



NOVA MEDICAL
SCHOOL



CHARACTERIZATION OF NOVEL FACTORS REGULATING NEURAL STEM CELL PROLIFERATION, TEMPORAL FATE AND NEURONAL DIVERSITY

GRAÇA SUSETE COSTA DE CARVALHO MARQUES

Tese para obtenção do grau de Doutor em Mecanismos de Doença e Medicina Regenerativa

Doutoramento em associação entre:

Universidade NOVA de Lisboa (Faculdade de Ciências Médicas | NOVA Medical School)

Universidade do Algarve (UAlg)

Junho, 2021



NOVA MEDICAL
SCHOOL



CHARACTERIZATION OF NOVEL FACTORS REGULATING NEURAL STEM CELL PROLIFERATION, TEMPORAL FATE AND NEURONAL DIVERSITY

Graça Susete Costa de Carvalho Marques

Orientadores: Catarina de Cértima Fernandes Homem, Investigadora Principal, CEDOC,
FCM|NMS/UNL

**Tese para obtenção do grau de Doutor em Mecanismos de Doença e Medicina
Regenerativa**

Doutoramento em associação entre:

Universidade NOVA de Lisboa (Faculdade de Ciências Médicas | NOVA Medical School)

Universidade do Algarve (UAlg)

Junho, 2021

The chapters presented in this thesis refer to work performed under the supervision of Doctor Catarina Homem. I hereby state that I have fully participated in the conception and execution of the experimental work, in the interpretation of the collected data and in the writing of the manuscript. The work was approved by the Ethical Committee of NMS|FCM-UNL (109/2021/CEFCM)

The experimental work was performed at CEDOC, Chronic Diseases Research Centre, Nova Medical School|Faculdade de Ciências Médicas da Universidade Nova de Lisboa, New York University, New York and New York University, Abu Dhabi.

FUNDING

Fundação para a Ciência e a Tecnologia (FCT, Portugal)

PD/BD/114253/2016 (to Graça Susete Costa de Carvalho Marques) and IF/01265/2014/CP1252/CT0004.

With support from Fundo Social Europeu (SE) and Programa Operacional Capital Humano (POCH)

European Research Council (ERC)

H2020-ERC-2017-STG-GA 759853-StemCellHabitat, under the European Union's Horizon 2020 research and innovation programme

Wellcome Trust and Howard Hughes Medical Institute

HHMI-208581/Z/17/Z-Metabolic Reg SC fate

EMBO

H2020-EMBO-3311/2017/G2017 (EMBO Installation grant)

CONGENTO

CONGENTO LISBOA-01-0145-FEDER- 022170 (to CEDOC's Fly facility)

New York University, Abu Dhabi

NYU Abu Dhabi Research Institute grant to the NYUAD Center for Genomics and Systems Biology (ADHPG-CGSB).

The scientific content of the present thesis originated:

Peer-reviewed papers

Marques, G.S., Teles-Reis, J., Konstantinides, N., Brito, P.H., and Homem, C.C.F. (2021). Fate transitions in *Drosophila* neural lineages: a single cell road map to mature neurons. *BioRxiv* 2021.06.22.449317 (uploaded to bioRxiv)

Oral communications in scientific conferences:

Marques, G.S.; Homem, C.C. Identifying novel factors regulating neural stem cell proliferation and temporal fate. 2019 Portuguese *Drosophila* Meeting, 13th and 14th September 2019. Batalha, Portugal.

Marques, G.S.; Homem C.C. Novel factors regulating neural stem cell proliferation and temporal fate. 26th European *Drosophila* Research Conference, 5th-8th September 2019. Lausanne, Switzerland.

Poster presentations in scientific conferences:

Marques G.S.; Homem, C.C. Identification of new regulators of neuroblast proliferation and neuronal fate. 2nd NMS Symposium – Chronic Diseases and Translational Science 2018. 18th and 19th October 2018. Lisbon, Portugal.

Marques, G.S.; Homem, C.C. Understanding neuroblast fate regulation throughout development. 2018 Portuguese *Drosophila* meeting. 7th and 8th September 2018. Tomar, Portugal.

Marques, G.S.; Homem, C.C. 2017. Characterization of Novel Factors Regulating Neural Stem Cell Proliferation, Temporal Fate and Neuronal Diversity. 2017 Portuguese *Drosophila* meeting. 8th and 9th September 2017. Tomar, Portugal.

Oral communications in internal scientific meetings:

Marques, G.S.; Homem, C.C. Identifying novel factors in the regulation of neural stem cell proliferation and differentiation. Casual Friday – CEDOC, 8th November 2019. Lisbon, Portugal.

Marques, G.S.; Homem, C.C. Identification of new regulators of neural stem cell proliferation and temporal fate. ProRegeM Annual Meeting 2018, 13th and 14th December 2018. Gambelas, Faro, Portugal.

Marques, G.S.; Homem CC. Temporal Regulation of Neural Stem Cells in *Drosophila melanogaster*. 4th Annual Meeting of ProRegeM, 14th and 15th December 2017. Lisbon, Portugal.

Marques, G.S.; Homem, C.C. Temporal Regulation of Neural Stem Cells in *Drosophila melanogaster*. II Retiro de Estudantes Doutoramento do Centro de Estudos de Doenças Crónicas (CEDOC). 22th-24th September 2017. Lisbon and Cerdeira Village – Lousã, Portugal.

Marques, G.S.; Homem, C.C. 2017. Characterization of Novel Factors Regulating Neural Stem Cell Proliferation, Temporal Fate and Neuronal Diversity. 3rd Annual Meeting of ProRegeM. 9th and 10th January 2017. Gambelas, Faro, Portugal.

Para os meus pais Licínia e Armando
e para a minha irmã Licínia Isabel.

Table of Contents

LIST OF TABLES	iv
LIST OF ACRONYMS AND ABBREVIATIONS	vii
ACKNOWLEDGMENTS	xiii
INTRODUCTION	1
1. Brain development and neuronal diversity	3
2. <i>Drosophila</i> as an animal model	3
2.1. Life Cycle	5
3. Neuroblast generation in <i>Drosophila</i>	6
4. Asymmetric cell division in NBs.....	10
5. NB Lineages	11
6. Spatial patterning in <i>Drosophila</i>	14
6.1. Embryonic body segmentation	14
6.2. Spatial patterning in <i>Drosophila's</i> brain.....	16
7. Temporal transcription factors – temporal identity	18
7.1. Embryonic TTFs.....	18
7.2. Larval TTFs.....	21
8. NB lifespan	22
9. Parallels to mammals.....	24
10. Defects caused by misregulation of proliferation and lineage progression	27
11. Neuronal generation and maturation	28
12. Transcriptomics at the single-cell level	30
13. <i>Drosophila</i> Toolkit.....	32
13.1. Gal4/UAS	33
13.2. RNAi interference	33
13.3. Temporal and regional gene expression targeting	34
13.4. Gal4 technique for real-time and clonal expression	35
OBJECTIVES	37
CHAPTER I	41
I.1. Introduction	43
I.2. Materials and Methods	45
I.3. Results	46
I.3.1. Selecting new candidate regulators of neuroblast proliferation and lifespan	46
I.3.2. Candidate 1: Stat92E	49

I.3.3.	Candidate 2: TFIIIF α	52
I.3.3.1.	Knock down of TFIIIF α leads to longer-lived neuroblasts	52
I.3.3.2.	Knock down of other members of the RNA PolII transcriptional machinery increases neuroblast lifespan.....	54
I.3.3.3.	Knock down several members of the RNA Polymerase II transcriptional machinery alters NB lineage progression	57
I.3.4.	Gap genes as candidates for neuroblast fate and lifespan regulation	59
I.4.	Discussion.....	62
CHAPTER II		65
II.1.	Introduction	67
II.2.	Materials and Methods	69
II.3.	Results	74
II.3.1.	Neuroblast lineage labeling to identify new fates	74
II.3.2.	Selection of timepoints to assess within the animal development	77
II.3.3.	Assess the transcriptome of neuroblast lineages through scRNA-Seq ...	79
II.3.4.	Bioinformatic Setup of transcriptomic data of <i>Drosophila</i> 's developing brain	82
II.4.	Discussion.....	88
CHAPTER III		91
III.1.	Introduction	93
III.2.	Materials and Methods	96
III.3.	Results	102
III.3.1.	Atlas of early fates in NB lineages in CB and VNC	102
III.3.2.	Validation and characterization of imGMCs as a transition state between type I NBs and GMCs	106
III.3.3.	Identification of candidate regulators of neuronal lineage progression by differential gene expression analysis.....	108
III.3.4.	Transcriptomic differences distinguish cells with different degrees of differentiation	112
III.3.5.	Transcriptomic differences distinguish neurons by age.....	115
III.3.6.	Neurons in phase 2 of maturation transcribe the neurotransmitter gene ChAT, but do not translate it into protein	117
III.3.7.	Initiation of ChAT translation is coordinated with animal developmental stage rather than neuron age	120
III.4.	Discussion.....	123
CONCLUDING REMARKS		127
CONCLUDING REMARKS		129
SUPPLEMENTARY DATA		133

Gene ontology analysis	135
REFERENCES	141

LIST OF TABLES

Table I.1. Summary of phenotypes obtained when knocking down the expression of candidate genes specifically in neuroblasts	48
Table II. 1. Number of cycles used for the cDNA amplification step in the 10x Genomics protocol.....	72
Table II. 2. Summary of bioinformatic filters used for each scRNA-Seq dataset and respective outputs	73
Table II. 3. Summary of bioinformatic settings used for each scRNA-Seq dataset and respective number of identified differentiation states.....	74
Table II. 4. Developmental timings of the larval timepoints used for scRNA-Seq analysis	78
Table II. 5. Amount of material (brains and cells) processed for scRNA-Seq	80
Table II. 6. Preliminary Summary according to Cell Ranger output	81
Table S1. Gene ontology analysis of group 1 - genes identified in the neuronal population after analysis with Monocle	135
Table S2. Gene ontology analysis of group 2 - genes identified in the neuronal population after analysis with Monocle	137

LIST OF FIGURES

Figure 1. Life cycle of <i>Drosophila</i> .	6
Figure 2. Regions of the <i>Drosophila</i> central nervous system in embryo and larvae.	8
Figure 3. Asymmetric division of neuroblasts.	11
Figure 4. Types of asymmetric cell division in neuroblasts.	13
Figure 5. Regions of the embryonic <i>Drosophila</i> brain.	17
Figure 6. Neuroblast temporal identity during embryonic stages.	18
Figure I. 1. Experimental design to access candidate impact in neuroblast lifespan regulation.	47
Figure I. 2. RNAi and dominant negative variants of Stat92E do not consistently increase type I and type II neuroblast lifespan.	50
Figure I. 3. Stat92E is expressed in glia cells but not in NBs.	52
Figure I. 4. RNAi of TFIIFa increases type I neuroblast lifespan.	53
Figure I. 5. RNAi of TFIIFa increases type II neuroblast lifespan.	54
Figure I. 6. RNAi of different members of the transcriptional machinery increase type II neuroblast lifespan.	56
Figure I. 7. RNAi of different members of the RNA Polymerase II transcriptional machinery alter type II neuroblast lineages in wandering 3 rd instar larvae.	58
Figure I. 8. tll and otd gap genes are expressed in neuroblasts and neurons in wandering 3 rd instar larvae.	61
Figure II. 1. Neuroblast and neuron cell populations can be temporally and spatially labeled and sorted by FACS.	76
Figure II. 2. Quality control and characterization of clustering (105h ALH dataset).	84
Figure II. 3. Expression profile of neural cell markers in the 105h ALH dataset.	87
Figure III. 1. Type I and type II NB lineages identified by scRNA-Seq in wandering larvae brain.	103
Figure III. 2. Characterization of NB-imGMCs-GMCs transition in type I lineages.	107
Figure III. 3. Differentially expressed markers throughout the different differentiation states in NB lineages.	109
Figure III. 4. Transcriptomic dynamics of differentiation in type I and type II NB lineages.	114

List of figures

Figure III. 5. Transcriptomic differences in neurons identify 2 phases of maturation.. 116

Figure III. 6. ChAT expression in neurons with different ages..... 119

Figure III. 7. Neuronal maturation occurs in 3 phases possibly regulated at the translation level. 122

LIST OF ACRONYMS AND ABBREVIATIONS

AbdA	Abdominal-A
AD	Abu Dhabi
AL	Antennal lobe
ALS	Amyotrophic Lateral Sclerosis
Ana	Anachronism
<i>Antp</i>	<i>Antennapedia</i>
<i>ap</i>	<i>apterous</i>
APF	After puparium formation
aPKC	Atypical protein kinase C
Ase	Asense
AUC	Area under the ROC curve
Baz	Bazooka (also, Par3)
<i>bcd</i>	<i>bicoid</i>
BCL6	B-cell lymphoma 6
BDSC	Bloomington Drosophila Stock Center
bHLH	basic helix-loop-helix proteins
Brat	Brain tumor
BSA	Bovine serum albumin
Bsh	Brain-specific homeobox
<i>cad</i>	<i>caudal</i>
<i>cas</i>	<i>castor</i>
Casz1	castor zinc finger 1
CB	Central brain
cDNA	complementary DNA
ChAT	Choline acetyltransferase
Chinmo	Chronologically inappropriate morphogenesis
CNS	Central nervous system
<i>cpx</i>	<i>complexin</i>
CX	Central complex
CycE	Cyclin E
D	Deutocerebrum
D	Dichaete
Dap	Dacapo
<i>dati</i>	<i>datilografo</i> (also known as <i>Lin29</i>)
DDRTree	Discriminative dimension reduction tree
dINP	developing INP
DNA	Deoxyribonucleic acid
DNAB	DNA binding
DnaJ-1	DnaJ-like-1
Dpn	Deadpan
DSHB	Developmental Studies Hybridoma Bank
<i>Dsp1</i>	<i>Dorsal switch protein 1</i>
dsRNA	Double-stranded RNA
<i>E(spl)mγ-HLH</i>	<i>Enhancer of split mγ, helix-loop-helix</i>

List of acronyms and abbreviations

EcRB1	Ecdysone receptor B1
EGFP	Enhanced green fluorescent protein
Eip93F	Ecdysone-induced protein 93F (also known as E93)
<i>elav</i>	<i>embryonic lethal abnormal visual system</i>
<i>EloC</i>	<i>Elongin C</i>
<i>ems</i>	<i>empty spiracles</i>
Erm	Earmuff
Ets	E26 transformation-specif
Ey	Eyeless
FACS	Fluorescence activated cell sorting
FBS	Fetal bovine serum
FDS	Fat body derived signal
FGF	Fibroblast growth factor
FLP	Flippase recombinase
Foxo	Forkhead box, sub-group O
FRT	Flippase recombination target
FSC-A	Forward scatter-area
fTRG	FlyFos transgenome
Gad1	Glutamatic acid decarboxylase 1
GEM	Gel bead-in-emulsions
GFP	Green fluorescent protein
GMC	Ganglion mother cell
<i>grh</i>	<i>grainyhead</i>
GRHL1-3	Grainyhead-like 1-3
Gsb	Gooseberry
GTF2F1	General transcription factor IIF subunit 1
GTF2F2	General transcription factor IIF subunit 2
G-TRACE	Gal4 technique for real-time and clonal expression
Gai	G protein α i subunit
Ham	Hamlet
<i>Hat1</i>	<i>Histone acetyltransferase 1</i>
<i>hb</i>	<i>hunchback</i>
Hey	Hairy/E(spl)-related with YRPW motif
Hh	Hedgehog
Hid	Involution defective
<i>HmgD</i>	<i>High mobility group protein D</i>
Hsp26	Heat shock protein 26
Hsp67Bc	Heat shock gene 67Bc
Hsp68	Heat shock protein 68
Hsp83	Heat shock protein 83
Hsromega	Heat shock RNA omega
IGF2BP1	insulin like growth factor 2 mRNA binding protein 1
<i>Ikzf1</i>	<i>Ikaros/Znfn1a1</i>
ILP	Insulin like peptide
IM14	Immune induced molecule 14 (also known as Daisho2, Dso2)
imGMC	immature ganglion mother cell
imINP	Immature INP
Imp	IGF-II mRNA-binding protein

INP	Intermediate neural progenitor
InR	Insulin receptor
Insc	Inscuteable
IPC	Intermediate Progenitor Cell
JAK/STAT	Signal transducer and activator of transcription
<i>jigr1</i>	<i>jing interacting gene regulatory 1</i>
<i>Klu</i>	<i>Klumpfuss</i>
<i>kni</i>	<i>knirps</i>
<i>koi</i>	<i>klaroid</i>
<i>Kr</i>	<i>Krüppel</i>
<i>l(3)neo38</i>	<i>lethal (3) neo38</i>
<i>l'sc</i>	<i>lethal of scute</i>
<i>lola</i>	<i>Longitudinals lacking</i>
MB	Mushroom body
mid	midline
mINP	Mature INP
Mira	Miranda
mRNA	Messenger RNA
msi	musashi
mTORC1	mechanistic target of rapamycin complex 1
<i>nAChRalpha2</i>	<i>nicotinic acetylcholine receptor alpha2</i>
NB	Neuroblasts
nerfin-1	nervous fingers-1
<i>Nos</i>	<i>Nanos</i>
<i>nrv3</i>	<i>nervana 3</i>
NSC	Neural stem cell
<i>nSyb</i>	<i>neuronal synaptobrevin</i>
NT	Neurotransmitter
Nub	Nubbin
NY	New York
NYU	New York University
<i>oc</i>	<i>ocelliless (also known as orthodenticle, otd)</i>
OL	Optic lobe
<i>otd</i>	<i>orthodenticle (also known as ocelliless, oc)</i>
P	Protocerebrum
Paip2	polyA-binding protein interacting protein 2
PAN	Posterior Asense-negative
PBS	Phosphate buffered saline
PC	Principal component
PCA	Principal component analysis
PCR	Polymerase chain reaction
Pdm1	POU-domain protein 1 (also known as Nubbin, Nub)
Pdm2	POU-domain protein 2
PH3	phospho-histone H3
PI3K	phosphatidylinositol 3-kinase
Pins	Partner of inscuteable (also known as raps)
Pnt	Pointed
PolII	Polymerase II

List of acronyms and abbreviations

POU2F1	POU class 2 homeobox 1
POU2F2	POU class 2 homeobox 2
Pros	Prospero
PTEN	phosphatase and tensin
pum	pumilio
QC	Quality control
Repo	Reverse polarity
RFP	Red fluorescent protein
RHG	Reaper/Hid/Grim dependent apoptosis
RNA	ribonucleic acid
RNAi	RNA interference
RNA-Seq	RNA-Sequencing
ROC	Receiver operating curve
Rpr	Reaper
<i>sbb</i>	<i>scribbler</i>
scATAC-Seq	Single cell Assay for Transposase-Accessible Chromatin Sequencing
scRNA-Seq	Single cell RNA-Sequencing
Sirt2	Sirtuin 2
smFISH	single molecule fluorescence in situ hybridization
<i>Smox</i>	<i>Smad on X</i> (also known as <i>Smad2</i>)
<i>spen</i>	<i>split ends</i>
Sqz	Squeeze
Stat92E	Signal-transducer and activator of transcription protein at 92E
Stip1	Stress induced phosphoprotein 1 (also known as Hsp70/Hsp90 organization protein, HOP)
<i>Su(var)205</i>	<i>Suppressor of variegation 205</i>
Svp	Seven-up
SWI/SNF	Switch/Sucrose non-fermenting
SYNCRIP	synaptotagmin binding cytoplasmic RNA interacting protein
Syp	Syncrip
T	Tritocerebrum
Taf	TBP-associated factor
tap	target of Poxn
TARGET	Temporal and regional gene expression targeting
TF	Transcription factor
TFIIB	Transcription factor IIB
TFIID	Transcription factor IID
TFIIFα	Transcription factor IIF α
TFIIFβ	Transcription factor IIF β
TFIIS	Transcription factor IIS
<i>tll</i>	<i>tailless</i>
TOR	Target of rapamycin
Toy	Twin of Eyless
TRiP	Transgenic RNAi Project
<i>trol</i>	<i>terrible reduced optic lobes</i>
ts	Temperature sensitive
<i>t-SNE</i>	<i>t</i> -distributed stochastic neighbor embedding

TTF	Temporal transcription factor
UAS	Upstream activating sequence
UMAP	Uniform manifold approximation and projection
UMI	Unique molecular identifier
VACHT	Vesicular acetylcholine transporter
VDRC	Vienna Drosophila Research Center
VGlut	Vesicular glutamate transporter
Vmat	Vesicular monoamine transporter
VNC	Ventral nerve cord
<i>vnd</i>	<i>ventral nervous system defective</i>
VT	Vienna Tiles
Wor	Worniu
WT	Wild type
<i>zld</i>	<i>zelda</i> (also known as <i>vfl</i>)

ACKNOWLEDGMENTS

Em primeiro gostaria de agradecer à Doutora Catarina Homem por me ter recebido no seu laboratório, dando-me a oportunidade de desenvolver o meu projeto de doutoramento. Agradeço-lhe a sua cuidada supervisão, bem como o interesse que sempre demonstrou no desenvolvimento deste trabalho. Agradeço ainda a confiança que sempre demonstrou no meu trabalho e a forma como me ajudou a evoluir como investigadora.

To Doctor Claude Desplan, for inviting me to his lab, and for allowing me to perform a crucial part of my project there. I would also like to thank him for all the enthusiasm he has always shown regarding my project. I am also thankful to Doctor Claude's groups, both in New York and Abu Dhabi, for welcoming me and for their interest in helping and discussing my project. I would like to thank particularly to Nikos for his help and for always being available for my questions. I also want to thank Doctor Katarina Kapuralin for her help with my experiments and for sharing her know-how with me. Queria também agradecer ao Filipe pelo interesse que sempre demonstrou em discutir o meu trabalho e pelas suas sugestões para o melhorar. Agradeço também à Lígia por ter sido uma companhia de aventuras em Abu Dhabi.

Ao José e à Patrícia (Grácio), por me terem acompanhado na saga da bioinformática. Por tudo o que aprendemos juntos e pela forma como nos fomos ajudando a resolver os problemas e que tanto contribuíram para melhorar o nosso trabalho.

Aos meus colegas e amigos, atuais e antigos, do grupo Proliferation and Fate Regulation of Stem Cells. Especialmente à Márcia, à Andreia, à Margarida, ao Tiago, à Rita e à Eunice, agradeço pelos bons momentos passados no laboratório, sem os quais tudo teria sido mais difícil. Ou certamente, pelo menos mais aborrecido! Agradeço também o interesse que sempre demonstraram pelo meu trabalho e todas as discussões científicas que tanto contribuíram para o desenvolvimento deste projeto. Agradeço ainda a todos a forma como sempre se mostraram prestáveis para me ajudar e a forma entusiasta como o faziam.

Aos meus colegas do ProRegeM, Liliana, Daniela, Maria João, Nuno e Pedro. Por tudo o que passámos juntos enquanto alunos de doutoramento: muito trabalho, fortes gripes, mas também bons momentos de descontração. Pelos almoços, pelos passeios e pelo esforço que sempre demonstraram por nos mantermos em contacto.

Acknowledgments

Às minhas grandes amigas Clara e Juliana, por estarem sempre por perto. Pelos nossos jantares “quinzenais” e outros encontros ocasionais. Agradeço a amizade e a partilha de alegrias e frustrações que tanto contribuíram para aligeirar os momentos mais difíceis e para recarregar baterias!

Agradeço também à Inês e à Maria, por continuarem a estar presentes e a contribuir com apoio e animação sempre que necessário.

Por fim, agradeço aos meus pais, Armando e Licínia, e à minha irmã Licínia Isabel, pelo vosso apoio, preocupação com o meu bem-estar e por sempre torcerem pelo meu sucesso.

Aos restantes amigos e colegas, por de uma forma ou de outra, me terem apoiado ao longo desta fase da minha vida.

ABSTRACT

A brain is formed from a small number of neural stem cells (NSCs), that give rise to the large neuronal diversity in the adult. For the brain to properly function it is important that the right neurons are generated at the right time. The *Drosophila* brain has proven to be a great model to understand how the brain develops; it contains a small set of NSCs, the neuroblasts (NBs), that generate invariable neuronal lineages in precise developmental windows, thus making it a perfect system to study the mechanisms of neural differentiation and cellular diversity. The transcriptional profile of NBs and their lineages are essential to specify the neuron type generated. After neurons are formed their neuronal terminal features, essential for neuron function, start being expressed and are maintained throughout neuron life.

Thus, for proper brain development to occur, NBs must express the correct sequence and combination of transcription factors (TFs) throughout brain development. Moreover, the differentiation process from NB to neurons also needs to be tightly regulated so that the right neurons can be generated. Though many advances have been made in this area, the complete set of genes that regulates lineage progression is still not known, neither are the full set of mechanisms that lead to proper neuronal maturation and function.

NBs divide to generate neural cells in very defined windows during development, and once their lineages are complete, they disappear. The timings for NB disappearance differ between NB sub-types, and although some of the molecular mechanisms that drive final disappearance, as apoptosis or autophagy are known, it is still unclear how these mechanisms are temporally coordinated with brain development.

In **Chapter I** of this thesis, I focused on central brain (CB) NBs which disappear in early pupal stages. Terminal differentiation of these NBs is known to be regulated in part by a pulse of the hormone ecdysone, however, it is unknown if there are also temporal TFs fine tuning this process. Thus, to identify possible TF regulators of NB timely disappearance, I selected a set of candidates and assessed the impact that knocking them down in NBs had in their lifespan and proliferation. Only one of the tested candidates, TFIIF α , a member of the general RNA Polymerase II (PolII) transcriptional machinery, has proven to have an impact in NB lifespan, as TFIIF α knock down led to the presence of longer-lived pupal NBs. Moreover, down regulation of TFIIF α levels in NBs also led to defects in lineage differentiation and formation of ectopic NB-like cells already during larval stages. Other members of the RNA PolII transcriptional machinery

such as TFIIF β , TFIIB and TFIIIS have shown a similar phenotype, ultimately indicating that the transcriptional machinery may play an important role in regulating NB lineages.

In **Chapter II**, I describe the generation of a comprehensive single cell RNA sequencing (scRNA-Seq) atlas, composed of 6 smaller datasets ranging from late larval stages to early pupa. The goal of this effort was to generate a comprehensive pool of transcriptomic data of the developing *Drosophila* brain, comprising the several differentiation states within a NB lineage. These datasets include CB and ventral nerve cord (VNC) NBs and the respective lineages generated within restricted sequential temporal windows. To generate this data, I took advantage of a conditional genetic labeling strategy, that allowed to include NBs and their progeny, up until young neurons with less than 12.5h. Though, the full integration of all datasets is not completed yet, this atlas represents an important pool of information for future studies regarding a large variety of biological questions, from NB proliferation to differentiation.

Finally, in **Chapter III**, we selected one of the previously obtained datasets, collected at 105h after larval hatching, in order to study NB lineage differentiation, as well as neuronal maturation. We aimed to identify TFs that might be regulating fate changes within a NB lineage differentiation process. Interestingly, we have identified a transition state between type I NBs and their daughter cells, the ganglion mother cells (GMCs), that we called immature GMCs (imGMCs). This transitory cell state, seems to be transcriptionally different enough from NBs and GMCs to allow its independent identification; moreover, it shows that the fate change between NBs to GMCs is not immediate, but requires a lag period. By comparing the different cell differentiation states within a lineage, we were able to identify genes that are differentially expressed at each state, which makes them likely candidates for playing an important role in NB proliferation and differentiation.

Furthermore, the fact that these datasets only include very young neurons, allowed us to study the early steps of neuronal maturation. Our findings revealed that neuron maturation starts very quickly after neuron formation, and sub-divide the process of maturation into 3 distinct phases based on the dynamics of transcription with or without translation of mature neuronal features. Phase 1 is composed by immature neurons that have yet to start expressing mRNA of mature neuronal features; Phase 2 includes neurons that start transcribing but not translating maturation markers such as neurotransmitter genes; Phase 3 of neuron maturation, which is coordinated with the animal developmental stage, is when not only mRNA, but also the protein of more mature neuronal features is expressed.

Ultimately, the findings resulting from the work presented in this thesis have contributed to a better understanding of neuronal lineage progression and neuronal maturation. Moreover, the data generated in this study, represents a valid and useful resource for the scientific community interested in brain development and neuronal diversity.

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

RESUMO

O cérebro é formado por um pequeno número de células estaminais neuronais que dão origem à grande diversidade neuronal no adulto. Para que o cérebro funcione adequadamente, é importante que os neurónios certos sejam formados no momento certo. O cérebro de *Drosophila* provou ser um ótimo modelo para estudar o desenvolvimento cerebral pois contém um pequeno conjunto de células estaminais neuronais, os neuroblastos (NBs), que geram linhagens neuronais invariáveis ao longo de janelas de desenvolvimento precisas, tornando-se um sistema perfeito para estudar os mecanismos de diferenciação neuronal e diversidade celular. Os perfis transcricionais dos NBs e das suas linhagens são essenciais para especificar o tipo de neurónios formados. Depois dos neurónios serem formados, as suas características terminais neuronais, as quais são essenciais para o funcionamento do neurónio, começam a ser expressas e são mantidas ao longo da vida do neurónio.

Assim, para que desenvolvimento cerebral ocorra corretamente, os NBs devem expressar a sequência e combinação corretas de fatores de transcrição ao longo do desenvolvimento do cérebro. Além disso, o processo de diferenciação desde os NBs até aos neurónios também precisa de ser regulado de forma precisa para que os neurónios certos possam ser formados. Embora muitos avanços tenham sido feitos nesta área, o conjunto completo de genes que regula a progressão da linhagem ainda não é conhecido, o mesmo acontece com o conjunto completo de mecanismos que levam a uma maturação e função neuronal adequadas.

Os NBs dividem-se para gerar células neurais em janelas temporais bem definidas durante o desenvolvimento e uma vez que suas linhagens estejam completas, desaparecem. A altura em que os NBs desaparecem difere consoante os subtipos de NBs e embora alguns dos mecanismos moleculares que levam a este desaparecimento já sejam conhecidos, como é o caso da apoptose ou autofagia, ainda não está claro a forma como esses mecanismos são coordenados temporalmente relativamente ao desenvolvimento cerebral.

O **Capítulo I** desta tese, tem como foco os NBs do cérebro central, os quais desaparecem nos estadios iniciais de pupa. Sabe-se que a diferenciação terminal desses NBs é regulada, em parte, por um pulso da hormona ecdisona, no entanto, não se sabe se também existem fatores de transcrição temporais a regular este processo. Assim, para identificar possíveis fatores de transcrição responsáveis por regular o desaparecimento atempado dos NBs, selecionei um conjunto de candidatos e avaliei o impacto que diminuir a sua expressão tem nos NBs relativamente ao seu tempo de vida

e proliferação. Apenas um destes candidatos, o TFIIF α , um membro da maquinaria de transcrição associada à RNA Polimerase II (PolII), provou ter um impacto no tempo de vida dos NBs, pois a diminuição de TFIIF α levou a que alguns NBs sobrevivessem em pupa durante mais tempo. Além disso, a diminuição dos níveis de TFIIF α em NBs também levou a defeitos na diferenciação de linhagem neuronal e na formação de células ectópicas semelhantes a NBs já em larva. Outros membros da maquinaria transcricional associada à RNA PolII, como o TFIIF β , o TFIIIB e o TFIIIS, mostraram um fenótipo semelhante, indicando que a maquinaria transcricional pode desempenhar um papel importante na regulação das linhagens neuronais.

No **Capítulo II** encontra-se descrita a criação de um atlas abrangente de sequenciação do transcriptoma de células individuais (scRNA-Seq), composto por 6 conjuntos de dados de menores dimensões, que vão desde os últimos estádios larvares até ao início da pupa. Este esforço teve como objetivo gerar um conjunto abrangente de dados transcripcionais do cérebro em desenvolvimento de *Drosophila*, incluindo os vários estados de diferenciação dentro de uma linhagem neuronal. Este conjunto de dados inclui NBs do cérebro central e do cordão nervoso ventral, bem como as respetivas linhagens geradas dentro das janelas temporais sequenciais e restritas selecionadas. Para gerar estes dados, usei uma estratégia de marcação genética condicional, que permitiu incluir os NBs e toda a sua descendência até aos neurónios mais novos com menos de 12,5h. Embora a integração de todos os conjuntos de dados ainda não esteja concluída, este atlas representa um importante conjunto de informação para estudos futuros sobre uma grande variedade de questões biológicas, que vão desde a proliferação de NBs até à diferenciação neuronal.

Por fim, no **Capítulo III**, foi selecionado um dos conjuntos de dados obtidos anteriormente, correspondente às 105h após a formação da larva, para estudar a diferenciação da linhagem de NBs, bem como a maturação neuronal. O objetivo desta análise consistiu em identificar fatores de transcrição que potencialmente estivessem a regular as transições entre as diferentes fases de diferenciação de uma linhagem de NBs. Curiosamente, foi identificado um estado de transição entre os NBs do tipo I e as suas células-filhas, as células-mãe ganglionares (GMCs), o qual designámos GMCs imaturas (imGMCs). Este estado de diferenciação transitório, parece ser suficientemente diferente a nível transcricional quer dos NBs quer das GMCs, de forma a permitir sua identificação independente; além disso, a presença de imGMCs mostra que a diferenciação de NBs para GMCs não é imediata, requerendo um período de transição. Ao comparar os diferentes estados de diferenciação celular de uma linhagem, pudemos identificar genes que são expressos diferencialmente em cada estado, o que

os torna prováveis candidatos a desempenhar um papel importante na proliferação e diferenciação dos NBs.

Além disso, o fato deste conjunto de dados incluir apenas neurónios muito jovens, permitiu o estudo das primeiras etapas da maturação neuronal. Os resultados revelaram que a maturação do neurónio começa rapidamente após a formação do mesmo e levaram à divisão do processo de maturação em 3 fases distintas com base na dinâmica transcricional de genes associados a características neuronais mais maduras, acompanhada ou não de tradução das mesmas. Consequentemente, a Fase 1 é composta por neurónios imaturos que ainda não começaram a expressar mRNA de características neuronais maduras; a Fase 2 inclui neurónios que começam a transcrever, mas não a traduzir marcadores de maturação, como genes associados a neurotransmissores; na Fase 3 da maturação neuronal, a qual é coordenada com o estadio de desenvolvimento do animal, existe expressão de mRNA, mas também de proteína, de características neuronais mais maduras.

Em última análise, os resultados do trabalho apresentado nesta tese contribuíram para um melhor entendimento da progressão da linhagem neuronal e da maturação dos neurónios. Adicionalmente, os dados gerados neste estudo, representam um recurso válido e útil para a comunidade científica interessada no desenvolvimento cerebral e na diversidade neuronal.

Palavras-chave: Sequenciamento de RNA de células individuais (scRNA-Seq), *Drosophila*, maturação neuronal, Neuroblastos, Cérebro, Diferenciação de linhagens neuronais, Reguladores de destino neuronal; Tempo de vida do neuroblasto, Proliferação do neuroblasto.

INTRODUCTION

INTRODUCTION

1. Brain development and neuronal diversity

During brain development, neural stem cells (NSCs) proliferate in order to originate all the cells that will populate the brain of the adult animal.

Throughout time, NSCs divide to self-renew and simultaneously give rise to more differentiated cells that will ultimately become neurons or glia. In order for the limited number of NSCs to generate the wide variety and number of neurons in the adult brain, several levels of regulation have to occur. Though the mechanisms that regulate neural development still remain largely unsolved, it is known that in part, neuronal diversity is achieved by layers of temporal and spatial regulation that contribute to endow the cell with different fates and ultimately ensure that the right neurons are formed at the right time. During neurogenesis, it is important that the mechanisms that control cell proliferation are tightly controlled to avoid developmental defects. For instance, misregulation of the number and type of cells that are formed may lead to abnormalities, such as tumors or neurodevelopmental diseases (Homem and Knoblich, 2012; Homem et al., 2015).

These characteristics are recapitulated in *Drosophila* NSCs, the neuroblasts (NBs), which I will use in this study to identify possible regulators of NSC proliferation, lifespan and fate, as well as to better understand neuronal lineage differentiation regulation.

2. *Drosophila* as an animal model

Drosophila melanogaster (herein simply *Drosophila*), commonly known as the fruit or vinegar fly, is considered a very powerful animal model, having been used in the laboratory for over a century. It is a human commensal species that can be found in rotting fruit as well as other types of decaying matter (Hales et al., 2015).

Drosophila has been reared in laboratory and used as a research organism since the early beginnings of the 20th century, and its great boost in popularity occurred with the work of Thomas Hunt Morgan on chromosome inheritance (Morgan, 1910). These findings would eventually earn him the Nobel Prize in Physiology and Medicine in 1933, which together with the rest of his work contributed to Morgan being known until our days as the founder of *Drosophila* genetics. Other contributions from Morgan's lab include

Introduction

deducing the linear organization of genes in a chromosome (Sturtevant, 1913) and establishing the use of polytene chromosomes in *Drosophila* and their subsequent use for mapping (Bridges, 1935); ultimately, all these studies helped to establish *Drosophila* as a useful model organism that it is now used by several hundreds of groups all over the globe.

More groundbreaking discoveries took place with the use of *Drosophila*, some leading to 6 more investigators receiving the Nobel Prize until 2020; the latest of each being the 2017 Nobel Prize in Physiology and Medicine attributed jointly to Jeffrey C. Hall, Michael Rosbash and Michael W. Young with work exploring the molecular mechanism of the circadian rhythm.

Several aspects contributed for *Drosophila* to be so vastly used as an animal model. For instance, *Drosophila* can be easily cultured in the lab, with low expenses associated with its rearing; it also has a short life cycle, and it can generate high numbers of offspring. Moreover, it has a compact genome with only 4 chromosomes. Another advantage of *Drosophila* is being easy to genetically manipulate, which allows researchers to use a bigger number of genetic tools to study a wider range of biological questions.

Although *Drosophila* is a simple model, several processes are conserved from vertebrates, ranging from body axes patterning, organogenesis, cell proliferation and nervous system development. For instance, a study published in 2001 found that 77% of human disease genes were conserved in *Drosophila* (Reiter et al., 2001). Ultimately, using a simple system such as *Drosophila*, combined with the large amount of genetic tools that are available, allow for the study of genes and mechanisms that cause diseases in humans, making it a great asset to understand more complex systems (Mohr and Perrimon, 2019; Ugur et al., 2016). Conveniently, the Bloomington Drosophila Stock Center (BDSC) curates a stock collection with several fly lines that are associated with human diseases or health-related mechanisms and pathways (<https://bdsc.indiana.edu/stocks/hd/index.html>).

A field of knowledge that has greatly benefited from using *Drosophila* as a model animal is neurobiology. The important concepts and tools that were developed to study *Drosophila* since around 1910 have greatly contributed to multiple discoveries related to the development of the nervous system, such as NSC maintenance and differentiation, neurogenesis, neuron arborization, behavior, learning and memory. A thorough chronology of the most relevant genes and tools discovered in *Drosophila* until the beginning of this century can be found in (Bellen et al., 2010). For instance, *Drosophila* has proven to be a powerful model to study neurodegenerative disorders such as

Alzheimer's Disease, Parkinson's Disease, Huntington's Disease and Amyotrophic Lateral Sclerosis (ALS). Moreover, other complex neurodegenerations such as traumatic brain injury and mitochondrial dysfunction have also benefited from being studied using a simpler model such as *Drosophila* (Bulus et al., 2020; McGurk et al., 2015).

Furthermore, efforts from a large group of scientists allowed the generation of a detailed atlas reflecting the circuitry of most neurons in the adult central brain (CB) of *Drosophila*. By consulting these data within the available interactive tool, one is able to explore brain wiring and connections between neurons (Xu et al., 2020). Considering the high genetic convergence observed in *Drosophila* when compared to more complex animals, this "connectome" might prove to be a useful tool in unraveling brain function.

Ultimately, using *Drosophila* as a genetic animal model has the potential to revolutionize our understanding in areas such as neurobiology, behavior, evolution, and gene regulation.

2.1. Life Cycle

As previously mentioned, one of the main advantages of working with *Drosophila* is its short life cycle, as it allows the fast generation of progeny to be used in experiments.

Under optimal laboratory conditions, meaning 25°C and 60% humidity, *Drosophila* takes approximately 10 days to develop from embryo to adult. Briefly, females lay their eggs in the food, and after approximately 22h the eggs will hatch (Markow et al., 2009) (**Figure 1**). Next, *Drosophila* goes through 3 larval stages, called instars, during which the animal mainly stays in the food. The 1st and 2nd instar stages take approximately 24h each, while the 3rd instar stage takes at least 48h (Bakker, 1959) (**Figure 1**). At this point, the larvae stop feeding and start to wander until they pupariate. The pupal stages, when metamorphosis occurs, span through approximately 4 days, at the end of which, the eclosion of the adult animal takes place (Bainbridge and Bownes, 1981) (**Figure 1**). During the adult stages, females typically mate with several males. Consequently, a female may produce approximately 3000 eggs during its lifetime, being able to lay up to around 100 eggs per day while at fecundity peak, which happens within the first 3-4 days of adult life (Klepsatel et al., 2013; Shapiro, 1932). Once the eggs are laid, *Drosophila*'s life cycle restarts (**Figure 1**).

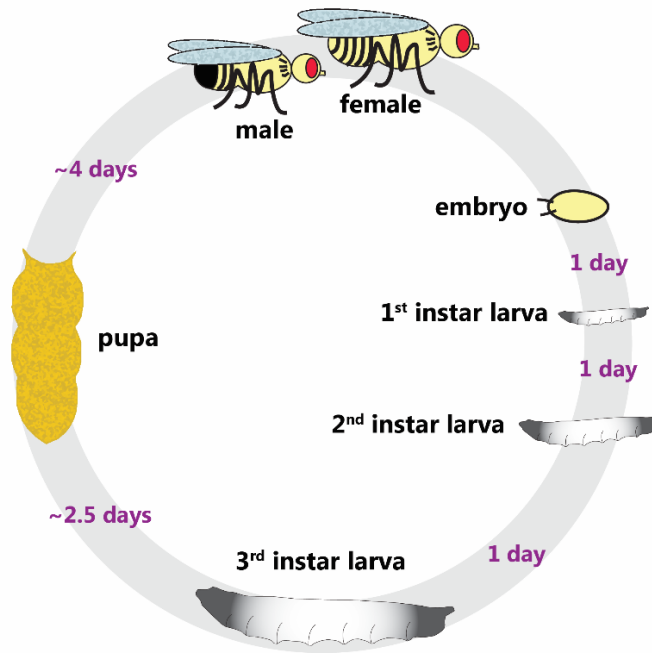


Figure 1. Life cycle of *Drosophila*.

When maintained at 25°C, *Drosophila* life cycle is completed in approximately 10 days. Adults mate and females lay eggs, where the embryo develops for approximately 1 day. The embryo hatches and spends approximately 5 days as larva (1 day as 1st instar larva, 1 day as 2nd instar larva and ~2.5 days as 3rd instar larva). Afterwards the larva pupate and goes through metamorphosis during approximately 4 days, after which the adult animal ecloses and the life-cycle re-starts.

3. Neuroblast generation in *Drosophila*

It is important to understand at a basic and functional level how brain development occurs and avoid possible defects.

Drosophila NSCs, the NBs, are a good model to study the mechanisms that regulate neural cell development and fate, precisely because these pathways are conserved, yet simplified when compared to mammals. Moreover, NBs form invariable lineages that can be genetically tracked, making it easier to identify defects in them.

Drosophila has different types of NBs, with characteristic lineages, that populate different parts of the brain, which will be described in more detail in section “5. NB Lineages”.

Each NB has a specific identity attributed by the expression of transcription factors (TFs) in a temporal and spatial dependent manner. This means that, not only NBs in different areas of the brain may have different identities, but also, that the “same” NB can express different TFs throughout time and consequently go through different temporal identities. Ultimately, such mechanisms will allow the small number of NBs in the *Drosophila* brain to generate thousands of neurons and glia cells. Spatial and temporal patterning will be

described in more detail in sections “**6. Spatial patterning in *Drosophila***” and “**7. Temporal Transcription factors – temporal identity**”.

NBs are first generated during embryogenesis at the ectodermal layer of the germ band, in what can be called the first wave of neurogenesis. Cells from the ectodermal layer start by acquiring a neural fate and consequently differentiate from their neighbor cells by lateral inhibition signals mediated by Notch (reviewed in (Artavanis-Tsakonas and Simpson, 1991)). At this point, NBs delaminate basally towards the mesodermal layer where they increase in diameter and become properly distinguishable from the ectoderm (Hartenstein and Campos-Ortega, 1984). Shortly after, NBs start dividing asymmetrically and originating ganglion mother cells (GMCs). In turn, each GMC will divide symmetrically in terms of size, to originate post-mitotic neurons or glia (Fredieu and Mahowald, 1989; Hartenstein et al., 1987).

Curiously, studies first performed in the grasshopper and later translated to flies found that the identity of a particular NB was determined by its position in the neuroepithelium, and that each NB is formed at a specific time (Doe and Goodman, 1985).

In terms of spatial organization, NBs can delaminate from different regions of the embryo. The embryonic *Drosophila* central nervous system (CNS) is composed by the brain and the ventral nerve cord (VNC), equivalent to the vertebrate spinal cord (**Figure 2A**). The brain is divided in 3 regions called neuromeres: protocerebrum, deutocerebrum and tritocerebrum (**Figure 2A**). The VNC is divided in several neuromeres or hemisegments: 3 subesophageal (S1-3), 3 thoracic (T1-3), 7 abdominal (A1-7) and 3 terminal neuromeres (A8-10) (Crews, 2019) (**Figure 2A**). Within each hemisegment, NBs are organized in rows and columns, which originated an intuitive naming system based on their position; for instance NB7-3, refers to NB in row 7 and column 3 (Broadus et al., 1995). The neuroectoderm of the VNC will give rise to thoracic and abdominal NBs, while the procephalic neuroectoderm will originate central brain NBs (Doe, 2017).

In the embryonic brain there are approximately 100 NBs that delaminate from the neuroectoderm on either side (Urbach and Technau, 2003a).

In the embryonic VNC, 30 NBs are generated within each abdominal and thoracic bilateral hemisegments that will originate lineages. A thorough clonal analysis, following each lineage, allowed to identify all neural cell types generated by these NBs in the embryo in each segment, amounting to a total of approximately 400 neural cells in abdominal hemisegments, and around 500 in thoracic hemisegments (Schmid et al., 1999).

