Neurogenic Lineage Decisions with Single Cell Resolution

DISSERTATION
zur Erlangung des akademischen Grades

Doctor of Philosophy

(Ph.D.)

eingereicht an der
Lebenswissenschaftlichen Fakultät der Humboldt-Universität zu Berlin

Von Ana Veloso, M.Sc.

Präsident (komm.) der Humboldt-Universität zu Berlin

Prof. Dr. Peter Frensch

Dekan der Lebenswissenschaftlichen Fakultät der Humboldt-Universität zu Berlin

Prof. Dr. Dr. Christian Ulrichs

Gutachter/innen

1. Dr. Robert Zinzen
2. Prof. Dr. Markus Landthaler
3. Prof. Dr. Ann Ehrenhofer-Murray

Tag der mündlichen Prüfung: 21.02.2022

Hiermit erkläre ich, die Dissertation selbstständig und nur unter Verwendung der angegebenen Hilfen und Hilfsmittel angefertigt zu haben. Ich habe mich anderwärts nicht um einen Doktorgrad beworben und besitze keinen entsprechenden Doktorgrad. Ich erkläre, dass ich die Dissertation oder Teile davon nicht bereits bei einer anderen wissenschaftlichen Einrichtung eingereicht habe und dass sie dort weder angenommen noch abgelehnt wurde. Ich erkläre die Kenntnisnahme der dem Verfahren zugrunde liegenden Promotionsordnung der Lebenswissenschaftlichen Fakultät der Humboldt Universität zu Berlin vom 5. März 2015. Weiterhin erkläre ich, dass keine Zusammenarbeit mit gewerblichen Promotionsberaterinnen/Promotionsberatern stattgefunden hat und dass die Grundsatze der Humboldt Universität zu Berlin zur Sicherung guter wissenschaftlicher Praxis eingehalten wurden.

Berlin, den

_____

Unterschrift

1 ABSTRACT

1.1 GERMAN ABSTRACT

Die embryonale Neurogenese in *Drosophila* ist eine hochgradig koordinierte Abfolge von Zellschicksalsentscheidungen, die viele Ähnlichkeiten mit der Entwicklung des Nervensystems in Wirbeltieren aufweist. Diese Zellschicksalsentscheidungen sind räumlich und zeitlich koordiniert und hängen von der unmittelbaren zellulären Umgebung der neurogenen Zellen ab. Im *Drosophila*-Embryo delaminieren die Neuroblasten in fünf zeitlich getrennten Wellen vom Entwicklungsstadium 8 bis 11. Diese Zellen entstehen an stereotypischen Positionen in jedem Segment und sind entlang zweier räumlicher Achsen angeordnet: der dorsoventralen Achse, die das Neuroektoderm in drei Spalten unterteilt (ventrale, intermediäre und laterale Spalte), sowie sieben Zellreihen entlang der anteroposterioren Achse, die spezifische Markergene (z. B. *en* und *wg*) exprimieren.

Die Neuroblasten teilen sich und bringen stereotype Linien hervor, die durch die Zelltypen, Positionen und Verbindungen, die sie bilden, definiert sind. Obwohl diese Zelllinien stereotyp sind und die Zellen charakteristische Zellmorphologien, Projektionspfade, Verzweigungsmuster und Ziele aufweisen, sind die molekularen Mechanismen, die diese Merkmale bestimmen, noch weitgehend unbekannt. Jahrzehnte der Genetik haben einige Faktoren aufgedeckt, die für viele der Entscheidungen, die einzelne differenzierende Neuroblasten treffen, notwendig sind, aber ein Verständnis der einzelnen neurogenen Linien im Verlauf der Neurogenese auf Genomebene war bis vor kurzem *in vivo* unmöglich.

Die mRNA-Sequenzierung einzelner Zellen ermöglicht es uns nun, ein umfassendes Verständnis der Transkriptome einzelner Zellen in komplexen Systemen zu gewinnen. Ich habe diese Techniken genutzt und verbessert, um die Transkriptomdynamik zu untersuchen, die wichtige Schicksalsentscheidungen in der frühen Entwicklung des Nervensystems begleitet. Ich habe neurogene Stammzellen und ihre Nachkommenschaft aus eng-gestaffelten Embryonen analysiert, um aufeinanderfolgende Wellen der Neuroblasten-Delamination sowie ihre Proliferations- und Differenzierungsbahnen zu erfassen. Ich habe die Transkriptomdynamik einzelner neurogener Zellen und Zellpopulationen untersucht, wobei ich mich besonders darauf konzentriert habe, wie sich räumliche und zeitliche Variationen auf die Entscheidungen über das Zellschicksal auswirken. Mein Ziel ist es, zu entschlüsseln, wie sich Zellen unterscheiden, wenn Entscheidungen getroffen werden, die für die Entwicklung des Nervensystems essenziell sind. Dieses Wissen ist von unschätzbarem Wert für die Entwicklung von Modellen für die *in vivo*-Mechanismen, die es einzelnen Zellen im Nervensystem ermöglichen sich zu spezifizieren, differenzieren, projizieren und miteinander zu verknüpfen.

Ich habe Transkriptomdaten einzelner Zellen von Zehntausenden von Neuroblasten von der anfänglichen Abgrenzung bis zum Beginn der Differenzierung in Neuronen und Glia erzeugt. Ich war

in der Lage, spezifische neurogene Populationen und ihre Genexpressionsprofile entlang ihrer Differenzierungspfade zu identifizieren. Dies ermöglichte es mir, verschiedene Ebenen der Neuroblastenentwicklung zu erforschen, wie z. B. die räumliche, zeitliche und stammbaumspezifische Regulierung und spezifische Neuroblastenteilmengen. Ich war in der Lage, die komplizierten zeitlichen Achsen, die das entstehende embryonale Nervensystem formen, teilweise zu entschlüsseln – ein Prozess, der von der Fliege bis zum Menschen konserviert ist. Die Einzelzell-Transkriptomik hat die Identifizierung von lokalisierten Markern und sogar von spezifischen Neuroblasten ermöglicht. Dieses Verständnis kann nun mit Informationen über die individuelle Zellschicksale kombiniert werden, die diese Neuroblasten hervorbringen, wie z. B. ihre spezifischen neuronalen und glialen Schicksale. Durch die Aufdeckung ihres charakteristischen Repertoires an Signal-, Leit- und Adhäsionsmolekülen versuche ich nun, das Verhalten neurogener Linien *in vivo* durch gezielte Störung vorhersagbar zu verändern.

1.2 English Abstract

Embryonic neurogenesis in *Drosophila* is a highly coordinated sequence of cell fate decisions that bears many similarities to nervous system development in vertebrate organisms. These cell fate decisions are coordinated in space and time and depend on the immediate cellular environment of neurogenic cells. In the *Drosophila* embryo neuroblasts delaminate in five temporally separate waves from developmental stage 8 through 11. These cells emerge in stereotypical positions in each segment and are arranged along two spatial axes: the dorsoventral axis that divides the neuroectoderm into three columns (ventral, intermediate and lateral columns), as well as seven rows of cells along the anteroposterior axis that express specific marker genes (e.g. *en* and *wg*).

The neuroblasts divide and give rise to stereotypic lineages, which are defined in terms of the cell types, positions and connections they establish. Even though, these lineages are stereotypical with cells exhibiting characteristic cell morphologies, projection paths, branching patterns and targets, the molecular cues that establish these features are still largely unknown. Decades of genetics have uncovered several factors necessary for many of the decisions that individual differentiating neuroblasts make, but a genome-level understanding of individual neurogenic lineages as neurogenesis unfolds has been impossible *in vivo* until recently.

Single cell mRNA sequencing now allows us to gain a thorough understanding of the transcriptomes of individual cells in complex systems. I leveraged and improved these techniques to dissect the transcriptome dynamics that accompany major fate decisions in early nervous system development. I analyzed neurogenic stem cells and their progeny from tightly staged embryos that I collected to capture consecutive waves of neuroblast delamination, as well as their proliferation and differentiation trajectories. I have explored the transcriptome dynamics of individual neurogenic cells and cell populations with special focus on how spatial and temporal variation impacts cell fate decisions. My aim is to unravel how cells differ as decisions crucial for nervous system development are being made. This knowledge will be invaluable in developing models for the in vivo mechanisms that allow individual cells in the nervous system to specify, differentiate, project, and connect.

I generated single cell transcriptome data from tens of thousands of neuroblasts from initial delimitation to the beginning of differentiation into neurons and glia. I have been able to identify specific neurogenic populations and their gene expression profiles along their differentiation trajectories. This allowed me to explore distinct layers of neuroblast development, such as spatial, temporal and lineage specific regulation and to identify specific neuroblast subsets. I have been able to partially resolve the intricate temporal axes that shape the emerging embryonic nervous system, a process that is conserved from flies to humans. In fact, single-cell transcriptomics has allowed for the identification of localized markers and even of specific neuroblasts. This understanding can now be combined with information about the specific cell fates these neuroblasts produce, such as their specific neuronal and glial fates. By

uncovering their characteristic repertoires of signaling, guidance and adhesion molecules, I am now attempting to predictably alter the behavior of neurogenic lineages *in vivo* by targeted perturbation.

2 ACKNOWLEDGEMENTS

First, I want to thank my supervisor Robert Zinzen for the opportunity of developing this project in his lab. For all the scientific discussions, I learned a lot from you. I am very grateful for the opportunity to learn from my mistakes and for all your support throughout these four years.

I would like to thank the members of my thesis committee, Uwe Ohler and Altuna Akalin for all the discussions and guidance. I really appreciate the time you took to give me suggestions on how to proceed.

I am very grateful for the opportunity to learn from Vedran Franke on how to perform scRNAseq analysis, I want to thank you for taking the time to teach me. Philipp Wahle for the help starting this project, your support and guidance were a great help.

Laura Wandelt for all the help and work in the validation experiments performed.

I would like to thank all the former and current members of the Lab: Ana Guimarães, Alexander Glahs, Philipp Wahle, Alessandra Zappulo, Konstantinos Papadakis, Liene Astica, Sarah Ugowski, Rúben Abreu, Anna Monaco, Fiona Kerlin and Miriam Faxel. It has been a pleasure to share this experience with you.

Thank you, Sami and Liene for being here until the end, for all the support and fun times.

Thank you Ana Guimarães and Samantha Mendonsa for all the laughs and support, for always being there with a bottle of wine and a night on Ana's sofa.

Obrigada às meninas das conspirações, Diana, Sara e Joana por estarem sempre disponíveis para qualquer problema, piada ou situação caricata que aconteceu nestes quatro anos.

Obrigada Ana Guimarães por me teres recebido em Berlim. Obrigada por todas as sessões de terapia. Todas as conversas no lab, as piadas secas, todos os abraços e todo o teu apoio.

Thank you, Jake for all the support and kindness. Thank you for always having a supportive word and encouraging me all the way. For the fun times that helped dealing with the PhD frustrations.

Obrigada à minha família por sempre me apoiarem e estarem sempre presentes. Aos meus pais por me ensinarem que nada se consegue sem trabalho, é sem dúvida a lição que me motiva e incentiva todos os dias. Obrigada por me darem todas as oportunidades que nunca tiveram e todo o apoio que me continuam a dar. À minha irmãzinha, Bia, por sempre me lembrares que estou a fazer o que escolhi, por seres a pestinha mais querida. Pelo amor incondicional muito obrigada.

2.1 CONTRIBUTIONS

I would like to acknowledge every contribution to the work present in this dissertation.

The MDC Sequencing platform generated all the libraries for scRNAseq and performed all the NGS sequencing. I would like to specifically thank Caroline Braeuning for generating all of the 10X libraries.

The MDC FACs facility, Dr. Hans-Peter Rahn and Caroline Braeuning for the training and guidance.

The bioinformatics facility for all the help and support solving every problem throughout this project. I would like to specifically acknowledge Vedran Franke for mentoring me in the scRNAseq analysis and performing some of the PiGx analysis present in this dissertation.

Laura Wandelt for performing some of the *in situ* stainings used in this work (referenced in the figures when appropriate).

Anna Vlot (Uwe Ohler Lab) for the discussions, collaborations and insights in the identification of individual neuroblasts.

3 ABBREVIATIONS

Abl	Abelson tyrosine kinase
ac	achaete
AC	anterior commissure
Antp	antennapedia
ap	apterous
AP	anteroposterior
aPKC	atypical protein kinase C
ase	asense
ato	atonal
AUC	area under the recovery curve
babo	baboon
baz	bazooka
beat	beaten path
bHLH	basic helix-loop-helix
Bio	biotin
BLRP	biotin ligase recognition peptide
BSA	bovine serum albumin
cas	castor
cato	cousin of atonal
ChIP	Chromatin immunoprecipitation
Con	connectin
CycE	cyclin E
dan	distal antenna
danr	distal antenna-related
dap	dacapo
daw	dawdle
DIG	digoxigenin
Dlar	leukocyte-antigen-related-like
Dlg	discs large 1
dlp	dally-like
dpn	deadpan
dpp	decapentaplegic
Dr	Drop
drl	derailed
DSP	dithio-bis-[sulfosuccinimidyl propionate]
DTT	dithiothreitol
DV	dorsoventral
EGF	Epidermal growth factor
elav	embryonic lethal abnormal vision
EMT	Epithelial- mesenchymal transition
en	engrailed
ena	enabled
E(spl)	Enhancer of split
eve	even skipped
FACS	fluorescence-activated cell sorting
FasII	fasciclin II
FITC	fluorescein isothiocyanate
fra	frazzled
frac	faulty attraction
Gαi	G protein α i subunit
GMC	ganglion mother cell
GO	gene ontology
grh	grainy head
gro	groucho
gsb	gooseberry
hAF	hours after fertilization
hb	Hunchback
hh	Hedgehog
hkb	huckebein
HVG	highly variable genes
IC	intermediate column
Ilp4	insulin-like peptide 4
ind	intermediate neuroblast defective
insc	inscuteable
ISH	*in situ* hybridization

ISN intersegmental nerve
klu klumpfuss
Kr Kruppel
LC lateral column
lgl lethal (2) giant larvae
l(sc) lethal of scute
MeOH methanol
MNB midline derived neuroblast
mRNA messenger RNA
MZT maternal to zygotic transition
NB neuroblast
NC nuclear cleavage cycle
NES normalized enrichment score
Net Netrin
NGS next generation sequencing
N^{ICD} Notch intracellular domain
nmo nemo
NP neuronal precursor
Nrg Neuroglian
oc ocelliless
opa odd paired
pan pangolin
PBS phosphate buffered saline
PC posterior commissure
PCA principal component analysis
pdm POU domain protein
PNS Peripheral nervous system
pins partner of inscuteable
pnt pointed
pon partner of numb
pros prospero
ptc patched
put punt
rdx roadkill
repo reversed polarity
retn retained
RGE reversed graph embedding
rho rhomboid
RNA ribonucleic acid
sc scute
scRNAseq single cell RNA sequencing
Sdc syndecan
sens senseless
seq sequoia
sim singleminded
sli slit
SN segmental nerve
sog short gastrulation
SoxN SoxNeuro
sprt sprite
svp seven up
TN transverse nerve
TP time point
TSA tyramide signal amplification
TSS transcription start site
UMI unique molecular identifiers
VC ventral column
vn vein
VNC ventral nerve cord
vnd ventral nervous system defective
wg Wingless
wor worniu
ZGA zygotic genome activation

A note on nomenclature: Convention in the *Drosophila* field is:

- *Species* and *gene names* are generally italicized, while proteins are not italicized.
- Protein are generally capitalized, whereas capitalization of the gene name is determined by the official flybase.org reference. (Capitalization in the beginning was determined by whether the

first isolated mutation was dominant or recessive, this has not been a reliable indicator for a long time, especially considering computed genes, CGs.)

4 TABLE OF CONTENTS, FIGURES AND TABLES

4.1 INDEX OF CONTENTS

4.2 Index of Figures

4.3 Index of Tables

5 INTRODUCTION

5.1 EMBRYONIC DEVELOPMENT: FROM TOTIPOTENCY TO DIFFERENTIATED CELLS

Embryonic development starts with the fertilization of the female germ cell by a sperm, forming the zygote, a single diploid cell. The zygote is a totipotent cell that is capable of originating every other cell of the organism. After successive cell divisions (i.e. ***cleavage)*** regulated by maternally deposited proteins and mRNA, nuclear transcription is initiated. This switch from maternally supplied to the embryo's own genetic information has been commonly known as the ***mid-blastula transition*** – a term originating from the first studies of this phenomenon in frogs (*Xenopus leavis*)(Gerhart 1980). Today, this crucial step is generally referred to as the *maternal-to-zygotic transition* (MZT), or as *zygotic genome activation* (ZGA) and it occurs in different animals at vastly different times: For example, ZGA in the mouse embryo occurs as early as the 2-cell stage (Schultz 1993), while in *D. melanogaster* the zygotic genome does not activate fully until nuclear cleavage cycle 14 (NC14), which corresponds to an embryo containing >6 000 nuclei (Hamm and Harrison 2018).

ZGA generally precedes gastrulation, which describes the phenomenon of structural changes within the multi-cellular embryo. Coordinated cell movements due to, for example, apical-basal constriction, intercalation, and asymmetric division rearrange the embryo in a characteristic fashion, bringing cells into contact and allowing for the emergence of layers, structures, and specific cellular interactions. These rearrangements are instrumental in establishing the germ layers of metazoan embryos, such as meso-, endo-, and ectoderm, and allows them to effectively communicate with each other via morphogen gradients and signaling pathways (Ephrussi and St Johnston 2004; Hong et al. 2008; Gavis and Lehmann 1992).

The establishment of the embryo's body axes is a crucial step in early embryonic development and often precedes ZGA. In *Drosophila*, for example, both the anteroposterior and the dorsoventral axes are established even before fertilization during oogenesis.

5.1.1 The anteroposterior axis

The anteroposterior axis is established primarily by directed transport of proteins and messenger RNAs (mRNAs) in the developing egg. The egg is physically connected to an array of 15 nurse cells. Proteins and mRNAs produced in the nurse cells are exported into the egg, where an elaborate transport mechanism involving a microtubule network and molecular motors recognize peptides and sequences on proteins and mRNAs to transport them to the anterior or posterior poles.

Among the two most intensively studied anteroposterior axis determinants are *nanos*, *hunchback (hb),* and *bicoid*. Nanos protein is localized at the posterior pole of the egg (Figure 5-1A), where it acts as a translational inhibitor of a subset of mRNAs, including Hunchback – while *hunchback* RNA is

ubiquitous throughout the egg, local translational inhibition by Nanos results in an anteroposterior concentration gradient of Hunchback protein (Gavis and Lehmann 1992). Bicoid achieves an anteroposterior gradient by a different mechanism: *bicoid* RNA is supplied to the egg from the nurse cells, where its 3' UTR is recognized by adapter proteins that interact with molecular motors and transport the message to the anterior pole, where it is effectively anchored. Local translation results in a sharp anteroposterior Bicoid protein gradient (Ephrussi and St Johnston 2004)(Figure 5-1A).

The anteroposterior (AP) axis established in the egg will regulate the partition of the embryo into segments. The Bicoid gradient activates gap genes that are expressed in broad partially overlapping domains (Figure 5-1A) (Ephrussi and St Johnston 2004). Gap genes are also transcription factors and different gap protein combinations and concentrations activate pair rule genes which divide the embryo into regions and establish pair rule gene expression (Figure 5-1A) (Alberts, Johnson, and Lewis 2002). Pair rule genes are expressed in stripes and encode transcription factors that will in turn activate segment polarity genes (Alberts, Johnson, and Lewis 2002). Together, Gap and pair rule genes produce a segmented embryo, where each segment is endowed a specific identity along the anteroposterior axis (Figure 5-1A).

Segment polarity gene interactions establish different expression domains within the segment and play an important role in cell specification in these domains (Bhat 1999). These genes are not only transcription factors (e.g. Engrailed (En)) but also signaling molecules (e.g. Hedgehog (Hh) and Wingless (Wg)) that activate signaling pathways in specific striped domains of the embryo (Figure 5-1A) (Bhat 1999).

5.1.2 The dorsoventral axis

In contrast to the anteroposterior axis, the chief determinant of the *Drosophila* dorsoventral (DV) axis is the transcription factor Dorsal and regulation of its nuclear entry. In short, the oocyte's pronucleus migrates along the future dorsal side of the egg chamber, where it produces the *gurken* messenger RNA. Locally translated, the Gurken protein is secreted into the perivitelline space between oocyte and follicle cells and interacts with the Torpedo receptor displayed by the follicle cells surrounding the egg chamber (Hong et al. 2008). This interaction blocks the synthesis of the Pipe protease, which is therefore limited to the future ventral side. A pipe-induced proteolytic cascade ensues in the perivitelline space between, which results in the graded activation of the Toll receptor (homolog of the human TGFβ) by cleaved Spaetzle, with peak levels in the ventral-most position (Hong et al. 2008). The transcription factor Dorsal (homolog to NFκβ) is ubiquitously present in the egg, as is its inhibitor Cactus (the homolog to Iκβ), which prevents Dorsal from entering nuclei. Toll activation leads to degradation of Cactus, which therefore enables Dorsal to enter nuclei in a graded fashion with peak levels of nuclear Dorsal in the ventral-most nuclei of the syncytial embryo (Hong et al. 2008). Once inside the nucleus, Dorsal acts as a transcription factor that regulates genes in a concentration-

dependent manner. While some genes are repressed by Dorsal and therefore confined to the most dorsal domains of the embryo (e.g. *race, tup, dpp*), other genes are activated by the Dorsal transcription factor (Hong et al. 2008). Several Dorsal activation thresholds have been postulated, including genes that are activated by the highest levels of nuclear dorsal and are therefore confined to the ventral-most regions (e.g. *snail, twist*), others appear to require lower levels of Dorsal and are therefore expressed more laterally (e.g. *vnd, ind, msh, sog, soxN*; Figure 5-1B) (Hong et al. 2008). However, it should be noted that while Dorsal is the chief dorsoventral determinant, it is generally accepted that it does not act alone; rather, Dorsal initiates an intricate network of dorsoventral activators and repressors that act combinatorially to delineate dorsoventral expression territories.

One such territory is the neurogenic ectoderm that emerges in lateral domains of the early (stage5/6) embryo on either side of the embryo. The neurectoderm is marked by expression of genes such as *short gastrulation* (*sog*) and *sox Neuro* (*soxN*) and is delimited ventrally by the presumptive mesoderm and dorsally by the dorsoectoderm, which gives rise to epidermis and the extraembryonic membranes (amnioserosa). The neurectoderm gives rise to every part of the embryonic nervous system which is the focus of this study.

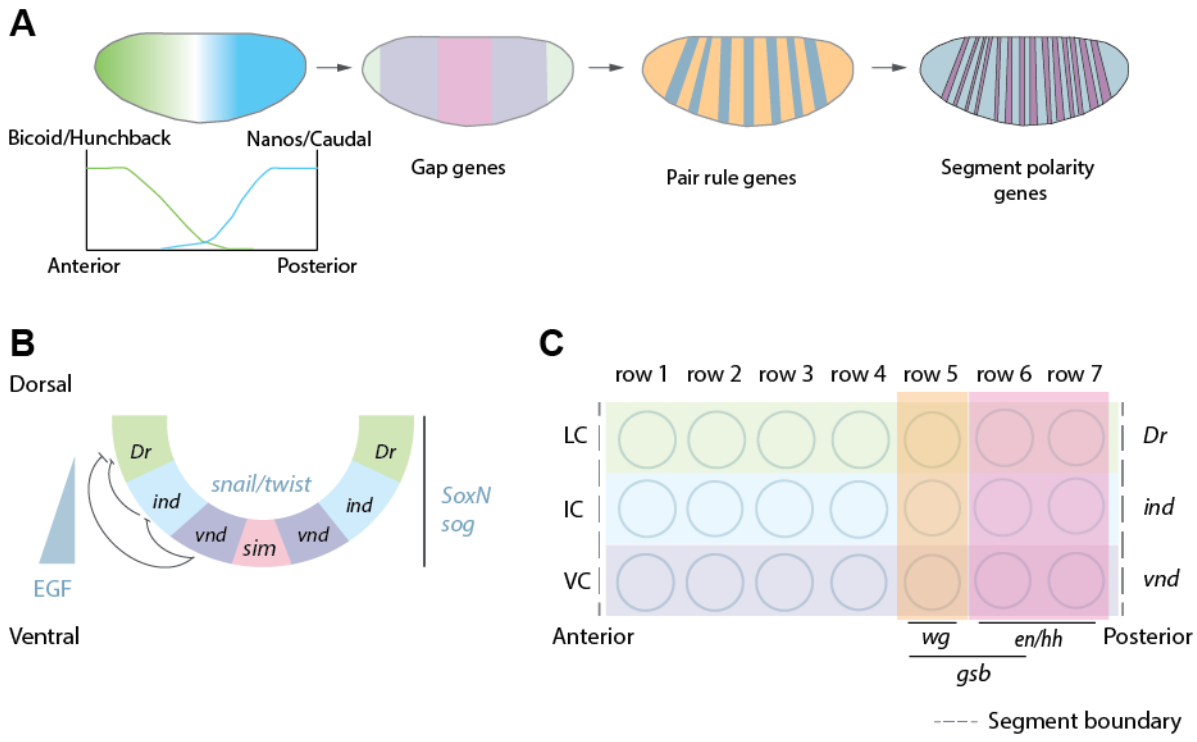


Figure 5-1 Early patterning and establishment of neurogenic domains

(A) Anteroposterior axis establishment. Bicoid/Hunchback and Caudal/Nanos gradients are deposited with peak levels at opposing poles of the embryo to regulate the expression of different gap genes along their concentration gradients. Together, these genes regulate pair rule gene expression, which in turn regulate segment polarity genes. Embryo schematics are lateral views, anterior left. For simplicity a stage 5 embryo morphology is shown. **(B)** Establishment of the dorsoventral domains of the neuroectoderm. A dorsoventral nuclear concentration gradient of the transcription factor Dorsal (not shown) produces specific dorsoventral expression domains. High

Dorsal levels activate genes such as *snail* and *twist* in the mesoderm (internalized at stage 7, not shown) and relegate expression of *SoxN* and *sog* as well as the columns specific transcription factors *(vnd* in the ventral column, *ind* in the intermediate column and *Dr* in the lateral column) to the neurogenic ectoderm in more lateral regions. The ventral midline expressing *sim* emerges at the interphase. EGF signaling emanates from the ventral midline and the ventral column and activates and/or maintains the expression of *vnd* and *ind* reinforcing these column identities. **(C)** Overview of a segment with markers. Indicated are the 3 dorsoventral columnar domains (horizontal) and 7 anteroposterior rows (vertical), as well as several signaling molecules expressed.

5.2 Origin of a neurogenic territory in *Drosophila melanogaster*

The embryonic nervous system of the fruit fly derives entirely from the neurectoderm in a manner that bears intriguing similarities to neural tube development in higher metazoans including mouse and human. One such similarity is the division of the neuroectoderm into dorsoventral domains along the dorsoventral axis characterized by the expression of specific transcription factors. In *Drosophila,* these domains are referred to as *neurogenic columns* and are specified with the onset of cellularization (embryonic stage 5) just dorsal to the presumptive mesoderm and ventral to the dorsal ectoderm. These swaths of cells on either side of the embryo express pan-neurectodermal marker genes such as *sog* and *soxN* (Figure 5-1B), but can be further subdivided into four columnar domains from ventral to dorsal:

(i) The ventral-most domain is a single row of cells called the ***mesectoderm*** which expresses the basic helix-loop-helix (bHLH) transcription factor *singleminded* (*sim*). The mesectoderm forms a sharp border to the ventral presumptive mesoderm at stage 5 and as the mesoderm invaginates and curls into the embryo's interior with the initial morphogenetic movements (stages 6 & 7), the two mesectodermal rows of cells come together at the ventral midline (Kearney et al. 2004).

(ii) The second domain – the ***ventral column*** – is a several cell wide strip that overlaps the mesectoderm and abuts the mesoderm at its ventral border. It is marked by the expression of the NK-homeodomain transcription factor *ventral nervous system defective* (*vnd*) which is expressed starting at stage 5 (Crews 2019).

(iii) The third domain – the ***intermediate column*** – is a slightly narrower strip of cells just dorsal of the *vnd*-expression domain; it expresses the transcription factor *intermediate neuroblasts defective* (*ind*) starting at stage 6 (Crews 2019).

(iv) The dorsal-most neurogenic column abuts the intermediate column and is referred to as the ***lateral column***. The lateral column is marked by expression of the transcription factor *Drop* (*Dr*) starting at stage 7 (Crews 2019).

It should be noted that while the genetic details of how these columns are established are not completely understood, the expression of the columnar markers is in part regulated by the activity of Dorsal in combination with other regulators, such as the Epidermal growth factor (EGF) pathway. Its receptor – EGFR – is ubiquitously expressed, but its ligands (Spitz and Vein) are spatial restricted. The

potent EGF ligand Spitz is effectively restricted first to the ventral column (stage 5/6) and then refines to the ventral midline (by stage 7/8) and the less potent ligand Vein remains confined to the ventral region of the neuroectoderm assuring restricted activity of Spitz and Vein and the localized activity of the EGF pathway.

In terms of establishment of the neuroectodermal columns only Dorsal is required to activate *vnd* expression but EGF is needed to maintain it at later stages. *Ind* is similarly regulated by both signaling pathways but, in this case, both Dorsal and EGF are required to initiate *ind* expression as both Dorsal and EGFR mutants result in a complete loss of *ind* (J B Skeath 1998). *Dr* on the other hand only requires Dorsal to be expressed (J B Skeath 1998)– in this region EGF is not active and missexpression of *Egfr* in the lateral column domain leads to loss of *Dr* expression, likely due to the expanded *ind* expression that has been shown to inhibit *Dr*.

A cross-regulatory mechanism between the transcription factor markers has also been described: In a regulatory network termed *ventral dominance*, it has been established that more ventral transcription factors repress more dorsal ones (i.e. Vnd represses *ind* and *Dr*, whereas Ind represses *Dr*)(Cowden and Levine 2003). This scheme may also hold for the bHLH transcription factor encoded by *sim*: though *sim* is initially co-expressed with *vnd* at stage 5, by the time the mesectoderm has converged at the ventral midline, *vnd* expression has been eliminated from the mesectoderm; however, the repression of *vnd* by Sim may be indirect.

As outlined below, each of these four columns will give "birth" to a series neurogenic stem cells – called neuroblasts in *Drosophila* – in a segmentally repeated fashion. These neuroblasts divide asymmetrically and give rise to self-renewing daughter cells as well as ganglion mother cells (GMCs). Ganglion mother cells in turn divide once – and only once – to give rise to differentiating neurons and/or glia. Importantly, the pattern of neuroblasts and their specific marker genes that emerge from each segment and column is largely stereotypic in that the place and timing of neuroblast delamination dictates the potential of its progeny with respect to what cell types are formed and how they project and connect.

5.3 EARLY NEUROGENESIS IN *DROSOPHILA* – MECHANISMS, CONSERVATION AND SOME OPEN QUESTIONS

Neurogenic stem cells, generally called "***neuroblasts***" in *Drosophila* emerge in early stages of embryonic development from the neurogenic columns after cells belonging to a proneural cluster have specified the central cell as the neuroblast-to-be via Notch/Delta signaling in a process generally known as lateral inhibition. Neuroblasts become round and large in size, they lose contact with their surrounding cells and "*delaminate*" in a process akin to the epithelial-mesenchymal transition (EMT) (Arefin et al. 2019). The delamination process spans five (5) distinct delamination waves over roughly

300 minutes of developing waves starting at late developmental stage 8 and concluding by late stage 11. Neuroblasts emerge in a segmentally repeated manner so that similar sets of neuroblasts emerge in each segment within the three columns (ventral, intermediate and lateral) and seven rows (from anterior to posterior). Most neuroblasts are identified by a 2-number naming convention (e.g. NB.X-Y) according to the position they occupy in this checkboard pattern: the first number (X) identifies the neuroblast's position along the anteroposterior axis and the second number (Y) refers to their position along the dorsoventral axis. For example, the neuroblasts emerging in the 7th row and 1st column (i.e. the ventral column) would be labeled NB7-1 (Figure 5-2B). However, it should be noted that this labeling is based on neuroblast positioning of the final neuroblast set at late stage 11, so that the number does not necessarily inform about the neuroblast's temporal origin. Similarly, each row gives rise to between three (3) and six (6) neuroblasts, so that the second number does not give information about the column of origin: while all neuroblasts NB.X-1 are from the ventral column, neuroblasts labeled NB.X-2 may delaminate from the ventral column or the intermediate column.

5.3.1 Proneural genes and lateral inhibition

In *Drosophila,* the first neuroblasts emerge by late stage 8 of development. Neuroblasts delaminate from the neuroectoderm, where groups of cells – proneural clusters – emerge that bear the potential to generate neuroblasts due to the expression of proneural genes. Proneural genes are transcription factors of the bHLH class and are necessary to initiate the development of neuronal cells. There are four members of this family that are involved in embryonic neuroblasts selection, *achaete (ac), scute (sc), lethal of scute (l(sc))* and *asense* (*ase*) (Bertrand, Castro, and Guillemot 2002; González et al. 1989). Different combinations of *ac*, *sc* and *l(sc)* are expressed in all the proneural clusters (Figure 5-2B); the fourth gene, *ase* is expressed in all neuroblasts of the CNS (Jarman et al. 1993). The activity of these transcription factors seems to be at least partially redundant as loss of any one of these four factors does not prevent neuroblast formation (Jimenez and Campos-Ortega 1990). The specific expression of these factors is important with respect to the later stages of neurogenesis: for example, it has been shown that expression of *sc* and *ac* is necessary for correct specification of the RP2 neuron (progeny of neuroblast 4-2), indicating that the proneural genes regulate distinct targets that modulate neuronal cell fate (Lai et al. 2005; J B Skeath and Doe 1996).

The acheate-scute gene family is expressed in the proneural clusters of the neuroectoderm and one of their roles is to restrict their own expression to one cell of the cluster. This selection process is called ***lateral inhibition*** (Artavanis-Tsakonas, Rand, & Lake, 1999; Figure 5-2A). Lateral inhibition is initiated by the activation of the Notch ligand, Delta. Expression of Delta in the future neural progenitors activates Notch (the transmembrane protein) in the neighboring cells, inducing Notch cleavage and release of notch intracellular domain (N^{ICD})(Campuzano and Modolell 1992; Hartenstein and Wodarz 2013). N^{ICD} enters the nucleus activating enhancer of split complex (E(Spl)) genes that will in turn inactivate the

expression of proneural genes while maintaining the Notch pathway active, thereby reinforcing the selection of the neuronal progenitor – the neuroblast – while the other cells of the proneural cluster acquire an ectodermal progenitor fate (Campuzano & Modolell, 1992; Hartenstein & Wodarz, 2013; Figure 5-2A). While Delta-Notch signal is bidirectional between all neighboring cells within the proneural cluster, the central cell generally expresses more Delta and becomes the neuroblast (Bertrand, Castro, and Guillemot 2002).

The role of proneural genes can be divided into two phases; an early phase with low levels of expression in all cells with the potential to become neuroblasts and a second where high levels of expression are achieved due to low levels / absence of Notch activity in the selected cell, culminating in the irreversible cell commitment. After neuroblast commitment the activity of the achaete-scute family decreases before neuroblasts start to divide, so the impact of their activity on neuronal cell fate must be dependent on the activation of a regulatory cascade they help initiate which will affect neuronal-differentiation programs later on in development (Cubas et al. 1991; James B. Skeath and Carroll 1991).

5.3.2 Segment polarity genes and neuroblast formation

As described above, the anteroposterior axis establishment in the egg (section 5.1.1) will later induce the expression of segment polarity genes. These interact to establish different expression domains within each segment and play a crucial role in cell specification. Segment polarity genes are either transcription factors directly and cell-autonomously driving gene regulatory networks (e.g. Engrailed), or signaling molecules acting cell-autonomously and -non-autonomously to affect cell fate (e.g. Hh and Wg) (Bhat 1999).

These genes divide each segment consisting of ~7cells in the anteroposterior direction into smaller domains (rows; Figure 5-1C). For example, *engrailed* is expressed in the posterior region of each segment (row 6 and 7) and *wingless* is expressed in the row of cells directly anterior to it (row 5; Figure 5-1C). Patched (Ptc) has been described to be expressed in rows 2-5 of neuroblasts (Bhat 1996) and gooseberry in rows 5 and 6 (Figure 5-1C). Only En and Wg label non-overlapping domains of the anteroposterior axis of the embryonic segment and exclusive markers for the other rows are still unknown.

The domain-specific expression of these genes is maintained by regulatory interactions between them. Cells expressing Hh (*en*-positive cells) activate Wg in the adjacent domain and in turn, secreted Wg maintains the activity of Hh in its neighbors (Swarup and Verheyen 2012). Secreted Wg binds to Frizzled (Wg receptor) in the *hh*-positive cells and activates the Wnt pathway, maintaining engrailed expression. The two gradients formed by Wg and Hh are thought to maintain the identity of the segment cells.

The importance of these genes in neuroblast fate is clear as absence of one or more of them results in failure to form specific neuroblasts. One example is *wg*, where loss of function results in failure to form

neuroblasts not only from the wg-positive row five, but also from adjacent rows. Wg mutations affect the identity of all neuroblasts in row 4 and 6 where, for example, the cells in the spatial position of neuroblast NB4-2 acquire the fate of the adjacent neuroblast NB3-2. This might be due to the expansion of *runt* from rows 2-3 in wild type embryos to row 4 in *wg* mutant embryos (Chu-LaGraff and Doe 1993), as well as absence of *slp1* and *slp2* (both activated by wg). *Slp* mutants recapitulate the *wg* loss of function phenotype seen in NB4-2 without affecting the neuroblasts from row 6 (Bhat, Van Beers, and Bhat 2000). For row 6 neuroblasts, mutations in *wg* might affect the Hh pathway, as in *hh* mutants – just like in *wg* mutants – row 6 neuroblasts fail to form, suggesting that these pathways interact to regulate the formation of these neuroblasts (Bhat 1999). Loss of function of any of the segment polarity genes result in failure of neuroblasts to form, or in fate changes of the neuroblasts in a row-specific manner.

5.3.3 Temporal regulation

The process of proneural gene expression followed by neuroblast selection occurs in waves. It is initiated in late stage 8 of embryonic development when the first nine neuroblasts per segment emerge (delamination wave S1), which is followed by four more rounds of delamination. This process occurs from stage 8 to stage 11 and it adds an intricate temporal dimension to neuroblast development.

5.3.3.1 Developmental time

Developmental time represents embryonic age and is often referred to in terms of "hrs after fertilization" (hAF) or morphological stages of embryogenesis. Neuroblasts delamination and proliferation occur between developmental stage 8 and 11, i.e. from ~3h 30min to ~8h 30min after fertilization. Considering that gene expression is a very dynamic process – particularly at early developmental stages – neuroblasts that emerge at earlier time points (first delamination waves) will delaminate within a different gene expression environment compared to later neuroblasts, which will affect neuroblast fate.

5.3.3.2 Lineage time

Drosophila neuroblast divisions and the progeny they give rise to are regulated by a series of temporal transcription factors (Figure 5-2E). The described temporal transcription factors cascade for the embryonic ventral nerve cord (VNC) neuroblast consists on the sequential expression of *Hunchback → Kruppel → POU domain protein 2 → Castor (Hb→Kr→Pdm2→Cas;* Figure 5-2E, F). These factors are intrinsically regulated as isolated neuroblasts grown in cell culture express the same factors in the correct order (Grosskortenhaus et al. 2005).

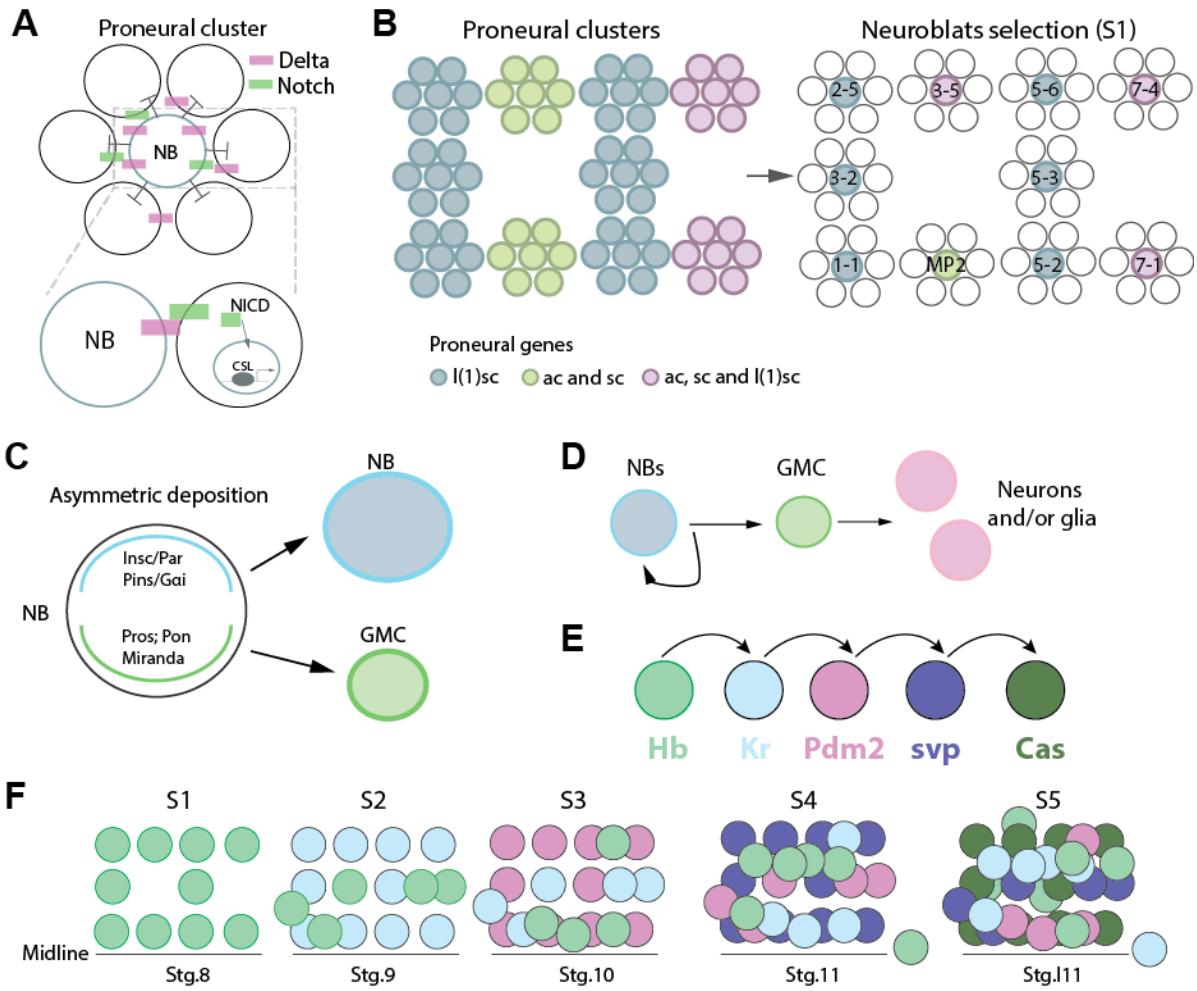


Figure 5-2 Neuroblast emergence, progression and division

(**A**) A neuroblast is selected from within a proneural cluster by lateral inhibition via Notch signaling. The neuroblast fate is maintained by activating proneural genes in one cell and reinforcing Notch in the surrounding cells. (**B**) Proneural genes are differently expressed among the proneural clusters within a segment. After neuroblast selection the proneural gene continues to be expressed in the neuroblast and modulate its progeny fate. Shown are the presumptive S1 neuroblasts and their respective names. **(C)** Asymmetric deposition of some of the known factors. Insc/Par; Pins/Gαl localize at the apical side, whereas Pros, Pon and Miranda locate to the basal side of the neuroblast. Asymetric distribution is required for asymmetric division into a self-renewing neuroblast and a smaller ganglion mother cell. (**D**) All but 1 embryonic neuroblast undergo a type I division mode, where a neuroblast divides asymmetrically giving rise to a neuroblast and a smaller ganglion mother cell (GMC) that will then divide only once more into two daughter cells (glia, neuron, or both). (**E**) Neuroblasts progress through a temporal cascade of transcription factors while self-renewing upon asymmetric divisions. (**F**) Waves of neuroblast delamination occur from developmental stage 8 to 11. As neuroblast delaminate, they express the first member of the temporal cascade and they will progress through this cascade (see E) with increasing asymmetric divisions. Neuroblasts born later will initiate the same cascade, but starting later upon delamination.

The first transcription factor, *Hunchback (Hb),* is necessary to maintain the neuroblasts in a young state (state with high proliferative potential) and *Hb* downregulation is necessary for cascade progression. Cytokinesis is required for the transition from *Hb* to *Kr*. In mutants for genes that affect the cell cycle progression, such as *string* and *pebble* that lead to G2-arrest and no-cytokinesis respectively,

neuroblasts never stop expressing Hb and do not progress to expressing *pdm* or *cas* (Grosskortenhaus et al. 2005).

In the absence of Hb, the progeny of the neuroblasts that emerge during the *hb* expression window are lost. The loss of Hb leads to an earlier expression of late factors which results in duplication of the progeny of the later factors. Surprisingly, mutations in *hb* have no further effect on the progression of the expression cascade from *Kr* to *pdm* to *cas*. However, if neuroblasts are deficient for Kr, the neurogenic progeny of Kr-positive ganglion mother cells is lost, whereas *Kr* overexpression results in respective duplications (Grosskortenhaus et al. 2005).

The temporal factors interact with each other to regulate the progression of the expression cascade: *hb* overexpression activates *Kr* and represses *pdm* and *cas*; *Kr* overexpression activates *pdm*, represses *cas* and has no effect on *hb*; Pdm activates *cas* (Grosskortenhaus et al. 2005). These interactions regulate the cascade progression but cannot be solely responsible for it as mutations have only slight effects on the cascade progression. The complete regulatory network driving the temporal factors is still unknown, but other genes are known to play a role. One example is *seven-up (svp),* which is expressed with Kr and plays a role in repressing *hb* while promoting *Kr* expression (Crews 2019).

Furthermore, the competence window of individual factors is regulated by other complexes. *Distal antenna/Distal antenna-related (Dan/Danr)* regulate the position of Hb in the nucleus, for example. Misexpression of *Dan* has been shown to extend the Hb competence window by blocking Hb movement to the nuclear lamina (Crews 2019). The Kr competence window is regulated in some lineages by the Polycomb Repressor Complexes and this regulation seems to be exclusive for lineages that give rise to motorneurons as it was not observe for neuroblasts with other progenies (Crews 2019). The temporal window[1] a neuroblast belongs to also helps define its progeny's position within the embryo, as the first-born neurons tend to occupy deeper layers and have longer projections, whereas later born neurons occupy more superficial layers and possess shorter projections.

5.3.4 Waves of Delamination

The selection of neuroblasts from proneural clusters occurs in five phases/waves (Figure 5-2F). In each wave a new set of neuroblast emerges, which contributes to the diversity of the progenitor pool. Lineages of neuroblasts that occupy the same spatial domains (anteroposterior and dorsoventral) with only temporal separation are often extremely distinct. For example, NB6-1 and NB7-1 belong to the ventral column and to the *en* domain but are temporally offset: NB7-1 emerges in the *first* and NB6-1 in the *third* delamination wave from the same spatiotemporal position (Figure 5-3A). Though these neuroblasts only differ in the time of delamination, NB7-1 gives rise to a glial cell and a host of important

[1] Temporal window represents the temporal transcription factor that the neuroblast expresses.

motoneurons with highly specific projection and connection behaviors, NB6-1 produces almost exclusively interneurons. To illustrate these stark differences more clearly:

- The motoneurons derived from NB7-1 have a complex projection pattern defasciculate and innervate nearly half of the larval muscles (~13 of 30) intra-segmentally. In addition, NB7-1 produces several local interneurons, some of which are axon-less to modulate local neuronal activity. The interneurons that produce axons also have a complex branching behavior with some crossing the anterior, others the posterior commissure; once across the midline, some contralateral projections are directed anteriorly, others posteriorly. Finally, NB7-1 reliably produces a nerve root glial cell. Ergo, NB7-1 produces a highly diverse, but precise sets of neurons, including one specific glia cell, a set of interneurons with and without projections, as well as a bundle of motoneurons that coordinate the contraction of almost half of the larva's body musculature throughout all larval muscle layers (Figure 5-3C).
- In contrast, NB6-1 delaminates later (S3) and it primarily produces three distinct groups of *intersegmental interneurons*. While NB6-1 has not ever been observed to generate proper glia, it has been reported to generate a single intrasegmental motoneuron, but only in the 1st thoracic segment (T1) (Figure 5-3B).
- It should also be noted that the manner in which the projection extend is regulated is not yet understood. Both neuroblasts generate interneurons that extend axons contralaterally, but while NB6-1 derived interneurons can project long distances spanning several segment boundaries, NB7-1 derived interneurons remain intra-segmental (Torsten Bossing et al. 1996).

Neuroblasts like NB6-1 and NB7-1 must be subject to several additional layers of regulation that control the exact progeny of individual neuroblasts necessary to form a functioning nervous system. To understand why the lineages of closely related neuroblasts are distinct in terms morphological features, as well as developmental and physiological behavior (Figure 5-3), we have to consider that "time of birth" likely plays a major role . The exact role of the delamination waves has not been described but it is a process that effectively generates regulatory diversity. The combination of delamination waves and lineage time produces a scenario where at any given point in neurogenesis neuroblasts are endowed with distinct sets of regulatory factors (Figure 5-2E/F).

