NBs (Suppl. Fig. 3A).

We next compared type II NBs and descendants dINPs (Fig. 9B). Interestingly, we found that there was a large overlap between the TFs/DNAB markers identified in Type I NBs vs. Type I GMCs and Type II NBs vs. dINPs comparisons (Fig. 9C). This suggests that many lineage developmental programs are conserved between the two proliferation modes. Furthermore, when analysing the markers specifically upregulated in dINP vs. Type II NBs (dINP>NBII) it was possible to identify genes such as *Sp1*, *erm*, *pnt* and *D* (data not shown), all of which have been described to have a role in type II lineages (Álvarez and Díaz-Benjumea, 2018) (Weng et al., 2010) (Zhu et al., 2011) (Bayraktar and Doe, 2013).

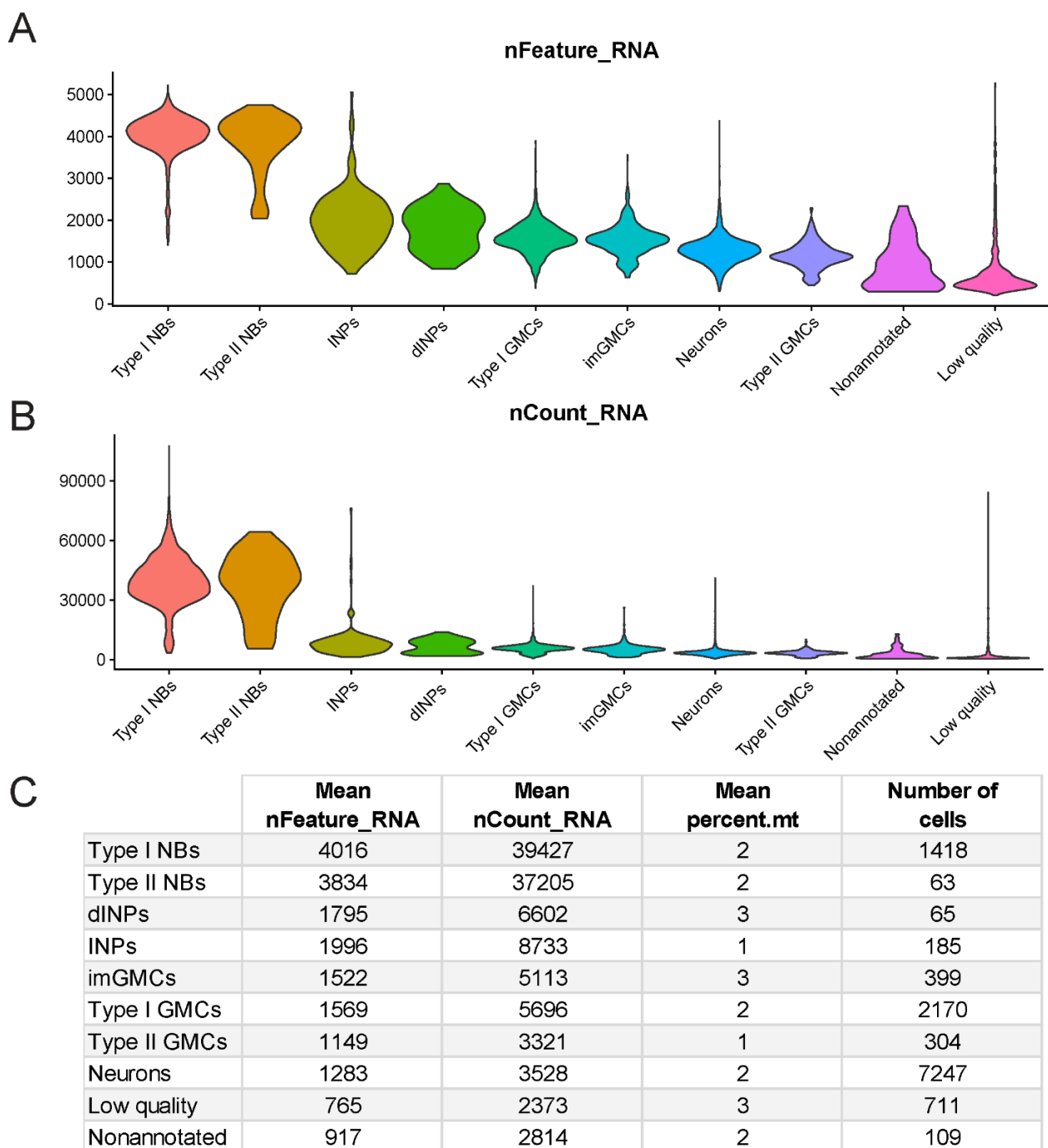


Fig. 8 Quality control for the major cluster groups

A) Violin plots show the cells' distribution according to the number of genes expressed (nFeature_RNA) and the number of RNA molecules (nCount_RNA) for the major cluster groups. Major cluster groups were organized in descending scale order (left to right). **B)** Table summarizes the quality control metrics mean and number of cells of the major cluster groups.

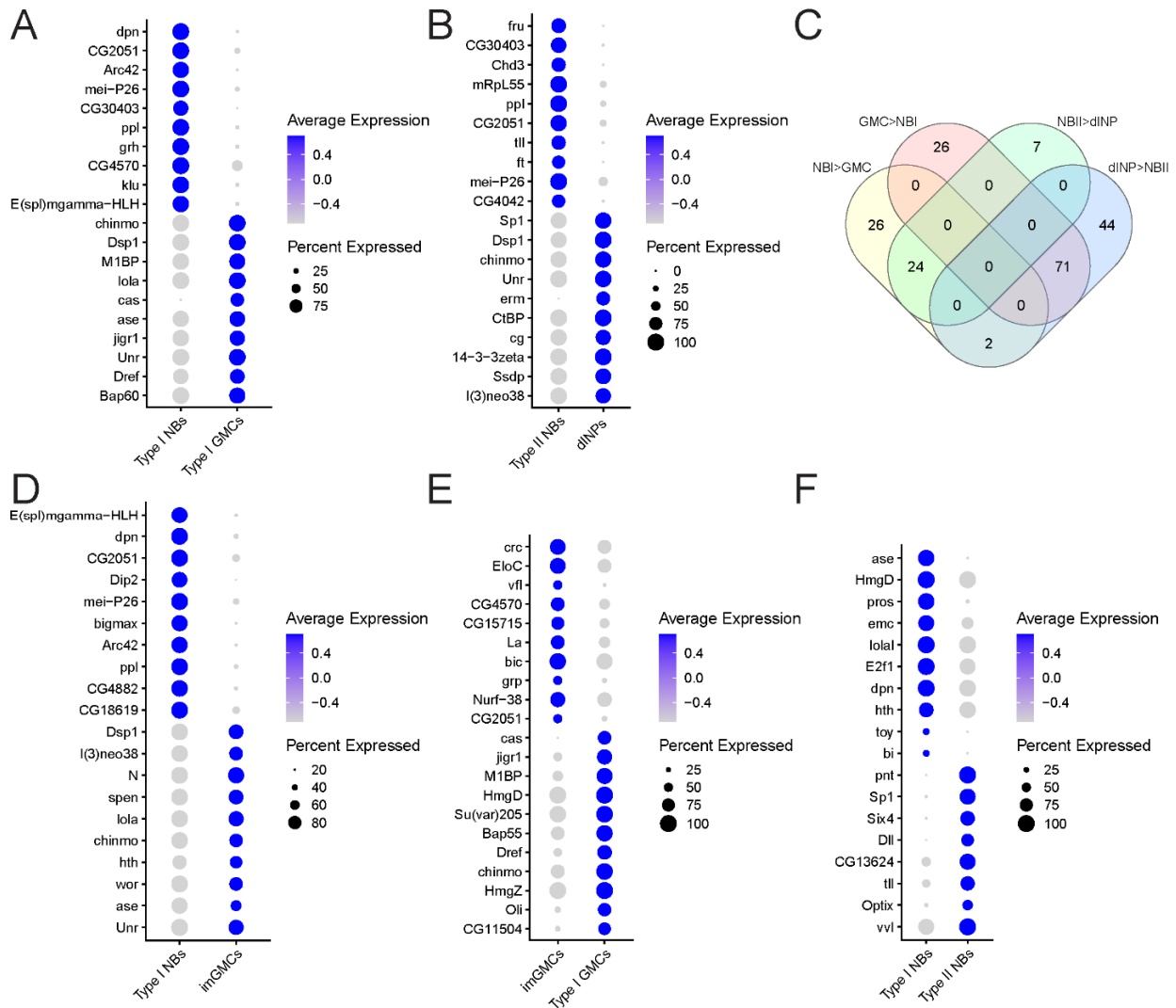


Fig. 9 Major cell group comparison identifies gene expression changes across NB lineage development and proliferation mode

Dot plots show the top 10 TF/DNA binding protein marker genes for each major cell group comparison. The following comparisons were made **A**) Type I NBs vs. Type I GMCs **B**) Type II NBs vs. dINPs **D**) Type I NBs vs. imGMCs **E**) imGMCs vs. Type I GMCs **F**) Type I NBs vs. Type II NBs. Major cell group average scaled expression level is indicated by the colour scale. Dot dimension informs about the proportion of cells per major cell group that express a certain gene. **C**) Venn diagram shows the overlap of TF/DNA binding protein marker genes identified to be upregulated in Type I NBs vs. Type I GMCs (NBI>GMC), upregulated in Type I GMCs vs. Type I NBs (GMC>NBI), upregulated in Type II NBs vs. dINPs (NBII>dINP), and upregulated in dINPs vs. Type II NBs (dINP>NBII).

Subsequently, we characterized the previously unidentified NB to GMC transition state, imGMCs. When compared to type I NBs, genes such as *E(spl)mgamma-HLH*, *CG2051* and *meiotic P26* (*mei-P26*) were found to be downregulated in imGMCs (Fig. 9D). *E(spl)mgamma-HLH*, is a Notch target gene from the Hes/ Enhancer-of-split family (Kageyama and Ohtsuka, 1999) (Zacharioudaki *et al.*, 2012). Given that the Notch pathway is inhibited in GMCs after cell division through the asymmetric segregation of Numb, this could explain the quick downregulation of *E(spl)mgamma-HLH* in imGMCs. Furthermore, genes identified to be upregulated in type I GMCs vs. type I NBs, such as

Dorsal switch protein 1 (Dsp1) and *lola*, were already found to be upregulated in the imGMCs when compared to type I NBs (Fig. 9D). Afterwards, we compared imGMCs with type I GMCs and found that genes such as *cryptocephal* (*crc*), *Elongin C* (*EloC*) and *CG4570* to be expressed in higher levels in the imGMCs vs. type I GMCs (Fig. 9E). These genes were also found in higher levels in type I NBs vs type I GMCs (NBI>GMC) (data not shown). Interestingly, in a genetic screen to identify molecular players involved in NB self-renewal, both the knockdown of *crc* and *EloC* driven by *Inscutable-GAL4* (NB driver) resulted in altered phenotypes. Downregulation of *crc* resulted in lethality and in the case of *EloC* lineage defects were identified, such as shorter lineages and larger GMCs, and also lethality (Neumüller *et al.*, 2011). It is thus possible that these genes could have a role in the NB to GMC transition. Moreover, genes which were identified as GMC>NBI enriched genes, such as *jigr1* and *M1BP*, were shown to be in lower levels in imGMCs (Fig. 9E), pertaining to the idea that indeed these cells are in a transitory state.

Lastly, we compared type I and II NBs to try to identify differences between the two proliferation modes. In the group of TFs/DNAB we found that genes such as *ase*, *High mobility group protein D* (*HmgD*), *pros*, *extra macrochaetae* (*emc*), *lola like* (*lolal*), *E2f1*, *dpn*, *hth* were found in higher levels in type I NBs (Fig. 9F). On the other hand, genes such as *pnt*, *Sp1*, *Six4*, *Dll*, *tll*, *Optix* were upregulated in type II NBs (Fig. 9F). These results are consistent with what had been previously observed, except for *HmgD*, *lolal*, *E2f1*, and *dpn* which had been reported as pan-NB markers (Yang *et al.*, 2016). Further experiments are required to validate these results. Furthermore, Yang *et al.* identified *tll* to be depleted in antennal lobe type I NBs, and in higher levels in type II NBs but also in type I MB NBs. Interestingly, *Sp1* has been shown to be important for the initial specification of type II NB lineages in the embryo (Álvarez and Díaz-Benjumea, 2018). The continuous expression throughout development in type II lineages suggests that it could play a role in the maintenance of the type II proliferation mode or to confer unique features to its neural descendants.