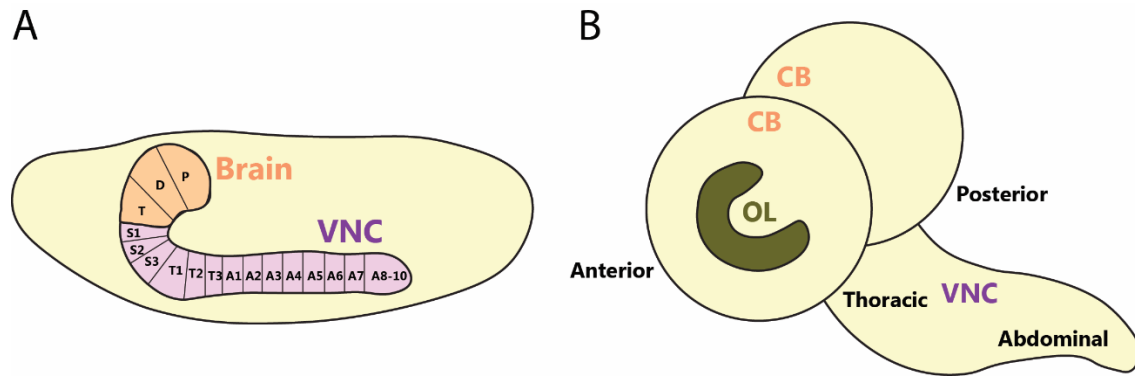


Figure 2. Regions of the *Drosophila* central nervous system in embryo and larvae.

(A) Embryonic central nervous system (CNS) is composed by the brain and ventral nerve cord (VNC); brain is divided in 3 regions called neuromeres: protocerebrum (P), deutocerebrum (D) and tritocerebrum (T); VNC is divided in the following neuromeres or hemisegments: subesophageal (S1-3), thoracic (T1-3), abdominal (A1-7) and terminal neuromeres (A8-10).

(B) Larval CNS is divided in central brain (CB), optic lobe (OL) and VNC; CB is populated by neuroblasts (NBs) in both the anterior and posterior sides; VNC is populated by NBs both in the thoracic and abdominal regions.

Thoracic NBs seem to maintain their number from embryo to the beginning of post-embryonic stages, *i.e.*, after larval hatching (Prokop and Technau, 1991), however, several abdominal NBs are lost by apoptosis still in embryonic stages (Abrams et al., 1993; Peterson et al., 2002; White et al., 1994).

Embryonic NBs enter quiescence at the end of embryogenesis and survive until larval stages where they initiate a second wave of neurogenesis. One of the mechanisms that is involved in NB entry in quiescence is the glycoprotein Anachronism (Ana), which is produced by glia (Ebens et al., 1993). At quiescence exit, Ana is repressed by *terrible reduced optic lobes (trol)*, which is also expressed in glia, and re-activates NB division via Fibroblast Growth Factor (FGF) and Hedgehog (Hh) signaling (Park et al., 2003).

Exit of quiescence occurs during larval stages and is triggered by the beginning of food intake and consequent activation of the insulin receptor (InR) and target of rapamycin (TOR) pathways (Chell and Brand, 2010; Sousa-Nunes et al., 2011). The insulin pathway acts as a sensor for nutrition, and when the larvae start to feed, it activates phosphatidylinositol 3-kinase (PI3K) and Akt. At the same time TOR can be activated either by PI3K/Akt or after detecting the presence of amino acids and ATP, increasing cell growth and proliferation (Russell et al., 2011). Essentially, when nutrients are available, TOR activation in the fat body leads to activation of PI3K and TOR pathway in glial cells through a signal termed fat body derived signal (FDS). In turn, glial cells release insulin-like peptides (ILPs) that will target quiescent NBs. ILPs will bind to InR in NBs, activating PI3K/Akt, and the circulating amino acids in the NBs will activate TOR. Together, InR and TOR pathways will inhibit forkhead box, sub-group O (Foxo) and

stimulate NB proliferation, consequently leading to their re-entry in cell cycle (Hietakangas and Cohen, 2009; Homem and Knoblich, 2012).

There are only a few NBs escaping this arrest in cell cycle at the end of embryogenesis and proliferate continuously while the remaining NBs are in a quiescent state. The first case are 4 NBs, called the mushroom body (MB) NBs, whose neurons project into the mushroom body neuropile (Ito and Hotta, 1992; Ito et al., 1997). Contrarily to other CB NBs, MB NBs divide independently of nutrient conditions and PI3K activity. Instead, MB NBs proliferation is regulated by the expression of *Eyless* (*Ey*), which bypasses the need for nutrient availability allowing proliferation even when nutrients are low (Sipe and Siegrist, 2017). The other case is the lateral lineage of the antennal lobe (AL) (Ito and Hotta, 1992).

In post-embryonic stages a second wave of neurogenesis commences (Prokop and Technau, 1991; Truman and Bate, 1988).

In the larval brain, NBs are distributed in 3 main regions, instead of 2 regions as in the embryonic CNS: the CB, the VNC and the optic lobe (OL) (**Figure 2B**). Until recently, the common view in the field was that NBs in the OL only originated in the larva brain, deriving from a group of 30-40 progenitor cells per brain hemisphere that are part of the newly hatched larvae (Egger et al., 2007; Hofbauer and Campos-Ortega, 1990). These progenitor cells are neuroepithelial cells that divide in a symmetric way in order to amplify their numbers. Then, a wave of neurogenesis transiently inhibits Notch in these progenitors and converts them into NBs, thus causing a shift to asymmetric division (Ceron et al., 2001; Egger et al., 2010). However, a more recent work described that the optic lobe epithelium produces NBs already during embryonic stages. These NBs enter quiescence and remain in the brain pos-embryonically, meaning that, the OL neuroepithelium starts producing neurons and glia more than 60h prior to what had been previously described (Hakes et al., 2018).

The OL is not part of the scope of this thesis and, for that reason, will not be thoroughly described, except for punctual mentions on specific matters.

The second wave of neurogenesis ends at pupal stages, with NBs disappearing, either because they die or terminally differentiate (Homem et al., 2014; Maurange et al., 2008; Truman and Bate, 1988; White and Kankel, 1978). However, the timings and mechanisms that lead to NB disappearance vary in different types of NBs, but this will be approached in more detail in section **“8. NB Lifespan”**.

Overall, *Drosophila*'s CNS suffers major remodeling throughout development. The brain of the adult fruit fly is composed of 10% of cells that were generated while the animal was an embryo and then remodeled during larval stages. The remaining 90% of the cells were generated during larval and pupal stages, making the studies of these developmental periods of most importance to understand brain development (Maurange and Gould, 2005; Truman and Bate, 1988).

This thesis will focus on NBs and their lineages on late larvae and early pupae stages.

4. Asymmetric cell division in NBs

One of the main features of NBs is their asymmetric mode of division, where the 2 daughter cells have different sizes and fates after mitosis. Thus, usually, 1 of the daughter cells retains the stem cell fate and the other inherits differentiation determinants and consequently becomes more differentiated (Horvitz and Herskowitz, 1992; Knoblich, 2008). Moreover, NBs, unlike other *Drosophila* stem cells such as germline, do not require an extrinsic signal to self-renew. Larval NBs are capable of dividing even without an external axis of orientation, and thus are able to divide even in culture (Ceron et al., 2006).

The division process is essentially composed of 4 steps: set up of polarity axis, orientation of the mitotic spindle, asymmetric localization of the cell fate determinants in the NB and their consequent segregation to the 2 daughter cells in a differential manner (Homem and Knoblich, 2012).

Embryonic NBs inherit their polarity from the neuroectodermal cells, from which they delaminate basally. NBs apically retain the Par complex, which is composed by Bazooka (Baz, also known as Par3), Par-6 and atypical protein kinase C (aPKC) (Yoshiura et al., 2012). In the next embryonic divisions, as well as in larval ones, NBs set up their polarity axis by aligning it relatively to the axis of the previous division (Rebollo et al., 2007, 2009). The process of NB division itself has been thoroughly reviewed (Knoblich, 2008). Briefly, Baz, establishes spindle orientation by recruiting Inscuteable (Insc), which in turn will recruit the Pins/Mus/Gai complex to the apical side. At this point, the mitotic spindle is oriented towards the apical-basal axis via Mud. Then, the Par complex manages the localization of cell fate determinants in the basal cortex through successive events of phosphorylation. These fate determinants are Numb, Prospero (Pros) and brain tumor (Brat), and upon division, they are segregated to the daughter GMC, promoting differentiation. Conversely, the other daughter cell will lack these cell fate determinants

and thus will self-renew and remain a stem cell. This mechanism is represented in **Figure 3**.

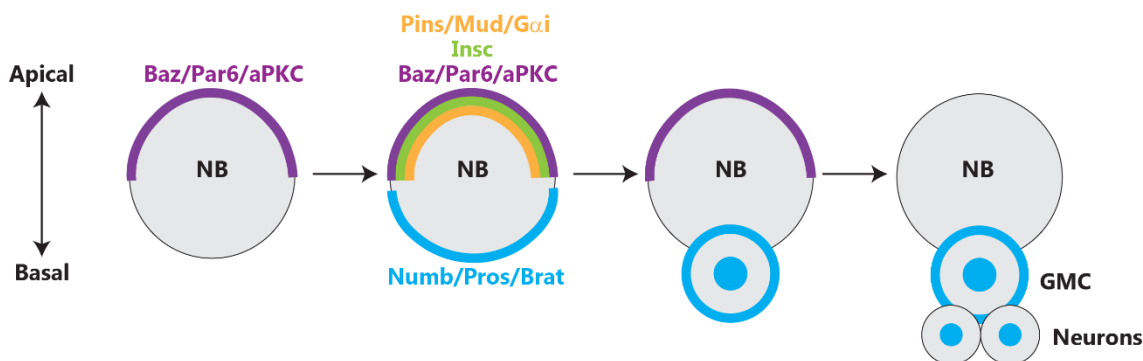


Figure 3. Asymmetric division of neuroblasts.

Neuroblasts (NBs) divide in an asymmetric way, self-renewing and originating a more differentiated daughter cell. The Par complex (Baz, Par6 and aPKC; purple) localizes on the apical side of the NB and recruits Insc (green), which in turn will recruit the Pins/Mud/Gai complex (orange). Mud orients the mitotic spindle and establishes the apical-basal axis. The Par complex directs the cell fate determinants Numb/Pros/Brat to localize basally by a series of phosphorylation events. Upon division, the cell fate determinants are segregated asymmetrically to the ganglion mother cell (GMC), leading to differentiation rather than self-renewal.

Overall, the process of asymmetric cell division is controlled intrinsically in the NBs and needs to be tightly regulated to avoid defects in proliferation, which will be addressed in section “**10. Defects caused by misregulation of proliferation and lineage progression**”.

5. NB Lineages

There are different types of NBs, characterized by their type of division, the specific lineage characteristics, as well as their position in the brain.

Type I NBs (Figure 4A) are the most numerous and are localized in all 3 main regions of the larval CNS. Type I NBs are identified by expressing a characteristic set of TFs, among which there is the basic helix-loop-helix proteins (bHLH) Deadpan (Dpn) and Asense (Ase), and the zinc finger protein Worniu (Wor) (Ashraf et al., 2004; Bier et al., 1992; Brand et al., 1993; Jarman et al., 1993). Type I NBs divide asymmetrically to self-renew and generate a smaller intermediate neural progenitor called GMC. The GMC has a limited mitotic potential and consequently divides once in a size symmetric way, to generate 2 neurons or glia cells. GMCs are characterized by expressing the cell fate determinant Pros in the nucleus, a homeodomain protein that is synthesized in the NB and polarized to the basal cell cortex and upon division is inherited by the GMC, favoring differentiation (Chu-Lagraff et al., 1991; Doe et al., 1991). The asymmetric distribution of

Pros in the NB is also regulated by the protein Miranda (Mira). After being inherited by the GMC, Mira is degraded, Pros is released from the cortex and consequently proceeds to the nucleus (Ikeshima-Kataoka et al., 1997; Schuldt et al., 1998; Shen et al., 1997). After the GMC divides, originating neurons or glia. Neurons are characterized by expressing the RNA binding protein Elav, a product of the *embryonic lethal abnormal visual system (elav)* gene; though Elav has been initially thought to be specific to neurons (Robinow and White, 1991), it has been proven later on to also be transiently expressed in NBs and glia (Berger et al., 2007). It has been described for some GMCs in CB to divide in a Notch-dependent way, originating 1 Notch^{ON} neuron and its Notch^{OFF} sibling, which attributes different fates to the neurons and consequently contributes to neuronal diversity (Bai et al., 2010).

On the other hand, glia cells express the *reversed polarity* gene (*repo*), which encodes a homeodomain protein (Xiong et al., 1994).

Another type of NSCs in *Drosophila's* CNS is **type II NBs (Figure 4B)**, which are localized in the posterior side of the CB, and there are 8 per brain lobe: 6 pairs located in the dorsomedial area of the CB, and 2 pairs located more laterally. Similarly to type I NBs, type II NBs also express Dpn and Wor, however, they do not express Ase; this characteristic, together with the fact that they are located on the posterior side of the CB led to them initially being known as posterior Ase-negative (PAN) NBs (Bowman et al., 2008). Another difference between lineages of type II NBs and the most common type I NBs is the fact that the former has a transient amplifying pool of intermediate progenitors. Unlike type I NBs, type II NBs do not express Pros and thus will not segregate it to its smaller daughter cell which will continue to express NB-specific markers. Opposite to GMCs that will undergo their final division, these cells retain a proliferative capacity, attributing them a profile of intermediate neural progenitors (INPs). In turn, the INPs will go through a maturation period, after which they will divide a limited number of times in a way that is asymmetric in terms of the molecular components attributed to each daughter cell. Thus, one cell will remain an INP and the other daughter cell will retain fate determinants such as Pros and Mira and become a GMC (Bayraktar et al., 2010; Bello et al., 2008; Boone and Doe, 2008; Izergina et al., 2009). The developmental profile of an INP is restricted by Earmuff (Erm), a fate regulator that activates the termination of proliferation in a Pros-dependent manner while preventing dedifferentiation via Notch (Weng et al., 2010). Once an INP generates a GMC, the lineage progression will resume as for type I, with the GMC dividing and giving rise to 2 neurons or glial cells. Another TF that allows a certain distinction between type I and type II lineages is Pointed (Pnt), a member of the E26 transformation-specific (Ets) family that is characterized by a winged

helix-turn-helix DNA binding domain (Ets domain). Pnt is only expressed in type II NBs, immature INPs (imINPs) and newly mature INPs (mINPs), thus making it specific to type II lineages (Zhu et al., 2011).

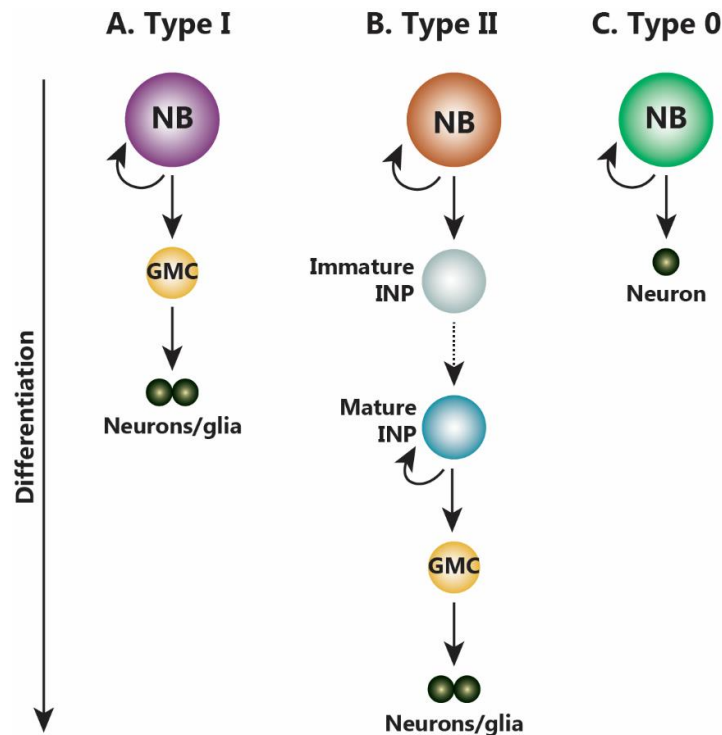


Figure 4. Types of asymmetric cell division in neuroblasts.

(A) Type I neuroblasts (NBs; purple) divide asymmetrically, self-renewing and originating a ganglion mother cell (GMC; yellow); GMCs divide once, to generate two neurons or glia (dark green).

(B) Type II NBs (brown) divide asymmetrically, self-renewing and originating an immature intermediate neural progenitor (INP, light blue), that after maturing (blue), divides asymmetrically to self-renew and to originate a GMC; the GMC divides once to generate two neurons or glia.

(C) Type 0 NBs (green) divide asymmetrically, self-renewing and directly originating a neuron.

Besides type I and type II, another type of division has been identified, and has been named as **type 0** (Figure 4C). However, so far, this mode of division has only been observed in NB lineages from the embryonic VNC and early OL (Baumgardt et al., 2014; Bertet et al., 2014). These lineages initially undergo a type I division progression, but at a certain point undergo a switch in their division mode. After this switch, NBs will generate neurons directly, without going through the intermediate GMC fate (Baumgardt et al., 2009; Karcavich and Doe, 2005). The switch from type I to type 0 division mode occurs upon activation of the cyclin-dependent kinase inhibitor Dacapo (Dap). In turn, Dap expression is activated by the temporal TFs *castor* (*cas*) and *grainyhead* (*grh*) and by the Hox gene *Antennapedia* (*Antp*). Ultimately, this cascade will allow NBs to generate neurons directly (Baumgardt et al., 2014).

Type I NB lineages within the CB and VNC give rise to 30-150 neurons each, that will populate the adult brain. However, the fact that type II lineages have an extra amplifying step (INPs) and that MB lineages have a longer lifespan, allows these lineages to contribute to the adult brain with a higher number of neural progeny, which can amount to approximately 200 neurons per lineage (Yu et al., 2013).

Several key regulators of NB proliferation and differentiation state progression within the lineages have already been identified. One example is the previously mentioned Dap, which is required for the type I to type 0 switch. Other examples are TFs such as Dpn and Wor that are associated specifically with NBs or intermediate progenitors with a higher proliferative potential. However, there is still a lot to be discovered regarding NB lineage progression and differentiation. Though we already know specific markers of each differentiation state within a NB lineage, that knowledge does not yet allow unequivocal identification of all states and could be still further improved. For instance, the INPs go through a maturation process during which different genes such as *Klumpfuss* (*Klu*) are turned on or off (Janssens et al., 2014), but several other genes are surely also at play.

Consequently, it is important to better understand the molecular players in NB lineage proliferation and differentiation, and how they are responsible for regulating fate changes throughout the lineage.

6. Spatial patterning in *Drosophila*

Spatial organization occurs at several levels in *Drosophila* to allow the formation of all the complex structures that will compose the adult fly. In this next section, I will briefly go through a few important hallmarks of spatial patterning of the general embryo, subsequently focusing on the NBs.

6.1. Embryonic body segmentation

Early in development, the embryo is organized in a series of repeated body segments. This segmentation is made possible by the sequential expression of a series of genes that will end up determining an antero-posterior axis. The 2 main types of genes that make this organization possible are the maternal-effect genes and the zygotic genes. Maternal-effect genes are expressed during oogenesis and basically give the embryo the initial organizational cues. Zygotic genes only start being expressed after fertilization

and will take on the cues provided by maternal genes to continue the antero-posterior organization. There are 2 classes of zygotic genes: segmentation genes, and homeotic selector genes. While the first is responsible for setting the correct number of segments, the latter regulate the pathways that allow each segment to acquire its identity. Segmentation genes can be further divided into gap, pair-rule, and segment polarity genes. The names of these 3 classes of genes are relatively self-explanatory and describe the phenotype obtained upon mutation of the genes within each class. Meaning that mutations in gap genes lead to gaps in the arrangement of segments; embryos with mutated pair-rule genes suffer a deletion of alternate segments; and deleting segment polarity segments leads to reversed polarity of patterns within segments, for instance (reviewed in (Gaul and Jäckle, 1990)). Although there are several genes within each of these classes, I will introduce only a selection that will be important for this thesis.

One of the most widely known maternal genes is *bicoid* (*bcd*). Bicoid protein sets an anterior-posterior gradient, where its expression represses *caudal* (*cad*; homeobox gene) messenger RNA (mRNA) in the anterior side of the embryo, thus leading to a Cad gradient (Frohnhofer and Nüsslein-Volhard, 1986; Macdonald and Struhl, 1986). Another important maternal gene is *hunchback* (*hb*), its mRNA starts by being homogeneously distributed throughout the embryo. However, its translation in the posterior side is later repressed by Nanos (Nos), which is also present due to maternal contribution (Irish et al., 1989). Consequently, a Hb gradient is generated and will contribute to the synchronization of body organization (Porcher et al., 2010).

Another class of genes that will be mentioned in this thesis is the one of gap genes. An example of a gene that is classified as such is *hb*, also previously mentioned due to its maternal contribution. Ultimately, both maternal and zygotic inputs will be important to establish the boundaries between the domain of expression of the different gap genes. More specifically, Hb will act through a gradient with different concentration thresholds that will repress other gap genes (Yu and Small, 2008). Within gap genes, *hb* is involved in the patterning of the trunk. In the trunk area of the embryo, there are 2 phases of gap gene regulation: an early mRNA regulation of gap genes led by feed-forward maternal gradients and a late regulation when gap proteins start to accumulate allowing gap-gap interactions. Other examples of trunk gap genes are *Krüppel* (*Kr*) and *knirps* (*kni*) (reviewed in (Jaeger, 2011)).

Some of the gap genes involved in body organization have also been described to have other roles at later stages in development, such as in neurogenesis. For instance, the head gap genes *tailless* (*tll*) and *orthodenticle* (*otd*, also known as *ocelliless*, *oc*). More

specifically, it has been described that head gap genes are necessary for the proper function of the proneural gene *lethal of scute* (*l'sc*); consequently loss-of-function of *tll* or *otd* lead to the loss of specific subsets of NSCs (Younossi-Hartenstein et al., 1997). Furthermore, *tll* has been described to be part of a temporal cascade in optic lobe NSCs, thus being responsible for the patterning of progenitor cells and consequent generation of diverse neural progeny (Li et al., 2013).

6.2. Spatial patterning in *Drosophila's* brain

When it comes to brain development, one NB lineage always develops in the same spatial location within the brain. This type of stereotypical behavior makes it easier to study brain development and organization.

A comprehensive work developed by the Technau lab almost 2 decades ago, thoroughly described the spatiotemporal pattern of generation of the approximately 100 NBs in each side of the embryonic brain. Thus, they accompanied early brain segmentation within the 3 main areas that subdivide the insect brain: tritocerebrum, deutocerebrum and protocerebrum (**Figures 2A, 5**). Such identification of independent neurons allowed the attribution of a systematic nomenclature based on the complete NB array, which is complete by embryonic stage 11, and that is aimed to portray the topological and functional characteristics of each NB. Briefly, this nomenclature starts by naming each NB depending on the area where they are located within the main groups: tritocerebrum, deutocerebrum and protocerebrum; with the latter being further subdivided into 3 more groups: anterior, central and posterior protocerebrum (Urbach et al., 2003). Furthermore, the expression of the dorsoventral patterning gene *ventral nervous system defective* (*vnd*) allowed to also divide the previous 5 groups into ventral and dorsal groups (Urbach and Technau, 2003b). The combination of all these positional cues then allowed to number NBs within each group towards a dorso-ventral axis, from anterior-to-posterior and ventral-to-dorsal (Urbach et al., 2003). Moreover, Technau lab further detailed the expression of over 30 molecular markers such as the head gap genes *hb*, *tll* and *empty spiracles* (*ems*) within the NB population. Several of these genes are specific to subsets of NBs and, ultimately, each of the 106 identified NBs presents a specific combination of these genes that is stereotypical and reflects their spatial position. The identification of these reproducible patterns of gene expression is important not only to identify each independent NB but will also help to understand phenotypes that might occur when those genes are being abnormally expressed (Urbach and Technau, 2003a).

Later studies by both the Lee and Ito labs focused on neuronal lineages in the adult brain. Using clonal analysis, they labeled individually the roughly 100 NB lineages (per lobe) at the beginning of larval stages and then analyzed the progeny in the adult. For this study they focused on the central part of the brain, equivalent to the CB, and excluded the OL; this region is called the cerebrum and it is estimated to have approximately 15000 neurons in each side of the midline. Though there was no reliable way to accurately follow NBs from embryo to their progeny in the adult, NB position and the fact that there is no neuronal migration in CB could be used to identify the different clones. The different clones had stereotypical characteristics regarding clonal size, neurite projection, and the specific area that they will innervate. The Lee and Ito studies combined were able to identify 114 different clonal lineages, a few more than the 106 lineages identified in the embryo by the Technau lab. Nonetheless, this inconsistency might be explained by the embryonic screen not including a few NB populations located at the most lateral part of the larval cerebrum. Other explanations might be that some embryonic NBs may originate more than 1 larval lineage or that the same NB might change fate and originate different clonal units (Ito et al., 2013; Yu et al., 2013).

Overall, a lot has already been achieved in describing the positional cues that characterize NBs throughout development. Though most neuronal clones in the adult seem to have been associated with the NB that originated them in earlier stages, the full characterization of those NBs is still not completed. Meaning that the complete molecular profile that NBs assume throughout time, depending on their position, is still not known.

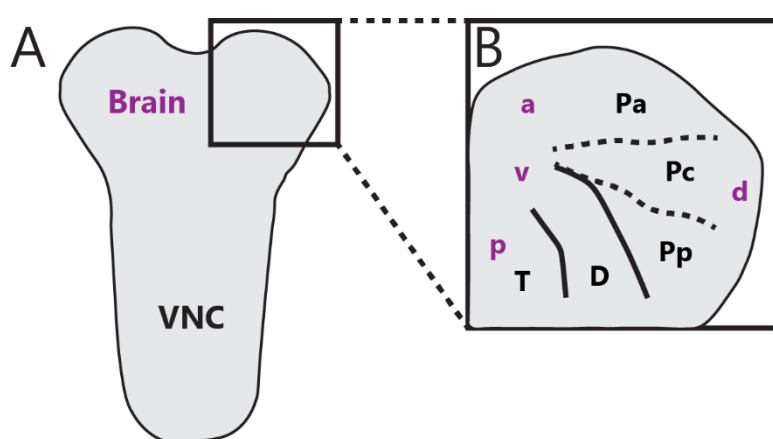


Figure 5. Regions of the embryonic *Drosophila* brain.

(A) schematic representation of the ventral view of a flat preparation, after the head had been dorsally opened.

(B) Focus on the right half of the head preparation showing the main regions of the embryonic *Drosophila* brain: there are 3 main regions separated with full lines: tritocerebrum (T), deutocerebrum (D) and protocerebrum; the protocerebrum is subdivided into 3 more groups: anterior (Pa), central (Pc) and posterior (Pp) protocerebrum. Each group is further divided in ventral (v) and dorsal (d); a, anterior; p, posterior.

7. Temporal transcription factors – temporal identity

When it comes to brain development, the importance of spatial regulation in generating diversity of neural cells has already been introduced. However, the vast number and variety of neural progeny are not due to spatial regulation alone, with another important aspect regulation occurring at the temporal level.

All NBs sequentially express a series of temporal TFs (TTFs). This temporal regulation allows NBs to present molecular and cellular diversity throughout development, ultimately allowing them to generate the neuronal diversity required in the adult brain. The first studies identifying TTFs were performed in embryonic VNC NBs, and that is still the best characterized system nowadays (Doe, 2017).

There are several levels of temporal regulation in NBs, which can differ among lineages in a more or less accentuated manner, to generate all the neuronal diversity of the adult *Drosophila* brain. I will include next an overview of some important regulation pathways, but a more thorough review of the windows of expression of TTFs in several NB lineages can be found in (Doe, 2017).

7.1. Embryonic TTFs

The first TTF that is expressed in most VNC NBs is Hb, an Ikaros-type zinc finger protein (**Figure 6**). The expression of this TF attributes an early-born identity to NBs generated during its window of expression; these NBs will generate GMCs and neurons that will also express Hb. Moreover, different levels of Hb expression within the daughter GMCs induce different neuronal fates, which further contributes to generate the enormous neuronal diversity that is required for proper brain development (Isshiki et al., 2001; Novotny et al., 2002; Pearson and Doe, 2003).

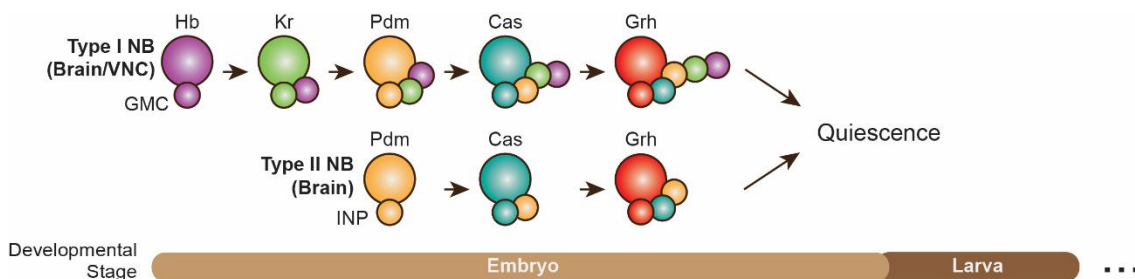


Figure 6. Neuroblast temporal identity during embryonic stages.

Neuroblasts (NBs) express a series of temporal transcription factors (TTFs). This series starts at embryonic stages and determines the temporal TF that is going to be expressed next. Most type I NBs from the brain and the ventral nerve cord (VNC) start by expressing Hunchback (Hb; purple), followed by, Kruppel (Kr, green), Pdm1/2 (Pdm; yellow), Castor (Cas, blue), and Grainyhead (Grh, red). In type II NBs (only in the brain), the series starts later in embryogenesis, skipping Hb and Kr, and starting directly with Pdm; the rest

of the embryonic series is identical to type I NBs. At the end of embryogenesis, NBs enter quiescence until larval stages.

The second TTF being expressed in most embryonic VNC lineages is Kr, a zinc-finger which expression increases after Hb is downregulated (**Figure 6**). Similar to Hb, Kr is both necessary and sufficient to establish the respective temporal identity fate, and is also expressed in the progeny that is generated within this temporal window (Isshiki et al., 2001; Tran and Doe, 2008). The switch from Hb to Kr temporal fate is controlled by the expression of the nuclear receptor Seven-up (Svp) (Kanai et al., 2005).

The third TTF that is expressed in most NB lineages is commonly referred as Pdm (**Figure 6**). This TTF refers to a pair of co-expressed POU-domain proteins with a redundant function, namely POU-domain protein 1 (Pdm1, also known as Nubbin, Nub) and POU-domain protein 2 (Pdm2). Contrarily to Hb and Kr, which are associated with early-born neuronal identities, Pdm attributes its neuronal progeny a late-born identity (Grosskortenhaus et al., 2006; Isshiki et al., 2001).

The next TTF in the temporal series that is also expressed in multiple lineages is Cas (encoded by the gene *cas* also known as *ming*), a zinc finger protein (**Figure 6**). Once Cas is activated it will inhibit the previous TTF Pdm. Moreover, like the previously mentioned TTFs, Cas is also expressed in the progeny of NBs generated within its window of expression (Brody and Odenwald, 2000; Cui and Doe, 1992; Isshiki et al., 2001; Mellerick et al., 1992).

The last TTF to be activated in multiple embryonic NB lineages is the bHLH protein Grh (**Figure 6**). It is activated in the embryo before NBs enter quiescence and it is maintained in larval NBs. However, despite Grh being expressed in post-embryonic NBs, it is no longer expressed in their progeny except for newly-born GMCs, meaning that it is never detected in post-mitotic neurons (Almeida and Bray, 2005; Brody and Odenwald, 2000; Cenci and Gould, 2005).

Moreover, the zinc-finger protein Squeeze (Sqz) and the transcription cofactor Nab have been described to further divide the Cas temporal window, though not directly regulating neither Cas nor the next TTF Grh. This led to the proposal of a new level of regulation referred as “subtemporal”. Thus, subtemporal genes are characterized by not regulating temporal genes, but acting downstream of them and consequently leading to further partition of the already known temporal windows (Baumgardt et al., 2009).

Interestingly, when analyzing the late embryonic distribution of Hb, Pdm and Cas, it was possible to detect that they were expressed in layers among the postmitotic neurons.

Thus, Hb is expressed in early-born neurons localized within deeper layers, while neurons in intermediate layers express Pdm and late-born neurons in more superficial layers express Cas (Kambadur et al., 1998).

Though, most type I NBs from CB and VNC follow the Hb > Kr > Pdm > Cas > Grh temporal cascade (Brody and Odenwald, 2000; Doe, 2017; Grosskortenhaus et al., 2005; Isshiki et al., 2001; Urban and Mettler, 2006), there are exceptions of later-born NBs that have been described to bypass the initial temporal fate Hb and even Kr, and starting the temporal series immediately in Kr or Pdm (Benito-Sipos et al., 2010; Tsuji et al., 2008). In type II NBs, this series starts later, directly with the expression of Pdm, meaning it skips Hb and Kr expression (**Figure 6**) (Walsh and Doe, 2017).

Ultimately, brain development relies on tight regulation between spatial and temporal cues. For instance, the misexpression of the TTF Hb leads to the generation of ectopic neurons with an early-born identity, yet, the number of neurons generated varies depending on the NB lineage (Isshiki et al., 2001). Since it is known that each NB has a specific position in the brain (Urbach and Technau, 2003a, 2003b; Urbach et al., 2003), this shows that this phenotypic variance is probably related to the location of each NB, proving that to generate the proper progeny at the right time, both spatial and temporal cues have a crucial role. Moreover, the combination of both spatial and temporal patterning helps to attribute independent characteristics to NBs and their lineages, which tell them apart from all others.

A more recent study further dissects the mechanisms by which temporal and spatial regulation might be integrated (Sen et al., 2019). By identifying the genomic targets of the TTF Hb in 2 different embryonic NBs (NB5-6 and NB7-4) during its competence window the authors determined that the targets were different. This means that the integration between spatial and temporal cues does not occur using an independent specification model in which spatial and temporal TFs bind to their targets in an independent way, with the TTF targets being the same in different cells, and ultimately allowing for combined spatial and temporal TFs to determine the identity of the NB. Conversely, the authors suggest a model where the specification of NB identity occurs sequentially, with spatial TFs regulating chromatin states that will determine Hb DNA-binding in a specific manner. This model was further supported by the fact that the spatial TF Gooseberry (Gsb), required to establish the identity of the NB5-6 and not of the NB7-4, was associated with open chromatin in the first NB, but not in the latter (Sen et al., 2019).

Another known case where integration of spatial and temporal cues occurs is in the OL. These NBs originate GMCs and then 2 neurons with different fates (Notch^{ON} and Notch^{OFF}). In turn, neurons with Notch^{ON} only take advantage of temporal cues, while Notch^{OFF} neurons use both temporal and spatial information, generating neurons specific to a certain region (Erclik et al., 2017).

7.2. Larval TTFs

Curiously, the most studied temporal regulators of NB identity in larval stages are RNA-binding proteins and not TFs. IGF-II mRNA-binding protein (Imp) and Syncrin (Syp) are responsible for creating opposing temporal gradients of expression in larval stages, encoding NSC age and consequently specifying several cell fates. These gradients have been well described in MB NBs as well as type II NBs, presenting similar characteristics in both. Imp is expressed in a high-to-low pattern, thus specifying early cell fates, while Syp is expressed in a low-to-high manner, endowing the cells with a late fate (Liu et al., 2015; Ren et al., 2017). For instance, in the MB, Imp is required to specify the early-born γ neuron identity, simultaneous expression of Imp and Syp, specify the intermediate identity of α'/β' neurons and finally, Syp expression determines the late-born α/β neuronal identity (Liu et al., 2015).

In thoracic NBs, at a later postembryonic phase, the TTF Cas stops being expressed at approximately 50h after larval hatching (ALH) and there is a burst of expression of Svp from approximately 40h-60h ALH, with timings varying among different NB lineages. Both Cas and Svp are responsible for regulating another temporal switch in the identity of neuronal progeny. This switch occurs between 2 BTB-zinc finger proteins: Chronologically inappropriate morphogenesis (Chinmo) and Broad Complex (Br-C). In thoracic NBs, the Chinmo/Br-C switch occurs at approximately 60h ALH, meaning that early-born neurons (deeper layers) express Chinmo, and late-born neurons (superficial layers) express Br-C (Maurange et al., 2008). Another interesting case in thoracic lineages occurs with the Hox gene *Antennapedia* (*Antp*): a temporal gradient of Antp generates adult motor neuron diversity, thus suggesting that Antp might be a target of NB TTFs (Baek et al., 2013).

Moreover, the levels of Chinmo expression have also been described to regulate the different fates in MB neurons, as *chinmo* translation is regulated by Imp and Syp (Liu et al., 2015).

In type II lineages TTFs responsible for temporal regulation were identified not only at the NB level but also in INPs. Regarding type II NBs, in a first phase that goes from 0h to approximately 60h ALH there is expression of the previously mentioned Imp and Chinmo, but also the RNA-binding protein Lin-28. This early phase is subdivided by other TTFs, as Dichaete (D), Cas and Svp. Afterwards, from approximately 60h to 120h ALH there is a late second phase that is characterized by the expression of Syp and the Ecdysone receptor B1 (EcRB1); this late phase is also divided into Br-C and the DNA-binding protein E93 (also known as Ecdysone-induced protein 93F, Eip93F). It has been described that the switch from Chinmo/Imp to Syn/Br-C/E93 is triggered by Svp by inducing the expression of the Ecdysone receptor. Such mechanism seems to also be in place for most lineages of CB type I NBs (Syed et al., 2017). Also, manipulations of the levels of Imp/Syp in type II NBs led to alteration in the number of cells presenting each neuronal fate. For example, by decreasing Syp, the number of early neuronal classes increase, but late neuronal classes decreased (Ren et al., 2017).

Regarding INPs, it has been determined that more than being an important asset in amplifying the lineages, INPs are also capable of generating neuronal diversity themselves. To achieve that, INPs, have their own cascade of sequential TTFs: D > Grh > Ey (Pax6 family), where D is expressed in the first half of the lineage, Ey is expressed on the last one and Grh is expressed in the middle, overlapping both D and Ey windows of expression. Within the temporal series, these TTFs show feedback repression between them, meaning that each TTF is repressed by the next one. Moreover, young INPs generate neurons expressing D or the TF Brain-specific homeobox (Bsh), while old INPs will either give rise to neurons that express the TF Twin of Eyeless (Toy) or to glia, which will express the glial TF Reverse polarity (Repo) (Bayraktar and Doe, 2013; Wang et al., 2014).

Despite all advances that have already been achieved in unravelling the TTFs that regulate neuronal diversity, there are still several players that remain unknown.

8. NB lifespan

NBs disappear before the adult stage, exiting mitotic proliferation during the first 2 days of pupal stages (White and Kankel, 1978). The timings and mechanisms by which NBs exit cell cycle vary depending on the subtype of NB, although, most NBs stop proliferating and disappear until 24h after puparium formation (APF) (Homem et al., 2014).

In the CB, type I NBs start decreasing in size during pupal stages and disappear in a terminal symmetric division (Homem et al., 2014; Maurange et al., 2008). Cell cycle exit in type I NBs is triggered by a burst of Pros translocates to the nucleus prior to NB disappearance (Maurange et al., 2008). Type I NBs located on the posterior side disappear around 16h APF, while type I NBs in the anterior side survive until a bit later, with their terminal division occurring around 24h APF (Homem et al., 2014). Type II NBs also decrease in size during pupal stages, becoming the same size as their daughter INPs at approximately, 16h APF, consequently disappearing (Homem et al., 2014).

Reduction of cell size in both type I and type II NBs, followed by terminal differentiation have been shown to be caused by the induction of oxidative phosphorylation and the mitochondria respiratory chain. The irreversible cascade that leads to the exit of cell cycle is induced by a pulse of ecdysone, a steroid hormone and by the Mediator complex, that cause a change in metabolism at the larval to pupal transition (Homem et al., 2014). The mediator multi subunit complex is composed of almost 30 subunits that are involved in binding TFs to RNA Polymerase II (PolII) (Conaway and Conaway, 2013; Malik and Roeder, 2010; Napoli et al., 2012). Ecdysone is produced by the prothoracic gland and its synthesis is tightly associated with nutrition, being an important coordinator of major metamorphic changes in *Drosophila* (Colombani et al., 2005; Handler, 1982; Koyama and Mirth, 2021).

MB NBs are the last subtype of NBs in CB to disappear: they are mitotically active until approximately 72h APF by when they start to decrease in size. At 78h APF there is a decrease in mitotic activity, and by approximately 96h APF MB NBs have already disappeared. The fact that MB NBs are present until roughly 10h prior to eclosion, makes them the last subtype of NBs to disappear. On the contrary to type I NBs, that have a similar division mode to MB NBs, MB NBs disappearance is not regulated by the presence of nuclear Pros. Instead, MB NB elimination occurs after downregulation of insulin/PI3K signaling and Foxo nuclear localization. Consequently, 2 mechanisms may be activated: autophagy and caspase-dependent cell death led by RHG-dependent apoptosis (process regulated by the proapoptotic proteins Reaper (Rpr), *Involution defective* (Hid) and Grim). These pathways work together in a way that both will compensate each other to assure MB NB elimination (Siegrist et al., 2010).

In the VNC, NB disappearance occurs at different times depending on their location. During embryonic stages there are 30 NBs per hemisegment, however, at larval stages, NB number has already decreased: in thoracic hemisegments there are roughly 23 NBs and in the abdomen that number decreases even more with the central abdominal

hemisegments presenting 3 NBs (Truman and Bate, 1988). Consequently, the biggest decrease in NB number occurs by programmed cell death in abdominal NBs still during late embryonic stages, and is regulated by the proapoptotic gene *rpr* (Abrams et al., 1993; Peterson et al., 2002; White et al., 1994). The remaining larval abdominal NBs divide for 22h to 40h and end up disappearing by apoptosis during the late larval instar stage due to a pulse of Abdominal-A (AbdA), a Hox/Homeotic protein (Bello et al., 2003). Interestingly, programmed cell death seems to have an important role in the developing nervous systems of other animals: in *Caenorhabditis elegans*, approximately 10% of cells disappear by programmed cell death, but this number increases to 50-90% in mammals, with *Drosophila* having an intermediate profile between the two. This prevalence seems to indicate that programmed cell death has an increasing role as the complexity of the animal also increases (Rogulja-Ortmann et al., 2007).

Thoracic NBs are mitotically active for approximately 4 days and undergo a decrease in cell size until they make a slower terminal division at approximately 120h ALH producing 2 daughter cells, similar in size, in a similar way than type I NBs in the CB (Bello et al., 2008; Maurange et al., 2008).

Regardless of the subtype of NB, their lifespan never exceeds pupal stages, and until relatively recent years proliferation in the adult brain of *Drosophila* had not been reported (Fernández-Hernández et al., 2013; Simões and Rhiner, 2017; Von Trotha et al., 2009). However, nowadays it is known that in the adult *Drosophila* brain there is a specific cell type that presents adult NB characteristics at least after mechanical injury, proving the potential for neurogenesis in the adult medulla cortex of the *Drosophila* OL. Furthermore, acute brain damage caused to this specific region further promotes neurogenesis. These findings were very promising in validating *Drosophila* as a model to study not only adult neurogenesis, but also regeneration and plasticity (Fernández-Hernández et al., 2013).

To this date, the exact mechanism that regulates the timing for NB to exit mitotic proliferation is not known. However, the fact that the mitotic proliferation does not end at the same time for all different NBs subtypes, seems to indicate that cell cycle exit is not simply occurring as an immediate response to a global mechanism.

9. Parallels to mammals

Brain structure has been studied for over a century, with the histological studies performed by Cajal representing a revolutionary hallmark on the study of neurons (Glickstein, 2006).

Despite the obvious striking differences between *Drosophila* and mammalian regarding size and morphology, there are several conserved aspects and pathways between them. Like in *Drosophila*, mammalian neuroepithelial cells also go through a period of symmetric divisions to allow cortex expansion. However, once neurogenesis starts, they become radial glia, a structure that then expands both apically and basally. Moreover, upon division, radial glia divides to self-renew and to generate a neuron or an intermediate progenitor cell (IPC). In turn, IPCs divide only once to generate a pair of neurons. In some mammals such as primates, the number of neurons is further amplified by another type of progenitor cell called the outer radial glia. These cells result from asymmetric divisions of radial glia, and in turn will also divide asymmetrically to generate IPCs. Overall, in terms of the neurogenesis, in “simpler” mammals such as rodents, it tends to follow a type I division mode, while in more “complex” mammals such as primates, neurogenesis reflects more a type II lineage (reviewed in (Homem et al., 2015)).

In mammals, the type and number of the generated neurons are also determined following temporal cues. Interestingly, several of TTFs identified in *Drosophila* have homologs in mammalian, which role is also associated with temporal regulation and even neurogenesis. Such similarity between *Drosophila* and mammals further strengthens the power of using *Drosophila* as an animal model.

One of the mammalian orthologs of the TTF Hb is Ikaros, a zinc finger protein expressed in retinal progenitor cells. Like Hb, Ikaros is more expressed earlier in cortical progenitors during neurogenesis, being sufficient to set earlier neuronal fates (Alsiö et al., 2013; Elliott et al., 2008).

The *Drosophila* TTF Cas also has a related TTF in vertebrates, named castor zinc finger 1 (Casz1). In the mouse genome, *Casz1* is located downstream of *Ikaros/Znfn1a1* (*Ikzf1*), with *Casz1* being expressed during mid-to-late stages in retina progenitor cells. *Casz1* represses the early-born fate and is sufficient to establish the mid-to-late temporal identity in murine retina cells, similar to what happens with Cas in *Drosophila* (Mattar et al., 2015).

Mammalian orthologs of the temporal regulator Imp, such as insulin-like growth factor 2 mRNA binding protein 1 (IGF2BP1), are frequently associated with brain glioma by promoting tumorigenesis (Liu et al., 2020; Xue et al., 2019; Zhan et al., 2021). Interestingly, despite *Drosophila*'s Imp being a larval temporal regulator, a recent study seems to indicate that in human cerebral cortex *IGF2BP1* has a higher expression during the prenatal period than the postnatal one (Shin et al., 2021). Though this seems to

indicate that IGFP2BP1 expression is mainly embryonic in humans, the high-to-low tendency in time is the same as in *Drosophila*.