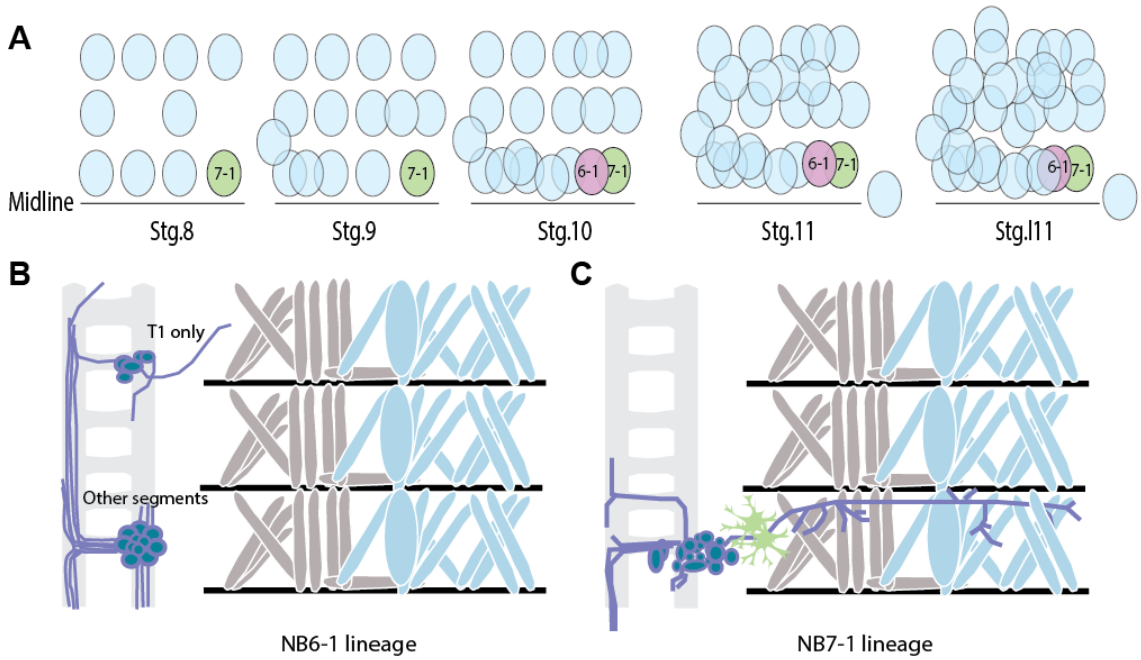


Figure 5-3 Lineage specificity of neuroblasts

(**A**) Schematic of a segmental neuroblast set along development. NB7-1 emerges at stg.8 (1st delamination wave, S1) while NB6-1 only delaminates at stg.10 (S3), both emerge from the same dorsoventral (ventral column) and anteroposterior domains. (**B**) Schematic of NB6-1 progeny: this neuroblast gives rise to primarily interneurons (dark blue) in the ventral nerve cord (light grey), producing both ipsi- and contralateral projections; it also gives rise to a motorneuron in thoracic segment 1 (T1). (**C**) Schematic of NB7-1 progeny: this neuroblast produces several, primarily contralaterally projecting interneurons. It also produces motorneurons that innervate several muscles (light blue and dark grey) from the lateral body wall, as well as glia (green).

5.3.5 Asymmetric Neuroblast division

Neuroblasts, once born, divide asymmetrically creating a larger and a smaller cell in a so-called "*type I*" division pattern – the larger cell constitutes the self-renewing neuroblast, while the smaller daughter cell constitutes the ganglion mother cell (Figure 5-2C). This process is regulated by neuroblast polarity and asymmetric deposition of well-studied factors. When neuroblasts delaminate from the neuroectodermal epithelium, they lose contact with the other cells from the epithelium in a epithelial-mesenchymal type transition, while proteins from the Par complex stay localized to the apical cortex (Wodarz and Huttner 2003; F. Yu, Kuo, and Jan 2006). This complex establishes the apical-basal polarity and it consists of Bazooka (Baz), Par-6 and the atypical Protein Kinase C (aPKC). This PAR-complex is responsible for the basal localization of cell fate determinants in the ganglion mother cell and it co-localizes with Inscutable (Insc, a protein that regulates the spindle orientation) to maintain neuroblast polarity (F. Yu, Kuo, and Jan 2006; Wodarz and Huttner 2003).

During mitosis, Insc/Par localize to the apical cortex and recruit another protein complex comprised of Partner of Inscutable (Pins) and the heterotrimeric G protein subunit (Gαi) that are involved in the control of spindle orientation as loss of Pins or overexpression of Gαi results in aberrant spindle orientation and mis-localization of basal determinants (F. Yu et al. 2002; Schaefer et al. 2001).

The Par complex regulates the basal localization of cell fate determinants, such as Prospero (Pros), Numb and their adaptor proteins Miranda and Partner of Numb (Pon), involving two tumor suppressor proteins Disc Large (Dlg) and Lethal (2) giant larva (Lgl) (Wodarz and Huttner 2003; F. Yu, Kuo, and Jan 2006). Mutations in Dlg and Lgl lead to loss of basal localization of cell fate determinants but there is a recovery of basal localization at ana- and telophase resulting in the segregation of the cell fate determinants to the GMC – *telophase rescue* (Ashraf and Ip 2001; Cai, Chia, and Yang 2001). This rescue is not observed in triple mutants for *snail*, *worniu (wor)* and *escargot* and even though the triple-mutant results in loss of *insc* expression, *insc* is not sufficient to explain the phenotype as loss *insc*-mutants show telophase rescue. Hence, there must be (an)other target(s) of Snail, Wor and Escargot that modulate(s) the segregation of ganglion mother cell determinants to the basal cortex (Wodarz and Huttner 2003; F. Yu, Kuo, and Jan 2006)

As mentioned above, besides the asymmetric localization of factors there is also a difference in cell size between the neuroblast and the ganglion mother cell. During anaphase the mitotic spindle moves closer to the basal cortex which means that the spindle asymmetric localizes the anaphase plate off-center thereby creating a smaller GMC from the basal side compared to the neuroblast (Kaltschmidt et al. 2000; Spana and Doe 1995).

5.3.5.1 Types of neuroglioblast division mode

In each hemi-segment, 30 neuroblasts produce in excess of 300 neurons and ~30 glia, with the exact numbers differing slightly depending on segmental identity. Among the ventral nerve cord neuroblasts are mostly pure neuroblasts (i.e. neuroblasts that only divide into neurons), one glioblast (that gives rise to only glia, called *LG*) and neuroglioblasts that produce both neurons and glia. In the *Drosophila* embryo there are six neuroglioblats per segment, NB1-1 (in abdominal segments), NB2-2 (in thoracic segments), NB1-3, NB2-5, NB5-6, NB6-4 and NB7-4 (Sasse, Neuert, and Klämbt 2015; Altenhein, Cattenoz, and Giangrande 2016). Among the neuroblasts with glial potential are several distinct division types:

- **Type 1 neuroglioblats** (i.e. NB1-1, NB2-2 & NB5-6) first produce ganglion mother cells that will divide into one neuron and one glia cell. After a type 1 neuroblast has generated all its glial offspring, the neuroblast continues dividing giving rise to only neurons. Only the abdominal NB5-6 produces an equal number of glia and neurons (four of each), whereas thoracic NB5-6, as well as all NB1-1 and NB2-2 produce more neurons than glia (Sasse, Neuert, and Klämbt 2015; Crews 2019; Altenhein, Cattenoz, and Giangrande 2016).
- ***Type 2a neuroglioblasts*** (e.g. NB1-3) first produce *neuroglial* ganglion mother cells, then switch to a pure glial lineage. NB1-3 belongs to mode 2a, as its lineage comprises more glia than neurons and it presumably switches from a neurogenic or neurogliogenic to a pure

gliogenic phase (Sasse, Neuert, and Klämbt 2015; Crews 2019; Altenhein, Cattenoz, and Giangrande 2016).

- ***Type 2b neuroglioblasts*** (e.g. NB2-5- & NB7-4) first produce *neurogenic* ganglion mother cells and later switch to producing gliogenic ganglion mother cells (Sasse, Neuert, and Klämbt 2015; Crews 2019; Altenhein, Cattenoz, and Giangrande 2016). For example, NB2-5- and NB7-4-derived glia cells are only produced in later stages indicating a type 2b division pattern.
- ***Type 3 neuroglioblasts*** (e.g. NB6-4) divide into a glial and a neuronal precursor that later give rise only to glia *or* neurons. The only currently known example is NB6-4 in thoracic segments, which produces two glia cells and four to six neurons (Altenhein, Cattenoz, and Giangrande 2016).
- Finally, the only known embryonic ***Type 3b glioblast*** is the LG neuroblast, which produces a purely glial lineage. This glioblast divides into two progenitors that divide into a total of eight to twelve glia cells (Altenhein, Cattenoz, and Giangrande 2016).

5.3.6 Neuroblast progression

Ventral nerve cord neuroblasts generally display a division pattern referred to as *type I* (with the sole exception of the MP2 neuroblast) and most neuroblast will later progress to a division *type 0*. Type I simply refers to a division pattern of neuroblasts dividing asymmetrically giving rise to a larger self-renewing neuroblast and a smaller ganglion mother cell, whereas type 0 indicates the absence of further regenerative-proliferative potential in the terminal neuroblast, which produces a final set of two similarly sized ganglion mother cell daughters. Following this final type 0 division the neuroblasts stop dividing (Crews 2019). It is assumed that there are factors that modulate the two division types and that regulate the switch from type I to 0. Though some putative regulators of the different stages have been described, this process remains to be elucidated further.

5.3.6.1 Early factors

There are groups of transcription factors that are expressed in all the neuroblasts. Among them, are members of the Snail zinc finger family (i.e. *snail, worniu*), the SoxB family (*SoxN* and *Dichaete*) and the proneural gene *asense.* These transcription factors play a crucial role in maintaining the neuroblasts in a proliferative state. Another group of genes that is important for the early stages of neuroblast progression are the early temporal transcription factors (*hb, Kr* and *pdm2*).

Previous work exploring the specific lineage of NB5-6T (thoracic segments) has shown that mutations in these genes result in lower neuroblast proliferation (Bahrampour et al. 2017). All of these factors are able to stimulate the expression of cell cycle genes such as *Cyclin E* (*CycE*), *string* (*stg*) and the E2f transcription factor (*E2f1*) that promotes cell proliferation and division and inhibits the cell cycle

inhibitor *dacapo* (*dap*). Downregulation of these early factors is necessary for the transition from type I to type 0 divisions.

5.3.6.2 Type I to type 0 transition

The transition from type I to type 0 requires alterations in the cell cycle genes and proliferative state of the neuroblasts. Mutations in the cell cycle regulators (*CycE* and *E2f1*) result in a decrease of neuroblast proliferation while missexpression results in an increase of neuroblasts and ganglion mother cells due failure to progress to type 0 divisions and resulting in overproliferation. As a neuroblast progresses, the expression of the early factors will decrease, as the late temporal transcription factor *cas* starts to be expressed, as well as *grainy head* (*grh)* and *Antennapedia* (*Antp)*. These genes induce *dap* expression that will in turn inhibit the cell cycle genes (*CycE* and *E2f1*), thereby triggering the final type 0 neuroblast division (Bahrampour et al. 2017; Baumgardt et al. 2014).

Notch signaling has also been reported to have a role in this transition, as *Notch* mutants result in more neurons (Ap neurons from the NB5-6T) suggesting a loss of type 0 in favor of a type I division. This is potentially modulated by the transcription factor Sequoia (Seq) which decreases in abundance with neuroblast progression (Gunnar et al. 2016). In early neuroblast progression Seq inhibits Notch, but as Seq levels decrease Notch becomes more active, thereby inducing E(spl) transcription factors that in turn inhibit cell cycle genes (Gunnar et al. 2016). Notch also activates *dap,* thus further inducing the transition to type 0. The gene expression changes that accompanied neuroblast progression affect the ganglion mother cell: in a type I daughter cell, *CycE* and *E2f1* are inhibited by *pros,* this inhibition still allows for another cell division as *dap* levels are low. In a type 0, the same cell cycle genes are inhibited by dap preventing the ganglion mother cell from further dividing and inducing its differentiation (Baumgardt et al. 2014).

As mentioned above after the transition to type 0, neuroblasts will stop dividing and there are three ways a neuroblast life is ended: (i) it exits cell cycle and it enters a quiescent state (most thoracic neuroblast); (ii) it exits cell cycle and it goes through apoptosis (e.g. NB5-6T); (iii) the lineage its terminated by apoptosis (e.g. NB7-3A) (Baumgardt et al. 2014).

5.3.7 Ganglion mother cell division – modulation of cell fate

Once formed, ganglion mother cells divide to produce two distinct daughter cells that often adopt different cell fates, such as neuron and glia cell, or motorneuron and interneuron. The mechanisms that modulate these distinct cell fates are largely unknown. However, it has been described that Notch plays a role in distinguishing the two daughter cells of a ganglion mother cell, called GMCa and GMCb (Karcavich 2005). When the ganglion mother cell divides, Numb is often asymmetrically localized to one of the daughter cells. Presence of Numb inhibits the membrane localization of Sanpodo (*spdo*, required for Notch signaling), leading to the loss of Notch signaling. This Numb-positive/Notch-

negative cell is referred to as GMCb, while the cell that has Notch activity is called GMCa (Crews 2019). Mutations in *numb* result in Notch signaling in both daughters and therefore two GMCa cells, whereas blocking of Notch results in two GMCb cells. While Notch signaling is crucial for differential fate acquisition, how exactly Notch modulates the resulting fates is unclear because neurons (motor- or interneurons) and glia can derive from either GMCa or GMCb (Karcavich 2005; Crews 2019).

Other genes have also been shown to be important for determining specific GMCa vs. GMCb fates, such as *even-skipped* (*eve*) and *klumpfuss* (*klu*). Eve is expressed in the first GMC (GMC1) of some lineages (e.g. NB1-1, NB4-2 and NB7-1). For example, in NB4-2, eve is expressed in GMC1a and the RP2 motorneuron it yields, but not in GMC1b and the resulting RP2sib. Loss of Eve in GMC1a results in failure to form the RP2 motoneuron, and concomitant duplication of the RP2sib neuron. *Klu* on the other hand is expressed in the second GMC (GMC2) produced by NB4-2 and limited to GMC2b. Accordingly, loss of *klu* results in duplication of GMC2a (Karcavich 2005).

The mechanisms that drive cell fate specification in ganglion mother cells is largely unknown. Exploring when and how cell fates of ganglion mother cell progeny are determined is an important step in understanding the emergence and safeguarding of cell type diversity in the nervous system. A prerequisite for gaining this understanding is the identification of specific markers of individual ganglion mother cells and their offspring, as well as a detailed understanding of their gene expression complements.

5.3.8 General Anatomy of the Drosophila embryonic nervous system

The central nervous system in the trunk of the *Drosophila* embryo is organized in a ladder-like structure, that contains glia and the cell bodies of neurons, as well as their axonal projections. A stain with an Antibody called HRP, which recognizes a glycoprotein moiety specific to axons (Finkelstein et al. 1990), is commonly used to reveal the general central nervous system architecture in insects (e.g. Figure 5-4A), as it reveals the position and routing of nerves like the bilaterally symmetric main tracts running beneath the ventral surface of the embryo to either side of the midline, as well as nerves such as the intersegmental (ISN) as it project from the "ladder" dorsally towards the somatic musculature in a segmentally repeated pattern.

The central nervous system is generally described as being composed of neuromeres (Figure 5-4), which are segmental, structural hallmarks (individual neurons may transverse multiple of these segment-size units). A neuromere contains the lateral tracts, as well as two commissures – an anterior commissure (AC) and a posterior commissure (PC), which are the channels through which axons traverse and connect the left and right central nervous system tracks (Figure 5-4). Similar to the organization of the neuroblast in the segment, the axonal projections and cell body positions are stereotypically repeated in each segment.

The motoneurons extend their axons outside of the central nervous system through three primary nerves, segmental (SN), ISN and transverse (TN) and the sensory neurons extend into the central nervous system through the SN and ISN (Figure 5-4). Of the 270 interneurons, 50% project anteriorly, 20% project posteriorly and 30% stay within the neuromere and the mechanisms driving this projection behavior remain poorly understood. Additionally, 69% of interneurons cross the midline and even this decision is not fully understood (Rickert et al. 2011). Neurons that cross the midline are referred as commissural or contralateral as they project in the opposite side of the neuromere of their cell body; neurons that do not cross the midline are called ipsilateral neurons (Figure 5-4B).

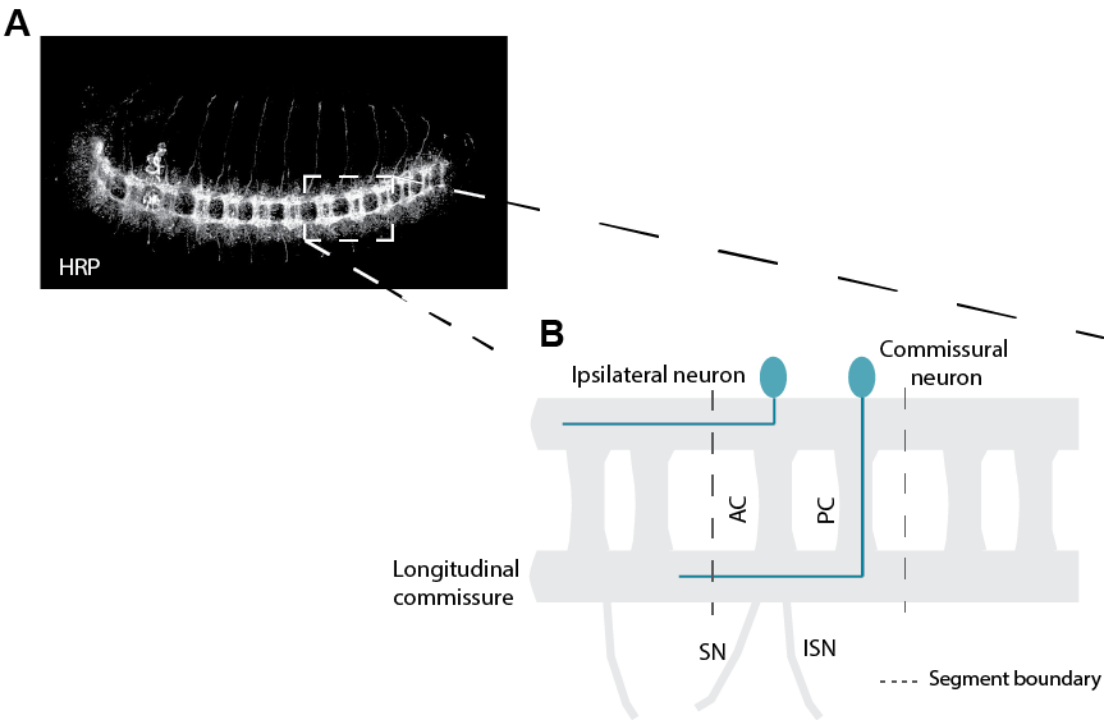


Figure 5-4 Structure of the embryonic nervous system.

(**A**) HRP-stained Embryo (stg. 13, ventrolateral view, anterior left) labels axonal projections. Neuromere units are the double-rung structures along the ventral nerve cord, with anterior (AC) and posterior commissures (PC) connecting the the main anteroposterior bundles of the central nervous system. Fine nerve bundles can be seen projecting from the ventral nerve cord dorsally within each segment towards the somatic musculature (not shown). (**B**) Schematic of the nervous system's neuromere units. Commissures are connecting nerve bundles; anterior (AC) and posterior commissures (PC) connect the two longitudinal commissures on either side of the midline running along the anteroposterior axis. The ISN and SN nerve bundles project out of the ventral nerve cord and dorsally. Types of neuronal projections are exemplified: axons of commissural neurons cross the midline through an AC or PC (projecting contralaterally), while ipsilateral axons project along the longitudinal commissure of its cell body.

5.3.9 Glial Lineages

Glia cells have been classified based on their function and the contacts they establish. Detailed descriptions of some of the neuroblasts that give rise to the different glial classes as well as their markers have been produced (Sasse, Neuert, and Klämbt 2015; Beckervordersandforth et al. 2008).

5.3.9.1 Surface-associated glia

The class of surface-associated glia consists of three glia types: (i) perineural glia, (ii) sub-perineural glia, and (iii) channel glia. The perineural glia and hemocytes produce a lamellar barrier sheet around the CNS. Underneath the perineural glia, lie the sub-perineural and channel glia, which form the blood (hemolymph)–brain barrier covering the longitudinal axons running along the ventral nerve cord.

5.3.9.2 Cortex-associated glia

Cortex-associated glia ensheath individual neuronal cell individually and are thought to provide support to the neurons. They associate with the sub-perineural glia and are likely facilitating nutrient and gas exchange between the neurons and the hemolymph.

5.3.9.3 Neuropil-associated glia

The class of neuropil-associated glia constitute the most variable group of *Drosophila* glia and can be divide into several types: (i) longitudinal (ensheating and wrapping) glia, (ii) astrocyte-like, and (iii) anterior midline glia.

(i) Longitudinal glia are derived from the longitudinal neuroblast and act as cellular cues for axonal growth in early neurogenesis.
They can be further subdivided into *ensheathing* and *wrapping glia.* Ensheathing glia form a thin layer around the neuropil (a dense network of nerve fibers, extensions and synapses), whereas wrapping glia wrap around nerves of the PNS.

(ii) Astrocyte-like glia are multi-purpose glia and serve several roles including neurotransmitter clearance and metabolic support. The name derives from their morphological and functional similarity to astrocytes in vertebrates.

(iii) Anterior midline glia are further subdivided into two groups: AMG and PMG.
AMGs ensheath axons that cross the midline; they play a roles in mediating several signaling pathways and secrete neurotrophic molecules that maintain axonal growth.
PMGs are a transient cell type with unknown function and undergo apoptosis during embryonic development.

5.3.10 Neuronal Lineages

The average hemi-segment contains 41 motoneurons derived from neuroblasts of the ventral, intermediate and lateral columns, an additional three midline-derived motoneurons and around 270 interneurons. The same neuroblast gives rise to the two groups of neurons and many GMCs divide into one of each neuronal type.

5.3.10.1 Motoneurons

Motoneurons in *Drosophila* are characterized phenomenologically by their projections; their molecular profile – with the exception of a few specific marker genes – remains largely unknown.

Drosophila motoneurons project along three major nerves the ISN (intersegmental nerves) that innervate the muscles in the next posterior segment, the SN (segmental nerves) that innervate muscles in the same segment and the TN (transverse nerve) that runs along the segment border to innervate a specific muscle called VT1 (Landgraf and Thor 2006; Kohsaka et al. 2012; Arzan Zarin and Labrador 2019) (Figure 5-5).

The ISN can be further divided into three groups based on the muscles they innervate: projections along the entire ISN to dorsal muscles; ISNb motoneurons that project to a set of ventral muscles and ISNd motoneurons that project to a different group of ventral muscles.

Similarly, the SN motoneurons are sub-divided into SNa motoneurons that innervate lateral domains of muscles and SNc motoneurons that project to ventral domains (Landgraf and Thor 2006; Kohsaka et al. 2012)(Figure 5-5).

Importantly, the regulation of motorneuron projections is still poorly understood, but several molecules have been uncovered. For example, dorsally-projecting motoneurons express *even-skipped* (*eve*) and *grain* (*grn*). Both transcription factors appear to be crucial for dorsal projection behavior, as mutants fail to innervate dorsal muscles (Landgraf and Thor 2006; Kohsaka et al. 2012). On the other hand, ventrally projecting neurons express *Nkx6* and *HB9,* and Nkx6 then modulates the expression of *Islet* and *Lim3* (Kohsaka et al. 2012). These two genes together with *ventral veins lacking (vvl)* regulate the specification between ISNb and ISNd (Thor et al. 1999).

What sets the specific motoneurons apart is still largely unknown and an in-depth understanding of their specific molecular repertoires is necessary to uncover the molecular mechanisms that modulate their morphology and function.

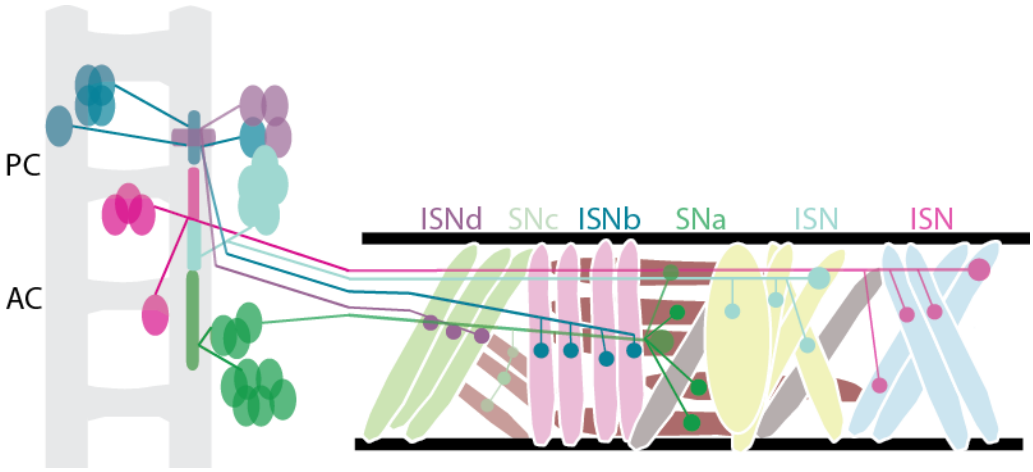


Figure 5-5 Morphology of Motorneurons and their projections

Motoneurons are divided into distinct groups based on their projections. Intersegmental (ISN) that can be divided into three subgroups, dorsal projecting (ISN) and ventral projecting (ISNb and ISNd); segmental (SN) that are divided into lateral external (SNa) and ventral external (SNc). PC – posterior commissure, AC – anterior commissure.

5.3.10.2 Interneurons

The interneurons are a bit of a "catch-all" class of cells of the *Drosophila* nervous system. They are characterized loosely based on molecular fingerprint and morphology – i.e. interneurons express neuronal markers such as *embryonic lethal abnormal vision* (*elav)*, intertwine with other neurons (primarily in the vicinity of the ventral nerve cord) and do not connect to muscles. No general mechanisms have been described regarding to how they develop or what modulates their projections; in fact, these mechanisms may not exist for interneurons generally, as it is likely that they can be subdivided into functional subclasses.

It is generally assumed that the mechanisms of axonal guidance are the same for all neurons (described below, section 5.3.11). Briefly, there are approximately 270 interneurons per segment and around 70% send their axons contralaterally, while ~30% stay ipsilateral (Rickert et al. 2011). Half of the interneurons have ascending (anteriorly projecting) axons and 20% have descending (posteriorly projecting) ones, and the predominance of ascending projections might be related with receiving and integrating signals from the brain (Rickert et al. 2011).

It should also be noted that many interneurons with exceedingly short axonal projections and many entirely axon-less interneurons have been described. Their role remains subject to heavy speculation, but it is generally assumed that they modulate local neuronal activity via feedback loops.

5.3.11 Regulation of projections – connections and guidance

There are two models that describe recognition between a specific muscle fiber and the axonal projection of its motoneuron: '*Relative Balance*' and '*Lock and Key*'.

(i) *Relative balance* describes a mechanism in which alterations in levels of cell surface proteins result in biased connections. Finding the correct target depends on the balance of attractive and repulsive cues and initially transient connections may be strengthened or weakened and resolve. Some evidence supports this model, such as broad expression of cell adhesion molecules like Fasciclin II (FasII).

(ii) The *Lock and Key* mechanism describes a decidedly more concrete targeting interaction where each motor neuron-muscle pair expresses a complementary set of molecular cues. For example, Netrin-B (NetB) is a cell adhesion molecule containing multiple laminin- and laminin-type EGF-like domains that interact with either the Unc-5 or Frazzled (Fra) receptor. NetB is only expressed in three of the 40 (3/40) embryonic muscle cells and mutations in NetB only effect the connections of the motor neurons that innervate these specific muscle cells. Several protein families have been described to play a role in guiding axons in the embryo to the proper muscle, but the regulatory mechanisms for individual neurons remain largely unknown (Landgraf and Thor 2006).

5.3.11.1 Netrins

There are two netrins (NetA and NetB) that are expressed in the midline, in a subset of neurons, in dorsal and ventral muscles and in longitudinal patches of the epidermis (Mitchell et al. 1996). These

netrins can potentially interact with two receptors, Unc-5 and Frazzled. Unc-5 is expressed in ISN, SNa as well as in glia positioned along these axons. Mutants for this receptor lead to ISN improperly crossing the segment boundary and SNa loose some of its branching; in contrast, ectopic expression of Unc-5 in the CNS neurons results in repulsion from the midline (Kolodziej et al. 1996; Labrador et al. 2005). One the other hand, Frazzled activity depends on the interaction of other factors such as Abelson (Abl), Trio and Enabled (Ena) (Forsthoefel et al. 2005) and mutations in *frazzled* result in early stalling of axonal projections. A comprehensive body of work has shown that Netrins are bifunctional guidance molecules capable of guiding axonal and glial projections by attraction as well as repulsion. While the repulsive effect of Netrin is mediated by Unc-5 (or a combination of unc-5/Frazzled to prevent segment boundary crossing), its attractive effect is mediated by Frazzled.

5.3.11.2 Slit and Robo

Interactions between Slit (sli) and Robo are important to control the midline crossing. Slit is an extracellular matrix protein produced by the midline and glia and it is detected by Robo present in the axonal growth cones (Kidd, Bland, and Goodman 1999). There are three Robo proteins, Robo (sometimes referred to as Robo1 for clarity), Robo 2 and Robo 3.

Robo1 is broadly expressed and it is critical for midline repulsion. Neurons with the highest Robo1 levels generally are longitudinal neurons that avoid the midline altogether. Robo2 is expressed in ventrally projecting motoneurons and is needed for guidance to the correct target muscles; it is thought this role is mediated by a protein domain that distinguishes it from the other Robo proteins (Rajagopalan et al. 2000). Importantly, the midline-crossing decision is likely regulated by the relative levels of attractive cues and repulsion via Slit/Robo and both Robo1 and Robo2 play a role in driving commissural neurons out of the midline, as only Robo1/2 double mutants show axonal projections to linger in the midline (Rajagopalan et al. 2000). Robo3 is expressed early in the aCC motoneuron but no defects due to absence of Robo 3 have been described (Rajagopalan et al. 2000).

5.3.11.3 Semaphorins and plexins

Semaphorins can be divided into transmembrane (Sema-1a, Sema-1b and Sema-5c) and secreted (Sema-2a and Sema-2b), though neither Sema-1b nor Sema-5c are expressed in the embryonic nervous system (Pasterkamp 2012; H.-H. Yu et al. 1998). Two *plexins* interact with semaphorins: Plexin A, which can interact with the transmembrane semaphorin Sema-1a and Plexin B, which can interact with the secreted semaphorins Sema-2a and Sema-2b (Ayoob, Terman, and Kolodkin 2006; Winberg et al. 1998). Interestingly, plexins are transmembrane receptors that contain an intracellular GTPase activity; upon interaction with semaphorins, these plexins may alter actin, microtubule and cell adhesion behavior.

Sema-1a and PlexinA act as a repulsive guidance cue and this is also mediated by Perlecan (a proteoglycan) that is present in motor axon trajectories and pathway choice points (Cho et al. 2012). Mutants in PlexinA show guidance defects in ISNb and SNa (Winberg et al. 1998). Sema-1a can also act

as a receptor where it affects motor axon defasciculation as mutants show axon pathfinding defects in ISNb and SNa. Sema-2a mutants show ectopic muscle contacts in ISNb, SNa and TN. PlexinB mutants display a similar phenotype to PlexinA (Ayoob, Terman, and Kolodkin 2006), the two plexins (A and B) have different effectors and while A can substitute B, the opposite is not true (Ayoob, Terman, and Kolodkin 2006).

5.3.11.4 Receptor protein phosphatases and heparan sulphate proteoglycans

Receptor protein phosphatases (RPTPs) are a group of eight identified proteins, where six play a role in axonal guidance in *Drosophila* (Sun et al. 2001). One such protein is leukocyte-antigen-related like (DLAR) that is necessary to innervate the ISNb targets. Abl and Ena interact with Dlar and are downstream effectors for the ISNb guidance (Wills et al. 1999). Heparan sulphates, Syndecan (Sdc) and Dally-like (Dlp) are the Dlar ligands (Fox and Zinn 2005).

5.3.11.5 Beat and Sidestep families

Beat (beaten path) is a family with 14 members, most of which are expressed in the nervous system (Pipes et al. 2001); they are transmembrane proteins that belong to the immunoglobulin superfamily. The members of this family have specific roles to play in selective defasciculation of motor axons. Sidestep functions as a chemo-attractant and most of the eight members of this family where shown to interact with beat proteins (Aberle 2009). For example, Beat-Ia is a receptor in selected axons and interacts with sidestep present in the cellular membranes along the axonal trajectory, affecting ISNb and ISNd branching (Fambrough and Goodman 1996; Sink et al. 2001).

5.3.11.6 TGFβ

The TGFβ superfamily can be divided into three groups: TGFβs, BMPs and Activins (Massagué 1998). One member of the activin group is Dawdle (Daw) that is required for axonal guidance – it is produced in the muscle and glia and it signals through the receptor complex Baboon (Babo) and Punt (Put) present in motor axons (Parker et al. 2006; Serpe and O'Connor 2006). There is also evidence for a role of BMP in axonal guidance, *faulty attraction (frac)* activity is dependent on BMP signaling and its required for ISNb and SNa guidance (Miller et al. 2011).

5.3.11.7 Cell adhesion

Most cell adhesion molecules have a rather mild effect on axonal guidance but some studies have shown that misexpression can effect motor neuron projections (Arzan Zarin and Labrador 2019). This is the case for FasII, where misexpression in neurons increases fasciculation in all motor nerves (D. M. Lin and Goodman 1994). Neuroglian (Nrg) is also important for proper projection behaviors: it is expressed throughout the nervous system and mutants show stalling in the ISN and ISNb axons (Bieber et al. 1989).

5.3.11.8 Transcriptional regulation of guidance and target finding

Transcription factors known to be expressed in subsets of motoneurons regulate (directly or indirectly) the expression of pathfinding cues. This regulation can probably explain the individual path- and target finding of each neuron. However, current technical limitations have only allowed the study of groups of cells systematically, or the assessment of a very limited number of factors with specificity, so that the neuron- specific molecular mechanisms are still largely unexplored.

Zfh1 is a transcription factor expressed in all motoneurons and *zfhI* mutants display projections defects, particularly ISN and SN axons fail to leave the central nervous system (Zarin et al. 2014; Arzan Zarin and Labrador 2019). Zfh1 regulates the expression of *unc-5*, *beat-Ia* and *FasII*, thereby regulating a significant portion of molecules involved in chemo-attraction and -repulsion, explaining the observed projection phenotypes (Layden et al. 2006).

The specific transcription factors for the dorsal and ventral motoneurons (see section 5.3.10.1) also regulate the expression of the guidance proteins of the different families described above. The dorsal ISN express *eve* and *grain,* and these factors are needed for correct axonal pathfinding. Eve was shown to regulate *unc-5, beat-Ia, FasII and Nrg* (Zarin et al. 2012, 2014).

Neurons that form the ISNb and ISNd express several transcription factors, including *Hb9, Nkx6, Islet, Lim3, Oli* and *Ubx* and mutants in any of these genes produce phenotypes in axon pathfinding. These phenotypes can in part be explained by the regulation of *Fas3* and *Robo2* by Nkx6, as well as the requirement of Islet for the expression of *frazzled* in the ISNb (Arzan Zarin and Labrador 2019). It has also been shown that Ubx regulates the expression of *wnt4*, a repellent cue present in muscle and ISNb (Inaki et al. 2007).

Interestingly, ISNb expresses *vvl* while ISNd does not, this combined with still undiscovered factors might explain the separation of these two classes of neurons and why they have distinct projections. As more expression differences are found between distinct neurons, we should be able to unravel the molecular mechanisms that drive their specific axonal projections and targets.

5.4 *DROSOPHILA* AS A MODEL ORGANISM

While certainly not suited for addressing questions regarding vertebrate inventions, such as the origin of the neural crest, there are many striking similarities between *Drosophila* and vertebrate development despite >700 million years of evolutionary divergence when it comes to fundamental biological principles of metazoan development. This includes crucial aspects of early neurogenesis as outlined above.

Early researchers in development, cell biology and genetics had adopted *Drosophila* as an experimental model because of key features such as ease of handling, crossing and maintenance, the power of replicating observations using large numbers of organisms, as well as because mutations are

easily induced and maintained. While these advantages are important factors for *Drosophila* remaining a prime model organism, many additional aspects, available resources and experimental strategies are constantly being developed and improved upon, which keeps *Drosophila* at the forefront of developmental biology. Among them are efficient and reliable generation of mutant and transgenic lines and experimental strategies allowing for the up- or down-regulation of specific endogenous and exogenous genes, sometime with pristine precision. While transgenesis has long been possible, innovations such as TALENs and CRISPR/Cas9 over the past few years have made precision genome engineering a reality (Gratz et al. 2015; Lee et al. 2015)

Additionally, the *Drosophila* genome was not only one of the first sequenced genomes, but it remains one of the best-annotated genomes, structurally as well as functionally. Not only have dozens of Drosophilid species and hundreds of inbred *Drosophila melanogaster* lines been sequenced, but huge consortium efforts have generated reagents to modulate gene activity and a plethora of freely available functional genomics data. One of the newest efforts in the *Drosophila* community is the FlyCellAtlas initiative founded in 2017, which seeks to resolve cell type diversity and developmental trajectories of the entire organism with single cell resolution. My project fits well within this effort as I am leveraging recent advances in single cell sequencing to understand diversity and developmental pathways in the emerging nervous system in the fruit fly embryo.

Genomically, *Drosophila* is and remains a premier research organism among complex diploid metazoan model systems, because of our deep understanding of genome sequence, structure and function. The *Drosophila* genome is relatively small with a size of *only* ~144 Mb (i.e. ~0.05% the size of the human genome). Current estimates place the number of coding genes at slightly more than half of the human gene complement (~14 000 vs. ~20 400, respectively). However, this number somewhat underestimates true complexity, as there are likely ~40 000 coding isoforms in *Drosophila* and the total estimated number of non-coding genes adds an additional 4 000 potentially regulated genes to the mix. Furthermore, it has been amply demonstrated that counting genes is a poor proxy for estimating regulatory complexity when it comes to spatio-temporal expression and combinatorial gene expression logic.

Therefore, the fruit fly remains highly relevant for the study of human disease: a recent assay found that ~65% – 70% of all disease-relevant mutations mapped in humans have an identifiable ortholog in the fly that may serve as an entry point into modeling disease (Ugur, Chen, and Bellen 2016) and some homologous transcription factors have been shown to be able to functionally replace one another in fly and human (Park et al. 1998; Halder, Callaerts, and Gehring 1995). The relevancy for *Drosophila* as a model for human developmental biology is very well illustrated by early neurogenesis: not only are crucial regulatory nodes in the gene regulatory networks of fly trunk neurogenesis and vertebrate neural tube development conserved (Weiss et al. 1998), but so are signaling pathways and many specific regulatory edges (Gilbert 2000).

6 AIMS OF THE THESIS

The overall aim of this thesis is to characterize the transcriptomes of individual neuroblasts and to place them within the embryo in terms of both, space (i.e. along the anteroposterior and dorsoventral axes) and time (i.e. developmental and lineage time) and to use this knowledge to mechanistically explain distinct cell fate behaviors.

6.1 CHARACTERIZATION OF SPECIFIC NEUROBLAST POPULATIONS ACROSS DEVELOPMENT

Using tightly staged embryo collections, I aimed to capture the entire process of neuroblast delamination. I managed to develop several improvements for the sequencing of a targeted cell population and I was able to use scRNAseq data to identify neuroblasts as they temporally progress. This allowed me to discern new factors along the lineage time gene expression cascade as well as factors that may be involved in specific neurogenic cell type decisions (e.g. neuro-glial decisions).

6.2 IDENTIFICATION OF NEUROBLAST POPULATIONS BASED ON THEIR SPATIAL EMBRYO POSITIONS

Neuroblasts delaminate from distinct regions in the neurectoderm and several markers of specific anteroposterior and dorsoventral positions are known. I aimed to use this information to spatially assign the sequenced neuroblasts according to their anteroposterior and dorsoventral origin. This opens new venues for investigating spatial regulation of neurogenesis – the ultimate aim is to identify individual neuroblast identities and to extract specific gene expression features that render them and their lineages distinct from one another.

6.2.1 Exploring dorsoventral position to understand the cell fate determining mechanisms

The neurectoderm of the early embryo is divided into three major columnar domains along the dorsoventral axis that can be identified by the expression of specific transcription factors. By separating neuroblast transcriptomes based on expression of these markers, I was able to explore similarities and distinctions in columnar regulation of the transcriptome throughout time. I aimed to explore transcription factors, cell adhesion molecules, as well as signaling pathways and to understand their regulation and regulatory impact.

6.2.2 Exploring anteroposterior position to understand domain-specific lineage distinctions

The neuroblasts in the *Drosophila* embryo delaminate in a segmentally repeated manner and are organized in stereotypical rows along the anteroposterior axis in each segment. I was able to identify cells from different domains (rows) along the anteroposterior axis due to the expression of segment polarity genes. This allows to separate cells based on anteroposterior position and studying what sets

these neuroblasts apart spatially in order to explore how these differences determine cell fates along the anteroposterior axis.

6.3 IDENTIFY INDIVIDUAL NEUROBLASTS

Using the information obtained from the previous aims to assign temporal and spatial identities along the anteroposterior and dorsoventral axis should allow me to thoroughly explore spatio-temporal transcription patterns. I was able to combine temporal and spatial information to identify and assign individual neuroblast identities. This identification opens the way for further discovery of unique cell fate determinant genes, such as transcription factors, cell adhesion and signaling molecules. I aim to understand how individual neuroblast transcriptomes differ from one another and to unravel the mechanisms by which individual neuroblasts produce distinct progeny in terms of cell types, projections and connections. I built on this information to functionally test and validate regulatory models *in vivo* and engineering cellular fates by applying a variety of genetic tools (e.g. Gal4-UAS and CRISPR).

7 RESULTS

7.1 scRNAseq of embryonic cell subpopulations

7.1.1 Background and state of the art

Single cell genomics is a fast-growing approach for addressing a variety of biological questions, including the regulatory dynamics that define cell types and states. Single cell studies are capable of capturing the incredible diversity of complex cell populations and revealing a multiplicity of cellular identities that were not appreciable previously (Macosko et al. 2015; Vitak et al. 2017; Alemany et al. 2018; Raj et al. 2018; Karaiskos et al. 2017). Additionally, single cell studies now allow us to investigate the progression from progenitors to differentiated cells with unprecedented resolution (Macosko et al. 2015; Vitak et al. 2017; Alemany et al. 2018; Raj et al. 2018; Karaiskos et al. 2017). In the last few years several platforms have been developed making single cell sequencing not only possible, but widely available. Among these platforms are microfluidic droplet-based methods such as DropSeq (Macosko et al. 2015) and commercial solutions such as 10X Genomics (10xGenomics 2019), which capture cells in nano-liter sized droplets, where cells are lysed and their RNA is barcoded and reverse-transcribed. Alternatives to droplet-based sequencing includes sequential indexing methods (e.g. SCI-seq (Vitak et al. 2017)) which utilize iterative barcoding and random re-arraying of cells to build unique cellular identifiers. These approaches allow for the sequencing of a large number of cells (10s to 100s of thousands at a reasonable sequencing depth per cell (100s to 1000s of detected genes), with droplet-based methods usually trading the lower throughput for the higher detection depth.

On the other hand, plate-based single cell sequencing methods profile individual cellular transcriptomes in dedicated micro-wells (e.g. SMARTseq and SMARTseq2 (Picelli et al. 2013)); the result is transcriptome information of great depth per cell, but at the cost of severely reduced throughput (often 100s to 100s of cells per experiment) and a greatly increased financial commitment per cell.

Single cell analysis is therefore not a specific methodology, but rather a conceptual approach to a biological problem – choosing the right experimental platform and parameters is of considerable importance to best address a particular biological question. For example, if the sample under study is complex and contains a wide variety of individual cell types, cellular throughput may become more important compared to the study of few, well-defined cell types or treatments. Similarly, if the sample is precious and hard to obtain, then recovery of material might be an issue and single cell sequencing approaches may differ drastically in this respect. If the cells under study change quickly – either developmentally or in response to a treatment course – chemical fixation may be desirable and not all single cell sequencing platforms can accommodate fixation.

In this thesis, I will primarily utilize the 10X approach to single cell sequencing for three key reasons: (i) it allows for reliable sequencing with reasonably-high throughput of thousands of cells per experiment, (ii) it does so at significant per-cell read depth generally quantifying the expression of several thousands of genes per cell, and (iii) it offers high input recovery, usually of around 40%. As neuroblast isolation required significant investment of time and resources on my part, input recovery (i.e. the proportion of input cells for which sequence information is actually obtained) was a major consideration.

7.1.2 Considerations for single cell sequencing of individual neuroblasts.

The central nervous system of the *Drosophila* embryo develops in a highly stereotypic manner from neuroblasts, which give rise to ~300 neurons and ~30 glia per hemi-segment. Most embryonic neuroblasts in the *Drosophila* embryo are so-called *type I* neuroblasts – a term that describes their proliferation pattern: Type I neuroblasts divide asymmetrically to self-renew and to give rise to a ganglion mother cell (GMC). The GMC divides once to produce two progeny cells (neurons and/or glia). Importantly, the developmental trajectories from neuroblast to differentiated neurons and glia are largely fixed and segmentally repeated (Torsten Bossing et al. 1996; Schmidt et al. 1997; Bhat 1999). Neuroblasts delaminate in 5 successive phases from the beginning of developmental stage 8 to the end of stage 11 producing a total of ~30 neuroblasts per hemi-segment. This means that each embryo in the developmental window we are interested in may contain anywhere from ~150 – ~450 neuroblasts, which represents less than 5% of the total number of cells in the embryo.

My project aims to create a cell atlas of the early developing *Drosophila* nervous system in order to better understand the specific developmental trajectories that distinct neuroblasts embark on. It is feasible that by accurately delineating individual neuroblasts and by understanding what molecular programs differentiate them, I will be able to develop mechanistic models of how individual neuroblasts form distinct progeny, such as glial cells and inter- or motor neurons showing specific projections and connections.

In a first feasibility test at unraveling neuroblast diversity in the developing embryo, I attempted to identify neuroblasts among single cell sequencing data from whole embryos according to established methods. As no single cell sequencing sets at the desired stages were available (stage 8 - 11), I collected embryos for 1h and aged them for another 3h 30min to capture the beginning of neurogenesis (primarily stage 8-9). To obtain single cell-resolved transcriptome data, I adjusted published protocols (Karaiskos et al., 2017, Alles et al., 2017) to sequencing on the 10X Genomics platform. In short, collected embryos were dechlorinated, dissociated using a dounce homegenizer, the cell suspension was filtered to obtain single cells, and the resulting suspension was examined microscopically to assure that intact single cells or nuclei were predominant (for a detailed protocol see Materials and Methods, section 9.1.5). After preparing RNA-sequencing libraries using the standard 10X Genomics Single Cell 3' Reagent Kit (v2)

procedure on the Chromium™ platform (10xGenomics 2019), I ran the single cell transcriptome libraries on an Illumina Next Generation Sequencer (NGS) and processed the resulting reads (the general NGS run statistics for all experiments in this study are summarized in Supplementary Table S 1). After basic sample quality control, alignment and clean up using PIGx (Wurmus et al. 2018)(see methods section 9.2.1 for detail description). Seurat (Stuart et al. 2019) was used to apply further quality thresholds: I required at least 200 *Drosophila* genes to define a cell for further consideration.

When clustering the resulting 7694 cells by their 12 principal components and visualizing the resulting clustering after dimensional reduction in 2D UMAP space, 11 clusters are evident that separate according to cell type identity (Figure 7-1A). All major cell types are readily identifiable, including a cluster containing 364 cells classified as the ventral nerve cord based on expression of highly variable genes (see Figure 7-1A, the discontinuous VNC cluster is outlined in green). This cluster harbors the neuroblasts I am interested in, but represents only ~4,7% of cells overall, which agrees with my assessment that the neuroblasts should contribute <5% of the embryo.

Upon extracting cells of the VNC population and re-clustering them (Fig. 7-1 B), I could confirm that this cluster clearly encompasses my cells of interest, as they highly and specifically express typical nervous system markers including the neuroblast marker *worniu*, the neuronal marker *elav* and the glial marker *gcm* (*glial cells missing*). It should be noted that these markers regionally separate within the UMAP projection of the VNC cluster, indicating that the transcriptomes of the VNC cluster cells, though still largely similar, are already in the process of differentiating along their neuro-glial trajectories. It should also be noted, that the cells identified here originate from stage 8-9 embryos, which is early in terms of neurogenesis: at these stages only the first two to three rounds of neuroblasts delamination have occurred, which means that the cells present should primarily represent neuroblasts and their primary GMC derivatives of the very first neuroblast delamination.

The challenge in my project revolves around accurate, in-depth cartography of the early nervous system – I aim to identify all neuroblasts and their immediate descendants. After sequencing whole embryos (Figure 7-1A-B), it became apparent that *in-silico* dissection of neuroblasts from whole embryo data is not a viable solution, simply because a yield of ~360 cells per sequencing run of ~10 000 cells will not allow for the desired neuroblast sampling depth.