In conclusion, here we have shown that a number of genes are differentially expressed between the major cell groups identified in the dataset. Part of these have previously been demonstrated to drive NB lineage development, validating the potential of this approach. It is thus possible that the additional uncharacterized genes identified here could have roles for example in stem cell maintenance or differentiation.

However, although this approach seems to be a solid strategy for identifying discrete changes in gene expression along NB lineage development, it is possible that some information may be lost. For example, what is the order to which genes are activated along differentiation? In the recent years there has been a surge of bioinformatic tools to infer this type of dynamic processes. Strategies for this include ordering cells on the basis of their similarity in order to infer timely changes. This is

termed pseudotemporal ordering (Tritschler *et al.*, 2019). It is thus possible that we could extract more information from this dataset, and which could be attempted in future studies.

Type I neuroblast subclustering analysis unveils transcriptional heterogeneity

NB identity has been shown to play a fundamental role in shaping CNS diversity. To identify differences between type I NBs, these cells were isolated and re-analysed. Strikingly, it was possible to identify 21 type I NB subclusters with different sets of marker genes (Fig. 10A and 10C). Gene ontology enrichment analysis was performed on the markers of all clusters to understand the nature of the genes identified. As for the molecular function, there was a consistent identification of DNA binding and transcription factor activity in the top 6 results (Fig. 10B). Regarding the biological process, there was an enrichment for genes involved in the regulation of transcription, and also for nervous system and neuroblast development (Fig. 10B), which was consistent with the identity of the cells analysed here.

Previous studies have allowed the construction of marker maps for the different NB lineages (Urbach, 2003b) (Doe, 1992) (Broadus *et al.*, 1995) (Lacin *et al.*, 2014) (Urbach *et al.*, 2016) (Lacin and Truman, 2016) and the identification of multiple heterogeneously expressed genes (Brody *et al.*, 2002) (Yang *et al.*, 2016) (Kuert *et al.*, 2014). Of the 78 transcription factor markers that we identified in the type I subclustering analysis, 31 had already been described to be distinctively expressed among NBs (suppl. file 2). This included a variety of spatial patterning genes. The gap genes *ems* and *hkb*, pair-rule genes *runt* (*run*), *slp1* and *slp2*, segment polarity genes *mirror* (*mirr*), *gsb*, *invected* (*inv*), and homeotic genes *Antp*, *lab* were all identified as subcluster markers (Fig 9C and Suppl. file 1 and 2). Additionally, the columnar gene *ind* was also shown to be differentially expressed between type I NBs (Suppl. file 1 and 2). As others have reported, we also identified genes associated with eye formation (Urbach, 2003b). This included *dachshund* (*dac*), *twin of eyeless* (*toy*) and the sine oculis (SIX) family member *Optix* (Fig. 10C and Suppl. file 1). Interestingly, another gene belonging to the SIX family, *Six4*, was also found in the subcluster markers (Fig. 10C). Moreover, besides *slp1* and *slp2*, it was also possible to identify three additional members of the forkhead box TF family. This was *forkhead* (*fkh*), *forkhead domain 96Cb* (*fd96Cb*) and *crocodile* (*croc*) (Fig. 10C and Suppl. file 1 and 2).

Surprisingly, two clusters, 17 and 20, had a similar transcriptomic profile to neurons and GMCs, respectively. For example, cluster 17 had the sodium potassium pump subunit *nervana 3* (*nrv3*) as a cluster marker (Fig. 10C). Cluster 20 had genes such as *jigr1* and *insb* as markers I (Suppl. file 1), both of which were found to be upregulated in GMCs when compared with type I NBs. It is possible that these cells could be type 0 NBs. Given that these progenitors generate neurons immediately after division, genes related to differentiation could potentially be in a poised state in type 0 NBs in order to trigger a fast fate acquisition. Type 0 proliferation mode has not yet been identified in post-

embryonic CB or VNC lineages. Alternatively, even though these cells originally clustered with type I NBs and are *dpn+*, it is possible that these clusters might be the result of a technical artefact. For example, this could have been caused by the generation of doublets. This happens when two cells are included inside the same droplet during scRNAseq library preparation. We identified 37 transcription factor markers which were enriched in these clusters and not in the others.

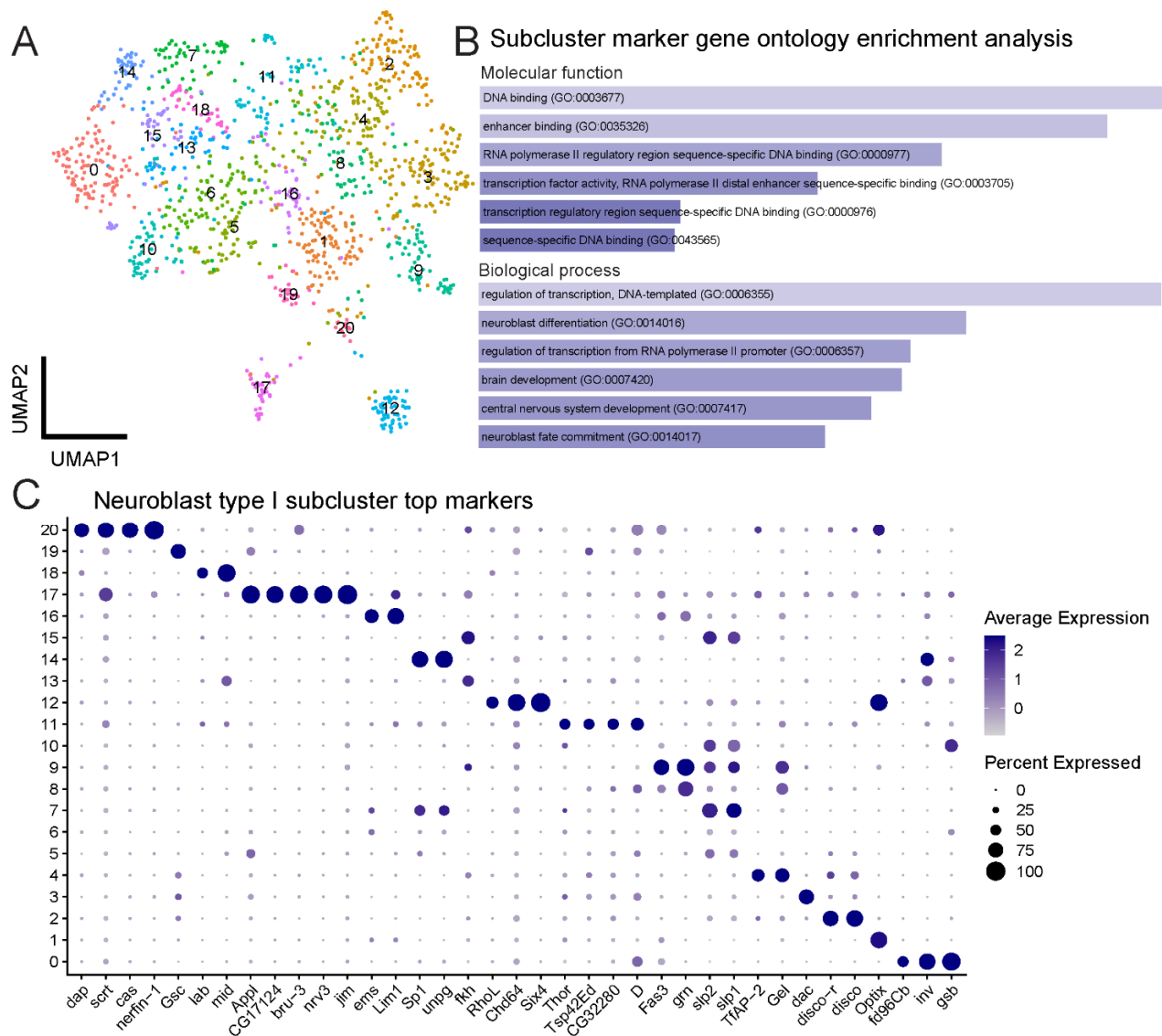


Fig. 10 Subclustering analysis identifies transcriptional heterogeneity within type I NBs

A) UMAP shows the different type I NB subclusters. **B)** Gene ontology enrichment analysis for all subcluster marker genes. Bar graph shows the combined score value for FlyEnrichr gene ontology enrichment analysis. Only the top 6 GO terms are shown for the molecular function and biological process gene ontology. **C)** Dot plot shows the top subcluster marker genes (pct.1 > 0.5 and pct.2 < 0.2, only the top 5). Subcluster average scaled expression level is indicated by the colour scale. Dot dimension informs about the proportion of cells that express a subcluster marker.

Furthermore, we also found 9 transcription factors which, to the best of our knowledge, have not been described to be differentially expressed among NBs (suppl file 1 and 2). This included genes such as *anterior open*, *doublesex-Mab related 99B*, *grain* (grn), *goosecoid*, *no ocelli*, *E2f1*, *proboscipedia* (suppl file 1 and 2).

Moreover, outside of the TF gene group, it was possible to identify membrane protein genes heterogeneously expressed in type I NBs. This included members of the immunoglobulin superfamily such as *defective proboscis extension response 17* (dpr17), *Lachesin* (Lac), *babos*, *Amalgam* (Ama), *Fasciclin 3* (Fas3) (Fig. 10C and Suppl. file 1). It is widely known that this superfamily plays important roles in axonal guidance and wiring specificity (Tan *et al.*, 2015) (Rougon and Hobert, 2003). It is possible that the expression of these molecules is initiated at the stem cell level and inherited by the neuronal progeny to influence the early steps of neuronal development. Additionally, it is also possible that they play role in regulating adhesion between the NB and GMC to maintain lineage structure and positioning in CNS. We also identified the membrane protein gene *wunen-2* (wun2) in the subcluster markers (Suppl. file 1). This lipid phosphate phosphatase has been found to be important for primordial germ cell survival in *Drosophila*, but is also described to be expressed in part of the embryonic NBs (Starz-Gaiano, 2001) (Hanyu-Nakamura *et al.*, 2004). The role of *wun2* in this subset of NBs remains unstudied.

To validate some of the subcluster markers identified here and gain information about the location of the different post-embryonic NB subtypes, we analysed the expression pattern of a number of protein traps. We focused on markers for which tools were available. Fkh, run, croc, optix, grn, mirr and Six4 were all shown to be heterogeneously expressed between larval type I NBs (Fig. 11A, Suppl. Fig 5A). Interestingly, fkh, mirr, run and Six4 were detected in the CB and VNC NBs, while croc, grn and Optix were only detected in CB type I NBs (Fig. 11A and C). Within the CB or VNC we also found an expression spatial bias. For example, croc was expressed in a subset of progenitors on the anterior side of the CB and not on the posterior side (Fig. 11C). Furthermore, we observed type II lineages and examined the expression of these protein traps. Type II lineages were identified by the presence of a large number of small *mira*⁺ cells adjacent to the NBs (INPs). However, given that type I NB lineages also present a few adjacent *mira*⁺ cells, which are most likely imGMCs, without additional markers stained to identify INPs in our samples it was not possible to identify all 8 type II lineages per lobe unambiguously. Due to this we could only assess if a gene was expressed in type II lineages, but could not confirm if it was expressed in all. The additional staining of *dpn* or *ase* in future experiments should allow the assignment of the expression patterns to all specific lineages. Interestingly, Optix and Six4 were expressed in type II NBs (Fig. 11C), consistent with the results above (Fig. 9F). Additionally, run was also present in type II NBs (Fig. 11C). Surprisingly, Optix and run were not expressed in all type II lineages. It is possible that these differences could play a role in the variability of the progeny derived from the different type II progenitors. For example,

Ren *et al.* have shown that type II NB DL1 generate ensheathing glia while DM2, DM3, DM5, and DM6 produce both Astrocyte-like glia and ensheathing glia (Ren *et al.*, 2018). Furthermore, it was not clear whether Six4 was expressed in all type II lineages due to the reasons mentioned above.

Overall, a number of transcription factors identified in this subclustering analysis was previously shown to be heterogeneously expressed in embryonic neural progenitors. This included multiple spatial patterning genes, such as *ems*, *hkb*, *run*, *slp1* and *slp2*, *mirr*, *gsb*, and others (suppl. file 2). This result points to the previously suggested idea by Lacin and Truman, that the lineage identity to which each NB is endowed upon birth is conserved during metamorphosis (Lacin and Truman, 2016). Even though the embryonic common temporal patterning programs are interrupted and a new one is started (Maurange *et al.*, 2008) (Syed *et al.*, 2017), the transcription factor code of each lineage seems to not be reprogrammed in the second wave of embryogenesis.