Other TTFs, though not having a temporal role in fate determination that has been identified so far, seem to be related to cell proliferation. Next, are a few examples.

Kr also has orthologs in mammals, such as the zinc finger protein B-cell lymphoma 6 (BCL6). BCL6 is a transcription repressor and a proto-oncogene with a crucial role in lymphocyte differentiation, survival and function, for instance protecting B cells from undergoing apoptosis after suffering DNA damage (Jardin et al., 2007). Though BCL6 does not seem to have a role in fate regulation, it plays an important part in repressing proliferation in B-cell-lymphoma, where among other things, blocks B cell terminal differentiation, which ultimately makes it a promising therapeutic target (Cardenas et al., 2017).

The human orthologs of Nub (Pdm) also have described roles in cancer and neurodegenerative diseases. For instance, the TF POU Class 2 Homeobox 2 (POU2F2), like BCL6, has a role in regulating B cell proliferation. Consequently, POU2F2 can promote tumorigenesis by causing a metabolic shift to aerobic glycolysis (Yang et al., 2021). Moreover, POU2F2 has also been described as promoting proliferation and motility in other types of cancer, such as lung cancer (Luo et al., 2021). Another homologue of Nub, POU2F1 (POU Class 2 Homeobox 1) was also identified as being associated with a specific profile of late-onset Alzheimer's disease (Jiang et al., 2019).

One of the mammalian orthologs of Syp is the synaptotagmin binding cytoplasmic RNA interacting protein (SYNCRIP). In rat hippocampal neurons, SYNCRIP is part of the mRNA granules, mainly in cell bodies but also in dendrites, where they can be transported bi-directionally (Bannai et al., 2004). Moreover, recent studies showing that SYNCRIP targets have a strong component related to neuronal differentiation, are leading towards a new model where SYNCRIP post-transcriptionally controls the transition between neural progenitor cell to a differentiated neuron (Khudayberdiev et al., 2020).

The *Drosophila grh* also has 3 highly conserved orthologs, Grainyhead-like 1-3 (GRHL1-3). GRHL proteins are more commonly associated with epithelial biology and physiological conditions such as neural tube defects. Despite not having a role in determining cell temporal identity, GRHL factors have an important role in suppressing oncogenic cellular dedifferentiation, by chromatin binding and facilitation of epigenetic

alterations. Moreover, GRHL proteins further enable tumor conditions by increasing cell migration (Sundararajan et al., 2020)

Overall, there are several cases where *Drosophila* TTFs have orthologs with a described role in setting temporal identity, and it seems that at least to a certain point, there is conservation of the temporal identity cascade from *Drosophila* to mammals.

10. Defects caused by misregulation of proliferation and lineage progression

When progenitor cells fail to proliferate properly, neurodevelopmental defects may be originated, such as tumors and severe alterations in brain size, as microcephaly and megalencephaly.

It has been shown that loss-of-function of genes involved in the control of the fate of stem cells' progeny may lead to overproliferation and eventually to cancer. In *Drosophila* NBs, such phenotypes can be caused by mutations in *pins* (also known as *raps*), *numb*, *mira* and *pros*. Therefore, not only do the brains from larvae with these mutant genotypes increase drastically in size, but also, become invasive in other tissues and consequently kill the host. Furthermore, these immortal tumor cells could be transplanted to new hosts for years, with the host developing characteristics of malignant carcinomas after weeks (Caussinus and Gonzalez, 2005). Overall, this highly proliferative potential observed in *Drosophila* recapitulates what happens in mammalian tumors regarding carcinogenic potential and the capacity to invade other organs (metastasis).

Though, in a general way it is believed that cancer is generated based on defects in cell proliferation itself, tumors can also be caused by defects in differentiation (Harris, 2004).

For instance, *Osa*, a component of the chromatin-remodeling SWI/SNF (Switch/Sucrose non-fermenting) complex, ensures proper lineage progression consequently avoiding tumor formation. Loss of *Osa* leads to dedifferentiation of imINPs into NBs, resulting in transplantable brain tumors. Another key component in this process is the protein Hamlet (Ham), which is a target of the SWI/SNF complex in type II NBs and is directly induced by *Osa*. Both *Osa* and Ham have been shown to play a key role in limiting the number of divisions in transit-amplifying cells and consequently ensure proper temporal patterning (Eroglu et al., 2014). Interestingly, homologues of Ham in humans, such as Evi1, have been linked to tumorigenesis, including in humans (Fog et al., 2012; Wieser, 2007). This indicates that this mechanism might also be conserved in tumor formation in

humans, further strengthening the usefulness of *Drosophila* as an animal model for NSC proliferation.

Besides tumors, defects in the control of the proliferation of either NSCs or progenitor cells may also lead to brains with severely reduced or increased sizes. In mammals, such defects are termed microcephaly and megalencephaly, respectively.

Microcephaly occurs when there is a decrease in head circumference without severe effects on brain structure. Genes implicated in this defect are involved in centrosome biogenesis and maturation, as well as spindle regulation, which consequently affects the early symmetric divisions during the expansion of neuroepithelial cells (Gilmore and Walsh, 2013; Homem et al., 2015). On the other side, in megalencephaly, brains have an increased brain size and weight when compared to normal. Such phenotype can be caused by mutations in the tumor suppressor gene phosphatase and tensin homologue (PTEN). Studies in mice revealed that lack of PTEN led to increased cell proliferation and decreased cell death, resulting in enlarged and overall abnormal brains. Studies using neurosphere cultures have also shown that the higher proliferation of *Pten* NSC mutants was associated with shorter cell cycles (Groszer et al., 2001).

The proper regulation of NSC proliferation is critical for brain development, highlighting the importance of studying its pathways.

11. Neuronal generation and maturation

Neurons are generated during 2 main waves of neurogenesis. The first wave of neurogenesis occurs during embryonic stages and originates the larval CNS. The second wave of neurogenesis occurs mostly during larval stages and originates the cells that will populate the adult brains and therefore comprise the adult CNS (Ito and Hotta, 1992; Ito et al., 2013; Pereanu and Hartenstein, 2006; Truman and Bate, 1988; Yu et al., 2013).

In the fully developed adult brain, neurons generated from the same NB are always stereotypically organized and innervate the same area, thus representing one of the levels of organization in the brain. Areas with interactions between neurites that are denser than the adjacent areas are defined as neuropils and are used as a proxy to identify different brain circuits. To understand brain organization and function, it is important to study how NBs generate neuropils, and how they interact with each other at the circuitry level. On the contrary to what happens in the OL, in the CB neurons do not

migrate, therefore, the different clonal units that compose a specific neuropile are built regionally. If neurons only project within a single neuropil, they are designated as local interneurons, however, if neurons project through different neuropils, thus linking them, they are called projection neurons. One of the most studied neuropils in the *Drosophila* CB is the previously mentioned MB; other examples of neuropils are the AL, and the central complex (CX). These particular neuropils have clear boundaries and a specific and easily identifiable morphology, thus justifying why they are the most studied ones (Yu et al., 2013). Though finding boundaries in further divisions within neuropils might be relatively arbitrary, further studies have been made to understand the neuroarchitecture of *Drosophila*'s adult brain (Jenett et al., 2012; Rein et al., 2002; Wolff et al., 2015).

Each neuron itself must express a combination of effector molecules to recognize other neurons and form stereotyped connections during circuit assembly. A fully differentiated functional neuron requires the expression of features such as surface genes, ion channels, genes encoding neurotransmitter (NT) synthesis (Hobert, 2016).

Neuron terminal features are determined by several layers of TF regulation. As already described, at the level of the NSC, a series of spatial and temporal TFs ensure the formation of diverse neuron fates throughout development (Homem et al., 2015). A second class of TFs has been proposed to specify subtypes of neurons, with different combinations of genes resulting in different neuronal identities (Briscoe and Novitch, 2008; Landgraf and Thor, 2006). Lastly, a layer of transcriptional regulation is done by terminal selectors genes, which activate the expression of terminal differentiation genes that define and maintain neuronal attributes such as NT expression, neurite morphology and electrophysiological properties (Allan and Thor, 2015; Hobert, 2016).

Thus, neuron and brain function critically rely in the regulation of fate and terminal feature expression in a temporally accurate and coordinated manner. Though many programs that define cell-type specificity have been described, it remains unclear how these different sets of wiring programs are temporally controlled to ensure neuron maturation.

Terminal selectors have been identified and thoroughly studied in *C. elegans*, however, their identification in *Drosophila* is still in its preliminary steps. Nonetheless, some studies have already been made to identify possible regulators of terminal features in neurons, mostly in OL (Konstantinides et al., 2018; Kurmangaliyev et al., 2020).

Interestingly, it has been recently shown in *Drosophila* using single cell RNA-Seq (scRNA-Seq) that the transcriptomic information of neurons regarding their age is lost within 15h after their birth, resulting in a convergence of the transcriptomic profile in older,

more mature neurons (Özel et al., 2020). The study suggests that on the onset of synaptogenesis there might be a trigger, possibly ecdysone-related, that upregulates cell-type-specific cell-surface molecules, which in turn are necessary for synaptic specificity. Until this point, the upregulation of cell-surface molecules is accompanied by transcriptional diversity, which will be lost later upon transcriptomic convergence (Özel et al., 2020). Another scRNA-Seq study further supports this finding describing that the expression of pan-neuronal genes with synaptic functions or associated with synapse formation, such as adhesion molecules, are upregulated during mid-pupal stages. The authors further suggest the synchronous expression of these pan-neuronal genes might be regulated by a common regulatory mechanism acting at that developmental time, such as an ecdysone pulse, since members of that transcriptional cascade are expressed specifically at this same timepoint (Kurmangaliyev et al., 2020). Synaptogenesis is described to start around P45 in the OL (Özel et al., 2015, 2019) and after 60h APF in the CB (Muthukumar et al., 2014).

Ultimately, a neuron must go through a series of steps until it is fully mature, and that changes seem to be portrayed until a certain point in the transcriptional differences that are characteristic of the early hours of neuronal development. Therefore, in the future it will be interesting to study the mechanisms that regulate neuronal maturation and the specific timings that might be associated with it.

12. Transcriptomics at the single-cell level

Many findings regarding neurogenesis are only possible through analysis at the single cell level, using features such as transcriptomic information. Examples of such cases are the 2 studies described in the previous section, which identified transcriptional differences in young neurons that would disappear in older ones (Özel et al., 2020) and suggested that the onset of synaptogenesis could be associated with an ecdysone pulse (Kurmangaliyev et al., 2020; Özel et al., 2020). Other powerful analysis can be seen in a recent study of the transcriptome of the adult *Drosophila* brain throughout time, covering almost all cell types identified in a normal brain (Davie et al., 2018). Another interesting study, concerned the adult *Drosophila* ovary, leading to the discovery of new expression patterns throughout oogenesis (Jevitt et al., 2020).

Consequently, the analysis of transcriptomic data, mostly on a single cell level, poses itself as a valuable tool to study a specific cell type identity, the cells that make up a specific organ or tissue or even regulatory mechanisms, among others.

Any RNA sequencing (RNA-Seq) experiment usually follows a similar pipeline: cell labeling, cell sorting, library preparation, sequencing and bioinformatic analysis. Moreover, when working with limited amounts of RNA, library preparation also includes transcriptome amplification; several protocols have been developed and compared in order to optimize such crucial steps, resulting in efficient amplifications without significant bias (Faherty et al., 2015; Sun et al., 2013). Ultimately, such analysis will result in a snapshot description of the population being analyzed.

For several years, the method of choice to obtain such profiles was bulk RNA-Seq. However, these experiments have the disadvantage of representing the weighted average from the total population. Regardless of the organism in question, there is always cell-to-cell variation, even within the same organ or tissue. Even the “same” cell can change its characteristics throughout time, for instance as it differentiates or matures. Consequently, in more complex and heterogenous populations, the contribution of rarer cell types may be overlooked (Avila Cobos et al., 2018).

In recent years there have been several advances in techniques to study cells at a single cell level, and scRNA-Seq became a tool of choice. The first scRNA-Seq study based in next-generation sequencing was published in 2009, with the authors reporting the detection of more genes than previous sequencing methods (Tang et al., 2009). Ultimately, getting a more complete profile of the different cells in a certain organism might help to identify new cell types or differentiation states.

Of course, scRNA-Seq still has its limitations: for instance, the small amount of starting material may lead to the amplification of noise. Also, highly variable genes may overpower the counts, biasing the abundance relatively to other genes. Biological variation between different samples or batches, may also cause issues during analysis. Nonetheless, these issues can be mostly overcome, or at least decreased during bioinformatic analysis (Hwang et al., 2018).

During analysis, another important step is to perform good quality control to identify and preferably remove all cells with low quality, either because they are dead or dying or do not represent single cells, as any of the cases can result in compromised data interpretation. The removal of low quality cells is a valid concern among the scientific community, and efforts have been made to optimize data analysis (Ilicic et al., 2016).

There are several bioinformatic packages available for analyzing scRNA-Seq data that greatly improve data exploration. Such packages range from identifying sources of

heterogeneity in a single dataset to integrating several datasets (Satija et al., 2015; Stuart et al., 2019).

Ultimately, scRNA-Seq provides the transcriptomic profile of a cell at a specific point in time. Thus, to have a grasp of the temporal evolution of a specific tissue or cell type, would require the collection of several distinct timepoints. However, several tools have been proposed to overcome this limitation and allow us to look at datasets in a more fluid way. For instance, it is possible to compare the cells of a dataset and order them according to their pseudotime. This method takes advantage of individual cell asynchronous progression and uses gene expression changes to order cells along a certain trajectory such as lineage differentiation (Trapnell et al., 2014). Another interesting type of analysis is RNA velocity. By analysing the balance between nascent (unspliced) mRNAs and mature (spliced) mRNAs in single cells, this method estimates an RNA velocity vector for each cell. This allows predicting the future state of every cell in a scale of hours, making it compatible with the timescales of cellular fate changes and lineage development (La Manno et al., 2018).

Data from single cell Assay for Transposase-Accessible Chromatin Sequencing (scATAC-Seq) can also be integrated with other single-cell technologies, for instance with scATAC-Seq, that focus on chromatin accessibility and consequently, originate datasets with a more diverse and complete range of single cell information (Cusanovich et al., 2018; Hao et al., 2020).

Analysis based in scRNA-Seq acquired data is very useful to better understand different organisms and have already proven to be very efficient and insightful in *Drosophila*, as emphasized at the beginning of this section. Of course, any finding discovered at a bioinformatical level based on transcriptomic analysis would still require further validation *in vivo*.

13. *Drosophila* Toolkit

There are multiple genetic tools that make it possible to study neurogenesis in *Drosophila*. The work developed within the scope of this thesis required genetic manipulations such as restricting gene expression to specific cell populations and temporal windows, permanently label a cell lineage or even knockdown the expression of a gene. For that reason, I will now introduce a brief overview of the genetic tools I have used in my project.

13.1. Gal4/UAS

When studying a specific gene, it is useful to be able to control its expression, and one way that can be achieved is through the Gal4/Upstream activating sequence (UAS) system. This system was designed by Brand and Perrimon (Brand and Perrimon, 1993) and one of its main advantages is allowing the expression of a target gene, *i.e.* a gene of interest, in specific tissues or cell types in embryos, larvae and adults. Another advantage is that this is a two-part system, as it uses the target gene and its transcriptional activator in different fly lines, ensuring that they remain viable. Consequently, only after crossing both lines, the system gets activated. Another advantage of this system is that it generates lines that express a transcriptional activator instead of a single target gene. Consequently, it is possible to “drive” the expression of the target gene in specific cell types or tissues, simply by using a line that expresses the appropriate activator (Brand and Perrimon, 1993).

The Gal4/UAS system has several applications and has benefited from initiatives like the Janelia Farm Gal4 collection that curates thousands of transgenic lines which expression patterns are restricted to very specific and small subsets of cells in the adult brain (Pfeiffer et al., 2008).

The Gal4/UAS system allows cell or tissue-specific expression of transgenes. Additionally, it has even suffered specific modifications taking advantage of beneficial characteristics such as the dependence of Gal4 to temperature (Duffy, 2002).

During the work developed within the scope of this thesis, the Gal4/UAS system was immensely used, in combination with all the tools described next. This system allowed me for instance to downregulate the expression of a gene in a specific population (*e.g.* by using a driver for *ase*, I am only affecting type I NBs and not type II). It also allowed me to restrict the expression of a fluorescent protein during restricted temporal windows by benefiting of Gal4 dependence on temperature.

13.2. RNAi interference

In order to better understand gene function, it is common to knock down its expression. That can be achieved by introducing a double-stranded RNA (dsRNA) that will interfere with the function of the gene in question, a technique called RNAi interference (RNAi) (Kennerdell and Carthew, 1998; Misquitta and Paterson, 1999). The introduced dsRNA will target endogenous mRNA and consequently lead to its degradation before being translated (Montgomery et al., 1998). Moreover, to make this method more stable,

dsRNA is expressed as an extended hairpin-loop RNA that will interfere specifically with its target gene and consequently “silence” it (Kennerdell and Carthew, 2000).

This is a very useful approach that allows to better study the function of specific genes. Nowadays, there are available libraries that significantly cover the genes coding for proteins of interest. One of such examples are the transgenic RNAi lines available on the Vienna Drosophila Research Center (VDRC), which according to the most recent data in their website, cover 12,671 (91%) protein coding genes in *Drosophila* throughout their comprehensive collections (https://stockcenter.vdrc.at/control/library_rnai). Another example is the Transgenic RNAi Project (TRiP), that also presents a similar coverage of over 12,000 RNAi stocks in *Drosophila* (<https://fgr.hms.harvard.edu/trip-in-vivo-fly-rnai>).

Within the scope of this project, RNAi was used in combination with the Gal4/UAS system, mainly to study a gene function in NB lineage proliferation and lifespan. Briefly, UAS-RNAi lines of target genes were combined with Gal4 lines to induce the expression of the RNAi in specific cell types, usually either type I or type II NBs, thus allowing the study of the effect of a gene in a more particular context.

13.3. Temporal and regional gene expression targeting

At this point, I have already referred the advantages of regulating gene expression in a spatial manner, either by associating it with a fluorescent protein to better detect its pattern of expression or by silencing a gene in a specific cell type or tissue. However, often is also useful to control gene expression in a temporal manner.

Such purpose can be achieved using the temporal and regional gene expression targeting (TARGET) system. The TARGET system takes advantage of the multiple Gal4 and UAS transgenic *Drosophila* lines that are already available and combines it with a temperature sensitive (ts) Gal80 protein. The authors that developed this method showed that when cloning a tubulin promoter-Gal80ts construct into *Drosophila*, that construct was in fact able to regulate Gal4 in a temporally sensitive manner throughout different stages of development. In a conventional Gal4/UAS system, the Gal4 can be used to bind to its UAS counterpart to activate the transcription of a certain effector gene located downstream. When using the TARGET system, the authors have shown that the ubiquitous tubulin-Gal80ts was expressed at 19°C and repressed the expression of the UAS transgene. However, at 30°C, the tubulin-Gal80ts transgene became inactive, thus allowing the Gal4 to “drive” the expression of the UAS transgene associated with the gene of interest (McGuire et al., 2003, 2004).

Ultimately, being able to “drive” the expression of a gene to a specific cell type or tissue, and then further restricting its window of expression to a specific time, is a very useful tool. One possible application of such system is rescue experiments; interestingly, the first study using the TARGET system features it as an efficient way to rescue a memory defect in *Drosophila*, both at a temporal and spatial level (McGuire et al., 2003). Another application of the TARGET system is to knockdown a certain gene with temporal and spatial sensitivity, to avoid lethality or simply to have a more focused look on its role. But ultimately, it is an efficient, yet relatively simple way of studying when and where a certain gene product is used.

The TARGET system was a crucial tool for a major part of the work developed within the scope of this thesis. It was used to temporally restrict GFP expression in type I and type II NBs in CB and VNC during specific temporal windows of larval and pupal stages.

13.4. Gal4 technique for real-time and clonal expression

When analyzing NB lineages, it is common to use NB-specific genes to drive the expression of a fluorescent protein, such as green fluorescent protein (GFP). Usually, the progeny will no longer express the fluorescent protein themselves, but they will inherit it from their progenitors, allowing to track the lineage. Nonetheless, this means that eventually, in older cells, the fluorescence may no longer be visible which would make it harder to appropriately follow the complete lineage.

A solution for that is the Gal4 technique for real-time and clonal expression (G-TRACE), which provides permanent temporal and spatial labeling, as it also allows to trace expression patterns within cell lineages at a genomic level. This purpose is achieved by combining the Gal4/UAS system with the FLP/FRT (flippase recombinase/flippase recombination target) system and also with fluorescent reporters (Evans et al., 2009).

The G-TRACE system is in part similar to the standard Gal4/UAS system, as it requires the expression of Gal4 to initiate the reporter expression in the lineage.

Also, like the standard Gal4/UAS system, G-TRACE allows for the determination of the real-time expression of whatever gene is associated with the Gal4. However, G-TRACE also allows for the progeny of these cells to continue to express the reporter, which does not happen in the standard system. To achieve this, Gal4 will activate a FLP recombinase that will be responsible for removing a FRT-flanked termination cassette. This “FLP-out” cassette is strategically inserted between a Ubiquitin-*p63E* (*Ubi-p63E*) promoter and an

enhanced green fluorescent protein (EGFP) open reading frame, meaning that when it is removed, EGFP expression is maintained permanently in the generated progeny, regardless of Gal4 expression on those cells. G-TRACE is also designed to include, if necessary, a red fluorescent protein (RFP) reporter that would portray real-time expression, and combined with the EGFP reporter can allow to distinguish real-time from lineage expression (Evans et al., 2009).

The fact that not only the cells that express the Gal4 but also their derived lineage can be genetically tracked also led to this type of assay being called “permanent lineage tracing”. Such characteristics make this a powerful tool that can be used to study alterations that occur within a lineage, for instance regarding differentiation or cell fate determination (Bayraktar and Doe, 2013).

Within the scope of this work, permanent labeling was used in combination with the temperature sensitive TARGET system, to more specifically label NB lineages with fluorescence. Thus, GFP expression was driven in NBs starting at specific timepoints, and then maintained throughout the lineage forever, allowing a more thorough tracking of older neurons.

OBJECTIVES

Objectives

The main goal of this project is to understand how NB proliferation, fate and lifespan are controlled. Moreover, I am also interested in the neural lineage differentiation process, as a way to understand, not only how neuronal identity and fate are determined, but also how the neuron itself matures. The study of these aims will contribute to the field of brain development, as only by understanding the mechanisms that control NSC proliferation and differentiation we will be able to avoid defects such as tumor formation or neurodevelopmental diseases.

In this project I will address the following more specific aims within each chapter:

- I. **Identify possible TF regulators of NB proliferation and lifespan.** This will help me to identify the TFs that might be regulating NB proliferation and ultimately the end of their mitotic phase during pupa stages, which leads to the end of neurogenesis.
- II. **Sequence the transcriptome of CB and VNC lineages generated within sequential temporal windows of larval and pupal stages to create a comprehensive scRNA-Seq atlas of neuronal lineages of the developing *Drosophila* brain.** These transcriptomic datasets will be a useful resource for a variety of studies aimed to better understand neural development, such as the identification of new TTFs by comparing data from different timepoints.
- III. **Focus on a particular timepoint and identify regulator TFs important for neural lineage progression. Study the initial stages of neuronal maturation.** This will help me to identify several TFs with a likely role in regulating each differentiation state within a neuronal lineage (e.g. NB, GMC, neuron, etc.). Moreover, it will also allow me to better understand the timings and phases required for neuronal maturation.

CHAPTER I

Identifying regulators of NB lifespan

Authors' contributions

Graça Marques and Doctor Catarina Homem designed the experiments. Graça Marques performed the genetic experiments, analyzed the data and wrote the chapter, under the supervision of Doctor Catarina Homem.

CHAPTER I: Identifying regulators of NB lifespan

I.1. Introduction

During larval stages, *Drosophila* goes through a second wave of neurogenesis, when NBs generate most of the neurons that will populate the adult brain. NB lineages are stereotypical in a spatial and temporal way, meaning that the same NBs will always originate the same type and number of neurons at a specific time (Ito and Hotta, 1992; Ito et al., 2013; Yu et al., 2013). Most larval CB NBs divide asymmetrically every 1.3h, self-renewing and generating a more differentiated cell, the GMC, that will divide once more giving rise to neurons or glia (Homem et al., 2013). Consequently, throughout larval stages, each NB will divide dozens of times, always re-growing to its original size (Homem et al., 2014).

During their lifespan, NBs express a series of TTFs that determine their temporal fate; these fates are also inherited by their progeny, thus determining the fate of the neuron that is generated, or even determining the generation of glial cells (Doe, 2017). Ultimately, this temporal regulation plays a pivotal role in creating the diversity of neuronal cells in the adult brain.

Type I and type II NBs in the CB start decreasing in size during pupal stages and eventually go through a terminal symmetric division and terminally differentiate (Maurange et al., 2008). Consequently, at 24h APF, most NBs in the CB have already disappeared (except for MB NBs) (Homem et al., 2014). However, the temporal TFs that end NB proliferation and lead to their disappearance are not yet entirely known.

An important effort to study NB shrinkage and consequent disappearance was performed with a genome-wide RNAi screen (Homem et al., 2014). In this screen, a transgene encoding firefly luciferase and RNAi lines from the VDRC KK collection were expressed together specifically in type II NBs. Then, luciferase activity was measured in 1-day-old adults and compared to the wild type (WT) control. Though these transgenes are no longer expressed in the adult, luciferase signal can still be detected, thus allowing this comparison. Overall, 9,700 RNAi lines were tested, resulting in a coverage for approximately 70% of the *Drosophila* coding genome. In cases where the gene had no impact in NB lifespan, no change of luciferase activity was expected; eventually, a decrease in luciferase activity was also expectable, as RNAi can cause NB cell death or mitotic arrest. Conversely, an increase in luciferase signal indicates that in some way the tested gene might be influencing NB lifespan regulation. Although most of the tested RNAi lines revealed no alteration or decrease in luciferase signal, there was a minority

of genes that showed an increase of signal. For instance, there was a subset of 388 genes that showed at least a 4-fold increase of luciferase signal towards the control, indicating a strong effect in NB lifespan (Homem et al., 2014).

Moreover, the previous study also generated bulk transcriptomic data for NBs from wandering larvae and 2h APF (Homem et al., 2014). The comparison of how a certain gene was expressed in NBs in larvae vs. pupae would indicate if its presence seemed to be particularly relevant in one of those stages specifically.

Consequently, by analyzing the previously described datasets I was able to select a set of 11 candidates that either showed an increased NB lifespan (upon RNAi of a specific gene) or a decrease/increase expression from larva to pupae. This selection was also guided by literature research on the functions associated with the candidate genes.

Hence, the selected candidates were: Stat92E (Signal-transducer and activator of transcription protein at 92E), tap (target of Poxn), TFIIFa (Transcription factor IIFa), Hey (Hairy/E(spl)-related with YRPW motif), mid (midline), IM14 (Immune induced molecule 14, also known as Dso2 (Daisho2)), nerfin-1 (nervous fingers-1), Sirt2 (Sirtuin 2), Hsp67Bc (Heat shock gene 67Bc), Notch, Stip1 (Stress induced phosphoprotein 1, also known as HOP (Hsp70/Hsp90 organization protein)).

To better access a possible effect in regulating NB lifespan, all candidates were independently knocked down specifically in NBs. Next, 24h pupa brains were analyzed for the presence of NBs which at this point should no longer be present. Consequently, if upon analysis at 24h APF there were still NBs present, this would indicate that the candidate in question was likely to have a role in regulating NB lifespan.

On a second approach, I started from the interesting observation that when comparing the genes that regulate temporal series with the spatial TFs regulating body segmentation, there are overlapping TFs, such as Hb and Kr. This led to the hypothesis that maybe temporal and spatial cascades are regulated by the same genes and that these pathways could be re-purposed.

Supporting this idea, tll and otd, which are considered to be head gap genes due to their important role in the establishment of that segment of the animal, were described as being expressed in the procephalic neuroectoderm, as well as in subsets of brain NBs, thus having a crucial role in patterning the cephalic ectoderm and therefore the brain (Urbach and Technau, 2004; Younossi-Hartenstein et al., 1997). Moreover, tll has already been described as having a role in controlling a switch between NB stages in the optic lobe (Apitz and Salecker, 2015; Li et al., 2013). For these reasons I have decided

to further include the following segmentation genes in my candidate gene list for possible regulators of NB proliferation: Kr, tll, otd, kni, cad, bcd.

In this case, I started by testing reporter lines for these candidates to prove if they were being expressed during pos-embryonic stages. Only when that was the case, I proceeded to express RNAi of the candidates in NBs specifically, in order to study their impact in NB lifespan.

Among all tested candidates TFIIF α showed the most promising preliminary results in regulating NBs proliferation and ultimately their lifespan, as knocking it down specifically in NBs led to abnormal lineages in larvae and to longer-lived NBs in pupae.

I.2. Materials and Methods

Fly strains

w¹¹¹⁸ was used as wild-type (WT) for immunofluorescence experiments. The following Gal4-driver lines were used: UAS-*dicer2*;ase-Gal4UAS-CD8::*GFP*; (Type I NBs) (Harzer et al., 2014), UAS-*dicer2*;wor-Gal4ase-Gal80;UAS-CD8::*GFP* (Type II NBs) (Homem et al., 2014), *Repo*-Gal4 (glia driver; Bloomington line #7415), *GMR55B12*-Gal4 (glia driver – a gift from S. Schirmeier) and *R9D11*-Gal4 (imINPs driver – a gift from J. Knoblich)

The following RNAi lines were obtained from the BDSC: UAS Stat92E RNAi (#33637), UAS tap RNAi (#34985), UAS TFIIF α RNAi (#35428, #57412), UAS Hey RNAi (#31898), UAS mid RNAi (#50681), UAS IM14 RNAi (#57048), UAS nerfin-1 RNAi (#55395), UAS Sirt2 RNAi (#32482), UAS Hsp67Bc RNAi (#35452), UAS N RNAi (#33611, #33616), UAS Stip1 RNAi (#32979 #34002), UAS tll RNAi (#34329) and UAS otd RNAi (34327). The following RNAi lines were obtained from VDRC: UAS TFIIF α RNAi (#110225), UAS TFIIF β RNAi (#110569), UAS TFIIB RNAi (#106688), UAS TFIIS RNAi (#103718). GFP-tagged lines of segmentation genes were obtained from the BDSC: Kr (#56152), tll (#30874), otd (#66393), kni (#66588), cad (#30875) and bcd (#29018, #30869 and #67654). The Stat92E^{Y711F} and Stat92E^{R442P} dominant negative lines were a gift from E. Bach (Ekas et al., 2010). The JAK/STAT (Janus kinase/Signal transducer and activator of transcription) pathway *in vivo* GFP reporter (Bach et al., 2007) was a gift from J. Knoblich.

Fly crosses were set up at 29°C to increase UAS-Gal4 expression and fluorescence intensity. White pupae were collected and aged for 24h at 29°C. GFP-tagged lines were developed at 25°C.

Antibodies and Immunohistochemistry

The following primary antibodies were used: rabbit anti-Mira (1:200) and guinea pig-anti Ase (1:100) were both a gift from J. Knoblich, mouse anti-Prospero (1:100) and mouse anti-ELAV (1:100) were purchased at Developmental Studies Hybridoma Bank (DSHB), and rabbit anti-phosphor-histone H3 (PH3; 1:1000) was purchased from Millipore. Fluorescent secondary antibodies (Molecular Probes) were used at 1:1000.

Larval and 24h APF brains were dissected and fixed in 4% paraformaldehyde for 20min at room temperature; afterwards, the brains were washed 3 times with phosphate buffered saline (PBS) with 0.1% Triton X-100. Fixed brains were blocked for 20min in PBS with 0.1% Triton X-100 and 1% Normal Goat Serum (blocking solution) and incubated with the primary antibodies over-night at 4°C. Next day, brains were washed 3 times, blocked 20min and incubated with the secondary antibodies for 2h at room-temperature. Afterwards, brains were washed 3 times, incubated 10min in PBS, mounted in aqua/polymount (Polysciences, Inc) and imaged.

Immunofluorescent images were acquired either on a LSM710 or LSM880 microscope (Carl Zeiss GmbH). Adobe Photoshop was used to adjust brightness and contrast. Quantification of NB number and size was performed manually using the imaging software Fiji (ImageJ); data were analyzed using GraphPad Prism 6 (GraphPad Software, Inc); statistical differences were analyzed using Student's *t*-test, considering a *P*-value<0.05 as statistically significant.

I.3. Results

I.3.1. Selecting new candidate regulators of neuroblast proliferation and lifespan

During pupal stages, most CB NBs start decreasing in size and eventually terminally differentiate (Homem et al., 2014). However, the mechanisms that lead to the end of NB proliferation and subsequent disappearance is not yet fully understood.

For that reason, I aimed at identifying possible regulators of NB lifespan. To do that I started by independently knocking down a group of pre-selected candidates, specifically in NBs. Next, I analyzed the brain of the animals 24h APF and checked whether NBs were still present (**Figure I.1**). If that was the case, it would be an indication that the candidate gene was likely to have a role in regulating NB lifespan.

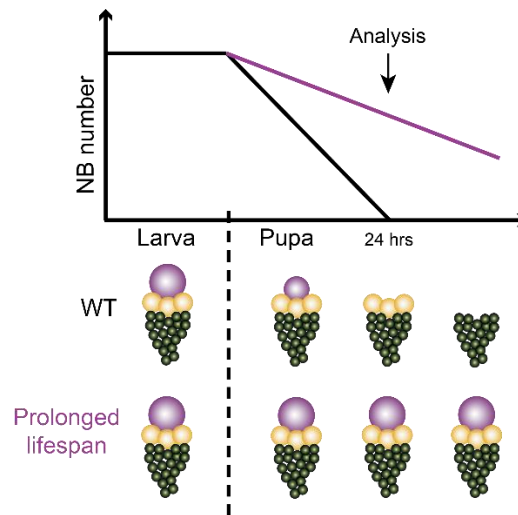


Figure I. 1. Experimental design to access candidate impact in neuroblast lifespan regulation.

Each candidate gene was knocked down independently in neuroblasts (NBs) and brains were analyzed 24h APF, a timepoint where usually NBs are no longer present in the brain. If the candidate gene has a role in NB lifespan regulation, its knock down in NBs may result, for instance, in longer lived NBs that would still be present 24h APF.

To select the candidates to test I focused on the hits resulting from a previous genome-wide screen to identify possible regulators of NB lifespan (Homem et al., 2014). Briefly, in this screen luciferase and RNAi of different genes were expressed in type II NBs and afterwards luciferase activity was accessed in the adult animal. Levels of luciferase activity superior to the ones in WT animals indicated that the associated gene could have a role in regulating NB lifespan. Moreover, I also consulted 2 datasets of bulk transcriptomic datasets: one collected from NBs of wandering larvae and the other from NBs at 2h APF (Homem et al., 2014). By comparing the expression of any gene within these 2 datasets that would indicate if its expression varied between larvae and pupae (**Table I.1**). For instance, if a gene had a higher expression during larval stages that could indicate that it was important for NB regulation in those stages, rather than in pupal ones.

Based on the combination of these criteria I selected and tested a small group of 11 candidates: Stat92E, tap, TFIIIF α , Hey, mid, IM14, nerfin-1, Sirt2, Hsp67Bc, Notch and Stip1.

Table I.1. Summary of phenotypes obtained when knocking down the expression of candidate genes specifically in neuroblasts

Candidate	Fold change Larva to pupae	Line tested	Presence of longer-lived NBs	
			Type I NBs	Type II NBs
Stat92E	1.03	RNAi (BL33637)	Yes	No
		Dominant Negative (Stat92E ^{Y711F})	No	No
		Dominant negative (Stat92E ^{R442P})	No	No
		Dominant Active	No	No
tap	1.28	RNAi (BL34985)	No	No
TFIIIFα	0.91	RNAi (BL35428)	Yes	Yes
		RNAi (BL35412)	Yes	Yes
		RNAi (110225KK)	Yes	Yes
Hey	1.23	RNAi (BL31898)	ND	No
mid	2.03	RNAi (BL50681)	ND	No
IM14	0.28	RNAi (BL57048)	ND	No
nerfin-1	2.85	RNAi (101631KK)	No	No
		RNAi (BI55395)	ND	No
Sirt2	0.83	RNAi (103790-KK)	No	No
		RNAi (BL32482)	ND	No
Hsp67Bc	5.06	RNAi (103974KK)	ND	No
		RNAi (BL35452)	ND	No
Notch	1.87	RNAi (BL33611)	ND	No
		RNAi (BL33616)	ND	No
Stip1	0.93	RNAi (BL32979)	No	No
		RNAi (BL35002)	ND	No

ND – Not determined

To test this list of candidates I selected BDSC or VDRC RNAi lines according to their availability, knocked down each gene independently in NBs and checked the brain of these animals at 24h APF (**Figure I.1**). In order to facilitate analysis, I analyzed type I and type II NBs independently using different drivers: UAS-*dicer2*;ase-Gal4UAS-*CD8::GFP*; and UAS-*dicer2*;wor-Gal4ase-Gal80;UAS-*CD8::GFP*, respectively. At 24h APF most NBs have already disappeared in a WT situation, therefore, if the knock down of any candidate resulted in more NBs than expected in a WT brain, that phenotype of longer-lived NBs was considered to indicate that gene had a role in regulating NB lifespan. The results of this analysis are summarized in **Table I.1**. In many cases, and for the same genotype, the results were variable with occasional longer-lived NBs, enhancing the need to test other independent RNAi lines. Though all candidates have been tested for the presence of longer-lived type II NBs, for the sake of time or for the

lack of promising preliminary results, some of them were not tested for a phenotype in type I NBs. Overall, knocking down most of these 11 candidates did not seem to have a significant effect in terms of longer-lived NBs. However, there were 2 genes that seem to have more NBs when they were knocked down when compared to WT brains 24h APF: *Sat92E* and *TFIIFα* (**Table I.1**). Though it had already been described that *Stat92E* is involved in stem cell fate (Amoyel and Bach, 2012), that was not the case for *TFIIFα*. Being a member of the general RNA PolII machinery, *TFIIFα* is associated with more general functions during transcription, such as recruiting RNA PolII to the initiation complex (Hahn, 2004), and to preventing transient pausing (Price et al., 1989; Rossignol et al., 1999). Since these 2 genes were the only ones that showed a consistent phenotype, I analyzed them in more detail.

I.3.2.Candidate 1: *Stat92E*

A previous genome-wide *in vivo* RNAi screen to identify NB proliferation regulators performed by our group indicated *Stat92E* as a candidate regulator of NB proliferation (Homem et al., 2014). It has already been described that *Stat92E* is involved in stem cell fate (Amoyel and Bach, 2012). Thus, to confirm these results, I knocked down *Stat92E* by RNAi specifically in NBs and compared the size and number of type I and type II NBs 24h APF, a timepoint where usually most NBs have already terminally differentiated.

In WT pupae, 24h APF, there are in average 9.2 ± 2.417 type I NBs per lobe (anterior side; **Figure I.2A,F**) with 6.156 ± 0.1992 nm (**Figure I.2E**), and 0 type II NBs (**Figure I.2G**). Knock down of *Stat92E* in type I NBs significantly increased the number of type I NBs to 16.80 ± 1.744 (**Figure I.2B,F**); however, the size of these type I NBs was approximately the same as WT NBs (6.429 ± 0.4796 nm; **Figure I.2E**). Knock down of *Stat92E* in type II NBs led to the presence of type II NBs at 24h in some pupae. Nevertheless, their number highly varied from 0 to 6 type II NBs per lobe (**Figure I.2H**), having a size of approximately 5.461 ± 1.102 nm. These data suggested that after knocking down *Stat92E*, the lifespan of type I and type II NBs increased.

To confirm if *Stat92E* indeed has a role in NB lifespan regulation, I directly interfered with *Stat92E* activity by expressing two different dominant negative variants of *Stat92E*. Both of these dominant negative variants resulted from amino acid substitutions: in *Stat92E*^{Y711F}, the critical tyrosine in position 711 was replaced by phenylalanine, and in *Stat92E*^{R442P} the arginine in position 442 was replaced by proline. These dominant

negatives resulted in an inability to form an activated Stat92E dimer and in an incapacity to maintain normal dimer-DNA interactions, respectively.

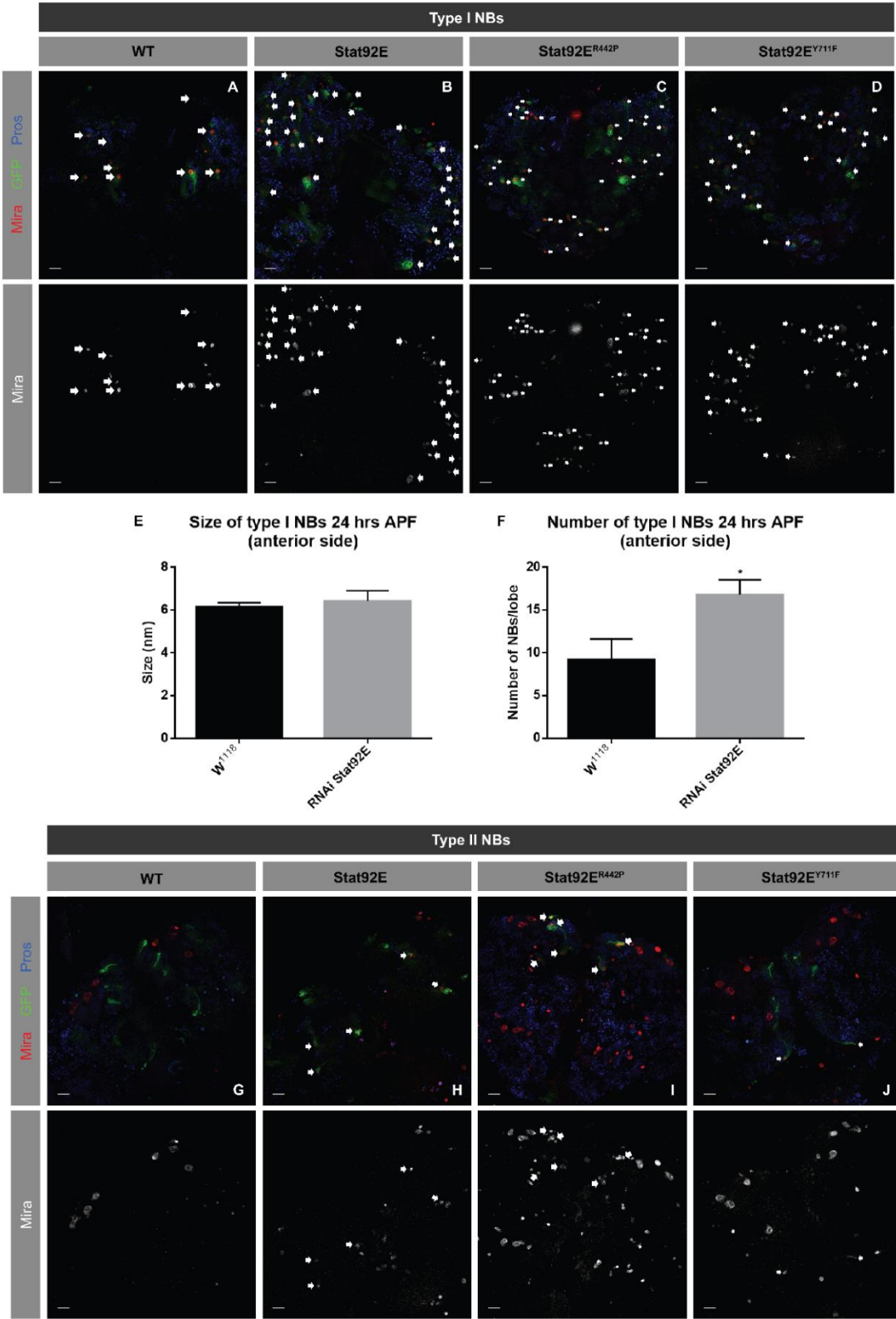


Figure I. 2. RNAi and dominant negative variants of Stat92E do not consistently increase type I and type II neuroblast lifespan.
(A-D, G-J) Fixed brains 24 hrs APF. Type I (A-D) and type II (G-J) neuroblast (NB) driver Gal4 lines driving the expression of the indicated RNAi or dominant negative variant. (A, G) WT, (B, H) RNAi of Stat92E, (C,

I) Stat92E^{R442P} dominant negative variant, (D, J) Stat92E^{Y711F} dominant negative variant. Mira, red; CD8::GFP, green; Pros, blue. Arrows indicate type I NBs (A-D) or type II NBs (G-J). (E, F). Scale bars represent 20 μ m.

(E,F) Quantification of type I NB diameter (E) and type I NB number (F) in the anterior side of brains from 24 hrs pupae of the WT and Stat92E RNAi. Error bars represent $\pm$ SEM.

In contrast to Stat92E RNAi, expression of Stat92E dominant negative caused no changes in the number of type I NBs (**Figure I.2C,D**). Expression of the dominant negative Stat92E in Stat92E^{R442P} increased the number of type II NBs to 2-4 per lobe (**Figure I.2I**). Similarly, Stat92E^{Y711F} dominant negative also presented longer-lived type II NBs (3-7 type II NBs per lobe) (**Figure I.2J**). The phenotypes caused by dominant negative variants of Stat92E are more variable than those caused by Stat92E RNAi. Consequently, this inconsistency does not allow me to make a robust conclusion regarding the effect of Stat92E in NB lifespan.

To understand what might explain the variability between the number of NBs that were still present 24h APF, I used 10XStat92E-GFP, an endogenous fluorescent reporter, to assess in which cells Stat92E was expressed. This analysis showed that in 3rd instar larval brains, only glial cells but not NBs expressed Stat92E (**Figure I.3**, arrows). Consequently, any effect that Stat92E might have in NB lifespan is likely occurring indirectly through glia. Despite being usually associated with an insulator role in the brain, glia has been shown to also be important for brain development through the regulation of NB proliferation (Kanai et al., 2018). However, it is odd that if Stat92E is only expressed in glia, that its knockdown in NBs also leads to a phenotype.