7.1.3 Protocol development for neuroblast-specific sorting

An efficient alternative for *in-silico* dissection of embryos would be the targeted sequencing of individual neuroblasts. To achieve neuroblast isolation followed by single cell sequencing, several considerations have to be addressed, one of which is the *purification strategy*. Over the years, several purification strategies have been established for the bulk-sequencing of isolated cell types, some of which maintain the integrity of the cells. These methods include sorting of fluorescence-marked cells using

UAS-GFP responders and cell-type specific *Gal4 drivers* (Domsch et al. 2021), *INTACT* (Deal and Henikoff 2010), or *BiTS* (Bonn et al. 2012) and *MARIS* (Hrvatin et al. 2014).

I decided against relying on a UAS-Gal4 bipartite marker system, for two reasons:

(1) at the beginning of my project, we did not yet have any specific pan-neuroblast drivers, and

(2) the Gal4-UAS response is often weak in embryos and invokes a time delay.

This response delay is hard to quantitate but appreciable, which would not bode well for assessing developmental transcriptome dynamics. INTACT uses the same Gal4-UAS system to tissue specifically express the biotin ligase BirA, as well as a ubiquitously expressed modified nuclear pore complex component that bears a biotin ligase recognition peptide (BLRP) (Deal and Henikoff 2010). Tissue-specifically biotinylated nuclear pores allow for much fast and efficient purification of tagged nuclei, with the caveats that INTACT cannot be used to sort cells but *only* nuclei and that any leakiness of the Biotin ligase would be detrimental to the assay.

Instead, I decided to establish a neuroblast-isolation approach akin to BiTS / MARIS, which relies on endogenous markers: wild type embryos would be histochemical stained using neuroblast-specific 1° (primary) antibodies and visualized using fluorescently labeled 2° (secondary) antibodies, followed by fluorescent-activated cell sorting (FACS) for neuroblast purification. While the lab has established these methods for bulk ChIP-seq (Bonn et al. 2012), ATAC (Glahs 2021) and transcriptomics (McCorkindale et al. 2019), using such material in single cell sequencing is new territory and brings new challenges such as fixation compatibility.

For antibody stains to be possible, the biological material must be chemically cross-linked (fixed). This was in fact not a drawback, as I considered fixation to be a requirement anyway. Fixation *'freezes'* the regulatory state of the cell, which is crucial because neuroblast development proceeds exceedingly quickly and the FACS procedure takes hours to complete. Traditionally, we use formaldehyde as a chemical cross-linker (Bonn et al. 2012), which is a short-distance cross linker that primarily acts to covalently link amine groups to each other. However, formaldehyde was not an option here because of compatibility with the 10X sequencing platform: Reverse crosslinking after formaldehyde treatment requires extended and excessive heat prior to encapsulation of cells with beads on the 10X microfluidic chip. As this heat treatment would destroy cell integrity (Evers et al. 2011), I investigated alternative fixation methods that could be compatible with staining, FACS and 10X-sequencing. Though denaturation by heat or acid are well-established histochemical fixation methods, neither was a viable choice as the integrity of the cell and of nucleic acids would be equally compromised (i.e. acid-shearing). Alternative aldehydes were not an option either, as all of them share the same reversal problem as formaldehyde.

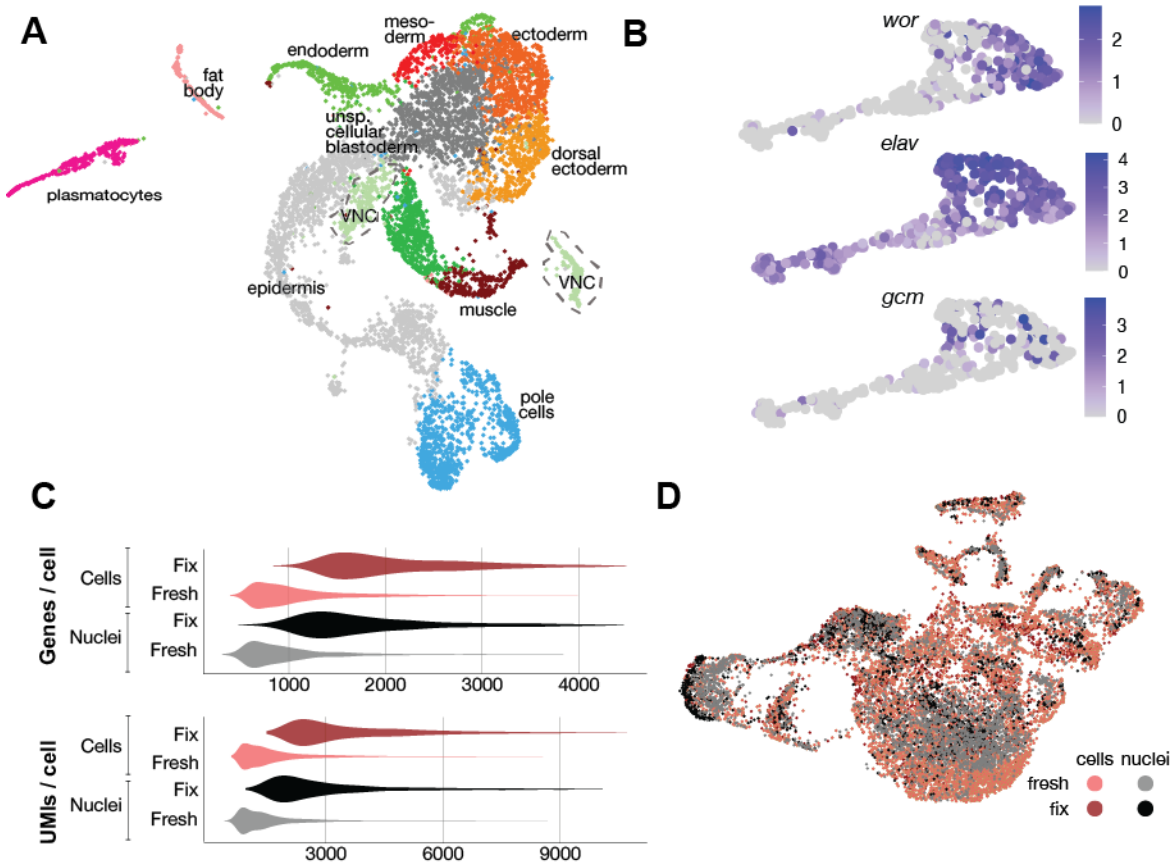


Figure 7-1: Assessment of whole embryo scRNAseq conditions.

(**A**) UMAP clustering of whole embryo scRNAseq data shows the major cell populations present in the embryo at early stages of development (stg. 8/9); only a small population (<5% of total cells) corresponds to the nervous system (VNC, note that the cluster is spatially separated into 2 regions, see stippled outline). (**B**) Re-clustering of the VNC cluster in (A); relative expression levels of marker genes for neuroblasts (*wor*), neurons (*elav*) and glia (*gcm*) are shown. (**C**) Comparison of fresh vs. fixed material and nuclei vs. cells in terms of identified genes per cell (top) and unique reads detected per cell (bottom). Cell preparations perform slightly better than nuclei and fixed material performs better than fresh material. (**D**) t-SNE representation of all conditions shows a good overlap of all populations indicating that all conditions capture the same populations.

A logical option I considered was methanol (MeOH), which fixes by denaturation and desiccation. Reversal is simple and effective as it only requires rehydration and the procedure is perfectly compatible with 10X transcriptomics as we demonstrated previously (Alles et al. 2017; Karaiskos et al. 2017). To validate that fixation did not affect single cell sequencing results, I wanted to directly compare fixed and fresh material side-by-side. Furthermore, I considered that the physical preparation for FACS isolation often disrupts the embryonic tissue not just to cells, but down to the size of nuclei. While nuclei have been used in single cell sequencing studies (Habib et al. 2017), the fact remains that nuclear isolation could shift RNA balance in favor of newly transcribed and nascent mRNA. To further assess if sequencing of single nuclei vs single cells measurably affects the resulting data, I prepared indexed 10X libraries for four samples in parallel and sequenced them together: (i) fresh unfixed cells, (ii) MeOH-fixed cells, (iii) fresh unfixed nuclei, and (iv) MeOH-fixed nuclei.

When comparing per-cell number of genes and unique transcripts (as measured by counting unique molecule identifiers, UMIs), cells do perform better than nuclei as might be expected: For shallowly sequenced samples, I detected 1078 or 2133 genes per fresh or fixed cells, respectively, identified by 1655 or 3628 UMIs. In comparison I was able to detect ~10% fewer genes (1016 genes for fresh and 1823 for fixed nuclei) and ~10% fewer UMIs (1578 UMIs for fresh and 3039 for fixed nuclei) (Figure 7-1C, compare red to grey). These numbers also show, however, that fixed materials allowed identification of a significantly larger number of UMIs and genes than fresh materials (Figure 7-1C, compare lighter to darker coloration) – the reason for this is not entirely clear, though it is conceivable that an initial denaturation of proteinaceous cellular components by alcohols partially strips away RNA-binding proteins and increases accessibility for the enzymes and reagents that are part of the library preparation process after rehydration. Importantly, when co-clustering all four samples each treatment contributes similarly across all clusters (Figure 7-1D). This indicates that despite some variation in the number of transcripts identified, the biological distinction among cells remains largely unaffected. Therefore, it is warranted to aim for the preparation of single-cell rather than single-nucleus suspensions, but it will not be necessary to avoid nuclei in the preparation: Nuclei may yield fewer genes per event but my indications are that they deliver faithful information nonetheless. On the other hand, it is advisable to use fixed material not only to assure cleaner and more reliable resolution of neuroblast development in staged collections, but also because fixation should allow for higher sampling depth per cell (Figure 7-1C-D).

However, I was unable to successfully stain MeOH-fixed embryos for any of several neuroblast markers, including Worniu, Deadpan, or Prospero, thus rendering simple MeOH fixation mute. The likely reason for this failure to stain is that denaturation by MeOH destroys the antibody-reactive epitopes. Therefore, I reasoned it might be possible to use a reversible crosslinker before MeOH fixation to stabilize the necessary epitopes. I focused on the homo-bi-functional fixative DSP (*dithio-bis-[sulfosuccinimidyl propionate]*) (Attar et al. 2018; Xiang et al. 2004), a less common fixative than aldehydes that is sometimes referred to as *Lomant's reagent*. DSP contains two reactive *N-hydroxysuccinimydyl* groups that – like formaldehyde – react with 1° and 2° amines to form covalent bonds. *Unlike* formaldehyde, DSP also contains a substantial 'spacer arm' between the two functional groups (which allows it to initiate crosslinks at longer distances than aldehydes) with a central *dithio-linkage* (S–S). The di-thio (S-S) link can be reduced and cleaved easily by treatment with low amounts (30-50m*M*) of the reducing agent dithiothreitol (DTT) at mild temperatures (25-37°C) and slightly basic pH (~8.5) within a reasonably short period of time (~30 min)(Attar et al. 2018).

Published studies have used DSP as a fixative in *Drosophila* embryos for chromatin immunoprecipitation (ChIP) (Aoki et al. 2014), where intact embryos were incubated with a mix of DSP (2-5m*M* in aqueous solution) and heptane. Presence of an organic solvent like heptane is well known to facilitate transport of the polar fixative across the hydrophobic vitelline membrane that surrounds the

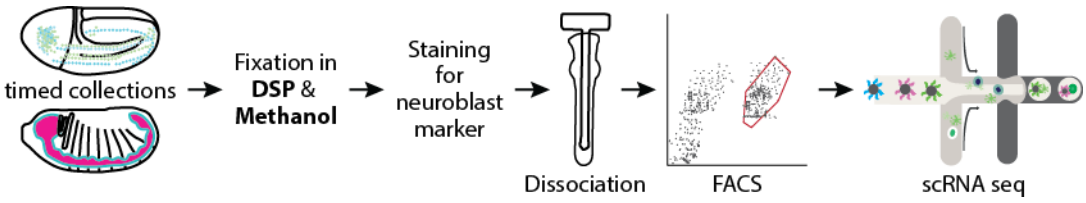


Figure 7-2: Schematic representation of DSP-fixed neuroblast isolation by FACS.

Tightly collected embryos that capture specific waves of neuroblast delamination were collected and fixed using a combination of DSP and methanol. The embryos are later rehydrated and stained with appropriate 1° antibodies (such as the neuroblast-marker mouse-anti-Worniu) and 2° antibodies (such as anti-mouse-Alexa555). After mechanical dissociation using a dounce homogenizer and filtering for intact single cells and nuclei, the neuroblasts are sorted by FACS using a combination of size scatter measurements and fluorescent intensity for DAPI (DNA) and 2° antibodies. Purified cells are inspected for integrity on a microscope prior to single-cell transcriptome library preparation (10X Chromium) and NGS sequencing

embryo. Indeed, I found that embryos fixed with a DSP could be stained for neuroblast markers such as Worniu, However, embryos fixed in this manner (see materials and methods, section 9.1.2.2) need to be stored at 4°C and should *not* be frozen to preserve cellular integrity in an aqueous environment. Unfortunately, this procedure preserves RNA integrity for only a few days (Attar et al., 2018). It is possible that DSP-fixation is insufficient to completely abrogate RNase activity, or that DSP is slowly being reduced even at 4°C. Therefore DSP-only fixation is not compatible with my experimental plan, which requires collection, immediate fixation and storage of embryos over days-to-weeks for later pooling and neuroblast extraction by FACS.

To improve RNA stability while maintaining DSP-mediated preservation of epitopes, I used an initial fixation in 1mg/ml DSP (~2.5m*M*) mixed 1:1 with heptane, followed by dehydration and freezer storage in methanol (Figure 7-2). This allowed for longer-term storage with no significant RNA degradation observed even after more than eight weeks of storage at -20°C. When ready, the embryos can easily be rehydrated in RNase-free phosphate buffered saline (PBS) and stained with antibodies (Figure 7-2). I was able to reliably stain DSP-MeOH fixed embryos with anti-Worniu for example (e.g. see Figure 7-3A), which became my standard stain for marking and sorting neuroblasts.

It should be noted that while staining embryos after DSP-MeOH fixation with anti-Worniu antibodies generally worked well and gave a robust signal indistinguishable from the reported neuroblast pattern (Arefin et al. 2019), staining of isolated cells *after* embryo homogenization generally failed. A possible reason for this behavior is that the homogenization process allows for partial reduction of DSP crosslinks at room temperature and subsequent unavailability of antibody epitopes, but as the successful stains of whole embryos allowed the additional opportunity to quality-control obtained staining patterns on intact embryos, I did not explore this issue further.

In order to separate embryonic neuroblasts from the remainder of the embryo, I developed a FACS-protocol that would robustly produce single neuroblast cells and nuclei (Figure 7-3B; a detailed description of the FACS considerations are available in the *Materials and Methods*, section 9.1.6). In

short, stained DSP/MeOH-fixed embryos were dounce-homogenized, the cell-slurry was separated from large embryonic debris by centrifugation and the remaining homogenate was filtered using 20nm-mesh cell strainers to remove cell clumps, taking care not to remove neuroblasts. The resulting cell suspension was loaded onto a BD AriaIII™ or Phusion™ cell sorter in a PBS buffer and 0.04% RNase-free bovine serum albumin (BSA) to aid cells and nuclei remain in suspension during the sorting procedure. My sorting strategy (Figure 7-3B) included to first gate-out large aggregates and small cellular debris using the areas produced by forward and side scatter of particles (Figure 7-3B, top). Furthermore, I excluded duplicates by gating on forward scatter area versus width (Figure 7-3B, middle). Finally, I used DAPI fluorescence to assure that sorted particles correspond to cells or nuclei, as well as 2° antibody fluorescence (primarily an Alexa555 conjugate) to purify neuroblasts by detection of the anti-Worniu antibody (Figure 7-3B, bottom).

Microscopic inspection of the material before and after sorting visually demonstrates sorting necessity: While Worniu-positive neuroblasts are few and far between before sorting (Figure 7-3C, top), neuroblast purity is generally above 95% after a single round of sorting (Figure 7-3C, bottom).

Importantly, I could not observe any significant RNA degradation after DSP-fixation, MeOH storage, staining, dissociation, and FACS purification of neuroblasts. The Bioanalyzer™ trace that separates RNA by size clearly shows a predominant peak pair at ~1 800 nucleotides (Figure 7-3D), which is the expected behavior for intact *Drosophila* RNA extracts[2]. I was therefore able to produce single cell transcriptome sequencing libraries from sorted neuroblasts that allowed detection of 10 000s of unique transcripts (UMIs; Figure 7-3E, top) mapping to 1 000s of genes (Figure 7-3E, bottom) per cell (for a comprehensive table of basic sequencing statistics across all libraries, see Supplementary Table S 1).

[2] Note that the 28S rRNA of insects contains a hidden break, causing the 28S and 18S peaks to largely coincide instead of creating independent peaks as in vertebrates that would allow for calculation of a RIN (RNA integrity number).

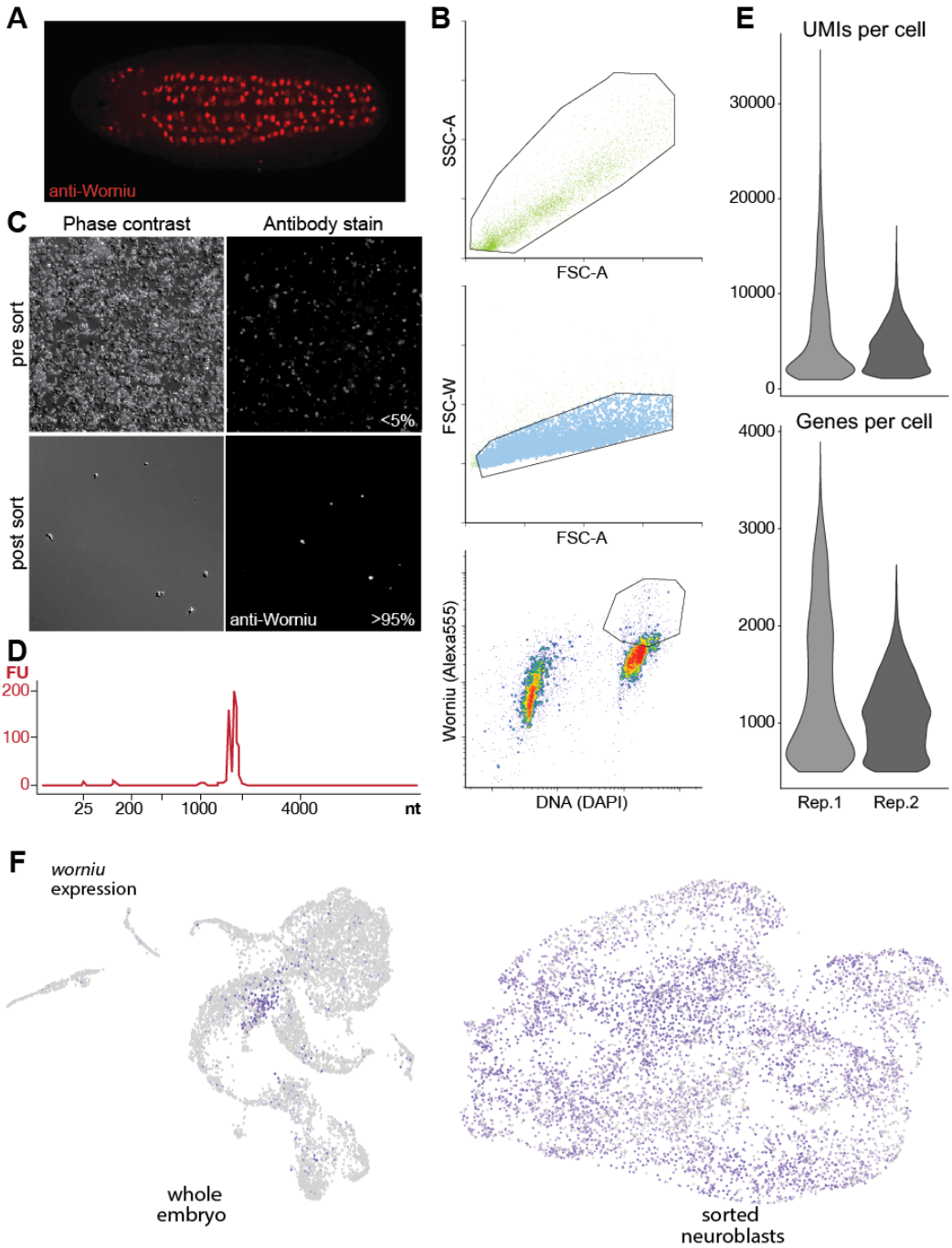


Figure 7-3: Fixation and single cell sequencing of sorted neuroblasts.

(**A**) Anti-Worniu labels all neuroblasts of the stg. 8/9 embryo. The ventral view (anterior left) shows the 3 adjacent lateral columns of neuroblasts above and below the midline. (**B**) Gating scheme for neuroblast sorting by FACS. Forward by size scatter area gating (top) eliminates large debris, forward scatter width by area gating (middle) selects against doublets, DAPI separates particles based on DNA content, whereas the Alexa555 fluorescence separates particles based on the neuroblast marker Worniu (bottom). (**C**) Microscopy of dissociated cells pre- (top) and post-sort (bottom) shows successful purification of Wor-positive cells from <5% to >95% purity. (**D**) The Bioanalyzer plot shows that RNA integrity is maintained and no RNA degradation is detectable after DSP-MeOH fixation of embryos, 20°C storage, dounce homogenization and FACS. (**E**) Violin-plots show the number of UMIs (top) and genes (bottom) per cell for two replicates of sorted neuroblasts. (**F**) Plot of dimensionally reduced scRNAseq data by UMAP from whole embryos (left) and sorted neuroblasts (right) for the same developmental timepoint; purple indicates the level of expression of the neuroblast marker Worniu. Note the drastic proportional increase in neuroblasts among the sequenced cells.

7.1.4 Method outlook and application

The drastic difference my sorting protocol makes can be easily appreciated in side-by-side comparison of dimensionally reduced UMAP plots showing whole embryo versus sorted neuroblast data (Figure 7-3F). While neuroblast-expressed marker gene RNA is detectable in almost all of the sorted neuroblasts (Figure 7-3F, right), this is true only for a tiny fraction of cells from whole embryos (Figure 7-3F, left). Accordingly, my new sorting protocol increases the number of detected neuroblasts from a few hundred to several thousand per experiment, which emphasizes the additional sampling depth we stand to gain by sorting our cell type of interest.

It should be noted that a major drawback of sorting is not in data quality, but in inherently limiting the data by design: We are implementing a hard filter by requiring expression of a sorting marker (here: Worniu protein). For example, I will not be able to gain insights into any neuroblast populations that may have gone undetected so far because it is not labeled by Worniu. Additionally, as Worniu protein presence wanes with increasing neuro-glial differentiation, more differentiated neuroblast progeny will rapidly disappear from my analysis so that my insights will be limited to my specific cell population of interest: neuroblasts and their immediate GMC descendants.

However, I believe that the power of sorting for a specific cell population is specifically in focusing the analysis and in sacrificing breadth for depth. As the following chapters will show, I was able to assemble a large dataset of >42 000 neuroblast transcriptomes, which would have been prohibitively expensive with whole-embryo sequencing (a >20-fold increase in cost assuming neuroblasts represent ~5% of sequenced cells). This sampling depth was instrumental in delineating specific territories and neuroblast populations and therefore allowed insights into the mechanisms driving neuroblast specification that would not have been possible otherwise.

My protocol currently requires reliable antibodies for the cell population of interest and access to a FACS machine. If the antibody detects a nuclear protein, extraction of nuclei is possible in addition to cells. Further modifications, such as the genetic encoding of lineage markers (e.g. a cell type-specific enhancer driving a site-specific recombinase to remove a stop-cassette from a marker gene, thus permanently marking the cell type and its descendent cells) would further expand the applicability of this protocol. The protocol presented here and used in the following chapters has the potential to enable major advances in terms of understanding the mechanisms that delineate specification of defined cell populations.

7.2 Embryonic Drosophila neuroblasts across early development

Neuroblasts are pluripotent neurogenic *Drosophila* stem cells that have the potential to give rise to neurons and glia. In the embryo, these stem cells delaminate from an epithelial sheet of cells called the neurectoderm. Groups of cells – the so-called proneural clusters – express proneural genes endowing

cells within the clusters with the *potential* to become neuroblasts. Proneural clusters initiate a selection process based on reciprocal signaling referred to as *lateral inhibition* that selects one cell from each cluster as a neuroblast, whereas the remaining cells will become part of the epidermis (epidermoblasts) (Hartenstein and Wodarz 2013). Lateral inhibition involves Delta-Notch signaling between adjacent cells, where the eventual neuroblast at the arrangement's center inhibits the surrounding cells from adopting a neurogenic fate while reinforcing its own identity (Arefin et al. 2019; Hartenstein and Wodarz 2013; Egger, Chell, and Brand 2008) (for more details see section 5.3.1). In each hemi-segment, a total of ~30 neuroblasts delaminate over a total of five delamination waves across *Drosophila* embryonic development starting at developmental stage 8 and lasting until stage 11. At any point during these stages of early neurogenesis, ~150 to ~840 neuroblasts will be present in the embryonic trunk region to generate a structure known as the ventral nerve cord – a structure that is part of the embryos' central nervous system and analogous in function to the spinal cord in vertebrates. (Hartenstein and Wodarz 2013; Egger, Chell, and Brand 2008).

Proper neuroblast development is under several layers of regulation, both intrinsic and extrinsic. Examples of these regulatory inputs are the expression of spatially-specific neuroblast marker combinations along the anteroposterior and dorsoventral axes (S. Lin and Lee 2012), as well as a temporal cascade of transcription factors that – as far as we know – most of the ventral nerve cord neuroblasts progress through after delamination. This cascades initiates with the expression of *hunchback (hb)*, which initiates *Krueppel (Kr)*, which initiates *paired domain (pdm)*, which initiates *Castor (Cas)* (*Hb→Kr→Pdm→Cas*)(Egger, Chell, and Brand 2008). Additionally, at least Pdm and Cas repress earlier genes in this cascade, promoting the lineage progression. The spatio-temporal identity of a neuroblast affects its transcriptome dynamics and defines the progeny it gives rise to (Grosskortenhaus et al. 2005; Bhat 1999; Weiss et al. 1998; J B Skeath 1998)

Once delaminated, neuroblasts divide successively and asymmetrically: they give rise to self-renewing neuroblasts and ganglion mother cells (GMCs). While the molecular machinery responsible for these asymmetric divisions is rather well-understood and asymmetric deposition of cell fate determinants that either maintain stem cell identity or induce differentiation have been described (Karcavich 2005; Hartenstein and Wodarz 2013; Kaltschmidt et al. 2000; F. Yu et al. 2002), the global gene expression dynamics and how neuroblast identity is transmitted to the GMC is not known (Crews 2019; Egger, Chell, and Brand 2008). Embryonic GMCs do not undergo transit amplification and only divide once producing two daughter cells: either a neuron and a glia cell, or two neurons (Crews 2019). Several neuronal and glial markers are known, but how these specific neuroblast- and GMC-intrinsic factors regulate progeny fate is still unclear.

In this section, I will explore sc-RNAseq data of sorted neuroblasts that capture the five waves of neuroblast delamination. I will focus on:

(i) identification of cellular diversity and cell populations across early neurogenesis,
(ii) understanding the dynamics of these cell populations across development, and
(iii) exploring the transcriptomic signatures that determine neuroblast identity.

7.2.1 Single cell sequencing of neuroblasts reveals distinct subpopulations

In order to validate the feasibility of identifying distinct neuroblasts from a pool of cells sorted based on the expression of the Worniu marker, I started with a pilot-run by collecting and processing embryos corresponding to the first time point (TP1). For TP1, I collected embryos for 1h and aged the collected embryos for a further 3h40, yielding a collection bin of 3h40 – 4h40. TP1 primarily consists of stage 8/9 embryos according to visual inspection of collected, aged and fixed embryos (Supplementary Figure S 1), thereby capturing the first waves of neuroblast delamination. These embryos were fixed and processed as outlined in section 7.1.3 (for a detailed protocol see 9.1.2.2). After sequencing, I used the the PiGx scRNAseq pipeline (Wurmus et al. 2018) for read-quality control, genome alignment, and generation of the digital gene expression matrix. Next, I employed Seurat (v3.1, (Stuart et al. 2019)) for quality control of single-cell parameters, which included setting thresholds that needed to be met to retain cells as part of the data set. Among these threshold conditions were a minimum number of identified genes (500), and a maximum mitochondrial read percentage (10%), as significantly higher percentages tend to indicate cells in distress. Clustering of the remaining 7221 cells for the earliest timepoint (TP1) by the twelve most significant principle components revealed eight clusters (Figure 7-4A). Based on the most distinctively expressed genes per cluster (i.e. highly variable genes, HVGs – for a comprehensive list, see Supplementary Table S 3) and screening HVGs for known molecular and developmental roles, I could discern at least four major populations among the clusters:

7.2.1.1 Population 1: midline/brain

A population characterized primarily by the expression of *castor* (*cas,* the last member of the canonical embryonic temporal transcription factor cascade). According to our prior knowledge, this gene should not be found in the trunk neuroblast at this early stage of development (stage 8/9), which is why I performed *in situ hybridization* for *castor* using appropriately staged embryos. The detected expression of *castor* (Figure 7-4B) indicates that these cells should be ***midline-derived or brain neuroblasts*** as *cas* expression is observed in the midline and in the anterior region of the embryo (Figure 7-4B). This *castor*-positive population is comparatively small, accounting for ~ 6% of all TP1 neuroblasts, which further supports the midline identity of this population as the midline/brain neuroblasts should account for a smaller population than the remaining trunk neuroblasts at TP1. The role of *cas* in the midline remains unexplored, but might be distinct from its described role as the last member of the temporal cascade of transcription factors (for discussion, see appendix 11.1).

7.2.1.2 Population 2: glial precursors

The cluster labeled as ***glia precursors*** is marked by expression of definitive glial marker genes, such as the homeodomain transcription factor *repo* (*reversed polarity*) and the zinc finger transcription factor *gcm* (*glial cells missing*) (Sasse, Neuert, and Klämbt 2015; Altenhein, Cattenoz, and Giangrande 2016)(Figure 7-4C). It should be noted that these cells do not actually represent differentiated glia cells, as proper glia are neither present at this stage of development, nor do they express Worniu. Rather, this cluster contains Wor-positive neuroblasts that have begun to express signature marker genes. Expression of these glial markers might indicate a potential developmental trajectory towards glial fate. *In situ* hybridization indicates that at this developmental stage only one or two cells per segment expresses *gcm* (stained embryo in Figure 7-4C), which (based on its dorsoventral position) belongs to the lateral column-domain of the neurectoderm. The lateral column is known to produce the vast majority of glia in the embryo (Altenhein, Cattenoz, and Giangrande 2016).

7.2.1.3 Population 3: neural precursors

I was also able to identify a population of ***neural precursors*** (***NPs***, Figure 7-4A) based on expression of neuronal marker genes, such as *nerfin-1* (*nervous fingers 1*) and *fne* (*found in neurons*). As with the glia population above, these cells cannot yet represent differentiated neurons; rather, they likely are neuroblasts that adopted a developmental trajectory towards a neuronal fate. For example, the protein encoded by *nerfin-1* (Figure 7-4D) is a zinc finger transcription factor that regulates early axon guidance (Stratmann, Ekman, and Thor 2019) and neuronal maintenance (Froldi et al. 2015), while *fne* encodes a ribosomal binding protein that is known to be present in the cytoplasm of all neurons (Samson and Chalvet 2003). Indeed, I was able to clearly identify neuroblasts that express neuronal markers in a subset of cells in appropriately staged embryos. Expression of *nerfin-1* is readily detectable in neuroblasts to either side of the ventral midline, likely corresponding to the ventral column. (Figure 7-4D).

7.2.1.4 Population 4: neuroblasts

Finally, a group of five clusters labeled as neuroblasts (***NBs***, Figure 7-4A) is recognizable. Together, these clusters form the largest group and have higher levels of the early neuroblast marker *deadpan* (*dpn*) compared to the three other group clusters described above (midline, glial, and neuronal precursors) (Figure 7-4E). *Dpn* is a temporally highly restricted marker that is often used to specifically distinguish neuroblasts from their immediate ganglion mother cell descendants – antibodies against Dpn protein stain neuroblasts (Dpn-positive), but not their GMC progeny (Dpn-negative) (Southall and Brand 2009). It is therefore feasible that these five (5) clusters collectively represent the least mature of the sorted neuroblasts.

The fact that this population contains five sub-clusters is likely driven by factors other than neurogenic identity, such as spatial origin along the anteroposterior or dorsoventral axes of the embryo

and the markers that Seurat identified as HVGs do not allow for a more specific neuroblast classification, but are indicative of spatial identity. For example, *Obp99a*, an odorant binding protein (Wang et al. 2010) is enriched in one of the NB-clusters (Figure 7-4A, red cluster), but has no known role in embryonic neuroblasts. Obp99a is expressed primarily in a restricted neurectodermal pair-rule-like pattern from developmental stage 8 encompassing several (maybe all) neurogenic columns (Figure 7-4F). It is feasible that cells in the red cluster primarily derive from the *Obp99a*-expressing stripes (Figure 7-4F). Similarly, *odd-paired* (*opa*) is a pair rule gene and transcription factor that plays a role in defining segmental identity (Cimbora and Sakonju 1995) and is primarily found in a distinct ventral set of neuroblasts at stage 9 (Figure 7-4G). Expression of *opa* is enriched in a small neuroblast cluster (Figure 7-4A, purple cluster). High *Obp99a* and *opa* expression appears to largely not overlap in my single cell data and the expression domains *in situ* appear largely exclusive (Figure 7-4A, and compare F with G). This indicates that these two clusters are likely to be immature neuroblasts from spatially distinct origins and that it is their spatial character that leads to their clustering apart.

However, in how far this spatial identity shapes the developmental trajectories of the neuroblasts remains unclear and further reinforces the need for an in-depth investigation into what characterizes these cells and how they change along development.

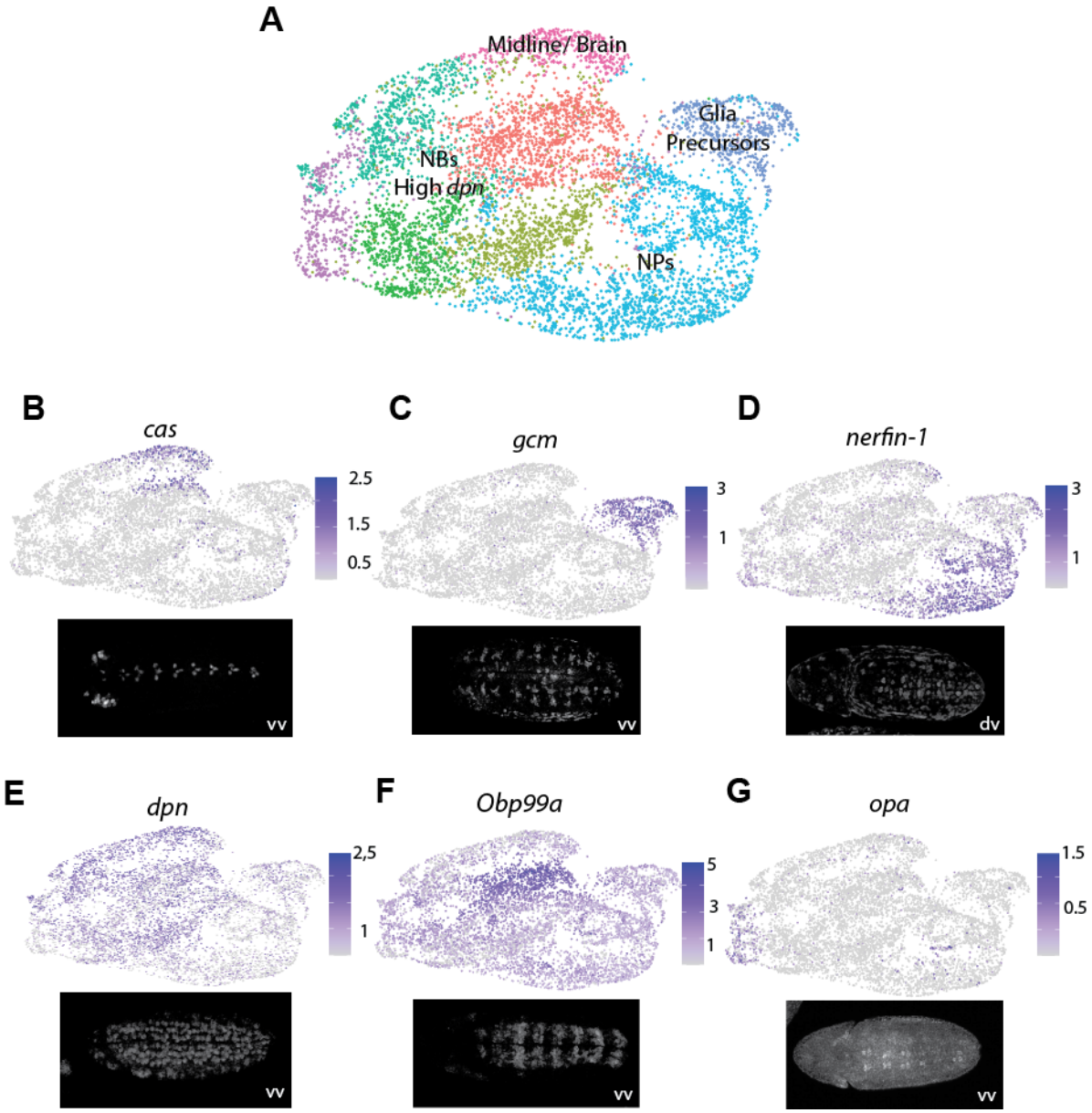


Figure 7-4 scRNAseq of early *Drosophila* neuroblasts

(**A**) UMAP of the transcriptomes of Wor-positive sorted cells extracted from TP1 embryos capturing the first two waves of neuroblast delamination. 4 major populations are identified: neuroblasts ("NBs"), glia progenitors ("Glia"), neural progenitors ("NPs") and midline ("midline/brain"). (**B-G**) Marker validation of cluster identity. (**B**) Expression of cas marks the presumed midline- and pre-cephalic-derived cells, (**C**) gcm the glial and (**D**) nerfin-1 the neuronal precursors. (**E**) dpn labels the 5 NBs clusters. Obp99a (**F**) and opa (ISH performed by Laura Wandelt) (**G**) are enriched in sub-clusters of the NB population. Embryo insets for each panel show the expression pattern of the corresponding gene in situ; anterior left in ventral (B, C and E) or dorsal views (D, F and G).vv - ventral view; dv – dorsal view.

7.2.2 Major neuroblast subpopulations across timepoints

To capture the entire process of delamination, early neuroblast renewal and ganglion mother cell development, I collected embryos in six 1-hour time-bins (TP1 – TP6, encompassing developmental stages 8 – 11, Figure 7-5A). Collections were performed in biological replicates and microscopic assessment of embryo age distributions were as expected (Supplementary Figure S 1).

As outlined in section 7.1.3 and described for TP1 above, pooled time point collections were stained, FACS sorted for Worniu (Wor-positive) and sequenced (Figure 7-2). Microscopic evaluation of FACS-sorted cells indicated high levels of purity in excess of 90%. Sequencing data was processed using the PiGx scRNAseq pipeline (Wurmus et al. 2018) for read quality control and genome alignment (for more details see Materials and Method section 9.2.1). The resulting digital gene expression matrix was explored using Seurat (version 3.1 (Stuart et al. 2019), see Materials and Method section 9.2.2 for more details). In short, I first performed quality assessment of all samples, considering general statistics such as the number of genes per cell and the number of UMIs, as well as the percentage of mitochondrial reads. A given cell was excluded if less than 500 genes were detected, or >10% of the aligned reads were mitochondrial. Initial inspection of the replicates after dimensional reduction and clustering served as quality control to ensure that clustering was driven by cell-intrinsic parameters rather than replicate identity. Since clusters consisted of cells from both replicates without noticeable bias (Supplementary Figure S 2) and since principle component analysis (PCA) of all datasets in pseudo-bulk showed strong co-clustering of time point replicates with temporal identity accounting for the largest of variance (PC1 ~30%, see Supplementary Figure S 2), I concluded that (1) the replicates faithfully characterize the temporal progression of neuroblasts and that (2) my sorting scheme allows for reliable sorting of consistent sets of neuroblasts. Therefore, I merged the replicates of each timepoint and used replicate identity for batch correction where necessary, thereby limiting residual variation introduced by sample preparation (for more details see method section 9.2.2.2). After combining the replicates of each timepoint, all samples were combined into a single digital gene expression matrix. The resulting UMAP plot of the dimensionality reduced transcriptome information (Figure 7-5B) shows populations characterized by their spatio-temporal identity.

Clusters for each of the three major column identities were identified due to enrichment in the characteristic transcription factors (*vnd* for the ventral column, *ind* for the intermediate column, and *Dr* for the lateral column). I also identified neuroblast populations not directly associated with distinct dorsoventral columns in the trunk of the embryo, such as:

(i) *a midline neuroblast cluster* (Figure 7-5B), as determined by the expression of genes such as singleminded (*sim*), described as the master regulator of the midline (Kasai, Stahl, and Crews 1998) and *rhomboid* (*rho*), a protease specifically expressed in the midline (during the stages of this data) that is involved in EGF signaling, growth regulation and cell survival (Sturtevant, Roark, and Bier 1993).

(ii) a *sensory complex progenitor cluster* (Figure 7-5B) due to the expression of specific markers such as *sens* (*senseless*), *ato (atonal)*, and *cato* (*cousin of atonal*), three transcription factors that have been demonstrated to be pivotal for the differentiation of the sensory neurons and development of the peripheral nervous system in general (Singhania and Grueber 2014; zur Lage and Jarman 2010).

The remaining clusters were named according to the expression of the canonical temporal transcription factors (i.e. the lineage time factors *Hb→Kr→Pdm2→Cas*). The intermediate neuroblast

cluster (labeled “intermediate NBs”) is enriched in *pdm2* and the late neuroblast cluster (labeled “late NBs") is enriched in *cas* (Supplementary Figure S 3*)*. Furthermore, the late neuroblast cluster exhibits higher levels of *dpn* expression compared to the neural progenitor (NP) and glial progenitor clusters, thus indicating that neural progenitors and glia progenitors are further along the differentiation path and likely represent the GMC progeny of neuroblasts. This notion is corroborated by increased expression of neuronal and glia markers such as *elav*, *nerfin-1, fne* and *repo*. Accordingly, in multiplex situ hybridization for *dpn*, *wor*, *gcm* and *nerfin-1* to mark neuroblasts, neuroblasts + ganglion mother cells, ganglion mother cells with glial and ganglion mother cells with neuronal character indicate that the assignment logic due to marker behavior in clustered single cells holds true (Supplementary Figure S 6).

When stratifying the UMAP plot by timepoint (Figure 7-5C), the assigned cluster identity (based on enriched genes and lineage time markers) comports with the sampled timepoints: later neuroblast populations (intermediate and late neuroblast clusters, Figure 7-5B) as well as neural progenitors (NPs) are mostly comprised of cells from later time points (primarily TP5 and TP6, Figure 7-5C), whereas the neuroblasts and ventral nerve cord primordium should represent less differentiated cells and are primarily comprised of cells from earlier timepoints (TP3 and TP4, Figure 7-5C).

To further understand the complexity of the populations identified (Figure 7-5B), I decided to extract two populations and to re-cluster them to obtain a more detailed understanding of the cellular diversity within individual neuroblast populations. For this, I focused on the midline cluster, as well as on the neuronal precursor clusters.

I extracted the 1397 cells of the midline cluster, and re-clustered these cells to identify further separation of the midline population. This resulted in ten sub-clusters, of which one stood out specifically by expression of marker HVGs such as *glec* (a carbohydrate binding protein) and *wrapper (*a protein involved in axon-ensheathing and glial apoptosis) (Kearney et al. 2004). Both genes are glial markers, which indicates that this cell population represents the midline glia precursors (Figure 7-5D, dark green cluster); as these cells cluster more closely with the midline rather than the glia precursor population indicates that the midline remains molecularly separate and distinct from the three main neurogenic columns. Another cluster appears to represent the neuronal precursors of the midline (Figure 7-5D, blue cluster), as indicated by expression of genes such as *Kr*, that in the context of the midline neuroblasts labels only the neuroblasts that give rise to neurons (Kearney et al. 2004). The remaining clusters do express known midline markers but further specification was not possible.

Cells of the neuronal precursor population express well known neuronal markers such as *fne*, *nerfin-1* and *fasII,* which supports that they are further along the differentiation pathway. This population refines into ten sub-clusters. By examining the HVGs that drive sub-clustering of neurogenic precursors, it appears that clustering is to some degree driven by anteroposterior origin of the cells.

For example, when considering cells that express the segment polarity genes *en* (*engrailed*) or *wg* (*wingless*), then cluster 8 stands out for characteristically high *en* expression and cluster 5 for high *wg* expression. It stands to reason that cluster 5 *versus* cluster 8 cells are originating in different segmental domains along the anteroposterior axis of the embryo. It should be noted that the typical segment in *D. melanogaster* embryos stretches approximately seven cells, with *wg* labeling the fifth segmental row (row 5, see Figure 5-1C for segment schematic) and *en* labelling rows six and seven (row 6 and 7, see Figure 5-1C). This leaves cell rows 1-4 without unique markers that allow for their identification.

Unequivocal identification of additional anteroposterior rows is instrumental to assign individual neuroblasts to their spatial origin and connect their gene expression programs to their developmental trajectories. Single cell analysis offers a unique opportunity where it should be possible to identify cells of the segment domains. Once identified, we should be able to explore these cells and contrast them on a genome-wide transcriptome level with the anterior and posterior segmental domains. HVGs such as the unstudied computed gene *CG42342* constitute potential candidates for identifying the *wg-negative, en-negative* domains (Figure 7-5E). Indeed, I was able to show that *CG42342* is a distinct and exclusive marker for neurogenic rows 1 and 2, a region in between *wg* (row 5) and *en* (row 6-7) (see section 7.3.1 and Figure 7-8 for details).

It should be emphasized that clustering of single cells is driven by the transcriptomic signature across thousands of genes. These signatures cannot be simply accounted for (in most cases) by a simple feature such as anteroposterior position, as they are the result of a multitude of complex, overlapping components. Clustering aims to represent the transcriptomic similarity among sets of cells according to a limited set of principle components. Therefore, clusters can only ever approximate the real *in vivo* cellular identities, particularly if these identities are shaped by complex, overlapping conditions such as developmental time, lineage time, anteroposterior origin, dorsoventral origin, immediate cellular environment, and specific cell fate decisions made. Nonetheless, the analysis above shows that understanding the drivers of the resulting clustering can allow me to not only assign expected identities to clusters, but also to place cells into their spatiotemporal *in vivo* context.

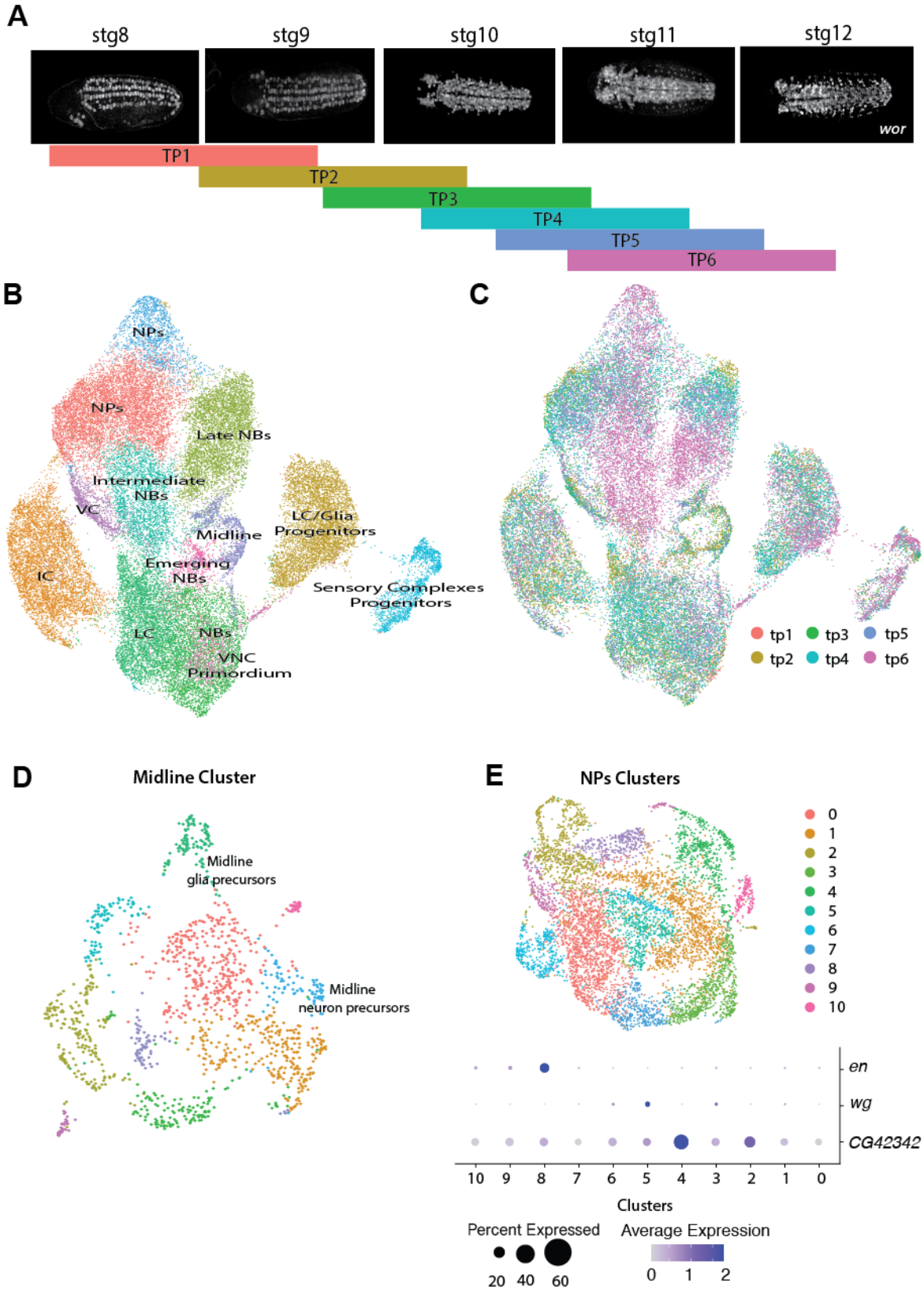


Figure 7-5 Neural progenitors scRNAseq across developmental time.