There were also some additional markers which distinguish NB progenitors in the embryo, and which have been shown to be also conserved in the larva, that we did not recover as cluster markers. This included for example the columnar genes *vnd* and *msh*. This could have been potentially caused by two reasons. First, the *GAL4* line used in this study (VT201094) to isolate cells was not expressed in every NB lineage. Second, expression levels or the number of cells expressing these genes could have been below the filters used for the tests performed.

In the marker validation and localization experiments we did not map marker expression to a single NB lineage resolution as it has been previously done in the larval brain (Pereanu, 2006). This was not incorporated in this work due the short duration nature of master theses. Both in the larval CB and VNC, the identification of every NB lineages has been achieved through the staining of Neurotactin, which labels all secondary lineages and SATs (secondary axonal tracts) (Pereanu, 2006). The low variability in the positioning of each lineage in the developing CNS and the stereotyped orientation to which SATs project has allowed the mapping of all NB lineages (Pereanu, 2006) (Truman, 2004). This will be addressed in further studies.

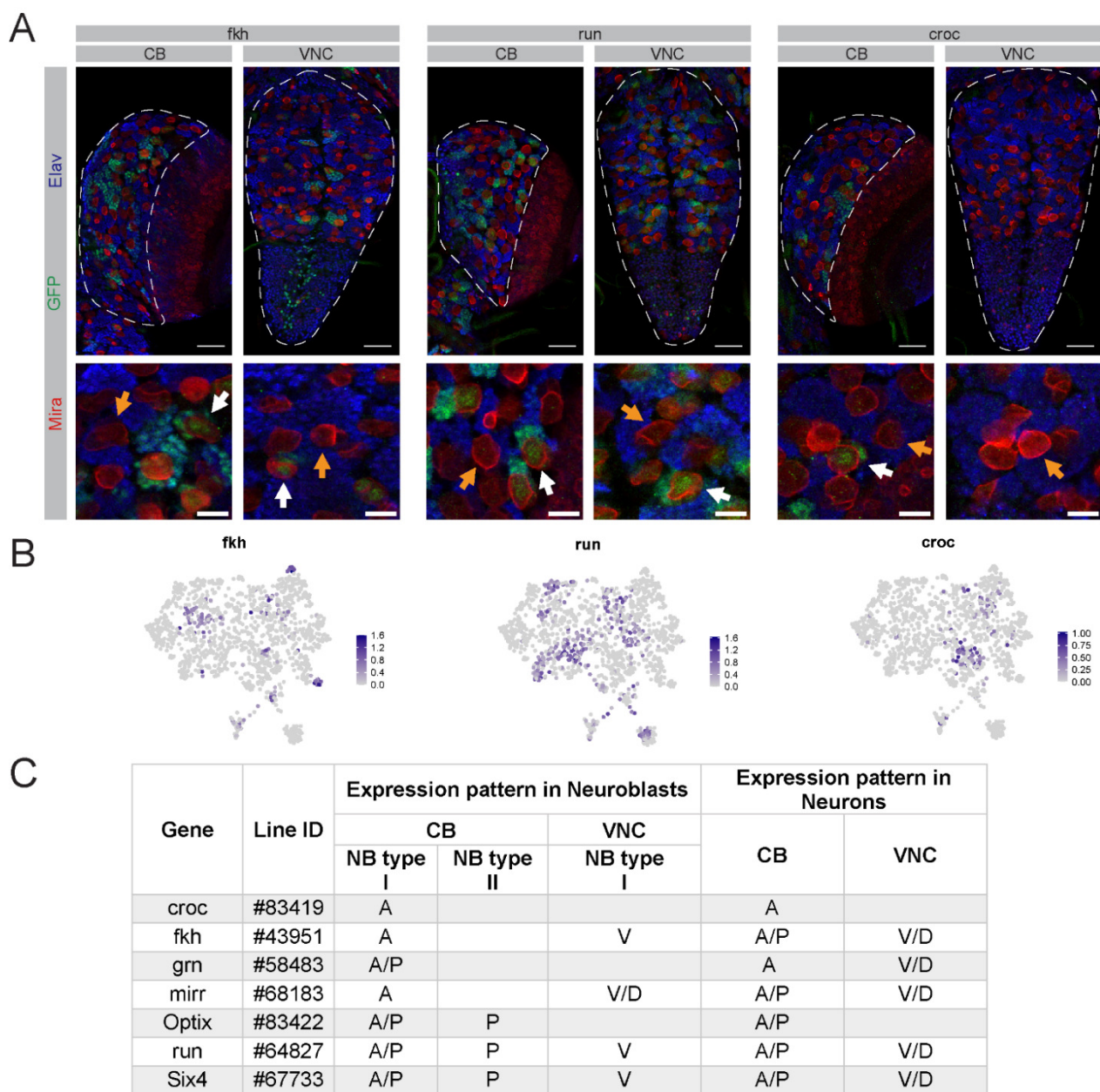


Fig. 11 Validation and localization assessment of type I NB subcluster marker genes

A) Confocal microscopy images of the CNS of protein trap lines for *fkh*, *run* and *croc*. Staining was done for mira - NBs (red), GFP (green), elav - neurons (blue). For each genotype is shown a z-stack slice of the superficial layer of the anterior and ventral side of the CB and VNC, respectively (n=5). White arrow and orange arrow indicate a positive NBs and a negative NBs, respectively, for protein trap expression. Scale bar – 30 microns. Below an inset of each image is shown; scale bar – 10 microns. **B)** UMAP plot shows type I NB subclusters and marker normalized expression per cell (*fkh*, *run* and *croc*) **C)** Expression pattern of different type I NB subcluster markers, in NBs and neurons. A – Anterior; P – Posterior; V – Ventral; D – Dorsal.

Neuronal cluster markers reveal differences in maturation state and fate identity

In the initial clustering analysis, the neuronal major cell group already presented significant divisions, 24 clusters (Fig. 7A and B). To identify further transcriptional diversity within this cell class we have thus not performed a subclustering analysis, and analysed the markers of these 24 neuronal clusters (Fig. 12A). To understand the type of genes that characterize the different neuronal clusters we performed a gene ontology enrichment analysis on the markers of all these clusters. As for the molecular function, there was an enrichment for genes with DNA binding function, excitatory extracellular ligand-gated ion channel activity, 3',5'-cyclic-nucleotide phosphodiesterase activity (Fig. 12B). Strikingly, regarding the biological process, genes related to axon guidance and axonogenesis were found to be enriched in the markers of the neuronal clusters (Fig. 12B), which is characteristic phenomenon for secondary neurons in the larval stage (Pereanu, 2006).

When observing the markers for the different neuronal clusters it was possible to identify neurons within different maturation states. Cluster 2, which was closer to GMCs expressed the Notch target gene *Hairy/E(spl)-related with YRPW motif* (Hey) in the great majority of cells (Fig. 12A). Hey has previously been shown to be expressed in NotchON early born neurons in larval NB lineages (Monastirioti *et al.*, 2010). Besides Hey, other known Notch target genes were also identified, these included *E(spl)m7-HLH* and *E(spl)m6-BFM* (Fig. 12A). Interestingly, a lncRNA CR44272 had a very similar pattern of expression as Hey (Fig. 12A). Whether the NotchON hemilineage fate is influenced by lncRNA regulation is unknown. Furthermore, it is also known that some early born neurons die as a result of programmed cell death (Pinto & Teixeira *et al.*, 2016). Indeed, the neuronal cluster 11 very specifically expressed the proapoptotic genes *reaper* (*rpr*), *grim* and *sickle* (*skl*) (Fig. 12A). Consistent with previous reports that both NotchON and NotchOFF hemilineages can undergo apoptosis (Lin and Lee, 2012), around half of the cells in cluster 11 expressed Hey (Fig. 12A).

In neuronal clusters which are farther from GMCs we observed a downregulation of Notch signalling and upregulation of genes related to neuronal maturation (Fig. 13A). This included for example the neuronal *Synaptobrevin* (*nSyb*), a v-SNARE critical for neurotransmitter vesicle exocytosis (Fig. 13A). Furthermore, genes involved in neurotransmitter activity and biosynthesis, such as *Vesicular acetylcholine transporter* (VACHT) or *Choline acetyltransferase* (Cha) also start to be expressed (Fig. 13A).

Another group of proteins which are strongly associated with neuronal maturation due to their role in neuronal excitability and synaptic transmission are ion channels. Interestingly, many genes belonging to this group were found to be expressed in neuronal clusters (Fig. 13C). This included genes such as *Sh*, which encodes a voltage-gated potassium channel subunit, and multiple genes encoding subunits of the nicotinic Acetylcholine Receptor (nAChR) (Fig. 13C). Strikingly, the clusters

5 and 35 which are farther away from GMCs in the UMAP plot, and therefore poised to be more differentiated, were the neuronal clusters with more ion channels marker genes (Fig. 13C).

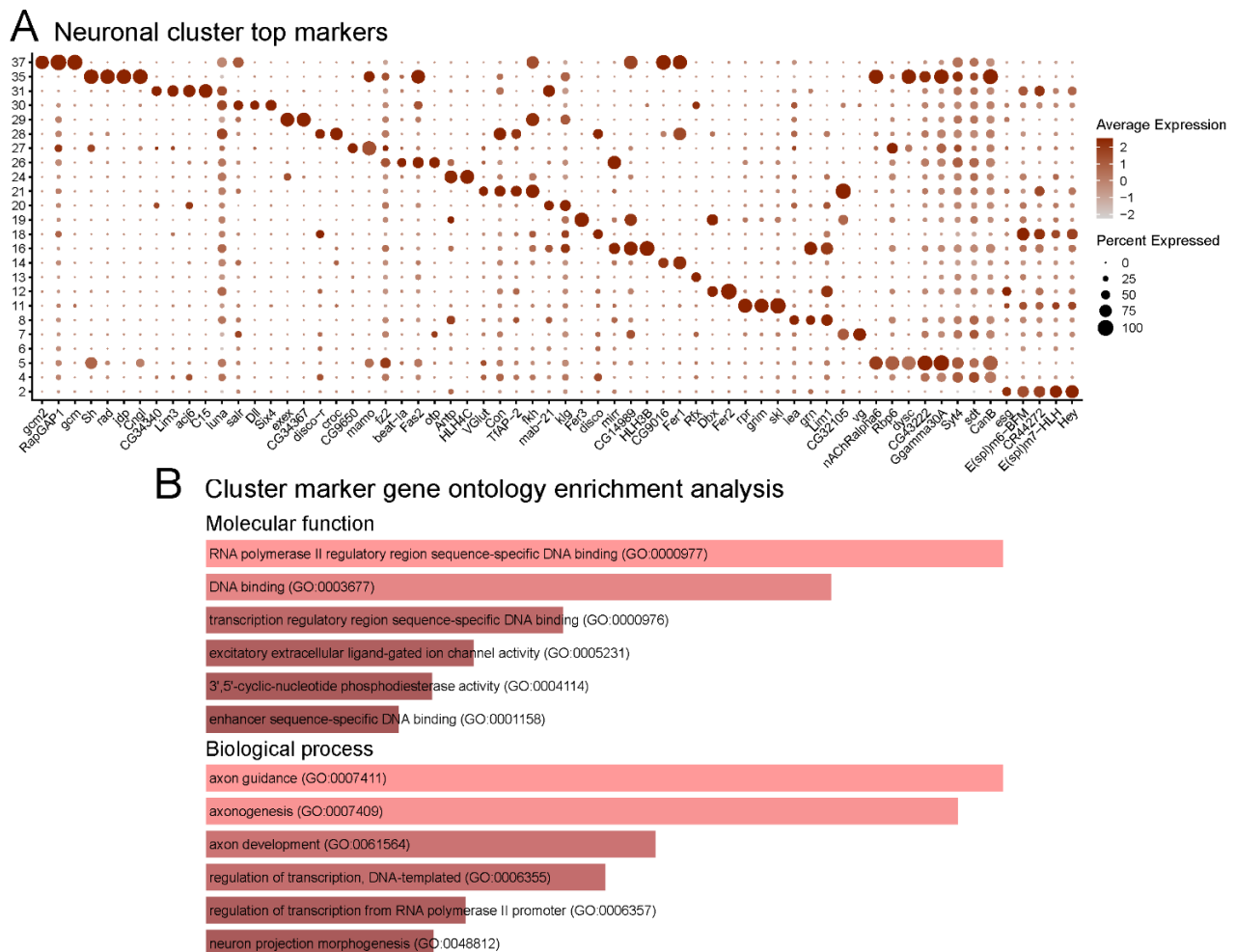


Fig. 12 Neuronal cluster gene expression diversity reveal differences in neuronal maturation and fate identity

A) Dot plot shows the top neuronal cluster marker genes (pct.1 > 0.5 and pct.2 < 0.2, only the top 5). Neuronal cluster average scaled expression level is indicated by the colour scale. Dot dimension informs about the proportion of cells per cluster that express the neuronal markers **B)** Gene ontology enrichment analysis for all neuronal cluster marker genes. Bar graph shows the combined score value for FlyEnrichr gene ontology enrichment analysis. Only the top 6 GO terms are shown for the molecular function and biological process gene ontology.