At this point I do not know how to explain that, eventually the phenotype obtained with the RNAi line could be explained by off targets while the inconsistent results obtained with the dominant negative variants may be dismissible. However, since NBs also have the potential to generate glia cells another hypothesis might be that, by knocking down Stat92E in precursor cells, somehow that will also affect glia, which in turn might misregulate NB proliferation potential.

To further exploit the possibility that Stat92E is a factor secreted by glia to regulate NB proliferation, I interfered with Stat92E pathway in glia. To do this, I crossed the Stat92E^{R442P} dominant negative with two different glial drivers, Repo and GMR55B12, and analyzed NB lifespan. Although in this case I did not separately analyzed type I and type II NBs, in the dominant negative pupae crossed with either of these glia drivers, at 24h APF there were a maximum of 2 NBs per lobe (posterior side) compared to 0 NBs in WT brains, which is not a very promising result. Hence, it seems unlikely that Stat92E is directly affecting NB lifespan. Due to the variability of the results, the observed

phenotypes where longer-lived NBs were still present could possibly be explained by off targets or eventually because Stat92E knock down might have an indirect role by impacting another member within its pathway. In order to get to a robust conclusion on this matter, further experiments would have to be performed. However, due to the unlikelihood of Stat92E to play a role in NB lifespan, I decided to move forward and test *TFIIFα*, the other candidate gene that showed a phenotype with longer-lived NBs.

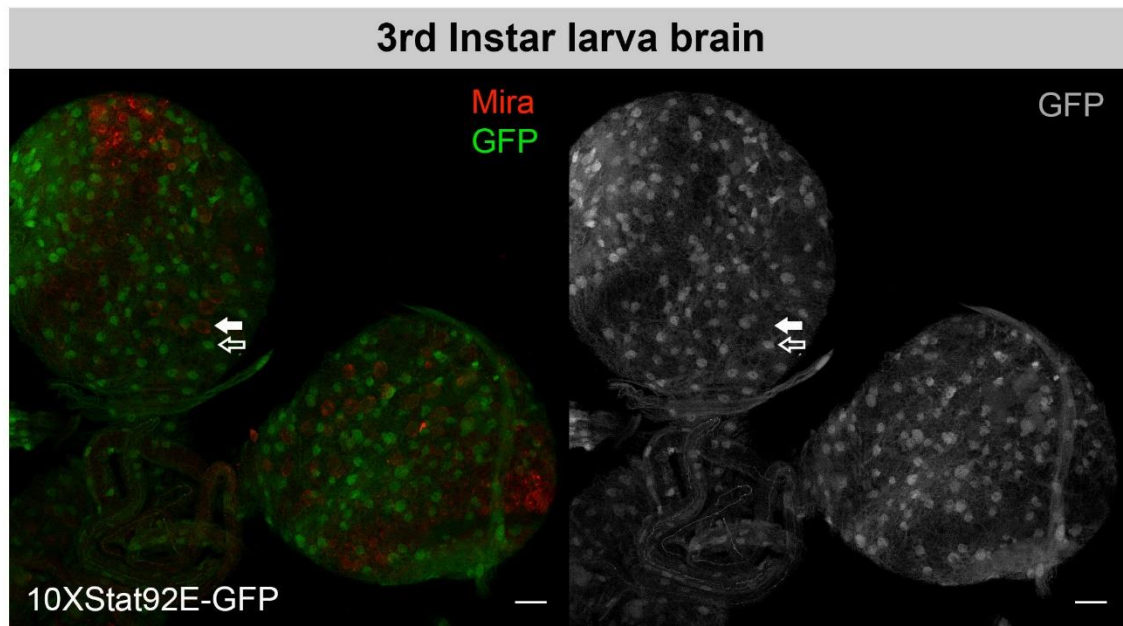


Figure I. 3. Stat92E is expressed in glia cells but not in NBs.

Fixed brain of 3rd instar larva. 10xStat92E-GFP reporter indicates Stat92E expression. Mira, red; Stat92E-GFP, green. Filled arrow indicates neuroblasts (NB), outlined arrow indicates glia. Scale bars represent 20μm.

I.3.3.Candidate 2: TFIIFα

I.3.3.1. Knock down of TFIIFα leads to longer-lived neuroblasts

From the initial list of 11 candidates, the second gene that showed a phenotype with longer-lived NBs was *TFIIFα* (**Table I.1**). *TFIIFα* is a member of the general RNA PolII transcriptional machinery with an important role in transcription initiation and elongation (Gong et al., 1993), such assisting RNA PolII recruitment to the pre-initiation complex (Hahn, 2004).

I started by knocking *TFIIFα* down by RNAi using BDSC stock #35428. In WT brains 24h APF, there were in average 5.400 ± 0.5099 type I NBs per lobe (anterior side; **Figure I.4A,C**) with an average size of $5.963 \pm 0.1148\mu\text{m}$ (**Figure I.4D**); at this time point, there were no longer any type II NBs present (**Figure I.5A,D**). Knock down of *TFIIFα* in type I NBs significantly increased the number of type I NBs to 43.40 ± 3.108 (**Figure I.4B,C**).

Additionally, their size, was also significantly increased to $7.927 \pm 0.35\mu\text{m}$ (**Figure I.4D**). Moreover, preliminary data showed that by using PH3, a mitotic marker, it was possible to determine that these type I NBs seemed to be still dividing, though in a lower ratio than the WT ones. Knock down of TFIIIF α in type II NBs resulted in the presence of an average of 5.250 ± 1.109 type II NBs per lobe (**Figure I.5B,D**), with an average size of $6.510 \pm 0.3679\mu\text{m}$ (**Figure I.5E**). These data suggests that, after knocking down TFIIIF α , the lifespan of type I and type II NBs is increased.

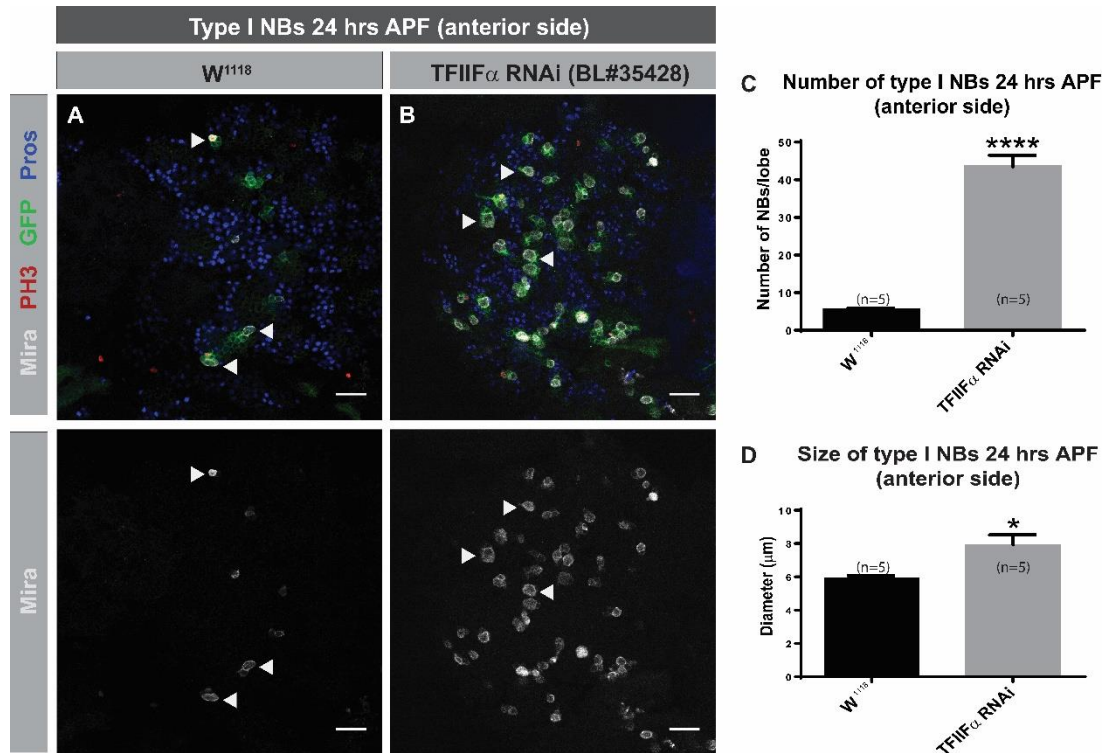


Figure I. 4. RNAi of TFIIIF α increases type I neuroblast lifespan.

(A,B) Fixed brains 24h APF. Type I neuroblast (NB) driver Gal4 line driving the expression of the indicated RNAi. (A) WT, (B) RNAi of TFIIIF α (BDSC stock #35428). Miranda (Mira; NB marker), white; PH3 (mitotic marker), red; CD8::GFP, green; Prospero (Pros; neuronal marker), blue. Arrowheads point to representative type I NBs. Scale bars represent 20 μm .

(C,D) Quantification of the (C) number and (D) diameter/size of type I NBs. Error bars represent $\pm\text{SEM}$; statistically significant differences were calculated using t-test (* $P \leq 0.05$, **** $P \leq 0.0001$).

To confirm if TFIIIF α indeed has a role in NB lifespan regulation, I used a different RNAi line (BDSC stock #35412). Knock down of TFIIIF α expression with this second line in type II NBs showed that, similarly to what happened with the previous RNAi line, there were still an average of 4.778 ± 0.7222 type II NBs per lobe (**Figure I.5C,D**), with $6.562 \pm 0.1399\mu\text{m}$ (**Figure I.5E**).

These consistent results led me to conclude that TFIIIF α indeed influences NB lifespan.

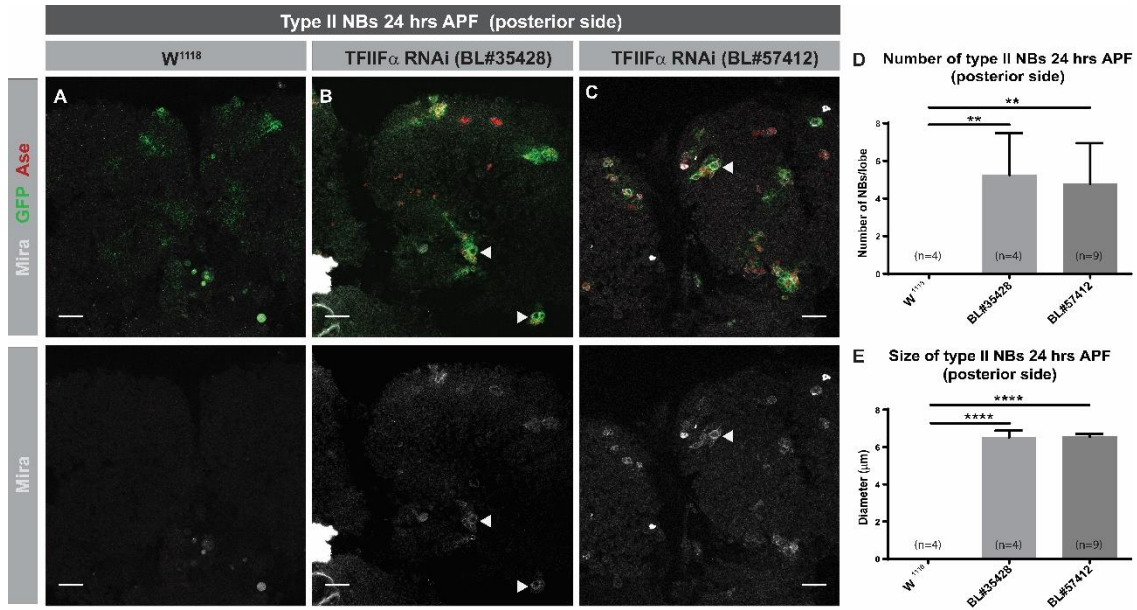


Figure I. 5. RNAi of TFIIIF α increases type II neuroblast lifespan.

(A-C) Fixed brains 24h APF. Type II neuroblast (NB) driver Gal4 line driving the expression of the indicated RNAi. (A) WT, (B) RNAi of TFIIIF α (BDSC stock #35428), (C) RNAi of TFIIIF α (BDSC stock #57412). Mira (NB marker), white; CD8::GFP, green; Ase (INP marker in type II lineages), red. Arrowheads point to representative type II NBs. Scale bars represent 20μm.

(D,E) Quantification of the (D) number and (E) diameter/size of type II NBs. Error bars represent $\pm$ SEM; statistically significant differences were calculated using *t*-test $**P \leq 0.01$, $****P \leq 0.0001$)

Since TFIIIF α is just one of the subunits of TFIIIF, the next step was to assess whether the down regulation of TFIIIF β (Transcription factor IIF β), the other subunit of this complex, also increased NB lifespan. This way, I could determine if NB lifespan regulation is a specific role of the α subunit, or if it is regulated by the complete TFIIIF complex. Moreover, testing other TFs that are part of the RNA PolII machinery would indicate if NB lifespan regulation is controlled only by the TFIIIF complex or if other TFs are also involved.

I.3.3.2. Knock down of other members of the RNA PolII transcriptional machinery increases neuroblast lifespan

To determine if NB lifespan regulation is a specific role of the α subunit, or if it is regulated by the TFIIIF complex as a whole, I also tested TFIIIF β (VDR line #110569), which is the other subunit of TFIIIF.

I also tested other members of the transcriptional machinery with more specific functions: TFIIIB (Transcription factor IIB; VDR line #106688), which is only present during transcription initiation, and TFIIIS (Transcription factor IIS; VDR line #103718), which is an elongation factor, thus is not present during initiation.

I also tested a third RNAi for TFIIF α (VDRC line #103718) to further confirm the phenotype.

As previously stated, in WT brains 24h APF, there are no longer type II NBs present (**Figure I.6A,F**). However, when I knocked down TFIIF α in type II NBs, similarly to what had happened with the 2 previous RNAi lines (**Figure I.5**) at 24h APF there were still 5 (out of 8) type II NBs present per lobe, with an average size of $6.573 \pm 0.2267\mu\text{m}$ (**Figure I.6 B,F,G**). Knocking down TFIIF β , the other subunit of TFIIF, also led to an increase of the number of type II NBs to an average of 6.833 ± 1.400 NBs per lobe, with an average of $7.601 \pm 0.1512\mu\text{m}$ 24h APF (**Figure I.6C,F,G**).

Similar to the knockdown of the 2 subunits of TFIIF, the knockdown of TFIIIB, led to the presence of approximately 3.500 ± 0.5000 type II NBs per lobe with an average size of $7.662 \pm 0.2587\mu\text{m}$ at 24h APF (**Figure I.6D,F,G**). Moreover, when I knocked down TFIIIS, there were still 3.923 ± 0.5713 type II NBs present per lobe with a diameter of $5.803 \pm 0.1232\mu\text{m}$ at 24h APF (**Figure I.6E-G**).

Preliminary data for the knock down of TFIIIB and TFIIIS in type I NBs also show a consistent increase in number and size when compared to WT (data not shown).

Overall, though this is only a small subset of the TFs that compose the RNA PolII machinery, data indicates that at least TFIIF α , TFIIF β , TFIIIB and TFIIIS seem to have a role in regulating NB lifespan.

In the future it would also be interesting to determine if the role of TFIIF (and the other tested TFII genes) is specific to NSCs or if it affects other highly proliferative cells. An interesting experiment could be knocking down TFIIF α in the wing disc, by using an *apterous* (*ap*) driver.

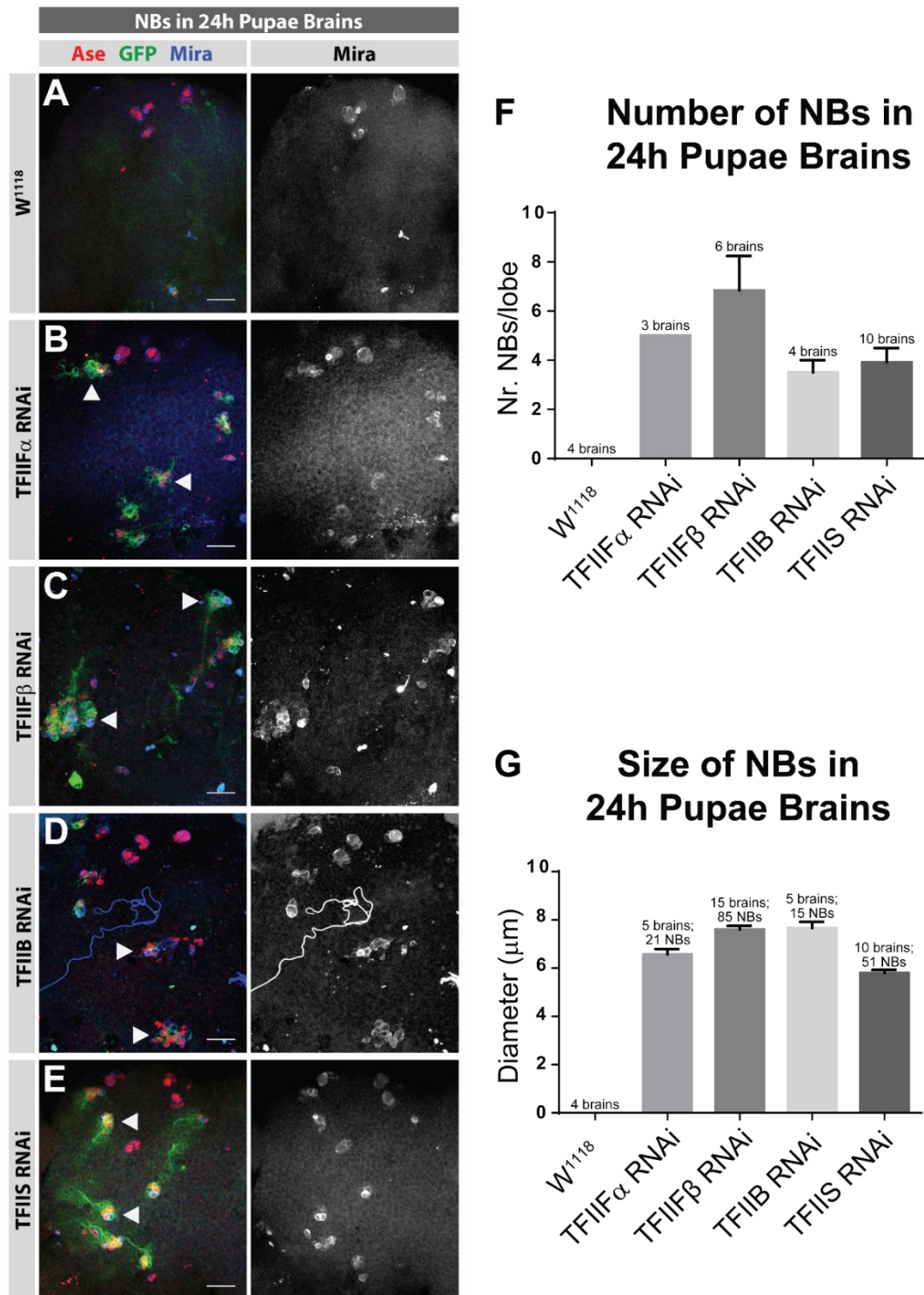


Figure I. 6. RNAi of different members of the transcriptional machinery increase type II neuroblast lifespan.

(A-E) Fixed brains 24h APF (maximum projections). Type II driver Gal4 line driving the expression of the indicated RNAi. (A) WT, (B) RNAi of TFIIF α (VDR stock #110225), (C) RNAi of TFIIF β (VDR stock #110569), (D) RNAi of TFIIB (VDR stock #106688), (E) RNAi of TFIIS (VDR stock #103718). Mira (neuroblast, NB marker), blue; Ase (INP marker in type II lineages), red; CD8::GFP, green. Arrowheads point to representative type II NBs. Scale bars represent 20 μ m.

(F,G) Quantification of the number (F) and diameter/size (G) of type II NBs. Error bars represent $\pm$ SEM.

I.3.3.3. Knock down several members of the RNA Polymerase II transcriptional machinery alters NB lineage progression

As I have previously shown, the existence of longer-lived NBs 24h APF upon TFIIF α knock down is an indication that TFIIF α has a role in regulating the end of the proliferative life of NBs. Consequently, I was also interested in determining whether TFIIF α has a role in regulating NB proliferation at earlier stages. For that reason, I decided to knock down TFIIF α in NBs, but this time assess the animals at an earlier timepoint, namely wandering 3rd instar larvae. At this developmental stage, in WT brains, NBs are still proliferating normally and generating stereotypical lineages. Consequently, if the absence of TFIIF α leads to lineage defects, that would be a strong indicator that TFIIF α is a regulator not only of NB lifespan but also of NB proliferation. At this point, I decided to focus only on type II NB lineages and to test TFIIF β and TFIIIS as well.

In a qualitative analysis it is possible to state that when either TFIIF α , TFIIF β or TFIIIS are knocked down, the lineage organization seems to be altered in wandering 3rd instar larvae. As visible in **Figure I.7A**, in the control L3 brain the lineages present a stereotypical organization with a sequence of cells with increasing degrees of differentiation. Therefore, in one tip there is the type II NB, which expresses Dpn, followed by a succession of cells with increasing levels of differentiation (INPs also express Ase) until the neurons, that only express Pros (Pros staining not shown). However, when I knocked down TFIIF α , TFIIF β or TFIIIS (**Figure I.7B-D**), the lineages were obviously more disorganized, with cells that were less differentiated (Dpn⁺ and/or Ase⁺) far away from the NB, namely, in positions where usually there are only neurons. Unfortunately, quantitative analysis was not possible as, in some cases, affected lineages would be too altered and eventually intertwined with neighboring lineages, thus making it impossible to tell them apart and accurately present the number of cells belonging to each lineage.

Altogether, knock down of several members of the transcriptional machinery leads to altered lineages in larval stages indicating that these TFs have a role in regulating proliferation. These early defects in NB proliferation may be what is ultimately causing the longer-lived phenotype at 24h APF.

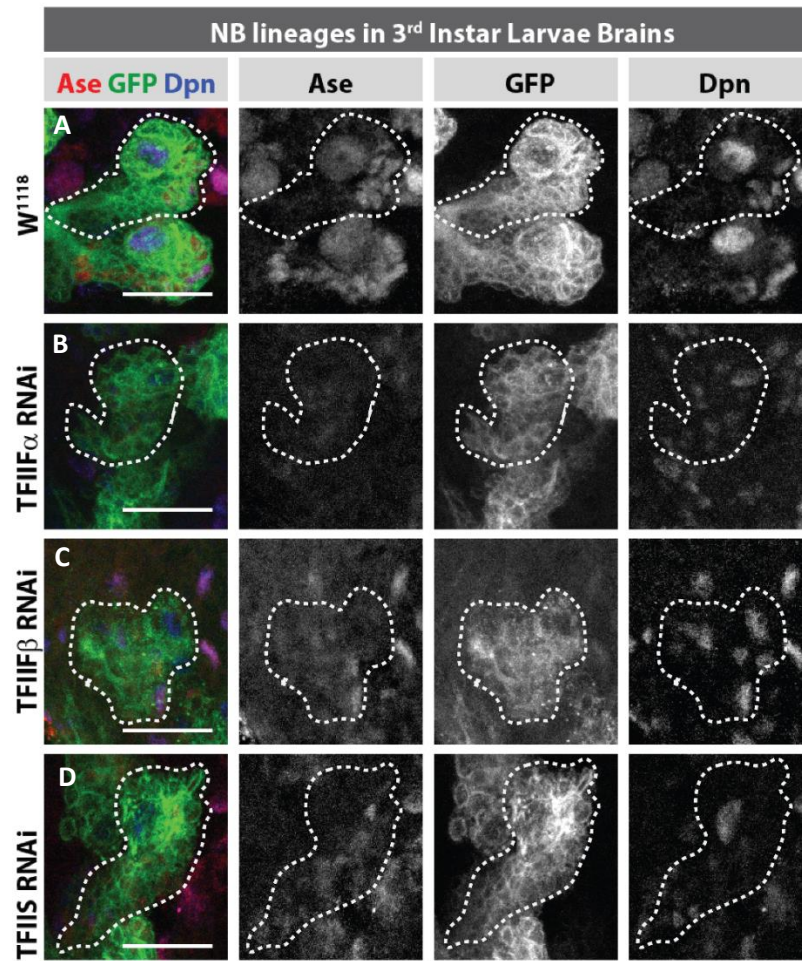


Figure I. 7. RNAi of different members of the RNA Polymerase II transcriptional machinery alter type II neuroblast lineages in wandering 3rd instar larvae.

(A-D) Fixed brains of wandering 3rd instar larvae (maximum projections). Type II driver Gal4 line driving the expression of the indicated RNAi. **(A)** WT, **(B)** RNAi of TFIIF α (VDRRC stock #110225), **(C)** RNAi of TFIIF β (VDRRC stock #110569), **(D)** RNAi of TFIIS (VDRRC stock #103718). Dpn (neuroblast, NB marker), blue; Ase (INP marker in type II lineages), red; CD8::GFP, green. Lineages are counterstained in white. Scale bars represent 20 μ m.

Considering that there seems to be something wrong with the differentiation process, I took a better look at what might be causing this misregulation. Under normal circumstances, a type II NB divides asymmetrically, originating 1 NB and 1 INP. So, when we have more than one NB cell per lineage that is usually because either (i) the NB did not differentiate (by dividing symmetrically instead of asymmetrically) or (ii) because the resulting INP was not able to maintain its fate and de-differentiated back to a NB identity. This means that the defect can either start in the NB or in the INP.

Since any of the primary NBs in the knocked down lineages, seems to be altered, this means that the defect does not appear to be directly in this primary NB. So, I decided to investigate the second hypothesis.

To test the second hypothesis, I knocked down TFII α in imINPs using a *R9D11*-Gal4 driver and checked if the lineage was affected. Results showed there were no extra NBs in the lineage, nor abnormal numbers of imINPs (average number imINPs = 27.56 in TFII α RNAi vs. 26.36 in WT; at least 3 independent brains were analyzed). These results seem to indicate that no dedifferentiation of INPs into NBs is occurring.

Data so far seems to prove neither of my initial hypothesis. However, more experiments would be necessary to corroborate these results. For hypothesis (i) a more thorough analysis of asymmetric segregation of fate determinants could be performed and for hypothesis (ii) a driver for mINPs could be tested.

At this point it seems likely that TFII α has in fact a role in regulating type I and type II NB lifespan, and also their overall proliferation. However, since TFII α is a member of a very general machinery, so far, I cannot exclude the possibility that this regulation occurs at a more general level and not in a way that is specific to NSCs.

I.3.4. Gap genes as candidates for neuroblast fate and lifespan regulation

In a parallel effort to the previously described candidate testing, I was aiming to identify possible temporal regulators of NB fate (TTFs). Within that scope, I also wanted to test if those candidate TTFs had a regulating role at the time of NB disappearance, thus controlling their lifespan.

NB fate specification is known to be determined by both temporal and spatial cascades. Body segmentation is one of the first and most important spatial determinations to occur, during which gap genes have a crucial role in determining the anterior-posterior axis. Hb and Kr, not only are gap genes with an important role in body segmentation, but they are also temporal TFs that determine NB fate throughout time. This prompts the hypothesis that there might be other gap genes that are TTFs regulating NB lifespan.

To determine if there were other gap genes expressed in NBs or in their lineages that could be acting in larval stages, I selected several gap genes and analyzed their expression in wandering 3rd instar larval brains. I used several fly lines of GFP-tagged gap genes, namely, Kr (BDSC stock #56152), tll (BDSC stock #30874), otd (BDSC stock #66393) and kni (BDSC stock #66588). I have also selected 2 other TFs that are maternally inherited: cad (BDSC stock #30875) and bcd (BDSC stocks #29018, #30869 and #67654).

I started by dissecting the brains of wandering 3rd instar larvae from all the previously mentioned GFP tagged lines and used GFP expression as a proxy for gene expression. I used Miranda to identify NBs and Pros or Elav to identify neuron (Pros is usually used to identify younger neurons, while only mature neurons express Elav). If NBs expressed one of these candidates at a certain time point, they would be expected to pass this fate to their progeny. A shift in the position of the neurons expressing these candidates regarding the distance to the NB would mean that their expression was temporally regulated: for example, in earlier timepoints they should be closer to the NB, while as time passes, if the expression of the candidates stopped, as new neurons would be formed, they would no longer express GFP, and the GFP⁺ neurons would be further away from the NB.

The only candidates that showed GFP expression during wandering 3rd instar larval stages were *tll* and *otd*. As it is shown in **Figure I.8A,B**, *tll* is expressed in most NBs as well as in the group of neurons closer to it (younger neurons). Moreover, there are few GFP⁺Pros⁺ neurons (**Figure I.8A**) and some overlap between Elav and *tll* neurons (**Figure I.8B**). *Otd* is also expressed in most NBs as well as in the closest/younger neurons (**Figure I.8C,D**). Among the GFP⁺ neurons, there were a few closer to the NB that are Pros⁺ (**Figure I.8C**) and older neurons (further away from the NB) that express Elav (**Figure I.7D**). Summing up, both *tll* and *otd* are expressed in NBs and neurons; moreover, both these gap genes are expressed in patches of both younger and older/mature neurons.

Considering that *tll* and *otd* are being expressed in wandering larva, a clear indication that they are temporal TTFs would be if their expression was specific of this stage. Therefore, I assessed younger larvae (2nd instar) and pupa 24h APF to check if these candidates are always expressed or if their expression is temporally regulated. Preliminary results indicate that at these distinct timepoints there were several NBs and neurons expressing *tll* and *otd*, indicating that these genes do not seem to be temporally regulated, at least within the resolution of the analyzed time range.

To further test if *tll* or *otd* have a role in NB lifespan I tested RNAi lines for both *tll* and *otd* (BDSC stocks#34329 and #34327, respectively). I independently knocked down both genes in type II NBs, and for both cases there were no type II NBs present 24h APF. At this point, though not promising, these results would still need to be confirmed with a different RNAi line, but due to time limitations I did not pursue that. Nonetheless, with the available results, it seems that neither *tll*, nor *otd* have a role in regulating NB lifespan. However, I cannot exclude the possibility that they still might be regulators of NB fate.

Consequently, it would be interesting to determine if knocking down these two genes in NBs would lead to defects in lineages during larval stages, such as altered fates.

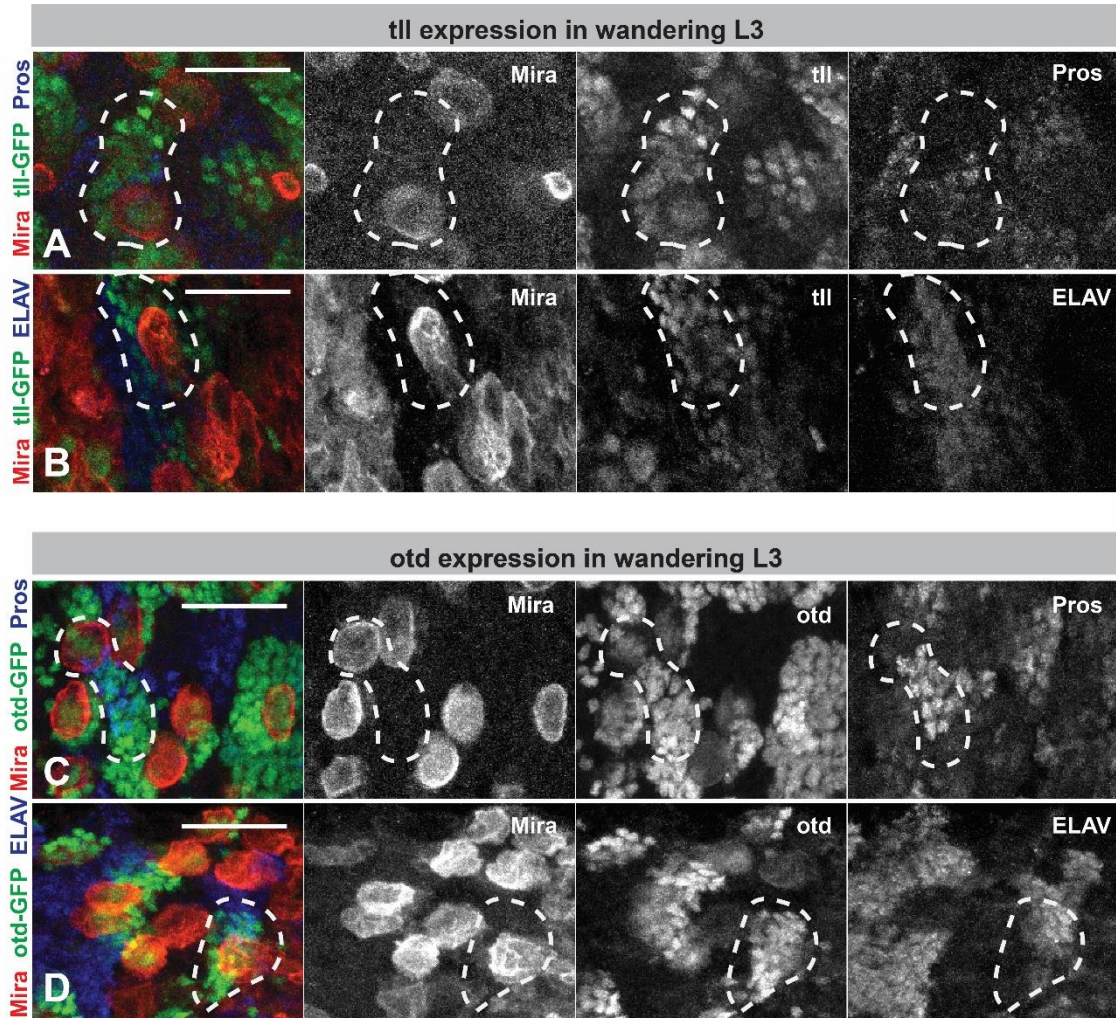


Figure I. 8. *tll* and *otd* gap genes are expressed in neuroblasts and neurons in wandering 3rd instar larvae.

Fixed brains of wandering 3rd instar larvae (L3; maximum projections).

(A,B) *tll* expression (GFP tagged line, BDSC stock #30874; (A) Pros staining, (B) Elav staining.

(C,D) *otd* expression (GFP tagged line, BDSC stock #66393; (C) Pros staining, (D) Elav staining. Mira (NB marker), red; *tll*-GFP or *otd*-GFP, green; Pros or Elav, blue.

Lineages are countoured in white. Scale bars represent 20 μm.

Additionally, despite not having shown any expression in wandering 3rd instar larvae, it would also be interesting to address whether *Kr*, *kni*, *cad* and *bcd* were expressed in other developmental stages. For instance, expression of any of these genes during an earlier larval stage or early pupal stages could indicate a possible role in NB proliferation or lifespan during those stages.

Overall, there were several open questions that remained unanswered regarding a possible role of segmentation genes as temporal regulators of NB fate and lifespan. These questions were not further pursued due to the lack of encouraging results. However, it would be interesting to address them in the future.

I.4. Discussion

Finding TFs that act as master regulators for NB proliferation regulation is crucial to understand how the correct neurons are generated at the appropriate time, as well as which are all the existing neuronal fates. The work presented in this chapter has mainly focused on the final stages of NB proliferation, when NBs division ends and they terminally differentiate. Within that scope I selected several candidates and knocked them down independently in NBs and checked if that resulted in extended lifespan of NBs. The only candidate that showed a clear phenotype was TFIIF α , a subunit of TFIIF, which is a member of the general RNA PolII transcription machinery. Further testing of other members of this machinery showed that TFIIF β (the other subunit of TFIIF), TFIIB and TFIIIS also had a similar phenotype. By knocking down these TFs specifically in NBs I was able to determine that they influence NB lineage fate and lifespan.

TFIIF, TFIIB and TFIIIS are important members of the transcriptional machinery. These specific TFs act during specific periods of the transcription process: TFIIF acts both during transcription initiation and elongation, but TFIIB only acts during initiation, while TFIIIS acts during elongation. Despite their differences, they all seem to facilitate the start of transcription. It has been described that TFIIF prevents transient RNA PolII pausing (Price et al., 1989; Rossignol et al., 1999). TFIIB is described as having a role mainly in initiation, and through interaction with TFIIF can influence start site selection (Freire-Picos et al., 2005; Khaperskyy et al., 2008). On the same note, TFIIIS is also able to stimulate transcription re-initiation after the RNA polymerase had been paused (Reines et al., 1989). Thus, it seems that the decreasing of pausing periods might be related to NB proliferation and lifespan. Hence, it makes sense that by knocking down any of these factors, transcription of a certain gene/TF relevant for regulation of NB proliferation is impaired. To the best of my knowledge, there are no studies so far correlating these members of the transcriptional machinery to fate regulation, as their effects are more often studied regarding their impact on transcription itself. However, a study has proposed that the phosphorylation of GTF2F1 (general transcription factor IIF subunit 1), the human ortholog of TFIIF, suppresses cancer cell proliferation via mTORC1 (mechanistic target of rapamycin complex 1) (He et al., 2019). Though not a connection

to fate determination, there seems to be a link in regulating highly proliferative cells. On the other hand, a study using neuroblastoma cells showed that overexpression or knock down of GTF2F2 (general transcription factor IIF subunit 2), the human ortholog of TFIIIF β , positively regulated neurite outgrowth (Tong et al., 2013). Though, once again, not a link to cell proliferation or fate, it is interesting to identify a possible connection with neural cell specification.

Studies performed by others have shown that knocking down several Tafs (TBP-associated factors; subunits of TFIID) resulted in reduced NB proliferation and defective cell polarity without affecting NB survival (Neves and Eisenman, 2019). However, I performed preliminary tests for Taf6, Taf10 and Taf11, and although I did not specifically look for polarity or lineage defects, the results indicated that knocking down these TFs do not seem to influence NB life span, at least at a consistent level. At this point, I cannot exclude the possibility that different Tafs might have complementary functions, and that I simply have not tested the ones that might actually have a role in regulating NB lifespan.

Overall, so far, I can conclude that the transcriptional machinery influences NB lineage fate and lifespan. However, it is not yet known if this happens through a mechanism regulated by the complete machinery, or if different members have specific functions. In the future, it will be interesting not only to further characterize the NB lineages after knocking down other members of the RNA PolII machinery, but also, to further study the mechanism that each of these candidates use to regulate NB life span.

The remaining candidates that were tested, did not show any striking or consistent phenotype regarding NB lifespan. However, I cannot exclude a potential effect in lineage progression as that was not assessed. In any case, it was not completely surprising that genes as *nerfin-1* and *Hey* did not result in a phenotype when knocked down in NBs as they are mainly associated with neuronal differentiation and not with NB specifically (Monastirioti et al., 2010; Xu et al., 2017). Both these candidates are related to Notch, which also plays an important role in asymmetric division and neuronal fate determination (Skeath and Thor, 2003). Preliminary data shows that Notch knockdown seems to cause an early NB disappearance in type II NBs (data not shown), which is consistence with Notch having an important role in maintaining stemness in these NBs (Li et al., 2016b).

Another candidate for NB proliferation and lifespan regulation was Stat92E. Unfortunately it presented variable phenotypes regarding NB lifespan, with the results ultimately indicating that its possible effect in NBs might occur through glia. Nonetheless, it would be interesting to further test other members of the JAK/STAT pathway, as a

recent study states that it has an essential role in regulating MB neurogenesis through CycE (Cyclin E), a G1/S cell cycle regulator (Du and Wang, 2020).

Furthermore, and taking into account the negative results regarding NB lifespan that I obtained for most of my candidates, it seems that comparing gene expression from larva to pupae at the transcriptional level, as summarized in **Table I.1** is not a suitable strategy to identify candidates to regulate NB proliferation and lifespan. For instance, Hsp67Bc and nerfin-1 showed the biggest fold-changes from larva to pupae (5.06 and 2.85, respectively) and resulted in no longer-lived NBs. Conversely, TFIIIF α had a fold-change of 0.91 and its knock down resulted in defective NB lineages and in the presence of longer-lived NBs.

Regarding the possibility that some segmentation genes might also be regulating NB proliferation and fate at later stages, I ultimately focused on tll and otd, as those were the ones that showed expression in larval stages. However, results so far show that their knockdown does not seem to increase NB lifespan. Interestingly, tll had been identified as a TTF in the OL (Li et al., 2013), however, my results indicate that at least from late 2nd instar to early pupae that does not seem to be the case in the CB.

Though I did not test the effects of knocking down or overexpressing tll in larval NBs, recent studies show that high levels of tll lead to tumorigenesis, further describing tll as necessary for type II NB identity and consequent lineage progression (Hakes and Brand, 2020).

Ultimately, there is still a lot to learn regarding NB proliferation regulation, as well as the mechanisms that lead to their disappearance in pupal stages. While I was not able to unveil any of these mechanisms during my thesis, I have contributed to the identification of candidates, such as TFIIIF α , that might be playing a role in these processes.

CHAPTER II

**Generating a transcriptomic atlas of the
Drosophila developing brain**

Authors' contributions

Graça Marques and Doctor Catarina Homem designed the experiments.

Graça Marques performed the immunofluorescence experiments.

Graça Marques, Doctor Nikolaos Konstantinides (NYU-NY) and the NYUAD Center for Genomics and Systems Biology (ADHPG-CGSB) performed the Chromium single cell protocol (10x Genomics).

Graça Marques, Patrícia Grácio and José Teles-Reis performed bioinformatic analysis of the scRNA-Seq data. Doctor Nikolaos Konstantinides (NYU-NY) and Doctor Nizar Drou (CGSB Core Bioinformatics, NYU-AD) provided assistance during bioinformatic analysis of the data.

Graça Marques wrote the chapter, under the supervision of Doctor Catarina Homem.

CHAPTER II: Generating a transcriptomic atlas of the *Drosophila* developing brain

II.1. Introduction

For the approximately 200 NBs that exist in the CB to be able to generate the 30,000 neurons of different types present in the adult animal (Ito et al., 2013), the same NB must generate different neuronal fates (Homem and Knoblich, 2012). This can be achieved by the sequential expression of a series of TTFs. A more detailed description of the temporal cascade can be found in the Introduction of this thesis (section “**7. Temporal transcription factors – temporal identity**”), but I will briefly summarize it here.

The temporal series starts at embryonic stages and presents a few variations depending on the type of NBs. In most cases, including the type I NBs from CB and VNC, that will be addressed in this chapter, the sequence is Hb > Kr > Pdm > Cas > Grh (Brody and Odenwald, 2000; Doe, 2017; Grosskortenhaus et al., 2005; Isshiki et al., 2001; Urban and Mettler, 2006). In type II NBs, this series starts later, directly with the expression of Pdm, thus skipping Hb and Kr expression (Walsh and Doe, 2017). At the end of embryogenesis NBs enter quiescence until larval stages, by when they re-express Cas (together with Dichaete (D) in type II NBs (Bayraktar and Doe, 2013)), and then Seven-up (Svp) (Maurange et al., 2008). Each TF not only determines the expression of the following TTF, but is also inherited by the GMC or INP (for type I or type II NBs, respectively), and therefore determines the type of neuron that is going to be formed (Grosskortenhaus et al., 2005; Hirono et al., 2017; Homem and Knoblich, 2012; Isshiki et al., 2001). Interestingly, type II NBs daughter cells, the INPs, have their own series of TTFs that consist in D > Grh > Ey, with slightly overlapping temporal windows of expression. (Bayraktar and Doe, 2013).

Imp and Syp are also important temporal regulators of NB identity in larval stages, creating opposing temporal gradients of expression. Imp specifies early cell fates and Syp specifies late fates, leading to the generation of neurons with specific identities depending on the Imp/Syp gradient that is at place (Liu et al., 2015; Ren et al., 2017).

However, the vast diversity of neurons in the adult *Drosophila* brain indicates that additional temporal TFs exist. Currently, not all neuronal fates are known, nor the identity of all TTF regulators. The generation of the right neurons at the right time is crucial for proper brain development; disruption of these mechanisms may lead to incorrect cell numbers or defects in cell differentiation, and therefore to neurodegenerative disorders (Homem and Knoblich, 2012; Homem et al., 2015).

Furthermore, the importance of temporal regulation of neurogenesis is also evident in mammals, with progenitor cells originating different types of neurons throughout different stages of cerebral cortical development (Desai and McConnell, 2000). For instance, Ikaros, the mammalian ortholog of the TTF Hb is described to be sufficient to set earlier neuronal fates (Alsiö et al., 2013; Elliott et al., 2008). Another example is Casz1, an ortholog of Cas in mammals, that is sufficient to establish the mid-to-late temporal identity in retina progenitor cells in mice (Mattar et al., 2015). Consequently, like in *Drosophila*, TTFs seem to have role in generating neuronal diversity in mammals, with several mechanism being quite conserved among them. This highlights the convenience and significance of using a simpler animal model, such as *Drosophila*, to better understand complex mechanisms, such as how the expression of different TFs throughout time can lead to a huge variety of neurons during brain development.

A useful way to better understand cell identity and fate choices is by studying their transcriptome. Throughout the years, bulk RNA-Seq and more recently, scRNA-Seq have proven to be valuable tools to study NBs and neurons, for the most diverse goals, such as spatial organization, neuronal diversity and temporal profile (Berger et al., 2012; Davie et al., 2018; Homem et al., 2014; Konstantinides et al., 2018; Özel et al., 2020).

As TTF expression is dynamic and sequential, meaning that each TTF is expressed during a specific temporal window, I aimed to compare the transcriptome of NBs at different timepoints. Consequently, my goal was to generate enough transcriptomic data to allow the construction of a comprehensive atlas of the *Drosophila* developing brain, featuring the transcriptomic profile of NBs and their progeny throughout larval and pupal neurogenesis. Therefore, I obtained 6 independent scRNA-Seq datasets representing CB and VNC neuronal lineages generated during sequential temporal windows ranging approximately from 60h ALH to 2h APF. In order to achieve this, I optimized the strategy to label NBs and their progeny specifically in the CB and VNC, as well as selecting suitable temporal windows to analyze. Afterwards, I used the Chromium system (10x Genomics) for the scRNA-Seq experiments and subsequently performed bioinformatic analysis of all timepoints. Each independent timepoint was analyzed until all expected cell types and differentiation states had been satisfactorily annotated. In this chapter, I will describe the main steps required for protocol optimization as well as the standard pre-processing bioinformatic analysis until cell annotation was successfully performed for each timepoint. In **Chapter III**, I will focus on one of the timepoints (105h ALH) and present a more detailed analysis, including some biological experiments.

The integration of all generated timepoints into a more complete atlas is still undergoing final optimizations and for that reason will not be presented in this thesis.