(**A**) worniu staining of embryos depicting the 5 waves of neuroblast delamination from stg. 8–12. Below: schematic representation of the stages that each timepoint encompasses (for quantification, see Supplementary Figure S 1). (**B, C**) UMAP of all cells from all timepoints labeled by cluster (B) or collection timepoint (C). (**D**) Sub-clustering of midline cluster from (B). (**E**) Sub-clustering of neural progenitor (NP) cluster from (B). The Dot-plot of 3 anteroposterior genes across 10 NP sub-clusters shows enrichments in specific sub-clusters, indicating a degree of clustering by anteroposterior position: *CG42342* is enriched in cluster 2 and 4, *en* in cluster 8 and *wg* in cluster 5.

7.2.3 scRNAseq data across embryonic neuroblasts shows early spatial identity/character give way to cell type identity

Previous work based on dye labeling of specific neuroblast and tracing these cells and their progeny demonstrated that the ~30 neuroblasts per embryonic hemi-segment have distinct developmental trajectories (Schmidt et al. 1997; Torsten Bossing et al. 1996). This work conclusively demonstrated that individual neuroblasts give rise to distinct progeny that will establish specific neural and glial cell types. In fact, the developmental trajectories of most neuroblasts has been described in sufficient detail so that we are not only aware of how many glia, motor-, and interneurons a particular neuroblast gives rise to, but we have a good understanding of where individual neurons will project and how they will connect (Schmidt et al. 1997; Torsten Bossing et al. 1996).

To illustrate the wealth of information available for some embryonic neurogenic stem cells, consider the neuroblast known as "NB4-2" as an example: NB4-2 delaminates from the intermediate column in the 2nd delamination wave (S2) at developmental stage 9 and can be identified unequivocally as it expresses *ind* and *hkb* (*huckebein*) upon delamination, later also *Klu* (*Klumpfuss*) (Doe, Smouse, and Goodman 1988; Yang et al. 1997). NB4-2 will not produce any glia, but it will give rise to at least ten, and possibly as many as 16-22 neurons. Among these neurons are at least four specific motor neurons, the rest being interneurons. The pioneer neuron RP2 is one of these and is among the best-studied embryonic motor neurons in *Drosophila*: RP2 projects its axon anteriorly along the midline and above the intersegmental motor nerve (ISN) into the adjacent segment, before it sharply turns ipsilaterally by stage 15 to contact the body wall muscle fibers 2, 9, and 11. The bulk of NB4-2 progeny are interneurons. As many as 19 have been reported and they differ significantly in terms of axonal projection behavior. Only 2-3 of the interneurons have substantial projections into the contralateral connective. However, most interneurons appear axon less and remain closely associated with larger motoneurons, where they may act to modulate motorneuron activity. The *RP2-sib* cell is one such interneuron that extends a very short axon into the local neuropil, where it presumably modulates neuronal network behavior.

However, despite this phenomenological wealth of information with respect to the types of cells NB4-2 and almost every other neuroblast gives rise to, it is still largely unknown what the specific molecules and mechanisms are that modulate their developmental trajectories so precisely and the open questions range from extremely broad to extremely specific. For example, even after decades of studying early neurogenesis, we still do not understand the rather general phenomenon that neuro-glial decisions are extremely skewed: while the vast majority of embryonic glia derive from the lateral column, a few are

generated by the ventral column, but *none* emerge from the intermediate column. Similarly, the mechanisms for more specific decisions remain unstudied (or at least not-understood), such as the discrepancy between sister neuroblasts such as NB7-1 and NB6-1 (see Figure 5-3). Both NB7-1 and NB6-1 delaminate from the ventral column and within the same anteroposterior domain (both express *en*), albeit temporally offset (delamination waves S1 and S3, respectively) and have very distinct lineages both in terms of cell type and cell morphology (see section 5.3.4 and Figure 5-3).

The underlying mechanisms and individual factors that shape the developmental trajectories of each of the ~30 hemi-segmental neuroblasts are still largely unknown and remain to be elucidated.

To begin tackling this question, I started by isolating the neuroblasts from the single cell data. As mentioned before, *Worniu* is not an exclusive neuroblast marker, but also labels the ganglion mother cells. To focus on the diversity within the true progenitor cells, I selected cells positive for *dpn* (*deadpan*), a marker that is exclusively expressed in neuroblasts and is not detectable in ganglion mother cells. This filter left me with a set of 15 896 proper neuroblasts, which I queried further to delineate and characterize individual, functional groups of neuroblasts.

When analyzing the isolated *dpn*-positive neuroblasts, it became clear that spatial origin and functional specification were major determinants of identity (Figure 7-6A). I could identify several clusters corresponding to the *midline* (1 cluster), the three main dorsoventral columns (8 clusters), a late neuroblast cluster as well *brain* and *sensory complex* clusters (1 each). Cells of these clusters were captured during sorting due to their expression of Worniu, as were two additional clusters for which I have not been able to assign a distinct spatial or temporal identity (Figure 7-6A, clusters #12 and #14) – cells within these clusters express known neuroblast genes, but are not further defined by the expression of specific spatial identity or cell type markers as far as I can tell. However, I was able to positively identify the other 2 clusters based on known markers.

For example, *toy* (Figure 7-6A, C) is a transcription factor important for the development of the central nervous system (Furukubo-Tokunaga et al. 2009) and is expressed predominantly in brain neuroblasts. The sensory complex progenitors express very specific markers, such as *sens (senseless)* and *cato* (Figure 7-6A/C), a transcription factor involved in peripheral nervous system development (zur Lage and Jarman 2010; Singhania and Grueber 2014).

Midline neuroblasts were identified by the presence of midline-expressed genes such as the so-called "master regulator" of midline fate: *singleminded (sim)*, a bHLH transcription factor that has been shown to play a pivotal and instructive role in midline identity and nervous system patterning (Nambu et al. 1991). Other midline markers are present as well, such as the signaling *sog,* and the intermembrane protease, *rho* (Figure 7-6A/C), which modulates EGF signal by cleaving and liberating the membrane-anchored EGFR ligand *spitz* (*spi*) (Sturtevant et al., 1993). Expression of both is maintained

by Sim and together, these genes are crucial components for the midline acting as a signaling center to pattern the emerging trunk central nervous system (Zinzen et al. 2006).

As my main interest lies in characterizing and understanding the neuroblasts in the trunk of the embryo that give rise to the central nervous system, I concentrated primarily on neuroblasts belonging to the column clusters. The bulk of the trunk central nervous system emerges from neuroblasts delaminating from abutting domains in the neurectoderm. To either side of the midline three strips of cells – referred to as *neurogenic columns* – emerge from ventral to dorsal: the ventral column (VC), intermediate column (IC), and lateral column (LC) (also see Figure 5-1). My clustered neuroblast data not only reveals these neurogenic dorsoventral columns, but I detect that their transcriptomes effectively sub-stratify them into (Figure 7-6A):

- three ventral column clusters,
- three intermediate column clusters, and
- two lateral column clusters.

The sub-stratification of columnar clusters is attributable, at least in part, to spatial identity along the anteroposterior axis. For example, ventral column cluster #3 (Figure 7-6A, labeled VC3) is enriched in *wg* expression, which is an anteroposterior marker that labels only one row (i.e. row 5) of neuroblasts along the anteroposterior axis in each segment (see Figure 5-1).

The *late neuroblast* cluster cells was assigned based on the expression of neuronal markers (e.g. *fne* and *vvl*, Figure 7-6C). Expression of such *de facto* differentiation markers indicates that these cells are further along the differentiation path. The expression of glia markers is excluded from this cluster but the equivalent late neuroblasts with glial fate are found in the cluster labelled LC2, where glial markers such as *repo* are enriched, primarily in cells from the latest collection timepoint, TP6, in the right part of this cluster (Figure 7-13A/B, Supplementary Figure S 5).

However, a general temporal trend is readily observable when displaying the cells by time point they are derived from (Figure 7-6B): Temporal progression within the neurogenic column clusters is apparent, as cells from early collections are primarily present towards the left and cells from older collections are concentrated further to the right (see arrows in Figure 7-6B). Apparently, dorsoventral identities converge into a common late neuroblast state as transcriptomic differences due to spatial origin are superseded by emerging cell type identity. (Figure 7-6A, B). Arrows in Figure 7-6A indicated the presumed developmental progression and it should be noted that the lateral column arrow splits to account for a portion of lateral column-derived neuroblasts giving rise to the majority of embryonic glia, as glial markers are predominantly enriched in TP6 cells of the LC2 cluster.

Altogether, my data shows complex stratification of neuroblast identity according to dorsoventral, anteroposterior, and temporal identity. This indicates, that it might be possible to establish a virtual gene expression *grid*, which could allow to confidently assign spatiotemporal neuroblast identities according to gene expression combinations. By extension, we should then be able to query gene expression

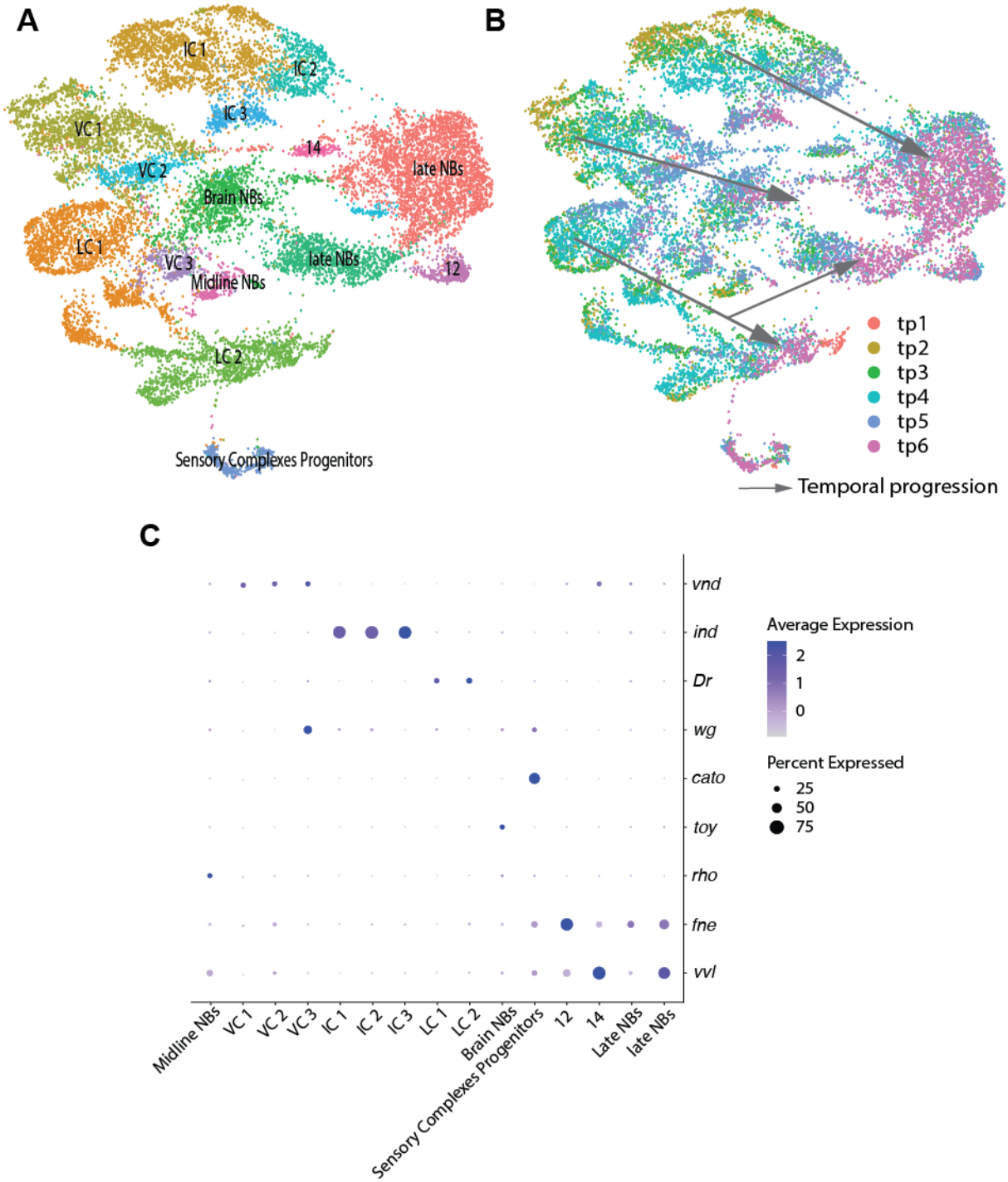


Figure 7-6 Neuroblasts populations reveal high diversity and temporal dynamics.

(**A**) UMAP of all dpn-positive cells from all datasets, shows diversity in the dorsoventral neuroblast populations as each of the column populations (VC, IC and LC) are subdivided into several clusters. (**B**) UMAP of all dpn-positive cells labeled by timepoint shows developmental time progression from columnar populations into a common late neuroblast cluster (progression shown by arrows). (**C**) Dot-plot showing major makers for each population.

complements to better understand the behaviors and developmental trajectories of individual neuroblasts – this is the main impetus of this thesis and will be explored.

7.2.4 Diversity within the GMC populations

Sorting of cells based on worniu expression allowed me to capture not only the NBs but also the intermediate progenitor state, GMCs. Briefly, neuroblasts divide asymmetrically giving rise to a new neuroblast and a GMC. The identity of these cells is defined by the asymmetric deposition of cell fate determinants, apical proteins (e.g. Par Complex, Inscutable) maintain the neuroblast fate and basal proteins (e.g. Prospero and Numb) define the GMC fate (for detailed description see section 5.3.7).

The neuroblast will maintain its proliferative state and be able to divide again generating more GMCs that will give rise to different daughter cells. As development progresses the neuroblast expression profile changes and so do the transcripts that are transferred to the daughter cells directly impacting the progeny they give rise to. For example, the expression of temporal transcription factors is transmitted from the neuroblast to the GMC.

The GMC on the other end will divide into two daughter cells that will differentiate into neurons and/or glia. The two cells that result from a GMC divison (GMCa and GMCb) are distinguish by the activity of Notch, GMCa has Notch activity and GMCb expresses Numb that inhibits Notch. These two cells will have different fates, but Notch is not sufficient to specify which cell it will be as GMCa differentiate into motor-, interneurons or glia. The specific mechanisms that drive cell fate are unknown but it is possible that other factors are asymmetrically transmitted to these cells and are driving their fate.

Considering that no exclusive markers for GMCs have been described, cells with absence of neuroblast specific markers (e.g. *dpn*) while still expressing *wor* were identified as GMCs. This selection is not perfect as scRNAseq data has a significant number of dropouts and the absence of neuroblast markers might be a technical problem and not a biological one. Nonetheless the cells seem to cluster based on the expression of *dpn* and *wor* (Supplementary Figure S 6A/B), this separation also seems to be accompanied by the expression of later genes in *wor*-positive, *dpn*-negative cells, further validating the identity of these cells further along the differentiation path (Supplementary Figure S 6C/D).

When characterizing each individual timepoint, I was able to identify neuronal and glia precursors clusters, these clusters were combined and only cells with absence of neuroblast specific markers (*dpn*) were maintained for further analysis.

This allowed me to identify ~3100 cells as GMCs. After clustering it was clear that they separate into two major populations, glia and neuronal precursors (Figure 7-7A). Within each of these populations there was further separation, based on the top variable genes of each cluster, I could identify populations enriched in terms such as developing glia and neuron development indicating the earlier stages of cell differentiation. There were also more specific terms, I could identify the brain cells and VNC cells for the neuronal precursors. For the glia population more specific cell types could be identified based on the expression of specific cell markers such as *CG6218*, a known marker for neuropile associated glia (Figure 7-7A).

Looking at the developmental time in the GMCs showed a progression along development with the clusters characterized by earlier terms (e.g. developing glia) mapping to earlier times and more differentiated terms such as neuropile glia mapping to the later timepoints. An interesting observation is that glia identity emerges in early times of development, glia specific genes are expressed in early timepoints of the data (TP2, developmental time stg.9/10) while neuronal markers are only seen at TP5 (developmental stage 11) (Supplementary Figure S 7). At this point in development there are no glia or neurons, so there must be a reason why the progenitors are expressing specific markers such as *gcm*, the glia master regulator and *nerfin-1*, a transcription factor involved in axon guidance. These could be potential regulators of GMC daughter cell fate and their presence is required along with the Notch pathway to drive specific cell fates.

Exploring differential gene expression between the two GMCs populations allowed me to identify specific markers, *CG6218* a known marker for some glia subtypes and *NimC4* a phagocytic receptor necessary to establish glia cells phagocytic ability as well as *CG3036* a potential anion transporter which expression has been described in surface glia. For the neuronal precursors, there was an enrichment of *salm* known to mediate neuronal cell fate in the PNS. I also identified *CG8407*, a gene with no known role to be specific to the NPs. Through *in situ* stainings, I was able to show that *CG3036* labels two lateral *gcm*-positive cells per segment (Figure 7-7D), indicating that this gene labels a subset of cells that express glial markers even though its expression is not exclusive as *CG3036* also labels more ventral cells along the same anteroposterior domain (Figure 7-7D). Stainings for *CG6218* indicate that most of the cells positive for this marker overlap with *gcm* and *wor* making it a good marker for the glia precursor population (Figure 7-7C). The *in situ* staining for *CG8407* shows that this gene seems to label all the progenitor cells as it overlaps with most *wor*-positive cells while being enriched in *nerfin-1*-positive cells indicating it is expressed in neuronal precursors cells as predicted even though it is not exclusive (Figure 7-7E).

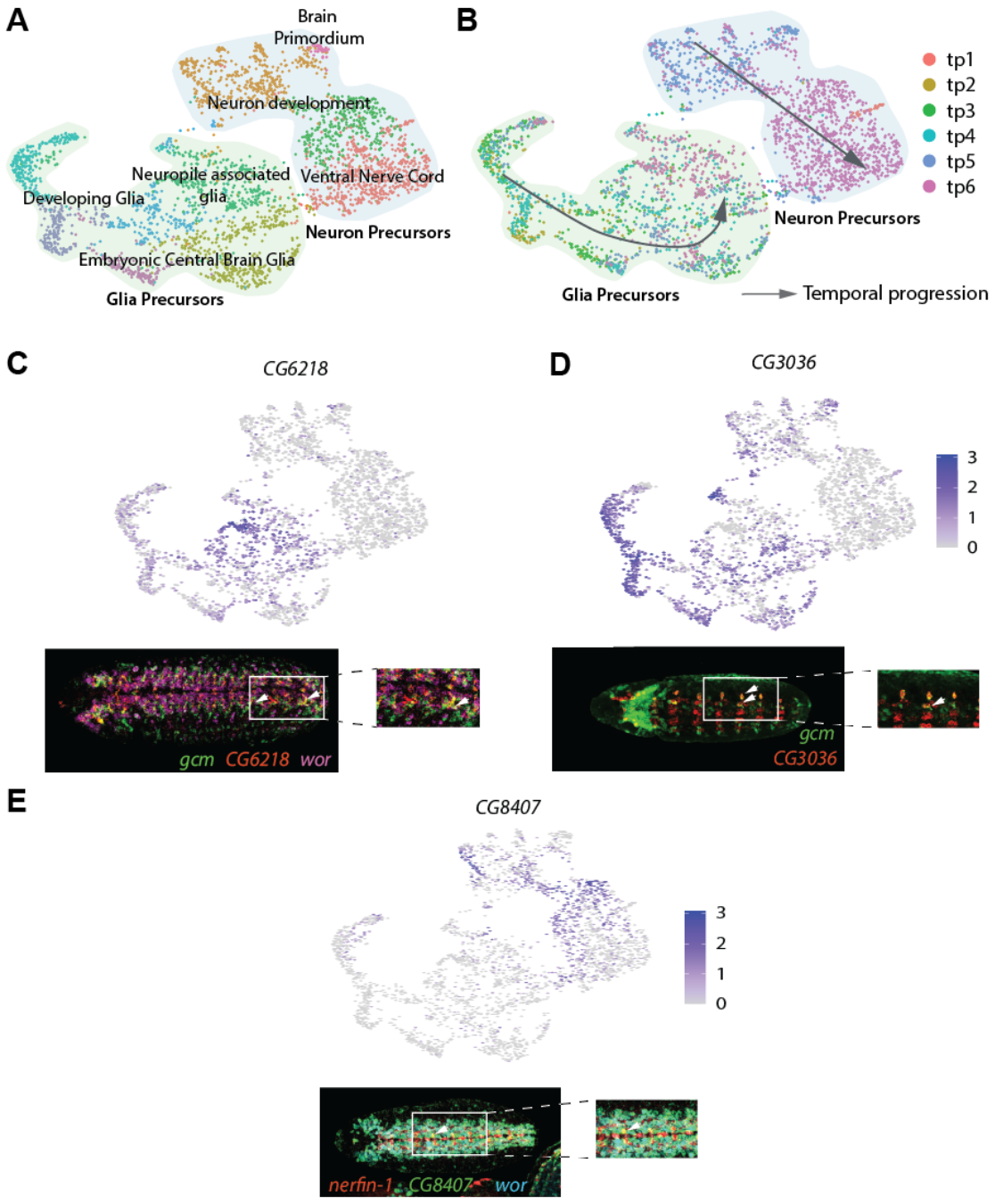


Figure 7-7 GMCs diversity

(**A**) UMAP of GMCs with cluster identity; blue cluster indicates likely neuronal precursors and green likely glial precursors. (**B**) same plot as (A), but labelled by timepoint. (**C**) UMAP of glial ganglion mother cells, relative expression level per cell indicated for *CG6218*, which is enriched in glial precursors. The *CG6218 in situ* hybridization below demonstrates overlap between CG6218 and a subset of cells expressing the glial marker *gcm* in the wor-positive neuroblast/ganglion mother cell population (arrows). (**D**) as (C), but for *CG3036*. *in situ* hybridization shows co-expression of *CG3036* and *gcm* in two cells per segment (arrow). (**E**) as (C, D), but for *CG8407*. UMAP demonstrates enriched expression in ganglion mother cells with neuronal character; *in situ* hybridization confirm co-expression with the neuronal marker *nerfin-1* at stage 11; *wor* expression confirms that the *CG8407*-expressing cells are not yet differentiated neurons. (arrow).

The VNC neuroblasts are divided into a checkboard pattern defined by two spatial axes, the dorsoventral and anteroposterior axis (see Figure 5-1C). The dorsoventral axis has been extensively studied and has been introduced previously (see section 5.2). Briefly, a local transcription, translation and signal exchange between the egg *in statu nascendii* and the overlying follicle cells triggers a serine-protease cascade that results in graded activation of the Toll receptor in the oocyte membrane with peak levels at the ventral side. Toll activation causes degradation of the maternally deposited protein Cactus, releasing the transcription factor Dorsal. Dorsal enters the nucleus, thus establishing a ventral-to-dorsal nuclear concentration gradient. While Dorsal activates genes ventrally along a concentration gradient, it can also act as a repressor, thereby limiting some genes to the dorsal side of the egg where Dorsal remains absent from nuclei. Among these genes is, for example, Dpp (decapentaplegic), which is a ligand for the BMP pathway. The result are two opposing concentration gradients: The Dorsal transcription factor morphogen with peak levels ventrally and lower levels laterally, and a Dpp gradient with peak levels dorsally. Among the target genes of Dorsal is the transcription factor *snail* (a transcriptional repressor), which requires peak-Dorsal levels for activation and restricts several neurogenic genes that can be activated by lower Dorsal levels to lateral regions. Among Dorsal target genes are EGF signaling molecules (e.g. *rho* and *vn*), as well as BMP antagonists (e.g. *sog/chordin*). The neurogenic region therefore emerges in lateral domains within opposing morphogen gradients of Dorsal/EGF ventrally and BMP signaling dorsally. By the time the mesodermal region has invaginated and the ventral-most neurogenic line of cells – the mesectoderm – has come to rest at the ventral midline, the Dorsal gradient is no longer consequential, but the midline has become the EGF and anti-BMP signaling center. By this time, major fate decisions have already been made, as the opposing Dl and Dpp gradients have subdivided the neurogenic regions into four lateral "column" strips of cells: the mesectoderm, ventral column, intermediate column, and lateral column, marked by expression of *sim, vnd, ind* and *Dr*, respectively. These genes are more than simply markers – they are instrumental for the specification of columnar identity of the neuroblasts that will emerge from the column domains. For example, mutations in *vnd* causes an expansion of the intermediate domain, leading to a loss in ventral column-derived neuroblasts and an increase in neuroblasts with intermediate column character (J B Skeath 1998; Weiss et al. 1998).

The other major spatial embryo axis is the anteroposterior axis, where a well-studied gene regulatory network of maternally supplied and localized transcription factors define major 'territories' via the deployment of the so-called gap genes. Gap genes cross-regulate to allow for stripy expression of pair rule genes. Expression of the pair rule genes coordinate the segmentation of the embryonic abdomen into para-segments along *hedgehog (hh)* and *wingless* (*wg*) signaling boundaries (Bhat 1999). Importantly for the emerging nervous system, segment polarity genes divide each segment into smaller

domains (called rows). The best studied segment polarity genes include *wg, hh, ptc, gooseberry (gsb)* and *engrailed (en)*, which have been shown to interact with each other and cross-regulate to allow the tight domains of expression that are required for neuronal specification to emerge. The exact mechanisms through which these genes interact in the neuroblasts have not been described, but it is generally assumed that these mechanisms mirror interactions in the epidermis, where the segment polarity genes regulate each other expression. Wg is expressed in row 5 and activates the Wnt pathway in the adjacent domains; Wg is known to induce the expression of *engrailed* and *hh* in the adjacent domain (*en*-positive and *hh*-positive), and hh signaling in turn induces *wg* expression in its neighboring cells (Chu-LaGraff and Doe 1993; Bhat 1999; Swarup and Verheyen 2012) In the neuroectoderm, *wg* mutants show loss of neuroblasts from the adjacent rows (4 and 6) (Chu-LaGraff and Doe 1993), which highlights the cell non-autonomous effects of row 5 expressed wg.

The combination of both axes results in a checkboard pattern where each segmental region (i.e. *dorsoventral x anteroposterior*) may have distinct signaling and transcription factor inputs that will regulate these regions' expression profiles and set the developmental trajectory of neuroblasts that delaminate – while it is known that spatial origin of a neuroblast is determinative for the developmental fate of its progeny, a proper understanding about the molecular mechanisms that regulate this remain largely unknown. In this section I will explore:

(i) aspects of spatial neuroblast identity as revealed in the single cell expression data,

(ii) new markers for distinct spatial populations,

(iii) differential expression of signaling pathways in the DV domains.

7.3.1 Anteroposterior clustering reveals new domains of gene expression

The anteroposterior domain is divided into subgroups by the expression of segment polarity genes. The known markers for exclusive domains include *engrailed* which is expressed in row 6 and 7. These two rows can be separated based on the presence of *gsb* in row 6, but not 7. Row 5 is labeled by *wg*, but no exclusive markers have been described for rows 1-4.

In order to identify neuroblasts from rows 1-4, I searched for markers not expressed in either the wg or en domains. I first validated the non-overlapping expression of *wg* and *en* in row 5 and 6/7. *Wg* and *en* are exclusively expressed and both are absent from rows 1-4 at all stages of the data (Figure 7-8C). I could then select cells that express either of these genes in all time points and generate a set of *wg*-positive (row 5) or *en*-positive (rows 6 and 7) cells. Differentially expressed genes between these two populations included *CG12496* and *slp2* in the *wg*-positive domain, as well as *inv* and roadkill (*rdx)* in the *en*-positive domain (Supplementary Figure S 8).

One of the markers that emerged for the *en* domain was *CG42342*. Upon validating it by *in situ* hybridization (*ISH)*, I saw that this gene labels a new subset of the AP domain, encompassing rows 1-2

(Figure 7-8C). While *CG42342* is primarily expressed in rows 1-2, it appears to also be expressed at much lower levels in the *en*-positive population, but not in the *wg*-positive population.

It should be noted that *CG42342* was a candidate for expression in the *en*-positive domain at first due to expression in *en*-positive cells to a higher degree than in *wg*-positive cells, but the *in situ* hybridization pattern is clear: *CG42342* is a novel marker for neurogenic rows 1 and 2. The reason for *CG42342's* initial identification in the *en*-positive domain (row6/7) is apparent in the dot-plot (Figure 7-8A), where a small proportion of *en*-positive cells also has detectable levels for *CG42342* and which is not seen for the *wg*-positive population. This may be explained by the complexity of the neuroblasts in later timepoints of development (Figure 5-2F): at the first two delamination waves the neuroblasts are neatly placed in rows and columns, but with additional delamination waves neuroblasts intermingle. It is there for feasible that neuroblasts at the segment interphase express both genes to some degree.

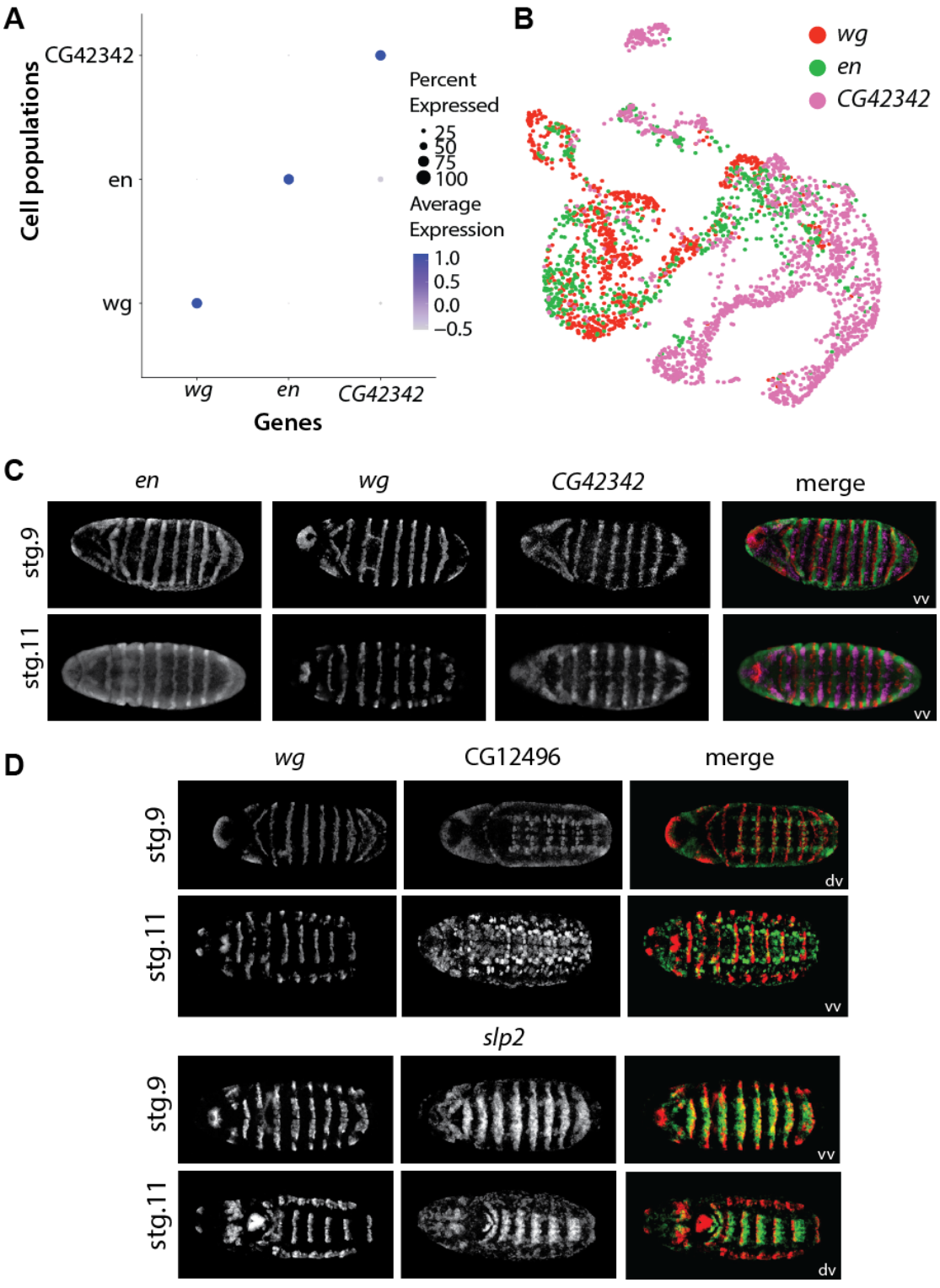
After defining these three domains – row1/2 neuroblasts marked by *CG42342*, row 5 neuroblasts marked by *wg*, and row 6/7 neuroblasts marked by *en*, I compared them by differential gene expression analysis and identified potential new AP markers (Figure 7-8D-F, see Supplementary Table S 4). While differential gene expression between these three populations does not necessarily indicate either exclusive or pan-territorial expression, I could prioritize candidates based on published expression data (Hammonds et al. 2013). For *wg*-positive cells two candidate genes were especially promising, *CG12496* and *slp2*. CG12496 has not yet been intensively studied and has no known function. ISH shows that *CG12496* expression coincides with *wg* staining (Figure 7-8D), but is also expressed in all cells of the lateral column. The pattern of expression is maintained throughout all developmental stages captures in my data (Figure 7-8D). Another gene that emerged as enriched in the wg domain, was *slp2* a transcription factor necessary for the specification of NB4-2 (Bhat, Van Beers, and Bhat 2000). In the embryonic nervous system, *slp2* is likely to play a developmentally distinct role in a subset of neuroblasts, as early segment-wide expression (stage 9) refines to the neurogenic domain by stage 11. Intriguingly, *slp2* at the later stages does overlap with, but is not exclusive to the *wg*-positive row 5; instead, expression also stretches anteriorly, likely into adjacent rows 3/4 (Figure 7-8D).

For the *en* domain, I found *inv* to be highly enriched, its pattern of expression perfectly overlaps with *en* (Figure 7-8E and Supplementary Figure S 8). As *inv* is a transcription factor functionally redundant with *en*, my observation points to redundancy in mediating the engrailed response. Another candidate is *rdx,* a regulator of the hedgehog signaling pathway (Kent, Bush, and Hooper 2006). Rdx likely acts to regulate Hh signaling in the *en*-domain, but potentially also beyond, as it is expressed in a subset of the *en* domain but also in cells of the adjacent domains (Figure 7-8D).

Finally, for the *CG42342* domain, a good potential marker was *drl (derailed)*, an atypical tyrosine kinase receptor, it is involved (in other contexts) in axon guidance (Yoshikawa et al. 2003). During

embryogenesis it is expressed in all cells of the LC and *CG42342* domain. Its expression only emerges in later developmental stages (Figure 7-8F and Supplementary Figure S 8).

I was able to identify new genes that have antero-posterior spatially restricted genes that are enriched in distinct domains of the neurogenic segment. On the one hand, this is significant for my goal of being able to categorize my single cell transcriptomes in terms of specific regions (and in some cases time) of neuroblast origin. On the other hand, each marker is interesting in its own right in terms of what role it may play in regulating neurogenesis. Of note is the identification of *CG42342:* this gene labels a new (non-overlapping) domain of the *Drosophila* neurogenic segment and allows access to a cellular population that was previously inaccessible and further subdivides the segment into domains with known markers (Figure 7-8G).


A
Cell populations
CG42342
en
wg
Percent Expressed
25
50
75
100
Average Expression
1.0
0.5
0.0
−0.5
wg
en
CG42342
Genes
B
wg
en
CG42342
C
en
wg
CG42342
merge
stg.9
stg.11
vv
vv
D
wg
CG12496
merge
stg.9
stg.11
dv
vv
slp2
stg.9
stg.11
vv
dv

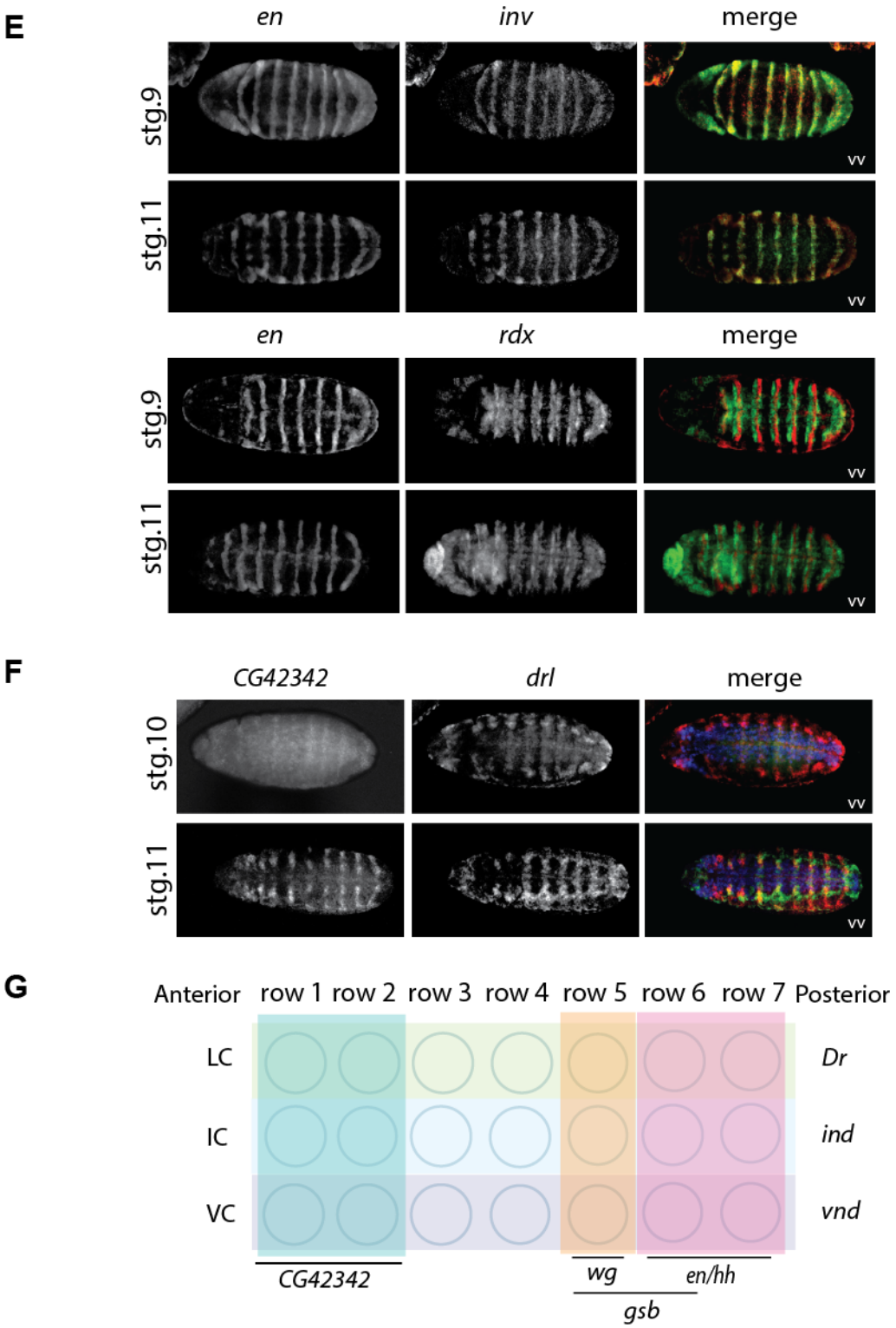


Figure 7-8 Identification of new Anteroposterior domains

(**A**) Dot-plot shows three genes (*en*, *wg*, and *CG42342*) that are expressed largely exclusively in separate cell populations. (**B**) UMAP of cells identified by the expression of the three genes in (A). (**C**) Multiplex *in situ* hybridization staining of *en*, *wg* and *CG42342* shows that they are expressed in mutually exclusive domains along the anteroposterior axis. (**D**) *in situ* hybridization for potential new markers for the *wg* domain. *CG12496* (top) and *slp2* (bottom) are enriched in the *wg*-positive cells. However, the expression domain of *slp2* expands beyond the *wg* expression. (**E**) *in situ* hybridization for potential markers for the *en* domain. Both, *inv* and *rdx* overlap with *en* expression. Expression *inv* and *en* coincide almost perfectly, *rdx* is expressed in the ventral cells of the engrailed domain (corresponding to the emerging nervous system), as well as in more posterior cells. (**F**) ISH stain of genes enriched in the *CG42342*-positive cells. *drl* is co-expressed with *CG42342* in the more ventral rows of

neuroblasts but it seems to be expressed in the entirety of the dorsal column. (**G**) Schematic of neuroblasts and their markers within one hemi-segment of a *Drosophila* embryo. vv- ventral view; dv - dorsal view.

7.3.2 Neuroblast diversity in dorsoventral subpopulations

The division of the neuroectoderm into three columns has been described extensively ((J B Skeath 1998; Weiss et al. 1998) also see section 5.1.2) and is marked by the expression of three column-specific transcription factors: *vnd* in the ventral column, *ind* in the intermediate column and *Dr* in the lateral column. As indicated in the introduction, among the most fascinating aspects of this columnar subdivision are

(i) the fact that each column brings forth a specific set of neuroblasts (Weiss et al. 1998; McDonald et al. 1998), that

(ii) the columns appear to be shaped by dorsoventrally distributed signaling gradients that aid in their patterning (J B Skeath 1998)(which will be explored in more detail in section 7.3.2.2), and

(iii) that this organization appears to be evolutionarily conserved from insect to vertebrates (Weiss et al. 1998).

7.3.2.1 Exploring dorsoventral columnar neuroblast diversity.

In order to identify the neuroblasts belonging to the three dorsoventral columns, I selected cells that were positive for the transcription factors *vnd*, *ind,* or *Dr* across all timepoints (Figure 7-9A). It should be noted that selecting columnar cells based on these markers – while definitive – will necessarily miss many columnar cells simply because of high false-negative rates with respect to marker gene detection. The extent of this issue will be considerable given the drop-out rate for single cell studies (including sequencing using the 10x platform) and the issue will be exacerbated if a particular marker gene is lowly expressed. Further columnar genes such as *vein* (*vn), rho* and *brk* in the ventral column could serve as additional markers, but I elected not to include then in my selection filters as (a) their expression does not always completely overlap with the main column markers and (b) their expression quickly becomes dynamic with advancing development (Hammonds et al. 2013). The main column markers are expressed at distinct levels and *vnd* identified the fewest cells as would be expected, given its low expression level. Overall, I identified 6 595 cells exclusively expressing either of the three columnar marker genes (1 214 *vnd*-positive; 3 260 *ind*-positive; 2 121 *Dr*-positive).

Clustering this subset of trunk neuroblasts shows that columnar identity (i.e. dorsoventral position) is a main driver of clustering as the three marker genes occupy distinct and coherent domains in the resulting UMAP with limited overlap (Figure 7-9A). Some overlap is expected, especially later in

development, as columnar domains produce similar cell types. After clustering, I performed differential gene expression analysis between these three domains and I was able to identify multiple new markers for the dorsoventral populations (Figure 7-9B). *Ilp4* (insulin like peptide 4) (Figure 7-9B) was enriched in the *vnd*-positive cells of the ventral column, while *CG10479* – a computed gene for which no function has yet been described – was enriched in the *Drop*-positive cells of the lateral column. *In situ* stainings clearly validate the exclusive columnar expression of both of these genes. *Ilp4* is exclusive to the ventral column at early developmental stages (Figure 7-9C, stage 9); however, it should be noted that ventral column specificity of *Ilp4* is lost later (Figure 7-9C, stage 11). *CG10479* is exclusively expressed in the lateral column at all developmental stages under investigation, as co-expression of *CG10479* with the most lateral column of *worniu*-positive neuroblasts from stage 9 to 11 demonstrates (Figure 7-9C).

Interestingly, no new exclusive spatially restricted markers could be identified for the intermediate column even though several genes are enriched in this domain, such as *CG10035*. Their expression it is not specific for the intermediate column, but appears to be higher compared to the abutting ventral or lateral column cells (see Supplementary Figure S 9). The intermediate column cells appear to exhibit a generally intermediate transcriptome, between the neighboring ventral and lateral columns, with very few exclusively expressed genes such as *ind*.

When labeling the columnar cells by collection time point, a progression from early to late can be observed in the UMAP plot (Figure 7-9D), illustrating progressive neuroblast maturation. Specifically, the columnar cells appear to arrange along two axes: one axis separating the columnar domains (VC, IC, LC, Figure 7-9A, D), and one characterized by collection time point (roughly early on the left towards later on the bottom right, see arrows in Figure 7-9D). It is also noteworthy that with time, the columnar identities increasingly intermingle, indicating that the transcriptomic differentiation signatures outcompete those due to columnar origin – this is further supported by the expression of later markers such as *fne* and *nerfin-1* (Supplementary Figure S 10).

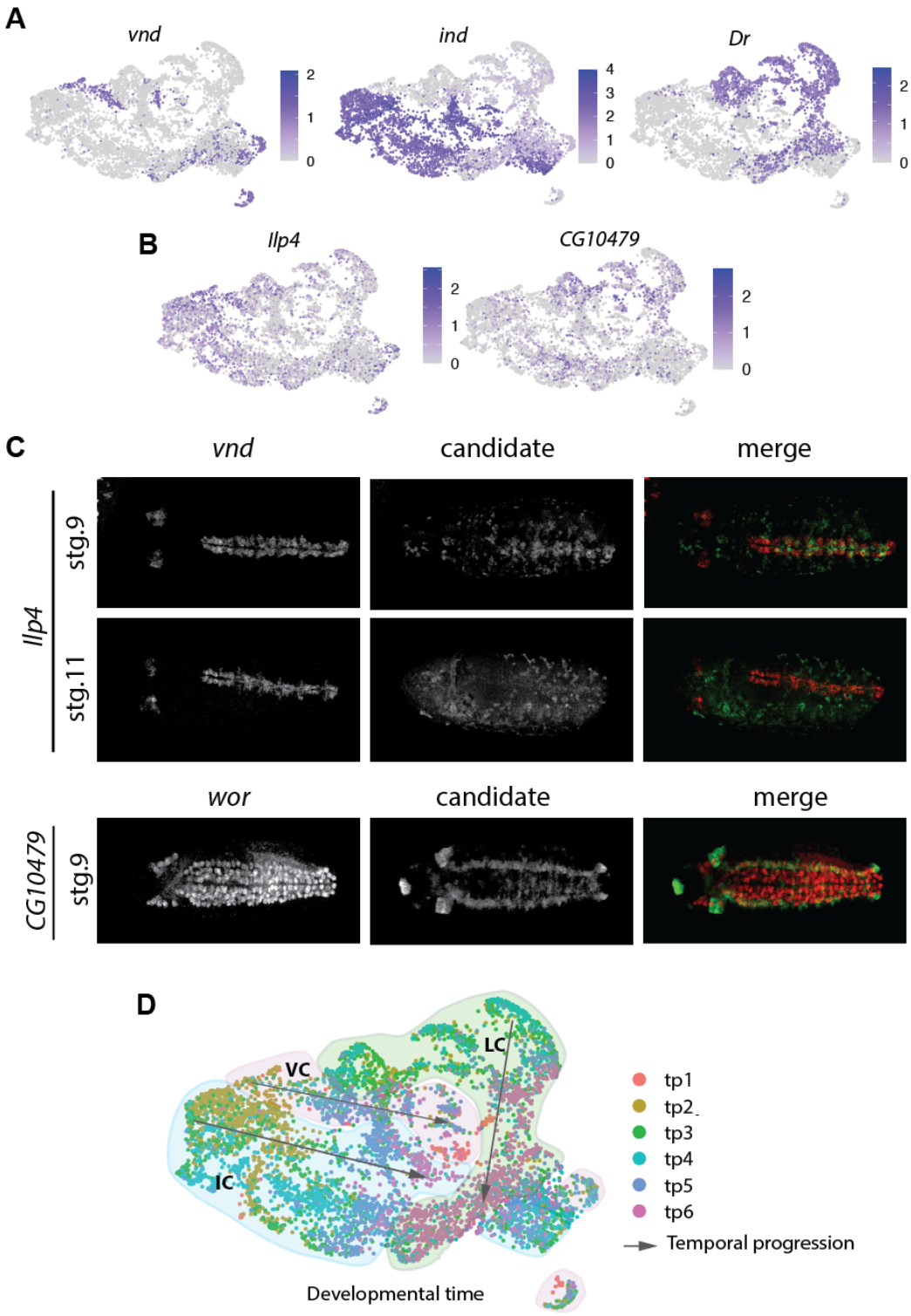


Figure 7-9 Major neural progenitor populations are identified across developmental time

(**A**) UMAP of columnar neurogenic populations (identified by markers vnd (for ventral column, VC), ind (for intermediate column, IC) and Dr (for lateral column, LC)) shows expression of these markers in cells from all timepoints. (**B**) UMAP plots as in (A), showing the relative expression of new dorsoventral markers discovered via differential gene expression between VC, IC , and LC cells. Ilp4 (insulin like peptide 4) is enriched in the ventral column and CG10479 is enriched in the lateral column. (**C**) Staining of the new columnar markers. Markers and stages as indicated; shown are ventral views, anterior left. (**D**) UMAP labeled by timepoint shows a temporal progression within column populations, indicated by arrows.