The immunoglobulin and cadherin super families, mostly composed by adhesion proteins and secreted molecules play an important role in axonal development and in the organization of neuronal circuits (Tan *et al.*, 2015) (Ranscht, 2000). Interestingly, multiple genes which belong to these super families were identified in neuronal cluster markers (Fig. 13D). These included the immunoglobulin *beaten path* (beat) family, the *sidestep* (side) family, *Dscam*, and the cadherin super family gene *Cadherin-N* (CadN) (Fig. 13D). Clusters 5 and 35 were shown to have the highest number of cluster markers belonging to these two super families (Fig. 13D). Indeed, these two clusters seem to

correspond to older and more mature neurons given that these groups of genes are inherently linked to neuronal maturation.

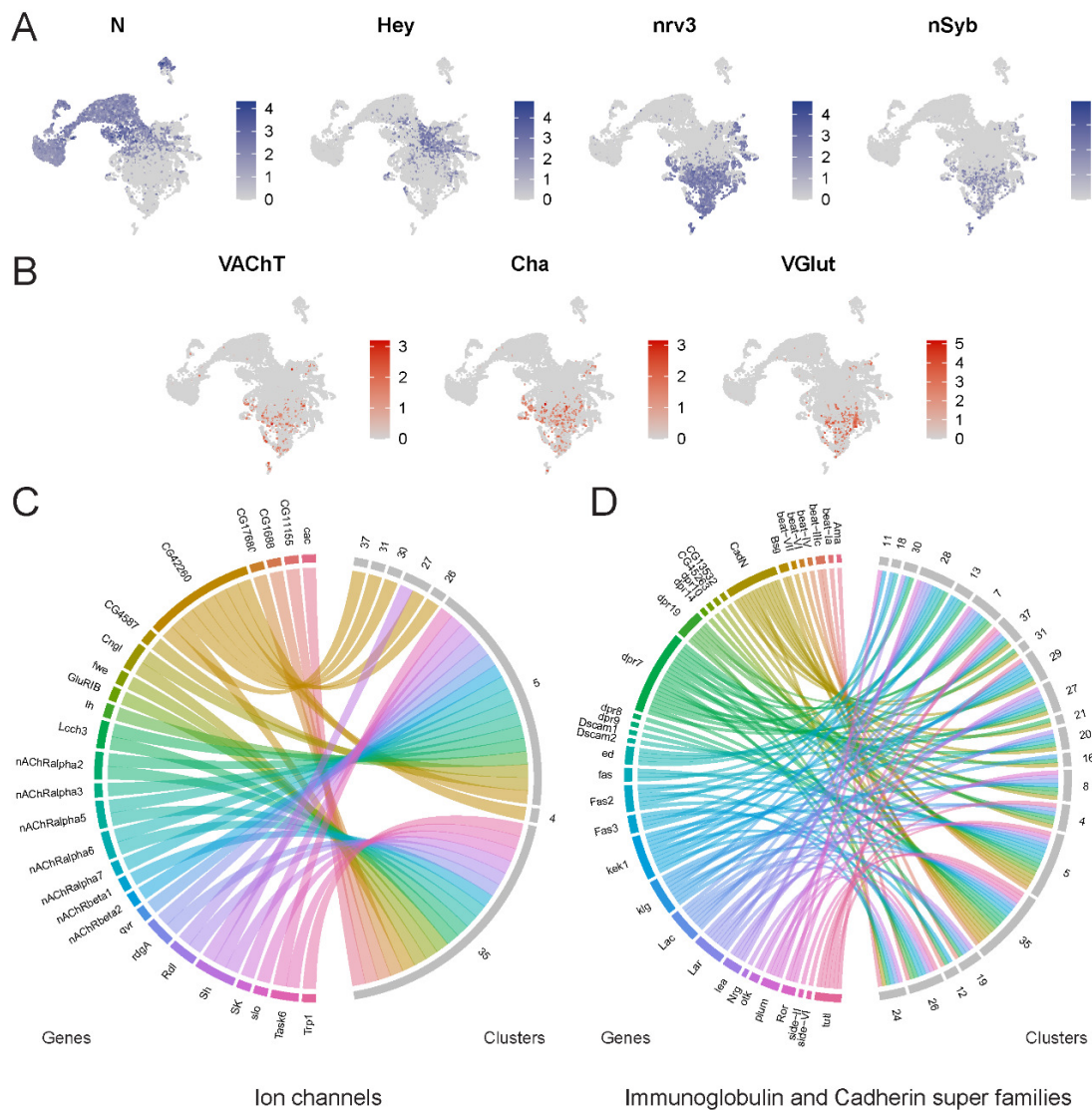


Fig. 13 Early born and maturation neuronal genes

A) and **B)** UMAP plot shows which cells express early born neuronal genes *Notch* and *Hey*, and maturation associated markers *nrv3*, *nsyb* and neurotransmitter activity genes *VACht*, *Cha*, *VGlut*. Normalized expression levels per cell are indicated by the colour scale. Chord diagrams show the correspondence between **C)** Ion channels cluster markers **D)** Immunoglobulin and Cadherin super families cluster markers, and the clusters in which they are differentially expressed (Code for visualization was adapted from Allen *et al.*, 2020).

This analysis revealed that neurons can be distinguished in terms of their maturation state. We have identified clusters which seem to correspond to NotchON early born neurons. We did not identify any obvious marker genes which could potentially identify NotchOFF early born neurons. Furthermore, clusters which are more distant to GMCs upregulate genes related to neuronal maturation such as

nSyb. Moreover, we have identified a number of ion channel, adhesion proteins, and secreted molecules which already start to be expressed. Indeed, an 18hr developmental window seems to be enough for NBs to generate neurons and for these to start expressing genes critical for neuronal function.

Besides the differences in maturation it was also possible to distinguish neuronal clusters by differential expression of a small set of TFs (Fig. 12A). Some of these TFs, such as *fkf*, *croc*, *Six4*, *Optix*, *mirr*, *run* were previously identified to be heterogeneously expressed between type I NBs (Fig. 10C and Suppl. file 1) (Lacin and Truman, 2016). This strongly suggests that these TF differences are the result of these neurons being descendants of specific NB lineages. Indeed, when re-observing the immunostaining for these markers, positive NBs and neurons (*elav*⁺) are often together in a characteristic NB lineage shape (Fig 11A, Suppl. Fig 5A). Interestingly, a few of these TFs were not identified in the type I NB subclustering analysis. This suggests that they could be upregulated only in neurons to specify a given fate as a result of the inheritance of molecular players from the NB or GMC. For example, in the cOPC, *svp* is only expressed in neurons as a result of the action of spatial and temporal factors, *Vsx* and *hth*, respectively (Erclik *et al.*, 2017). In our analysis, genes such *abnormal chemosensory jump 6* (*acj6*), *C15*, *maternal gene required for meiosis* (*mamo*), *Helix loop helix protein 4C* (*HLH4C*), *48 related 2* (*Fer2*) (Fig. 12A), were found to be heterogeneously marker genes between neurons but not in type I NBs. Indeed, *acj6* has previously been shown in the VNC to be upregulated only in neurons of specific NB lineages (Lacin and Truman, 2016). *Mamo* has also recently been demonstrated in the mushroom body lineages to be upregulated only in neurons to generate a specific fate (Liu *et al.*, 2019).

An interesting observation was seen for cluster 37, where *glial cells missing* (*gcm1* and 2) were present as cluster markers (Fig. 12A). Both these genes have previously been suggested to be “*bona fide* cell type selectors” for specifying the glial fate by inducing the expression of *repo* (Allan and Thor, 2015). However, a large number of cells in this clusters were positive for the pan-neuronal marker *Elav* (~84% expressed *elav*) and negative for *repo* (~5% expressed *repo*). It is possible that *repo* is only induced later in development and not in the 18hr window frame captured in this dataset. However, *gcm* has also been shown to be expressed in neurons (Cattenoz and Giangrande, 2016). For these reasons, we have decided to label these cells as neurons, but further experiments will be required to define their identity.

Overall, in this analysis, neurons were distinguished in two axes: a neuronal maturation axis and a NB lineage origin axis. First, neurons were found to be different according to their maturation stages, which was identified by the expression of early born markers and genes related to neuronal function by different groups of cells. Secondly, we have identified a number of heterogeneously expressed TFs between neuronal clusters which seem to be related to NB lineage origin. Here, we have not

addressed how both these axes can work together. Very likely these NB lineage TF codes influence the expression of specific neuronal maturation genes in order to determine, for example, neurite target projection sites or arborization complexity. Interestingly, many of the TFs identified here have also recently been shown to characterize neuronal clusters in single-cell studies of the adult brain and ventral nerve cord (Davie *et al.*, 2018) (Allen *et al.*, 2020). This includes genes such as *fkf*, *HLH4C*, *C15*, *Fer2*, *acj6* and others. This points to the idea, as similarly suggested by Allen *et al.*, that these genes are not only critical for neuronal development, but also for fate identity maintenance throughout adult life (Allen *et al.*, 2020).

Conclusion and future studies

Deciphering how gene regulatory networks drive fate commitment of neural progenitors into an incredible array of diverse neurons, is a fundamental question which may bring many biomedical benefits such as the finer design of regenerative therapies. In the past decades, neurodevelopmental work in *Drosophila* has contributed substantially to the answer of this enigma, however many details still remain unclear. Recent advances in single-cell methodology now offer the opportunity to push this question forward and to identify additionally involved genes. In this work we built a single-cell atlas for *Drosophila* larval NB lineages. Through the use of Seurat's bioinformatic tools we have succeeded in the identification of the characteristic NB lineage cell types. Later, through the comparison of the multiple cell type groups we have identified a battery of genes that are distinctively expressed in the different levels of lineage differentiation. Further genetic studies will allow the assessment of whether the unstudied genes identified here play a role in, for example, NB self-renewal, lineage progression or fate specification.

We have further studied the heterogeneity between type I NBs through a subclustering analysis. We found that, in accordance with previous reports, larval type I NBs continue to express embryonic NB heterogeneous genes (Lacin and Truman, 2016). Thus, the quiescent period in between embryo and larva does not seem to induce a "reset" in the identity to which NB are endowed upon birth (Lacin and Truman, 2016). Next, we validated *in vivo* the expression of a few heterogeneous type I NB marker genes and analysed their expression patterns. Previous work in the developing VNC have established comprehensive single lineage marker maps for a part of the NB heterogeneous genes identified in this analysis (Lacin and Truman, 2016). It would be of great value to see these maps extended with the additional markers that we validated to be expressed in the VNC type I NB, *fkf* and *six4*. *Mirr* and *run* have already been included (Lacin and Truman, 2016). Furthermore, to the best of our knowledge there has not been any work that has created a similar marker map for larval CB lineages. This will be a great addition for the tracing and genetic manipulation of CB lineages. At the moment, the developmental rules governing the innervation of the majority of adult neuropile correspondents of CB larval lineages remain poorly characterized. Current studies are mostly focused in the type II, mushroom body, and antennal lobe lineages. Recently, Awasaki *et al.* have created a genetic system which allows the fluorescent labelling of the neuronal descendants of a given NB subset in the adult brain (Awasaki *et al.*, 2014). Through the use of a NB subset-specific *GAL4* line, a cascade of recombinases are activated in the NB and immortalize the expression of a neuronal-specific driver. In conjunction with twin-spot MARCM, if the mitotic recombination event happens at the NB level, it will label two hemilineages with one fluorophore and the rest of the lineage with a second different fluorophore (Awasaki *et al.*, 2014). Future studies where this system is driven by the promoter regions of heterogenous type I NB genes identified in this study should be valuable for understanding which CB lineages express specific players. The latter being a critical pre-requisite

for the manipulation of certain lineages and for functionally studying the role of NB heterogeneously expressed genes.