In any case, both the independent datasets and the combined version represent relevant transcriptomic data compilations that can be used to assess various aspects regarding brain development. For instance, this data can be used to study NB proliferation and differentiation, identify new TTFs or even to study neuronal specification and diversity.

II.2. Materials and Methods

Fly strains

Temporal labeling of NBs and their lineage was achieved by crossing females ;,VT012709-Gal4 (VDRC#201094; nearest genes: Koi and CG15236; CB/VNC NB driver) with ;UAS-CD8::GFP; *tub*-Gal80ts males.

For larvae timepoints, fly crosses were set up at 25°C, allowing ~12h-egg lays. First instar larvae were synchronized upon hatching and transferred to 18°C to inactivate CD8::GFP expression. CD8::GFP expression was activated 18h prior to the dissection timepoints at 60h, 69h, 78h, 87h, 96h or 105h ALH, as summarized in **Table II.2**. For the pupae timepoint, fly crosses were kept at 25°C and flipped every day. The vial with the laid eggs was immediately shifted to 18°C and shifted to 25°C after 10-15 days. After 16h at 25°C all white pupae were collected to allow synchronization at 0h pupae. White pupae were kept at 25°C for 2 more hours after which animals were dissected (2h APF). Larvae and pupae labeling was done according to the previously described timepoints and the animals were used for scRNA-Seq.

Antibodies and Immunohistochemistry

The following primary antibodies were used: rabbit anti-Mira (1:200) was a gift from J. Knoblich, and mouse anti-Prospero (1:100) was purchased at DSHB. Fluorescent secondary antibodies (Molecular Probes) were used at 1:1000.

Whole larval brains were dissected and fixed in 4% paraformaldehyde for 20min at room temperature; subsequently, brains were washed 3 times with PBS with 0.1% Triton X-100. Fixed brains were blocked for 20 min in PBS with 0.1% Triton X-100 and 1% Normal Goat Serum (blocking solution) and incubated with the primary antibodies over-night at 4°C. The next day, brains were washed 3 times with blocking solution, blocked for 20min

in blocking solution and incubated with the secondary antibodies for 2h at room temperature. Afterwards, brains were washed 3 times, incubated 10min in PBS, mounted in aqua/polymount (Polysciences, Inc) and imaged.

Immunofluorescent images were acquired on a LSM710 microscope (Carl Zeiss GmbH). Adobe Photoshop was used to adjust brightness and contrast.

Cell Dissociation and Cell Sorting

Third instar larvae (60h, 69h, 78h, 87h, 96h, 105h ALH) and 2h APF animals were collected and dissected in supplemented Schneider's medium (10% fetal bovine serum (Sigma), 20mM Glutamine (Sigma), 0.04mg/mL L-Glutathione (Sigma), 0.02mg/mL Insulin (Sigma) Schneider's medium (Sigma)). After dissection, brains were transferred to Chan & Gehring solution (Chan and Gehring, 1971) supplemented with 2% FBS, and washed once. After this, brains were enzymatically dissociated in Chan & Gehring solution 2% FBS with 1mg/mL Papain (Sigma) and 1mg/mL Collagenase I (Sigma) for 1h at 30°C. Afterwards, brains were washed once with Chan & Gehring 2% FBS solution and twice with supplemented Schneider's medium. After these washing steps, brains were resuspended in PBS 0.1% BSA (Bovine serum albumin; Sigma) and mechanically disrupted using a pipette tip. The cell suspension was filtered through a 30µL mesh into a 5mL FACS tube (BD Falcon) and immediately sorted by fluorescence activated cell sorting (FACS; FACS Aria II, BD for the experiments at NYU, New York and FACS Aria III for all the remaining experiments) (Berger et al., 2012; Harzer et al., 2014). GFP⁺ NBs and their lineage were collected in a drop of PBS 0.1% BSA. Since NBs represent a lower percentage of sorted cells when compared to neurons, they were sorted separately to assure an enrichment of less differentiated cells in the final pool. Cells were resuspended in 0.1% BSA at a final concentration of approximately 400cells/µL and immediately processed according to the 10x Genomics protocol.

The 69h and 105h ALH timepoints were processed at New York University, New York (NYU-NY), and the 60h, 69h (repetition), 78h, 87h and 96h ALH and 2h APF were processed at NYU, Abu Dhabi (AD).

10x Genomics experimental procedure

For each of the timepoints (60h, 69h, 78h, 87h, 96h and 105h ALH and 2h APF), at least 25k of sorted cells (NB lineages) were used to construct single-cell libraries; libraries

were obtained using Chromium Single Cell 3' reagent Kits v2 (10x Genomics) standard protocol. All timepoints were processed at NYU, with the 69h and 105h AHL being processed in the NY facilities, and the remaining ones (including a repetition for the 69h ALH timepoint) being processed at the AD facilities. Cells were equally divided into 2 samples (duplicates) and loaded in 2 wells of a Single Cell A Chip, aiming for an estimated target cell recovery of at least ~7k cells. Cells were then partitioned into nanoliter-scale Gel Bead-In-EMulsions (GEMs) and reverse-transcribed using an Eppendorf Mastercycler pro Thermal Cycler (Eppendorf), set for 53°C during 45min, 85°C for 5 minutes and hold at 4°C. Post reverse transcription incubation, GEMs were then broken and the complementary DNA (cDNA) was recovered and cleaned using Silane DynaBeads (Thermo Fisher Scientific). The next step consisted in amplifying the cDNA, by incubating the samples in a Thermal Cycler programmed for 98°C for 3min, 10 cycles of 98°C for 15sec, 67°C for 20sec and 72°C for 1min, followed by 72°C for 1min and hold at 4°C. The amplified cDNA was then cleaned using SPRIselect and quantified using a TapeStation/BioAnalyzer (Agilent Technologies). Moreover, the amplified cDNA was fragmented, end-repaired and A-tailed by incubating in a Thermal Cycler at 32°C for 5min, 65°C for 30min and hold at 4°C; next, the cDNA went through a double-sided size selection using SPRIselect. Subsequently, the samples went through adaptor ligation, by incubating in a Thermal Cycler at 20°C for 15min, after which there was a new SPRIselect cleanup step. Afterwards, samples were attributed independent indexes and amplified by polymerase chain reaction (PCR) using a Thermal Cycler set for 98°C for 45sec, "x" cycles at 98° for 20 sec (the "x" number of amplification cycles was determined for each sample independently and is summarized in **Table II.1**), 54°C for 30sec and 72°C for 20sec, followed by 72°C for 1min and hold at 4°C. The generated library went through a new double-sided size selection by SPRIselect and run on a TapeStation/BioAnalyzer for quality control and quantification.

Table II. 1. Number of cycles used for the cDNA amplification step in the 10x Genomics protocol

Location	Timepoint	# Amplifying Cycles (cDNA)
NYU-NY	69h ALH_NBs	13
	69h ALH_progeny	15
	105h ALH_1	14
	105h ALH_2	14
NYU-AD	60h ALH_1	16
	60h ALH_2	16
	69h ALH_1	12
	69h ALH_2	12
	78h ALH_1	11
	78h ALH_2	13
	87h ALH_1	12
	87h ALH_2	12
	96h ALH_1	12
	96h ALH_2	12
	2h APF_1	12
	2h APF_2	12

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

Quality Assessment

Quantification and quality of cDNA and respective libraries generated for 10x Genomics data were assessed with a BioAnalyzer (Agilent Technologies; experiments performed at NYU-AD) or with a TapeStation (Agilent Technologies; experiments performed at NYU-NY) following the standard protocols available at Agilent Technologies.

scRNA-Seq Raw Datasets

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

scRNA-Seq Dataset Pre-processing

The filtered gene matrices obtained after Cell Ranger processing were analyzed with R package Seurat 3.1.5. Filtering thresholds regarding the minimum number of feature counts and the percentage of mitochondrial genes to include was optimized independently for each timepoint, with help from the R Package “scater”. The final filtering parameters and the resulting number of cells and features are indicated in **Table II.2**.

Table II. 2. Summary of bioinformatic filters used for each scRNA-Seq dataset and respective outputs

Timepoint	Mito Filtering	Minimum Features (per cell)	Nr. cells	Nr. features
60h ALH	NP	NP	NP	NP
69h ALH ^a	5% (NYC); 15% (AD)	200	5,000	14,927
78h ALH	5%	528	12,195	14,908
87h ALH	10%	500	11,717	24,668
96h ALH	20%	622	10,296	8,235
105h ALH	20%	200	12,671	10,250
2h APF	15%	600	11,472	12,722

NP – not processed (due to low cell number); NYC – sample processed in NYU, New York City; AD – samples processed in NYU, Abu Dhabi.

^aWhen combining NYU and AD datasets, samples were integrated using an integrational dimensionality of 50.

Samples were normalized using the NormalizeData function, and the top 2,000 most variable features were then identified using the FindVariableFeatures function. Next, ScaleData was used to scale all genes; within this step, the percentage of mitochondrial genes was also regressed out to avoid artefacts in subsequent analysis.

scRNA-Seq Dataset Clustering

We performed a Principal Component Analysis (PCA), using the previously calculated top 2,000 most variable genes. Next, we used an elbow plot and a jackstraw approach to identify significant principal components (PCs). The number of PCs to be used on downstream analysis was selected for each timepoint independently, as to include only significant PCs ($p > 0.05$). Within the FindClusters function, the resolution parameter was also set independently for each timepoint in a way that allowed the identification of

smaller cell populations such as type II NBs and INPs. The number of PCs, resolution and the resulting number of clusters is summarized in **Table II.3**.

Table II. 3. Summary of bioinformatic settings used for each scRNA-Seq dataset and respective number of identified differentiation states

Timepoint	PCs	Resolution	Nr. Identified cell states ^a
60h ALH	NP	NP	NP
69h ALH ^b	78	1.1	10
78h ALH	57	3.1	11
87h ALH	78	3.3	13
96h ALH	50	4.1	9
105h ALH	51	1.55	10
2h APF	88	1.5	11

PC – principal components; NP – not processed (due to low cell number); NYC – sample processed in NYU, New York City; AD – samples processed in NYU, Abu Dhabi.

^aPossible identified differentiation cell states: Type I NBs, type II NBs, developing INPs (dINPs), INPs, immature GMCs (imGMCs), GMCs, type II GMCs, neurons, glia, Nervous system (unspecified), Nonannotated, Non-neural, Low quality.

Moreover, clusters were annotated into major groups corresponding to the different cell types or states identified in the datasets; resulting in up to 13 major clusters. This annotation was performed based on well described markers (dpn, ase, pros and pnt, among others), as well as based on relative cell localization within the uniform manifold approximation and projection (UMAP) plot.

II.3. Results

II.3.1. Neuroblast lineage labeling to identify new fates

NBs express a series of TTFs throughout time, that will be inherited by their progeny, and contribute to endow neurons with different fates (Doe, 2017). The neuronal fates that are generated during larval and pupal stages, which generate most of the neurons that will populate the adult brain, require more regulators than the ones described so far. One way to assess how NB proliferation and neuronal lineages change throughout time was to obtain snapshots of their transcriptome at different timepoints. To achieve that, I labeled NB lineages, from NBs to neurons, within sequential time windows, which allowed me to select close to single temporal fates. For this study, I will focus on CB and VNC lineages as they house the structures responsible for critical features such as memory-directed behavior and navigation (Cognigni et al., 2018; Seelig and Jayaraman, 2015).

Therefore, a labeling strategy that would fit my goal would have to respect the following:

- (1) ensure that the conditional genetic system results in a temporally controlled labeling;
- (2) use a driver that would only label CB and VNC NBs;
- (3) reflect a temporal window that would allow the labeling of enough neurons to sort and analyze, but still low enough to avoid different temporal fates to be included at once;
- (4) ensure that both NBs and their progeny could be successfully sorted by FACS.

First, for the conditional genetic system for temporally labeling NBs and their lineages, I used the previously introduced TARGET system (Introduction, section “**13.3. Temporal and regional gene expression targeting**”). This is a genetically controlled temperature sensitive method, based in the regulation of the UAS/Gal4 system by a temperature sensitive allele of Gal80 (McGuire et al., 2003, 2004); thus, at 19°C/18°C, the transcription of a certain gene of interest (in this case *CD8::GFP*) is repressed, however, when the temperature is shifted to 30°C, Gal80 will be degraded and, therefore, Gal4 will be expressed, driving the expression of the gene of interest. However, knowing that all the experiments would require dissecting the animals at very specific timings, which would be achieved after raising the animals at different combinations of inhibiting and permissive temperatures (18°C and 30°C), I have found it more useful to perform the activation of GFP labeling at 25°C than at 30°C. This would facilitate the calculation of the age of the animal, as is already described that growth is roughly twice as fast at 25°C than at 18°C (Powsner, 1935).

Second, regarding the candidate driver, I have selected 1 line from the Vienna Tiles (VT) Gal4 collection (VDRC line #201094) that would possibly restrict the expression of GFP to type I and/or type II NBs and their respective progeny (personal communication F. Bonnais). Since this line had not been previously published, I had to determine if it was actually only expressed in CB and VNC, and not in the OL, as intended.

Third, I tested a 18h window at 25°C to allow *CD8::GFP* activation and assessed the number of labeled neurons. Considering cell cycle duration of each differentiation state in a NB lineage (Homem et al., 2013), and assuming that GFP expression is immediately activated once the animals are shifted to 25°C, there could be a maximum of 20 labeled neurons, which is a reasonable number considering that there are approximately 200 NBs in the CB alone.

And fourth, I checked if I could successfully sort NBs and their progeny by FACS, in a similar way as previously done by (Berger et al., 2012; Harzer et al., 2014).

Therefore, I crossed ;UAS-CD8::GFP;*tub-Gal80^{ts}* males with VT-Gal4 females and kept their progeny at 18°C (permissive temperature). Afterwards, I transferred the respective progeny to 25°C when they were 3rd instar larvae, to activate the expression of CD8::GFP in CB and VNC NBs (**Figure II.1A**). Moreover, as a control, larvae were kept always at 18°C to confirm the efficiency of this system. Consequently, when larvae were always at 18°C, GFP expression was never activated (**Figure II.1B**), but when larvae went through a 18h-window at 25°C, NBs and neurons were labeled with GFP only in the CB and VNC (VNC not shown), and not in the OL (**Figure II.1C**). Moreover, I have determined that the duration of this temporal window allowed the labeling of approximately 10 neurons per NB. This number is lower than the theoretically predicted 20 neurons per NB, which suggests that GFP expression was not immediately activated after shifting the animals to 25°C. Nonetheless, I expect these numbers to be enough for a representative assessment of neuronal diversity and fate.

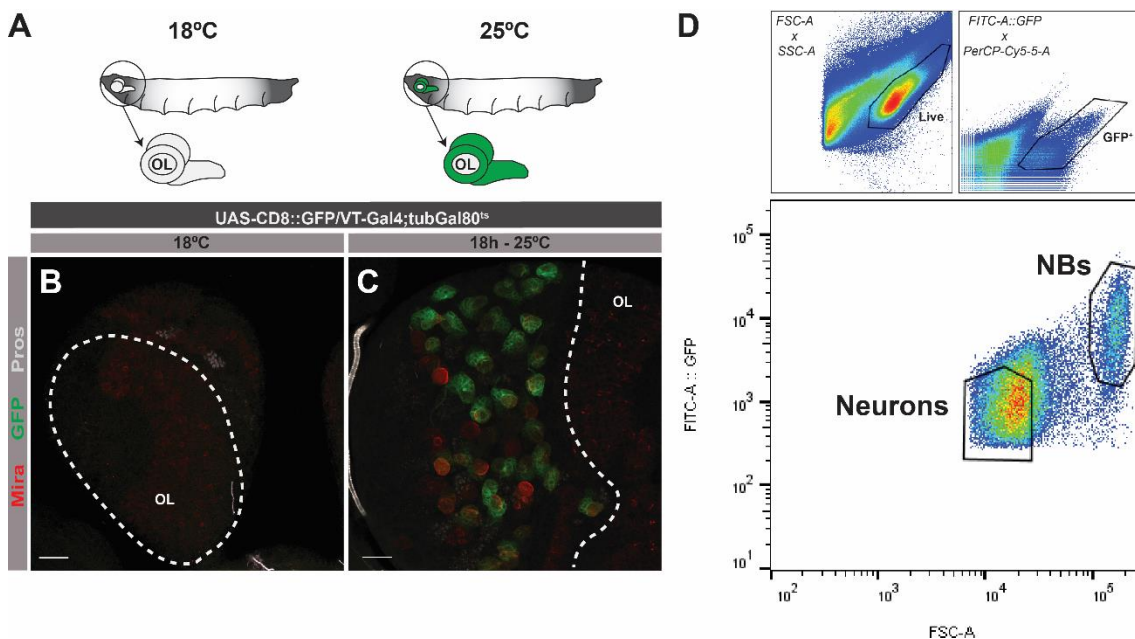


Figure II. 1. Neuroblast and neuron cell populations can be temporally and spatially labeled and sorted by FACS.

(A-C) TARGET system combined with VT-Gal4 line (VDRC ID #201094). (A) Schematic representation of the system. (B,C) Fixed third instar larvae brains developed at 18°C throughout all development (B) and after 18h at 25°C (C). Mira (neuroblast, NB marker), red, CD8::GFP, green; Pros (neuron marker), blue. Scale bars represent 20µm. Optic lobe (OL) is countered in a white dashed line.

(D) Flow cytometry dot plots representing the populations sorted by FACS. Neurons have a lower size (FSC-A, forward scatter-Area) and a lower intensity of GFP than NBs.

Next, I tested if I could sort both NBs and neurons by FACS. NBs are characterized by a higher size and a higher intensity of GFP expression than neurons (Berger et al., 2012; Harzer et al., 2014). As shown in **Figure II.1D**, both populations were clearly identifiable

in our sample. Moreover, it is important to mention that it was necessary to set independent gates for NBs and neurons instead of 1 more comprehensive gate within the GFP⁺ cells. This happens due to the great disparity in numbers between the 2 populations; as we have seen previously, each lineage appears to have 1 NB and approximately 10 neurons that are GFP⁺. This means that if I only had used 1 gate, there would be at least a 10-fold bias towards overrepresentation of neurons in comparison to NBs. For that reason, I considered it to be more advantageous to sort both populations independently. Taking all this into account, both NBs and neurons were successfully sorted.

Based on these results I determined that this conditional labeling system was suitable to label and sort CB and VNC NB lineages within restricted temporal windows.

II.3.2. Selection of timepoints to assess within the animal development

After optimization of the genetic labeling strategy, the next step was to select the timepoints to analyze at the transcriptomic level.

Since I was aiming to originate a collection of transcriptomic data from NB lineages of the developing brain, I focused mainly on the late larval stages and the first hours of pupa development. As the majority of the neurons in the adult brain are generated post-embryonically, these datasets should include a good amount of the neurons that will be present in the adult brain. A couple of the most known temporal fates described in post-embryonic stages are Chinmo and Br-C, with the shift from the former to the latter described to occur around 60h ALH (Maurange et al., 2008). With that in mind I selected 60h ALH as a starting point for the first temporal window to be labeled. Then, I decided to space the sequential times of dissection by 9h: 60h, 69h, 78h, 87h, 96h and 105h ALH.

As previously mentioned, during the 18h-window that the animals are at 18°C, there are approximately 10 neurons labeled per NB meaning that GFP expression is not immediately activated. Consequently, by distributing the timepoints every 9h, there will be a slight overlapping between the cells labelled at the beginning of the temperature shift (while GFP is still being activated) and cells labeled during the last hours of the previous timepoint.

Another aspect that might be important to mention is the fact that the duration of brain dissections and dissociation can take around 2h until cells are ready to be sorted,

meaning that during this time cells can continue dividing, thus slightly increasing the number of GFP⁺ cells more than what was initially calculated.

To be consistent throughout all the larval timepoints, larvae were always synchronized upon hatching and immediately shifted to 18°C to start developing until it was time to perform the temperature shift to activate GFP expression, which always took place 18h prior to dissection. Knowing that animals develop at different rates depending on the temperature, I have calculated timing using the previously established knowledge that animals grow approximately twice as fast at 25°C than at 18°C (Powsner, 1935). The timings for all larvae timepoints, with the appropriate 18°C/25°C relations, are summarized in **Table II.4**. For simplicity, from now on, I will refer to each timepoint by the age of dissection equivalent to development at 25°C.

Table II. 4. Developmental timings of the larval timepoints used for scRNA-Seq analysis

Inactive GFP		Active GFP	Animal age at dissection	
Time at 18°C (real time)	Equivalent if 25°C	Time at 25°C	Total (real time)	Equivalent if always at 25°C
84	42	18	102	60
102	51	18	120	69
120	60	18	138	78
138	69	18	156	87
156	78	18	174	96
174	87	18	192	105

Additionally, a last timepoint was obtained 2h APF to complete this sequence. The overall pipeline was similar as previously described, except for synchronization, which was performed at 0h APF instead of upon hatching. This alteration was necessary as by this time, hatching would have occurred too long ago to allow proper synchronization. Moreover, since the temperature shift to 25°C occurred still at larval stages, this would contribute to further dyssynchronisation upon the exact timing the animals would turn into pupae. Ultimately, this timepoint aims to represent a snapshot of NBs and their lineages during pupal stages; however, when selecting them, it was necessary to guarantee that I was activating GFP expression during a temporal window while NBs were still proliferating (because I was using a NB driver to express GFP). Moreover, this timepoint should be ideal to portray information of the temporal window where NBs might start suffering alterations that eventually will lead to their disappearance.

Ultimately, I have selected 7 temporal windows – 6 in larvae and 1 in pupae – that are aimed to portray NB lineage transcriptomic profile in the developing *Drosophila* brain.

II.3.3. Assess the transcriptome of neuroblast lineages through scRNA-Seq

After setting the labelling strategy and the timepoints to be analysed, the next step was to perform the scRNA-Seq experiments.

Doctor Claude Desplan (NYU-NY and AD) gave me the opportunity of going to his lab to perform a state-of-the-art scRNA-Seq technique using the Chromium system (10x Genomics). With this technique it is possible to process and analyze cells independently and obtain their individual transcriptomic profiles; moreover, it allows the analysis of a much higher number of cells than other single cell techniques, with the advantage of being less expensive. Consequently, I went to Doctor Claude Desplan Lab at NYU, New York City, and learned how to work with the 10x Genomics pipeline with the help of one of his PosDocs, Doctor Nikolaos Konstantinides (currently at École Normale Supérieure de Lyon, France).

The 10x Genomics pipeline, though complex, is well described by the manufacturer. Also, an important aspect that needed to be considered, was that samples had to be processed immediately after sorting, meaning that it was not possible to repeat the experiments several times until enough material was available for the transcriptomic experiment.

An alteration that seemed advantageous at this point was instead of simply sorting NBs and neurons as shown in **Figure II.1D**, to also include other cells with intermediate degrees of differentiation, such as INPs and GMCs. For that, we simply made the “neuron” gate more inclusive, but still kept it independent of the “NBs” gate to make sure there was a fair representation of this less numerous cell state.

To make sure that the 10x pipeline could be used for the purpose of my study, we decided to do an initial test with only 2 timepoints, for which we have selected 69h and 105h ALH. After determining if I could successfully obtain and analyze the transcriptomic data for these 2 timepoints, I would repeat the procedure for the remaining 5 timepoints.

We started by processing the 69h ALH timepoint as previously described, we dissected and dissociated the brains of 69h ALH larvae and then sorted them by FACS; the number of dissected brains and sorted cells is summarized in **Table II.5 – 69h ALH (NYC)**. The first step of the 10x Genomic pipeline consists in attributing barcodes to the transcriptome

of each cell, meaning that each cDNA within the same cell will have the same barcode. For this step, we have started with 2 samples, one corresponding to NBs and the other to their progeny, each with at least 15,000 cells (**Table II.5**). Both samples were resuspended to a concentration of approximately 400cells/ μ L and processed according to the specifications of the manufacturer.

Table II. 5. Amount of material (brains and cells) processed for scRNA-Seq

Timepoint	Nr. Dissected brains	Nr. of Sorted cells*		Nr. Cells used for scRNA-Seq*		
		NBs	progeny	NBs	progeny	Total
60h ALH	90	13,000	40,000	11,840	25,600	37,440
69h ALH (NYC)	268	16,900	271,980	16,900	17,225	34125
69h ALH (AD)	127	5,000	40,000	4,480	25,600	30,080
78h ALH	156	16,000	50,000	10,240	16,000	26,240
87h ALH	52	16,278	36,961	17,280	38,400	55,680
96h ALH	100	5,885	40,039	5,376	25,600	30,976
105h ALH	124	7,745	189,757	6,000	19,000	25,000
2h APF	55	22,019	39,967	17,600	25,600	43,200

*Estimated numbers; NYC – New York City; AD – Abu Dhabi.

The next steps entailed mostly cDNA amplification and library construction. The generated library had the following concentrations: 40.7pg/ μ L for NBs and 511pg/ μ L for their progeny. Though it seemed that the progeny sample had a good amount of material to submit for sequencing, unfortunately, that was not the case with the NB sample. There was no clear indication of what might have led to the low amount of the NB sample, as it was processed in parallel and in an identical way as the progeny sample. A possible explanation might be equipment malfunction upon amplification of the NB cDNA sample.

For the second sample that we tested, 105h ALH, we proceeded as previously described with one exception. After sorting, we mixed approximately 6,000 NBs with 19,000 neurons and divided them in 2 samples with a total of approximately 12,500 cells each (**Table II.5**, 105h ALH) that were resuspended as described for the 69h ALH timepoint. By having mixed samples comprising both NBs and their progeny, it would act as a backup in case something went wrong with one of the samples, as observed in the previous test. Samples were processed with the 10x Genomics protocol resulting in good quality libraries with 316pg/ μ L and 377pg/ μ L.

Once the libraries had been constructed for the total of 4 samples (2 per timepoint) they were sequenced using NovaSeq. Once the transcriptomes had been sequenced, a preliminary analysis should indicate how this pipeline fit within my experimental design and if it was advisable to proceed with the remaining timepoints. At this stage, we still

decided to submit the NB sample from the 69h ALH timepoint for sequencing, though it did not look promising.

After being sequenced, samples were processed through the Cell Ranger pipeline, a software used to analyse Chromium single cell data which allowed us to align the reads obtained after sequencing to the reference genome. As expected, the preliminary summary of the aligned data indicated that both samples from 105h ALH and the progeny sample from 69h ALH had good quality, with an estimated cell number close to 7,000 cells and with a good coverage of the genome (**Table II.6**). Regarding the NBs sample from 69h ALH, we did not get a reasonable coverage of the genome, with Cell Ranger only being able to identify very low numbers of mean reads per cell and of median genes per cells (79 and 9, respectively, **Table II.6**).

Table II. 6. Preliminary Summary according to Cell Ranger output

	Timepoint	Estimated Number of Cells	Mean Reads per Cell	Median Genes per cell
NY	69h ALH_NBs	9,325	79	9
	69h ALH_progeny	7,401	100,495	1,183
	105h ALH_1	6,259	60,319	1,409
	105h ALH_2	6,506	56,743	1,323
AD	60h ALH_1	377	990,992	709
	60h ALH_2	427	834,945	497
	69h ALH_1	4,618	78,392	988
	69h ALH_2	4,166	85,367	1,013
	78h ALH_1	13,088	24,711	1,054
	78h ALH_2	6,655	50,376	1,345
	87h ALH_1	14,340	25,074	1,259
	87h ALH_2	13,349	28,726	1,316
	96h ALH_1	4,871	71,789	1,202
	96h ALH_2	5,471	56,218	1,154
	2h APF_1	7,850	41,622	1,457
	2h APF_2	7,831	46,129	1,454

NY – New York City, AD – Abu Dhabi

Nonetheless, at this point, we were able to conclude that the 10x genomics pipeline seems to be suitable for my experimental design and we decided to move forward with the remaining timepoints (60h, 78h, 87h, 96h ALH and 2h APF). Moreover, since the 69h ALH timepoint was incomplete, as I was lacking information regarding the transcriptomic profile of NBs, I also decided to repeat that timepoint.

Therefore, I returned to Doctor Claude Desplan Lab, this time at NYU-AD, to process the remaining timepoints: 60h, 69h (repetition), 78h, 87h, 96h, 105h ALH and 2h APF. All timepoints were processed independently following the previously described pipeline and two samples were run in parallel; the number of dissected brains and processed cells is summarized in **Table II.5**. Sample quality was assessed in a BioAnalyzer, and revealed that each sample had a good profile, with concentrations ranging from 8.16ng/μL (96h ALH) to 46.80ng/μL (78h ALH). After sequencing the transcriptome of all samples, and using Cell Ranger to perform the alignment, I have determined that the protocol worked for all timepoints except for 60h ALH. All other samples had more than 4,100 cells, with a good coverage of the genome, which amounts for a good basis for downstream bioinformatic analysis (**Table II.6**). However, both samples collected at 60h ALH only allowed the identification of approximately 400 cells (**Table II.6**). This is not totally unexpected as the amount of cDNA that was generated prior to amplification was very low (36.50pg/μL and 26.46pg/μL). Even though the usage of a higher number of amplification cycles when compared to the other samples (**Table II.1**) allowed to achieve libraries with a concentration high enough for sequencing (17.08ng/μL and 11.59ng/μL, respectively), this only meant that the same few cells were being amplified more often, thus still leading to an outcome that was not representative.

Ultimately, I was able to obtain the transcriptome of 6 different timepoints representing the developing *Drosophila* brain: 69h, 78h, 87h, 96h and 105h ALH and 2h APF. Though processing of the 60h ALH timepoint was not successful, I consider that the remaining timepoints should still portray a considerable amount of information regarding *Drosophila* developing brain. At this point, the different datasets could be used for more specific bioinformatic pipelines, as will be described below.

II.3.4. Bioinformatic Setup of transcriptomic data of *Drosophila*'s developing brain

After having the transcriptomic data for the 5 larval stages (69h, 78h, 87h, 96h and 105h) and 1 pupal stage (2h APF), the next step was to bioinformatically analyze that data. Firstly, all timepoints were analyzed independently through a similar pre-processing workflow, where scRNA-Seq data went through quality control, normalization, scaling, detection of highly variable features, dimensionality reduction and annotation of the different cell types or differentiation states within the dataset. Only after this, the timepoints were used for more integrated and comparative studies among themselves.

Since the 105h ALH timepoint was the first one to be completely available, most optimizations of the analysis pipeline were performed with it. For that reason, I will now use it to describe a little more in detail the bioinformatic analysis that took place.

The first goal of the bioinformatic analysis was to perform cell clustering, basically a way to spatially organize cells in a plot, according to the similarity towards each other, which together with the features (genes) expressed by those cells, would allow us to identify different cell types or differentiation states. For this purpose we used Seurat, a R package designed for profiling transcriptomes which has become a gold standard in this type of analysis (Satija et al., 2015; Stuart et al., 2019). Hence, the first step consisted in reading the data that represents the output from the Cell Ranger pipeline. That data is stored in a counts matrix, where each number represents the number of molecules of each feature (gene) detected for each cell. After uploading the data, we created a Seurat Object where all this information, together with future analysis, will be stored. Next, the 2 samples corresponding to the 105h ALH timepoint were combined and we proceeded to quality control (QC). In **Figure II.2A**, it is possible to see the main parameters that were taken into consideration upon QC during our assessment. There was a high number of counts (UMIs) representing a good coverage of the genome. Usually, higher numbers of features are associated with the presence of doublets, however, in the 105h ALH dataset, the group of cells showing a higher number of genes being expressed is consistent with NBs, while the other group of cells with a lower number of expressed genes is consistent with their progeny, and for that reason no filter was applied. Moreover, because the presence of mitochondrial genes is often associated with damaged cells (Ilicic et al., 2016), cells with high percentage of such genes are usually filtered out. As shown in **Figure II.2A**, mitochondrial percentage is very low, which indicates that it should not be causing much impact in our dataset. Nonetheless, we tested 2 different mitochondrial filters of 5% and 20% and opted to include cells with less than 20% of mitochondrial genes, as a bigger filter (and consequent removal of more cells) did not seem to improve data quality.

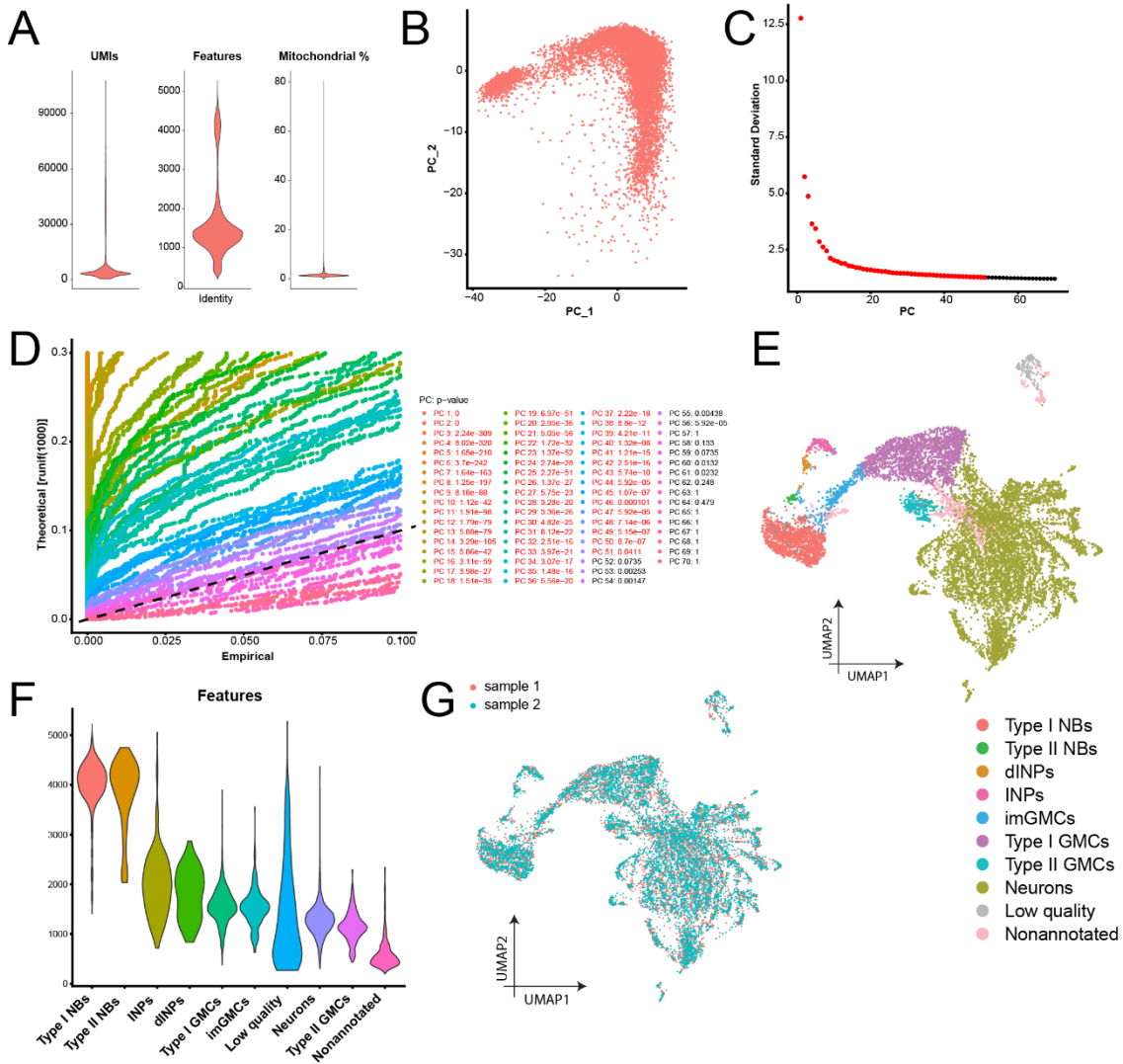


Figure II. 2. Quality control and characterization of clustering (105h ALH dataset).

(A) Quality control metrics showing unique molecular identifier (UMI) counts, feature counts and mitochondrial percentage in the initial dataset (before filtering).
 (B) Principal component analysis (PCA) on scaled data.
 (C) Elbow plot showing the ranking of principal components (PCs); PCs used for subsequent analysis are identified in red.
 (D) JackStraw plot showing the comparison of p -values for each PC; PCs used for subsequent analysis are identified in red.
 (E) UMAP (uniform manifold approximation and projection) plot of 12.7K cells composed of neuroblast (NB) lineages from central brain (CB) and ventral nerve cord (VNC).
 (F) Feature plot with the features (gene) counts for cell type or differentiation state.
 (G) UMAP plot showing the uniform distribution of both samples used in the analysis.

After filtering, the dataset should only portray good quality cells which will undergo further analysis and consists of 12,671 cells and 10,250 genes. This was followed by normalization, where each feature expression within a cell is normalized to its total expression, multiplied by a scaling factor and log-transformed. The next step consisted in identifying the genes that are highly variable throughout cells; it has been described that by focusing on those genes for subsequent analysis, such as PCA, is advantageous

in identifying true biological variability (Brennecke et al., 2013). Afterwards, data was scaled, by applying a linear transformation; moreover, at this point, we also removed the heterogeneity resulting from the contribution of mitochondrial genes. By focusing on the previously identified most variable genes, we performed a PCA to obtain a linear dimensional reduction of the data (**Figure II.2B**). Next, we determined the dimensionality of the dataset, meaning that we have selected which PCs were going to be used in the subsequent analysis. To choose the most robust and significant PCs we based our decision on the analysis of the Elbow plot and also in a JackStraw plot (**Figure II.2C,D**). An elbow plot gives the percentages of variance of each PC in a ranked visual manner, meaning that the most relevant PCs should be the ones located prior to the “elbow” of the plot (**Figure II.2C**). On the other hand, a JackStraw plot shows us the p -values for each PC, with robust PCs being associated with lower p -values (**Figure II.2D**). By analyzing both plots, we felt confident in selecting the first 51 PCs, as we were only including the most significant ones ($p > 0.05$).

To cluster the cells, we have performed modularity optimization using the Louvain algorithm; we have chosen a resolution of 1.55, a parameter that determines the granularity of the clustering, meaning that the highest the value, the more clusters there would be. These different clusters were visualized after non-linear dimensional reduction in a UMAP plot (McInnes et al., 2018), where we validated them to be appropriate to independently identify the less numerous cell states that should be present in our dataset, such as type II NBs and INPs. We have chosen to use a UMAP plot instead of other tools of dimensionality reduction such as t -distributed stochastic neighbor embedding (t -SNE) because it was described as being faster and resulting in cluster organization that was more meaningful at the biological level (Becht et al., 2019). Though more recent studies have shown that there is not necessarily any advantage of UMAP vs. t -SNE, as long as a suitable implementation is followed (Kobak and Linderman, 2021), we were satisfied with the graphical representation that it provided us.

After analyzing the expression pattern of the genes that characterize each cell state, further addressed in more detail in **Chapter III**, the different clusters were annotated concerning the cell state they represent (**Figure II.2E**). When analyzing the distribution of counts by each of the identified cell types or states (**Figure 2F**), we were able to confirm our initial assumption that cells with the higher number of counts corresponded to NSCs, while the group of cells with lower number of counts referred to their progeny, or eventually to cells with lower quality. This was expected as progenitor cells tendentially express a higher number of genes, which is linked to their potential to generate cellular diversity (Xia et al., 2020).

Moreover, as another step of QC, we checked how the 2 samples used in this dataset were distributed among the UMAP plot (**Figure II.2G**), finding that they presented very similar profiles which further strengthen the quality of our manipulation.

A more in-depth analysis of the 105h ALH timepoint is shown in **Chapter III**, however. However, I will now briefly highlight a few of the main characteristics of this timepoint, some of which contributed to the previous cell annotation presented in **Figure II.2E**. The analysis of the pattern of expression of specific markers in the cells composing our dataset, and its comparison with known markers, allowed not only to assess the identity of each cell (NB, GMC, neuron, etc), but also to obtain more information regarding their temporal fate or even spatial location. For instance, **Figure II.3A** shows that cells expressing *dpn* (red) are neural progenitors, such as type I and type II NBs, and INPs (compare with **Figure II.2E**). Conversely, neurons (**Figure II.2E**) are characterized by the expression of *elav* as shown in **Figure II.3B**. Knowing that at 105h ALH, NBs have already stopped expressing *Imp* and started expressing *Syp*, according to their temporal gradient, I also checked if that profile was reflected in our dataset. As can be seen by comparing **Figure II.3C and D**, most cells indeed are *Imp*⁻*Syp*⁺, being consistent with the temporal profile that was expected at this time. Finally, another interesting aspect to take into consideration is the possibility to identify cells based on their location. For instance, *Antp* has been associated with thoracic segments and also a few ventral ones (Carroll et al., 1986), meaning that we expect to find *Antp* expression in VNC cells and not in the CB. As shown in **Figure II.3E**, there are indeed both *Antp*⁺ and *Antp*⁻ cells, which should indicate the presence of cells from the VNC and CB, respectively. Moreover, if we focus for instance in type I NBs, only 366 (25.81%) of the 1,418 total cells within this cluster, presented counts for *Antp*, indicating that those should be VNC NBs and that the remaining 1,052 (74.19%) should be CB type I NBs. On the other hand, the cluster identified as type II NBs, a type of NBs that is only expressed in CB had no *Antp*, as expected, thus strengthening this analysis.

As for the remaining timepoints, their pre-processing was performed as previously described for the 105h ALH timepoint, with the appropriate alterations being made on filtering settings, PC and resolution selection, among other particularities, until proper cell annotation was achieved. The most relevant analysis features are summarized for all timepoints in **Tables II.2 and II.3**.

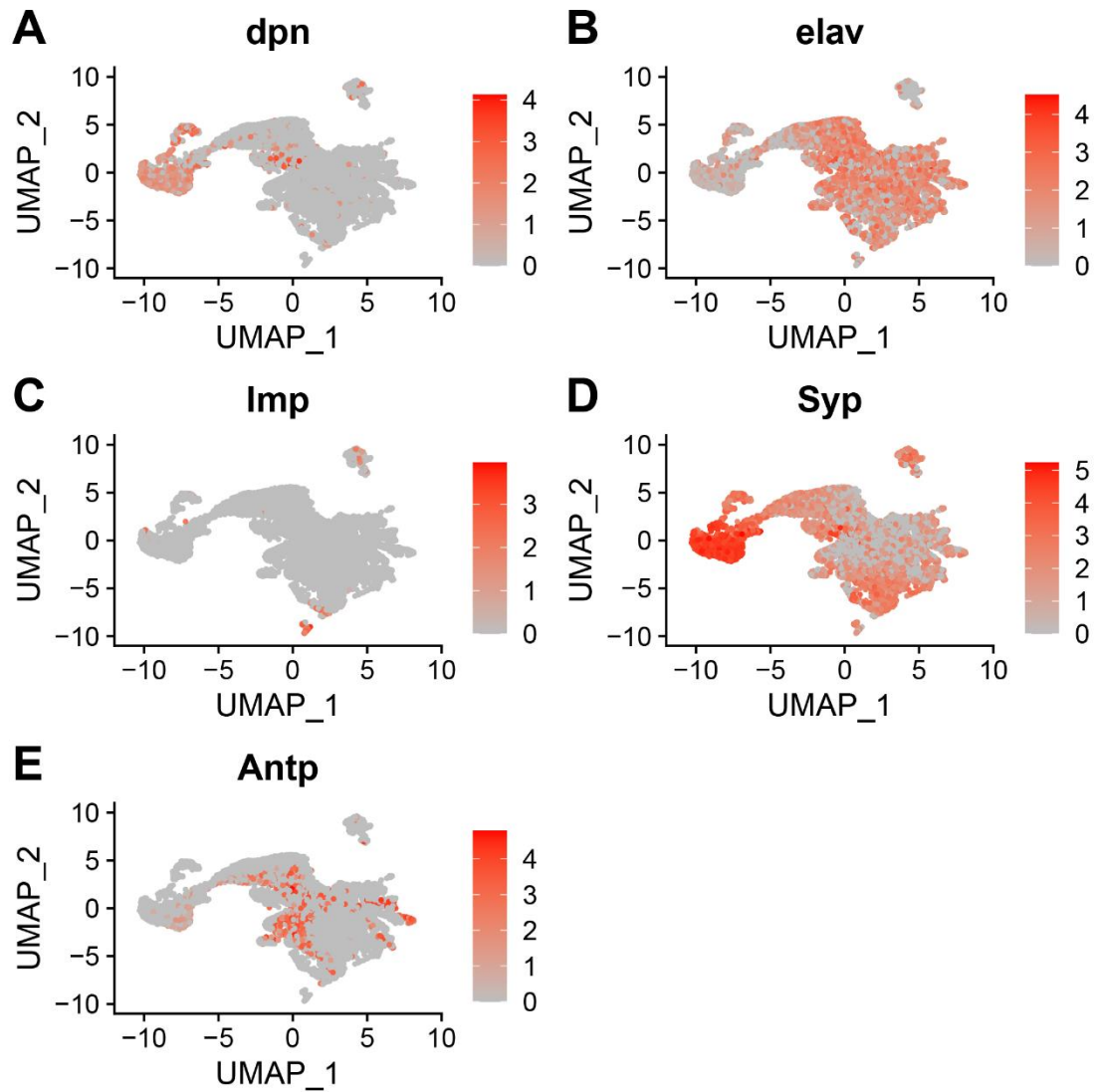


Figure II. 3. Expression profile of neural cell markers in the 105h ALH dataset.

Single cells colored according to the expression of specific genes within the previously calculated dimensional reduction shown in a UMAP (uniform manifold approximation and projection) plot: (A) *dpn*, (B) *elav*, (C) *Imp*, (D) *Syp*, (E) *Antp*. Scale ranges from grey (no expression) to red (increasing levels of expression) within data values.

Ultimately, each of these 6 datasets represents a snapshot of the developing *Drosophila* brain, having enormous potential in terms of possible bioinformatic analysis, exploring not only the inherent fate of the cell, but also how their transcriptome reflects its spatial and temporal identity.

Though each independent dataset holds value on its own, a lot of the potential of this study is the possibility to combine all datasets and explore their suitability to be used in temporal analysis.