7.3.2.2 Dorsoventral analysis indicates distinct signaling pathway deployment in neurogenic domains

As introduced previously, signaling pathways shape neurogenic patterning as specific ligands emanate from distinct positions in the embryo with respect to the developing nervous system. One such example is the Dpp signaling pathway. Dpp ligands are secreted from dorsal embryonic regions. Dpp is thought to help induce the dorsal ectoderm, but at least some evidence has been presented to indicate that it also helps pattern the dorsal most neurogenic column (the lateral column)(Esteves et al. 2014). Another example is the EGFR pathway that plays a role in neuronal specification in the *Drosophila* embryonic NBs (J B Skeath 1998). The EGF receptor has four ligands that activate the pathway – Spitz, Keren, Gurken and Vein – and one that inhibits it, Argos. Localized transcription of *vein* and *rhomboid* (a transmembrane protein that helps mediate EGFR signaling by processing Spitz to its active and secreted form) to the most ventral domains of the VNC helps to pattern the neuroectoderm into DV domains (J B Skeath 1998). This means that the ventral midline of the embryo acts as an EGF signaling center with demonstrated effects on the developing nervous system (J B Skeath 1998): mutations in the EGF receptor result in loss of intermediate column neuroblasts, abnormal identities of the ventral column neuroblasts, but with little to no measurable effect on the lateral column neuroblasts (J B Skeath 1998).

The fact that EGF ligands are secreted ventrally implies the existence of a morphogen gradient from ventral (high EGF signaling potential) to more dorsal (low signaling potential) regions in the developing nervous system. Differential morphogen activation could therefore result in a graded- or threshold response of EGF target genes purely based on a neuroblast's presence within the morphogen gradient. However, the recent literature suggests that signaling may often be regulated not only by the presence of ligands, but that precise combinations of ligands and their receptors in addition to transduction modulation can have crucial impact on signaling output (Antebi et al. 2018; Li and Elowitz 2019; Antebi et al. 2017).

Differential expression of EGF pathway components led me to explore it in the DV populations. For this purpose, I selected all cells that could be identified based on DV markers (*sim* for the midline, *vnd* for the VC, *ind* for the IC and *Dr* for the LC)(Figure 7-10C). These cells clustered based on DV identity: one VC cluster characterized by the expression of *vnd,* four IC and three LC clusters.

One IC cluster – labelled as "IC_row5" – is enriched in *ind* and *wg* suggesting these are IC neuroblast of row 5 (Figure 7-10B and Supplementary Figure S 11). The remaining three clusters were classified based on the expression of the temporal transcription factors, "IC_early" expresses *hb* and *Kr*, "IC_intermediate" expresses less *hb* and more *Kr* and "IC_late" express *Kr a*s well as *pdm2* and some *cas* (Figure 7-10B and Supplementary Figure S 11).

Three LC clusters were similarly identified by the presence of *Dr* and the temporal transcription factors: "LC_early" expresses mostly *hb*, "LC_intermediate" mostly *Kr* and "LC_late" *hb*, *Kr* and some *pdm2* (Figure 7-10 and Supplementary Figure S 11).

When a unique DV identity could not be assigned due to co-expression of columnar markers, (Figure 7-10B/C) the clusters were labeled as "Mix" and again separated into early, intermediate and late based on the temporal transcription factors (Supplementary Figure S 11). Cells seem to be arranged not only based on their spatial identity but also on their stage on lineage time.

To explore the dynamics of the EGF pathway in these populations, I investigated the expression of pathway components in the distinct clusters (Figure 7-10A). As previously described, I found *vn* and *rho* to be enriched in the most ventral populations (midline and VC clusters, Figure 7-10B/D). Some factors such as the ligand *edl (*a regulator of the ETS transcription factors *aop* (transcriptional repressor) and the ligand *spitz (spi)* are ubiquitously expressed in the DV populations.

Interestingly, *pointed* (*pnt*) – the main activating transcriptional effector of EGF is not expressed uniformly throughout the neurogenic columns. Rather, *pnt* is enriched in the midline and even more substantially in lateral column neuroblasts. Midline expression might be expected as it has been shown that the transcription factor *pnt* is required for glia-neuron interactions in the midline, a prerequisite of proper formation of the commissures of the embryonic nervous system (Klämbt 1993). Nonetheless, strong lateral column expression of *pnt* indicates that it could be instrumental in gliogenesis in the lateral column. Gliogenesis is primarily a feature of lateral column-derived neuroblasts and to a lesser degree of ventral column-derived neuroblasts. Neuroblasts from the intermediate column do not generate glia, which makes absence there more intriguing.

It is important to note that inferences about signaling based on transcriptomic data have significant limitations: signaling pathways are generally acknowledged to be primarily regulated at a post-transcriptional level, which is simply not reflected in gene expression data. For example, I found *spi* to be expressed ubiquitously throughout neuroblasts of the three columns, but that does not mean that Spi ligand activity is equally ubiquitous. Upon translation, Spi is membrane tethered and to be active it needs to be cleaved by Rho; therefore, though *spi* expression is not localized, effective signaling by Spitz should be due to midline restriction of *rho* expression. The insights we are able to gain from investigating signaling pathway components at a transcriptional level are therefore limited. However, valid hypotheses and conclusions can nonetheless be made, for example if a necessary signaling transducer or a modulator of signaling is found to be absent or present in a particular cell population.

Abundance of the repressive ETS transcription factor *aop* particularly in the midline is intriguing as it could be a mechanism to mitigate or even negate ETS signaling in the midline, i.e. in the very source of EGF signaling. Similarly, I can absolutely confirm the midline as the main EGF signaling center due

to its unique signature of *rho* and *vn* expression, but the ventral column may act to a lesser degree in a similar manner.

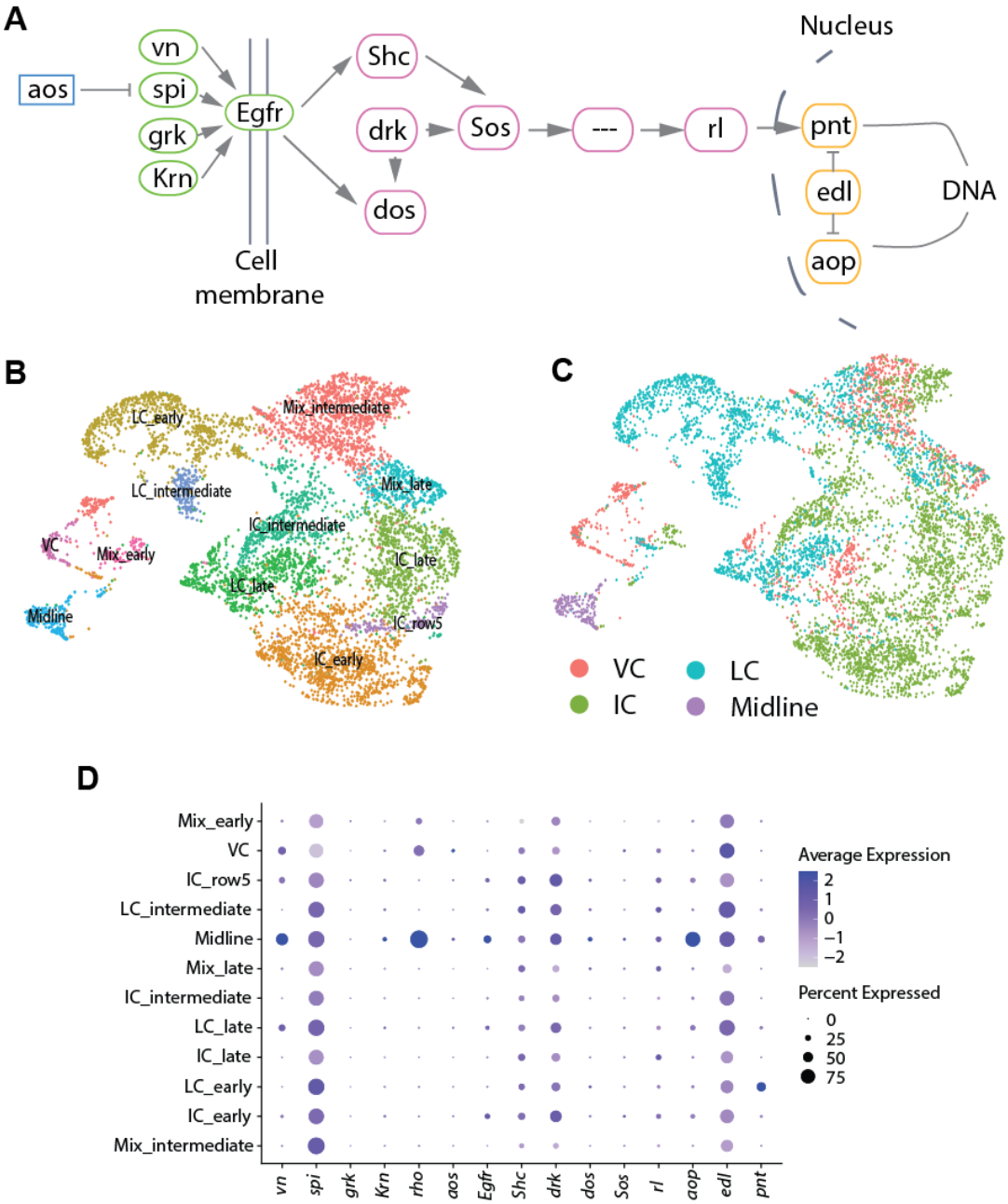


Figure 7-10 EGFR differential activity in the dorsal-ventral populations

(**A**) Schematic of the EGFR signaling pathway. (**B**) UMAP projection of dorsoventral populations including the midline cells with cluster labels by dorsoventral position and developmental state. (**C**) UMAP as in (B), but showing dorsoventral identity by columnar marker genes. (**D**) Dot-Plot showing the expression of the EGFR members (A) among the clusters in (B, C).

7.3.3 Combining information from both spatial axes

In an effort to identify markers for individual neuroblasts, I selected cells that express both dorsoventral (*vnd, ind* and *Dr*) and anteroposterior markers (e.g. *wg*). Considering that *wg* only labels one anteroposterior row of cells, the overlap of the two spatial axes should be exclusive for one or two

neuroblasts. Separating the cells that belong to the distinct columns and express *wg* revealed three well-defined clusters, that separate based on their columnar identity (Figure 7-11A/B).

Investigating the differentially expressed genes that separate these clusters revealed a few promising candidates, such as *CG2865*, a so-far unstudied gene of unknown function. *In situ* staining revealed that *CG2865* expression is restricted to the ventral column and not to a specific anteroposterior domain (Figure 7-11C). The comparison performed here does not exclude genes that are exclusive for the columnar domains as a whole as only one anteroposterior domain was considered. So, while *CG2865* should not play a role in anteroposterior nervous system patterning, it may very well play a role in determining the dorsoventral identity of neuroblasts. Interestingly, CG2865 has been proposed to contain a SERTA domain (Blum et al. 2021), which is a conserved motif commonly associated with proteins that interact with PhD-Bromodomains to regulate transcription.

Another interesting candidate was *Ptx-1*, a transcription factor expressed in the midgut and in parts of the emerging CNS. However, *Drosophila* Ptx-1 so far only has a described role in the gut (Vorbrüggen et al. 1997; Dutta et al. 2015). *In situ* staining showed that *Ptx-1* is specifically expressed in only a single cell per segment, and this cell is, indeed, located in the intermediate column. However, while it does colocalize with *wg* in the more posterior segments (starting at abdominal segment 4)(Figure 7-11D), it is expressed just posterior to the wg-positive rows in more anterior segments. *Ptx-1* expression is also temporally regulated only starting in later stages of neuroblast development (Figure 7-11D). This suggests that *Ptx-1* is a good marker for a specific neuroblast lineage and that is an interesting candidate to proceed with functional analysis.

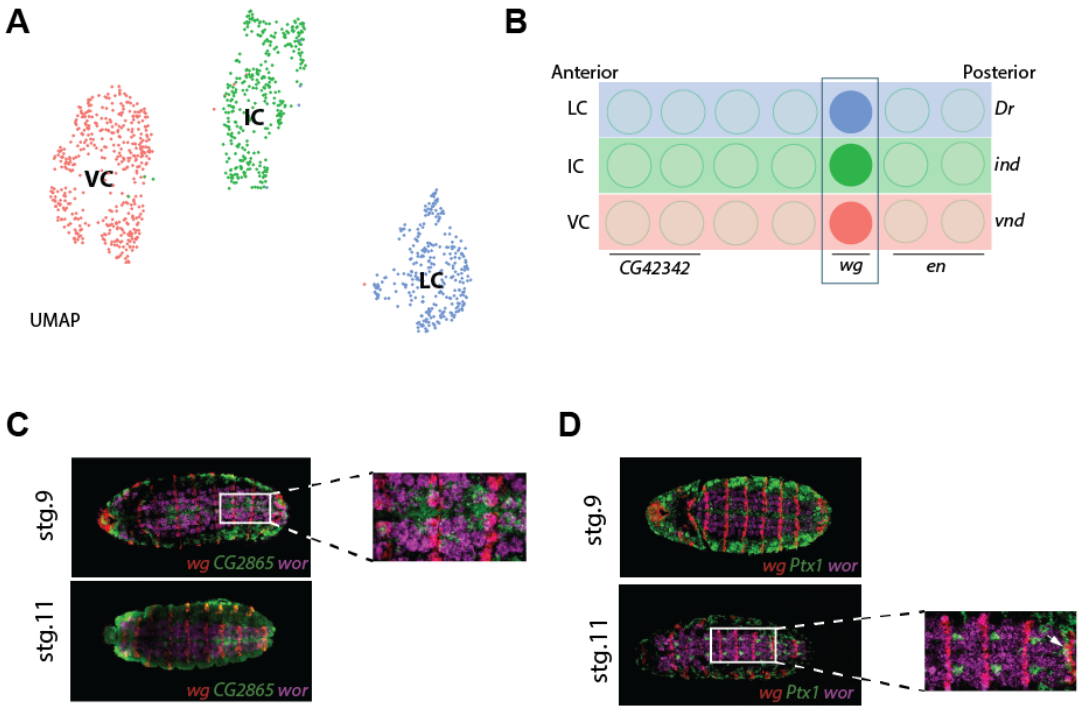


Figure 7-11 Combining anteroposterior and dorsoventral information reveals new and specific neuroblast markers

(**A**) UMAP showing dorsoventral clusters that express wingless; cells drawn from all TPs. (**B**) Schematic highlighting the neuroblast origin of the cells shown in (A). (**C**) Staining of CG2865 confirms enriched expression in the ventral column; though not specific for the wg+ row, expression is specific to the wg-positive ventral column cluster, compared to the equivalent clusters from intermediate column or lateral column. (**D**) Ptx-1 was found enriched in the IC cluster of wg-positive cells (see (A)) and is expressed in one cell per segment but it only overlaps with wg in the most posterior segments (arrowhead); in the anterior segments it seems to be expressed in the engrailed domain.

7.4 Characterization of the Glia Progenitors

The embryonic glia population has been described in great detail (Altenhein, Cattenoz, and Giangrande 2016; Beckervordersandforth et al. 2008; Sasse, Neuert, and Klämbt 2015) including comprehensive descriptions of the neuroblast that give rise to glia cells and the type of glia cells they will become. Glia identity is initiated by the expression of the master regulator *gcm* that will then differentially activate the expression of other glia markers such as *repo*, that further induces differentiation into glia cells and *ttk* to inhibit neuronal differentiation (Altenhein, Cattenoz, and Giangrande 2016).

In the *Drosophila* embryo there are eight neuroblasts that give rise to glia, NB1-1 (in thoracic segments) or NB2-2 (in abdominal segments), NB1-3, NB2-5, LG, NB5-6, NB6-4 and NB7-4 that will originate around 30 glia cells per hemi-segment (Altenhein, Cattenoz, and Giangrande 2016). Markers for each of these neuroblasts have also been described and even though these are not exclusive they can be used to identify these cells (Altenhein, Cattenoz, and Giangrande 2016; Beckervordersandforth et al.

2008). All of these neuroblasts except LG have mixed lineages as they give rise to both neurons and glia (Hartenstein and Wodarz 2013). In that sense, the LG is the only true ***glioblast*** as it gives rise to only glia cells.

According to clustering behavior, the glial population is among the most distinct across all time points (Figure 7-5B), indicating that this population has a strong, stable transcriptional signature. To further explore the progression of the glial population, I combined the cells mapping to a glial cluster at all timepoints to explore further differences within this population. As there is only one ***glioblast*** in the embryonic VNC most of the 5 111 cells in this "glial" population will be bi-potential. To identify the different neuroblast that have a glia progeny, I started by evaluating the expression of the described neuroblast markers, including *mirr* and *klu* (Schmidt et al. 1997; Torsten Bossing et al. 1996). This allowed me to identify five of the eight neuroblasts that give rise to glia cells.

7.4.1 Identification of the NB2-2

NB2-2 expresses *hkb* and *mirr* as it delaminates, then it adds *cas*, *klu* and *svp* (Weiss et al. 1998). With this list of known markers, I was able to identify one of the glia clusters as NB2-2 (Figure 7-12B). From all the neuroglioblasts, NB2-2 is the only one that expresses *hkb* making its identification from a pool of glia progenitors relatively simple (Figure 7-12A/B). The other known markers, *Klu* and *mirr,* are not exclusive to NB2-2, but are still enriched in this population (Figure 7-12B).

After assigning its identity, I explored genes that were enriched in this cluster and found exclusively expressed genes such as *Sp1* (Figure 7-12B). *Sp1* is a transcription factor that has been previously described to be expressed in the brain and involved in the specification of type-II neuroblasts (Álvarez and Díaz-Benjumea 2018). Another marker I find to characterize the NB2-2 cluster is *vein (vn),* an EGFR ligand known to be expressed in the midline and ventral column in early stages of embryonic development. *vn* has also been reported to be expressed in glia cells (Lanoue et al. 2000), though my data indicates that it may be an exclusive marker of the NB2-2 cluster. Beadex (*Bx*) was also exclusive to the NB2-2 cluster. *Bx* is a LIM-only protein that regulates that activity of LIM-homeodomain transcription factors such as *apterous (ap)*, which is known to play a role in axon pathfinding (Kairamkonda and Nongthomba 2014).

Finally, *CG8353* – an unstudied gene that may encode a deaminase involved in small molecule metabolism – was enriched in the NB2-2 population; however, it is not exclusive, as it is also expressed in NB7-4 and in the cluster identified as "Neurogenesis". An *in situ* of *CG8353* revealed that is expressed throughout the VNC and possibly is expressed in higher levels in the NB2-2 (Figure 7-12D, arrow). *Hkb* a known neuroblast marker that is required for proper neuronal specification (T Bossing, Technau, and Doe 1996) also emerged as a marker for NB2-2. *In situ* showed that its expression is restricted to a small group of cells per segment (Figure 7-12D, arrow).

7.4.2 Identification of the LG

The longitudinal neuroblast (LG) has been described as expressing *mirr* (Weiss et al. 1998) which is not enough to identify a neuroblast because *mirr* expression is not exclusive. However, considering that the LG lineage has been described in great detail and that this neuroblast is the only one that originates longitudinal glia (Beckervordersandforth et al. 2008), I identify these cells by exclusive expression of its lineage markers.

Some of the described markers for longitudinal glia in the *drosophila* embryo are

- *retained (retn)*, a transcription factor that is essential for cell shape and migration of longitudinal glia (Shandala, Kortschak, and Saint 2002; Shandala, Takizawa, and Saint 2003),
- *Connectin (Con)* a cell adhesion molecule expressed in some glia cells (Sasse, Neuert, and Klämbt 2015).
- *pnt* is a transcription factor *that* is expressed in the longitudinal glia and is known to be a target of the glia master regulator *gcm. Pnt* has been shown to direct glia differentiation (Shandala, Takizawa, and Saint 2003).
- *Alrm* is an ortholog for the human gene NRROS (negative regulator of reactive oxygen species), expressed in glial cells (Beckervordersandforth et al. 2008).
- *Htl* is expressed in astrocyte-like glia that derive from the LG where it modulates cell growth and migration (Stork et al. 2014).

All of these markers are highly enriched in one specific cluster labelled "LG" (Figure 7-12A/B). To verify the expression of these markers in the embryo neuroblasts I performed *in situ* stainings for *Con* and saw that it is enriched the *CG42342* anteroposterior domain and it overlaps with *gcm* indicating that is probably a glia precursor marker and not exclusive to the LG (Figure 7-12C, arrow).

By querying gene expression specific to this cluster, I identified a potential new marker for these cells – *Tina-1,* which encodes a predicted gamma-glutamylcyclotransferase, previously described as expressed in the head and heart of the adult fly (Leader et al. 2018). Its exclusive expression in the LG remains to be validated.

7.4.3 Identification of the NB5-6 neuroblast

NB5-6 expresses *wg* and *gsb* as it delaminates, it will later add *svp, Klu* and *cas* (Chu-LaGraff and Doe 1993; Cui and Doe 1992; J B Skeath et al. 1995). As NB5-6 is the only neuroblast from row 5 that gives rise to glia the presence of wg is sufficient to identify this neuroblast from a pool of neuro-glioblasts.

From Figure 7-12B, it is clear that the cluster identified as NB5-6 is enriched in the known markers, *wg, svp* and *gsb.* Besides the known markers I found *nub* (also known as *pdm1*, which has been described in some works as a member of the temporal cascade of transcription factors (Grosskortenhaus et al. 2005). If this is nub/pdm1's only role, it should not be enriched in a particular neuroblast as we see here.

Later in development, nub/pdm1 is required for wing formation (Ng, Diaz-Benjumea, and Cohen 1995) but the role in glia cells – if any – it has not been described. However, from *in situ* hybridizations of *wg, gsb* and *gcm,* it is clear that they overlap in only one neuroblast which I conclude to be NB5-6 (Figure 7-12E, arrow).

7.4.4 Identification of the NB6-4

Similar to the previous neuroblast, NB6-4 is the only neuroblast from row 6 that gives rise to glia. This allows me to identify it based on anteroposterior markers. A cell from this domain should express both *en* and *gsb.* The overlap of these genes was only found in one cluster labeled "NB6-4" (Figure 7-12A/B and F). *In situ* hybridization of these genes with the glia marker *gcm* labels only one neuroblast per segment, further supporting my identification procedure (Figure 7-12F, arrow).

7.4.5 Identification of NB7-4

NB7-4 belongs to an anteroposterior domain expressing the pair rule gene *en,* but this row does not express *gsb*, which should allow for identification of NB7-4 by the presence of *en* and absence of *gsb*. There are three clusters in the data that show expression of *en*. One is of NB6-4 as it also expresses *gsb. T*he second cluster was labeled "neurogenesis" because genes in this cluster show gene ontology (GO) term enrichment for *neuronal development*. Neuroblasts with neuro-glial potential produce a GMC that will give rise to one neuron and one glia cell; therefore, these cells may simultaneous express neuronal and glial markers. Hence, cells in the second cluster are likely ganglion mother cells that will give rise to neurons. The third *en*-positive cluster lacks *gsb* expression, as well as marked expression of glial markers, indicating that these cells are NB7-4. In the *in situ* showing *gsb, en* and *wor* expression, a cell positive for *en* and negative for *gsb* can be identified (Figure 7-12F, arrowhead), this is the NB7-4.

This cluster can be further characterized by expression of *sprite (sprt)*. In fact, *in situ* hybridization of the spatial markers shows overlap in only one lateral neuroblast (Figure 7-12G, arrowhead), indicating that only one cell labeled by *sprt* belongs to the en anteroposterior domain, this combined with the lateral position of this double positive cell further validates my identification of this neuroblast as NB7-4. The specific role of *sprt* in the NB7-4 lineage remains to be explored.

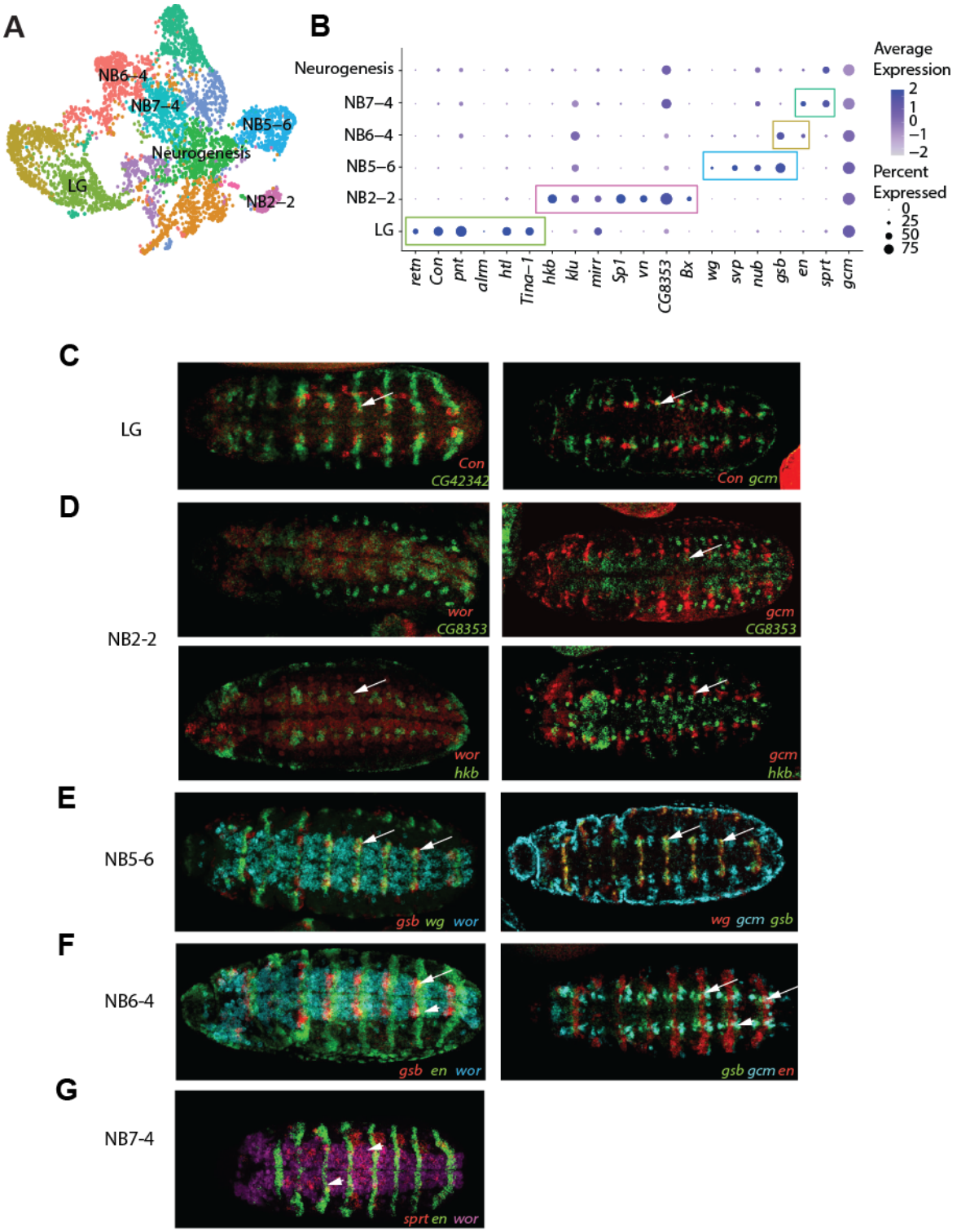


Figure 7-12 Identification of specific neuroglioblasts

(**A**) UMAP projection of cells positive for glial markers reveals clusters of individual neuroglioblasts (indicated by color). (**B**) Dot-Plot showing expression of specific markers for each of the neuroblast clusters identified in (A). (**C-G**) *In situ* staining of predicted markers for the different glioneuroblast clusters:(**C**) *Connectin (Con*) labels 1 – 2 neuroblasts in the CG42342 domain (arrow) – considering their lateral position and overlap with *gcm*, these cells are the LG and NB2-5 neuroblasts (ISH performed by Laura Wandelt). (**D**) The NB2-2 cluster showed enrichment for *CG8353* and *hkb*. *CG8353* shows broad expression at stg.11 (top panel,left) and overlap with *gcm* (top panel, right; arrow; ISH performed by Laura Wandelt), encompassing row 2, from which NB2-2 delaminates. Expression of *hkb* (bottom panel, left) is more specific for a subgroup of neuroblasts including NB2-2 (arrow). Co-stain of *hkb* and *gcm* (performed by Laura Wandelt) shows overlap in only one cell (bottom panel,

right, arrow) (**E**) NB5-6 is identified by expression of *wg, gsb* and *gcm* (arrows). (**F**) NB6-4 is identified by *en* and *gsb* expression as well as *gcm*. Absence of *gsb* and presence of *en* identifies the NB7-4 (arrowhead; performed by Laura Wandelt). (**G**) NB7-4 is precisely identified by a new marker, *sprt*, which only labels one neuroblast at the cross junction of the *en* domain and the lateral column (arrowhead).

Overall, I conclude that clustering of glia cells separates them in large part based on their neuroblast identity. The combination of described neuroblast specific markers together with markers for the corresponding cell lineages allowed me to identify populations of individual neuroblasts. This enables further transcriptome characterization and identification of additional markers. In the long term, this approach promises to make it possible to contrast and compare the average and individual transcriptomes of specific neuroblast identities, which is a promising approach for the identification and characterization of factors involved in neuroblast specification.

7.5 Developmental analysis reveals two distinct temporal axes

As outlined in the introduction (see section 5.3.3), embryonic neuroblasts develop along two partially independent temporal axes: *developmental time* and *lineage time*. Developmental time can be described simply as embryonic age at which the neuroblast is present; this is often measured in terms of "hrs after fertilization" (hAF), or in terms of morphological stages of embryogenesis (Campos-Ortega and Hartenstein 1985) – this temporal component is reflected in my data by collection time point (TP1-6).

Lineage time, on the other hand, is more complicated as it describes a given neuroblast's time after "birth" (i.e. delamination). Upon delamination, a neuroblast sequentially expresses a series of transcription factors, the so-called temporal transcription factors. In the embryo, the expression series initiates with *hb* (*hunchback*), then progresses to *Kr (Krüppel),* then *pdm2 (paired-domain 2)* and finally to *cas (castor)* (Hb→Kr→Pdm2→Cas). While not exclusive, it has been amply described that expression of early temporal transcription factors wanes as the next transcription factor becomes expressed and, indeed, several negative gene regulatory feedback loops among the lineage factors have been demonstrated (Bahrampour et al. 2017; Isshiki et al. 2001; Grosskortenhaus et al. 2005).

As shown in the analysis of *dpn*-positive neuroblasts (section 7.2.3), developmental time is readily discernible in low dimensional space (Figure 7-6B), which indicates that developmental time is a major component accounting for transcriptomic changes in neuroblast development – this is not surprising, but it bears noting that transcriptome dynamics in the isolated neuroblasts will only partially reflect neuroblast-specific development, while a large portion of the transcriptome dynamics will be due to general embryonic development. It has been demonstrated that a neuroblast's developmental trajectory is an amalgamation of spatial origin and developmental stage, as well as the age of the neuroblast (Karcavich 2005; Grosskortenhaus et al. 2005; Weiss et al. 1998; Bhat 1999)– which is to say that neuroblasts delaminated in the same relative position but in different delamination waves may have

strikingly different developmental outcomes. To better understand the contribution of lineage time in neuroblast progression, I aimed to disentangle developmental from lineage time. I reasoned that this might be feasible in a neuroblast population that is coherent and related. One population that fits these requirements are neuroblasts with glial potential.

7.5.1 Temporal progression of the glial precursors

As just described, neuroblasts and the ganglion mother cells they give rise to develop along two temporal axes – developmental time and lineage time. The lineage time transcription factors are transmitted from the neuroblast to the ganglion mother cell, where they help determine the fate of ganglion mother cell progeny. For example, *hb* is the first lineage factor expressed in a newly born neuroblast. The first asymmetric division will give rise to a ganglion mother cell and a self-renewing neuroblast, but while that neuroblast will exchange *hb* for *Kr* expression, the ganglion mother cell remains marked by *hb*, which plays a role in determining the development of the neurons and glia that the ganglion mother cell produces.

It has been shown that mutations in lineage time transcription factors directly impacts neuroblast progeny. For example, mutating *hb* leads to the loss of the specific neurons and glia that are derived from *hb*-positive ganglion mother cells, while *hb* misexpression results in an increase in the number of corresponding neurons and glia (Isshiki et al. 2001). Hence, the superimposition of lineage and developmental time produces a rather complicated situation in terms of neuroblast- and ganglion mother cell states in the developing embryo: while almost every neuroblast will progress through the expression cascade of lineage factors, at any given developmental time after delamination phase S1 there will be a diverse array of neuroblasts expressing different lineage transcription factors (Figure 5-2E/F). Moreover, the activity window for each factor varies significantly between neuroblasts, while some divide once while expressing *hb,* others divide two or three times (Averbukh et al. 2018).

To better understand how the temporal axes characterize the glia population, I started by investigating how developmental time is distributed throughout distinct clusters. For this purpose, the cells in the glia UMAP plot were labeled by their correspondent timepoint (Figure 7-13B).

To discern if there is an obvious relationship between lineage time and clustering, I labeled the cells based on expression of the temporal transcription factors (when multiple factors were expressed the one with highest normalized expression was selected) (Figure 7-13A). Cells positive for *hb* expression cluster together. Similarly, *Kr* and *pdm2* are expressed in the same regions of the UMAP plot and *Cas* is enriched primarily in the bottom right region of the same plot.

The analysis above underscores the pivotal role that these temporal axes are playing in neuroblast progression. To further explore the transcriptomic changes throughout the earliest stages of neurogenesis, I analyzed the neuroblast data using *Monocle2* (Trapnell et al. 2014). *Monocle*2 has a

function that allows for ordering cells in so-called pseudotime – an artificial construct that orders cells using the assumption that global gene expression changes gradually over time. *Monocle2* uses a machine learning approach (*Reversed Graph Embedding, RGE*) to extrapolate a principal progression graph that accounts for transcriptomic differences in the single cell data by constructing a minimal path in high-dimensional space (the dimensions being gene expression magnitudes). Cells are then placed along this graph according to their individual transcriptomes. This strategy of ordering single cells in pseudotime can ultimately allow identification of terminal and intermediate cell states, but also the gene expression dynamics that characterize cellular progression along the path(s).

The pseudotime graph constructed by *Monocle2* successfully managed to place cells according to collection time course – the pseudotime trajectory follows developmental time (Figure 7-13C, developmental time flow indicated by arrow from left to right), starting with earlier timepoints and ending in several termini. Some branching is observed, but differential gene expression analysis indicates very few genes to be expressed in a branch-specific manner. GO term analysis of the few distinct genes did not yield clear indicators and no specific identity could be assigned, cells are placed along pseudotime according with time and not the cluster (cell identity) they belong to (Supplementary Figure S 12). Pseudotime places cells along developmental time (Figure 7-13C), but also along lineage time as a progression from earlier to later temporal transcription factors can be observed (Figure 7-13D). It is also noteworthy that early neurogenesis factors such as *dpn* and *wor* are present throughout pseudotime with levels declining towards the end, when these cells are moving from a proliferative progenitor state to the beginning of differentiation.

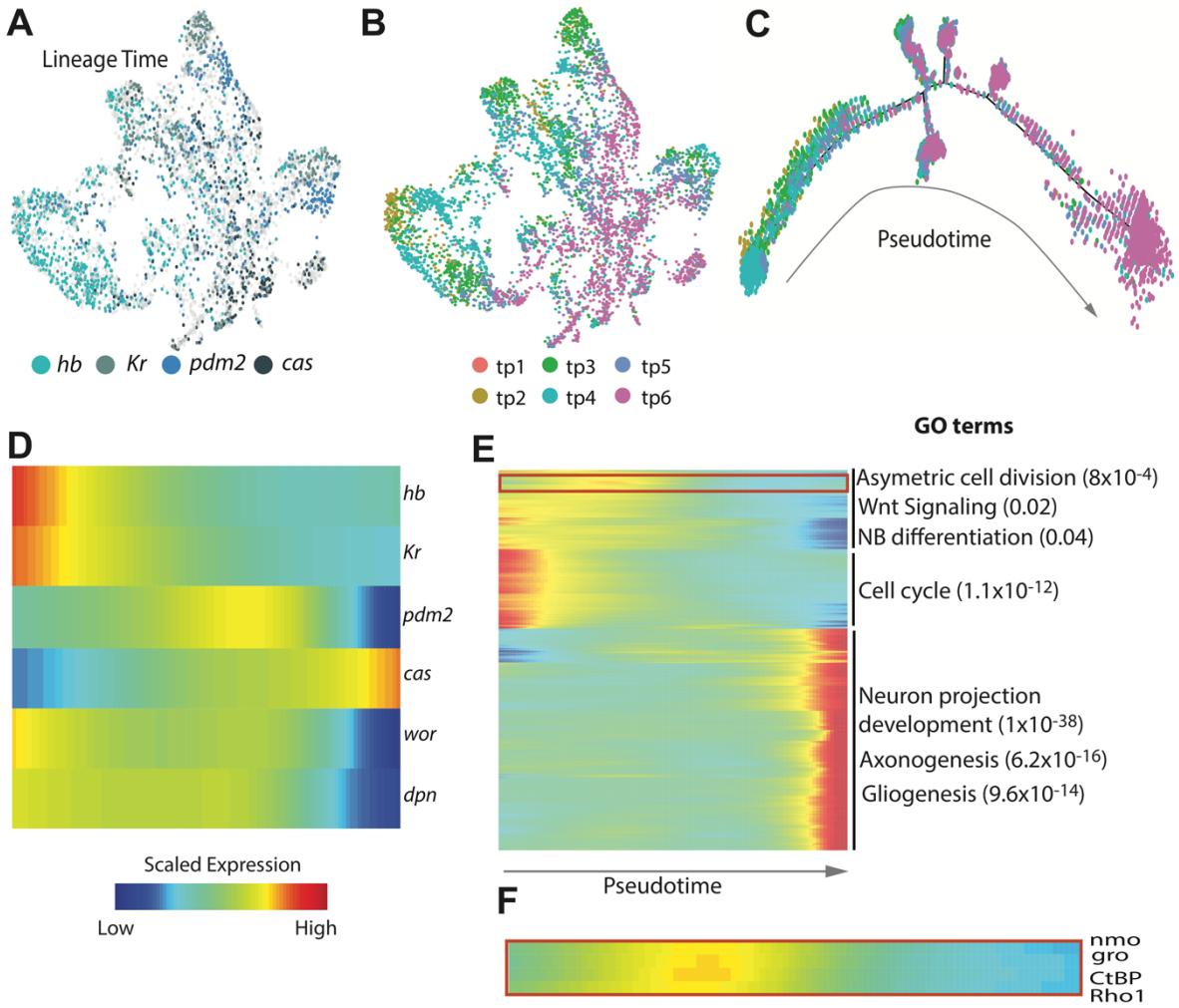


Figure 7-13 Glia progenitors in time

(**A**) UMAP labelled by expression of the temporal cascade of transcription factors shows lineage time progression across clusters (also compare with cluster calling in Fig 7-12). (**B**) UMAP of all cells expressing glial markers (selected by extracting the glial clusters from all timepoints) shows clusters with enrichment for differentiated glial cells. (**C**) Pseudo time plot of the glia population independently produces a progression from early to late time points; color coding as in (B). (**D**) Heatmap showing normalized expression of temporal lineage markers across pseudotime confirms sorting according to developmental progression. Note that expression of the general neuroblast markers dpn and wor wanes with pseudotime (**E**) Heatmap showing highly variable genes and their normalized expression profiles across pseudotime. GO-term enrichment across expression blocks is indicated on the right; enrichment p-values in parenthesis) (**F**) A closer look at the pseudotime distribution of differentially expressed genes reported to play a role in Wnt signaling.

Plotting the genes that significantly vary along the main pseudotime trajectory reveals sets of genes that are primarily expressed early, in the middle, or late along pseudotime. (Figure 7-13E). GO term analysis on these gene blocks revealed an enrichment in *cell cycle*-associated genes in the beginning of pseudotime. In contrast, genes associated with Wnt (*wingless, wg*) signaling appear to be enriched slightly later around the middle of pseudotime, as are genes involved in asymmetric cell division and neuroblast differentiation more generally. A large block of genes late in pseudotime is indictive of neuronal projection / axonogenesis, and glial differentiation, as might be expected (Figure 7-13E).

The block of early-enriched genes includes genes such as *Rcc1, Cdk2* and *Orc1*, all of which have been implicated in cell cycle regulation. As gene expression programs driving either proliferation or differentiation are often temporally separated (Crews 2019; Karcavich 2005), this indicates that early neuroblasts are primed for proliferation, thereby giving rise to ganglion mother cells and self-renewing neuroblast daughter cells. This also feasibly explains the lack of pseudotime branching at the earliest stages.

Interestingly, among the genes more strongly expressed around the middle of pseudotime and associated with the Wnt signaling pathway (Figure 7-13E) are *nemo* (*nmo*), *groucho (gro)* and *CtBP* (Figure 7-13F). The genes *nemo*, *groucho* and *CtBP* are part of the canonical Wnt pathway, which is activated by the Wg ligand (Supplementary Figure S 14A). Their enrichment indicates that *pan (pangolin)* may be acting as a transcriptional repressor, inhibiting the transcription of Wnt pathway targets – this mechanism has been described in the ectoderm (Bhat 1999) and *pan* is indeed detectable throughout the neuroblasts assessed here, but the effect of the Wnt pathway on embryonic neuroblasts has not yet been specifically studied. A specific subset of neuroblasts (i.e. row 5 neuroblasts) transcribe *wg,* which is secreted and can signal to the adjacent anteroposterior rows (Supplementary Figure S 14C). It has been described that Wg is not only important for the identity of neuroblasts belonging to row 5 (Bhat 1999)(see introduction section 5.1.1 for more details), but also for the formation of neuroblasts from row 4 and 6. In fact, *wg* mutants fail to form neuroblasts from rows 4 and 6 (Bhat 1999). The effect in row 4 neuroblasts is at least partially due to the loss of *slp1* and *slp2*, both downstream targets of Wg. Loss of function mutation in *slp* reproduce the *wg* mutant phenotype with respect to NB4-2 (Bhat, Van Beers, and Bhat 2000). For row 6 neuroblasts, the effects of *wg* mutations might be due to changes in the Hh pathway, as in both, *hh* mutants and in *wg* mutants, row 6 neuroblasts fail to form, arguing that these pathways interact to regulate the formation of row 6 neuroblasts (Bhat 1999). It is therefore feasible that inhibition of Wnt signaling as implied by the pseudotime trajectory may be specific for a subpopulation of neuroblasts (Supplementary Figure S 14A-C).

Most glia cells belong to the rows 1 and 2, where Wnt signaling is not active. This would explain the enrichment in transcripts for genes that inhibit the Wnt pathway (*nmo, gro* and *CtBP)* and it would delimit *en* expression in rows 1-2 acting to regulate anteroposterior separation and consequently modulate neuroblast identity. When investigating the expression of the major components of the Wnt pathway in the different anteroposterior domains (wg-positive (row 5), en-positive (rows 6-7) and CG42342-positive (rows1-2)) it is clear that the Wnt pathway inhibitors found in this pseudotime trajectory are significantly enriched in the CG42342 domain (Supplementary Figure S 14B). This further validates that the inhibition seen in the neuroglioblast population relates to their spatial identity (Supplementary Figure S 13).

However, it should be noted that neither CtBP, nor Gro are exclusively involved in Wnt signaling. Both proteins are more correctly classified as general transcriptional co-repressors that mediate negative gene regulation upon recruitment by sequence specific transcription factors. Gro, for example, has been invoked in *Decapentaplegic* (*Dpp*, aka. TGFβ, SMAD, or RTK) signaling, Notch signaling, as well as regulation by gap genes. Hence, their enrichment in the central pseudotime span may indicate a broad impact on gene regulation generally. A signaling link is further supported by high *nemo* expression over central pseudotime-spans – *nemo* is a Ser/Thr-kinase that has also been linked to Bmp and Notch signaling (Zeng et al. 2007; Verheyen et al. 2001).

Importantly though – and it is unclear if this function is signaling-related or not – *nemo* has also been linked to planar cell polarity and asymmetric cell divisions. *Nemo* is not unique in this respect. The GO terms "*asymmetric cell division*" and "*neuroblast differentiation"* are significantly associated with genes such as *pros*, *lnR, cnn, brat, Ptp69D, trn* and *lola* for example (Figure 7-13E), all of which are enriched during central pseudo time. Both GO terms are intriguing with respect to neuroblast development, as they relate to processes that are integral to neuroblast biology: neuroblasts divide asymmetrically into a large daughter neuroblast (self-renewal) and a ganglion mother cell. While the ganglion mother cell divides only once more to produce neurons and glia, the neuroblast undergoes several more asymmetric divisions, giving rise to consecutive ganglion mother cells as described (see section 5.3.5). Therefore, the genes identified here as enriched in central pseudotime warrant further study with respect to their specific roles in driving neuroblast proliferation and specification.

In contrast to the intermediate pseudotemporal period, GO terms such as *axonogenesis* and *neuron projection development* are associated with genes expressed highly along the later pseudotemporal trajectory (Figure 7-13E). This may suggest a support cell behavior among these neuroblasts with glial potential: in the classic vertebrate model glia primarily act as support and regulators of neuronal development and function, a role that has not been conclusively demonstrated in the *Drosophila* embryo (Sasse, Neuert, and Klämbt 2015). The GO terms *gliogenesis* and *axogenesis* were also enriched at the end of the pseudotime trajectory, which might indicate glia migration and projection guided by similar molecular cues as axon pathfinding (e.g. Netrin)(Sasse, Neuert, and Klämbt 2015).

Taken together, the enriched terms of this pseudotime trajectory show a cellular progression from division and proliferation of a progenitor state to the beginning of differentiation and migration of glial cells. A substantial number of genes is enriched across pseudotime progression, which deserve further exploration (for complete list of genes see Supplementary Table S 5). These genes may represent critical nodes in the regulatory network facilitating neuroblast specification and progression into glia and neurons.

7.6 GENE REGULATORY NETWORKS

To better understand the transcriptomic dynamics of the neuroblasts, it is important to not only look at which genes are expressed in each cell, but also at how these genes come to be expressed (i.e. *how are they regulated?*), how do they interact (*what do they regulate?*) and how they can be determining cell differentiation and fate (*what are their molecular roles?*). In order to explore gene regulatory networks in the neuroblasts, I employed the bioinformatic tool SCENIC (Aibar et al. 2017). This tool identifies co-expression modules (groups of genes that are expressed together) among single cell transcriptome data and evaluates these models against the expressed set of transcription factors and their expression signature.

The aim of SCENIC is to infer transcription factor inputs driving expression modules by exploring the genomic sequences surrounding each gene set member for enrichment of DNA motifs. It looks for transcription factor binding motifs that are enriched in the gene set, around the transcription start site (TSS) of the genes in the set. For this purpose, it uses a database of genome wide rankings for each motif. The motifs are then associated with a transcription factor (in the gene set) and the ones with a high normalized enrichment score (NES) are kept.

The gene sets that pass the NES threshold are identified as a 'regulon' of that specific transcription factor. Once regulons – expression cohorts of genes with a predicted regulatory signature – are identified, I can visualize where these 'regulons' are expressed, which is to say if these regulons represent a cluster-specific, a temporal, or a spatial signature in the neuroblast dataset. In order to find specific gene regulatory networks of the neuroblasts, I applied SCENIC to the cells with high *dpn* expression (Figure 7-6A) (for more details see Materials and Methods section 9.2.6). This allowed me to identify several regulons and their associated transcription factors (Figure 7-14).

For example, E2f1 is a transcription factor that has been reported to promote the cell cycle by activating genes important for the G1/S transition. The E2f1 motif was found to be a potential regulator of a cohort of 451 genes that are similarly expressed. Aside from *E2f1* itself, another member of this regulon is the *CycE,* a cyclin that is known to control the cell cycle during nervous system development (Crews 2019; Berger et al. 2005). The E2f1 regulon appears to be active in most *dpn*-positive neuroblasts (which are cycling cells) (Figure 7-14 A, B), which agrees with previous reports about this factor's role in the development of cycling cells (Duronio et al. 1995). The lowest activity of the E2f1 regulon is detected in the sensory complex progenitors, midline neuroblasts and LC2 neuroblasts. Though the roles of E2f1 and CycE in sensory complex neuroblasts and midline progenitors has not been investigated specifically, it is known that these cells do not divide as frequently compared to most other neuroblasts: midline and sensory complex neuroblasts divide in a type 0 pattern (i.e. they give rise to two daughter cells without generation of intermediate progenitor states) (Kearney et al. 2004; Hartenstein and Wodarz 2013), whereas most trunk neuroblasts exhibit the more typical type1 division pattern, where ganglion

mother cells are produced alongside self-renewing neuroblasts. The reduced proliferative potential in type 0 neuroblasts compared to type I neuroblasts might be mediated through reduced E2f1 and CycE levels. Regarding the LC2 cluster, a similar argument can be made. As previously described, the LC2 population represents late neuroblasts that approach their terminal division, thus accounting for the lower levels of gene expression for components of the E2f1 regulon. In fact, the LC2 population is enriched for *dacapo* expression, which is a kinase inhibitor that is required after that last mitosis for terminal differentiation of post mitotic cells ((Crews 2019; Baumgardt et al. 2014),Supplementary Figure S 15).