We have also additionally addressed the heterogeneity between neurons. Surprisingly, we identified neurons generated within an 18-hour window to express genes related to neuronal maturation, such as for neurotransmitter synthesis and ion channel activity. This was not expected given that secondary neurons have been described to stay in an immature state with simple projections until metamorphosis, and only then branch extensively and become fully mature (Truman, 1990) (Pereanu, 2006). Regarding the NB lineage origin related differences, it will be first important to map the poorly studied markers to the different NB lineages of the CB and VNC. This includes genes such as *C15*, *salr*, *Six4*, *CG34367*, *croc*, *orthopedia* (*otp*), *HLH4C*. It will also be interesting to see if they are restricted to just one hemilineage as it has been shown for other TFs, such as *Lim3* and *extra-extra* (*exex*) (Lacin *et al.*, 2014), and if Notch plays a role in the expression of these molecular codes. Furthermore, the continuous expression of some of these genes into adult neuronal populations (Davie *et al.*, 2018) (Allen *et al.*, 2020) suggests that they could potentially play a role in commanding the expression of sets of functional neuronal genes. Could these genes act as terminal selectors?

This study also presented some limitations. First, here we did not address how the temporal axis affects NB heterogeneity. Temporal patterning has been shown to be a critical intrinsic mechanism for the generation of neuronal diversity (Lin and Lee, 2012) (Li *et al.*, 2013). It is very likely that different NB lineages could exhibit different temporal gene expression cascades. As mentioned in the introduction, this has been shown for the optic lobe, tOPC lineages go through a different TF cascade than the rest of the OPC lineages. It is thus possible that in larval CB and VNC different lineages execute distinct TF cascades. Future studies where additional timepoints are sequenced could allow this to be studied. Secondly, due to the fact that we did not sequence all NB lineages in the developing CB and VNC, this work does not present a global atlas for all the different NB lineage programs. This was because the *GAL4* driver used to isolate NB lineages was not expressed in every single lineage. Future single-cell studies with different drivers, or the intersection with expression screen studies and marker maps may bridge and fill the missing information. Thirdly, due to the fact that we measured gene activity at the RNA level and significant post-transcriptional regulation is thought to play a role in NB lineages (Weng and Cohen, 2015) (Samuels *et al.*, 2020), results must always be interpreted with caution and validated. For example, we have observed that *chinmo* is expressed in NBs and GMCs (Fig. 9A) and neurons (data not shown). However, it has been shown that this gene is post-transcriptionally repressed and no protein is observed in the NBs, GMCs and newly born neurons in the late 3rd instar stage (Maurange *et al.*, 2008) (Dillard *et al.*, 2018). Furthermore, a major limitation of this study was that we did not establish a physical relationship between the type I NB subclusters and specific NB lineages in the developing CNS. This

was not attempted because we were not confident with attributing lineage identity to each NB subcluster without the aid of post-embryonic lineage marker maps. Currently, there are only available post-embryonic marker maps for VNC lineages. However, given that lineage identity seems to be conserved throughout metamorphosis, and that marker maps are available for the brain embryonic NB lineages, establishing a bridge between subclusters and physical lineages should be possible. Achieving this in future studies would be really useful, as knowing the transcriptome of each individual lineages could provide valuable insights into the mechanisms behind neuronal diversity,

With this work we hope to thrust functional genetic studies by providing a view into the transcriptional profile of NB lineages and to contribute for the dissection of the global gene networks behind neurodevelopment. Conquering the full understanding of how a small set of neural stem cells generates the vastly diverse nervous system will open many doors and opportunities, such as the finer design of regenerative medicine protocols and a better appreciation to where lies the origin of all animal behaviour.

References

- 10XGenomics (2020). Chromium Single Cell 3' Reagent Kits User Guide (v3 Chemistry).
- Alekseyenko, A.A., Gorchakov, A.A., Zee, B.M., Fuchs, S.M., Kharchenko, P.V., and Kuroda, M.I. (2014). Heterochromatin-associated interactions of *Drosophila* HP1a with dADD1, HIP1, and repetitive RNAs. *Genes Dev.* 28, 1445–1460.
- Allan, D.W., and Thor, S. (2015). Transcriptional selectors, masters, and combinatorial codes: regulatory principles of neural subtype specification: Regulation of neural subtype generation. *Wiley Interdiscip. Rev. Dev. Biol.* 4, 505–528.
- Allen, A.M., Neville, M.C., Birtles, S., Croset, V., Treiber, C.D., Waddell, S., and Goodwin, S.F. (2020). A single-cell transcriptomic atlas of the adult *Drosophila* ventral nerve cord. *ELife* 9, e54074.
- 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. *Development* 145, dev158394.
- Alyagor, I., Berkun, V., Keren-Shaul, H., Marmor-Kollet, N., David, E., Mayseless, O., Issman-Zecharya, N., Amit, I., and Schuldiner, O. (2018). Combining Developmental and Perturbation-Seq Uncovers Transcriptional Modules Orchestrating Neuronal Remodeling. *Dev. Cell* 47, 38-52.e6.
- Arefin, B., Parvin, F., Bahrampour, S., Stadler, C.B., and Thor, S. (2019). *Drosophila* Neuroblast Selection Is Gated by Notch, Snail, SoxB, and EMT Gene Interplay. *Cell Rep.* 29, 3636-3651.e3.
- Ashraf, S.I., and Ip, Y.T. (2001). Snail family in neuroblast division. *Development* 11.
- Atkins, M., Potier, D., Romanelli, L., Jacobs, J., Mach, J., Hamaratoglu, F., Aerts, S., and Halder, G. (2016). An Ectopic Network of Transcription Factors Regulated by Hippo Signaling Drives Growth and Invasion of a Malignant Tumor Model. *Curr. Biol.* 26, 2101–2113.
- Awasaki, T., Kao, C.-F., Lee, Y.-J., Yang, C.-P., Huang, Y., Pfeiffer, B.D., Luan, H., Jing, X., Huang, Y.-F., He, Y., et al. (2014). Making *Drosophila* lineage-restricted drivers via patterned recombination in neuroblasts. *Nat. Neurosci.* 17, 631–637.
- Baumgardt, M., Karlsson, D., Salmani, B.Y., Bivik, C., MacDonald, R.B., Gunnar, E., and Thor, S. (2014). Global Programmed Switch in Neural Daughter Cell Proliferation Mode Triggered by a Temporal Gene Cascade. *Dev. Cell* 30, 192–208.
- Bayraktar, O.A., and Doe, C.Q. (2013). Temporal patterning in intermediate progenitors increases neural diversity. *Nature* 498, 449–455.
- Bellen, H.J., Tong, C., and Tsuda, H. (2010). 100 years of *Drosophila* research and its impact on vertebrate neuroscience: a history lesson for the future. *Nat. Rev. Neurosci.* 11, 514–522.
- 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.

- Bertet, C., Li, X., Erclik, T., Cavey, M., Wells, B., and Desplan, C. (2014). Temporal Patterning of Neuroblasts Controls Notch-Mediated Cell Survival through Regulation of Hid or Reaper. *Cell* 158, 1173–1186.
- Bowman, S.K., Rolland, V., Betschinger, J., Kinsey, K.A., Emery, G., and Knoblich, J.A. (2008). The Tumor Suppressors Brat and Numb Regulate Transit-Amplifying Neuroblast Lineages in *Drosophila*. *Dev. Cell* 14, 535–546.
- Brand, A.H., and Perrimon, N. (1993). Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development* 118, 401–415.
- Brand, M., Jarman, A.P., Jan, L.Y., and Jan, Y.N. (1993). *asense* is a *Drosophila* neural precursor gene and is capable of initiating sense organ formation. *Development* 17.
- Broadus, J., Skeath, J.B., Spana, E.P., Bossing, T., Technau, G., and Doe, C.Q. (1995). New neuroblast markers and the origin of the aCC/pCC neurons in the *Drosophila* central nervous system. *Mech. Dev.* 53, 393–402.
- Broderick, S., Wang, X., Simms, N., and Page-McCaw, A. (2012). *Drosophila* Ninjurin A Induces Nonapoptotic Cell Death. *PLOS ONE* 7, e44567.
- Brody, T., Stivers, C., Nagle, J., and Odenwald, W.F. (2002). Identification of novel *Drosophila* neural precursor genes using a differential embryonic head cDNA screen. *Mech. Dev.* 113, 41–59.
- Cajal, S.R. (1999). *Texture of the Nervous System of Man and the Vertebrates* (Vienna: Springer Vienna).
- Carrillo, R.A., Özkan, E., Menon, K.P., Nagarkar-Jaiswal, S., Lee, P.-T., Jeon, M., Birnbaum, M.E., Bellen, H.J., Garcia, K.C., and Zinn, K. (2015). Control of Synaptic Connectivity by a Network of *Drosophila* IgSF Cell Surface Proteins. *Cell* 163, 1770–1782.
- Cattenoz, P.B., and Giangrande, A. (2016). Revisiting the role of the Gcm transcription factor, from master regulator to Swiss army knife. *Fly (Austin)* 10, 210–218.
- Chen, G., Ning, B., and Shi, T. (2019). Single-Cell RNA-Seq Technologies and Related Computational Data Analysis. *Front. Genet.* 10.
- Croset, V., Treiber, C.D., and Waddell, S. (2018). Cellular diversity in the *Drosophila* midbrain revealed by single-cell transcriptomics. *ELife* 7, e34550.
- 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.
- Dillard, C., Narbonne-Reveau, K., Foppolo, S., Lanet, E., and Maurange, C. (2018). Two distinct mechanisms silence *chinmo* in *Drosophila* neuroblasts and neuroepithelial cells to limit their self-renewal. *Development* 145, dev154534.
- Doe, C.Q. (1992). Molecular markers for identified neuroblasts and ganglion mother cells in the *Drosophila* central nervous system. *Development* 116, 855–863.
- Doe, C.Q. (2008). Neural stem cells: balancing self-renewal with differentiation. *Development* 135, 1575–1587.
- Doe, C.Q. (2017). Temporal Patterning in the *Drosophila* CNS. *Annu. Rev. Cell Dev. Biol.*

- Donlea, J.M., Pimentel, D., Talbot, C.B., Kempf, A., Omoto, J.J., Hartenstein, V., and Miesenböck, G. (2018). Recurrent Circuitry for Balancing Sleep Need and Sleep. *Neuron* 97, 378-389.e4.
- Egger, B., Boone, J.Q., Stevens, N.R., Brand, A.H., and Doe, C.Q. (2007). Regulation of spindle orientation and neural stem cell fate in the *Drosophila* optic lobe. *Neural Develop.* 2, 1.
- Erclik, T., Li, X., Courgeon, M., Bertet, C., Chen, Z., Baumert, R., Ng, J., Koo, C., Arain, U., Behnia, R., et al. (2017). Integration of temporal and spatial patterning generates neural diversity. *Nature* 541, 365–370.
- French, V. (2001). Insect segmentation: Genes, stripes and segments in ‘Hoppers.’ *Curr. Biol.* 11, R910–R913.
- Gallaud, E., Pham, T., and Cabernard, C. (2017). *Drosophila melanogaster* Neuroblasts: A Model for Asymmetric Stem Cell Divisions. In *Asymmetric Cell Division in Development, Differentiation and Cancer*, J.-P. Tassan, and J.Z. Kubiak, eds. (Cham: Springer International Publishing), pp. 183–210.
- Gilbert, S.F., and Barresi, M.J.F. (2016). The Genetics of Axis Specification in *Drosophila*. In *Developmental Biology*, p.
- Giraldo, Y.M., Leitch, K.J., Ros, I.G., Warren, T.L., Weir, P.T., and Dickinson, M.H. (2018). Sun Navigation Requires Compass Neurons in *Drosophila*. *Curr. Biol.* 28, 2845-2852.e4.
- Gold, K.S., and Brand, A.H. (2014). Optix defines a neuroepithelial compartment in the optic lobe of the *Drosophila* brain. *Neural Develop.* 9, 18.
- Grabe, V., and Sachse, S. (2018). Fundamental principles of the olfactory code. *Biosystems* 164, 94–101.
- Gu, Z., Gu, L., Eils, R., Schlesner, M., and Brors, B. (2014). circlize implements and enhances circular visualization in R. *Bioinformatics* 30, 2811–2812.
- Guan, Z., Saraswati, S., Adolfsen, B., and Littleton, J.T. (2005). Genome-Wide Transcriptional Changes Associated with Enhanced Activity in the *Drosophila* Nervous System. *Neuron* 48, 91–107.
- Hakes, A.E., Otsuki, L., and Brand, A.H. (2018). A newly discovered neural stem cell population is generated by the optic lobe neuroepithelium during embryogenesis in *Drosophila melanogaster*. *Development* 145.
- Hanyu-Nakamura, K., Kobayashi, S., and Nakamura, A. (2004). Germ cell-autonomous Wunen2 is required for germline development in *Drosophila* embryos. *Development* 131, 4545–4553.
- Harris, K.P., and Littleton, J.T. (2015). Transmission, Development, and Plasticity of Synapses. *Genetics* 201, 345–375.
- Heisenberg, M. (2003). Mushroom body memoir: from maps to models. *Nat. Rev. Neurosci.* 4, 266–275.
- Hobert, O. (2016). A map of terminal regulators of neuronal identity in *Caenorhabditis elegans*. *Wiley Interdiscip. Rev. Dev. Biol.* 5, 474–498.
- Hobert, O., and Kratsios, P. (2019). Neuronal identity control by terminal selectors in worms, flies, and chordates. *Curr. Opin. Neurobiol.* 56, 97–105.
- Holguera, I., and Desplan, C. (2018). Neuronal specification in space and time. *Science* 362, 176–180.
- Homem, C.C.F., and Knoblich, J.A. (2012). *Drosophila* neuroblasts: a model for stem cell biology. *Development* 139, 4297–4310.