II.4. Discussion

Previous studies of the *Drosophila* brain have provided comprehensive information regarding neuronal diversity and temporal regulation in different areas of the brain (Avalos et al., 2019; Davie et al., 2018; Konstantinides et al., 2018; Özel et al., 2020). However, to the best of my knowledge, no atlas generated so far focused specifically on the CB and VNC NB lineages in late larval and early pupal stages, a temporal window that fits within the second wave of neurogenesis, where most of the neurons that will populate the adult animal are being generated. It is important that the right neurons are formed at the right time, meaning that not only do they have to inherit the appropriate TTFs, but neuronal lineage differentiation must occur properly. Therefore, in order to better understand how neuronal formation and diversity is achieved, it is important to study neuronal lineages themselves and how they evolve throughout time. With the 5 larval (69h, 78h, 87h, 96h, 105h ALH) and 1 pupal (2h APF) datasets generated within the scope of this study, I consider having greatly contributed to fill the gap in information still existing regarding the later temporal window in which neurons are still being formed. All timepoints combined should portray NB lineages generated in the *Drosophila* brain, roughly during a 3-day period.

Overall, I have established a highly reliable pipeline from animal synchronization, to the labelling of specific neuronal population during restrict temporal windows, to the final output of the transcriptomic data. Nonetheless, it is important to highlight that the complete pipeline requires a high level of organization and coordination in order to respect the several steps involved. All steps in this process required precise sequential timings: 1st instar larvae collection, temperature shift to allow GFP activation, cell dissociation and sorting and immediate processing through 10x Genomics pipeline. Though this was very demanding in terms of time it was necessary to achieve the desired results.

Even though my initial plan featured transcriptomic analysis by bulk RNA-Seq, I had the opportunity to perform scRNA-Seq instead which is a more informative technique. By using scRNA-Seq, I was able to obtain transcriptomic data of individual cells, instead of pooled data that could be masking the contribution of smaller cell populations (as would have happened with bulk RNA-Seq). Moreover, each of our datasets seems to be robust enough to allow the identification of all expected cell types or differentiation states (NBs, INPs, GMCs and neurons). Throughout every independent dataset, a resolution of approximately 1 to 4 was enough to allow the annotation of the smallest populations (type II NBs and INPs), though there does not seem to be a direct relation between higher resolution and the number of cells being analyzed.

One of the caveats of this analysis is the fact that the final datasets do not reflect the biological ratio between the different cell states in a lineage. In order to allow for a good number of NBs to be present in our dataset, we had to sort them independently from the rest of their progeny, otherwise their number would be too diluted in the total number of cells, considering the ratio of approximately 1 NBs to 10 GFP⁺ neurons. Even though this should not be a problem for most transcriptomic characterizations, it is important to have this aspect into account if trying to establish any type of comparison that depends on accurate biological cell proportions. In any case, I do not expect this particularity to have influenced any of the analyses described in this thesis.

Another extremely useful purpose of the 6 datasets is the possibility to combine and compare them to better assess how NB lineages evolve throughout time at a transcriptomic level. Batch effects can occur in different experiments and may represent a disadvantage of this type of analysis. Batch effects seem to be having an impact once we combine the datasets processed at NYU-NY vs. AD. However, there are informatic tools that allow us to decrease that impact, which can be used to perform further improvements in our integrated datasets.

Though we have already performed preliminary studies after integrating the different datasets, unfortunately, the duration of my PhD was not sufficient to generate robust data to present and discuss in this thesis. For instance, in a very interesting study currently being developed in the lab, we are comparing the markers that characterize NBs in embryo, 1st and 3rd instar larvae and that ultimately are responsible for characterizing their spatial identity within the brain. Preliminary data indicates that there might be a major change in those markers from embryo to post-embryonic stages, as several of the TF combinations identified in the embryo are no longer featured in later stages.

Another interesting approach that is currently being pursued in the lab is trying to determine if by using transcriptomic profiles it is possible to recapitulate the biological temporal order by which the timepoints were obtained in a bioinformatic manner. In other words, we have merged all datasets and subset a single population of cells, namely CB type I NBs, and then used bioinformatic tools to assess if the transcriptomic information of these cells is enough to allow to determine their age relatively to each other. Preliminary data using Monocle3, a package that orders cells according to their pseudotime, meaning that cells are ordered within a biological process such as differentiation or maturation, showed that it was indeed possible to roughly order the datasets according to the age of the animal from which they were obtained. However,

further optimization is still required until a reliable cell trajectory is determined. Nonetheless, once such trajectory is set, it will be interesting to compare the different datasets/timepoints and identify genes that are differentially expressed throughout pseudotime or even identify new TTFs.

In the future, it would also be interesting to try to analyse the very last neurons being generated during the life of the animal, however that would be difficult using the current pipeline. Since I was selecting cells by their GFP intensity using a driver that was expressed in NBs and then inherited by their progeny, this means that I would have to start labelling the NBs at an age where they were still proliferating (most CB NBs start decreasing in size in pupal stages and disappear until 24h APF). Moreover, considering that GFP activation may take up to 9h to start in some NBs, this limits even further the age that the animal would have to be when shifted it to 25°C. Consequently, in order to allow the inclusion of the last neurons being generated in CB and VNC I would likely have to use a different driver.

These datasets represent an important starting point for future studies involving NB proliferation, lineage progression and neuronal generation. Moreover, it will be interesting to study how temporal and spatial information is portrayed at the transcriptomic level within the cells featuring these datasets. Another advantage of these datasets could be the possibility to combine them with other scRNA-Seq studies performed in other developmental stages (Avalos et al., 2019; Davie et al., 2018); or even to integrate it with other types of single cell analysis such as scATAC-Seq.

CHAPTER III

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

Authors' contributions

Graça Marques and Doctor Catarina Homem designed the project.

Graça Marques performed the genetic experiments and analyzed the data with Doctor Catarina Homem.

Ana Margarida Venda supplied de immunofluorescence images for Figure III.2A (imGMC) and Figure III.B (E(spl)my-HLH-GFP reporter).

Graça Marques and Doctor Nikolaos Konstantinides performed the scRNA-Seq experiment. Graça Marques and José Teles-Reis performed bioinformatic analysis using Seurat; Graça Marques performed the analysis using Monocle; Doctor Patrícia Brito performed the analysis using Velocity with assistance from Graça Marques and José Teles-Reis.

Graça Marques and Doctor Catarina Marques wrote the corresponding manuscript currently available at bioRxiv.

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

III.1. Introduction

Brain development is a highly complex process that requires the coordinated generation and maturation of thousands of different types of neurons and glia to ensure the formation of complex neuronal circuits. During brain development, a small number of NSCs gives rise to the large neuronal diversity in the adult. Generally, NSCs divide to originate intermediate progenitors that then generate differentiated neurons and glia. Neural lineages trees play pivotal roles in cell fate determination as transcriptional and molecular changes that occur at each step of lineage progression are inherited and ultimately determine neural cell fate specification. Indeed, the type of neuron generated by each neural lineage has been shown to be determined by the combination of several layers of TFs and signaling pathways acting at distinct steps of lineage progression. At the level of the NSC, a series of spatial and temporal TFs endow NSCs a dynamic identity which ensures the formation of diverse neuron fates throughout development (Homem et al., 2015). A second series of TFs, expressed at the level of intermediate neural progenitors, further expands neuronal fate (Briscoe and Novitch, 2008; Landgraf and Thor, 2006; Wang et al., 2014). Neuronal fate can be additionally regulated by Notch signaling, which has been shown to diversify neuron type by mediating binary fate decisions at the level of the differentiated neuronal cell (Cau and Blader, 2009; Kimura et al., 2008; Shin et al., 2007; Truman et al., 2010). After the differentiated neurons are born, they must still undergo a maturation period, during which they generate an axon and dendrites, form synapses and generate neuronal circuitries (Allan and Thor, 2015; Hobert, 2016). Thus, during maturation, neurons must coordinately express a combination of effector molecules such as cell surface genes, ion channels and NT receptors to ensure that neuronal wiring occurs only when their synaptic partners are formed (Kurmangaliyev et al., 2020). Several studies have transcriptionally analyzed mature adult neurons (Allen et al., 2020; Davie et al., 2018; Konstantinides et al., 2018; Özel et al., 2020), but very little is known about the mechanisms underlying the early stages of neuronal maturation (Bradke et al., 2020). Interestingly, it has been recently shown that the age-related transcriptomic diversity of neurons is partially lost as early as 15h after neuron birth, resulting in a transcriptomic convergence in mature neurons (Özel et al., 2020). This further indicates that the transcriptomic profiles involved in the initial phases of neuron maturation are quickly lost and can no longer be detectable in adult neurons.

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

To characterize the initial transcriptional changes in neural lineages required for neuronal formation and maturation, we have analyzed the transcriptomes of complete neural lineages in the developing *Drosophila* larval brain by single cell sequencing. *Drosophila*'s stereotypical neuronal lineages, the existence of many cell-type-specific markers, and precise developmental time points for neuron formation and synaptogenesis make it an ideal model to understand the genetic programs that regulate neural lineage progression and neuronal maturation. *Drosophila* has approximately 150,000 neurons in the entire brain, one-third in the CB and VNC and the remaining in the OL. The adult CB and VNC neurons are in their vast majority generated during larval stages, and these neurons form many key neural structures such as the MB and CX, which are, amongst other things, critical for memory-directed behavior and navigation, respectively (Cognigni et al., 2018; Seelig and Jayaraman, 2015). Neurons are produced from asymmetrically dividing NSCs called NBs. Upon asymmetric division, NBs give rise to an intermediate progenitor that usually divides once to give rise to two neurons or glia. Neurogenesis starts in the embryo (primary neurogenesis), when the neurons that form the embryo and larval CNS are formed (primary neurons). The second wave of neurogenesis (secondary neurogenesis) occurs in larval and early pupal stages when the large majority of adult neurons are formed (secondary neurons) (Truman and Bate, 1988). After secondary neurons are formed, they remain immature until mid-pupal stages when synaptogenesis starts (Hartenstein, 1993; Özel et al., 2015, 2019). The regulatory processes and gene networks involved in the formation of secondary neurons, particularly the ones from the CB and VNC, are however understudied. Previously generated bulk-seq datasets of the developing brain include both primary and secondary neurons and thus did not allow to distinguish adult neurons nor neurons with different ages. The available single-cell transcriptomic atlases did not cover the stages or the brain areas when CB and VNC adult neurons are generated (Allen et al., 2020; Avalos et al., 2019; Davie et al., 2018; Konstantinides et al., 2018; Kurmangaliyev et al., 2019, 2020; Özel et al., 2020); focused only on a sub-set of CB (type II lineages, not including type II NBs themselves) mixed with optic lobe lineages (Michki et al., 2021); or analyzed brains from the adult animal

which contain only post-mitotic fully mature neuronal cells, (Davie et al., 2018; Konstantinides et al., 2018). Thus, to characterize secondary neurogenesis and the mechanisms that regulate secondary neuron maturation we used a conditional genetic strategy to select CB and VNC NBs and their immediate progeny, including only 0h to 12.5h-old neurons (a time window prior to neuron transcriptomic convergence (Özel et al., 2020)). We took advantage of the previously obtained scRNA-Seq datasets described in **Chapter II** to assess these questions. Even though, we had successfully obtained 6 independent datasets, we have chosen to focus in only one in order to avoid transcriptional differences caused by the developmental age of the animal, rather than the age of each cell. Consequently, we performed a deeper analysis using the 105h ALH scRNA-Seq timepoint which encompasses the transcriptomes of ~12.6K single-cells of NBs and their daughter cells, ~7.2K of which were neurons.

This dataset allowed us to obtain the full transcriptome of all the described cell types in the developing CB and VNC, including type I and type II NBs, INPs, GMCs and young secondary neurons in their first phases of maturation. It additionally permitted the identification, characterization and validation of an intermediate differentiation state, the immature GMCs. It has also allowed us to identify several known as well as novel candidate regulators of lineage progression, as TFIIIF α , which was experimentally validated.

Having specifically included only young neurons, of defined ages, allowed us for the first time to transcriptionally characterize the initial phases of neuron maturation. Although synaptogenesis is only described to start days later at pupal stages, in (Chen et al., 2014; Hartenstein, 1993; Muthukumar et al., 2014; Özel et al., 2015, 2019) we have found that the young secondary larval neurons present in our dataset already transcribe several effector genes characteristic of mature neurons, such as ion channels and NT receptors. This was unexpected as synaptogenesis and NT signaling genes have been shown to be transcriptionally coordinated in other systems (Kratsios et al., 2015). We further show that NT genes, such as choline acetyltransferase (ChAT), are transcribed but not translated in young neurons and that translation of ChAT protein does not solely depend on neuron age but is rather coordinated with the animal developmental stage, only starting in pupal stages. We found that translation of ChAT initiates around stage P6-P7 (~48h APF), just prior to the start of synaptogenesis in the CB (~60h APF, (Muthukumar et al., 2014)).

Our results have allowed us to sub-divide the initial phases of secondary neuron maturation into 3 phases: a first phase when newly-born neurons do not transcribe NT

or effector genes; a second phase of maturation when neurons start transcribing, but not translating NT, ion channels and other neuron terminal feature genes; a third phase of neuron maturation when NT genes as ChAT start being translated in a coordinated manner with animal developmental stage.

III.2. Materials and Methods

Fly strains

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

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

w^{1118} was used as control for RNAi experiments. $\text{UAS-dicer2;ase-Gal4UAS-CD8::GFP}$ was used as a type I NBs driver (Harzer et al., 2014) and $\text{UAS-dicer2;wor-Gal4ase-Gal80;UAS-CD8::GFP}$ was used as a type II NB driver (Homem et al., 2014). The RNAi lines were crossed to either the control or type I and type II drivers; fly crosses for RNAi experiments were set up at 29°C to increase UAS-Gal4 expression and fluorescence intensity.

For other experiments, the lines RNAi-TK (empty) and $\text{UAS-3xFLAG-Cas9-VPR;pnt-Gal4}$ were crossed and used as control.

Antibodies and Immunohistochemistry

The following primary antibodies were used: guinea pig anti-dpn (1:1000) and rat anti-Ase (1:200) were both a gift from J. Knoblich, mouse anti-Elav (1:100) and mouse anti-ChAT4B1 (1:500) were purchased at DSHB. Rabbit anti-mira was obtained from

EurogenTec (non-commercial). Fluorescent secondary antibodies (Molecular Probes) were used at 1:1000.

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

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

Brain Dissociation and Cell Sorting

The brains of one hundred and twenty-four 3rd instar larvae (105h ALH) were dissected and dissociated according to what was described in **Chapter II** (Materials and Methods). NBs and their progeny were sorted by FACS as previously described (**Chapter II**, Materials and Methods). Cells were resuspended in 0.1% BSA at a final concentration of approximately 400cells/ μ L and immediately processed according to the 10x Genomics protocol.

10x Genomics experimental procedure

Approximately 25k of the previously sorted cells (NB lineages) were used to construct single-cell libraries; libraries were obtained using Chromium Single Cell 3' reagent Kits v2 (10x Genomics) standard protocol. Cells were equally divided into 2 samples (duplicates) and loaded in 2 wells of a Single Cell A Chip, aiming for an estimated target cell recovery of ~7k cells. The remaining pipelined was performed as described in **Chapter II** (Materials and Methods).

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

TapeStation

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

scRNA-Seq Raw Datasets

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

scRNA-Seq Dataset Pre-processing

The filtered gene matrices obtained after Cell Ranger processing were analyzed with R package Seurat 3.1.5. Only cells that had at least 200 unique feature counts were included in the analysis. Moreover, we only kept cells with a percentage of mitochondrial genes inferior to 20%. Higher percentages of mitochondrial genes are usually indicative of cell damage/rupture, and consequently, of altered overall transcriptional content (Ilicic et al., 2016). These initial quality control steps resulted in a dataset with 12,671 cells and 10,250 genes.

Samples were normalized using the NormalizeData function, and the top 2,000 most variable features were then identified using the FindVariableFeatures function. Next, ScaleData was used to scale all genes; within this step, the percentage of mitochondrial genes was also regressed out, in order to avoid artefacts in subsequent analysis.

scRNA-Seq Dataset Clustering

We performed a PCA, using the previously calculated top 2,000 most variable genes. Next, we used an elbow plot and a jackstraw approach to identify significant PCs. For the complete dataset analysis, we used the first 51 PCs, as not to include PCs that were not significant ($p > 0.05$). Within the FindClusters function, the resolution parameter was set to 1.55 as it resulted in a granularity that allowed the identification of smaller cell

populations such as type II NBs and INPs. The combination of these parameters originated 39 clusters.

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

RNA Velocity Dynamics

RNA velocity analysis was performed with the python version of Velocyto v.0.17.17 package (La Manno et al., 2018). We used the subcommand 'velocity run' to create a loom file for the cells that survived the filtering steps of Seurat pipeline using the *Drosophila melanogaster* genome annotation file (Drosophila_melanogaster.BDGP6.88.gtf) and the bam file with sorted reads that was estimated using the default parameters of the Cellranger software (10x Genomics). We masked repetitive regions using the genome expressed repetitive annotation file downloaded from UCSC genome browser. The loom file created separates molecule counts into 'spliced', 'unspliced' or 'ambiguous'. To estimate RNA velocity parameters we adapted the pipeline used in the analysis of the mouse hippocampus dataset from La Manno et al., 2018. We started by removing cells with extremely low unspliced detection requiring the sum of unspliced molecules per cell to be above the 0.4 threshold. We also selected genes that are expressed above a threshold of total number of molecules in any of the clusters requiring 40 minimum expressed (spliced) counts in at least 30 cells, after which we kept the top 3000 highly expressed and variant genes on the basis of a coefficient of variation CV vs mean fit that uses a nonparametric fit (Support Vector Regression). We applied a final filter to the dataset by selecting genes on the basis of their detection levels and cluster-wise expression threshold. This filter kept genes with unspliced molecule counts above a detection threshold of 25 minimum expressed counts detected over 20 cells, and with average counts of unspliced and spliced expression bigger than 0.01 and 0.08 respectively in at least one of the clusters. Finally, both spliced and unspliced counts were normalized for the cell size by dividing by the total number of molecules in each cell, and multiplying the mean number of molecules across all cells. All filtering steps resulted in a dataset of 12604 cells and 1086 t be used in the RNA velocyto analysis. For the preparation of the gamma fit we smooth the data using a kNN neighbors pooling approach (velocyto subcommand knn_imputation) and k=500 with calculations performed in the reduced PCA space defined by the top 99 principal

components. Velocity calculation and extrapolation to future states of the cells were performed under the assumption of constant velocity. Analysis pipeline can be obtained from the corresponding author.

Single-cell Trajectories

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

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

To construct single cell trajectories, we started by using the `differentialGeneTest` function to extract the genes distinguishing different clusters; these genes were then marked to be used for clustering in subsequent calls by using `setOrderFilter`. Afterwards, dimensions were reduced by using the Discriminative Dimension Reduction Tree (DDRTree) method, and finally, ordered using the `orderCells` function.

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

Gene annotation lists

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

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

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

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

Differential Expression between Clusters

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

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

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

Quantification of cells expressing a gene (mRNA)

To identify the number of neurons expressing genes necessary for neurotransmitter biosynthesis (*VGlut*, *VACHT*, *Gad1*, *Vmat*), we determined the number of cells with more than 0 counts for each specific gene or combination of genes. Percentages were calculated in relation to the total number of neurons in the dataset.

III.3.Results

III.3.1. Atlas of early fates in NB lineages in CB and VNC

To characterize the initial transcriptional changes that drive larval neural lineage progression and early maturation of secondary neurons we have specifically labeled, isolated and performed single cell transcriptome sequencing of larval CB and VNC NBs, intermediate progenitors and youngest neurons. Single cell transcriptome analysis allowed us to obtain a representation of every neural cell state, including transient or intermediate states, and different stages of neuron maturation. To ensure that only the lineages that originate adult neurons were analyzed we devised a conditional genetic strategy to temporally label neuronal lineages from *Drosophila* 3rd instar larvae and young pupae CB and VNC, which I have already described in **Chapter II**. Briefly, in order to fluorescently label and isolate these neural lineages, but not OL, we used the CB and VNC NB specific VT-Gal4 line#201094 (VT#201094; **Figure III.1A**). With this VT-Gal4 line, we drove the expression of CD8::GFP specifically in NBs, being consequently inherited by their daughter cells. To control the time window of GFP expression, we included a temperature sensitive tubGal80 in these animals, which allows GFP expression at 25°C, but represses GFP expression at 18°C. With this conditional genetic system, we could precisely control when NBs start expressing GFP and thus generate GFP-labeled progeny.

For this set of experiments, I focused on the 105h timepoint, as it was the first one having its pre-processing and clustering analysis successfully completed.

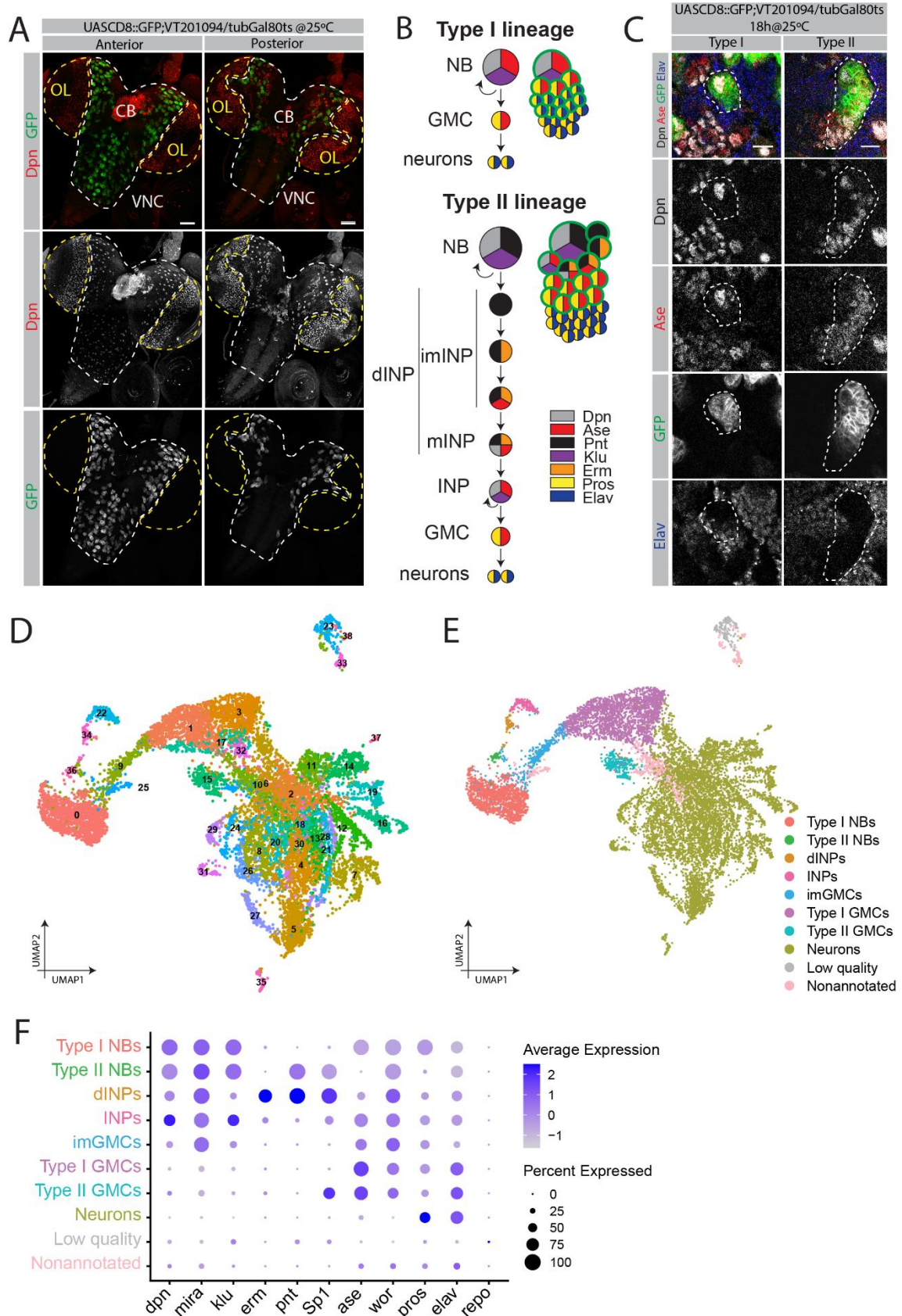


Figure III. 1. Type I and type II NB lineages identified by scRNA-Seq in wandering larvae brain.
 (A) Pattern of VT201094-Gal4 driver expression in the whole brain, showing GFP expression in NBs from CB and VNC, but not OL. Views from anterior and posterior sides; scale bar, 50µm.

- (B) Schematic representation of type I and type II lineages; cells colored by expression of *dpn* (grey), *ase* (red), *pnt* (black), *klu* (purple), *erm* (orange), *pros* (yellow) and *elav* (blue); green outline indicates the cells in which VT201094 expression is driven.
- (C) Close up of type I and type II NB lineages in the anterior side of CB; *dpn* (white), *ase* (red), GFP (green), *elav* (blue); scale bar, 10 μ m.
- (D) UMAP plot of 12.7K cells composed of NB lineages from CB and VNC.
- (E) Annotated cell types on the UMAP plot.
- (F) Dot plot showing the main genes that identify each cell type.

For our experimental set-up, precisely staged wandering 3rd instar larvae (equivalent to 105h ALH) were used: animals were raised at 18°C and shifted to 25°C for 18h prior to dissection to initiate GFP expression in NBs. This time window allows for NBs to divide 9-12 times, depending on lineage type, and for several GFP positive neurons to be generated. Neuronal lineages can be sub-divided in two types that differ in the structure of their lineage trees. Type I lineages are composed by NBs that divide every ~1.3h to self-renew and generate GMCs, which after ~4.2h divide to form neurons and glia (Homem et al., 2013) (**Figure III.1B**); consequently, the oldest neurons in our dataset can be at most ~12.5h-old (18h-1.3h-4.2h=12.5h). Type II NBs divide every ~1.6h to self-renew and generate an additional INP, which requires ~6.6h to mature and divide to self-renew and generate GMCs which then generate neurons and glia (**Figure III.1B**). Hence, in type I lineages, NBs, GMCs and the young neurons formed during this time window will be GFP⁺. In type II lineages, due to the extra differentiation step (INPs), this 18h labeling window allows for labeling of NBs, INPs and GMCs but fewer neurons (**Figure III.1B,C**). Because type II lineages are also a minority (16/~200), the majority of neurons represented in this dataset are of type I origin.

We sorted the labeled neural lineages by FACS, based on their size and GFP expression. Two samples were processed in parallel with the Chromium system (10x Genomics) and analyzed using the standard Seurat pipeline (Satija et al., 2015; Stuart et al., 2019). QC metrics attested an overall good quality of the samples, such as low percentage of mitochondrial genes as already shown in **Chapter II (Figure II.2A)** and allowed us to set appropriate filters, resulting in a library of 12,671 cells and 10,250 genes (**Table II.2**). Using Seurat's graph-based clustering approach, we identified 39 clusters with distinctively expressed marker genes, which were visualized in low dimensional space using the UMAP algorithm (McInnes et al., 2018) (**Figure III.1D**). Moreover, the 2 samples seemed to be evenly distributed through the different clusters, which strengthens the good quality of our dataset (**Figure II.2G**).

By checking the expression of known marker genes, we were able to identify and annotate the different cells that make up neuronal lineages (**Figures II.2E and III.1E,F**). NBs were globally identified by the expression of known NB markers *dpn*, *mira*, *wor* and

klu (**Figure III.1F**). Type I NBs were distinguished as they additionally co-express *ase* (cluster 0; **Figure III.1D-F**), while type II NBs do not (cluster 36; **Figure 1D-F**). Type II NBs and their lineage INPs and GMCs, were additionally identified by the expression of *Sp1* and *pnt* (**Figure III.1D-F**).

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

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

Neurons were identified based on the positive expression of *pros* or *elav*, but negative expression of *dpn*, *mira* and *wor* (clusters 2, 4-8, 11-14, 16, 18-21, 24, 26-31, 35, 37; **Figure III.1D-F**).

Finally, we identified a group of cells (cluster 9) as an intermediate state between type I NBs and GMCs. This annotation is supported, not only by the location of this cluster in the UMAP plot (between type I NBs and GMCs; **Figure III.1D,E**), but also because although transcriptional close to type I NBs, they have partially lost the expression of *dpn*, an essential marker of NBs. Namely, most of these cells (~92%), still express the NB marker *mira*, but only a third still express *dpn* (~35%) (**Figure III.1F**). As these cells also differ from GMCs, where both *Dpn* and *Mira* are not expressed, this led us to annotate them as immature GMCs (imGMCs; **Figure III.1E,F**). Notably, the presence of this transitional state indicates that the fate commitment of type I GMCs does not occur immediately after asymmetric NB division and is rather a progressive process.

In our analysis we did not identify any cluster enriched for the glial cell marker repo (**Figure III.1F**), indicating that no glial cells were included in our dataset. As neural lineages also generate glial cells (Doe, 2017), this absence might have been due to the

limited temporal window analyzed. We have additionally classified clusters characterized by the expression of ribosomal subunit genes or with no obvious expression of any of the previously mentioned markers as low quality (Ilicic et al., 2016) or nonannotated clusters (**Figure III.1E,F**).

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

III.3.2. Validation and characterization of imGMCs as a transition state between type I NBs and GMCs

To validate the existence of the transient imGMC population, we next carefully analyzed neuronal lineages *in vivo*. Based on their UMAP location, the population we called imGMCs, is transcriptionally “between” type I NBs and GMCs (**Figure III.1E**). The majority of imGMCs still express *mira*, but only one-third still express *dpn* (**Figure III.1F**). The partial absence of *dpn*, an important NB marker, indicates that imGMCs are no longer fully committed to the stem cell fate and supports these cells' transitional state into a more committed cell fate. Backing up the existence of imGMCs, occasions when Dpn is retained in the nucleus of the daughter cells after NB division have been described to sometimes occur in type I lineages of the anterior region and attributed to newborn GMCs (Boone and Doe, 2008). The presence of this intermediate cell state was also validated by us *in vivo* by immunofluorescence. Indeed, in some lineages, the daughter cells immediately next to type I NBs are Dpn⁺Mira⁺ (**Figure III.2A**, arrow). We could clearly distinguish between NBs and their daughter cells, as daughter cells have a smaller size when compared to their progenitor type I NBs. The low number of these double Dpn⁺Mira⁺ imGMCs is in accordance with our transcriptomic data, where these Dpn⁺Mira⁺ imGMCs should only represent ~33% of imGMCs. This proves that the bioinformatic analysis of our dataset can efficiently discriminate transient cell states, such as imGMCs (**Figure III.2B**).

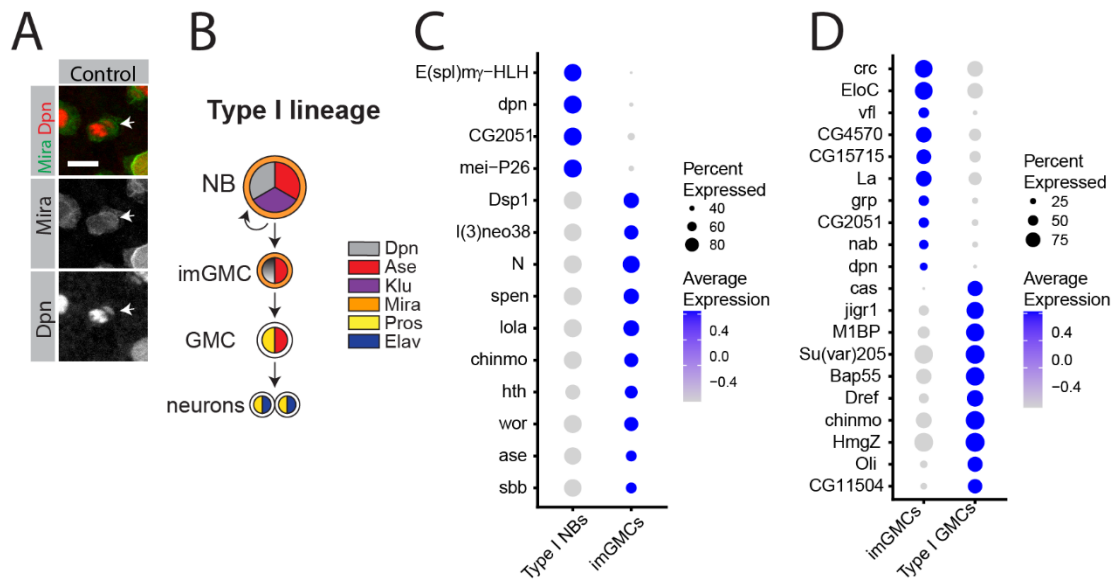


Figure III. 2. Characterization of NB-imGMCs-GMCs transition in type I lineages.

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

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

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

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

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

The comparison between imGMCs and type I NBs identified *E(spl)m γ -HLH* (*Enhancer of split m γ , helix-loop-helix*), a Notch target, as being downregulated in imGMCs (**Figure III.2C**). On the other hand, imGMCs have increased expression of several co-transcriptional repressors (as *Dsp1*, *l(3)neo38*, *spen* and *sbb*) and of genes involved in neuron differentiation (as *chinmo*, *lola*, *ase*) (**Figure III.2C**). A higher expression of transcriptional co-repressors and chromatin silencers is consistent with the required

shutdown of the stem cell program in imGMCs to transition to a more neuronal committed fate.

The comparison between imGMCs and mature GMCs has also identified several differentially expressed genes. As predicted from being a transitional state, imGMCs express higher levels of NB markers as *vfl* (or *zelda*, *zld*), *nab* and *dpn* (Reichardt et al., 2018) than their mature counterparts (**Figure III.2D**). imGMCs also express higher levels of *EloC* (**Figure III.2D**), involved in RNA PolII elongation, whose knockdown has been shown to cause a defect in GMCs (Neumüller et al., 2011). Interestingly, mature GMCs express higher levels of *Su(var)205* (or *HP1*; **Figure III.2D**) which has been recently shown to be recruited by Pros to promote heterochromatin compaction to promote terminal neuronal differentiation in GMCs (Liu and Song, 2020). Supporting their increased neuronal commitment, GMCs also express higher levels of several neuron differentiation genes than imGMCs (e.g., *jigr1* and *chinmo*; **Figure III.2D**).

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

III.3.3. Identification of candidate regulators of neuronal lineage progression by differential gene expression analysis

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

Type I NBs vs. Type I GMCs

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

Among the top TFs/DNAB genes that distinguish these two fates are known NB self-renewal factors as *dpn*, *grh*, *klu* and *E(spl)mγ* (Almeida and Bray, 2005; Berger et al.,

2012; Zacharioudaki et al., 2012) and genes involved in neuron differentiation as *jigr1* and *lola* (Southall et al., 2014; Yuva-Aydemir et al., 2015) (**Figure III.3A**). Notably, *grh* is identified as being more highly expressed in NBs vs. GMCs (**Figure III.3A**) but is not a top differentially expressed gene in the comparison of NBs vs. imGMCs (**Figure III.2C**). Consistently, *grh* expression has been previously observed *in vivo* in both NBs and in the GMC closest to the NBs (Cenci and Gould, 2005), which positionally corresponds to imGMCs, but is absent in older, more mature GMCs.

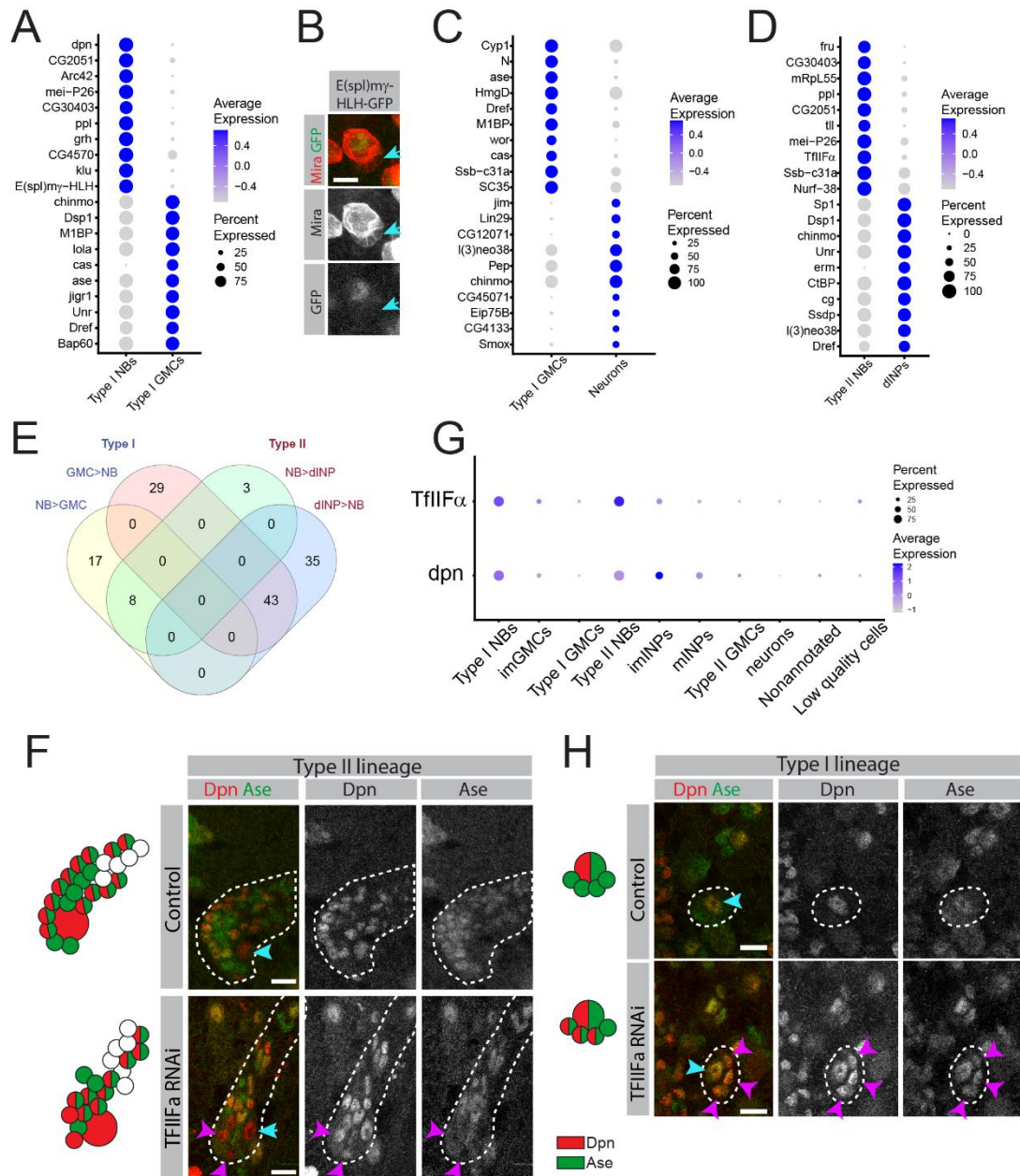


Figure III. 3. Differentially expressed markers throughout the different differentiation states in NB lineages.

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

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

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

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

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

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

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

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

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

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

Type I GMCs vs. neurons

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

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

Type II NBs vs. dINPs

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

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

To test whether TFIIF α is indeed important for the fate change from type II NBs to dINPs we expressed RNAi against it specifically in type II NBs and evaluated the impact in lineages *in vivo*. As already shown in **Chapter I**, knock down of TFIIF α in type II NBs leads to defective lineages containing ectopic Dpn⁺Ase⁻ NB-like cells (**Figure III.3F**, lower panels, pink arrowheads), a phenotype that is never seen in control brains (**Figure III.3F**, upper panels). In addition to this defective differentiation phenotype, knock down of TFIIF α in type II NBs results in disorganized lineages that frequently get mixed with each other at the level of (d)INPs/GMCs, thus preventing exact quantification of each

differentiation state. Although *Tfllfα* was not on the top 20 differentially expressed genes between type I NBs and GMCs, it is still more expressed in type I NBs than in GMCs (**Figure III.3G**). To test if *Tfllfα* has a conserved role in type I lineages, we knocked it down in type I NBs. In control type I lineages there is one type I NB (Dpn^+Ase^+), up to one imGMC (Dpn^+), and several GMCs (Dpn^+Ase^+ ; **Figure III.3H**, upper panels). However, RNAi of *Tfllfα* in type I NBs causes the formation of up to 3 cells that are Dpn^+Ase^+ like type I NBs, but have a smaller size (**Figure III.3H**, lower panels, pink arrowheads), suggesting that they are progeny that retained *Dpn* expression after NB division. This indicates that knock down of *Tfllfα* prevents proper lineage progression. In summary, *Tfllfα* RNAi leads to defects in type I and II NB lineage differentiation and super-numerary NB-like cells, thus showing that *Tfllfα* plays an important role in fate regulation (as it had already been shown in **Chapter II**).

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

III.3.4. Transcriptomic differences distinguish cells with different degrees of differentiation

The differential gene expression analysis nicely revealed an increasing level of neuronal differentiation from NBs to neurons. Remarkably, the UMAP plot itself (**Figure III.1E**) recapitulates the *in vivo* NB lineage progression order for both type I and type II lineages: NB → imGMCs → GMCs → neurons and NB → dINPs → INPs → GMCs, respectively. We have further validated the order of lineage progression in our dataset by predicting the future state of each cell using the RNA velocity method (La Manno et al., 2018) (**Figure III.4A**).

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

Figure III.4B,C). The type II lineage trajectory identified two clear GMC fate branches (**Figure III.4C**). However, manual analysis of the genes differentially expressed between both branches revealed that the same genes are expressed with only slightly different expression levels. We hypothesize that these 2 branches might be an artifact caused by the low number of cells available from type II lineages and to the lack of more differentiated cells (neurons) in the trajectory to allow a more complete comparison.

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

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

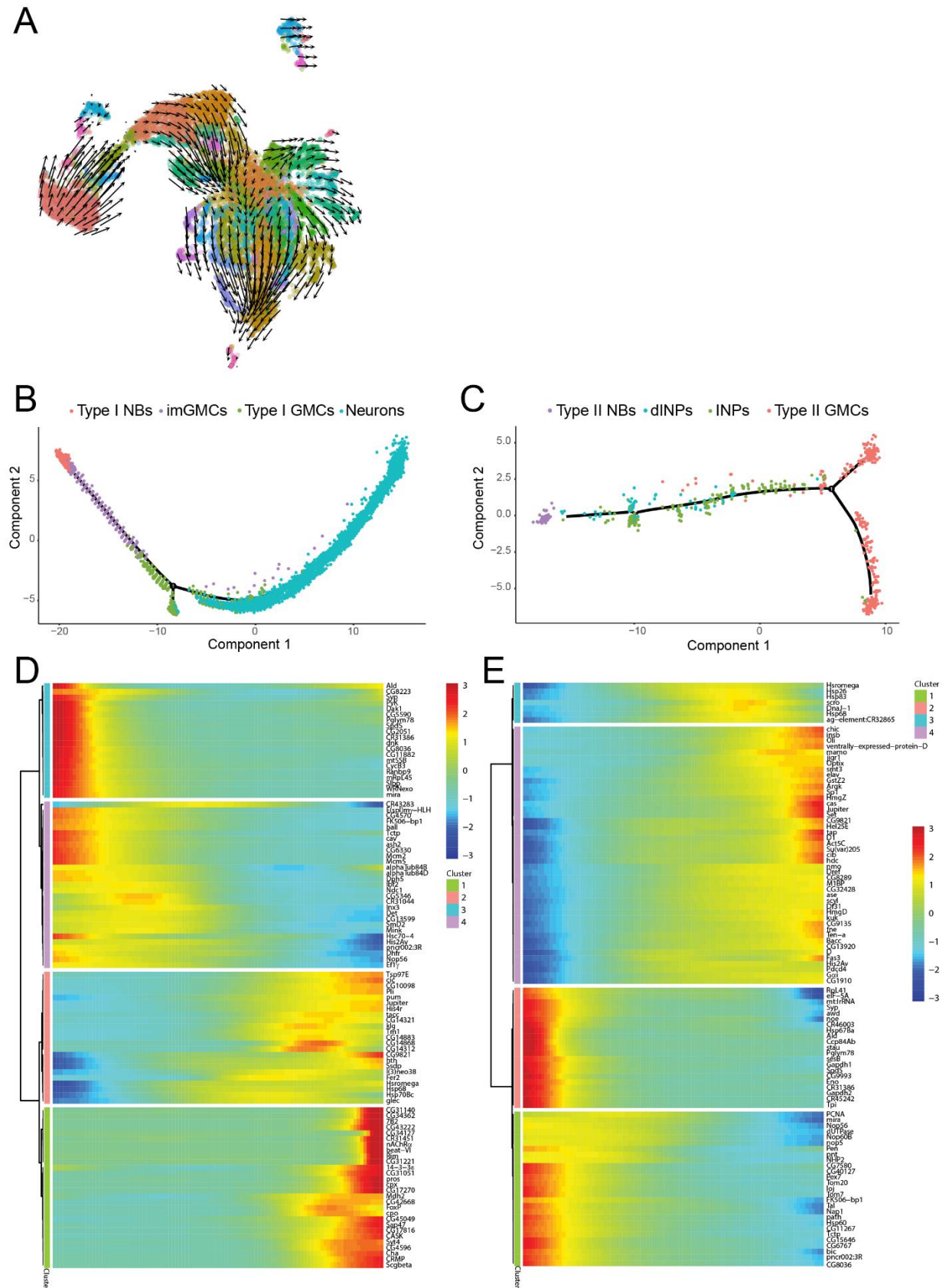


Figure III. 4. Transcriptomic dynamics of differentiation in type I and type II NB lineages.

(A) Velocity field projected on the UMAP plot; arrows indicate the average velocity at each region.
 (B and C) Cell trajectory for type I NB lineages (B) and type II NB lineages (C); cell ordering and trajectory defined based on pseudotime analysis using monocle; colored by cell differentiation state.

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

III.3.5. Transcriptomic differences distinguish neurons by age

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

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

revealed that the highest percentage amounted to 0.65% for the simultaneous presence of VGlut and VACHT.