Three other regulons were identified, which are identified by the transcription factors *wor, l(1)sc,* and *sna* are expressed in the early populations (labeled by their spatial identity) and their expression across clusters is highly similar (Figure 7-14 A). Furthermore, these regulons share a similar set of member genes. The reason is apparent when comparing the PWM motifs of the three transcription factors: all three have been reported to bind highly similar sequences (see Figure 7-14A, right), especially Worniu and Snail, which are in fact members of the same Zn-finger transcription factor protein family. Strong expression of the entire regulon gene sets, Worniu, Snail and l(1)sc, can be observed in the clusters primarily containing early neuroblast (Figure 7-14A, B) and all three transcription factors may be important regulators of gene expression within these clusters. For example, l(1)sc has been described as a regulator that is expressed in the proneural clusters and as having a role in selecting neuroblasts (Bertrand, Castro, and Guillemot 2002; González et al. 1989). The genes *sna* and *wor* are expressed in neuroblasts as well and have been reported to play a role in neuroblast identity and asymmetric cell division (Bahrampour et al. 2017); therefore, they should be expressed in the early cells that are dividing asymmetrically originating neuroblasts and ganglion mother cells.

More population-specific regulons were identified as well: the Gcm- and the Ocelliless (Oc)-regulon. The Gcm regulon characterizes the LC2 cluster with high specificity. As already indicated, my analysis leads me to conclude that neuroblasts in the LC2 cluster are relatively far developmentally, where cell cycle is being inhibit by *dap* and the neuroblast proliferation is ending. LC2 neuroblasts should have a more differentiated profile compared to earlier clusters and the *de facto* glial marker *gcm* is one such marker. Though the lateral column does give rise to neurons as well as glia, the vast majority of glia in the embryo trunk stem from lateral column-derived ganglion mother cells. It is likely that LC2 neuroblasts correspond to the cell population that will form gliogenic ganglion mother cells. Further analysis of this regulon indicates several reported Gcm targets as subject to regulation by Gcm in the early embryo as well: among them is *pnt* (*pointed*, discussed above), *repo (reverse polarity*, an often-used glial marker known to reinforce glial fate (Altenhein, Cattenoz, and Giangrande 2016)), *Argk* (a downstream *gcm* target and glial marker (Altenhein, Cattenoz, and Giangrande 2016)), and *Con* (*Connectin*, a cell adhesion molecule found in glia cells (Sasse, Neuert, and Klämbt 2015)). Importantly, several new genes with no prior implication for shaping glial identity were also found, such as *grn* (*grain,*

a transcription factor known to regulate the expression of adhesion molecules involved in axonal guidance (Garces and Thor 2006)), as well as a suspected anion transporter called *CG3036*. No glial function has previously been described to either *grn* or *CG3036*, I have shown that CG3036 is expressed in a subset of glial cells (Figure 7-7C) but their potential role remains to be investigated.

The Oc regulon is interesting as well, as it is associated with clusters encompassing neuroblasts of primarily ventral origin; this includes all ventral column clusters (VC1, VC2, VC3 and midline) in addition to the 'Brain NBs' cluster. Mutations in *oc* lead to neuromere fusion throughout the ventral notochord (for more details see introduction section 5.3.8), indicating a role regulating axonal projection and orientation (Finkelstein et al. 1990). This gene is expressed in the procephalic region and ventral regions of the neuroectoderm (Datta et al. 2018), which neatly comports with the clusters the Oc regulon is associated with: the brain neuroblasts, the VC clusters and the midline (Figure 7-14A/B). The motif identified by SCENIC (TAATCC, Figure 7-14A) is highly similar to the consensus site described for *oc* (Datta et al. 2018). Ten genes were identified by SCENIC as being directly regulated by *oc* via this motif, of which most are involved in cell adhesion (*CG45263* and *klg*), regulation of axonal targeting and projection (*comm2*, *Kank* and *tutl*), or chemoattraction and -repulsion (*sli*). How these and other potential *oc* targets (*CG45263, dve, nw*, and *Pvf3*) may be involved in mediating these processes in a ventrally-specific manner to shape nervous system development remains open for investigation.

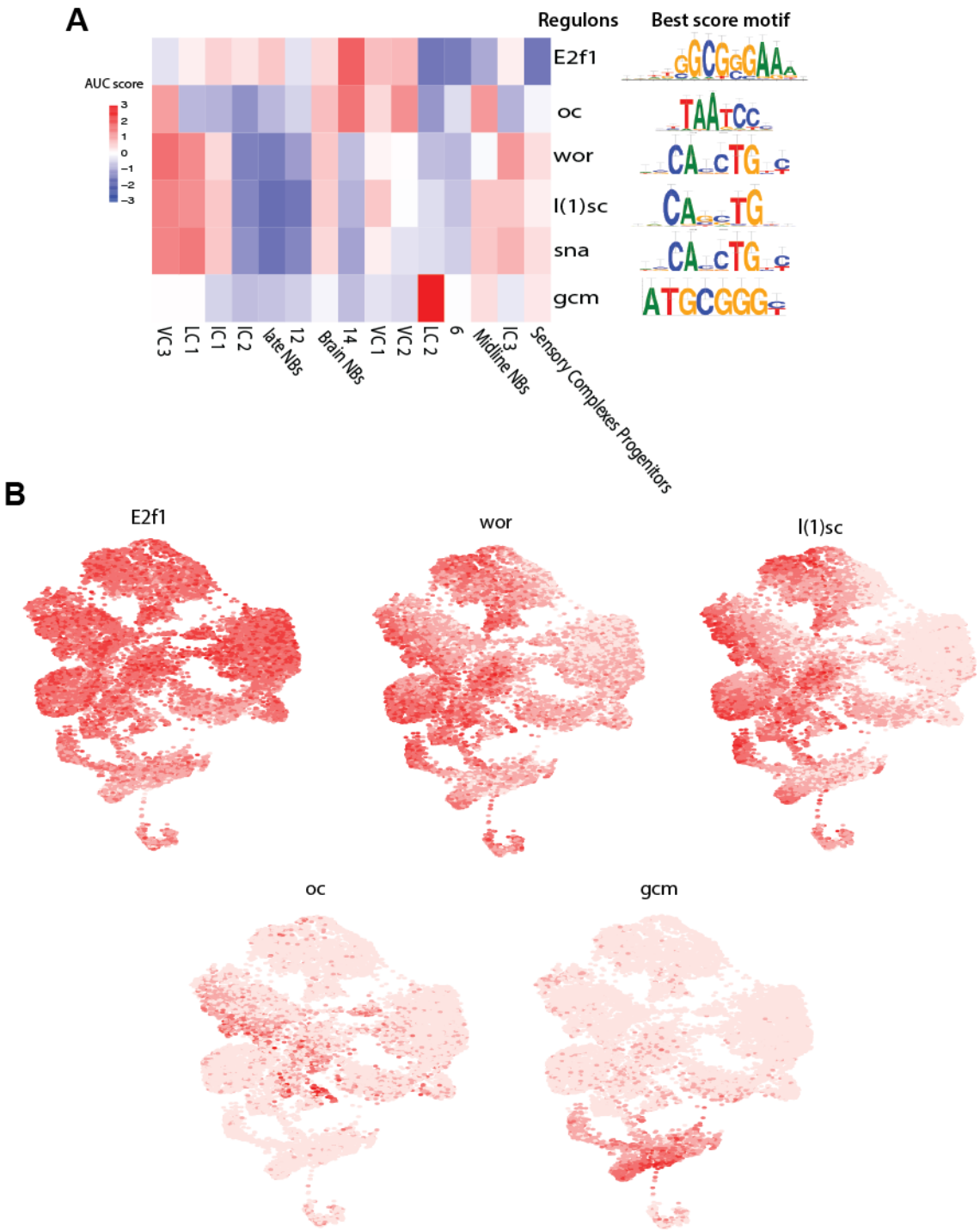


Figure 7-14 Gene regulatory networks in neuroblasts

(**A**) Heatmap showing AUC scores of the regulons in the different neuroblast clusters (compare to Figure 7-6A); the best-matching binding motif for each regulon (highest normalized enrichment score) is shown to the right. (**B**) Heatmap showing AUC score of each regulon/gene set per cell.

8 GENERAL DISCUSSION AND OUTLOOK

The main goal of this project was to unravel individual neuroblasts identities to understand the molecular mechanisms driving cellular diversity in the *Drosophila* embryonic nervous system. I developed a protocol that allows for the capture and sequencing of specific cellular populations based on FACS sorting. The fixation method established is compatible with single cell technologies and allows for antibody staining while preserving RNA integrity for long periods of time.

Applying this method, I generated a single cell atlas of embryonic neuroblasts which allowed me to uncover new spatial markers in the developing ventral nerve cord, markers for individual neuroblasts, characterize aspects of the gene regulatory networks involved in neuroblast development as well as identify potential new targets of transcription factors regulating these networks.

8.1 METHOD DEVELOPMENT

The recent developments in single cell technologies that allow for the generation of genome-wide data for millions of cells enable researchers to the explore specific subsets of cells within complex tissues. This single-cell view has allowed to better understand specific cell types and has uncovered a much more complex array of cell types and cell states that developing and differentiated tissues are composed of (e.g. Delile et al. 2019; Cao et al. 2019).

In order to explore the cellular diversity within the developing ventral nerve cord of the *Drosophila* embryo, I set out to develop a protocol that allows for the single cell sequencing of neuroblasts based on FACS sorting using antibody staining. This protocol allows for easy access to a specific cell population and significantly increases the coverage of the cells of interest allowing for a better understanding of small populations that are usually under sampled in a whole organism/tissue approach. A combination of different fixatives (DPS-MeOH) allowed for the stable storage of samples for long periods of time without noticeable RNA degradation while maintaining the antibody epitopes required for sorting. This was not possible using methanol or DSP alone (discussed in section 7.1.2). The protocol developed in this work allowed a significant increase in the number of neuroblasts captured by well over 2 orders of magnitude – from a few hundred to 10s of thousands when compared with whole embryo sequencing, which resulted in a drastic sampling depth increase by sorting a cell population of interest.

The protocol allows for sorting of both cells and nuclei, depending whether the epitope targeted by the antibody is nuclear or not. The use of antibodies targeting endogenous epitopes increases the speed of data generation as it does not rely on the generation of specific labelled lines while still allowing for the reliable sorting of a defined cell population. This method may also be compatible with a multitude of tags and could be applied to a series of existing tagged lines that are readily available from stock centers

(e.g. FlyORF (Bischof et al. 2013)). It also allows for further modifications such as the encoding of lineage markers, that will label not only the progenitor cells but also their progeny.

In summary, the protocol generated in this thesis establishes several important improvements in terms of gaining a single-cell understanding of complex tissues using high-throughput, droplet based assay platforms such as 10x Genomics:

(i) It allows for versatile sorting of specific cells without the generation of genetically tagged lines.

(ii) It results in a radical increase in the number of cells of interest captured compared with whole embryo single cell sequencing. Generating a high coverage atlas of the tissue/cells of interest, thus permitting the access to cellular populations that were previously inaccessible.

(iii) Fixation itself can be a crucial advantage as it arrests the cellular state in time. Not only is fixation important to resolve temporally dynamic processes such as development, but it also stabilizes molecules such as mRNA during time-intensive assays such as FACS. A major feature of my protocol is the combined use of easily-removable non-cross-linking fixatives (i.e. MeOH) and cross-linking fixatives whose effect is easily and efficiently reversed (i.e. DSP).

Together, these improvements allowed me to generate a single cell-resolved transcriptomic atlas of the neuroblasts involved in early neurogenesis in the *Drosophila* embryo, with high sampling depth and temporal resolution. The protocol developed here can be expected to be widely applicable to complex and temporally dynamic biological tissues.

8.2 DROSOPHILA NEUROBLASTS SEQUENCING

Drosophila embryonic neuroblasts are spatially arranged into a checkboard-like pattern characterized by the expression of dorsoventral and anteroposterior markers that have a role in proper neuroblast formation as well as their developmental trajectories. The mechanisms that further separate neuroblasts within the same spatial domain remain largely uncharacterized. Another dimension of neuroblast progression that helps explain the neuroblast diversity is the temporal regulation of these cells: Neuroblasts are regulated by two temporal axes, the *development time axis*, which refers to the time after fertilization, and the *lineage time axis* that is characterized by the expression of a series temporal transcription factors, whose expression directly impacts the neuroblast transcriptome and the specific lineage that emerges.

The method developed in this work allows us to unravel the regulatory inputs of neuroblast development and progression. I sorted cells based on the Worniu antibody that captures both the neuroblast and the intermediate progenitor state – ganglion mother cells (GMCs). To better understand the changes that occur during neuroblast development and capture both temporal axes, embryos were collected in successive one-hour collection timepoints to capture the entirety of early neuroblast development, from emergence until they stop dividing.

An atlas of the early embryonic nervous system development was generated which consists of a final, quality-controlled and filtered set of ~43 000 cells. This approximately corresponds to a 3-4[3] times coverage of each cell. This atlas allows for the identification of individual neuroblasts and their average transcriptomes. To guarantee that the observations made were due to biological and not technical variation two replicates per timepoint were generated and batch correction was employed when needed. This neuroblast atlas now allows exploration of long-standing questions regarding neuroblast diversity, general developmental trajectories and cell fate choices.

My data revealed that neuroblasts separate based on their spatial identity with the strongest signature being their dorsoventral identity, characterized by the expression of columnar markers, such as the transcription factors *vnd, ind* and *Dr* (Figure 7-6). While dorsoventral identity is especially strong early, in later timepoints I observed the emergence of the neuronal precursor cells or potential GMCs as they express either neuronal or glia markers as well as lower levels of neuroblast markers. This probably correspond to the transcriptomic shift that occurs from neuroblast to GMC when the proliferative state ends and cellular differentiation begins. There is a clear shift from spatial identity being the major driver of cell clustering in early timepoints to a cell fate identity in later timepoints where the expression of neuronal and glial markers drives the clustering.

Another very interesting observation is the presence of neuronal and glia markers in the progenitor cells (both neuroblast and ganglion mother cell) that has not been previously described. Nonetheless, both were clearly observable in the data and I validated my finding by *in situ* stainings in developing embryos. Though the neuronal and glia markers have described roles in differentiating and differentiated neurons and glia such as regulation of axonal projections, they should not yet be required in the precursor cells. Therefore, it is quite interesting to explore the potential reasons for their early expression: *Are these genes being expressed early but not translated? Will the gene products only be active in the daughter cells or do they have a different role in the progenitor cells? In the case of a ganglion mother cell producing both a neuron and a glial cell, are these genes asymmetrically inherited and how?* One way to start to explore this conundrum is to verify if the protein is actually present at this early stage or only the RNA detected here. If these genes are being translated, we could further explore their roles employing mutant strains and mis-expression assays and observing if phenotypes manifest as the nervous system develops.

Even though the ~30 *Drosophila* neuroblasts per hemi-segment are segmentally repeated, the diversity of the cellular lineages these cells produce is quite remarkable. The progeny of the embryonic neuroblasts is quite diverse both in terms of cellular fate but also cellular morphology with cells

[3] Coverage approximated considering the number of neuroblasts and GMCs that emerge from the ventral nerve cord. 30 neuroblast per hemi-segment in 14 segments results in ~840 neuroblasts. The GMC numbers are harder to determine as neuroblast generate anywhere from 3-12 GMCs and the exact number varies between segments. This results in around 10000 cells emerging from the ventral nerve cord (NBs and GMCs).

establishing very specific and complex connections (Weiss et al. 1998; McDonald et al. 1998). Very little is known of how these cells' morphologies are defined and even less of the molecular mechanisms that drive this process. The neuroblast atlas generated here now gives us a crucial tool to begin addressing how neuroblasts progress and diversify into distinct lineages. Our atlas captures two progenitor states, the neuroblast and the ganglion mother cell – we can therefore explore the diversity within each one of these progenitor states and understand the transcriptomic changes that drive their lineages.

8.2.1 Neuroblast progression

To better understand the diversity within the neuroblast population, I separated the single cell transcriptomes into pure neuroblasts (excluding from ganglion mother cells) by the expression of exclusive neuroblast markers (e.g. *dpn*). Analysis of the selected *dpn*-positive neuroblasts revealed diversity within the dorsoventral cellular populations as several clusters were identified for each of the columns. The separation of the columnar cells into distinct clusters can at least in part be explained by the presence of anteroposterior markers. This means that some of these clusters have distinct signatures of overlapping dorsoventral and anteroposterior markers allowing for the identification of smaller, more refined neuroblast populations. This strong spatial signature was especially apparent for cells harvested from the earliest timepoints (TP1-4). Cells from later timepoints converge into populations that are characterized by the expression of neuronal or glial markers. This indicates that neuroblast progression incorporates strong spatial signatures as neuroblasts delaminate and start dividing, but this spatial signature gives way to a differentiation signature characterized by markers of neurons and glia. This progression is also accompanied by a decrease of cell proliferation factors such as *CycE* and an increase of cell cycle inhibitors such as *dacapo*. I was able to clearly observe that the early transcriptome changes in neurogenesis – a process beginning with a very defined spatial identity, which is likely crucial for setting/defining specific lineages of neuroblasts in accordance with their local origin. These early neuroblasts show distinct gene programs indicative of proliferation to originate the ~330 cells that emerge from ~30 neuroblasts per hemi-segment with increasing lineage time, the cell cycle starts to be inhibited (see section 5.3.6) and differentiation pathways are upregulated to initiate the formation of neurons or glia.

The exact reason why neuronal and glia markers are expressed relatively early in neuroblasts still remains to be explored. So far, these genes have only been studied in fully differentiated cells and primarily at the protein level. The high coverage of this atlas leaves no doubt, however, that even early undifferentiated neuroblasts express genes generally associated with differentiated neurons and glia. It will be interesting to explore if these genes have different roles in the progenitor cells and if they are necessary for the correct specification of the neuroblast progeny along their developmental trajectories.

8.2.2 GMCs populations

The second class of progenitor cells captured in this data is the intermediate progenitor (ganglion mother cell, GMC), to explore the diversity within the GMCs, I selected cells that lack the expression of neuroblast exclusive markers (e.g. *dpn*). This selection method is not perfect, because failure to detect these markers is not necessarily proof-positive that they are not expressed (an issue generally referred to as the drop-out problem). However, due to the absence of exclusive GMC markers, this is the only available option to explore the GMC cell population. Nonetheless, it should be taken into consideration that some of the cells called 'GMCs' in this analysis could be neuroblasts.

I observed the increased presence of later neuronal and glial markers in this population, further validating their identity as GMCs (as they are further along the differentiation path). A striking feature is the presence of glia vs. neuronal markers, that separates these cells into two main populations: (i) neuronal precursors due to the expression of neuronal markers and (ii) glia precursors due to the expression of glia markers. Each of these 2 populations sub-clustered further, revealing clusters with '*brain*' and '*ventral nerve cord*' character among the neuronal precursors, as well as clusters with '*neuropile*' and '*brain*' character among the glia precursors.

Comparing the glia and neuronal precursors led to the identification of enriched genes such as *CG6218* for the glia and *CG8407* for the neuronal precursors. I showed that *CG6218* strongly overlaps with *gcm* (glia marker) indicating that this is indeed a good marker for this population. CG6218 is predicted to be a N-acetylglucosamine kinase, being involved in carbohydrate phosphorylation, this might be important to mediate interactions between glycans in the cell membrane or extracellular matrix and could potentially regulate cell morphology not only of glia cells but be mediating the axonal trajectories as other glycans such as heparan sulphates have been described to have a role in axonal projection in *Drosophila* (see section 5.3.11).

CG8407 on the other hand it is not exclusive to the neuronal precursors as it is expressed in all the progenitor cells; however, it does show a significant overlap with the neuronal marker *nerfin-1* explaining why I saw an enrichment of *CG8407* in the neuronal precursor population. Testing the role of these genes in the context of embryonic nervous system development might start to elucidate how cell fate decisions between neurons and glia are made. A good way to start would be to mutate these genes in the specific precursors populations and evaluate if the number of neurons or glia is altered. Mis-expression assays where the glia precursor-enriched genes are expressed in the neuronal precursors could also be performed to evaluate if these factors are enough to induce a specific cell fate or if they have a milder effect, impacting for example cell morphology.

As mentioned above, the most striking difference between glia and neuronal precursors is that the glia marker genes emerge in surprisingly early stages of neuroblast progression. Genes that have been associated with glial identity are already detectable at TP2 while neuronal markers emerge in the last

two timepoints (TP5-6). The glia master regulator, *gcm*, as well as some of its known targets (e.g. *repo* and *pnt*) are found in developmental times where no glia yet exist in the embryo, which indicates that *gcm* and its targets genes are not only expressed in the glia cells but already in the precursors that will give rise to glia. The reason why glia identity emerges before neuronal identity is unclear. *Why are glia markers expressed in progenitor cells that are mostly bipotential, i.e. the GMC divides originating a neuron and a glia cell*? While I am measuring RNA rather than protein, my data begs the questions "*Are glial markers expressed in neurons and glia after the final division?*" and "*Are ganglion mother cell-produced mRNAs for glial markers inherited by and translated in both daughter cells?*". The reason these questions are important is that the answer would help us understand the mechanisms that mediate cell fate decisions in the GMCs. One possible reason for the temporal delay between the expression of glia vs. neuronal markers is that the neuronal signature is '*stronger*' than the glial signature, so when neuronal genes are present the cell would adopt the neuron fate. It is important to note that the genes that determine neuron fate in this context have not been described, in this data I found the expression of several neuronal markers in these progenitors but which ones are pivotal for driving the neuronal fate is still unknown.

If the neuronal signature is indeed more determinative than the glial signature, it would make sense that glial fate determinants emerge earlier so that the glia cells can be selected and only later the neuronal fate of the sister cell is defined. This hypothesis could easily be tested. For example, if the expression of *gcm* is delayed until later stages, *will the glia cells still form as normal? Will there be more neurons being formed from the neuroglioblasts? Will the neuronal projections be altered?* These are just some of the questions that need to be answered to examine the impact of the observed expression delay on neuro-glial decisions. These questions can start to be explored by using conditional knockouts where the expression of *gcm* is delayed until after the neuroblasts emerge and start to divide – the temperature sensitive Gal80-Gal4 system would be an adequate tool to employ here. If glia cells are lost when *gcm* is temporarily delayed in specific neuroglial precursors, a strong argument could be made for the primacy of the neuronal differentiation path and the necessity to 'push' cells toward glial fate early. Further questions, such as the fate of *gcm*-target gene expression upon *gcm* delay would further elucidate the mechanisms that drive neurogliogenesis.

8.3 SPATIAL IDENTITY

In the early timepoints, spatial identity is the strongest driver of neuroblast clustering. To uncover which are the genes and mechanisms driving the very distinct lineages between different spatial domains, I looked at transcriptomic differences between the domains of both spatial axes (dorsoventral and anteroposterior).

8.3.1 Neuroblasts along the anteroposterior axis

The VNC neuroblasts are segmentally repeated (Figure 5-1) and different markers are expressed in the distinct rows of each segment. Selecting cells based on the expression of the known anteroposterior markers, *wingless* and *engrailed* allowed for the identification of a marker for a new anteroposterior domain, *CG42342*, which I determined to be exclusively expressed in rows 1 and 2. The identification of this marker makes a new subpopulation of neuroblasts accessible. Other genes were not exclusive but enriched in the anteroposterior domains such as *CG12496* in the *wg*-positive domain and *drl* in the CG42342-positive domain. Understanding if genes like these have a role in the development of the neuroblasts in the respective domains or in their specific lineages will bring us closer to understanding the full regulatory impact that the anteroposterior axis has in the nervous system development.

For example, *drl* is a receptor for the Wnt ligand, Wnt5, that has been shown to help modulate axonal guidance in the *Drosophila* embryo (Yoshikawa et al. 2003) and understanding how the spatially restricted expression of this receptor modulates the formation of the nervous system would be extremely interesting. Mis-expressing *drl* in other anteroposterior rows might alter the trajectory that specific neurons establish and change the connections these make to specific muscle cells. Eliminating *drl* from rows 1 and 2 might lead to the loss of specific nerves and result in severe phenotypes in the structure of the VNC. Understanding the specific genes involved in the regulation of each neuroblast lineage will elucidate the molecular mechanisms behind robustly organizing a complex tissue.

Exploring the role of the new markers found could better explain how the anteroposterior axis regulate neuroblast identity. The role of wingless has been extensively studied as it has a strong phenotype in the neuroblast of the row it is expressed in (row 5), as well as on the adjacent rows. Understanding the role of CG42342 in the establishment of neuroblast identity, not only in rows 1 and 2 but also in the adjacent domains could bring us closer to understand how the segment is separated (anteroposterior) into seven rows and how the different genes expressed here modulate neuroblast development and lineage. Currently I am exploring the role of a few potential markers such as CG42342: A knockout line is being generated to understand the impact of this gene in the neuroblast or their progeny formation and cell fate.

8.3.2 Neuroblasts along the dorsoventral axis and cross-sectioning

In order to explore the expression variances between the three dorsoventral columns, I started by separating cells based on their columnar identity (only cells that express the columnar markers were kept). This allowed me to directly compare what genes are differentially expressed between these three populations such as *Ilp4* which is restricted to the ventral column in early stages of the data or *CG10479* a new marker for the lateral column. Interestingly, no exclusive markers could be found for the intermediate column population, the genes enriched in these cells were also expressed to some degree

in one of the other columns (Supplementary Figure S 9). This seems to indicate that the intermediate column neuroblasts transcriptome is in-between the ventral and lateral column, expressing a mix of genes from those domains. Furthermore, I observed that the strong spatial signature is somewhat lost in later timepoints (TP5-6) where cells from the three columns seem to overlap, which is accompanied by the expression of later markers such as *fne* and *nerfin-1* (Figure 7-9D, Supplementary Figure S 10).

Further experiments are ongoing to specifically probe the role of a predicted ventral column marker, whose expression was validated by *in situ* hybridization. The new ventral column marker, *CG2865* (Figure 7-11C), is being explored using a CRISPR-Cas9 approach. A cross between a fly line that expresses two sgRNA that targets *CG2865* with a ubiquitous Cas9 line is being performed. The goal is to understand the role of *CG2865*, a gene with no known or predicted function has in the ventral column neuroblasts. Understanding the role of genes exclusive to the different spatial domains will potentially elucidate the differences found between the neuroblasts and their lineages that emerge from these domains.

When combining the two spatial axis, one gene emerged as a marker for the intermediate column and the *wg*-positive row, Ptx-1. *In situ* hybridization staining revealed that the overlap between *Ptx-1* and *wg* is not perfect and was only observed in the most posterior segments (starting with abdominal segment 4, A4), in the more anterior segments it seems to label more posterior cells indicating that it can be a marker for a neuroblast for the en-positive domain. To explore the role of Ptx-1 in the emergence and progression of this neuroblast we are performing mis-expression experiments where we drive *Ptx-1* expression in different domains of the VNC. Using a *UAS-Ptx1* line we want to drive the expression of this gene in all of the VNC by using a pan-neuronal driver (for e.g. *sca-Gal4*), depending on what we observe in this experiment we can narrow the expression of *Ptx1* to only one columnar domain of the VNC by using more targeted drivers (e.g. *vnd-Gal4* for the ventral column and *ind-Gal4* for the intermediate column). These experiments should clarify if this transcription factor plays a role in the specification of neuroblasts and if it is sufficient for inducing a specific neuroblast identity.

8.3.3 Spatial restriction of components of signaling pathways

Signaling pathway activity varies along the spatial axes. One such example is the EGF activity along the dorsoventral axis (see section 5.2 and 7.3.2.2). In the neuroblast data, I found differentially expressed components of the EGF pathway such as the exclusivity of *rho* and *vn* in the midline, which was expected based on published results. However, I also observed differential expression of the major transcriptional effector of this pathway, *pnt*. When the EGF pathway is activated, the transcriptional repressor Yan is destabilized (inhibited), whereas the transcriptional activator Pnt is activated and helps transcribe EGF target genes such as *argos, oc* and *tartan*. The specific effect of these genes depends on the tissue. Pnt was particularly enriched in the ventral and lateral column.

It makes sense to see it enriched in the ventral column as it has been described that EGF signaling is required for the maintenance of *vnd* expression in the embryo (J B Skeath 1998). The exact role of Pnt

in the lateral column, however, is less clear. P*ointed* has also been described as a target of the glia master regulator, *gcm*, and as necessary for glia-neuron interactions in the midline; therefore, it is possible that it also has a role in glial differentiation and function in the lateral column, from which most embryonic glia emerge. Interestingly, it has been described that EGF is not necessary for the emergence of lateral column neuroblasts and forcing EGF activity into the lateral column leads to the loss of lateral column neuroblasts likely due to the expansion of the *ind* expression domain (J B Skeath 1998). This may be less of a discrepancy than it would initially seem, given that *pnt* expression in the lateral column may not even be downstream of EGF signaling at all. Pointed is likely needed at later stages of neuroblast progression. It would be interesting to inhibit *pnt* transcriptional activation in the lateral column and examine how this would impact those neuroblasts and their lineages. Would Pointed inhibition affect the identity of the cells formed in the lateral column and would this effect be restricted to the glia cells that emerge from this domain or would it also affect the neurons.

8.4 IDENTIFICATION OF SPECIFIC NEUROGLIOBLASTS

Drosophila VNC neuroblasts share most of their transcriptome, the small distinctions between them largely being due to markers of the spatial domains they belong to. This high similarity between cells makes identifying individual neuroblasts a complicated task. Most of the described neuroblast markers are not exclusive for a single neuroblast and while combinations of known markers can allow for visual identification of specific neuroblasts by staining in some cases, the sparse nature of single cell data makes it very complicated to confidently identify a specific neuroblast. To make this task easier, I decided to separate the neuroblasts with a pure neuronal lineage (most of the neuroblasts) from neuroblasts at the top of an apparent mixed neuro-glial lineage (only 7 neuroblasts per hemi-segment) by selecting cells that express glia markers such as *gcm* and *repo*. After clustering the neuroglioblasts, I looked for known neuroblast markers, spatial markers and specific lineage genes. This set of genes allowed me to confidently identify clusters of five of the seven neuroblasts as well as potential new markers for them, some of these are not exclusive for a single cell (Figure 7-12), but are expressed in specific subsets of neuroglioblasts. Delineating the role of these genes will be an important first step in understanding what drives the distinct glia lineages that derive from these progenitor cells and it is a significant start in the description of the transcriptome of the individual neuroblasts.

The data generated in this project allowed for the identification of individual neuroblast transcriptomes and it is the first step to understand what sets the 30 neuroblasts apart from each other and what are the mechanisms that drive their unique lineages. Exploring the differentially expressed genes between the identified neuroblasts will further elucidate the role of unknown genes (e.g. see section 7.4) and further our understanding of how the complex *Drosophila* embryonic nervous system is formed.

8.5 GLIA PROGENITOR PROGRESSION ALONG DEVELOPMENT

To understand the temporal regulation of neuroblasts, I performed a pseudotime analysis in a subpopulation of neuroblasts – the neuroglioblasts – as this is the best-characterized population. The pseudotime analysis organized the cells along developmental time (Figure 7-13C) as well as lineage time (Figure 7-13E) reinforcing the role of both temporal axes in the regulation of neuroblast progression.

Exploring the genes that change along pseudotime and performing gene ontology (GO) term analysis for enriched genes at the different stages of pseudotime revealed a cellular progression from high proliferative state to differentiation. Cell cycle promoting genes are enriched in the beginning of pseudotime, which gives way to terms such as asymmetric division at intermediate stages, followed by axonogenesis and gliogenesis at the end of pseudotime.

One interesting pathway that emerged in the middle of this trajectory was the Wnt pathway that is known to be spatially regulated in the neuroblasts (see above), where the ligand (*wg*) is exclusively produced by cells of row 5 and will activate the Wnt pathway in cells from row 4 and 6 (Bhat 1999). This activation is necessary for the specification of the neuroblasts of these rows. How this pathway is exactly regulated in the neuroblasts is unknown but it assumed that is generally follows the canonical Wnt pathway, with Wnt ligands (like Wg) binding to the receptor Frizzled, leading to the phosphorylation of pangolin allowing it to be translocated to the nucleus where it activates Wnt target genes (Supplementary Figure S 14A). The exact targets of Wnt in the neuroblasts are also mostly unknown and seem to vary in different subsets of neuroblasts. For example, *en* is a target gene but only in row 6 while it is not expressed in row 4. The mechanisms that modulate these spatial differences in Wnt targets remain unexplored.

Interestingly in this specific subpopulation (glia progenitors) we see an enrichment for *inhibition* of the Wnt pathway, which is possibly related to the spatial origin of the neuroglioblasts. This enrichment is also observed when exploring the expression of the Wnt pathway components in different anteroposterior domains (Supplementary Figure S 14B-C) where the inhibitors are significantly enriched in the *CG42342* domain (rows 1- 2), from where most neuroglioblasts emerge (five out of seven), as well as the majority of glia cells. Wg is released by row 5 cells and signals to the surrounding cells; as such it is probable that the cells from rows 1 and 2 need to actively inhibit the activity of this pathway for the correct specification of the neuroblasts from this anteroposterior domain. Alternatively, it may be that temporal damping of Wnt signaling is a crucial prerequisite for the glial lineage decision. Molecular manipulation of the Wnt pathway in neuroblasts rows 1 and 2 could address these hypotheses: *Is Wnt activation in this spatial domain capable of altering the neuroblasts that emerge or their progeny?* From what has been previously described about the role of anteroposterior markers in neuroblast delamination it is likely that the fate of these neuroblasts would be altered, which could likely result in a loss of the majority of the glia cells as they emerge from the anteroposterior domain that lacks Wnt activity.

8.6 GENE REGULATORY NETWORKS IN THE NEUROBLASTS

Exploring how individual genes are differentially expressed between neuroblast subpopulations is important, as it allows us to access cellular populations that could not be previously identified or isolated at a genomic level. In order to explore how sets of genes are regulated together, I employed a gene regulatory network tool called SCENIC (Aibar et al. 2017). SCENIC found groups of genes and transcription factors that are differentially expressed in the neuroblast dataset. Some of the gene-transcription factor sets were globally expressed such as the gene set regulated by ***wor*** and ***l(1)sc***. This is to be expected and reassuring as all neuroblast express these genes. Another gene set that was active in most cells was the ***E2f1***, this regulates cell cycle in neuroblasts and it was most active in earlier cells and absent in cells that express the cell cycle inhibitor, *dacapo*; this observation further validates the quality of the analysis as this regulation of the cell cycle in neuroblasts has been previously described (Berger et al. 2005; F. Yu, Kuo, and Jan 2006; Baumgardt et al. 2014; Bahrampour et al. 2017)

More interesting were the gene sets that were enriched in specific neuroblast populations such as the ***gcm*** and ***oc.*** Gcm is the glia master regulator and SCENIC was able to identify several of its described targets such as *repo* and *Argk*, but it also found potential new targets such as *grain* and *CG3036*. *CG3036* was also identified as being enriched in the glia precursors population (Figure 7-7D). The expression of *CG3036* was validated by *in situ* and I was able to show that it overlaps with *gcm* in a subset of cells from the lateral column (Figure 7-7D). This further supports the finding that *CG3036* might be regulated by Gcm in a subpopulation of neuroblasts. Understanding the role of these genes in this population can further elucidate how the lineages of these neuroblasts are defined and the molecular mechanisms that regulate them.

One of the most interesting candidates found during this work was *CG42342* a marker for a new domain of the anteroposterior axis, potentially involved in cell adhesion. *CG42342* was also identified by SCENIC as being a potential target of Gcm (the glia master regulator), this suggests that CG42342 might have a role in glia specification and development as most of glia cells emerge from the rows of neuroblasts labeled by this *CG42342*. To explore the relevance of *CG42342* in the development of the neuroblast from this spatial domain or more specifically in glia development we are waiting for a knockout line to unravel if this mutation produces a nervous system phenotype and understand the molecular mechanisms behind it.

Another gene set was the *oc (o*celliless*)* regulon. Mutations in the *oc* transcription factor result in the collapse of the neuromeres (ladder structure) of the *Drosophila* embryonic VNC, but the mechanisms behind this phenotype are unknown. This gene set is exclusive to the midline and ventral column. SCENIC identified several potential targets of *oc* that are mostly involved with cell adhesion or regulation of axonal targeting. Besides the genes with known function, SCENIC also identified genes with unknown function such as *CG45263,* which does not yet have a defined role in the formation of the

VNC, as an *oc* target gene. Next steps in this analysis will include investigating the effect that mutating *oc* target genes has on VNC structure – some of them have already been described, such as *slit*. Mutations in *sli* result in a similar phenotype as the oc mutation (Bhat, Gaziova, and Krishnan 2007). The roles of other potential *oc*-target genes remain to be elucidated (such as that of *CG45263)* and I am currently creating the mutant fly line for this gene to verify if it also modulates the VNC structure.

8.7 FUTURE OUTLOOK

In this project, I have created an atlas of the early Drosophila VNC with unparalleled depth and coverage of each individual neuroblast. The added temporal dimension of the data further adds to its relevance to understand the incredibly dynamic and complex process that originates the embryonic nervous system. The data generated during this project constitutes a unique resource that has the potential to elucidate neuroblast specific markers and modulators of neuroblast fate. Some of these were validated, some are currently being explored but most remain to be studied. I believe that dissecting the specific role and expression of these genes will greatly increase our knowledge of how these cells generate their specific lineages and would get us closer to the main goal of obtaining the transcriptomic signature of each of the 30 neuroblasts of the *Drosophila* embryo hemi-segment.

My atlas of transcriptome dynamics in early neuroblasts can form the basis for unraveling the molecular mechanisms regulating neuroblast formation and progression as well as the mechanisms driving the specific lineages. Identifying the genes involved in these processes will allow for the construction of networks, beginning to understand individual transcription factors and their role in driving neuroblast fates, even starting to probe cell adhesion and their role modulating specific identities.

To completely explore the formation of the neuroblasts lineages and the formation of the specific connections they form, the data should be expanded and both neurons and glia should be collected from later timepoints of embryonic development. This would allow for tracking of the expression of specific factors from the neuroblast emergence until the establishment of neuronal connections and we could then analyze the dynamics of gene expression along this process. For this purpose, lineage tracing methodologies could be employed such as scGESTALT (Raj et al. 2018) a lineage tracing methodology that allows for lineage tracing for long periods of time. This methodology consists of Cas9 editing of a barcode. The specific benefit of scGESTALT is that it allows for two temporally separated rounds of barcode editing, thus allowing for cell tracking for longer periods of time. Applying this to the neuroblasts would create a traceable signature that would be transmitted to progeny allowing for the delineation of every neuroblast lineage. After identifying every cell from a lineage, we would have a detailed model of how each cell is formed and a better understanding of the gene networks that accompany cell fate specification.

In order to elucidate the connections that the neurons make to specific muscle cells a complementary dataset from muscle cells should be generated to reveal how specific motorneurons reach, stop and then establish neuromuscular junctions with specific cells. This is probably modulated by a complementary set of chemoattraction and repulsion cues as well as compatible cell adhesion molecules between axons and muscle cells, but the specific molecules that drive each connection remain unknown. This complementary dataset could elucidate the molecular mechanisms driving the incredibly distinct projection pattern of the neurons formed during embryonic development.

From decades of work in *Drosophila* embryonic development several similarities have been describe between it and vertebrate neuronal development such as the division of the VNC or neural tube in vertebrates into columnar domains characterized by the expression of specific transcription factors (Weiss et al. 1998). Unravelling molecular mechanisms in *Drosophila* that has a great experimental ease of genetics allows for in depth functional studies that can then be tested and validated in a vertebrate system.

In the vertebrate nervous system, neurons and glia emerge from a progenitor pool and even though specific markers for the different types of neurons and glia have been described the molecular mechanisms that are driving the specific lineages have not been fully characterized. After functionally validating some of the markers found in this work it would be interesting to explore the expression of the vertebrate orthologs and understand if their function or restricted expression is maintained between fly and vertebrates.

There are also significant differences between these organisms such as the temporal separation between the emergence of neurons and glia (Delaunay et al. 2008) seen in vertebrates where neurons are generated first and the same progenitor cells will later give rise to glia. Taking this into consideration it is likely that some of these genes have evolved to have a different role in the development of the nervous system and it would be fascinating to understand this evolution of gene function in distinct organisms.

Elucidating the mechanisms that drive nervous system formation and specification in *Drosophila* will build a general model than can be more easily tested in more complex model organisms and in this way advance our knowledge in the development of the nervous system and how its complexity arises.

9 METHODS

9.1 EXPERIMENTAL METHODS

9.1.1 Fly husbandry

Drosophila melanogaster flies were kept at 25°C with 60% humidity on standard fly food (192 g agar, 440 g sugar syrup, 1.6 kg malt extract, 360 g brewer's yeast, 200 g soy flour, 1.6 kg corn flour, 48 g methyl 4-hydroxybenzoate, and 125 ml propionic acid in 18L water). Fly stocks were transferred to new vials every 2-4 weeks.

The following genotype was used for this thesis: yellow-white - $y^1 w^{1118}; +; +$.

9.1.2 Embryo collection

9.1.2.1 Embryo collection and fixation for standard histochemistry

Embryos were collected on fresh apple juice agar plates and dechorionated in 6% (v/v) Na-hypochlorite (i.e. bleach) for 1min. After thorough washing, the embryos were fixed in 3.7% (v/v) formaldehyde in PBS. The formaldehyde solution and heptane were mixed 50% (v/v) and the embryos incubated for 12 min for antibody staining. For *in situ* hybridization (ISH), the formaldehyde concentration was increased to 9.25% (v/v) and the incubation to 25 min. The lower organic phase was discarded. The vitelline membrane was removed by shaking vigorously for 2 min in 50% (v/v) methanol. Finally, the embryos were washed three times with pure methanol and stored at -20 °C.

9.1.2.2 Embryo collection and fixation for single cell sequencing

Fresh flies (up to ~4 days old) were collected from standard maintenance vials and bottles used to set up collection cages. The agar plates were changed twice a day prior to collection and the cage was replaced every day. After two days in collection cylinders, embryos were collected on apple juice-agar plates with a dollop of yeast paste after three 1-hr pre-lays to purge older, retained embryos. After pre-lays, embryos were collected for 1-hr and aged for defined time spans (e.g. for 3h40, 4h25, 5h10, 5h55, 6h40, and 7h25) prior to dechorionation.

For dechorionation, embryos were incubated in 6% (v/v) Na-hypochlorite (i.e. bleach) for 1min followed by thorough washing with deionized water to remove bleach remnants. Next the embryos were transferred to a blue agar plate (~3% agar-agar with Coomassie brilliant blue R) for manual inspection of the embryos and removal of non-fertilized embryos, as well as excessively old embryos based on morphogenetic features to guarantee correct stage distribution (as quantified in Supplementary Figure S 1). Following this selection, embryos were washed from the plate into a centrifuge tube, wash buffer was removed by aspiration and samples were then fixed with DSP. Preparation time from removal of

collection plates to embryo incubation with fixative took ~40 minutes, which is a delay accounted for in my reported embryo ages (TP1-6 = 3h40, 4h25, 5h10, 5h55, 6h40, and 7h25 hrs AEL).

Fixation was performed with DSP alone according with Attar et al. 2018 or a mix of DSP and heptane. The DSP alone fixed embryos were incubated for 30 min in DSP (1mg/ml in PBS) and stored at 4°C. The DSP-heptane fixed embryos were incubated for 1-hr in a 1:1 solution of DSP (1mg/ml in PBS) and heptane, given that heptane is an organic solvent commonly used in formaldehyde fixation to aid passage of the aldehyde across the vitelline membrane. The embryos were placed on a rotary shaker with vigorous agitation for one hour in the DSP fixative solution followed by three washes with 100% MeOH followed by storage at -20°C. As outlined in section 7.1.3, embryos fixed as outlined could be stored for long periods of time without noticeable degradation in RNA quality.

9.1.3 Standard histochemical staining of embryos

9.1.3.1 Antibody staining

Fixed embryos (see above) were rehydrated in EtOH/PBT mixtures with increasing concentrations of PBT (3:1, 1:1, 1:3) After pre-blocking in several washes of PBT/Western Blocking Reagent (Roche, used as 5X), the primary antibody was incubated overnight in PBT/Western Blocking Reagent at 4°C. After extensive washes in PBT (PBS + 0.1% Triton X-100) to remove excess primary antibody, the embryos were pre-blocked as before and then incubated with the secondary antibody for at least four hours at room temperature (or overnight at 4°C) in PBT/Western Blocking Reagent (Roche). After extensive washes to remove access secondary antibodies, embryos were mounted in ProLong® Gold antifade mounting medium. All washing steps were with gentle agitation on a nutator.

9.1.3.2 In situ Hybridization

Fixed embryos were first cleared using a 1:1 EtOH:Xylenes mixture followed by extensive washes in PBT and post-fixation in 5% formaldehyde in PBT (0.1% Tween-20). The embryos were then hybridized with hapten-labeled antisense RNA probes. The haptens were either FITC (fluorescein), DIG (digoxigenin) or Biotin (Bio).

Probes were made from PCR products for an endogenous sequence with the T7 promoter at the 3'end of the gene (for detailed oligo list see Supplementary Table S 2). For fluorescent in situ hybridization (FISH), probes were detected with pre-absorbed sheep anti-FITC-POD, sheep anti-DIG-POD (both Roche) or mouse anti-Bio-HRP (abcam) primary antibodies (1:500), and visualized using the TSA Plus system (Perkin Elmer). Embryos were mounted in ProLong® Gold Antifade Reagent.

9.1.3.3 Microscopy

Imaging was performed using a confocal microscope (Leica SP8) equipped with 405 nm laser diode, white light laser, and hybrid detectors, with a 20× glycerol objective. For each image 60-100 slices were

acquired. Images were evaluated and projections were constructed using Fiji (ImageJ) for stack selection and z-projection.

9.1.4 Immunohistochemistry for single cell sequencing

DSP alone fixed embryos were resuspended in PBT and DSP/MeOH fixed embryos were rehydrated in EtOH:PBT solution with increasing concentrations of PBT (3:1, 1:1, 1:3), followed by blocking with 1% RNase-free BSA in PBT. The primary antibody was incubated overnight in BSA/PBT at 4°C. After 6 PBT washes, the secondary antibody was incubated overnight at 4°C. All steps were performed at 4°C and all reagents were cold and RNAse free. Embryos were inspected on a Leica DMi8 epifluorescence microscope to validate staining quality.

9.1.5 Embryo dissociation for sc-sequencing

After staining the DSP fixed embryos were immediately dissociated to limit cross link reversal and prevent RNA degradation. The dissociation protocol was adapted from Alles et al., 2017. The embryos were transferred to a 1ml ice cold dounce homogenizer and washed several times with ice cold PBS. The PBS was removed and 1ml of PBS-BSA at 0.04% was added followed by douncing with the loose pestle until embryos are cracked. The mix was transferred to a clean 1.5 ml Eppendorf and centrifuged in a swing-bucket rotor at 3000g for three minutes, the supernatant is discarded and the pellet is resuspended in 1ml of PBS-BSA 0.04% followed by a centrifugation at 40g for 3 minutes to remove large debris. The supernatant was then transferred to a new Eppendorf tube and centrifuged again at 1000g for 3 minutes. After removing the supernatant, the pellet was resuspended in 1ml of PBS-BSA with a 22Gx2" syringe 20 times. The cell suspension was then passed through a 20 μm filter to remove debris and cell clumps followed by centrifugation at 1000g for 3 minutes. The pellet was then dissolved in PBS-BSA and the cell suspension was passed through a cell strainer (Falcon 352235) before proceeding with FACS.

9.1.6 Fluorescence Automated Cell Sorting

Cells from DSP fixed anti-Worniu-stained embryos were sorted based on signal from a secondary antibody (Alexa 555 anti-rat, abcam). Cells were loaded onto a BD AriaIII™ cell sorter in a PBS buffer and 0.04% RNase-free BSA to promote that cells and nuclei remain in suspension during the sorting procedure. The gating strategy was designed to obtain isolated neuroblasts: (i) FSA-A vs. SSC-A to select cells and nuclei, the exact gating was determined by sorting different size particles and checking under a microscope. This gate allowed the removal of large debris. (ii) FSC-W vs. FSC-A was used to remove doublets. (iii) Alexa 555 vs. DAPI allowed the selection of high DAPI and high Alexa cells to select the neuroblasts. Cells were sorted into 80% MeOH to prevent RNA degradation and stored at -20°C until

10X library preparation. To quality control each sorting run, a small sample was collected to verify that an enrichment of at least 90% stained/unstained events (cells or nuclei) was obtained.

9.1.7 Single Cell Sequencing

Cells were sequenced using the 10X Chromium system (10xGenomics 2019). To prepare the cells for the encapsulation the fixation had to be reversed. For this purpose, the cells were centrifuged at 1000g for 3 minutes to pellet the cells, the supernatant was removed and the cells resuspended in PBS-BSA 0.04%. The sample was then incubated for 30 minutes at 37°C after adding Dithiotreiathol (DTT) to a final concentration of 50mM, which aids in reversal of cross links by reducing DSP's disulfide bridges. Afterwards the cells were centrifuged at 1000g for 3 minutes and resuspended in PBS-BSA 0.04% twice to assure the removal of DTT. The cells were counted using a Neubauer chamber and diluted to the desired concentration (here: 9000 cells/µl). The diluted sample was loaded on the 10X Controller to encapsulate the cells. Libraries were prepared following the standard 10X protocol and sequenced on an Illumina HiSeq4000 to obtain around 112 million reads per sample.

A table summarizing sequencing statistics per sample is available (Supplementary Table S 1).

9.2 Computational Methods

9.2.1 Sample quality control and alignment

Read quality control and alignment to the genome was performed using the *PiGx* pipeline (Wurmus et al. 2018). *PiGx* combines widely used tools in genomics into one simple, user friendly pipeline to obtain gene expression matrices from sequencing data and also generates basic analysis reports. More specifically, I used the *PiGx scRNAseq* pipeline, which first it aligns the reads to the genome using *STAR* (Dobin et al. 2013), then it performs read quality control using *fastQC* and *multiQC* (Andrews 2010; Ewels et al. 2016). It uses *dropbead* (Alles et al. 2017) to determine cell number. Its outputs are bam, bigwig files, gene expression matrix and a quality control report.