- Homem, C.C.F., Steinmann, V., Burkard, T.R., Jais, A., Esterbauer, H., and Knoblich, J.A. (2014). Ecdysone and Mediator Change Energy Metabolism to Terminate Proliferation in *Drosophila* Neural Stem Cells. *Cell* 158, 874–888.
- 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.
- Ikeshima-Kataoka, H., Skeath, J.B., Nabeshima, Y., Doe, C.Q., and Matsuzaki, F. (1997). Miranda directs Prospero to a daughter cell during *Drosophila* asymmetric divisions. *Nature* 390, 625–629.
- Ilicic, T., Kim, J.K., Kolodziejczyk, A.A., Bagger, F.O., McCarthy, D.J., Marioni, J.C., and Teichmann, S.A. (2016). Classification of low quality cells from single-cell RNA-seq data. *Genome Biol.* 17, 29.
- Ito, M., Masuda, N., Shinomiya, K., Endo, K., and Ito, K. (2013). Systematic Analysis of Neural Projections Reveals Clonal Composition of the *Drosophila* Brain. *Curr. Biol.* 23, 644–655.
- Janssens, D.H., Komori, H., Grbac, D., Chen, K., Koe, C.T., Wang, H., and Lee, C.-Y. (2014). Earmuff restricts progenitor cell potential by attenuating the competence to respond to self-renewal factors. *Development* 141, 1036–1046.
- Jenett, A., Rubin, G.M., Ngo, T.-T.B., Shepherd, D., Murphy, C., Dionne, H., Pfeiffer, B.D., Cavallaro, A., Hall, D., Jeter, J., et al. (2012). A GAL4-Driver Line Resource for *Drosophila* Neurobiology. *Cell Rep.* 2, 991–1001.
- Jukam, D., and Desplan, C. (2010). Binary Fate Decisions in Differentiating Neurons. *Curr. Opin. Neurobiol.* 20, 6–13.
- Kageyama, R., and Ohtsuka, T. (1999). The Notch-Hes pathway in mammalian neural development. *Cell Res.* 9, 179–188.
- Kohwi, M., and Doe, C.Q. (2013). Temporal fate specification and neural progenitor competence during development. *Nat. Rev. Neurosci.* 14, 823–838.
- Kolodkin, A.L., and Tessier-Lavigne, M. (2011). Mechanisms and molecules of neuronal wiring: a primer. *Cold Spring Harb. Perspect. Biol.* 3.
- Komori, H., Golden, K.L., Kobayashi, T., Kageyama, R., and Lee, C.-Y. (2018). Multilayered gene control drives timely exit from the stem cell state in uncommitted progenitors during *Drosophila* asymmetric neural stem cell division. *Genes Dev.* 32, 1550–1561.
- 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., Stolfi, A., Levine, M., and Hobert, O. (2012). Coordinated regulation of cholinergic motor neuron traits through a conserved terminal selector gene. *Nat. Neurosci.* 15, 205–214.
- Krzeptowski, W., Hess, G., and Pyza, E. (2018). Circadian Plasticity in the Brain of Insects and Rodents. *Front. Neural Circuits* 12, 32.
- Kudron, M.M., Victorsen, A., Gevirtzman, L., Hillier, L.W., Fisher, W.W., Vafeados, D., Kirkey, M., Hammonds, A.S., Gersch, J., Ammouri, H., et al. (2018). The ModERN Resource: Genome-Wide Binding Profiles for Hundreds of *Drosophila* and *Caenorhabditis elegans* Transcription Factors. *Genetics* 208, 937–949.

- Kuert, P.A., Hartenstein, V., Bello, B.C., Lovick, J.K., and Reichert, H. (2014). Neuroblast lineage identification and lineage-specific Hox gene action during postembryonic development of the subesophageal ganglion in the *Drosophila* central brain. *Dev. Biol.* **390**, 102–115.
- Kuleshov, M.V., Diaz, J.E.L., Flamholz, Z.N., Keenan, A.B., Lachmann, A., Wojciechowicz, M.L., Cagan, R.L., and Ma'ayan, A. (2019). modEnrichr: a suite of gene set enrichment analysis tools for model organisms. *Nucleic Acids Res.* **47**, W183–W190.
- Lacin, H., and Truman, J.W. (2016). Lineage mapping identifies molecular and architectural similarities between the larval and adult *Drosophila* central nervous system. *ELife* **5**, e13399.
- Lacin, H., Zhu, Y., Wilson, B.A., and Skeath, J.B. (2014). Transcription factor expression uniquely identifies most postembryonic neuronal lineages in the *Drosophila* thoracic central nervous system. *Development* **141**, 1011–1021.
- Landskron, L., Steinmann, V., Bonnay, F., Burkard, T.R., Steinmann, J., Reichardt, I., Harzer, H., Laurenson, A.-S., Reichert, H., and Knoblich, J.A. (2018). The asymmetrically segregating lncRNA cherub is required for transforming stem cells into malignant cells. *ELife* **7**, e31347.
- Lee, T. (2017). Wiring the *Drosophila* Brain with Individually Tailored Neural Lineages. *Curr. Biol.* **27**, R77–R82.
- Lee, C.-Y., Wilkinson, B.D., Siegrist, S.E., Wharton, R.P., and Doe, C.Q. (2006). Brat Is a Miranda Cargo Protein that Promotes Neuronal Differentiation and Inhibits Neuroblast Self-Renewal. *Dev. Cell* **10**, 441–449.
- Lee, Y.-J., Yang, C.-P., Miyares, R.L., Huang, Y.-F., He, Y., Ren, Q., Chen, H.-M., Kawase, T., Ito, M., Otsuna, H., et al. (2020). Conservation and divergence of related neuronal lineages in the *Drosophila* central brain. *ELife* **9**, e53518.
- Li, H., Shuster, S.A., Li, J., and Luo, L. (2018). Linking neuronal lineage and wiring specificity. *Neural Develop.* **13**, 5.
- Li, P., Yang, X., Wasser, M., Cai, Y., and Chia, W. (1997). Inscuteable and Staufen mediate asymmetric localization and segregation of prospero RNA during *Drosophila* neuroblast cell divisions. *Cell* **90**, 437–447.
- Li, X., Erclik, T., Bertet, C., Chen, Z., Voutev, R., Venkatesh, S., Morante, J., Celik, A., and Desplan, C. (2013). Temporal patterning of *Drosophila* medulla neuroblasts controls neural fates. *Nature* **498**, 456–462.
- Lin, S., and Lee, T. (2012). Generating neuronal diversity in the *Drosophila* central nervous system: Neuronal Diversity in *Drosophila*. *Dev. Dyn.* **241**, 57–68.
- Lin, S., Lai, S.-L., Yu, H.-H., Chihara, T., Luo, L., and Lee, T. (2010). Lineage-specific effects of Notch/Numb signaling in post-embryonic development of the *Drosophila* brain. *Development* **137**, 43–51.
- Liu, L.-Y., Long, X., Yang, C.-P., Miyares, R.L., Sugino, K., Singer, R.H., and Lee, T. (2019). Mamo decodes hierarchical temporal gradients into terminal neuronal fate. *ELife* **8**, e48056.
- Liu, Z., Yang, C.-P., Sugino, K., Fu, C.-C., Liu, L.-Y., Yao, X., Lee, L.P., and Lee, T. (2015). Opposing intrinsic temporal gradients guide neural stem cell production of varied neuronal fates. *Science* **350**, 317–320.

- Loedige, I., Stotz, M., Qamar, S., Kramer, K., Hennig, J., Schubert, T., Löffler, P., Längst, G., Merkl, R., Urlaub, H., et al. (2014). The NHL domain of BRAT is an RNA-binding domain that directly contacts the hunchback mRNA for regulation. *Genes Dev.* **28**, 749–764.
- Luecken, M.D., and Theis, F.J. (2019). Current best practices in single-cell RNA-seq analysis: a tutorial. *Mol. Syst. Biol.* **15**, e8746.
- Lundell, M.J., Chu-LaGriff, Q., Doe, C.Q., and Hirsh, J. (1996). The engrailed and huckebein Genes Are Essential for Development of Serotonin Neurons in the *Drosophila* CNS. *Mol. Cell. Neurosci.* **7**, 46–61.
- Macosko, E.Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, A.R., Kamitaki, N., Martersteck, E.M., et al. (2015). Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell* **161**, 1202–1214.
- Marshall, O.J., and Brand, A.H. (2017). Chromatin state changes during neural development revealed by in vivo cell-type specific profiling. *Nat. Commun.* **8**, 2271.
- Maurange, C., and Gould, A.P. (2005). Brainy but not too brainy: starting and stopping neuroblast divisions in *Drosophila*. *Trends Neurosci.* **28**, 30–36.
- Maurange, C., Cheng, L., and Gould, A.P. (2008). Temporal Transcription Factors and Their Targets Schedule the End of Neural Proliferation in *Drosophila*. *Cell* **133**, 891–902.
- McInnes, L., Healy, J., and Melville, J. (2018). UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *ArXiv180203426 Cs Stat.*
- 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.
- 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.
- Ninova, M., Tóth, K.F., and Aravin, A.A. (2019). The control of gene expression and cell identity by H3K9 trimethylation. *Development* **146**.
- Ninova, M., Godneeva, B., Chen, Y.-C.A., Luo, Y., Prakash, S.J., Jankovics, F., Erdélyi, M., Aravin, A.A., and Tóth, K.F. (2020). The SUMO Ligase Su(var)2-10 Controls Hetero- and Euchromatic Gene Expression via Establishing H3K9 Trimethylation and Negative Feedback Regulation. *Mol. Cell* **77**, 571-585.e4.
- Pereanu, W. (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., Konstantinides, N., and Desplan, C. (2016). Programmed cell death acts at different stages of *Drosophila* neurodevelopment to shape the central nervous system. *FEBS Lett.* **590**, 2435–2453.
- Prokop, A. (2015). Why fly? Website - <https://droso4schools.wordpress.com/why-fly/>
- Prokop, A., and Technau, G.M. (1991). The origin of postembryonic neuroblasts in the ventral nerve cord of *Drosophila melanogaster*. *Development* **111**, 79–88.