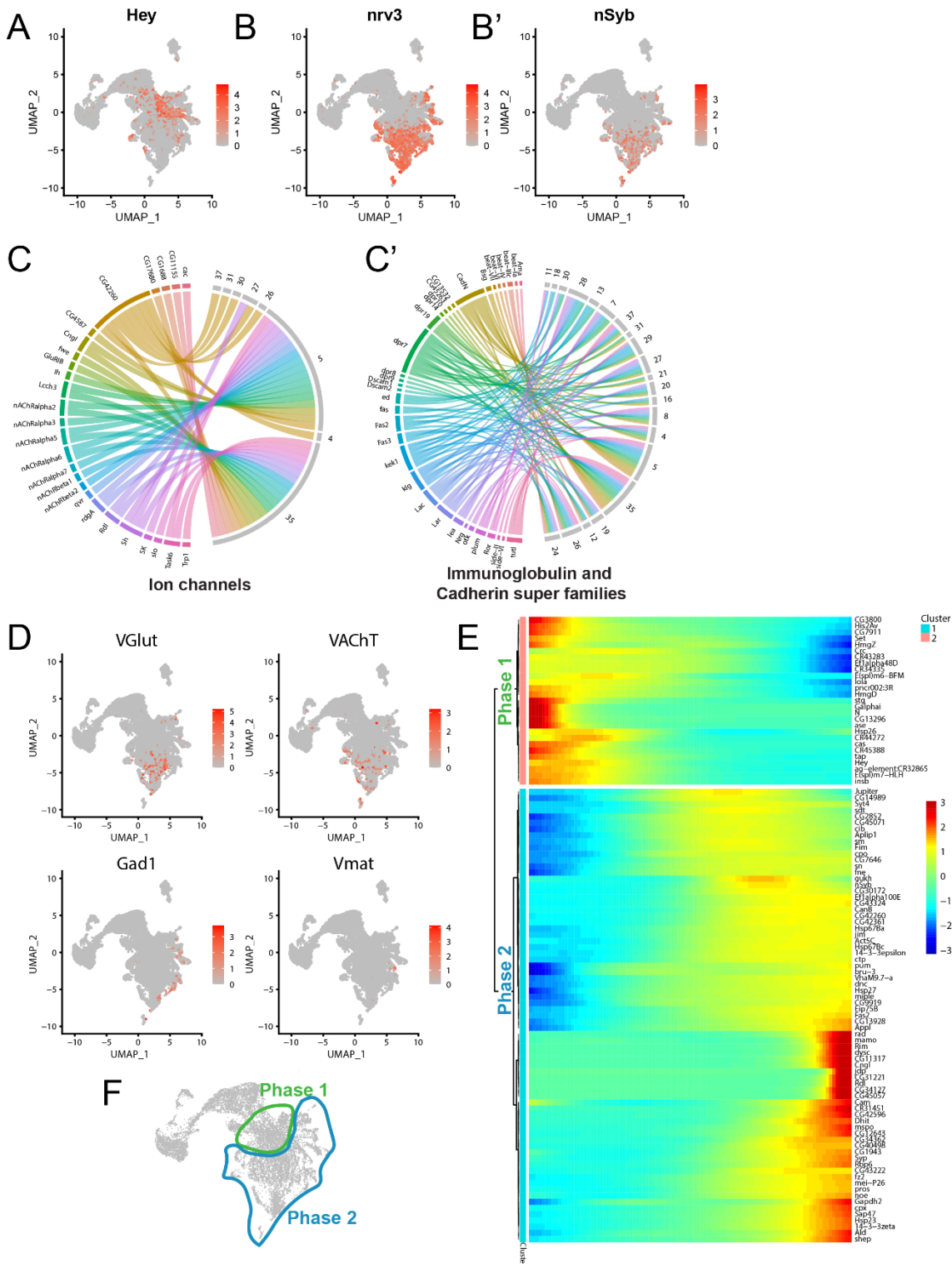


Figure III. 5. Transcriptomic differences in neurons identify 2 phases of maturation.
(A) Feature plot for the young neuronal markers (*Hey*).
(B) Feature plots for maturation markers (*nrv3*, (*B'*) *nSyb*).
(C,C') Chord diagrams showing the correspondence between Ion channels (C) and Immunoglobulin and Cadherin super families (C') and the clusters in which they are differentially expressed.
(D) Feature plots for genes necessary for neurotransmitter biosynthesis (*VGlut*, *VACHT*, *Gad1*, *Vmat*).

(E) Heatmap with the top 100 genes most differentially expressed throughout pseudotime for neurons; differential expression analysis performed with monocle.
 (F) Schematic representation of a UMAP plot showing the division of neurons according to the 2 phases of neuronal maturation proposed by us.

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

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

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

III.3.6. Neurons in phase 2 of maturation transcribe the neurotransmitter gene ChAT, but do not translate it into protein

The expression of NT genes, ion channels and other neuron functional genes in secondary neurons in larval stages was surprising as these neurons will only initiate

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

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

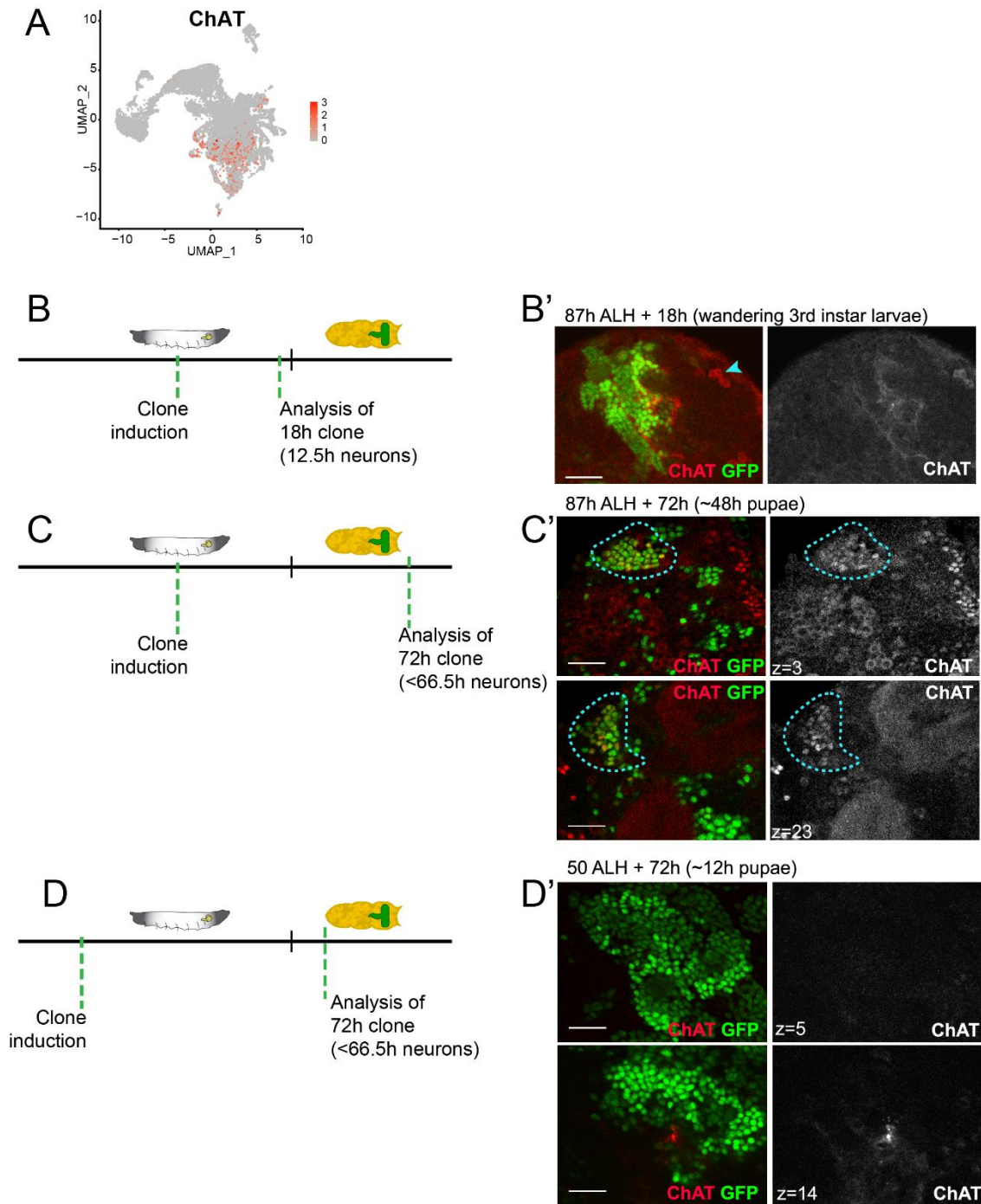


Figure III. 6. ChAT expression in neurons with different ages.

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

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

III.3.7. Initiation of ChAT translation is coordinated with animal developmental stage rather than neuron age

The presence of ChAT mRNA but the absence of protein suggests that there is a delay in the initiation of its translation in maturing neurons. To understand what determines translation initiation of ChAT in maturing neurons we tested the following two hypotheses: 1) translation of ChAT is dependent of neuron age and requires neurons to be older/of a certain age; 2) translation of ChAT is coordinated with animal developmental stage starting only when synaptogenesis begins in pupal stages. In order to test these hypotheses, we generated permanent labeling clones which allowed us to fluorescently label neuronal lineages, and thus neurons from birth, and follow them until a certain age and/or a specific developmental stage. In a 18h clone (starting at 87h ALH) with analysis at 105h ALH (wandering 3rd instar larva) there is no expression of ChAT protein in GFP labeled neurons (**Figure III.6B**). So, next, we allowed these neurons to age for a longer time by developing a 72h-clone, allowing the older neurons to be up to 66.5h old (**Figure 6C**), and analyzing these clones at 159h ALH (~48h APF). In these clones, co-expression of GFP and ChAT proteins is detected in some neurons (**Figure 6C'**). However, at this point, these results still fit both hypotheses, as neurons may be expressing ChAT because they are older than in the previous clones, or simply because the animal itself is older and closer to the onset of synaptogenesis. In order to uncouple neuronal age from the animal age, we generated a clone for the same duration of 72h but at an earlier time during animal development. We have thus induced clone formation at 50h ALH (**Figure 6D**). This new timeline still allows neurons to age up to 66.5h-old, but the animal itself, though still a pupae, would be only ~12h APF (122h ALH). Interestingly, in these clones there is no co-expression of GFP and ChAT (**Figure 6D'**). These results allowed us to conclude that ChAT protein translation does not start depending on the neuronal age itself, being rather dependent on the age of the animal. As we have designed this experiment to allow the neurons in these clones to undergo the larva to pupa transition, which is driven by a pulse of the steroid hormone ecdysone, these results further allow us to conclude that neither the larva to pupa transition, nor the ecdysone pupariation peak, are sufficient to initiate ChAT translation.

Based on these data, we propose that within neuronal maturation, there is a “phase 1” where newly-born neurons do not transcribe NTs and other terminal effector molecules, followed by a “phase 2” when neurons start transcribing NT, ion channels and other terminal feature genes. In “phase 2”, and using ChAT as an example, we show that its mRNA is however kept untranslated. Afterwards, approximately at 48h APF, a “phase 3” of maturation starts, when the neuron begins translation of ChAT in a coordinated

manner with animal and brain development (**Figure III.7A**). At this point, we can only suggest this model for neuroactive molecules, more specifically for ChAT, but we propose this to be a conserved mechanism for NT genes and possibly other molecules associated with more differentiated neuronal characteristics, such as ion channels. Interestingly several RNA binding proteins and translation regulators such as Syncrin (Syp), musashi (msi), pumilio (pum), brain tumor (Brat), or polyA-binding protein interacting protein 2 (Paip2), are expressed in phase 1 and phase 2 neurons (**Figure III.7B**), which might provide a molecular mechanism for translation inhibition of NT genes during larval and young pupal stages. Although the pupariation pulse of ecdysone does not seem to be the trigger of the molecular mechanisms that will ultimately drive NT translation, other important pulses of ecdysone occur during pupal development which might provide for the overarching temporal cue for translation initiation and phase 3 of neuron maturation.

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

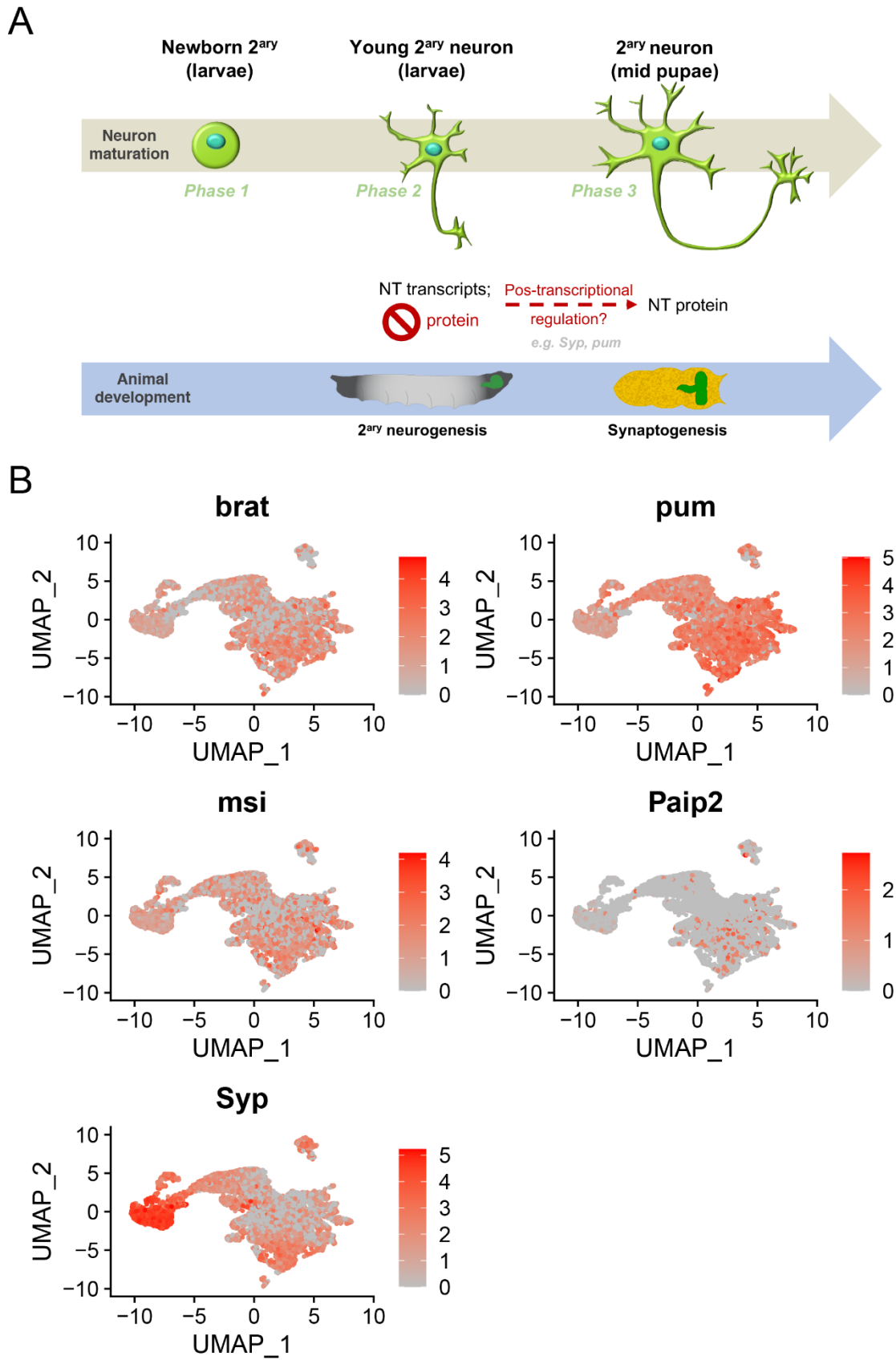


Figure III. 7. Neuronal maturation occurs in 3 phases possibly regulated at the translation level.

(A) Model for neuronal maturation.

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

III.4. Discussion

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

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

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

Interestingly, we did not identify any cluster that presented a clear glial identity (that would be identified for example by the expression of *repo*). One of the possible explanations for this is the fact that glial cells are described to be generated from CB and

VNC NBs during limited periods of time and those might not be featured within the labeling temporal window used for our dataset (Baek et al., 2013; Bayraktar and Doe, 2013; Enriquez et al., 2018; Truman et al., 2004).

By comparing the transcriptional profile of different cell states throughout differentiation we were able to identify possible candidates to regulate lineage differentiation. Such example was *TFIIF α* , a member of the RNA PolII machinery that was identified in our dataset as being differentially expressed in type I NBs vs. GMCs, which we have shown to be important to maintain proper NB larval lineage development *in vivo*. TFIIF α has important roles during transcription initiation, elongation and even pausing (Gong et al., 1993; Landick, 2006; Price et al., 1989; Rossignol et al., 1999). Based on its role in different steps of transcription, TFIIF α , either alone or associated with other TFs, might be influencing transcription start site selection (Freire-Picos et al., 2005; Khaperskyy et al., 2008) or preventing RNA PolII pausing (Price et al., 1989; Rossignol et al., 1999). Alterations in the usual rates by which these processes happen, could affect the transcription of genes responsible for NB differentiation, thus justifying the important role of TFIIF α in regulating lineage progression. In the future it might be interesting to study the mechanisms that directly or indirectly link TFIIF α to NB lineage regulation. Nonetheless, we have identified several other potential regulators of state progression within NB lineages and it would be interesting to validate them, as well as to understand their mechanisms of action. Ultimately, such studies would contribute to better understand not only NB proliferation, but also progression towards differentiation.

One of the most interesting findings in this study was determining that young neurons (<12.5h-old) have transcriptional profiles that reflect their different ages and consequently different degrees of maturation. Our analysis has led us to propose their sub-division into 3 phases of neuron maturation. A “phase 1”, composed by immature neurons that have yet to start expressing mRNA of more mature neuronal features. Followed by a “phase 2” of maturation, composed by the oldest neurons in our dataset, which start expressing mRNA of maturation markers such as genes involved in axonal development, neuronal circuit organization, neurotransmitter activity and ion channels. And finally, a “phase 3” of neuronal maturation where not only mRNA, but also the protein of more mature neuronal features are expressed. Using the NT gene ChAT as a case study, we have found that its mRNA, but not its protein, is expressed in the “phase 2” neurons in our dataset. We have shown that regardless of neuronal age, it is not until approximately 24h-48h APF that ChAT protein starts being expressed (“phase 3”). This indicates that ChAT translation depends on the age of the animal, rather than on the age of the neuron itself. Ultimately, such timeline seems to match the beginning of

synaptogenesis, which is described to start around 60h APF (Muthukumar et al., 2014). The early expression of mRNA of neuron maturation markers was unexpected since the neurons formed in larval stages, meant for the adult brain, only terminally differentiate and form dendritic and axonal branches in the pupal period (Dumstreit et al., 2003; Pereanu and Hartenstein, 2006; Truman, 1992). So, if ChAT protein is only expressed days after, why does its mRNA starts being expressed so much sooner? And how is its translation being inhibited? One hypothesis is that translation inhibitors might be acting during phases 1 and 2 of neuronal maturation, specifically keeping terminal neuronal genes untranslated and consequently keeping the neuron in a state that is not yet fully mature, until the time comes when their synaptic partners are formed and ready to connect. This hypothesis is supported by the fact that we can find several RNA binding molecules and translation inhibitors such as *Syp*, *pum* and *brat* expressed in the neurons in our dataset. Moreover, according to this hypothesis, in order to allow the translation of ChAT and other NT-associated molecules to be possible in “phase 3”, these translation inhibitors would then need to be downregulated or inactivated at the appropriate pupal stage, a model which would need to be confirmed *in vivo*. As upstream coordinators of this switch in post-transcriptional regulation and neuron maturation, we can envision hormones, which as systemic signals are ideally placed to coordinate animal and organ development. Although we have shown that the pupariation pulse of the hormone ecdysone is not sufficient to initiate ChAT translation, we do not exclude that other pupal pulses of ecdysone may provide such a function. Overall, a mechanism of temporal regulation of translation may provide an efficient responsive mechanism to ensure that protein synthesis can quickly initiate for the final phase of neuron maturation and synapse formation. Ultimately, the identification of these 3 different phases of neuronal maturation represents an important foundation for further studies to understand the mechanism and timelines that regulate neuronal maturation.

CONCLUDING REMARKS

CONCLUDING REMARKS

The work developed within the scope of this thesis aimed not only to identify candidate regulators of NB proliferation, lifespan, and fate, but also to characterize neuronal lineage differentiation in the developing *Drosophila* brain.

Consequently, I consider that the major contribution of this study to the field of neural development was the generation of comprehensive scRNA-Seq atlas of secondary neuronal lineages including their new-born neurons in 6 consecutive stages of *Drosophila* brain development. This atlas combines the single cell transcriptomic information of 6 independent datasets, ranging from shortly after 60h ALH until 2h APF (**Table II.4**). Moreover, all these datasets, focused on CB and VNC neuronal lineages, were robust enough to identify type I and type II NBs, INPs, type I and type II GMCs and neurons. Furthermore, this study performed the optimization of a robust genetic design to ensure timed collection of neural lineages and a subsequent practical and bioinformatic pipeline that can be adapted to other experimental designs.

Unfortunately, the duration of my PhD did not allow me to perform a robust analysis in all datasets, the amount of generated data will serve as basis for various future studies. One of such studies, already undergoing, aims at identifying temporal regulators of NB proliferation, which would hopefully help to complement the temporal cascade known so far (Doe, 2017). Such analysis could be performed by analyzing the specific subset of NBs (eventually further filtered to select specific NB populations) to identify genes whose transcriptional profile varies along the multiple time points of brain development. Since NB division in CB and VNC is not synchronized we could further order NBs by pseudotime using a package such as Monocle. Such analysis would be benefited by the partial overlap between datasets.

Though it was not possible to complete that type of integrated analysis for now, the study of a single dataset, namely 105h ALH, already resulted in several interesting findings. We were able to compare cells with different degrees of differentiation in the same lineages and successfully identify genes being differentially expressed between them (**Figures III.2 and 3**). Such genes represent strong candidates to regulate the differentiation states within a lineage.

For instance, TFIIIF α was identified as being significantly more expressed in Type II NBs than in dINPs, and later proven to be important for proper lineage development when knocked down specifically in type I and type II NBs (**Figure III.3D,F,H**). This result

Concluding remarks

integrates well with the phenotype of longer-lived NBs that I have detected when analyzing brains of animals 24h APF after TFII α knock down (**Figures I.4, I.5**). Meaning that eventually defects in the way the NB proliferates might lead to its longer permanence in the brain. Consequently, all data so far seems to indicate that TFII α has a role in regulating NB proliferation and lifespan. In the future, it would be interesting to further study the mechanisms that lead to these phenotypes, as well as to better understand its specificity (or not) to NSCs. For instance, in order to further dissect the mechanistic role of TFII α in proper lineage development and in longer-lived NBs, one could perform scRNA-Seq in a TFII α background and compare the lineages with our already available data, in order to identify possible misregulated mechanisms.

So far, one of the gaps in previous analysis of neuronal lineages has been the difficulty in analyzing the whole transcriptome of GMCs. To the best of my knowledge, no specific marker has been identified and attributed to GMCs specifically, meaning that they must be recognized by a combination of expression of several genes. This combined with the small proportion of GMCs in the lineage, makes it difficult to isolate these cells and study them separately without “contamination” from other cells. Thus, this study provides for the first time, transcriptomic data to allow the identification and characterization of the GMC population. Furthermore, by analyzing the genes that are differentially expressed when compared to their progenitor cells (NBs) and to their progeny (neurons) (**Figures III.2, III.3**), we have obtained a list of genes that seem to be turned on or off specifically in this differentiation state. A biological confirmation of these candidates and their functions would be an important step forward in the field. Moreover, since we have a population of GMCs in each of the 6 timepoints, it could also be interesting to evaluate how their transcriptome evolves throughout time and identify possible TTFs for this differentiation state. So far, temporal cascades have only been identified for NBs and INPs (Doe, 2017); though GMCs are known to inherit the TTF expressed in their progenitor NB, to the best of my knowledge, no TTF series has been described for GMCs specifically.

Another interesting characteristic of this study was the fact that each of these scRNA-Seq datasets portrays only the neuronal cells generated within a short temporal window. Hence, only relatively young neurons were included (younger than 12.5h). This is an important feature as it has been shown that young neurons have transcriptional divergence that is lost after 15h (Özel et al., 2020). Consequently, our datasets are uniquely advantageous to study the transcriptome of young neurons as our 18h-labeling was not enough to allow any generated neuron to become old enough to attain the transcriptional convergence. Our data is unique in that way, as other studies involving

Drosophila neural cells, analyzed all neurons present at a certain timepoint, which in a larval brain would include both newly generated secondary neurons as well as mature primary neurons (Allen et al., 2020; Avalos et al., 2019; Davie et al., 2018; Konstantinides et al., 2018; Özel et al., 2020). Furthermore, since the large majority of neurons that will populate the adult *Drosophila* brain are generated during larval and pupal stages (Ito and Hotta, 1992; Ito et al., 2013; Peraanu and Hartenstein, 2006; Truman and Bate, 1988; Yu et al., 2013), the 6 datasets generated in this study represent a window for the transcriptional differences that lead to the maturation of the neurons that will compose the adult brain.

Within this scope, it was interesting to find that neurons start expressing mature features soon after being born, although that only happens at the mRNA level. Using ChAT, a member of NT biosynthesis pathway, we have determined that the young neurons present in the 105h ALH dataset already have ChAT mRNA, while it was not being translated into protein yet (**Figure III.6**). These results are consistent with a very recent work in OL that shows the same to be true in immature larval neurons (Konstantinides et al., 2021). These data suggest that neurons go through several phases of maturation, with the onset of protein expression of genes related to more mature features only starting when the neuron is older. However, these genes such as ChAT and other NT related genes are only necessary later during development, when synaptogenesis starts at approximately 60h ALH in CB. Then, why are these genes being transcribed so soon? Though we do not know the answer to that question yet, we have a strong indication that protein expression is not dependent on neuronal age, but on animal age, and only starts around 48h APF, a later timepoint relatively close to synaptogenesis. Results indicate that no matter when the larval secondary neuron is generated, it will remain immature (Phase 1 or 2) until pupal stages, closer to 48h APF, when it becomes mature (Phase 3; **Figure III.7**). At this point, it would be interesting to understand if this also happens with other NT-associated molecules or even other genes associated with more mature functions, such as ion channels or adhesion molecules. A more detailed *in vivo* characterization of neuronal maturation, focused on the onset of mRNA and protein co-expression of more mature features could be achieved by combining single molecule fluorescence in situ hybridization (smFISH) with antibody labeling. Moreover, it would also be interesting to study the mechanisms that keep the neuron immature (in Phase 1 or 2) with its mRNA untranslated until pupal stages where the protein is closer to be required for brain function (Phase 3). We propose that mRNA is being kept untranslated due to the presence of RNA binding proteins or translation inhibitors, such as Syp and brat, which we found to be expressed in phase 1 or 2 of neuronal maturation. Other

Concluding remarks

studies have suggested that the upregulation of more mature neuronal features might be regulated by an ecdysone pulse close to synaptogenesis, since members of the ecdysone cascade are also upregulated at this point (Kurmangaliyev et al., 2020; Özel et al., 2020). It could also be interesting to analyze neurons aged during a longer temporal window, one that would allow us to also assess the transcriptome of neurons post-transcriptional convergence and thus have a better portrayal of the complete neuronal maturation process. This would help us to understand the mechanism that is regulating the translation of features related to more mature neuronal features; and at the same time, to understand the purpose of expressing mRNA much sooner than the time when it will be required, closer to synaptogenesis. Could it be a way of setting everything ready for a quicker activation once needed?

Taking all this into account, I consider the data generated within this study to represent a useful tool for anyone working in the areas of NSC proliferation and differentiation and overall brain development.

SUPPLEMENTARY DATA

Gene ontology analysis

Gene ontology analysis was performed in PANTHER with the statistical overrepresentation test for biological process. The analysis refers to the 2 groups of genes identified with the pseudotime analysis of neurons (Monocle) described in Chapter III.

Table S1. Gene ontology analysis of group 1 - genes identified in the neuronal population after analysis with Monocle

GO biological process complete	Fold Enrichment	P-value
Malpighian tubule tip cell differentiation (GO:0061382)	> 100	1.20E-04
neuron fate determination (GO:0048664)	> 100	1.20E-04
epithelial cell proliferation involved in renal tubule morphogenesis (GO:2001013)	> 100	1.70E-04
epithelial cell proliferation involved in Malpighian tubule morphogenesis (GO:0061331)	> 100	1.70E-04
epithelial cell proliferation (GO:0050673)	> 100	1.70E-04
R3/R4 cell differentiation (GO:0048056)	82.21	3.31E-04
glial cell differentiation (GO:0010001)	50.59	3.33E-05
neuroblast differentiation (GO:0014016)	48.12	3.84E-05
regulation of neural precursor cell proliferation (GO:2000177)	34.02	1.03E-04
cell fate determination (GO:0001709)	25.89	1.32E-06
regulation of neurogenesis (GO:0050767)	25.29	1.48E-06
gliogenesis (GO:0042063)	24.97	2.47E-04
neuron fate commitment (GO:0048663)	22.42	3.36E-04
regulation of cell development (GO:0060284)	20.3	4.20E-06
regulation of nervous system development (GO:0051960)	15.51	1.50E-05
cell fate commitment (GO:0045165)	15.5	2.06E-07
regulation of cell differentiation (GO:0045595)	14.35	2.80E-06
regulation of cell population proliferation (GO:0042127)	12.07	3.12E-04
regulation of multicellular organismal development (GO:2000026)	11.87	5.29E-05
positive regulation of macromolecule biosynthetic process (GO:0010557)	9.19	6.42E-06
positive regulation of RNA biosynthetic process (GO:1902680)	8.87	4.16E-05
positive regulation of transcription, DNA-templated (GO:0045893)	8.87	4.16E-05
positive regulation of nucleic acid-templated transcription (GO:1903508)	8.87	4.16E-05
positive regulation of cellular biosynthetic process (GO:0031328)	8.46	1.09E-05

Supplementary Data – Gene ontology analysis

positive regulation of biosynthetic process (GO:0009891)	8.45	1.11E-05
positive regulation of RNA metabolic process (GO:0051254)	8.43	5.49E-05
positive regulation of nucleobase-containing compound metabolic process (GO:0045935)	8.14	6.68E-05
regulation of transcription by RNA polymerase II (GO:0006357)	7.57	1.89E-07
regulation of RNA biosynthetic process (GO:2001141)	6.67	2.33E-08
regulation of transcription, DNA-templated (GO:0006355)	6.67	2.33E-08
regulation of nucleic acid-templated transcription (GO:1903506)	6.67	2.33E-08
generation of neurons (GO:0048699)	6.63	1.28E-05
regulation of cellular macromolecule biosynthetic process (GO:2000112)	6.32	7.67E-09
regulation of macromolecule biosynthetic process (GO:0010556)	6.3	8.02E-09
neurogenesis (GO:0022008)	6.26	1.93E-05
positive regulation of nitrogen compound metabolic process (GO:0051173)	6.22	7.78E-05
regulation of cellular biosynthetic process (GO:0031326)	6.06	1.28E-08
regulation of biosynthetic process (GO:0009889)	6.04	1.32E-08
regulation of RNA metabolic process (GO:0051252)	6.02	7.23E-08
regulation of developmental process (GO:0050793)	5.91	3.79E-04
regulation of nucleobase-containing compound metabolic process (GO:0019219)	5.84	1.02E-07
positive regulation of macromolecule metabolic process (GO:0010604)	5.67	1.39E-04
positive regulation of cellular metabolic process (GO:0031325)	5.61	1.49E-04
regulation of gene expression (GO:0010468)	5.26	6.93E-08
nervous system development (GO:0007399)	5.21	7.12E-05
positive regulation of metabolic process (GO:0009893)	5.12	2.61E-04
regulation of nitrogen compound metabolic process (GO:0051171)	4.68	2.84E-07
negative regulation of cellular process (GO:0048523)	4.67	4.96E-05
regulation of primary metabolic process (GO:0080090)	4.61	3.35E-07
animal organ development (GO:0048513)	4.6	1.72E-04
regulation of cellular metabolic process (GO:0031323)	4.32	7.23E-07
regulation of macromolecule metabolic process (GO:0060255)	4.24	8.98E-07
negative regulation of biological process (GO:0048519)	4.05	1.50E-04
system development (GO:0048731)	4.02	5.58E-05
regulation of metabolic process (GO:0019222)	3.94	2.14E-06
multicellular organism development (GO:0007275)	3.79	3.27E-06
cell differentiation (GO:0030154)	3.77	2.65E-04
cellular developmental process (GO:0048869)	3.74	2.81E-04
developmental process (GO:0032502)	3.37	3.97E-06
anatomical structure development (GO:0048856)	3.28	1.70E-05
regulation of cellular process (GO:0050794)	3.03	3.51E-07
regulation of biological process (GO:0050789)	2.82	1.06E-06
multicellular organismal process (GO:0032501)	2.67	6.95E-05
biological regulation (GO:0065007)	2.53	5.76E-06

Table S2. Gene ontology analysis of group 2 - genes identified in the neuronal population after analysis with Monocle

GO biological process complete	Fold Enrichment	P-value
cytoplasmic sequestering of protein (GO:0051220)	86.32	4.33E-04
cytoplasmic sequestering of transcription factor (GO:0042994)	86.32	4.33E-04
actin filament bundle assembly (GO:0051017)	30.83	1.77E-04
negative regulation of synaptic transmission (GO:0050805)	29.43	2.00E-04
regulation of neurotransmitter secretion (GO:0046928)	25.9	2.82E-04
regulation of neurotransmitter transport (GO:0051588)	25.9	2.82E-04
actin filament bundle organization (GO:0061572)	24.9	3.14E-04
cyclic-nucleotide-mediated signaling (GO:0019935)	24.9	3.14E-04
photoreceptor cell axon guidance (GO:0072499)	23.98	3.48E-04
short-term memory (GO:0007614)	23.12	3.84E-04
regulation of synaptic plasticity (GO:0048167)	22.72	4.04E-05
chaperone-mediated protein folding (GO:0061077)	22.48	4.31E-06
regulation of DNA-binding transcription factor activity (GO:0051090)	22.32	4.22E-04
maintenance of protein location (GO:0045185)	19.62	5.99E-04
associative learning (GO:0008306)	18.37	8.74E-05
synaptic vesicle exocytosis (GO:0016079)	15.69	1.55E-04
modulation of chemical synaptic transmission (GO:0050804)	15.2	1.69E-09
regulation of trans-synaptic signaling (GO:0099177)	15.2	1.69E-09
regulated exocytosis (GO:0045055)	14.15	2.26E-04
learning (GO:0007612)	13.7	2.54E-04
neurotransmitter secretion (GO:0007269)	13	5.20E-05
signal release from synapse (GO:0099643)	13	5.20E-05
response to ethanol (GO:0045471)	12.33	3.72E-04
olfactory behavior (GO:0042048)	12.16	3.91E-04
protein folding (GO:0006457)	12.08	2.38E-06
courtship behavior (GO:0007619)	11.66	4.55E-04
signal release (GO:0023061)	11.6	8.70E-05
chemical synaptic transmission (GO:0007268)	10.84	7.03E-09
anterograde trans-synaptic signaling (GO:0098916)	10.84	7.03E-09
chemosensory behavior (GO:0007635)	10.79	6.02E-04
trans-synaptic signaling (GO:0099537)	10.74	7.69E-09
response to heat (GO:0009408)	10.53	6.59E-04
synaptic vesicle cycle (GO:0099504)	10.37	1.44E-04
vesicle-mediated transport in synapse (GO:0099003)	10.37	1.44E-04
synaptic signaling (GO:0099536)	10.32	1.14E-08
regulation of neurotransmitter levels (GO:0001505)	9.81	1.85E-04
response to temperature stimulus (GO:0009266)	9.74	4.22E-05
neurotransmitter transport (GO:0006836)	9.63	2.01E-04

Supplementary Data – Gene ontology analysis

regulation of synaptic growth at neuromuscular junction (GO:0008582)	9.55	2.09E-04
cognition (GO:0050890)	9.32	1.22E-05
learning or memory (GO:0007611)	9.32	1.22E-05
response to alcohol (GO:0097305)	9.22	2.44E-04
regulation of neuromuscular junction development (GO:1904396)	9.07	2.63E-04
regulation of synapse assembly (GO:0051963)	8.99	2.73E-04
memory (GO:0007613)	8.92	2.83E-04
regulation of cell junction assembly (GO:1901888)	8.63	3.27E-04
synapse organization (GO:0050808)	8.58	2.04E-05
regulation of mRNA metabolic process (GO:1903311)	8.43	3.64E-04
cell-cell signaling (GO:0007267)	8.38	2.14E-08
regulation of synapse organization (GO:0050807)	8.09	1.13E-04
regulation of translation (GO:0006417)	7.39	6.50E-04
cell junction organization (GO:0034330)	7.16	1.79E-05
regulation of synapse structure or activity (GO:0050803)	6.78	2.87E-04
neuron projection morphogenesis (GO:0048812)	6.67	2.37E-07
cell projection morphogenesis (GO:0048858)	6.64	2.50E-07
plasma membrane bounded cell projection morphogenesis (GO:0120039)	6.64	2.50E-07
cell part morphogenesis (GO:0032990)	6.51	3.09E-07
neuron projection development (GO:0031175)	6.27	4.53E-07
axon guidance (GO:0007411)	6.17	1.54E-04
neuron development (GO:0048666)	5.98	2.37E-08
neuron projection guidance (GO:0097485)	5.97	1.87E-04
axonogenesis (GO:0007409)	5.91	6.73E-05
response to abiotic stimulus (GO:0009628)	5.78	3.15E-06
chemotaxis (GO:0006935)	5.77	2.30E-04
axon development (GO:0061564)	5.7	8.66E-05
cell morphogenesis involved in neuron differentiation (GO:0048667)	5.61	1.20E-05
behavior (GO:0007610)	5.45	6.71E-07
cellular component morphogenesis (GO:0032989)	5.37	7.78E-07
taxis (GO:0042330)	5.36	1.31E-04
cell morphogenesis (GO:0000902)	5.23	2.89E-06
cell morphogenesis involved in differentiation (GO:0000904)	5.04	2.92E-05
neuron differentiation (GO:0030182)	4.7	5.02E-07
plasma membrane bounded cell projection organization (GO:0120036)	4.58	4.45E-06
cell projection organization (GO:0030030)	4.5	5.38E-06
regulation of cellular component organization (GO:0051128)	4.29	1.58E-06
nervous system process (GO:0050877)	4.25	5.24E-05
generation of neurons (GO:0048699)	4.08	2.92E-06
regulation of cell communication (GO:0010646)	3.9	2.26E-06
regulation of signaling (GO:0023051)	3.9	2.26E-06
nervous system development (GO:0007399)	3.85	5.27E-07

Supplementary Data – Gene ontology analysis

neurogenesis (GO:0022008)	3.85	5.83E-06
system process (GO:0003008)	3.76	1.53E-04
cell development (GO:0048468)	3.74	3.04E-08
locomotion (GO:0040011)	3.58	4.55E-04
regulation of developmental process (GO:0050793)	3.55	2.47E-04
regulation of biological quality (GO:0065008)	3.34	4.51E-07
signaling (GO:0023052)	3.31	4.46E-06
cell communication (GO:0007154)	3.21	6.78E-06
cell differentiation (GO:0030154)	3.02	1.22E-06
cellular developmental process (GO:0048869)	3	1.39E-06
anatomical structure morphogenesis (GO:0009653)	2.84	2.03E-05
reproductive process (GO:0022414)	2.77	2.44E-04
system development (GO:0048731)	2.64	3.25E-05
negative regulation of cellular process (GO:0048523)	2.55	5.84E-04
reproduction (GO:0000003)	2.46	5.50E-04
anatomical structure development (GO:0048856)	2.4	1.31E-06
cellular component organization (GO:0016043)	2.4	1.36E-06
multicellular organism development (GO:0007275)	2.39	1.18E-05
developmental process (GO:0032502)	2.37	1.66E-06
cellular component organization or biogenesis (GO:0071840)	2.27	4.93E-06
regulation of cellular process (GO:0050794)	2.22	4.68E-08
response to stimulus (GO:0050896)	2.18	6.20E-05
biological regulation (GO:0065007)	2.14	2.62E-09
multicellular organismal process (GO:0032501)	2.12	1.66E-06
regulation of biological process (GO:0050789)	2.07	4.56E-07
cellular process (GO:0009987)	1.56	1.91E-06
biological_process (GO:0008150)	1.2	3.19E-04
Unclassified (UNCLASSIFIED)	0.16	3.19E-04

REFERENCES

REFERENCES

- Abrams, J.M., White, K., Fessler, L.I., and Steller, H. (1993). Programmed cell death during *Drosophila* embryogenesis. *Development* *117*, 29–43.
- Allan, D.W., and Thor, S. (2015). Transcriptional selectors, masters, and combinatorial codes: Regulatory principles of neural subtype specification. *Wiley Interdiscip. Rev. Dev. Biol.* *4*, 505–528.
- Allen, A.M., Neville, M.C., Birtles, S., Croset, V., Treiber, C.D., Waddell, S., and Goodwin, S.F. (2020). A single-cell transcriptomic atlas of the adult *Drosophila* ventral nerve cord. *Elife* *9*, 1–32.
- Almeida, M.S., and Bray, S.J. (2005). Regulation of post-embryonic neuroblasts by *Drosophila* Grainyhead. *Mech. Dev.* *122*, 1282–1293.
- Alsiö, J.M., Tarchini, B., Cayouette, M., and Livesey, F.J. (2013). Ikaros promotes early-born neuronal fates in the cerebral cortex. *Proc. Natl. Acad. Sci. U. S. A.* *110*.
- Álvarez, J.A., and Díaz-Benjumea, F.J. (2018). Origin and specification of type II neuroblasts in the *Drosophila* embryo. *Dev.* *145*, 1–10.
- Amoyel, M., and Bach, E. a. (2012). Functions of the *Drosophila* JAK-STAT pathway: Lessons from stem cells. *Jak-Stat* *1*, 176–183.
- Apitz, H., and Salecker, I. (2015). A region-specific neurogenesis mode requires migratory progenitors in the *Drosophila* visual system. *Nat. Neurosci.* *18*, 46–55.
- Artavanis-Tsakonas, S., and Simpson, P. (1991). Choosing a cell fate: a view from the Notch locus. *Trends Genet.* *7*, 403–408.
- Ashraf, S.I., Ganguly, A., Roote, J., and Ip, Y.T. (2004). Worniu, a snail family zinc-finger protein, is required for brain development in *Drosophila*. *Dev. Dyn.* *231*, 379–386.
- Avalos, C.B., Brugmann, R., and Sprecher, S.G. (2019). Single cell transcriptome atlas of the *drosophila* larval brain. *Elife* *8*, 1–25.
- Avila Cobos, F., Vandesompele, J., Mestdagh, P., and De Preter, K. (2018). Computational deconvolution of transcriptomics data from mixed cell populations. *Bioinformatics* *34*, 1969–1979.
- Bach, E.A., Ekas, L.A., Ayala-Camargo, A., Flaherty, M.S., Lee, H., Perrimon, N., and Baeg, G.H. (2007). GFP reporters detect the activation of the *Drosophila* JAK/STAT pathway in vivo. *Gene Expr. Patterns* *7*, 323–331.
- Baek, M., Enriquez, J., and Mann, R.S. (2013). Dual role for Hox genes and Hox co-factors in conferring leg motoneuron survival and identity in *Drosophila*. *Dev.* *140*, 2027–2038.
- Bai, X., Kim, J., Yang, Z., Juryneć, M.J., Akie, T.E., Lee, J., LeBlanc, J., Sessa, A., Jiang, H., DiBiase, A., et al. (2010). TIF1 γ Controls Erythroid Cell Fate by Regulating Transcription Elongation. *Cell* *142*, 133–143.
- Bainbridge, S.P., and Bownes, M. (1981). Staging the metamorphosis of *Drosophila melanogaster*. *J. Embryol. Exp. Morphol.* *Vol.66*, 57–80.
- Bakker, K. (1959). Feeding period, growth, and pupation in larvae of *Drosophila melanogaster*. *Entomol. Exp. Appl.* *2*, 171–186.
- Bannai, H., Fukatsu, K., Mizutani, A., Natsume, T., Iemura, S.I., Ikegami, T., Inoue, T., and Mikoshiba, K. (2004). An RNA-interacting protein, SYNCRIP (heterogeneous nuclear ribonuclear protein Q1/NSAP1) is a component of mRNA granule transported

References

- with inositol 1,4,5-trisphosphate receptor type 1 mRNA in neuronal dendrites. *J. Biol. Chem.* 279, 53427–53434.
- Baumgardt, M., Karlsson, D., Terriente, J., Díaz-Benjumea, F.J., and Thor, S. (2009). Neuronal Subtype Specification within a Lineage by Opposing Temporal Feed-Forward Loops. *Cell* 139, 969–982.
- 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). Combinatorial temporal patterning in progenitors expands neural diversity. *Nature* 498, 449–455.
- Bayraktar, O.A., Boone, J.Q., Drummond, M.L., and Doe, C.Q. (2010). *Drosophila* type II neuroblast lineages keep Prospero levels low to generate large clones that contribute to the adult brain central complex. *Neural Dev.* 5.
- Becht, E., McInnes, L., Healy, J., Dutertre, C.-A., Kwok, I.W.H., Ng, L.G., Ginhoux, F., and Newell, E.W. (2019). Dimensionality reduction for visualizing single-cell data using UMAP. *Nat. Biotechnol.* 37, 38–44.
- 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.
- Bello, B.C., Hirth, F., and Gould, A.P. (2003). A pulse of the *Drosophila* Hox protein Abdominal-A schedules the end of neural proliferation via neuroblast apoptosis. *Neuron* 37, 209–219.
- Bello, B.C., Izergina, N., Caussinus, E., and Reichert, H. (2008). Amplification of neural stem cell proliferation by intermediate progenitor cells in *Drosophila* brain development. *Neural Dev.* 3, 5.
- Benito-Sipos, J., Estacio-Gómez, A., Moris-Sanz, M., Baumgardt, M., Thor, S., and Díaz-Benjumea, F.J. (2010). A genetic cascade involving klumpfuss, nab and castor specifies the abdominal leucokinergetic neurons in the *Drosophila* CNS. *Development* 137, 3327–3336.
- Berger, C., Renner, S., Lürer, K., and Technau, G.M. (2007). The commonly used marker ELAV is transiently expressed in neuroblasts and glial cells in the *Drosophila* embryonic CNS. *Dev. Dyn.* 236, 3562–3568.
- 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.
- Bier, E., Vaessin, H., Younger-Shepherd, S., Lily Yeh Jan, and Yuh Nung Jan (1992). deadpan, an essential pan-neural gene in *Drosophila*, encodes a helix-loop-helix protein similar to the hairy gene product. *Genes Dev.* 6, 2137–2151.
- Bolus, H., Crocker, K., Boekhoff-Falk, G., and Chtarbanova, S. (2020). Modeling neurodegenerative disorders in *drosophila melanogaster*. *Int. J. Mol. Sci.* 21.
- Boone, J.Q., and Doe, C.Q. (2008). Identification of *Drosophila* type II neuroblast lineages containing transit amplifying ganglion mother cells. *Dev. Neurobiol.* 68, 1185–1195.