9.2.2 Sample quality control

The gene expression matrix was used to generate a *Seurat (v.3.1)* object. The first step was to check basic cell statistics, read number, UMI count and percentage of mitochondrial reads. Next I defined quality thresholds for each cell, minimum of 500 genes and less than 10% mitochondrial reads.

9.2.2.1 Normalization

Gene expression was normalized using scTransform (Hafemeister and Satija 2019), a regularized negative binomial regression that removes variation from sequencing depth and adjusts the variance based on genes of similar abundances, this method was selected instead of the standard log-

normalization included in the Seurat pipeline as it performes significantly better according to (Hafemeister and Satija 2019).

9.2.2.2 Dimensionality reduction and batch effect removal

Next, I defined the highly variable genes to perform dimensionality reduction and understand if there were any external factors (e.g. sample preparation, replicate) driving the clustering. As a rule, Seurat's batch effect correction was employed for mitochondrial reads and time point replicates when needed. Cell cycle was also inspected as a potential driving factor by scoring each cell based on cell cycle genes by converting the mouse cell cycle gene tables provided by Seurat to the fly orthologs – no significant effects were observed. After clustering the top 25 genes of each cluster were used for GO term analysis and to remove any contaminant cell population.

9.2.2.3 Differential gene expression analysis

The default differential gene expression analysis included in Seurat was applied, which is based on the non-parametric Wilcoxon rank test. By default, when determining differential expressed genes, it compares a cluster or a selected subset of cell with all the other cells in the dataset.

9.2.3 Selection criteria for the different subsets of the data

Every cell selection based on gene expression was performed on normalized gene expression as described in section 9.2.2. The range of each gene's expression values was taken into consideration; lower values were used for genes with generally lower expression and higher thresholds were employed for genes with higher expression levels:

(i) NBs (Figure 7-6) from the all data objects: only cells with values of expression for *dpn* above one were kept.

(ii) GMCs (Figure 7-7) all cells with *dpn* (neuroblast exclusive marker) values bellow one were kept.

(iii) Anteroposterior cell population (Figure 7-8) were selected based on the presence of segment polarity genes: *en* cells were kept with expression values above 1.1, for *wg* above 1.5 and for *CG42342* above 2.

(iv) Dorsoventral cells (Figure 7-9) were selected based on the presence and absence of the column-specific transcription factors (*vnd* for the ventral, *ind* for the intermediate and *Dr* for the lateral column); the values used for each gene were 1 for *vnd*, 2 for *ind* and 1 for *Dr*.

(v) For the dorsoventral with midline the above described object was merged with a midline population defined by levels of expression of *sog* above two and levels of *sim* above one with the absence of the other column markers.

(vi) Glia population (Figure 7-12 and Figure 7-13) was obtain by extracting every glia-assigned cluster from every timepoint. These clusters were assigned based on GO-term analysis and expression of glia markers such as *gcm* and *repo*.

9.2.4 Candidate selection

Candidates were selected based on differential gene expression analysis, focusing on the top 25 differentially expressed genes from each comparison. This gene list was explored for known roles and expression profiles. Genes with a known role in the specific population of interest were considered to validate the predictions and new genes (i.e. genes with no described role or no described role in the embryonic nervous system) were selected as candidates for new markers.

In the comparisons of spatially-restricted populations, the expression pattern was a determinant factor. Using the BDGP database (Hammonds et al. 2013) – a large-scale screen collecting *in situ* expression data in the *Drosophila* embryo – differentialy expressed genes were selected as potential candidates based on their observable expression pattern; genes that had a potential expression pattern restricted to the predicted spatial domain were selected for further testing.

9.2.5 Pseudotime

Pseudotime analysis was performed using Monocle2 (Trapnell et al. 2014), which aims at aligning cells along a continuous process, assuming a progressive change in transcriptome along this process (e.g. differentiation). The corresponding Seurat object was converted into a CellDataSet object (an object class used by Monocle2) and the default monocle workflow was used. A couple of choices are worthy to note: (i) clustering was performed without defining marker genes, (ii) differential gene expression was calculated based on the Seurat-defined clusters, (iii) dimensionality reduction for the trajectory was performed using "DDRTree".

After defining the pseudotime trajectory, the function plot_pseudotime_heatmap was used to obtain sets of genes that significantly change (q_value < $1x10^{-4}$) along pseudotime and understand which genes change together. These gene lists were used for GO-term analysis.

9.2.6 Gene regulatory Networks (SCENIC)

SCENIC (Aibar et al. 2017) was employed to probe the gene regulatory networks that impact neuroblasts (dpn-positive cells) and their sub-clusters. This analysis was performed using the available vignette from Aibar et al., 2017.

SCENIC takes a gene expression matrix and determines gene sets (called "regulons") based on DNA binding motifs. It performs this in three main steps:

(i) Building the gene regulatory networks – identification of co-expression modules of transcription factors and other genes.

(ii) Select potential direct-binding targets based on the presence of DNA binding motifs for the transcription factors. Only motifs with NES score above three were kept as defined in the SCENIC vignette.

(iii) Score the regulons using AUCell. AUCell takes the regulon and calculates gene set activity. Calculates the enrichment of the regulon as the area under the recovery curve (AUC) across the ranking of all genes in a particular cell. Genes are ranked by their expression value. AUCell represents the proportion of expressed genes in the signature and their relative expression values compared with the other genes in the cell.

10 REFERENCES

10xGenomics. 2019. "Chromium Single Cell V(D)J Reagent Kits with Feature Barcoding Technology for Cell Surface Protein,Document Number CG000186 Rev A."

Aberle, Hermann. 2009. "Searching for Guidance Cues: Follow the Sidestep Trail." *Fly* 3 (4): 270–73. https://doi.org/10.4161/fly.9790.

Aibar, Sara, Carmen Bravo González-Blas, Thomas Moerman, Vân Anh Huynh-Thu, Hana Imrichova, Gert Hulselmans, Florian Rambow, et al. 2017. "SCENIC: Single-Cell Regulatory Network Inference and Clustering." *Nature Methods* 14 (11): 1083–86. https://doi.org/10.1038/nmeth.4463.

Alberts, B, A Johnson, and J Lewis. 2002. "Drosophila and the Molecular Genetics of Pattern Formation: Genesis of the Body Plan." In *Molecular Biology of the Cell*.

Alemany, Anna, Maria Florescu, Chloé S. Baron, Josi Peterson-Maduro, and Alexander Van Oudenaarden. 2018. "Whole-Organism Clone Tracing Using Single-Cell Sequencing." *Nature* 556 (7699): 108–12. https://doi.org/10.1038/nature25969.

Alles, J, N Karaiskos, S D Praktiknjo, S Grosswendt, P Wahle, P L Ruffault, S Ayoub, et al. 2017. "Cell Fixation and Preservation for Droplet-Based Single-Cell Transcriptomics." *BMC Biol* 15 (1): 44. https://doi.org/10.1186/s12915-017-0383-5.

Altenhein, Benjamin, Pierre B. Cattenoz, and Angela Giangrande. 2016. "The Early Life of a Fly Glial Cell." *Wiley Interdisciplinary Reviews: Developmental Biology* 5 (1): 67–84. https://doi.org/10.1002/wdev.200.

Álvarez, José-Andrés, and Fernando J Díaz-Benjumea. 2018. "Origin and Specification of Type II Neuroblasts in the Drosophila Embryo." *Development* 145 (7). https://doi.org/10.1242/dev.158394.

Andrews, S. 2010. "FastQC: A Quality Control Tool for High Throughput Sequence Data." *[Online]*.

Antebi, Yaron E, James M Linton, Heidi Klumpe, Bogdan Bintu, Mengsha Gong, Christina Su, Reed McCardell, et al. 2018. "Dynamic Ligand Discrimination in the Notch Signaling Pathway." *Cell* 172 (4): 1184-1196.e24. https://doi.org/10.1016/j.cell.2017.08.015.

Antebi, Yaron E, James M Linton, Heidi Klumpe, Bogdan Bintu, Mengsha Gong, Christina Su, Reed McCardell, and Michael B Elowitz. 2017. "Combinatorial Signal Perception in the BMP Pathway." *Cell* 170 (6): 1184-1196.e24. https://doi.org/10.1016/j.cell.2017.08.015.

Aoki, Tsutomu, Daniel Wolle, Ella Preger-Ben Noon, Qi Dai, Eric C Lai, and Paul Schedl. 2014. "Bi-Functional Cross-Linking Reagents Efficiently Capture Protein-DNA Complexes in Drosophila Embryos." *Fly* 8 (1): 43–51. https://doi.org/10.4161/fly.26805.

Arefin, Badrul, Farjana Parvin, Shahrzad Bahrampour, Caroline Bivik Stadler, and Stefan Thor. 2019. "Drosophila Neuroblast Selection Is Gated by Notch, Snail, SoxB, and EMT Gene Interplay." *Cell Reports* 29 (11): 3636-3651.e3. https://doi.org/10.1016/j.celrep.2019.11.038.

Artavanis-Tsakonas, Spyros, Matthew D Rand, and Robert J Lake. 1999. "Notch Signaling: Cell Fate Control and Signal Integration in Development." *Science* 284 (5415): 770–76. https://doi.org/10.1126/science.284.5415.770.

Arzan Zarin, Aref, and Juan Pablo Labrador. 2019. "Motor Axon Guidance in Drosophila." *Seminars in Cell and Developmental Biology* 85: 36–47. https://doi.org/10.1016/j.semcdb.2017.11.013.

Ashraf, S I, and Y T Ip. 2001. "The Snail Protein Family Regulates Neuroblast Expression of Inscuteable

and String, Genes Involved in Asymmetry and Cell Division in Drosophila." *Development (Cambridge, England)* 128 (23): 4757–67.

Attar, Moustafa, Eshita Sharma, Shuqiang Li, Claire Bryer, Laura Cubitt, John Broxholme, Helen Lockstone, et al. 2018. "A Practical Solution for Preserving Single Cells for RNA Sequencing." *Scientific Reports* 8 (1): 1–10. https://doi.org/10.1038/s41598-018-20372-7.

Averbukh, Inna, Sen Lin Lai, Chris Q. Doe, and Naama Barkai. 2018. "A Repressor-Decay Timer for Robust Temporal Patterning in Embryonic Drosophila Neuroblast Lineages." *BioRxiv*, 1–19. https://doi.org/10.1101/354969.

Ayoob, Joseph C, Jonathan R Terman, and Alex L Kolodkin. 2006. "Drosophila Plexin B Is a Sema-2a Receptor Required for Axon Guidance." *Development* 133 (11): 2125–35. https://doi.org/10.1242/dev.02380.

Bahrampour, Shahrzad, Erika Gunnar, Carolin Jonsson, Helen Ekman, and Stefan Thor. 2017. "Neural Lineage Progression Controlled by a Temporal Proliferation Program." *Developmental Cell* 43 (3): 332-348.e4. https://doi.org/10.1016/j.devcel.2017.10.004.

Baumgardt, Magnus, Daniel Karlsson, Behzad Y. Salmani, Caroline Bivik, Ryan B. MacDonald, Erika Gunnar, and Stefan Thor. 2014. "Global Programmed Switch in Neural Daughter Cell Proliferation Mode Triggered by a Temporal Gene Cascade." *Developmental Cell* 30 (2): 192–208. https://doi.org/10.1016/j.devcel.2014.06.021.

Beckervordersandforth, Ruth M., Christof Rickert, Benjamin Altenhein, and Gerhard M. Technau. 2008. "Subtypes of Glial Cells in the Drosophila Embryonic Ventral Nerve Cord as Related to Lineage and Gene Expression." *Mechanisms of Development* 125 (5–6): 542–57. https://doi.org/10.1016/j.mod.2007.12.004.

Berger, Christian, S K Pallavi, Mohit Prasad, L S Shashidhara, and Gerhard M Technau. 2005. "A Critical Role for Cyclin E in Cell Fate Determination in the Central Nervous System of Drosophila Melanogaster." *Nature Cell Biology* 7 (1): 56–62. https://doi.org/10.1038/ncb1203.

Bertrand, Nicolas, Diogo S Castro, and François Guillemot. 2002. "Proneural Genes and the Specification of Neural Cell Types." *Nature Reviews Neuroscience* 3 (July): 517. https://doi.org/10.1038/nrn874.

Bhat, Krishna Moorthi. 1996. "The Patched Signaling Pathway Mediates Repression of Gooseberry Allowing Neuroblast Specification by Wingless during Drosophila Neurogenesis." *Development* 122 (9): 2921–32.

Bhat, Krishna Moorthi.. 1999. "Segment Polarity Genes in Neuroblast Formation and Identity Specification during Drosophila Neurogenesis," 472–85.

Bhat, Krishna Moorthi, Erik H. Van Beers, and Prema Bhat. 2000. "Sloppy Paired Acts as the Downstream Target of Wingless in the Drosophila CNS and Interaction between Sloppy Paired and Gooseberry Inhibits Sloppy Paired during Neurogenesis." *Development* 127 (3): 655–65.

Bhat, Krishna Moorthi, Ivana Gaziova, and Smitha Krishnan. 2007. "Regulation of Axon Guidance by Slit and Netrin Signaling in the Drosophila Ventral Nerve Cord." *Genetics* 176 (4): 2235–46. https://doi.org/10.1534/genetics.107.075085.

Bieber, Allan J, Peter M Snow, Michael Hortsch, Nipam H Patel, J.Roger Jacobs, Zaida R Traquina, Jim Schilling, and Corey S Goodman. 1989. "Drosophila Neuroglian: A Member of the Immunoglobulin Superfamily with Extensive Homology to the Vertebrate Neural Adhesion Molecule L1." *Cell* 59 (3): 447–60. https://doi.org/https://doi.org/10.1016/0092-8674(89)90029-9.

Bischof, Johannes, Mikael Björklund, Edy Furger, Claus Schertel, Jussi Taipale, and Konrad Basler. 2013.

"A Versatile Platform for Creating a Comprehensive UAS-ORFeome Library in Drosophila." *Development (Cambridge, England)* 140 (11): 2434–42. https://doi.org/10.1242/dev.088757.

Blum, Matthias, Hsin-Yu Chang, Sara Chuguransky, Tiago Grego, Swaathi Kandasaamy, Alex Mitchell, Gift Nuka, et al. 2021. "The InterPro Protein Families and Domains Database: 20 Years On." *Nucleic Acids Research* 49 (D1): D344–54. https://doi.org/10.1093/nar/gkaa977.

Bonn, Stefan, Robert P Zinzen, Alexis Perez-Gonzalez, Andrew Riddell, Anne-Claude Gavin, and Eileen E M Furlong. 2012. "Cell Type–Specific Chromatin Immunoprecipitation from Multicellular Complex Samples Using BiTS-ChIP." *Nature Protocols* 7 (5): 978–94. https://doi.org/10.1038/nprot.2012.049.

Bossing, T, G M Technau, and C Q Doe. 1996. "Huckebein Is Required for Glial Development and Axon Pathfinding in the Neuroblast 1-1 and Neuroblast 2-2 Lineages in the Drosophila Central Nervous System." *Mechanisms of Development* 55 (1): 53–64. https://doi.org/10.1016/0925-4773(95)00490-4.

Bossing, Torsten, Gerald Udolph, Chris Q Doe, and Gerhard M Technau. 1996. "The Embryonic Central Nervous System Lineages of Drosophila Melanogaster I . Neuroblast Lineages Derived from the Ventral Half of the Neuroectoderm" 64 (0240): 41–64.

Cai, Y, W Chia, and X Yang. 2001. "A Family of Snail-Related Zinc Finger Proteins Regulates Two Distinct and Parallel Mechanisms That Mediate Drosophila Neuroblast Asymmetric Divisions." *The EMBO Journal* 20 (7): 1704–14. https://doi.org/10.1093/emboj/20.7.1704.

Campos-Ortega, J. A., and Volker Hartenstein. 1985. *The Embryonic Development of Drosophila Melanogaster*.

Campuzano, Sonsoles, and Juan Modolell. 1992. "Patterning of the Drosophila Nervous System: The Achaete-Scute Gene Complex." *Trends in Genetics* 8 (6): 202–8. https://doi.org/https://doi.org/10.1016/0168-9525(92)90234-U.

Cao, Junyue, Malte Spielmann, Xiaojie Qiu, Xingfan Huang, Daniel M. Ibrahim, Andrew J. Hill, Fan Zhang, et al. 2019. "The Single-Cell Transcriptional Landscape of Mammalian Organogenesis." *Nature* 566 (7745): 496–502. https://doi.org/10.1038/s41586-019-0969-x.

Cho, Joong Youn, Kayam Chak, Benjamin J. Andreone, Joseph R. Wooley, and Alex L. Kolodkin. 2012. "The Extracellular Matrix Proteoglycan Perlecan Facilitates Transmembrane Semaphorin-Mediated Repulsive Guidance." *Genes and Development* 26 (19): 2222–35. https://doi.org/10.1101/gad.193136.112.

Chu-LaGraff, Q, and C Q Doe. 1993. "Neuroblast Specification and Formation Regulated by Wingless in the Drosophila CNS." *Science (New York, N.Y.)* 261 (5128): 1594–97.

Cimbora, D M, and S Sakonju. 1995. "Drosophila Midgut Morphogenesis Requires the Function of the Segmentation Gene Odd-Paired." *Developmental Biology* 169 (2): 580–95. https://doi.org/10.1006/dbio.1995.1171.

Cowden, J, and M Levine. 2003. "Ventral Dominance Governs Sequential Patterns of Gene Expression across the Dorsal-Ventral Axis of the Neuroectoderm in the Drosophila Embryo." *Dev Biol* 262 (2): 335–49.

Crews, Stephen T. 2019. "Drosophila Embryonic CNS Development : Neurogenesis , Gliogenesis , Cell Fate , and Differentiation" 213 (December): 1111–44.

Cubas, P., J. F. De Celis, S. Campuzano, and J. Modolell. 1991. "Proneural Clusters of Achaete-Scute Expression and the Generation of Sensory Organs in the Drosophila Imaginal Wing Disc." *Genes*

and Development 5 (6): 996–1008. https://doi.org/10.1101/gad.5.6.996.

Cui, X, and C Q Doe. 1992. "Ming Is Expressed in Neuroblast Sublineages and Regulates Gene Expression in the Drosophila Central Nervous System." *Development (Cambridge, England)* 116 (4): 943–52.

Datta, Rhea R., Jia Ling, Jesse Kurland, Xiaotong Ren, Zhe Xu, Gozde Yucel, Jackie Moore, et al. 2018. "A Feed-Forward Relay Integrates the Regulatory Activities of Bicoid and Orthodenticle via Sequential Binding to Suboptimal Sites." *Genes and Development* 32 (9–10): 723–36. https://doi.org/10.1101/gad.311985.118.

Deal, Roger B, and Steven Henikoff. 2010. "A Simple Method for Gene Expression and Chromatin Profiling of Individual Cell Types within a Tissue." *Developmental Cell* 18 (6): 1030–40. https://doi.org/https://doi.org/10.1016/j.devcel.2010.05.013.

Delaunay, Delphine, Katharina Heydon, Ana Cumano, Markus H Schwab, Jean-Léon Thomas, Ueli Suter, Klaus-Armin Nave, Bernard Zalc, and Nathalie Spassky. 2008. "Early Neuronal and Glial Fate Restriction of Embryonic Neural Stem Cells." *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience* 28 (10): 2551–62. https://doi.org/10.1523/JNEUROSCI.5497-07.2008.

Delile, Julien, Teresa Rayon, Manuela Melchionda, Amelia Edwards, James Briscoe, and Andreas Sagner. 2019. "Single Cell Transcriptomics Reveals Spatial and Temporal Dynamics of Gene Expression in the Developing Mouse Spinal Cord." *Development (Cambridge)* 146 (12). https://doi.org/10.1242/dev.173807.

Dobin, Alexander, Carrie A Davis, Felix Schlesinger, Jorg Drenkow, Chris Zaleski, Sonali Jha, Philippe Batut, Mark Chaisson, and Thomas R Gingeras. 2013. "STAR: Ultrafast Universal RNA-Seq Aligner." *Bioinformatics (Oxford, England)* 29 (1): 15–21. https://doi.org/10.1093/bioinformatics/bts635.

Doe, Chris Q, David Smouse, and Corey S Goodman. 1988. "Control of Neuronal Fate by the Drosophila Segmentation Gene Even-Skipped." *Nature* 333 (6171): 376–78. https://doi.org/10.1038/333376a0.

Domsch, Katrin, Julia Schröder, Matthias Janeschik, Christoph Schaub, and Ingrid Lohmann. 2021. "The Hox Transcription Factor Ubx Ensures Somatic Myogenesis by Suppressing the Mesodermal Master Regulator Twist." *Cell Reports* 34 (1): 108577. https://doi.org/10.1016/j.celrep.2020.108577.

Duronio, R J, P H O'Farrell, J E Xie, A Brook, and N Dyson. 1995. "The Transcription Factor E2F Is Required for S Phase during Drosophila Embryogenesis." *Genes & Development* 9 (12): 1445–55. https://doi.org/10.1101/gad.9.12.1445.

Dutta, Devanjali, Adam J Dobson, Philip L Houtz, Christine Gläßer, Jonathan Revah, Jerome Korzelius, Parthive H Patel, Bruce A Edgar, and Nicolas Buchon. 2015. "Regional Cell-Specific Transcriptome Mapping Reveals Regulatory Complexity in the Adult Drosophila Midgut." *Cell Reports* 12 (2): 346–58. https://doi.org/10.1016/j.celrep.2015.06.009.

Egger, Boris, James M. Chell, and Andrea H. Brand. 2008. "Insights into Neural Stem Cell Biology from Flies." *Philosophical Transactions of the Royal Society B: Biological Sciences* 363 (1489): 39–56. https://doi.org/10.1098/rstb.2006.2011.

Ephrussi, Anne, and Daniel St Johnston. 2004. "Seeing Is Believing: The Bicoid Morphogen Gradient Matures." *Cell* 116 (2): 143–52. https://doi.org/10.1016/s0092-8674(04)00037-6.

Esteves, Francisco F, Alexander Springhorn, Erika Kague, Erika Taylor, George Pyrowolakis, Shannon Fisher, and Ethan Bier. 2014. "BMPs Regulate Msx Gene Expression in the Dorsal Neuroectoderm of Drosophila and Vertebrates by Distinct Mechanisms." *PLoS Genetics* 10 (9): e1004625. https://doi.org/10.1371/journal.pgen.1004625.

Evers, David L, Carol B Fowler, Brady R Cunningham, Jeffrey T Mason, and Timothy J O'Leary. 2011. "The Effect of Formaldehyde Fixation on RNA: Optimization of Formaldehyde Adduct Removal." *The Journal of Molecular Diagnostics* 13 (3): 282–88. https://doi.org/https://doi.org/10.1016/j.jmoldx.2011.01.010.

Ewels, Philip, Måns Magnusson, Sverker Lundin, and Max Käller. 2016. "MultiQC: Summarize Analysis Results for Multiple Tools and Samples in a Single Report." *Bioinformatics (Oxford, England)* 32 (19): 3047–48. https://doi.org/10.1093/bioinformatics/btw354.

Fambrough, Douglas, and Corey S Goodman. 1996. "The Drosophila Beaten Path Gene Encodes a Novel Secreted Protein That Regulates Defasciculation at Motor Axon Choice Points." *Cell* 87 (6): 1049–58. https://doi.org/https://doi.org/10.1016/S0092-8674(00)81799-7.

Finkelstein, Robert, David Smouse, Theresa M. Capaci, Allan C. Spradling, and Norbert Perrimon. 1990. "The Orthodenticle Gene Encodes a Novel Homeo Domain Protein Involved in the Development of the Drosophila Nervous System and Ocellar Visual Structures." *Genes and Development* 4 (9): 1516–27. https://doi.org/10.1101/gad.4.9.1516.

Forsthoefel, David J, Eric C Liebl, Peter A Kolodziej, and Mark A Seeger. 2005. "The Abelson Tyrosine Kinase, the Trio GEF and Enabled Interact with the Netrin Receptor Frazzled in Drosophila." *Development* 132 (8): 1983–94. https://doi.org/10.1242/dev.01736.

Fox, A Nicole, and Kai Zinn. 2005. "The Heparan Sulfate Proteoglycan Syndecan Is an In Vivo Ligand for the Drosophila LAR Receptor Tyrosine Phosphatase." *Current Biology* 15 (19): 1701–11. https://doi.org/https://doi.org/10.1016/j.cub.2005.08.035.

Froldi, Francesca, Milan Szuperak, Chen Fang Weng, Wei Shi, Anthony T. Papenfuss, and Louise Y. Cheng. 2015. "The Transcription Factor Nerfin-1 Prevents Reversion of Neurons into Neural Stem Cells." *Genes and Development* 29 (2): 129–43. https://doi.org/10.1101/gad.250282.114.

Furukubo-Tokunaga, Katsuo, Yoshitsugu Adachi, Mitsuhiko Kurusu, and Uwe Walldorf. 2009. "Brain Patterning Defects Caused by Mutations of the Twin of Eyeless Gene in Drosophila Melanogaster." *Fly* 3 (4): 263–69. https://doi.org/10.4161/fly.10385.

Gallo, Steven M, Long Li, Zihua Hu, and Marc S Halfon. 2006. "REDfly: A Regulatory Element Database for Drosophila." *Bioinformatics (Oxford, England)* 22 (3): 381–83. https://doi.org/10.1093/bioinformatics/bti794.

Garces, Alain, and Stefan Thor. 2006. "Specification of Drosophila ACC Motoneuron Identity by a Genetic Cascade Involving Even-Skipped, Grain and Zfh1." *Development (Cambridge, England)* 133 (8): 1445–55. https://doi.org/10.1242/dev.02321.

Gavis, E R, and R Lehmann. 1992. "Localization of Nanos RNA Controls Embryonic Polarity." *Cell* 71 (2): 301–13. https://doi.org/10.1016/0092-8674(92)90358-j.

Gerhart, J C. 1980. "Mechanisms Regulating Pattern Formation in the Amphibian Egg and Early Embryo BT - Biological Regulation and Development: Molecular Organization and Cell Function." In , edited by Robert F Goldberger, 133–316. Boston, MA: Springer US. https://doi.org/10.1007/978-1-4684-9933-9_4.

Gilbert, Scott F. 2000. "Cell-Cell Communication in Development." In *Developmental Biology*.

Glahs, Alex. 2021. "Atlas of the Chromatin Landscape in the Developing Nervous System of Drosophila Melanogaster." *In Preparation*.

González, F, S Romani, P Cubas, J Modolell, and S Campuzano. 1989. "Molecular Analysis of the Asense Gene, a Member of the Achaete-Scute Complex of Drosophila Melanogaster, and Its Novel Role

in Optic Lobe Development." *The EMBO Journal* 8 (12): 3553–62. https://doi.org/10.1002/j.1460-2075.1989.tb08527.x.

Gratz, Scott J, C Dustin Rubinstein, Melissa M Harrison, Jill Wildonger, and Kate M O'Connor-Giles. 2015. "CRISPR-Cas9 Genome Editing in Drosophila." *Current Protocols in Molecular Biology* 111 (July): 31.2.1-31.2.20. https://doi.org/10.1002/0471142727.mb3102s111.

Grosskortenhaus, Ruth, Bret J. Pearson, Amanda Marusich, and Chris Q. Doe. 2005. "Regulation of Temporal Identity Transitions in Drosophila Neuroblasts." *Developmental Cell* 8 (2): 193–202. https://doi.org/10.1016/j.devcel.2004.11.019.

Gunnar, Erika, Caroline Bivik, Annika Starkenberg, and Stefan Thor. 2016. "Sequoia Controls the Type L>0 Daughter Proliferation Switch in the Developing Drosophila Nervous System." *Development (Cambridge)* 143 (20): 3774–84. https://doi.org/10.1242/dev.139998.

Habib, Naomi, Inbal Avraham-Davidi, Anindita Basu, Tyler Burks, Karthik Shekhar, Matan Hofree, Sourav R. Choudhury, et al. 2017. "Massively Parallel Single-Nucleus RNA-Seq with DroNc-Seq." *Nature Methods* 14 (10): 955–58. https://doi.org/10.1038/nmeth.4407.

Hafemeister, Christoph, and Rahul Satija. 2019. "Normalization and Variance Stabilization of Single-Cell RNA-Seq Data Using Regularized Negative Binomial Regression." *Genome Biology* 20 (1): 296. https://doi.org/10.1186/s13059-019-1874-1.

Halder, G, P Callaerts, and W J Gehring. 1995. "Induction of Ectopic Eyes by Targeted Expression of the Eyeless Gene in Drosophila." *Science (New York, N.Y.)* 267 (5205): 1788–92.

Hamm, Danielle C, and Melissa M Harrison. 2018. "Regulatory Principles Governing the Maternal-to-Zygotic Transition: Insights from Drosophila Melanogaster." *Open Biology* 8 (12): 180183. https://doi.org/10.1098/rsob.180183.

Hammonds, Ann S, Christopher A Bristow, William W Fisher, Richard Weiszmann, Siqi Wu, Volker Hartenstein, Manolis Kellis, Bin Yu, Erwin Frise, and Susan E Celniker. 2013. "Spatial Expression of Transcription Factors in Drosophila Embryonic Organ Development." *Genome Biology* 14 (12): R140. https://doi.org/10.1186/gb-2013-14-12-r140.

Hartenstein, Volker, and Andreas Wodarz. 2013. "Initial Neurogenesis in Drosophila." *Wiley Interdisciplinary Reviews. Developmental Biology* 2 (5): 701–21. https://doi.org/10.1002/wdev.111.

Hong, Joung-Woo, David A Hendrix, Dmitri Papatsenko, and Michael S Levine. 2008. "How the Dorsal Gradient Works: Insights from Postgenome Technologies." *Proceedings of the National Academy of Sciences of the United States of America* 105 (51): 20072–76. https://doi.org/10.1073/pnas.0806476105.

Hrvatin, Siniša, Francis Deng, Charles W O'Donnell, David K Gifford, and Douglas A Melton. 2014. "MARIS: Method for Analyzing RNA Following Intracellular Sorting." *PLOS ONE* 9 (3): e89459. https://doi.org/10.1371/journal.pone.0089459.

Inaki, Mikiko, Shingo Yoshikawa, John B Thomas, Hiroyuki Aburatani, and Akinao Nose. 2007. "Wnt4 Is a Local Repulsive Cue That Determines Synaptic Target Specificity." *Current Biology* 17 (18): 1574–79. https://doi.org/https://doi.org/10.1016/j.cub.2007.08.013.

Isshiki, Takako, Bret Pearson, Scott Holbrook, and Chris Q. Doe. 2001. "Drosophila Neuroblasts Sequentially Express Transcription Factors Which Specify the Temporal Identity of Their Neuronal Progeny." *Cell* 106 (4): 511–21. https://doi.org/10.1016/S0092-8674(01)00465-2.

Jarman, A P, M Brand, L Y Jan, and Y N Jan. 1993. "The Regulation and Function of the Helix-Loop-Helix Gene, Asense, in Drosophila Neural Precursors." *Development* 119 (1): 19–29. https://doi.org/10.1242/dev.119.1.19.

Jimenez, F, and J A Campos-Ortega. 1990. "Defective Neuroblast Commitment in Mutants of the Achaete-Scute Complex and Adjacent Genes of D. Melanogaster." *Neuron* 5 (1): 81–89.

Kairamkonda, Subhash, and Upendra Nongthomba. 2014. "Beadex Function in the Motor Neurons Is Essential for Female Reproduction in Drosophila Melanogaster." *PloS One* 9 (11): e113003. https://doi.org/10.1371/journal.pone.0113003.

Kaltschmidt, J A, C M Davidson, N H Brown, and A H Brand. 2000. "Rotation and Asymmetry of the Mitotic Spindle Direct Asymmetric Cell Division in the Developing Central Nervous System." *Nature Cell Biology* 2 (1): 7–12. https://doi.org/10.1038/71323.

Karaiskos, Nikos, Philipp Wahle, Jonathan Alles, Anastasiya Boltengagen, Salah Ayoub, Claudia Kipar, Christine Kocks, Nikolaus Rajewsky, and Robert P Zinzen. 2017. "The Drosophila Embryo at Single-Cell Transcriptome Resolution." *Science* 358 (6360): 194. https://doi.org/10.1126/science.aan3235.

Karcavich, Rachel E. 2005. "Generating Neuronal Diversity in the Drosophila Central Nervous System: A View from the Ganglion Mother Cells." *Developmental Dynamics* 232 (3): 609–16. https://doi.org/10.1002/dvdy.20273.

Kasai, Y, S Stahl, and S Crews. 1998. "Specification of the Drosophila CNS Midline Cell Lineage: Direct Control of Single-Minded Transcription by Dorsal/Ventral Patterning Genes." *Gene Expression* 7 (3): 171–89.

Kearney, Joseph B, Scott R Wheeler, Patricia Estes, Beth Parente, and Stephen T Crews. 2004. "Gene Expression Profiling of the Developing Drosophila CNS Midline Cells" 275: 473–92. https://doi.org/10.1016/j.ydbio.2004.08.047.

Kent, David, Erik W Bush, and Joan E Hooper. 2006. "Roadkill Attenuates Hedgehog Responses through Degradation of Cubitus Interruptus." *Development (Cambridge, England)* 133 (10): 2001–10. https://doi.org/10.1242/dev.02370.

Kidd, Thomas, Kimberly S Bland, and Corey S Goodman. 1999. "Slit Is the Midline Repellent for the Robo Receptor in Drosophila." *Cell* 96 (6): 785–94. https://doi.org/https://doi.org/10.1016/S0092-8674(00)80589-9.

Klämbt, C. 1993. "The Drosophila Gene Pointed Encodes Two ETS-like Proteins Which Are Involved in the Development of the Midline Glial Cells." *Development (Cambridge, England)* 117 (1): 163–76.

Kohsaka, Hiroshi, Satoko Okusawa, Yuki Itakura, Akira Fushiki, and Akinao Nose. 2012. "Development of Larval Motor Circuits in Drosophila." *Development Growth and Differentiation* 54 (3): 408–19. https://doi.org/10.1111/j.1440-169X.2012.01347.x.

Kolodziej, P A, L C Timpe, K J Mitchell, S R Fried, C S Goodman, L Y Jan, and Y N Jan. 1996. "Frazzled Encodes a Drosophila Member of the DCC Immunoglobulin Subfamily and Is Required for CNS and Motor Axon Guidance." *Cell* 87 (2): 197–204. https://doi.org/10.1016/s0092-8674(00)81338-0.

Kvon, EZ, T Kazmar, G Stampfel, JO Yáñez-Cuna, M Pagani, K Schernhuber, BJ Dickson, and A Stark. 2014. "Genome-Scale Functional Characterization of Drosophila Developmental Enhancers in Vivo." *Nature*, 91–95.

Labrador, Juan Pablo, David O'Keefe, Shingo Yoshikawa, Randall D McKinnon, John B Thomas, and Greg J Bashaw. 2005. "The Homeobox Transcription Factor Even-Skipped Regulates Netrin-Receptor Expression to Control Dorsal Motor-Axon Projections in Drosophila." *Current Biology*

15 (15): 1413–19. https://doi.org/https://doi.org/10.1016/j.cub.2005.06.058.

Lage, Petra I zur, and Andrew P Jarman. 2010. "The Function and Regulation of the BHLH Gene, Cato, in Drosophila Neurogenesis." *BMC Developmental Biology* 10 (1): 34. https://doi.org/10.1186/1471-213X-10-34.

Lai, Zhi-Chun, Xiaomu Wei, Takeshi Shimizu, Edward Ramos, Margaret Rohrbaugh, Nikolas Nikolaidis, Li-Lun Ho, and Ying Li. 2005. "Control of Cell Proliferation and Apoptosis by Mob as Tumor Suppressor, Mats." *Cell* 120 (5): 675–85. https://doi.org/10.1016/j.cell.2004.12.036.

Landgraf, Matthias, and Stefan Thor. 2006. "Development of Drosophila Motoneurons: Specification and Morphology." *Seminars in Cell and Developmental Biology* 17 (1): 3–11. https://doi.org/10.1016/j.semcdb.2005.11.007.

Lanoue, B R, M D Gordon, R Battye, and J R Jacobs. 2000. "Genetic Analysis of Vein Function in the Drosophila Embryonic Nervous System." *Genome* 43 (3): 564–73.

Layden, Michael J, Joanne P Odden, Aloisia Schmid, Alain Garces, Stefan Thor, and Chris Q Doe. 2006. "Zfh1, a Somatic Motor Neuron Transcription Factor, Regulates Axon Exit from the CNS." *Developmental Biology* 291 (2): 253–63. https://doi.org/https://doi.org/10.1016/j.ydbio.2005.12.009.

Leader, David P, Sue A Krause, Aniruddha Pandit, Shireen A Davies, and Julian A T Dow. 2018. "FlyAtlas 2: A New Version of the Drosophila Melanogaster Expression Atlas with RNA-Seq, MiRNA-Seq and Sex-Specific Data." *Nucleic Acids Research* 46 (D1): D809–15. https://doi.org/10.1093/nar/gkx976.

Lee, Han B, Zachary L Sebo, Ying Peng, and Yi Guo. 2015. "An Optimized TALEN Application for Mutagenesis and Screening in Drosophila Melanogaster." *Cellular Logistics* 5 (1): e1023423–e1023423. https://doi.org/10.1080/21592799.2015.1023423.

Li, Pulin, and Michael B Elowitz. 2019. "Communication Codes in Developmental Signaling Pathways." *Development (Cambridge, England)* 146 (12). https://doi.org/10.1242/dev.170977.

Lin, David M, and Corey S Goodman. 1994. "Ectopic and Increased Expression of Fasciclin II Alters Motoneuron Growth Cone Guidance." *Neuron* 13 (3): 507–23. https://doi.org/https://doi.org/10.1016/0896-6273(94)90022-1.

Lin, Suewei, and Tzumin Lee. 2012. "Generating Neuronal Diversity in the Drosophila Central Nervous System," no. September 2011: 57–68. https://doi.org/10.1002/dvdy.22739.

Macosko, Evan Z., Anindita Basu, Rahul Satija, James Nemesh, Karthik Shekhar, Melissa Goldman, Itay Tirosh, et al. 2015. "Highly Parallel Genome-Wide Expression Profiling of Individual Cells Using Nanoliter Droplets." *Cell* 161 (5): 1202–14. https://doi.org/10.1016/j.cell.2015.05.002.

Massagué, J. 1998. "TGF-β SIGNAL TRANSDUCTION." *Annual Review of Biochemistry* 67 (1): 753–91. https://doi.org/10.1146/annurev.biochem.67.1.753.

McCorkindale, Alexandra L, Philipp Wahle, Sascha Werner, Irwin Jungreis, Peter Menzel, Chinmay J Shukla, Rúben Lopes Pereira Abreu, et al. 2019. "A Gene Expression Atlas of Embryonic Neurogenesis in Drosophila Reveals Complex Spatiotemporal Regulation of LncRNAs." *Development* 146 (6). https://doi.org/10.1242/dev.175265.

McDonald, J A, S Holbrook, T Isshiki, J Weiss, C Q Doe, and D M Mellerick. 1998. "Dorsoventral Patterning in the Drosophila Central Nervous System: The Vnd Homeobox Gene Specifies Ventral Column Identity." *Genes & Development* 12 (22): 3603–12.

Miller, Crystal M, Nan Liu, Andrea Page-McCaw, and Heather T Broihier. 2011. "Drosophila Mmp2

Regulates the Matrix Molecule Faulty Attraction (Frac) to Promote Motor Axon Targeting in Drosophila." *The Journal of Neuroscience* 31 (14): 5335 LP – 5347. https://doi.org/10.1523/JNEUROSCI.4811-10.2011.

Mitchell, Kevin J, Jennifer L Doyle, Tito Serafini, Timothy E Kennedy, Marc Tessier-Lavigne, Corey S Goodman, and Barry J Dickson. 1996. "Genetic Analysis of Netrin Genes in Drosophila: Netrins Guide CNS Commissural Axons and Peripheral Motor Axons." *Neuron* 17 (2): 203–15. https://doi.org/https://doi.org/10.1016/S0896-6273(00)80153-1.

Nambu, J R, J O Lewis, K A Jr Wharton, and S T Crews. 1991. "The Drosophila Single-Minded Gene Encodes a Helix-Loop-Helix Protein That Acts as a Master Regulator of CNS Midline Development." *Cell* 67 (6): 1157–67. https://doi.org/10.1016/0092-8674(91)90292-7.

Ng, M, F J Diaz-Benjumea, and S M Cohen. 1995. "Nubbin Encodes a POU-Domain Protein Required for Proximal-Distal Patterning in the Drosophila Wing." *Development (Cambridge, England)* 121 (2): 589–99.

Park, M, C Lewis, D Turbay, A Chung, J N Chen, S Evans, R E Breitbart, M C Fishman, S Izumo, and R Bodmer. 1998. "Differential Rescue of Visceral and Cardiac Defects in Drosophila by Vertebrate Tinman-Related Genes." *Proceedings of the National Academy of Sciences of the United States of America* 95 (16): 9366–71. https://doi.org/10.1073/pnas.95.16.9366.

Parker, Louise, Jeremy E Ellis, Minh Q Nguyen, and Kavita Arora. 2006. "The Divergent TGF-β Ligand Dawdle Utilizes an Activin Pathway to Influence Axon Guidance in Drosophila." *Development* 133 (24): 4981–91. https://doi.org/10.1242/dev.02673.

Pasterkamp, R Jeroen. 2012. "Getting Neural Circuits into Shape with Semaphorins." *Nature Reviews Neuroscience* 13 (9): 605–18. https://doi.org/10.1038/nrn3302.

Picelli, Simone, Åsa K Björklund, Omid R Faridani, Sven Sagasser, Gösta Winberg, and Rickard Sandberg. 2013. "Smart-Seq2 for Sensitive Full-Length Transcriptome Profiling in Single Cells." *Nature Methods* 10 (11): 1096–98. https://doi.org/10.1038/nmeth.2639.

Pipes, G C Teg, Qing Lin, Stephanie E Riley, and Corey S Goodman. 2001. "The Beat Generation: A Multigene Family Encoding IgSF Proteins Related to the Beat Axon Guidance Molecule in Drosophila." *Development* 128 (22): 4545–52. https://doi.org/10.1242/dev.128.22.4545.

Raj, Bushra, Daniel E. Wagner, Aaron McKenna, Shristi Pandey, Allon M. Klein, Jay Shendure, James A. Gagnon, and Alexander F. Schier. 2018. "Simultaneous Single-Cell Profiling of Lineages and Cell Types in the Vertebrate Brain." *Nature Biotechnology* 36 (5): 442–50. https://doi.org/10.1038/nbt.4103.

Rajagopalan, Srikanth, Emmanuelle Nicolas, Valérie Vivancos, Jürg Berger, and Barry J Dickson. 2000. "Crossing the Midline: Roles and Regulation of Robo Receptors." *Neuron* 28 (3): 767–77. https://doi.org/https://doi.org/10.1016/S0896-6273(00)00152-5.

Rickert, Christof, Thomas Kunz, Kerri Lee Harris, Paul M. Whitington, and Gerhard M. Technau. 2011. "Morphological Characterization of the Entire Interneuron Population Reveals Principles of Neuromere Organization in the Ventral Nerve Cord of Drosophila." *Journal of Neuroscience* 31 (44): 15870–83. https://doi.org/10.1523/JNEUROSCI.4009-11.2011.

Samson, Marie-Laure, and Fabienne Chalvet. 2003. "Found in Neurons, a Third Member of the Drosophila Elav Gene Family, Encodes a Neuronal Protein and Interacts with Elav." *Mechanisms of Development* 120 (3): 373–83. https://doi.org/https://doi.org/10.1016/S0925-4773(02)00444-6.

Sasse, Sofia, Helen Neuert, and Christian Klämbt. 2015. "Differentiation of Drosophila Glial Cells."

Wiley Interdisciplinary Reviews: Developmental Biology 4 (6): 623–36. https://doi.org/10.1002/wdev.198.

Schaefer, M, M Petronczki, D Dorner, M Forte, and J A Knoblich. 2001. "Heterotrimeric G Proteins Direct Two Modes of Asymmetric Cell Division in the Drosophila Nervous System." *Cell* 107 (2): 183–94. https://doi.org/10.1016/s0092-8674(01)00521-9.

Schmidt, Hartmut, Christof Rickert, Torsten Bossing, Olaf Vef, Joachim Urban, and Gerhard M Technau. 1997. "The Embryonic Central Nervous System Lineages of Drosophila Melanogaster II . Neuroblast Lineages Derived from the Dorsal Part of the Neuroectoderm" 204: 186–204.

Schultz, R M. 1993. "Regulation of Zygotic Gene Activation in the Mouse." *BioEssays : News and Reviews in Molecular, Cellular and Developmental Biology* 15 (8): 531–38. https://doi.org/10.1002/bies.950150806.

Serpe, Mihaela, and Michael B O'Connor. 2006. "The Metalloprotease Tolloid-Related and Its TGF-β-like Substrate Dawdle Regulate Drosophila Motoneuron Axon Guidance." *Development* 133 (24): 4969–79. https://doi.org/10.1242/dev.02711.

Shandala, Tetyana, R. Daniel Kortschak, and Robert Saint. 2002. "The Drosophila Retained/Dead Ringer Gene and ARID Gene Family Function during Development." *International Journal of Developmental Biology* 46 (4 SPEC.): 423–30. https://doi.org/10.1387/ijdb.12141428.

Shandala, Tetyana, Kazunaga Takizawa, and Robert Saint. 2003. "The Dead Ringer/Retained Transcriptional Regulatory Gene Is Required for Positioning of the Longitudinal Glia in the Drosophila Embryonic CNS." *Development* 130 (8): 1505–13. https://doi.org/10.1242/dev.00377.

Singhania, Aditi, and Wesley B Grueber. 2014. "Development of the Embryonic and Larval Peripheral Nervous System of Drosophila." *WIREs Developmental Biology* 3 (3): 193–210. https://doi.org/https://doi.org/10.1002/wdev.135.

Sink, Helen, Edward Jay Rehm, Lee Richstone, Yolanda M Bulls, and Corey S Goodman. 2001. "Sidestep Encodes a Target-Derived Attractant Essential for Motor Axon Guidance in Drosophila." *Cell* 105 (1): 57–67. https://doi.org/https://doi.org/10.1016/S0092-8674(01)00296-3.

Skeath, J B. 1998. "The Drosophila EGF Receptor Controls the Formation and Specification of Neuroblasts along the Dorsal-Ventral Axis of the Drosophila Embryo." *Development* 125 (17): 3301–12.

Skeath, J B, and C Q Doe. 1996. "The Achaete-Scute Complex Proneural Genes Contribute to Neural Precursor Specification in the Drosophila CNS." *Current Biology : CB* 6 (9): 1146–52.

Skeath, J B, Y Zhang, R Holmgren, S B Carroll, and C Q Doe. 1995. "Specification of Neuroblast Identity in the Drosophila Embryonic Central Nervous System by Gooseberry-Distal." *Nature* 376 (6539): 427–30. https://doi.org/10.1038/376427a0.

Skeath, James B., and Scan B. Carroll. 1991. "Regulation of Achaete-Scute Gene Expression and Sensory Organ Pattern Formation in the Drosophila Wing." *Genes and Development* 5 (6): 984–95. https://doi.org/10.1101/gad.5.6.984.

Southall, Tony D., and Andrea H. Brand. 2009. "Neural Stem Cell Transcriptional Networks Highlight Genes Essential for Nervous System Development." *EMBO Journal* 28 (24): 3799–3807. https://doi.org/10.1038/emboj.2009.309.

Spana, E P, and C Q Doe. 1995. "The Prospero Transcription Factor Is Asymmetrically Localized to the Cell Cortex during Neuroblast Mitosis in Drosophila." *Development (Cambridge, England)* 121 (10): 3187–95.

Stork, Tobias, Amy Sheehan, Ozge E Tasdemir-Yilmaz, and Marc R Freeman. 2014. "Neuron-Glia Interactions through the Heartless FGF Receptor Signaling Pathway Mediate Morphogenesis of Drosophila Astrocytes." *Neuron* 83 (2): 388–403. https://doi.org/10.1016/j.neuron.2014.06.026.

Stratmann, Johannes, Helen Ekman, and Stefan Thor. 2019. "Branching Gene Regulatory Network Dictating Different Aspects of a Neuronal Cell Identity," no. March. https://doi.org/10.1242/dev.174300.

Stuart, Tim, Andrew Butler, Paul Hoffman, Marlon Stoeckius, Peter Smibert, Rahul Satija, Tim Stuart, et al. 2019. "Comprehensive Integration of Single-Cell Data Resource Comprehensive Integration of Single-Cell Data." *Cell* 177 (7): 1888-1902.e21. https://doi.org/10.1016/j.cell.2019.05.031.

Sturtevant, M. A., M. Roark, and E. Bier. 1993. "The Drosophila Rhomboid Gene Mediates the Localized Formation of Wing Veins and Interacts Genetically with Components of the EGF-R Signaling Pathway." *Genes and Development* 7 (6): 961–73. https://doi.org/10.1101/gad.7.6.961.

Sun, Qi, Benno Schindelholz, Matthias Knirr, Aloisia Schmid, and Kai Zinn. 2001. "Complex Genetic Interactions among Four Receptor Tyrosine Phosphatases Regulate Axon Guidance in Drosophila." *Molecular and Cellular Neuroscience* 17 (2): 274–91. https://doi.org/https://doi.org/10.1006/mcne.2000.0939.

Swarup, Sharan, and Esther M Verheyen. 2012. "Wnt/Wingless Signaling in Drosophila." *Cold Spring Harbor Perspectives in Biology* 4 (6). https://doi.org/10.1101/cshperspect.a007930.