- Pym, E.C., Southall, T.D., Mee, C.J., Brand, A.H., and Baines, R.A. (2006). The homeobox transcription factor Even-skipped regulates acquisition of electrical properties in *Drosophila* neurons. *Neural Develop.* 1, 3.
- 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.
- 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.
- Ren, Q., Awasaki, T., Wang, Y.-C., Huang, Y.-F., and Lee, T. (2018). Lineage-guided Notch-dependent gliogenesis by *Drosophila* multi-potent progenitors. *Development* 145, dev160127.
- Rieder, L.E., and Larschan, E.N. (2014). Wisdom from the Fly. *Trends Genet. TIG* 30, 479–481.
- Rives-Quinto, N., Komori, H., Janssens, D.H., Kondo, S., Dai, Q., Moore, A.W., and Lee, C.-Y. (2020). Continual inactivation of genes involved in stem cell functional identity stabilizes progenitor commitment. *BioRxiv* 2020.02.03.931972.
- Rossi, A.M., and Desplan, C. (2020). Extrinsic activin signaling cooperates with an intrinsic temporal program to increase mushroom body neuronal diversity. *ELife* 9, e58880.
- Rossi, A.M., Fernandes, V.M., and Desplan, C. (2017). Timing temporal transitions during brain development. *Curr. Opin. Neurobiol.* 42, 84–92.
- Rougon, G., and Hobert, O. (2003). New Insights into the Diversity and Function of Neuronal Immunoglobulin Superfamily Molecules. *Annu. Rev. Neurosci.* 26, 207–238.
- Salomon, R., Kaczorowski, D., Valdes-Mora, F., Nordon, R.E., Neild, A., Farbehi, N., Bartonicek, N., and Gallego-Ortega, D. (2019). Droplet-based single cell RNAseq tools: a practical guide. *Lab. Chip* 19, 1706–1727.
- Samuels, T.J., Järvelin, A.I., Ish-Horowicz, D., and Davis, I. (2020). Imp/IGF2BP levels modulate individual neural stem cell growth and division through myc mRNA stability. *ELife* 9, e51529.
- 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.
- Satija lab (2020). Seurat - Guided Clustering Tutorial. Website - https://satijalab.org/seurat/v3.1/pbm3k_tutorial.html
- Schmid, A., Chiba, A., and Doe, C.Q. (1999). Clonal analysis of *Drosophila* embryonic neuroblasts. 37.
- Sen, S., Cao, D., Choudhary, R., Biagini, S., Wang, J.W., Reichert, H., and VijayRaghavan, K. (2014). Genetic transformation of structural and functional circuitry rewires the *Drosophila* brain. *ELife* 3, e04407.
- Sen, S.Q., Chanchani, S., Southall, T.D., and Doe, C.Q. (2019). Neuroblast-specific open chromatin allows the temporal transcription factor, Hunchback, to bind neuroblast-specific loci. *ELife* 8, e44036.
- Shukla, V., Habib, F., Kulkarni, A., and Ratnaparkhi, G.S. (2014). Gene Duplication, Lineage-Specific Expansion, and Subfunctionalization in the MADF-BESS Family Patterns the *Drosophila* Wing Hinge. *Genetics* 196, 481–496.

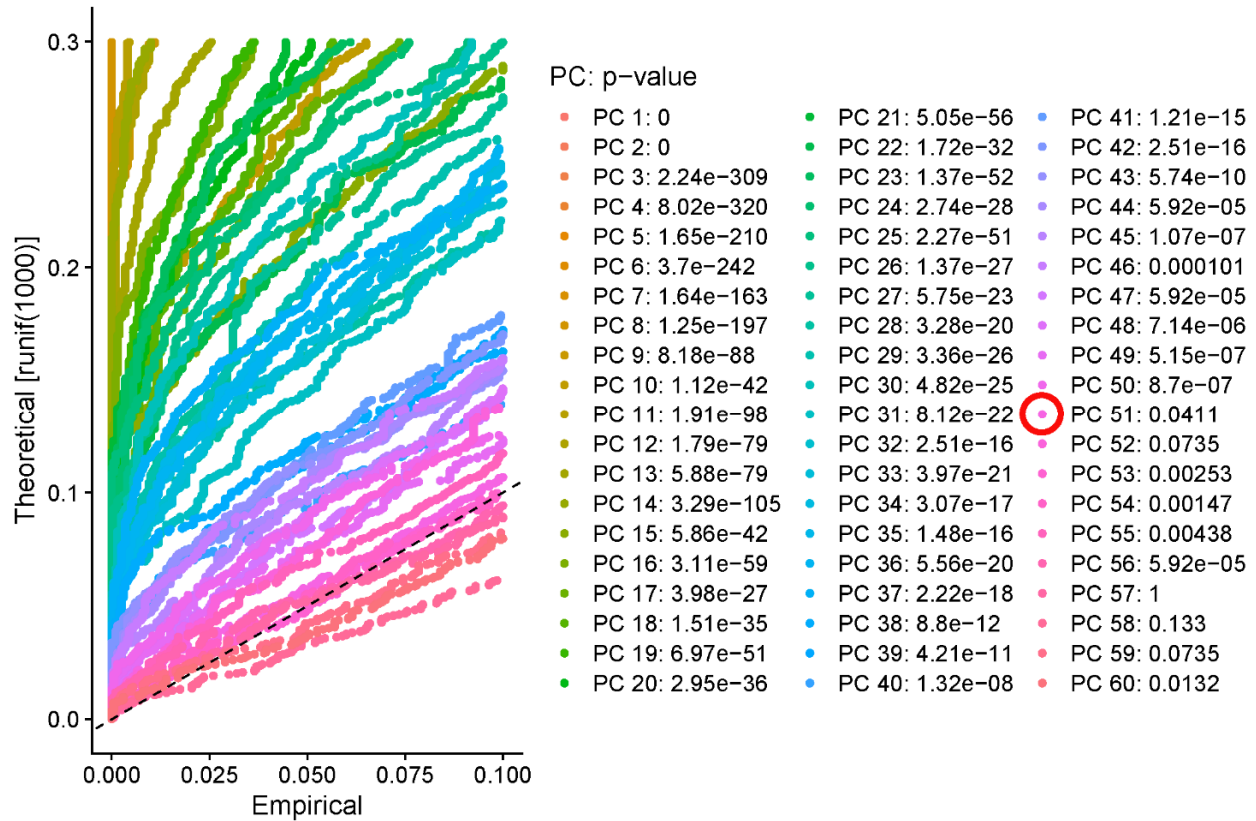
- Siebert, M., Banovic, D., Goellner, B., and Aberle, H. (2009). *Drosophila* motor axons recognize and follow a Sidestep-labeled substrate pathway to reach their target fields. *Genes Dev.* **23**, 1052–1062.
- Skeath, J.B. (1999). At the nexus between pattern formation and cell-type specification: the generation of individual neuroblast fates in the *Drosophila* embryonic central nervous system. *BioEssays* **10**.
- Slaidina, M., Banisch, T.U., Gupta, S., and Lehmann, R. (2020). A single-cell atlas of the developing *Drosophila* ovary identifies follicle stem cell progenitors. *Genes Dev.* **34**, 239–249.
- Soba, P., Zhu, S., Emoto, K., Younger, S., Yang, S.-J., Yu, H.-H., Lee, T., Jan, L.Y., and Jan, Y.-N. (2007). *Drosophila* Sensory Neurons Require Dscam for Dendritic Self-Avoidance and Proper Dendritic Field Organization. *Neuron* **54**, 403–416.
- 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.
- Starz-Gaiano, M. (2001). Wunen 2 activity guides germ cells. **9**.
- Stein, D.S., and Stevens, L.M. (2014). Maternal control of the *Drosophila* dorsal–ventral body axis. *Wiley Interdiscip. Rev. Dev. Biol.* **3**, 301–330.
- Svensson, V., Vento-Tormo, R., and Teichmann, S.A. (2018). Exponential scaling of single-cell RNA-seq in the past decade. *Nat. Protoc.* **13**, 599–604.
- Syed, M.H., Mark, B., and Doe, C.Q. (2017). Steroid hormone induction of temporal gene expression in *Drosophila* brain neuroblasts generates neuronal and glial diversity. **23**.
- Tan, L., Zhang, K.X., Pecot, M.Y., Nagarkar-Jaiswal, S., Lee, P.-T., Takemura, S., 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.
- Telley, L., and Jabaudon, D. (2018). A mixed model of neuronal diversity. *Nature* **555**, 452–454.
- Tirian, L., and Dickson, B.J. (2017). The VT GAL4, LexA, and split-GAL4 driver line collections for targeted expression in the *Drosophila* nervous system. *BioRxiv* 198648.
- Tritschler, S., Büttner, M., Fischer, D.S., Lange, M., Bergen, V., Lickert, H., and Theis, F.J. (2019). Concepts and limitations for learning developmental trajectories from single cell genomics. *Development* **146**, dev170506.
- Truman, J.W. (1990). Metamorphosis of the central nervous system of *Drosophila*. *J. Neurobiol.* **21**, 1072–1084.
- Truman, J.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*. *Dev. Camb. Engl.* **137**, 53–61.
- Tung, P.-Y., Blischak, J.D., Hsiao, C.J., Knowles, D.A., Burnett, J.E., Pritchard, J.K., and Gilad, Y. (2017). Batch effects and the effective design of single-cell gene expression studies. *Sci. Rep.* **7**, 39921.

- Uchida, O., Nakano, H., Koga, M., and Ohshima, Y. (2003). The *C. elegans* che-1 gene encodes a zinc finger transcription factor required for specification of the ASE chemosensory neurons. *Development* 130, 1215–1224.
- Urbach, R. (2003a). The pattern of neuroblast formation, mitotic domains and proneural gene expression during early brain development in *Drosophila*. *Development* 130, 3589–3606.
- Urbach, R. (2003b). Molecular markers for identified neuroblasts in the developing brain of *Drosophila*. *Development* 130, 3621–3637.
- Urbach, R., and Technau, G.M. (2003). Early steps in building the insect brain: neuroblast formation and segmental patterning in the developing brain of different insect species. *Arthropod Struct. Dev.* 32, 103–123.
- Urbach, R., Jussen, D., and Technau, G.M. (2016). Gene expression profiles uncover individual identities of gnathal neuroblasts and serial homologies in the embryonic CNS of *Drosophila*. *Development* 143, 1290–1301.
- Vogel, C. (2003). The immunoglobulin superfamily in *Drosophila melanogaster* and *Caenorhabditis elegans* and the evolution of complexity. *Development* 130, 6317–6328.
- Wangler, M.F., Yamamoto, S., and Bellen, H.J. (2015). Fruit Flies in Biomedical Research. *Genetics* 199, 639–653.
- Weiss, J.B., Von Ohlen, T., Mellerick, D.M., Dressler, G., Doe, C.Q., and Scott, M.P. (1998). Dorsoventral patterning in the *Drosophila* central nervous system: the intermediate neuroblasts defective homeobox gene specifies intermediate column identity. *Genes Dev.* 12, 3591–3602.
- Weng, R., and Cohen, S.M. (2015). Control of *Drosophila* Type I and Type II central brain neuroblast proliferation by bantam microRNA. *Dev. Camb. Engl.* 142, 3713–3720.
- 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.
- Wodarz, A., and Huttner, W.B. (2003). Asymmetric cell division during neurogenesis in *Drosophila* and vertebrates. *Mech. Dev.* 120, 1297–1309.
- Wolfram, V., Southall, T.D., Brand, A.H., and Baines, R.A. (2012). The LIM-Homeodomain Protein Islet Dictates Motor Neuron Electrical Properties by Regulating K⁺ Channel Expression. *Neuron* 75, 663–674.
- Yaghmaeian Salmani, B., and Thor, S. (2020). Genetic mechanisms controlling anterior expansion of the central nervous system. In *Current Topics in Developmental Biology*, (Elsevier), pp. 333–361.
- 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. *Development* 143, 411–421.
- Younossi-Hartenstein, A., Green, P., Liaw, G.-J., Rudolph, K., Lengyel, J., and Hartenstein, V. (1997). Control of Early Neurogenesis of the *Drosophila* Brain by the Head Gap Genes *stl*, *otd*, *ems*, and *btd*. *Dev. Biol.* 182, 270–283.

- Yu, H.-H., Awasaki, T., Schroeder, M.D., Long, F., Yang, J.S., He, Y., Ding, P., Kao, J.-C., Wu, G.Y.-Y., Peng, H., et al. (2013). Clonal Development and Organization of the Adult *Drosophila* Central Brain. *Curr. Biol.* **23**, 633–643.
- 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**, e1005264.
- Zacharioudaki, E., Magadi, S.S., and Delidakis, C. (2012). bHLH-O proteins are crucial for *Drosophila* neuroblast self-renewal and mediate Notch-induced overproliferation. *Dev. Camb. Engl.* **139**, 1258–1269.
- Zhu, S., Lin, S., Kao, C.-F., Awasaki, T., Chiang, A.-S., and Lee, T. (2006). Gradients of the *Drosophila* Chinmo BTB-Zinc Finger Protein Govern Neuronal Temporal Identity. *Cell* **127**, 409–422.
- Zhu, S., Barshow, S., Wildonger, J., Jan, L.Y., and Jan, Y.-N. (2011). Ets transcription factor Pointed promotes the generation of intermediate neural progenitors in *Drosophila* larval brains. *Proc. Natl. Acad. Sci.* **108**, 20615–20620.
- Zhu, S., Wildonger, J., Barshow, S., Younger, S., Huang, Y., and Lee, T. (2012). The bHLH Repressor Deadpan Regulates the Self-renewal and Specification of *Drosophila* Larval Neural Stem Cells Independently of Notch. *PLoS ONE* **7**.