- 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.
- Bradke, F., Di Giovanni, S., and Fawcett, J. (2020). Neuronal Maturation: Challenges and Opportunities in a Nascent Field. *Trends Neurosci.* 43, 360–362.
- 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* 119, 1–17.
- Brennecke, P., Anders, S., Kim, J.K., Kołodziejczyk, A.A., Zhang, X., Proserpio, V., Baving, B., Benes, V., Teichmann, S.A., Marioni, J.C., et al. (2013). Accounting for technical noise in single-cell RNA-seq experiments. *Nat. Methods* 10, 1093–1098.
- Bridges, C.B. (1935). Salivary chromosome maps: With a key to the banding of the chromosomes of *drosophila melanogaster*. *J. Hered.* 26, 60–64.
- Briscoe, J., and Novitsch, B.G. (2008). Regulatory pathways linking progenitor patterning, cell fates and neurogenesis in the ventral neural tube. *Philos. Trans. R. Soc. B Biol. Sci.* 363, 57–70.
- 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.
- Brody, T., and Odenwald, W.F. (2000). Programmed transformations in neuroblast gene expression during *Drosophila* CNS lineage development. *Dev. Biol.* 226, 34–44.
- Buhl, L.K., Jorquera, R.A., Akbergenova, Y., Huntwork-Rodriguez, S., Volfson, D., and Littleton, J.T. (2013). Differential regulation of evoked and spontaneous neurotransmitter release by C-terminal modifications of complexin. *Mol. Cell. Neurosci.* 52, 161–172.
- Cardenas, M.G., Oswald, E., Yu, W., Xue, F., MacKerell, A.D., and Melnick, A.M. (2017). The expanding role of the BCL6 oncoprotein as a cancer therapeutic target. *Clin. Cancer Res.* 23, 885–893.
- Carroll, S.B., Laymon, R.A., McCutcheon, M.A., Riley, P.D., and Scott, M.P. (1986). The localization and regulation of Antennapedia protein expression in *Drosophila* embryos. *Cell* 47, 113–122.
- Cau, E., and Blader, P. (2009). Notch activity in the nervous system: To switch or not switch? *Neural Dev.* 4, 36.
- Caussinus, E., and Gonzalez, C. (2005). Induction of tumor growth by altered stem-cell asymmetric division in *Drosophila melanogaster*. *Nat. Genet.* 37, 1125–1129.
- Cenci, C., and Gould, A.P. (2005). *Drosophila* Grainyhead specifies late programmes of neural proliferation by regulating the mitotic activity and Hox-dependent apoptosis of neuroblasts. *Development* 132, 3835–3845.
- Ceron, J., González, C., and Tejedor, F.J. (2001). Patterns of cell division and expression of asymmetric cell fate determinants in postembryonic neuroblast lineages of *Drosophila*. *Dev. Biol.* 230, 125–138.
- Ceron, J., Tejedor, F.J., and Moya, F. (2006). A primary cell culture of *Drosophila* postembryonic larval neuroblasts to study cell cycle and asymmetric division. *Eur. J. Cell Biol.* 85, 567–575.
- Chan, L.N., and Gehring, W. (1971). Determination of blastoderm cells in *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. U. S. A.* 68, 2217–2221.

References

- Chell, J.M., and Brand, A.H. (2010). Nutrition-responsive glia control exit of neural stem cells from quiescence. *Cell* **143**, 1161–1173.
- Chen, Y., Akin, O., Nern, A., Tsui, C.Y.K., Pecot, M.Y., and Zipursky, S.L. (2014). Cell-type-specific labeling of synapses in vivo through synaptic tagging with recombination. *Neuron* **81**, 280–293.
- Chu-Lagraff, Q., Wright, D.M., McNeil, L.K., and Doe, C.Q. (1991). The prospero gene encodes a divergent homeodomain protein that controls neuronal identity in *Drosophila*. *Development* **113**, 79–85.
- Cognigni, P., Felsenberg, J., and Waddell, S. (2018). Do the right thing: neural network mechanisms of memory formation, expression and update in *Drosophila*. *Curr. Opin. Neurobiol.* **49**, 51–58.
- Colombani, J., Bianchini, L., Layalle, S., Pondeville, E., Dauphin-Villemant, C., Antoniewski, C., Carré, C., Noselli, S., and Léopold, P. (2005). Antagonistic actions of ecdysone and insulins determine final size in *Drosophila*. *Science* (80-.). **310**, 667–670.
- Conaway, R.C., and Conaway, J.W. (2013). The Mediator complex and transcription elongation. *Biochim. Biophys. Acta - Gene Regul. Mech.* **1829**, 69–75.
- Crews, S.T. (2019). *Drosophila* embryonic CNS development: Neurogenesis, gliogenesis, cell fate, and differentiation. *Genetics* **213**, 1111–1144.
- Croset, V., Treiber, C.D., and Waddell, S. (2018). Cellular diversity in the *Drosophila* midbrain revealed by single-cell transcriptomics. *Elife* **7**, 1–31.
- Cui, X., and Doe, C.Q. (1992). Ming is expressed in neuroblast sublineages and regulates gene expression in the *Drosophila* central nervous system. *Development* **116**, 943–952.
- Cusanovich, D.A., Hill, A.J., Aghamirzaie, D., Daza, R.M., Pliner, H.A., Berletch, J.B., Filippova, G.N., Huang, X., Christiansen, L., DeWitt, W.S., et al. (2018). A Single-Cell Atlas of In Vivo Mammalian Chromatin Accessibility. *Cell* **174**, 1309-1324.e18.
- 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.
- Desai, A.R., and McConnell, S.K. (2000). Progressive restriction in fate potential by neural progenitors during cerebral cortical development. *Development* **127**, 2863–2872.
- Doe, C.Q. (2017). Temporal patterning in the *drosophila* CNS. *Annu. Rev. Cell Dev. Biol.* **33**, 219–240.
- Doe, C.Q., and Goodman, C.S. (1985). Early events in insect neurogenesis. I. Development and segmental differences in the pattern of neuronal precursor cells. *Dev. Biol.* **111**, 193–205.
- Doe, C.Q., Chu-LaGraff, Q., Wright, D.M., and Scott, M.P. (1991). The prospero gene specifies cell fates in the *drosophila* central nervous system. *Cell* **65**, 451–464.
- Du, L., and Wang, J. (2020). JAK/STAT pathway promotes *Drosophila* neuroblast proliferation via the direct CycE regulation. *BioRxiv* 2020.07.10.195875.
- Duffy, J.B. (2002). GAL4 system in *Drosophila*: A fly geneticist's Swiss army knife. *Genesis* **34**, 1–15.
- Dumstreij, K., Wang, F., Nassif, C., and Hartenstein, V. (2003). Early development of the *Drosophila* brain: V. Pattern of postembryonic neuronal lineages expressing DE-cadherin. *J. Comp. Neurol.* **455**, 451–462.
- Ebens, A.J., Garren, H., Cheyette, B.N.R., and Zipursky, S.L. (1993). The *Drosophila*

- anachronism locus: A glycoprotein secreted by glia inhibits neuroblast proliferation. *Cell* 74, 15–27.
- 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 Dev.* 2, 1–15.
- Egger, B., Gold, K.S., and Brand, A.H. (2010). Notch regulates the switch from symmetric to asymmetric neural stem cell division in the *Drosophila* optic lobe. *Development* 137, 2981–2987.
- Ekas, L.A., Cardozo, T.J., Flaherty, M.S., McMillan, E.A., Gonsalves, F.C., and Bach, E.A. (2010). Characterization of a dominant-active STAT that promotes tumorigenesis in *Drosophila*. *Dev. Biol.* 344, 621–636.
- Elliott, J., Jolicoeur, C., Ramamurthy, V., and Cayouette, M. (2008). Ikaros Confers Early Temporal Competence to Mouse Retinal Progenitor Cells. *Neuron* 60, 26–39.
- Enriquez, J., Rio, L.Q., Blazeski, R., Bellemin, S., Godement, P., Mason, C., and Mann, R.S. (2018). Differing Strategies Despite Shared Lineages of Motor Neurons and Glia to Achieve Robust Development of an Adult Neuropil in *Drosophila*. *Neuron* 97, 538–554.e5.
- 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.
- Eroglu, E., Burkard, T.R., Jiang, Y., Saini, N., Homem, C.C.F., Reichert, H., and Knoblich, J.A. (2014). SWI/SNF complex prevents lineage reversion and induces temporal patterning in neural stem cells. *Cell* 156, 1259–1273.
- Evans, C.J., Olson, J.M., Ngo, K.T., Kim, E., Lee, N.E., Kuoy, E., Patananan, A.N., Sitz, D., Tran, P.T., Do, M.T., et al. (2009). G-TRACE: Rapid Gal4-based cell lineage analysis in *Drosophila*. *Nat. Methods* 6, 603–605.
- Faherty, S.L., Campbell, C.R., Larsen, P.A., and Yoder, A.D. (2015). Evaluating whole transcriptome amplification for gene profiling experiments using RNA-Seq. *BMC Biotechnol.* 15, 17–19.
- Fernández-Hernández, I., Rhiner, C., and Moreno, E. (2013). Adult Neurogenesis in *Drosophila*. *Cell Rep.* 3, 1857–1865.
- Fog, C.K., Galli, G.G., and Lund, A.H. (2012). PRDM proteins: important players in differentiation and disease. *Bioessays* 34, 50–60.
- Fredieu, J.R., and Mahowald, A.P. (1989). Glial interactions with neurons during *Drosophila* embryogenesis. *Development* 106, 739–748.
- Freire-Picos, M.A., Krishnamurthy, S., Sun, Z.W., and Hampsey, M. (2005). Evidence that the Tfg1/Tfg2 dimer interface of TFIIF lies near the active center of the RNA polymerase II initiation complex. *Nucleic Acids Res.* 33, 5045–5052.
- Frohnhofer, H.G., and Nüsslein-Volhard, C. (1986). Organization of anterior pattern in the *Drosophila* embryo by the maternal gene bicoid. *Nature* 324, 120–125.
- Gaul, U., and Jäckle, H. (1990). Role of Gap Genes in Early *Drosophila* Development. *Adv. Genet.* 27, 239–275.
- Gilmore, E.C., and Walsh, C.A. (2013). Genetic causes of microcephaly and lessons for neuronal development. *Wiley Interdiscip. Rev. Dev. Biol.* 2, 461–478.
- Glickstein, M. (2006). Golgi and Cajal: The neuron doctrine and the 100th anniversary of the 1906 Nobel Prize. *Curr. Biol.* 16, R147–R151.

References

- Gong, D. wei, Horikoshi, M., and Nakatani, Y. (1993). Analysis of cDNA encoding drosophila transcription initiation factor TFIIF α (RAP74). *Nucleic Acids Res.* *21*, 1492.
- Grosskortenhaus, R., Pearson, B.J., Marusich, A., and Doe, C.Q. (2005). Regulation of temporal identity transitions in drosophila neuroblasts. *Dev. Cell* *8*, 193–202.
- Grosskortenhaus, R., Robinson, K.J., and Doe, C.Q. (2006). Pdm and Castor specify late-born motor neuron identity in the NB7-1 lineage. *Genes Dev.* *20*, 2618–2627.
- Groszer, M., Erickson, R., Scripture-Adams, D.D., Lesche, R., Trumpp, A., Zack, J.A., Kornblum, H.I., Liu, X., and Wu, H. (2001). Negative regulation of neural stem/progenitor cell proliferation by the Pten tumor suppressor gene in vivo. *Science* (80-.). *294*, 2186–2189.
- Hahn, S. (2004). Structure and mechanism of the RNA polymerase II transcription machinery. *Nat. Struct. Mol. Biol.* *11*, 394–403.
- Hakes, A.E., and Brand, A.H. (2020). Tailless/TLX reverts intermediate neural progenitors to stem cells driving tumourigenesis via repression of asense/ASCL1. *Elife* *9*, 1–21.
- 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. *Dev.* *145*.
- Hales, K.G., Korey, C.A., Larracuente, A.M., and Roberts, D.M. (2015). Genetics on the fly: A primer on the drosophila model system. *Genetics* *201*, 815–842.
- Hamid, R., Hajirnis, N., Kushwaha, S., Saleem, S., Kumar, V., and Mishra, R.K. (2019). Drosophila Choline transporter non-canonically regulates pupal eclosion and NMJ integrity through a neuronal subset of mushroom body. *Dev. Biol.* *446*, 80–93.
- Handler, A.M. (1982). Ecdysteroid titers during pupal and adult development in *Drosophila melanogaster*. *Dev. Biol.* *93*, 73–82.
- Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W.M., Zheng, S., Butler, A., Lee, M.J., Wilk, A.J., Darby, C., Zagar, M., et al. (2020). Integrated analysis of multimodal single-cell data. *BioRxiv* 2020.10.12.335331.
- Harris, H. (2004). Tumour suppression: putting on the brakes. *Nature* *427*, 201.
- Hartenstein, V. (1993). Atlas of drosophila development.
- Hartenstein, V., and Campos-Ortega, J.A. (1984). Early neurogenesis in wild-type *Drosophila melanogaster*. *Wilhelm Roux's Arch. Dev. Biol.* *193*, 308–325.
- Hartenstein, V., Rudloff, E., and Campos-Ortega, J.A. (1987). The pattern of proliferation of the neuroblasts in the wild-type embryo of *Drosophila melanogaster*. *Roux's Arch. Dev. Biol.* *196*, 473–485.
- Harzer, H., Berger, C., Conder, R., Schmauss, G., and Juergen, A. (2014). FACS purification of *Drosophila* larval Neuroblasts for next generation sequencing. *8*, 1088–1099.
- He, L., Fei, D.L., Nagiec, M.J., Mutvei, A.P., Lamprakis, A., Kim, B.Y., and Blenis, J. (2019). Regulation of GSK3 cellular location by FRAT modulates mTORC1-dependent cell growth and sensitivity to rapamycin. *Proc. Natl. Acad. Sci. U. S. A.* *116*, 19523–19529.
- Hietakangas, V., and Cohen, S.M. (2009). Regulation of tissue growth through nutrient sensing. *Annu. Rev. Genet.* *43*, 389–410.
- Hirono, K., Kohwi, M., Clark, M.Q., Heckscher, E.S., and Doe, C.Q. (2017). The Hunchback temporal transcription factor establishes, but is not required to maintain,

- early-born neuronal identity. *Neural Dev.* **12**, 1.
- Hobert, O. (2016). Terminal Selectors of Neuronal Identity. In *Current Topics in Developmental Biology*, (Academic Press Inc.), pp. 455–475.
- Hofbauer, A., and Campos-Ortega, J.A. (1990). Proliferation pattern and early differentiation of the optic lobes in *Drosophila melanogaster*. *Roux's Arch. Dev. Biol.* **198**, 264–274.
- 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., Reichardt, I., Berger, C., Lendl, T., and Knoblich, J.A. (2013). Long-term live cell imaging and automated 4D analysis of *Drosophila* neuroblast lineages. *PLoS One* **8**, 1–10.
- Homem, C.C.F., 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.
- Horvitz, H.R., and Herskowitz, I. (1992). Mechanisms of asymmetric cell division: Two Bs or not two Bs, that is the question. *Cell* **68**, 237–255.
- Hu, Y., Comjean, A., Perkins, L.A., Perrimon, N., and Mohr, S.E. (2015). GLAD: an Online Database of Gene List Annotation for *Drosophila*. *J. Genomics* **3**, 75–81.
- Hwang, B., Lee, J.H., and Bang, D. (2018). Single-cell RNA sequencing technologies and bioinformatics pipelines. *Exp. Mol. Med.* **50**.
- 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**, 1–15.
- Irish, V., Lehmann, R., and Akam, M. (1989). The *Drosophila* posterior-group gene nanos functions by repressing hunchback activity. *Nature* **338**, 646–648.
- Isshiki, T., Pearson, B., Holbrook, S., and Doe, C.Q. (2001). *Drosophila* neuroblasts sequentially express transcription factors which specify the temporal identity of their neuronal progeny. *Cell* **106**, 511–521.
- Ito, K., and Hotta, Y. (1992). Proliferation pattern of postembryonic neuroblasts in the brain of *Drosophila melanogaster*. *Dev. Biol.* **149**, 134–148.
- Ito, K., Awano, W., Suzuki, K., Hiromi, Y., and Yamamoto, D. (1997). The *Drosophila* mushroom body is a quadruple structure of clonal units each of which contains a virtually identical set of neurones and glial cells. *Development* **124**, 761–771.
- 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.
- Iyer, E.P.R., Iyer, S.C., Sullivan, L., Wang, D., Meduri, R., Graybeal, L.L., and Cox, D.N. (2013). Functional Genomic Analyses of Two Morphologically Distinct Classes of *Drosophila* Sensory Neurons: Post-Mitotic Roles of Transcription Factors in Dendritic Patterning. *PLoS One* **8**.
- Izergina, N., Balmer, J., Bello, B., and Reichert, H. (2009). Postembryonic development

References

- of transit amplifying neuroblast lineages in the *Drosophila* brain. *Neural Dev.* **4**, 1–13.
- Jaeger, J. (2011). The gap gene network. *Cell. Mol. Life Sci.* **68**, 243–274.
- 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.
- Jardin, F., Ruminy, P., Bastard, C., and Tilly, H. (2007). The BCL6 proto-oncogene: a leading role during germinal center development and lymphomagenesis. *Pathol. Biol.* **55**, 73–83.
- Jarman, A.P., Brand, M., Jan, L.Y., and Jan, Y.N. (1993). The regulation and function of the helix-loop-helix gene, *asense*, in *Drosophila* neural precursors. *Development* **119**, 19–29.
- 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.
- Jevitt, A., Chatterjee, D., Xie, G., Wang, X.F., Otwell, T., Huang, Y.C., and Deng, W.M. (2020). A single-cell atlas of adult *Drosophila* ovary identifies transcriptional programs and somatic cell lineage regulating oogenesis. *PLoS Biol.* **18**, e3000538.
- Jiang, S., Zhang, C.-Y., Tang, L., Zhao, L.-X., Chen, H.-Z., and Qiu, Y. (2019). Integrated Genomic Analysis Revealed Associated Genes for Alzheimer's Disease in APOE4 Non-Carriers. *Curr. Alzheimer Res.* **16**, 753–763.
- Kambadur, R., Koizumi, K., Stivers, C., Nagle, J., Poole, S.J., and Odenwald, W.F. (1998). Regulation of POU genes by *castor* and *hunchback* establishes layered compartments in the *Drosophila* CNS. *Genes Dev.* **12**, 246–260.
- Kanai, M.I., Okabe, M., and Hiromi, Y. (2005). Seven-up controls switching of transcription factors that specify temporal identities of *drosophila* neuroblasts. *Dev. Cell* **8**, 203–213.
- Kanai, M.I., Kim, M.-J., Akiyama, T., Takemura, M., Wharton, K., O'Connor, M.B., and Nakato, H. (2018). Regulation of neuroblast proliferation by surface glia in the *Drosophila* larval brain. *Sci. Rep.* **8**, 3730.
- Karcavich, R., and Doe, C.Q. (2005). *Drosophila* neuroblast 7-3 cell lineage: A model system for studying programmed cell death, notch/numb signaling, and sequential specification of ganglion mother cell identity. *J. Comp. Neurol.* **481**, 240–251.
- Kennerdell, J.R., and Carthew, R.W. (1998). Use of dsRNA-mediated genetic interference to demonstrate that *frizzled* and *frizzled 2* act in the wingless pathway. *Cell* **95**, 1017–1026.
- Kennerdell, J.R., and Carthew, R.W. (2000). Heritable gene silencing in *Drosophila* using double-stranded RNA. *Nat. Biotechnol.* **18**, 896–898.
- Khapersky, D.A., Ammerman, M.L., Majovski, R.C., and Ponticelli, A.S. (2008). Functions of *Saccharomyces cerevisiae* TFIIF during Transcription Start Site Utilization. *Mol. Cell. Biol.* **28**, 3757–3766.
- Khudayberdiev, S., Soutschek, M., Ammann, I., Heinze, A., Rust, M.B., Baumeister, S., and Schratt, G. (2020). The cytoplasmic SYNCRIP mRNA interactome of mammalian neurons. *RNA Biol.* 1–13.
- Kimura, Y., Satou, C., and Higashijima, S.I. (2008). V2a and V2b neurons are generated by the final divisions of pair-producing progenitors in the zebrafish spinal cord. *Development* **135**, 3001–3005.
- Klepsatel, P., Gálíková, M., De Maio, N., Ricci, S., Schlötterer, C., and Flatt, T. (2013).

- Reproductive and post-reproductive life history of wild-caught *Drosophila melanogaster* under laboratory conditions. *J. Evol. Biol.* 26, 1508–1520.
- Knoblich, J.A. (2008). Mechanisms of Asymmetric Stem Cell Division. *Cell* 132, 583–597.
- Kobak, D., and Linderman, G.C. (2021). Initialization is critical for preserving global data structure in both t-SNE and UMAP. *Nat. Biotechnol.* 39, 156–157.
- Konstantinides, A.M., R., A., E., L., D., Y.C., C., T., T., A., J., M.N., Ö., F., S., X., S., et al. (2021). A comprehensive series of temporal transcription factors in the fly visual system. Submitted 1, 1–10.
- 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.
- Koyama, T., and Mirth, C. (2021). Ecdysone Quantification from Whole Body Samples of *Drosophila melanogaster* Larvae. *BIO-PROTOCOL* 11, e3915.
- Kratsios, P., Pinan-Lucarré, B., Kerk, S.Y., Weinreb, A., Bessereau, J.L., and Hobert, O. (2015). Transcriptional coordination of synaptogenesis and neurotransmitter signaling. *Curr. Biol.* 25, 1282–1295.
- Kurmangaliyev, Y.Z., Yoo, J., Locascio, S.A., and Lawrence Zipursky, S. (2019). Modular transcriptional programs separately define axon and dendrite connectivity. *Elife* 8, 1–18.
- Kurmangaliyev, Y.Z., Yoo, J., Valdes-Aleman, J., Sanfilippo, P., and Zipursky, S.L. (2020). Transcriptional Programs of Circuit Assembly in the *Drosophila* Visual System. *Neuron* 108, 1045-1057.e6.
- Landgraf, M., and Thor, S. (2006). Development of *Drosophila* motoneurons: Specification and morphology. *Semin. Cell Dev. Biol.* 17, 3–11.
- Landick, R. (2006). The regulatory roles and mechanism of transcriptional pausing. *Biochem. Soc. Trans.* 34, 1062–1066.
- Li, H., and Tennessen, J.M. (2017). Methods for studying the metabolic basis of *Drosophila* development. *Wiley Interdiscip. Rev. Dev. Biol.* 6.
- Li, J., Chauve, L., Phelps, G., Briemann, R.M., and Morimoto, R.I. (2016a). E2F coregulates an essential HSF developmental program that is distinct from the heat-shock response. *Genes Dev.* 30, 2062–2075.
- 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.
- Li, X., Xie, Y., and Zhu, S. (2016b). Notch maintains *Drosophila* type II neuroblasts by suppressing expression of the *fez* transcription factor earmuff. *Dev.* 143, 2511–2521.
- Liu, P., Zhao, P., Li, B., Xu, D., and Wang, K. (2020). LncRNA PCAT6 Regulated by YY1 Accelerates the Progression of Glioblastoma via miR-513/IGF2BP1. *Neurochem. Res.* 45, 2894–2902.
- 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* (80-.). 350, 317–320.
- Luo, R., Zhuo, Y., Du, Q., and Xiao, R. (2021). POU2F2 promotes the proliferation and motility of lung cancer cells by activating AGO1. *BMC Pulm. Med.* 21, 117.
- Ma, X., Xu, L., Alberobello, A.T., Gavrilova, O., Bagattin, A., Skarulis, M., Liu, J., Finkel, T., and Mueller, E. (2015). Celastrol protects against obesity and metabolic dysfunction

References

- through activation of a HSF1-PGC1 α transcriptional axis. *Cell Metab.* 22, 695–708.
- Macdonald, P.M., and Struhl, G. (1986). A molecular gradient in early *Drosophila* embryos and its role in specifying the body pattern. *Nature* 324, 537–545.
- Malik, S., and Roeder, R.G. (2010). The metazoan Mediator co-activator complex as an integrative hub for transcriptional regulation. *Nat. Rev. Genet.* 11, 761–772.
- La Manno, G., Soldatov, R., Zeisel, A., Braun, E., Hochgerner, H., Petukhov, V., Lidschreiber, K., Kastrioti, M.E., Lönnerberg, P., Furlan, A., et al. (2018). RNA velocity of single cells. *Nature* 560, 494–498.
- Markow, T.A., Beall, S., and Matzkin, L.M. (2009). Egg size, embryonic development time and ovoviviparity in *Drosophila* species. *J. Evol. Biol.* 22, 430–434.
- Mattar, P., Ericson, J., Blackshaw, S., and Cayouette, M. (2015). A conserved regulatory logic controls temporal identity in mouse neural progenitors. *Neuron* 85, 497–504.
- 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.
- McGuire, S.E., Le, P.T., Osborn, A.J., Matsumoto, K., and Davis, R.L. (2003). Spatiotemporal Rescue of Memory Dysfunction in *Drosophila*. *Science* (80-.). 302, 1765–1768.
- McGuire, S.E., Roman, G., and Davis, R.L. (2004). Gene expression systems in *Drosophila*: A synthesis of time and space. *Trends Genet.* 20, 384–391.
- McGurk, L., Berson, A., and Bonini, N.M. (2015). *Drosophila* as an in vivo model for human neurodegenerative disease. *Genetics* 201, 377–402.
- McInnes, L., Healy, J., and Melville, J. (2018). UMAP: Uniform manifold approximation and projection for dimension reduction. *ArXiv*.
- Mellerick, D.M., Kassis, J.A., Zhang, S.D., and Odenwald, W.F. (1992). *castor* encodes a novel zinc finger protein required for the development of a subset of CNS neurons in *drosophila*. *Neuron* 9, 789–803.
- Michki, N.S., Li, Y., Sanjasaz, K., Zhao, Y., Shen, F.Y., Walker, L.A., Cao, W., Lee, C.-Y., and Cai, D. (2021). The molecular landscape of neural differentiation in the developing *Drosophila* brain revealed by targeted scRNA-seq and multi-informatic analysis. *Cell Rep.* 35, 109039.
- Misquitta, L., and Paterson, B.M. (1999). Targeted disruption of gene function in *Drosophila* by RNA interference (RNA-i): A role for *nautilus* in embryonic somatic muscle formation. *Proc. Natl. Acad. Sci. U. S. A.* 96, 1451–1456.
- Mohr, S.E., and Perrimon, N. (2019). *Drosophila melanogaster*: A simple system for understanding complexity. *DMM Dis. Model. Mech.* 12.
- 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.
- Montgomery, M.K., Xu, S., and Fire, A. (1998). RNA as a target of double-stranded RNA-mediated genetic interference in *Caenorhabditis elegans*. *Proc. Natl. Acad. Sci. U. S. A.* 95, 15502–15507.
- Mora, N., Oliva, C., Fiers, M., Ejsmont, R., Soldano, A., Zhang, T.T., Yan, J., Claeys, A., De Geest, N., and Hassan, B.A. (2018). A Temporal Transcriptional Switch Governs

- Stem Cell Division, Neuronal Numbers, and Maintenance of Differentiation. *Dev. Cell* 45, 53-66.e5.
- Morgan, T.H. (1910). Sex limited inheritance in drosophila. *Science* (80-.). 32, 120–122.
- Muthukumar, A.K., Stork, T., and Freeman, M.R. (2014). Activity-dependent regulation of astrocyte GAT levels during synaptogenesis. *Nat. Neurosci.* 17, 1340–1350.
- Napoli, C., Sessa, M., Infante, T., and Casamassimi, A. (2012). Unraveling framework of the ancestral Mediator complex in human diseases. *Biochimie* 94, 579–587.
- 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.
- Neves, A., and Eisenman, R.N. (2019). Distinct gene-selective roles for a network of core promoter factors in *Drosophila* neural stem cell identity. *Biol. Open* 8.
- Novotny, T., Eiselt, R., and Urban, J. (2002). Hunchback is required for the specification of the early sublineage of neuroblast 7-3 in the *Drosophila* central nervous system. *Development* 129, 1027–1036.
- Özel, M.N., Langen, M., Hassan, B.A., and Hiesinger, P.R. (2015). Filopodial dynamics and growth cone stabilization in *Drosophila* visual circuit development. *Elife* 4.
- Özel, M.N., Kulkarni, A., Hasan, A., Brummer, J., Moldenhauer, M., Daumann, I.M., Wolfenberg, H., Dercksen, V.J., Kiral, F.R., Weiser, M., et al. (2019). Serial Synapse Formation through Filopodial Competition for Synaptic Seeding Factors. *Dev. Cell* 50, 447-461.e8.
- Özel, M.N., Simon, F., Jafari, S., Holguera, I., Chen, Y.C., Benhra, N., El-Danaf, R.N., Kapuralin, K., Malin, J.A., Konstantinides, N., et al. (2020). Neuronal diversity and convergence in a visual system developmental atlas. *Nature* 589, 88–95.
- Park, Y., Rangel, C., Reynolds, M.M., Caldwell, M.C., Johns, M., Nayak, M., Welsh, C.J.R., McDermott, S., and Datta, S. (2003). *Drosophila* Perlecan modulates FGF and Hedgehog signals to activate neural stem cell division. *Dev. Biol.* 253, 247–257.
- Pearson, B.J., and Doe, C.Q. (2003). Regulation of neuroblast competence in *Drosophila*. *Nature* 425, 624–628.
- Pereanu, W., and Hartenstein, V. (2006). Neural lineages of the *Drosophila* brain: A three-dimensional digital atlas of the pattern of lineage location and projection at the late larval stage. *J. Neurosci.* 26, 5534–5553.
- Peterson, C., Carney, G.E., Taylor, B.J., and White, K. (2002). reaper is required for neuroblast apoptosis during *Drosophila* development. *Development* 129, 1467–1476.
- Pfeiffer, B.D., Jenett, A., Hammonds, A.S., Ngo, T.-T.T.B.B., Misra, S., Murphy, C., Scully, A., Carlson, J.W., Wan, K.H., Lavery, T.R., et al. (2008). Tools for neuroanatomy and neurogenetics in *Drosophila*. *Proc. Natl. Acad. Sci. U. S. A.* 105, 9715–9720.
- Pinto-Teixeira, F., Koo, C., Rossi, A.M., Neric, N., Bertet, C., Li, X., Del-Valle-Rodriguez, A., and Desplan, C. (2018). Development of Concurrent Retinotopic Maps in the Fly Motion Detection Circuit. *Cell* 173, 485-498.e11.
- Porcher, A., Abu-Arish, A., Huart, S., Roelens, B., Fradin, C., and Dostatni, N. (2010). The time to measure positional information: Maternal Hunchback is required for the synchrony of the Bicoid transcriptional response at the onset of zygotic transcription. *Development* 137, 2795–2804.
- Powsner, L. (1935). The Effects of Temperature on the Durations of the Developmental Stages of *Drosophila melanogaster*. *Physiol. Zool.* 8, 474–520.

References

- Price, D.H., Sluder, A.E., and Greenleaf, A.L. (1989). Dynamic interaction between a *Drosophila* transcription factor and RNA polymerase II. *Mol. Cell. Biol.* 9, 1465–1475.
- Prokop, A., and Technau, G.M. (1991). The origin of postembryonic neuroblasts in the ventral nerve cord of *Drosophila melanogaster*. *Development* 111, 79–88.
- 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.
- Rebollo, E., Sampaio, P., Januschke, J., Llamazares, S., Varmark, H., and González, C. (2007). Functionally Unequal Centrosomes Drive Spindle Orientation in Asymmetrically Dividing *Drosophila* Neural Stem Cells. *Dev. Cell* 12, 467–474.
- Rebollo, E., Roldán, M., and Gonzalez, C. (2009). Spindle alignment is achieved without rotation after the first cell cycle in *Drosophila* embryonic neuroblasts. *Development* 136, 3393–3397.
- Reichardt, I., Bonnay, F., Steinmann, V., Loedige, I., Burkard, T.R., Meister, G., and Knoblich, J.A. (2018). The tumor suppressor Brat controls neuronal stem cell lineages by inhibiting Deadpan and Zelda. *EMBO Rep.* 19, 102–117.
- Rein, K., Zöckler, M., Mader, M.T., Grübel, C., and Heisenberg, M. (2002). The *Drosophila* standard brain. *Curr. Biol.* 12, 227–231.
- Reines, D., Chamberlin, M.J., and Kane, C.M. (1989). Transcription elongation factor SII (TFIIS) enables RNA polymerase II to elongate through a block to transcription in a human gene in vitro. *J. Biol. Chem.* 264, 10799–10809.
- Reiter, L.T., Potocki, L., Chien, S., Gribskov, M., and Bier, E. (2001). A Systematic Analysis of Human Disease-Associated Gene Sequences In *Drosophila melanogaster*. 1114–1125.
- 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.
- Robinow, S., and White, K. (1991). Characterization and spatial distribution of the ELAV protein during *Drosophila melanogaster* development. *J. Neurobiol.* 22, 443–461.
- Rogulja-Ortmann, A., Lüer, K., Seibert, J., Rickert, C., and Technau, G.M. (2007). Programmed cell death in the embryonic central nervous system of *Drosophila melanogaster*. *Development* 134, 105–116.
- Rossignol, M., Keriél, A., Staub, A., and Egly, J.M. (1999). Kinase activity and phosphorylation of the largest subunit of TFIIF transcription factor. *J. Biol. Chem.* 274, 22387–22392.
- Russell, R.C., Fang, C., and Guan, K.L. (2011). An emerging role for TOR signaling in mammalian tissue and stem cell physiology. *Development* 138, 3343–3356.
- Salvaterra, P.M., and McCaman, R.E. (1985). Choline acetyltransferase and acetylcholine levels in *Drosophila melanogaster*: A study using two temperature-sensitive mutants. *J. Neurosci.* 5, 903–910.
- San-Juán, B.P., and Baonza, A. (2011). The bHLH factor deadpan is a direct target of Notch signaling and regulates neuroblast self-renewal in *Drosophila*. *Dev. Biol.* 352, 70–82.
- Satija, R., Farrell, J.A., Gennert, D., Schier, A.F., and Regev, A. (2015). Spatial reconstruction of single-cell gene expression data. *Nat. Biotechnol.* 33, 495–502.
- Schinaman, J.M., Giesey, R.L., Mizutani, C.M., Lukacsovich, T., and Sousa-Neves, R. (2014). The KRÜPPEL-Like Transcription Factor DATILÓGRAFO Is Required in Specific Cholinergic Neurons for Sexual Receptivity in *Drosophila* Females. *PLoS Biol.* 12.

- Schmid, A., Chiba, A., and Doe, C.Q. (1999). Clonal analysis of *Drosophila* embryonic neuroblasts: neural cell types, axon projections and muscle targets. *Development* 126, 4653–4689.
- Schuldt, A.J., Adams, J.H.J., Davidson, C.M., Micklem, D.R., Haseloff, J., St. Johnston, D., and Brand, A.H. (1998). Miranda mediates asymmetric protein and RNA localization in the developing nervous system. *Genes Dev.* 12, 1847–1857.
- Schulz, R., Bertrand, S., Chamaon, K., Smalla, K.H., Gundelfinger, E.D., and Bertrand, D. (2000). Neuronal nicotinic acetylcholine receptors from *Drosophila*: Two different types of α subunits coassemble within the same receptor complex. *J. Neurochem.* 74, 2537–2546.
- Seelig, J.D., and Jayaraman, V. (2015). Neural dynamics for landmark orientation and angular path integration. *Nature* 521, 186–191.
- 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, 1–26.
- Shapiro, H. (1932). THE RATE OF OVIPOSITION IN THE FRUIT FLY, *DROSOPHILA*. *Biol. Bull.* 63, 456–471.
- Shen, C.P., Jan, L.Y., and Jan, Y.N. (1997). Miranda is required for the asymmetric localization of prospero during mitosis in *Drosophila*. *Cell* 90, 449–458.
- Shin, J., Poling, J., Park, H.C., and Appel, B. (2007). Notch signaling regulates neural precursor allocation and binary neuronal fate decisions in zebrafish. *Development* 134, 1911–1920.
- Shin, J., Ma, S., Hofer, E., Patel, Y., Vosberg, D.E., Tilley, S., Roshchupkin, G. V., Sousa, A.M.M., Jian, X., Gottesman, R., et al. (2021). Global and regional development of the human cerebral cortex: Molecular architecture and occupational aptitudes. *Cereb. Cortex* 30, 4121–4139.
- Siegrist, S.E., Haque, N.S., Chen, C.H., Hay, B.A., and Hariharan, I.K. (2010). Inactivation of Both foxo and reaper Promotes Long-Term Adult Neurogenesis in *Drosophila*. *Curr. Biol.* 20, 643–648.
- Simões, A.R., and Rhiner, C. (2017). A cold-blooded view on adult neurogenesis. *Front. Neurosci.* 11, 327.
- Sipe, C.W., and Siegrist, S.E. (2017). Eyeless uncouples mushroom body neuroblast proliferation from dietary amino acids in *Drosophila*. *Elife* 6, 1–10.
- Skeath, J.B., and Thor, S. (2003). Genetic control of *Drosophila* nerve cord development. *Curr. Opin. Neurobiol.* 13, 8–15.
- Sousa-Nunes, R., Yee, L.L., and Gould, A.P. (2011). Fat cells reactivate quiescent neuroblasts via TOR and glial insulin relays in *Drosophila*. *Nature* 471, 508–513.
- Southall, T.D., Davidson, C.M., Miller, C., Carr, A., and Brand, A.H. (2014). Dedifferentiation of neurons precedes tumor formation in *lola* mutants. *Dev. Cell* 28, 685–696.
- Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., Hao, Y., Stoeckius, M., Smibert, P., and Satija, R. (2019). Comprehensive Integration of Single-Cell Data. *Cell* 177, 1888-1902.e21.
- Sturtevant, A.H. (1913). The linear arrangement of six sex-linked factors in *Drosophila*, as shown by their mode of association. *J. Exp. Zool.* 14, 43–59.
- Sun, M., Liu, L., Zeng, X., Xu, M., Liu, L., Fang, M., and Xie, W. (2009). Genetic interaction between Neurexin and CAKI/CMG is important for synaptic function in

References

Drosophila neuromuscular junction. *Neurosci. Res.* 64, 362–371.

Sun, Z., Asmann, Y.W., Nair, A., Zhang, Y., Wang, L., Kalari, K.R., Bhagwate, A. V., Baker, T.R., Carr, J.M., Kocher, J.P.A., et al. (2013). Impact of library preparation on downstream analysis and interpretation of RNA-Seq data: comparison between Illumina PolyA and NuGEN Ovation protocol. *PLoS One* 8.

Sundararajan, V., Pang, Q.Y., Choolani, M., and Huang, R.Y.-J. (2020). Spotlight on the Granules (Grainyhead-Like Proteins) - From an Evolutionary Conserved Controller of Epithelial Trait to Pioneering the Chromatin Landscape. *Front. Mol. Biosci.* 7, 213.

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. *Elife* 6, 1–23.

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

Tang, F., Barbacioru, C., Wang, Y., Nordman, E., Lee, C., Xu, N., Wang, X., Bodeau, J., Tuch, B.B., Siddiqui, A., et al. (2009). mRNA-Seq whole-transcriptome analysis of a single cell. *Nat. Methods* 6, 377–382.

Tong, C.W., Wang, J.L., Jiang, M.S., Hsu, C.H., Chang, W.T., and Huang, A.M. (2013). Novel genes that mediate nuclear respiratory factor 1-regulated neurite outgrowth in neuroblastoma IMR-32 cells. *Gene* 515, 62–70.

Tran, K.D., and Doe, C.Q. (2008). Pdm and Castor close successive temporal identity windows in the NB3-1 lineage. *Development* 135, 3491–3499.

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

Von Trotha, J.W., Egger, B., and Brand, A.H. (2009). Cell proliferation in the *Drosophila* adult brain revealed by clonal analysis and bromodeoxyuridine labelling. *Neural Dev.* 4.

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

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

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

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

Tsuji, T., Hasegawa, E., and Isshiki, T. (2008). Neuroblast entry into quiescence is regulated intrinsically by the combined action of spatial Hox proteins and temporal identity factors. *Development* 135, 3859–3869.

Ugur, B., Chen, K., and Bellen, H.J. (2016). *Drosophila* tools and assays for the study of human diseases. *DMM Dis. Model. Mech.* 9, 235–244.

Urbach, R., and Technau, G.M. (2003a). Molecular markers for identified neuroblasts in the developing brain of *Drosophila*. *Development* 130, 3621–3637.

Urbach, R., and Technau, G.M. (2003b). Segment polarity and DV patterning gene expression reveals segmental organization of the *Drosophila* brain. *Development* 130,

3607–3620.

Urbach, R., and Technau, G.M. (2004). Neuroblast formation and patterning during early brain development in *Drosophila*. *BioEssays* 26, 739–751.

Urbach, R., Schnabel, R., and Technau, G.M. (2003). The pattern of neuroblast formation, mitotic domains and proneural gene expression during early brain development in *Drosophila*. *Development* 130, 3589–3606.

Urban, J., and Mettler, U. (2006). Connecting temporal identity to mitosis: The regulation of Hunchback in *Drosophila* neuroblast lineages. *Cell Cycle* 5, 950–952.

Vogel, C., Teichmann, S.A., and Chothia, C. (2003). The immunoglobulin superfamily in *Drosophila melanogaster* and *Caenorhabditis elegans* and the evolution of complexity. *Development* 130, 6317–6328.

Walsh, K.T., and Doe, C.Q. (2017). *Drosophila* embryonic type II neuroblasts: Origin, temporal patterning, and contribution to the adult central complex.

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

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

White, K., and Kankel, D.R. (1978). Patterns of cell division and cell movement in the formation of the imaginal nervous system in *Drosophila melanogaster*. *Dev. Biol.* 65, 296–321.

White, K., Grether, M.E., Abrams, J.M., Young, L., Farrell, K., and Steller, H. (1994). Genetic control of programmed cell death in *Drosophila*. *Science* (80-.). 264, 677–683.

Wieser, R. (2007). The oncogene and developmental regulator EVI1: expression, biochemical properties, and biological functions. *Gene* 396, 346–357.

Wolff, T., Iyer, N.A., and Rubin, G.M. (2015). Neuroarchitecture and neuroanatomy of the *Drosophila* central complex: A GAL4-based dissection of protocerebral bridge neurons and circuits. *J. Comp. Neurol.* 523, 997–1037.

Xia, B., Yan, Y., Baron, M., Wagner, F., Barkley, D., Chiodin, M., Kim, S.Y., Keefe, D.L., Alukal, J.P., Boeke, J.D., et al. (2020). Widespread Transcriptional Scanning in the Testis Modulates Gene Evolution Rates. *Cell* 180, 248–262.e21.

Xiong, W.C., Okano, H., Patel, N.H., Blendy, J.A., and Montell, C. (1994). repo encodes a glial-specific homeo domain protein required in the *Drosophila* nervous system. *Genes Dev.* 8, 981–994.

Xu, C.S., Januszewski, M., Lu, Z., Takemura, S., Hayworth, K., Huang, G., Shinomiya, K., Maitin-Shepard, J., Ackerman, D., Berg, S., et al. (2020). A Connectome of the Adult *Drosophila* Central Brain. *BioRxiv* 2020.01.21.911859.

Xu, J., Hao, X., Yin, M.X., Lu, Y., Jin, Y., Xu, J., Ge, L., Wu, W., Ho, M., Yang, Y., et al. (2017). Prevention of medulla neuron dedifferentiation by nerfin-1 requires inhibition of notch activity. *Dev.* 144, 1510–1517.

Xue, J., Zhong, S., Sun, B. min, Sun, Q.F., Hu, L.Y., and Pan, S.J. (2019). Lnc-THOR silencing inhibits human glioma cell survival by activating MAGEA6-AMPK signaling. *Cell Death Dis.* 10, 866.

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

References

- Yang, R., Wang, M., Zhang, G., Li, Y., Wang, L., and Cui, H. (2021). POU2F2 regulates glycolytic reprogramming and glioblastoma progression via PDPK1-dependent activation of PI3K/AKT/mTOR pathway. *Cell Death Dis.* 12, 433.
- Yoshiura, S., Ohta, N., and Matsuzaki, F. (2012). Tre1 GPCR Signaling Orients Stem Cell Divisions in the Drosophila Central Nervous System. *Dev. Cell* 22, 79–91.
- 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 *tlh*, *otd*, *ems*, and *btd*. *Dev. Biol.* 182, 270–283.
- Yu, D., and Small, S. (2008). Precise Registration of Gene Expression Boundaries by a Repressive Morphogen in Drosophila. *Curr. Biol.* 18, 868–876.
- Yu, H., Awasaki, T., Schroeder, M.D., Long, F., Yang, J.S., He, Y., Ding, P., Kao, J., Wu, G.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.
- Zacharioudaki, E., Magadi, S.S., and Delidakis, C. (2012). bHLH-O proteins are crucial for Drosophila neuroblast self-renewal and mediate notch-induced overproliferation. *Development* 139, 1258–1269.
- Zhan, W.L., Gao, N., Tu, G.L., Tang, H., Gao, L., and Xia, Y. (2021). LncRNA LINC00689 Promotes the Tumorigenesis of Glioma via Mediation of miR-526b-3p/IGF2BP1 Axis. *NeuroMolecular Med.*
- Zheng, X., Zugates, C.T., Lu, Z., Shi, L., Bai, J.M., and Lee, T. (2006). Baboon/dSmad2 TGF- β signaling is required during late larval stage for development of adult-specific neurons. *EMBO J.* 25, 615–627.
- 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. U. S. A.* 108, 20615–20620.