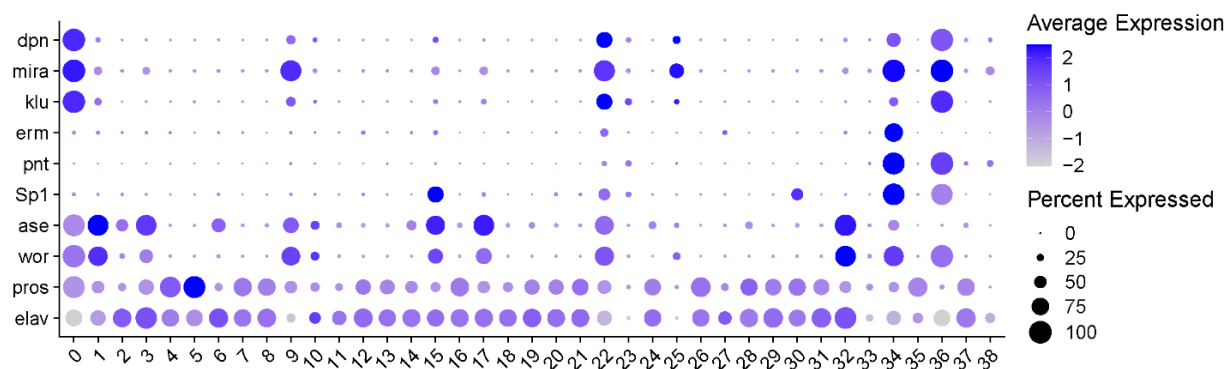
Supplementary information

Supplementary figures



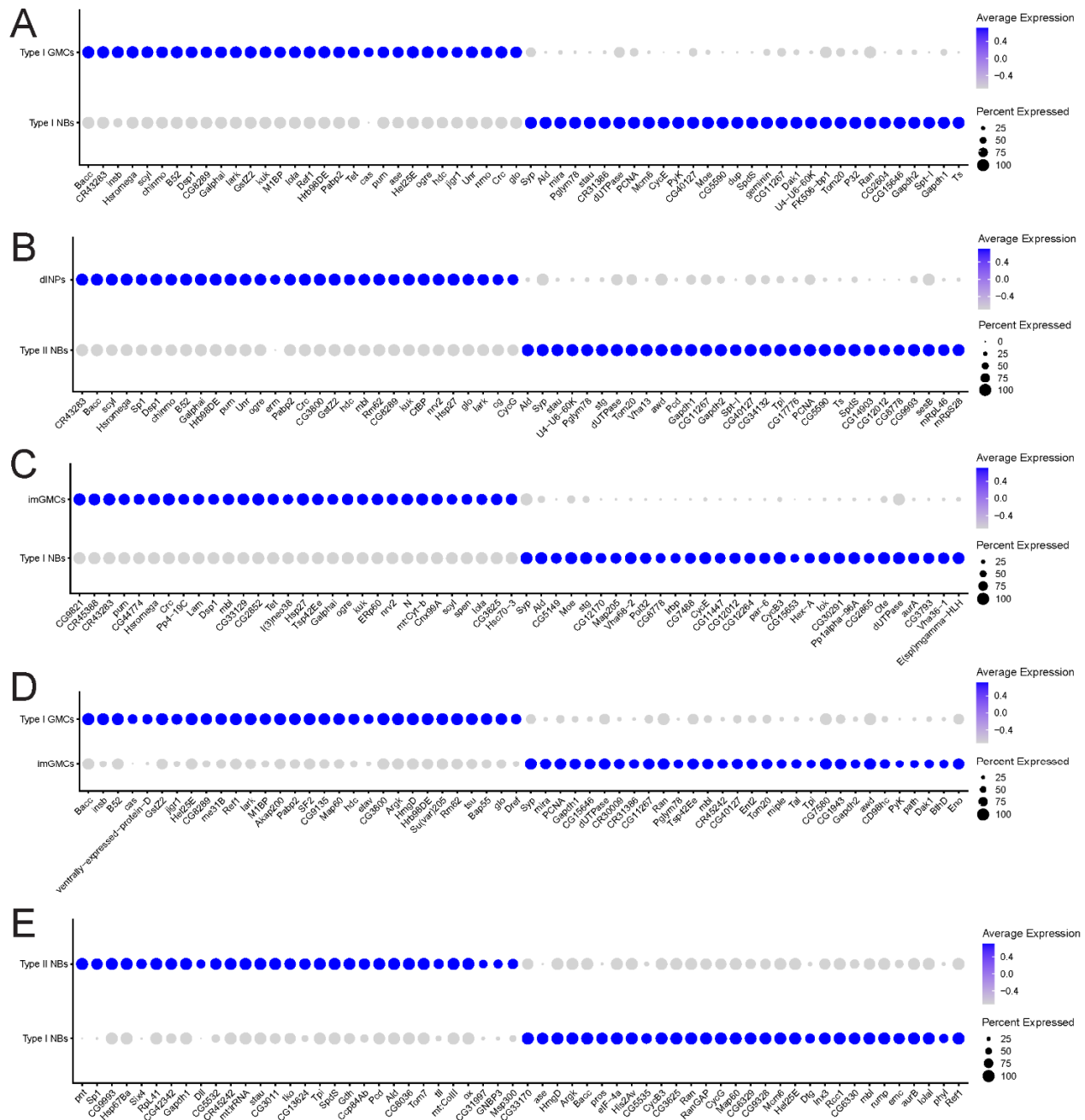
Suppl. Fig 1 Jackstraw results

Here are shown the outputs of the Jackstraw test used to decide on the number of PCs. As explained in Macosko *et al.*, and Seurat's documentation and tutorial, the jackstraw resampling test first picks random genes from a small percentage of the 2000 identified highly variable genes. Next, the genes in this matrix are "scrambled" and PCA is applied. This results in a "null distribution" which together with the initial PC loadings allows the calculation of a p-value which ranks the "association" of genes to PCs (Macosko *et al.*, 2015) (Satija lab, 2020). In the quantile-quantile plot (shown on the left), each line represents a different PC and indicates the quantile value comparison between a uniform distribution and the 2000 identified highly variable genes' p-values. PCs which coincide with the dashed line (uniform distribution) have higher p-values (Macosko *et al.*, 2015) (Satija lab, 2020). Furthermore, the Jackstraw calculates an overall p-value for all PCs, which we used to decide on the number of significant PCs (shown on the right) (Macosko *et al.*, 2015) (Satija lab, 2020). A red circle indicates the last chosen significant PC ($p < 0.05$), the rest were discarded.



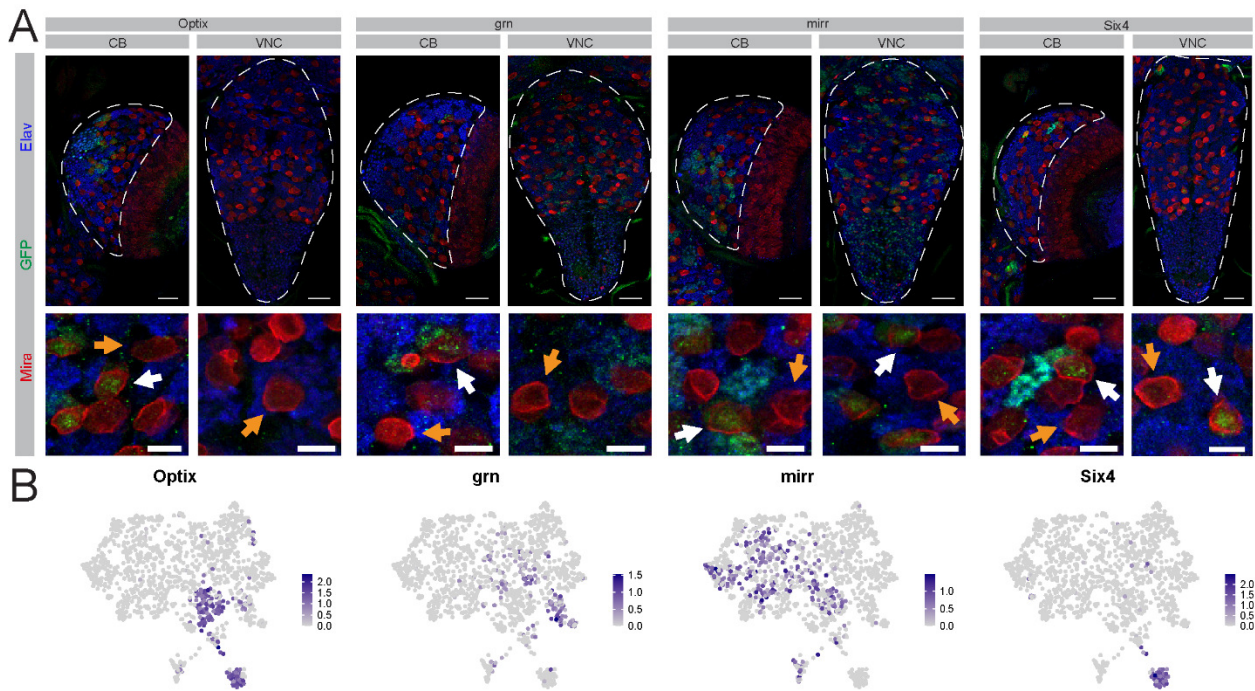
Suppl. Fig.2 Expression of NB lineage cell type specific markers in all clusters

Dot plot shows the expression of the NB lineage cell type specific markers (*dpm*, *mira*, *klu*, *erm*, *pnt*, *Sp1*, *ase*, *wor*, *pros*, and *elav*) used to group clusters into major cell groups. Cluster average scaled expression level is indicated by the colour scale. Dot dimension informs about the proportion of cells per cluster that express the NB lineage cell type specific markers.



Suppl. Fig.3 Cluster marker genes identified for each major cell group comparison

A) Dot plots show the expression of the top 30 marker genes for each major cell group comparison. The following comparisons are shown **A)** Type I NBs vs. Type I GMCs **B)** Type II NBs vs. dINPs **C)** Type I NBs vs. imGMCs **D)** imGMCs vs. Type I GMCs **E)** Type I NBs vs. Type II NBs. Major cell group average scaled expression level is indicated by the colour scale. Dot dimension informs about the proportion of cells per cluster that express the major cell group comparison marker genes.



Suppl. Fig.5 Additionally validated type I NB subcluster markers

A) Confocal microscopy images of the CNS of protein trap lines for *Optix*, *grn*, *mirr*, *Six4*. Staining was done for *mira* - NBs (red), GFP (green), *elav* - neurons (blue). For each genotype is shown a Z-stack slice of the superficial layer of the anterior and ventral side of the CB and VNC, respectively, with the exception of the CB for *grn*, which was rotated dorsally (n=5). White arrows and orange arrows indicate a positive NBs and a negative NBs, respectively, for protein trap expression. Scale bar – 30 microns. Below it is shown an inset of each image; scale bar – 10 microns. **B)** UMAP plot shows type I NB subclusters and marker normalized expression per cell (*Optix*, *grn*, *mirr*, and *Six4*).

Supplementary files

Excel file – ‘*Suppl._Excelfile_NBSubclusterMarkers.xlsx*’

Suppl. file 1 All type I NB subcluster markers

All identified marker genes for each NB subcluster are shown in this file. The column ‘*myAUC*’ shows the AUC “classifier” for each gene, which informs about the potential of a given marker gene. This value potentially can vary from > 0.5 to 1, where a value of 1 would indicate that gene is a perfect positive marker (Satija lab, 2020). The column ‘*avg_diff*’ shows the average log fold change for a given NB subcluster versus all other cells. ‘*Power*’ indicates the same as ‘*myAUC*’ but within a scale of 0 to 1. The columns ‘*pct.1*’ and ‘*pct.2*’ inform about the proportion of cells expressing the NB subcluster markers, in the cells of the cluster studied (*pct.1*) and in all other cells (*pct.2*).

Excel file - *Suppl._Excelfile_NBSubclusterMarkersTFs.xlsx*

Suppl. file 2 – ‘Transcription factor genes identified to be heterogeneously expressed in the type I NB subclustering analysis

All identified TF marker genes are shown in this table together with the information on which cluster they were found enriched. Here are also shown the references for the studies which have previously identified these TF to be differentially expressed between NBs. Two checkbox columns entitle ‘*Embryo*’ and ‘*Larva*’ can be found checked according to which developmental stage was studied in the papers referenced.