Thor, Stefan, Siv G E Andersson, Andrew Tomlinson, and John B Thomas. 1999. "A LIM-Homeodomain Combinatorial Code for Motor-Neuron Pathway Selection." *Nature* 397 (6714): 76–80. https://doi.org/10.1038/16275.

Trapnell, Cole, Davide Cacchiarelli, Jonna Grimsby, Prapti Pokharel, Shuqiang Li, Michael Morse, Niall J Lennon, Kenneth J Livak, Tarjei S Mikkelsen, and John L Rinn. 2014. "The Dynamics and Regulators of Cell Fate Decisions Are Revealed by Pseudotemporal Ordering of Single Cells." *Nature Biotechnology* 32 (4): 381–86. https://doi.org/10.1038/nbt.2859.

Ugur, Berrak, Kuchuan Chen, and Hugo J Bellen. 2016. "Drosophila Tools and Assays for the Study of Human Diseases." *Disease Models & Mechanisms* 9 (3): 235–44. https://doi.org/10.1242/dmm.023762.

Verheyen, E M, I Mirkovic, S J MacLean, C Langmann, B C Andrews, and C MacKinnon. 2001. "The Tissue Polarity Gene Nemo Carries out Multiple Roles in Patterning during Drosophila Development." *Mechanisms of Development* 101 (1–2): 119–32. https://doi.org/10.1016/s0925-4773(00)00574-8.

Vitak, Sarah A, Kristof A Torkenczy, Jimi L Rosenkrantz, Andrew J Fields, Lena Christiansen, Melissa H Wong, Lucia Carbone, Frank J Steemers, and Andrew Adey. 2017. "Sequencing Thousands of Single-Cell Genomes with Combinatorial Indexing." *Nature Methods* 14 (January): 302. https://doi.org/10.1038/nmeth.4154.

Vorbrüggen, G, R Constien, O Zilian, E A Wimmer, G Dowe, H Taubert, M Noll, and H Jäckle. 1997. "Embryonic Expression and Characterization of a Ptx1 Homolog in Drosophila." *Mechanisms of Development* 68 (1–2): 139–47. https://doi.org/10.1016/s0925-4773(97)00139-1.

Wang, Ping, Richard F Lyman, Trudy F C Mackay, and Robert R H Anholt. 2010. "Natural Variation in Odorant Recognition Among Odorant-Binding Proteins in Drosophila Melanogaster." *Genetics* 184 (3): 759–67. https://doi.org/10.1534/genetics.109.113340.

Weiss, J B, T Von Ohlen, D M Mellerick, G Dressler, C Q Doe, and M P Scott. 1998. "Dorsoventral

Patterning in the Drosophila Central Nervous System: The Intermediate Neuroblasts Defective Homeobox Gene Specifies Intermediate Column Identity." *Genes & Development* 12 (22): 3591–3602.

Wills, Zachary, Jack Bateman, Christopher A Korey, Allen Comer, and David Van Vactor. 1999. "The Tyrosine Kinase Abl and Its Substrate Enabled Collaborate with the Receptor Phosphatase Dlar to Control Motor Axon Guidance." *Neuron* 22 (2): 301–12. https://doi.org/https://doi.org/10.1016/S0896-6273(00)81091-0.

Winberg, M L, J N Noordermeer, L Tamagnone, P M Comoglio, M K Spriggs, M Tessier-Lavigne, and C S Goodman. 1998. "Plexin A Is a Neuronal Semaphorin Receptor That Controls Axon Guidance." *Cell* 95 (7): 903–16. https://doi.org/10.1016/s0092-8674(00)81715-8.

Wodarz, Andreas, and Wieland B. Huttner. 2003. "Asymmetric Cell Division during Neurogenesis in Drosophila and Vertebrates." *Mechanisms of Development* 120 (11): 1297–1309. https://doi.org/10.1016/j.mod.2003.06.003.

Wurmus, Ricardo, Bora Uyar, Brendan Osberg, Vedran Franke, Alexander Gosdschan, Katarzyna Wreczycka, Jonathan Ronen, and Altuna Akalin. 2018. "PiGx: Reproducible Genomics Analysis Pipelines with GNU Guix." *GigaScience* 7 (12). https://doi.org/10.1093/gigascience/giy123.

Xiang, Charlie C, Eva Mezey, Mei Chen, Sharon Key, Li Ma, and Michael J Brownstein. 2004. "Using DSP, a Reversible Cross-Linker, to Fix Tissue Sections for Immunostaining, Microdissection and Expression Profiling." *Nucleic Acids Research* 32 (22): e185–e185. https://doi.org/10.1093/nar/gnh185.

Yang, X., S. Bahri, T. Klein, and W. Chia. 1997. "Klumphuss, a Putative Drosophila Zn-Finger Transcription Factor, Acts to Differentiate between the Identities of Two Secondary Precursor Cells within One Neuroblast Lineage." *Doe, Hall* 234.

Yoshikawa, Shingo, Randall D McKinnon, Michelle Kokel, and John B Thomas. 2003. "Wnt-Mediated Axon Guidance via the Drosophila Derailed Receptor." *Nature* 422 (6932): 583–88. https://doi.org/10.1038/nature01522.

Yu, Fengwei, Chay T. Kuo, and Yuh Nung Jan. 2006. "Drosophila Neuroblast Asymmetric Cell Division: Recent Advances and Implications for Stem Cell Biology." *Neuron* 51 (1): 13–20. https://doi.org/10.1016/j.neuron.2006.06.016.

Yu, Fengwei, Chin Tong Ong, William Chia, and Xiaohang Yang. 2002. "Membrane Targeting and Asymmetric Localization of Drosophila Partner of Inscuteable Are Discrete Steps Controlled by Distinct Regions of the Protein." *Molecular and Cellular Biology* 22 (12): 4230–40. https://doi.org/10.1128/MCB.22.12.4230-4240.2002.

Yu, Hung-Hsiang, Houmam H Araj, Sherry A Ralls, and Alex L Kolodkin. 1998. "The Transmembrane Semaphorin Sema I Is Required in Drosophila for Embryonic Motor and CNS Axon Guidance." *Neuron* 20 (2): 207–20. https://doi.org/https://doi.org/10.1016/S0896-6273(00)80450-X.

Zarin, Aref Arzan, Jamshid Asadzadeh, Karsten Hokamp, Daniel McCartney, Long Yang, Greg J. Bashaw, and Juan-Pablo Labrador. 2014. "A Transcription Factor Network Coordinates Attraction, Repulsion, and Adhesion Combinatorially to Control Motor Axon Pathway Selection." *Neuron* 81 (6): 1297–1311. https://doi.org/https://doi.org/10.1016/j.neuron.2014.01.038.

Zarin, Aref Arzan, Amanda C Daly, Jörn Hülsmeier, Jamshid Asadzadeh, and Juan-Pablo Labrador. 2012. "A GATA/Homeodomain Transcriptional Code Regulates Axon Guidance through the Unc-5 Receptor." *Development* 139 (10): 1798–1805. https://doi.org/10.1242/dev.070656.

Zeng, Yi Arial, Maryam Rahnama, Simon Wang, Worlanyo Sosu-Sedzorme, and Esther M Verheyen. 2007. "Drosophila Nemo Antagonizes BMP Signaling by Phosphorylation of Mad and Inhibition

of Its Nuclear Accumulation." *Development (Cambridge, England)* 134 (11): 2061–71. https://doi.org/10.1242/dev.02853.

Zinzen, Robert P, Jessica Cande, Matthew Ronshaugen, Dmitri Papatsenko, and Mike Levine. 2006. "Evolution of the Ventral Midline in Insect Embryos." *Developmental Cell* 11 (6): 895–902. https://doi.org/10.1016/j.devcel.2006.10.012.

11 APPENDIX

11.1 Supplementary Note 1 - Midline Neuroblast temporal cascade

Midline neuroblasts are formed during the same period as the VNC neuroblast, but have distinct regulatory inputs (Kearney et al. 2004; Crews 2019; Zinzen et al. 2006) and division mode, similar to *type-0 (*i.e. *type-zero)* neuroblasts (see introduction section 5.3.5 for more details). These neuroblasts divide only once, giving rise to two daughter cells (Crews 2019) that differentiate directly and without self-renewal of the stem cell into intermediate progenitor states. With the sole exception of one midline-derived neuroblast – which is actually and confusingly called "*midline-derived neuroblast"* (which I will refer to only by its acronym MNB to distinguish it from other midline-derived neuroblasts) – these cells divide either into two neurons *or* two glia, not one of each. Only the MNB appears to produce the intermediate cell state called "ganglion mother cell" that gives rise to neurons (Kearney et al. 2004; Crews 2019).

These cells are under the same antero-posterior regulation as the trunk neuroblasts, but are dorsoventrally distinct in that they express midline markers such as *sim, rho, and sog* (Crews 2019; Kearney et al. 2004). Another important difference is that it has been demonstrated that the temporal cascade of transcription factors differs from more dorsally originating neuroblasts: midline-derived neuroblasts express the lineage factors *Kr, pdm2* and *cas*, but not in the same order and they do not seem to play the same role (Kearney et al. 2004). In midline neuroblasts, *Kr for example* is known to regulate cell fate and is limited to a subset of midline neuroblasts (Kearney et al. 2004).

To verify this distinct expression of the temporal factors in the midline, I performed *in situ* hybridization for several of the transcription factors of the canonical temporal cascade together with staining for *sim* expression as a midline marker, and *wor* as a neuroblast marker (Figure 11-1A). The earliest canonical temporal transcription factor *hb* is not expressed anywhere in the midline in early stages (Figure 11-1A first panel, stage 9); however, as there are no sim-positive wor-positive cells at stage 9 and given wor being a universal embryonic neuroblast marker, neuroblast delamination in the midline must start significantly later than in the columnar domains. However, I was able to observe sim-positive wor-positive cells by stage 10, *hb*-positive midline cells were not observed (Figure 11-1A). This shows, that the temporal transcription factor cascade in the midline does not commence with *hb* expression as in the columnar domains.

The second canonical temporal transcription factor, *Kr*, is expressed in some of the midline neuroblasts from the earliest stages of neuroblast delamination and this expression is maintained throughout neuroblast development (Figure 11-1B). To better understand the regulatory differences of *Kr* between the trunk neuroblasts and the midline neuroblasts, I explored possible differential enhancer

usage to regulate *Kr* expression in the midline vs other neuroblasts between the two populations. Using the REDFly Database (v.9.2.0, (Gallo et al. 2006)), I searched for regulatory elements for Kr and found a Vienna Tile element, VT22277 (Kvon et al. 2014), that drives specific expression in some of the midline cells (Figure 11-1D). This regulatory element has not yet been explored in terms of the mechanisms that drive its selective expression, but it could very well explain the distinct expression of *Kr* between the two neuroblast populations, in that a spatial *cis*-regulatory element drives stable expression of *Kr* in midline neuroblasts (possibly under control of midline-specific transcription factors such as sim), while the more canonical temporal regulation in columnar domains is regulated via temporally engaged cis-regulatory elements. and a difference in regulatory elements might explain the absence of the temporal transcription factors in these cells.

The last canonical temporal transcription factor is *cas*, which plays a major role in cell fate determination of late trunk neuroblasts' transition from a *type I* to *type 0* division mode (see introductory section 5.3.5) and cell cycle exit as it has been shown to be one of the determinants of the termination of neuroblast proliferation (Crews 2019). The specific role this factor plays in the midline neuroblasts is not clear, but it has been suggested that it controls the GMC formation of the MNB (Kearney et al. 2004). Visualization of *cas* expression together with *sim* and *wor* shows early expression of *cas* in a subset of midline cells before the neuroblasts emerge (Figure 11-1C, no co-expression with *wor* at stage 9). This strong *cas* expression is maintained in a subset of midline neuroblasts when these emerge.

I have shown that the canonical temporal transcription factors display an expression profile in the major neurogenic columns (i.e. ventral column, intermediate column, lateral column) that is clearly distinct from their midline expression, where

- *Kr* is expressed in a subset of midline neuroblasts that accordingly with previous work will originate neurons (Kearney et al. 2004),
- *cas* is expressed from early stages of development in a subset of midline cells before any neuroblasts in the main neurogenic columns express *cas*,
- and *hb* – the first temporal transcription factor in all main column neuroblasts – is not expressed in the midline neuroblasts.

Though these transcription factors have unique temporal expression profiles in the midline, they are still involved in cell fate determination. In the neuroblasts derived from the three main trunk columns, cell fate determination by these temporal transcription factors is achieved – at least in part – by sustained expression of the respective temporal transcription factors in the ganglion mother cells. In contrast, the midline neuroblasts may not require a temporal cascade of factors due to their division mode: unlike column-derived neuroblasts, midline neuroblasts do not go through successive divisions where they generate distinct progeny. After they delaminate, they simply divide once into two daughter cells. This indicates the intricacies and need for fine regulation of gene expression in the construction of a complex

nervous system, where the same genes might be regulated differently to play similar roles to achieve distinct outcomes.

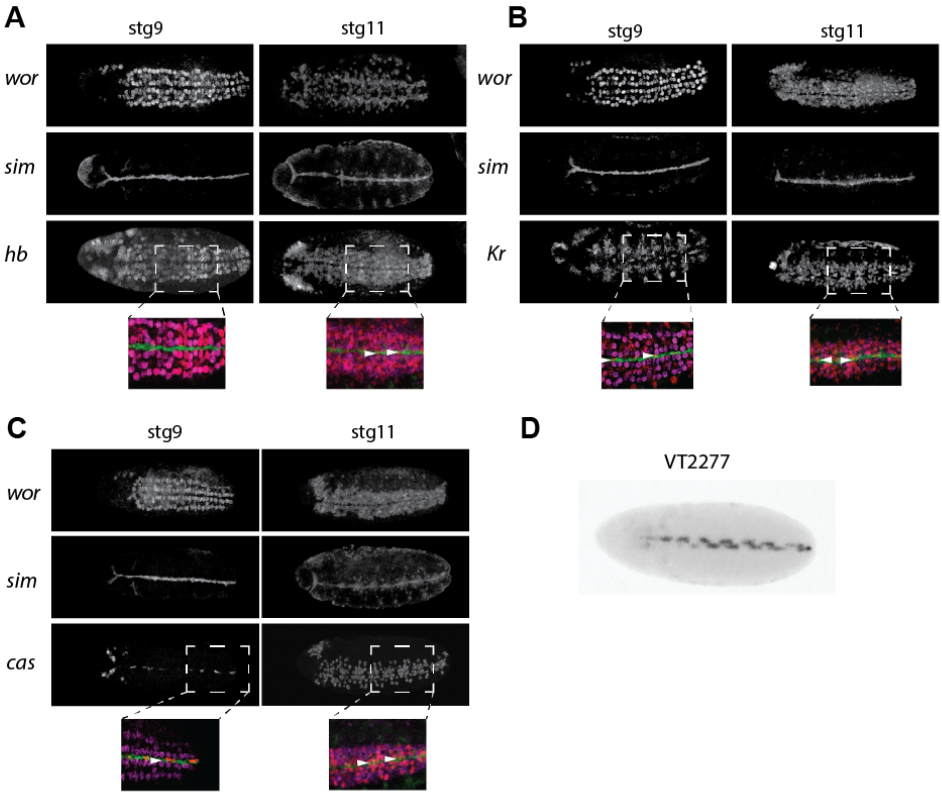
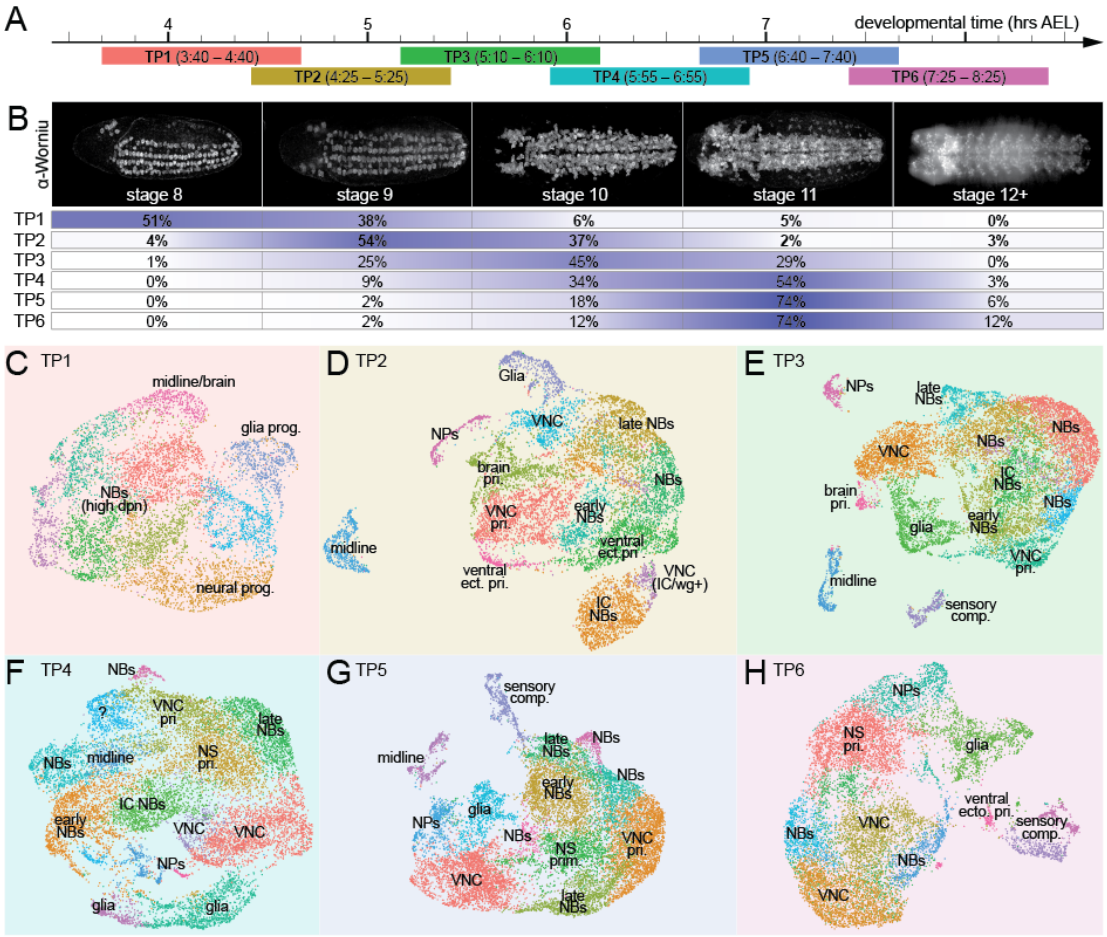


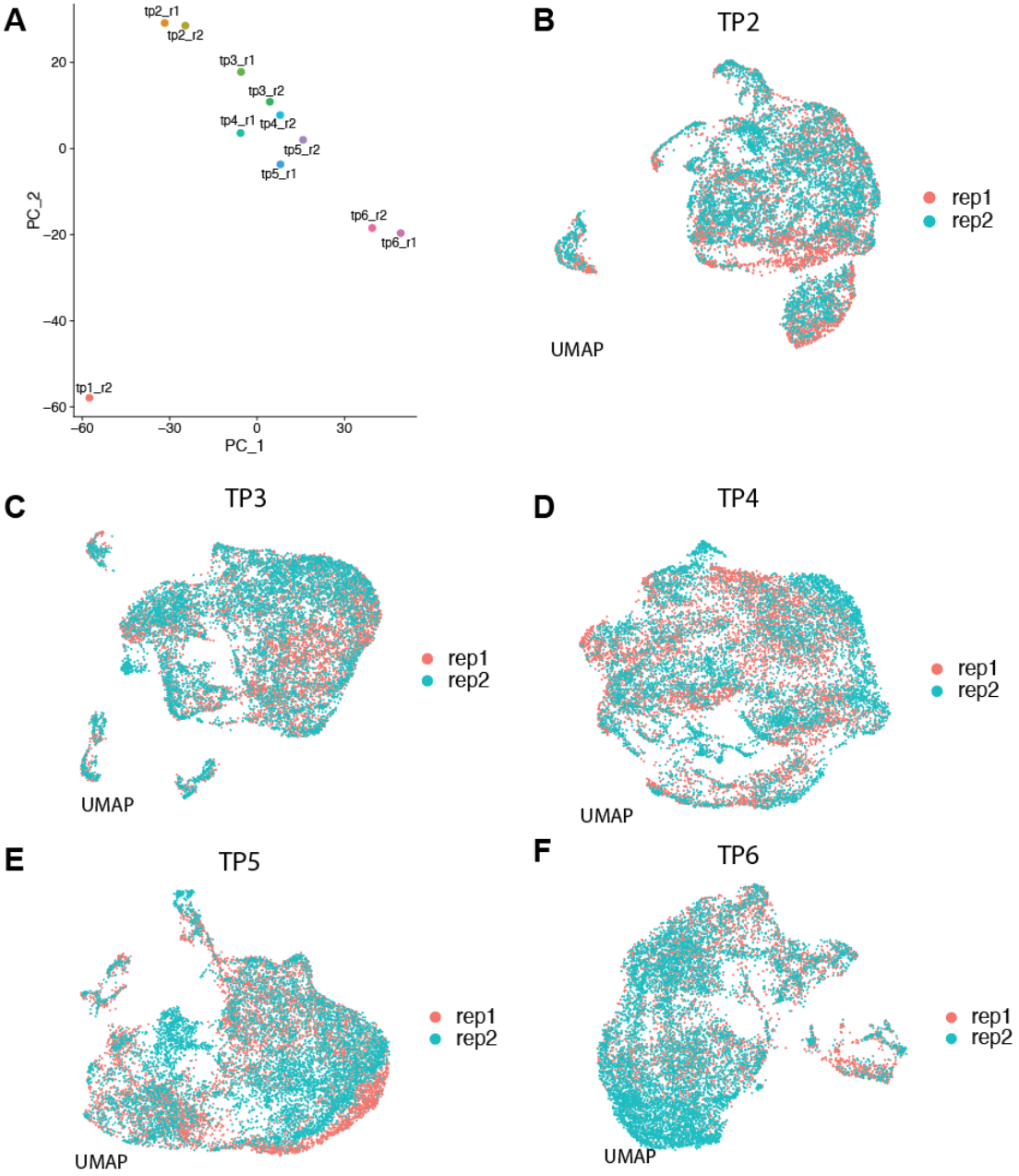
Figure 11-1 Temporal transcription factors in the midline

(**A**)–(**C**)Co-expression of temporal transcription factors (*hb*, *Kr* and *cas*) with *wor* (neuroblast marker) and *sim* (midline marker) shows early expression of late factors (*cas*) in the midline neuroblast and early factors (*hb*) only in late stages. (**D**) Specific Kr enhancer for the midline might explain the distinct pattern of expression in these neuroblast (image from Fly enhancers from the Stark lab (Gallo et al. 2006)).



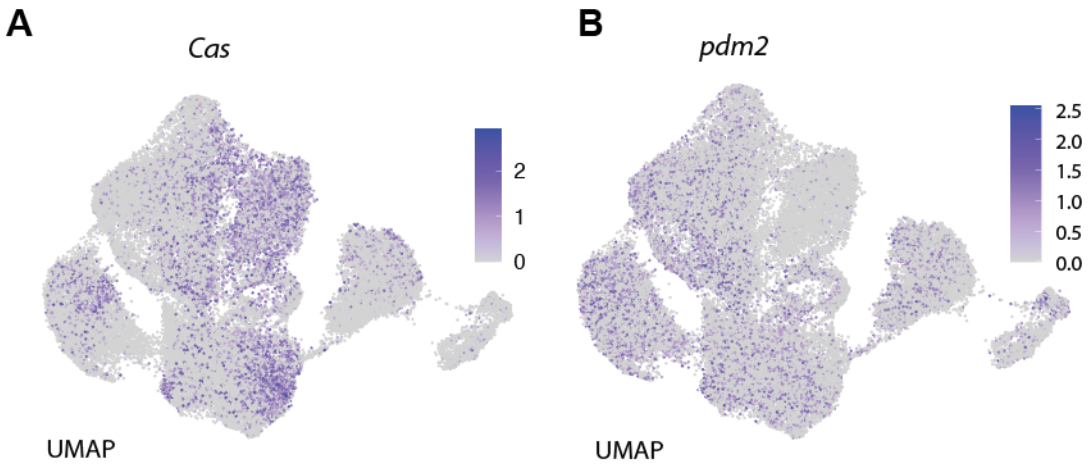
Supplementary Figure S 1 Embryo stage distribution along development

(**A**) Developmental timeline of the collection timepoints for the neuroblast collections. (**B**) Worniu stainings of the neuroblast for each stage of neuroblast delamination with percentage of each stage per timepoint. (**C**)-(**H**) Individual UMAPs of the six timepoints with cluster labels.



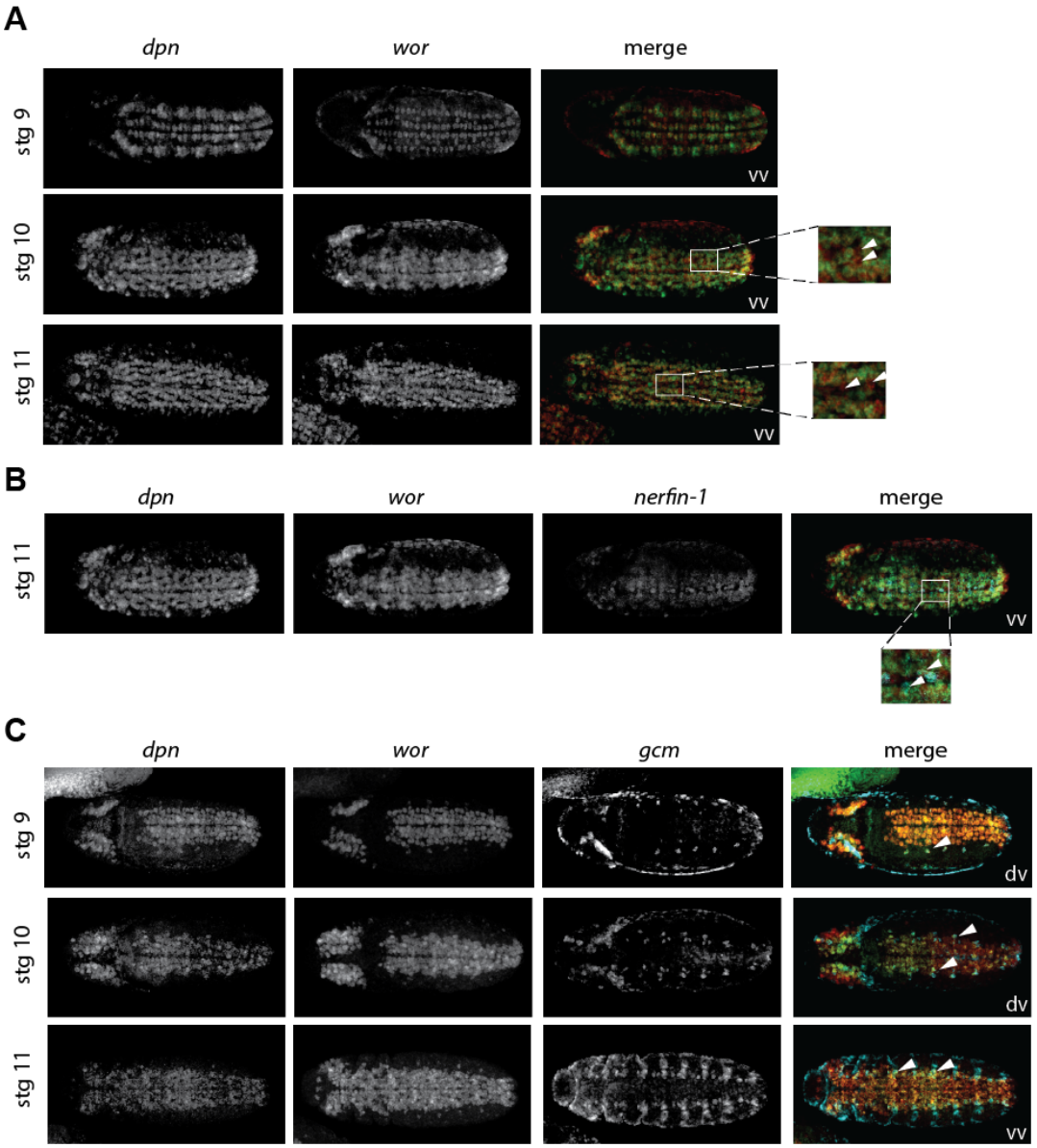
Supplementary Figure S 2 Reproducibility of replicates of scRNAseq data

(**A**) Pseudo-bulk analysis of all samples shows good reproducibility between replicates of the same timepoint and a progressive separation of samples according to the timepoint along PC_2. (**B-F**) UMAPs of each individual time point shows no batch effect between replicates.



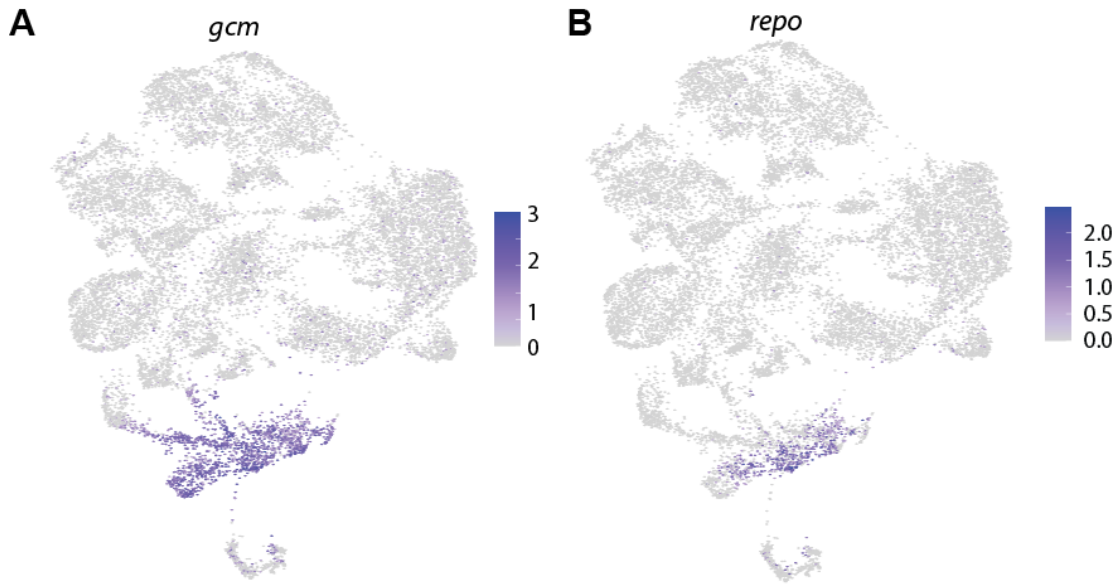
Supplementary Figure S 3 *Castor* and *Pdm2* expression across all cells

(**A**) UMAP with *cas* expression in the entirety of the neuroblast data shows enrichment in the late neuroblast cluster and VNC primordium (see Figure 7-5). (**B**) UMAP with *pdm2* expression in the neuroblast data shows expression in the Intermediate neuroblast cluster (see Figure 7-5).



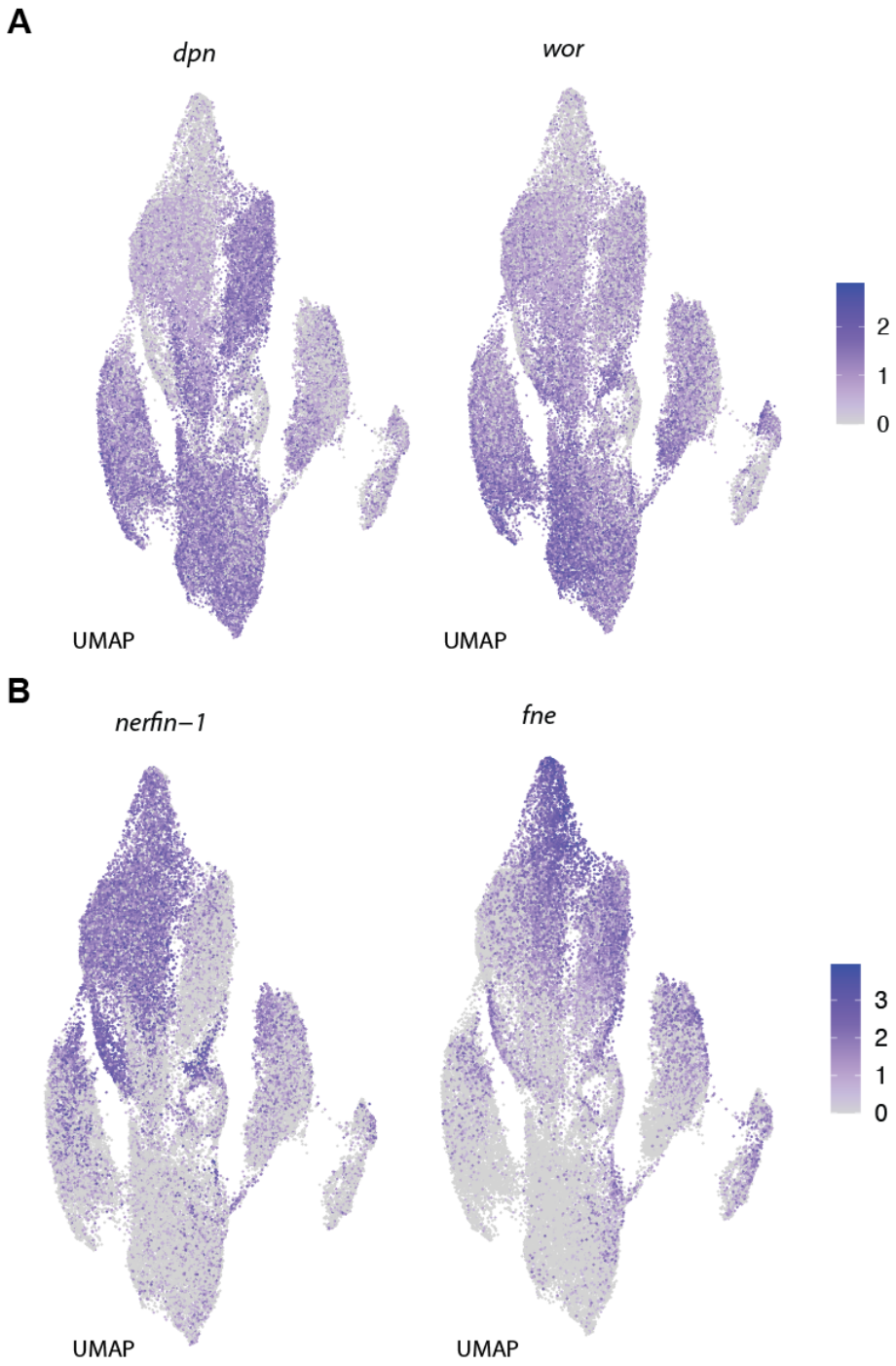
Supplementary Figure S 4 Expression of lineage markers in the progenitors

(**A**) Expression of *dpn* and *wor* along development shows a perfect overlap in early stages (top panel), in later time points (middle and bottom panel) cells that are only positive for *wor* emerge, these are the GMCs (see inset). (**B**) Expression of neuronal marker *nerfin-1* in progenitor cells. *Nerfin-1* is co-expressed with *dpn* and *wor* in later stages of neuroblast delamination. (**C**) Expression of the glia marker (*gcm*) along neuroblast delamination. Expression of *gcm* overlaps with *dpn* and *wor* from earlier stages of neuroblast delamination (stage 9, top panel); this overlap only increases as more neuroblast emerge (middle and bottom panels). All ISHs were performed by Laura Wandelt.



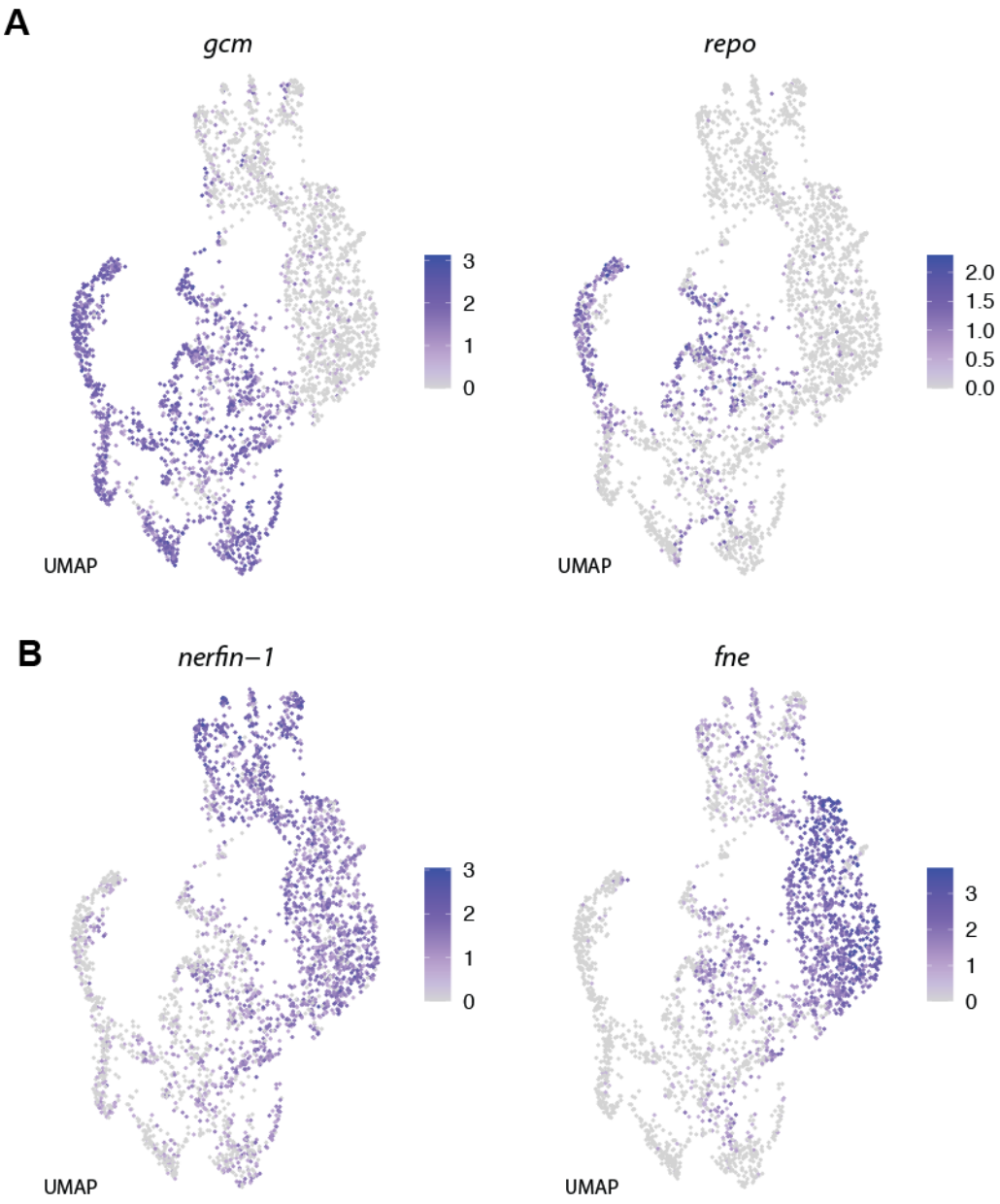
Supplementary Figure S 5 Identification of glia progenitors from the NB pool

(**A**) Expression of *gcm* in the NBs is restricted to the "LC" and "LC2" clusters (compare with Figure 7-6). (**B**) Expression of repo is restricted to cluster "LC2" indicating that these are glia progenitors as *repo* is a later marker of glia cells (see Figure 7-6).



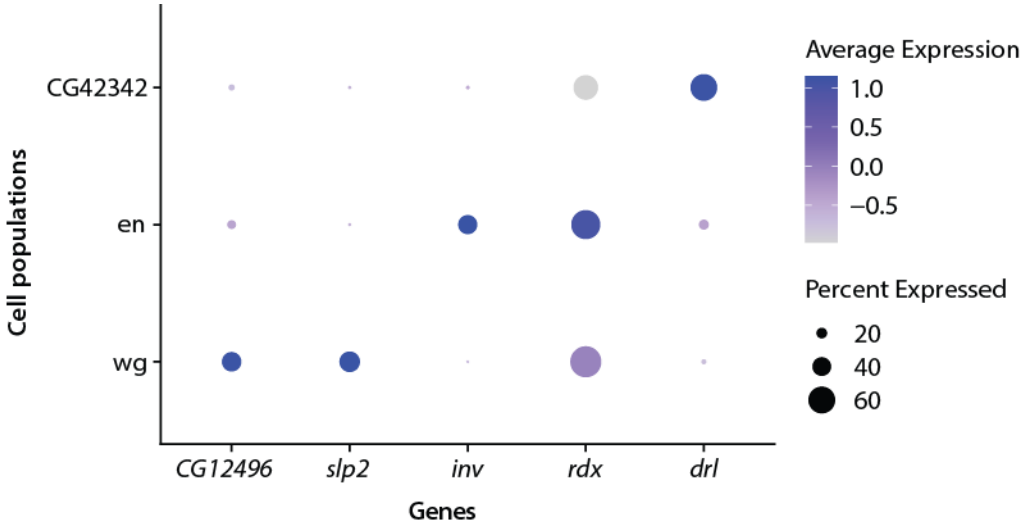
Supplementary Figure S 6 Neuroblast markers in all timepoints

(**A**) UMAP of all cells highlights the expression of *dpn* (left) and *wor* (right); *both markers are* expressed throughout the UMAP, though differences can be observed populations identified as NBs and NPs (compare with Figure 7-5). (**B**) Neuronal markers *fne* and *nerfin-1* show strong enrichment in the cluster identified as neuronal progenitors (NPs) (compare with Figure 7-5).



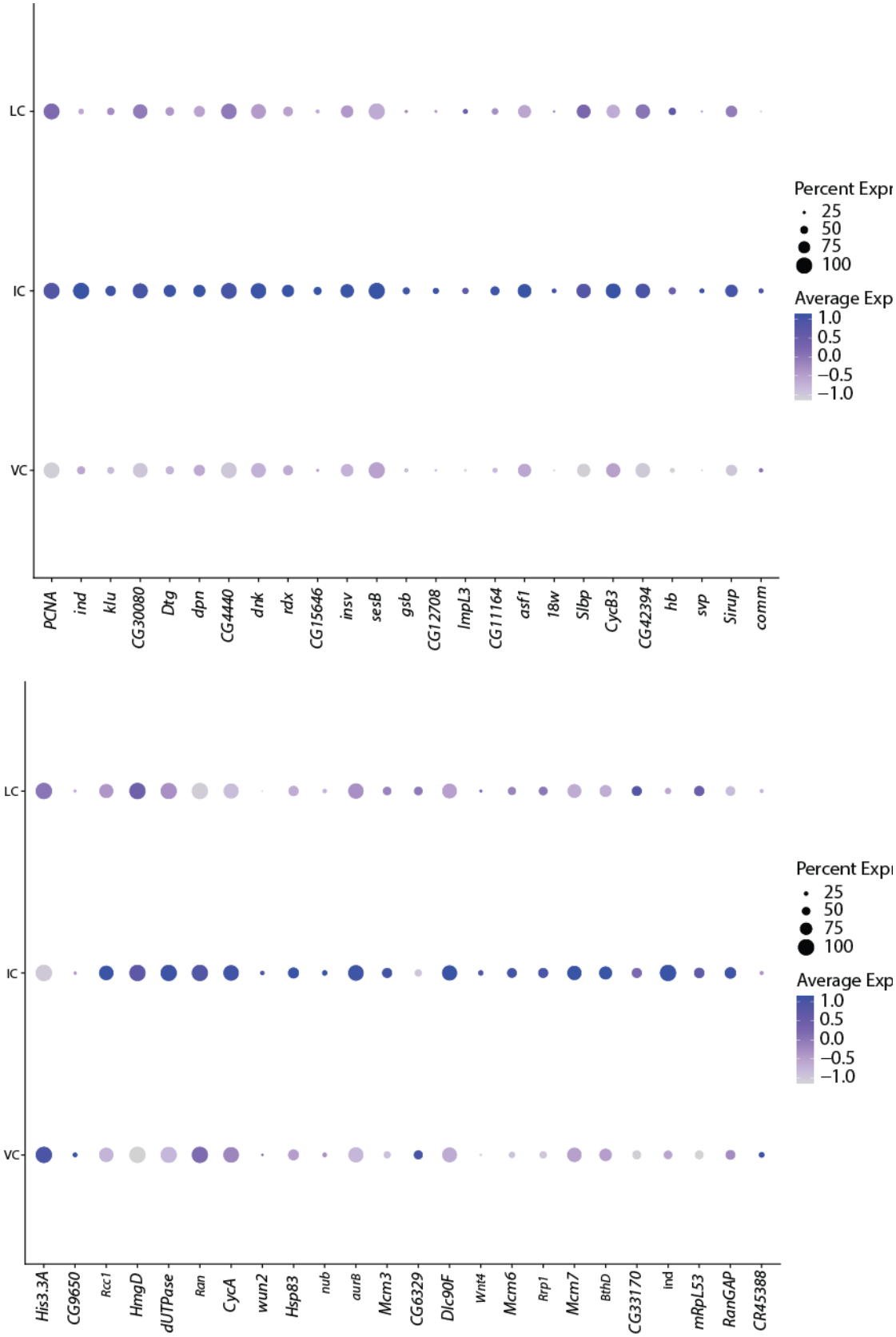
Supplementary Figure S 7 Identification of glia and neuronal precursors

(**A**) UMAP of expression of glia markers *gcm* and *repo* in the GMC population (see Figure 7-7*)*. (**B**) UMAP of expression of neuronal markers *nerfin-1* and *fne* in the GMC population (see Figure 7-7).



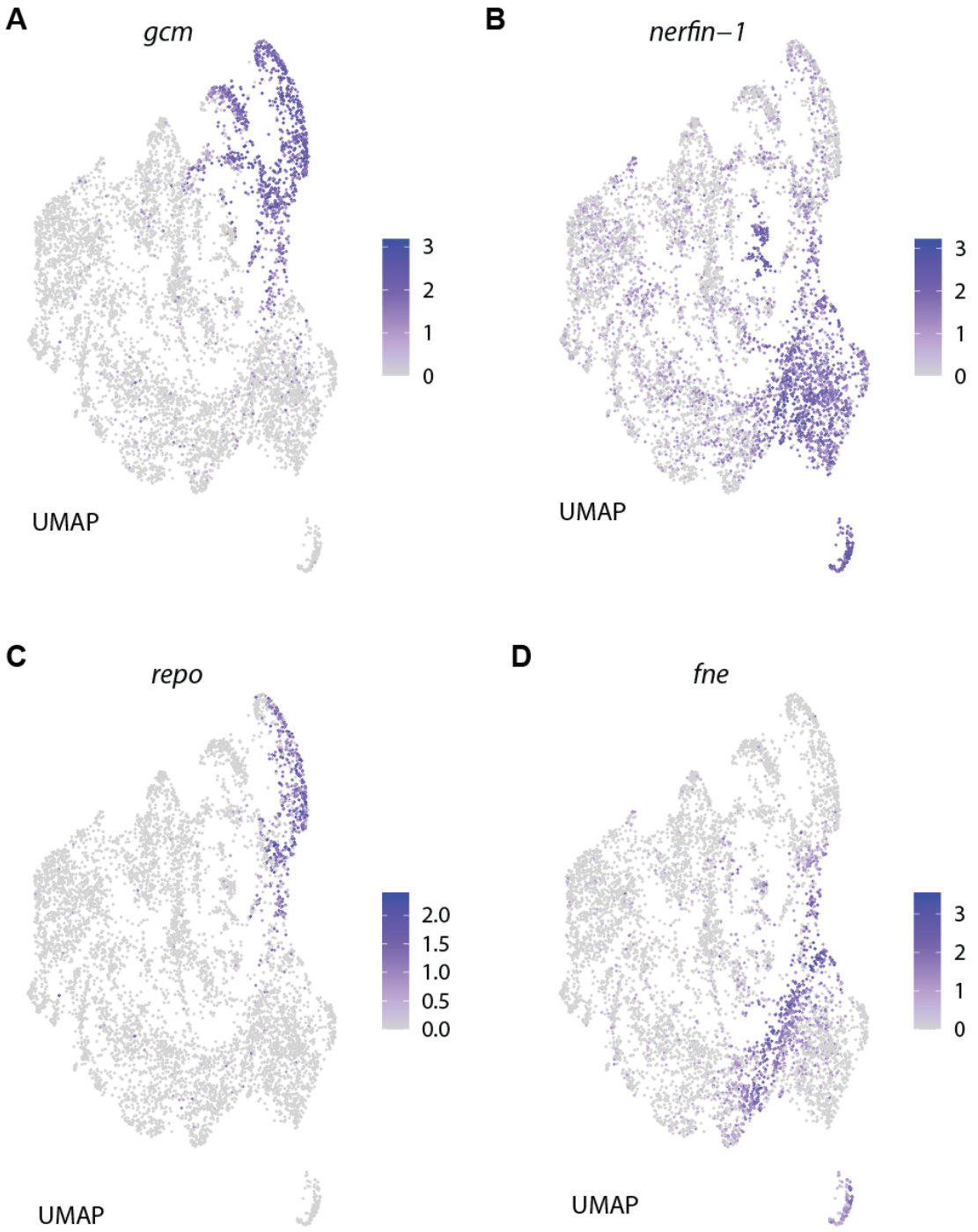
Supplementary Figure S 8 Expression of new anteroposterior markers in the different cell populations

Dot-plot showing the exclusive/enriched expression of the new markers tested for the anteroposterior domains.



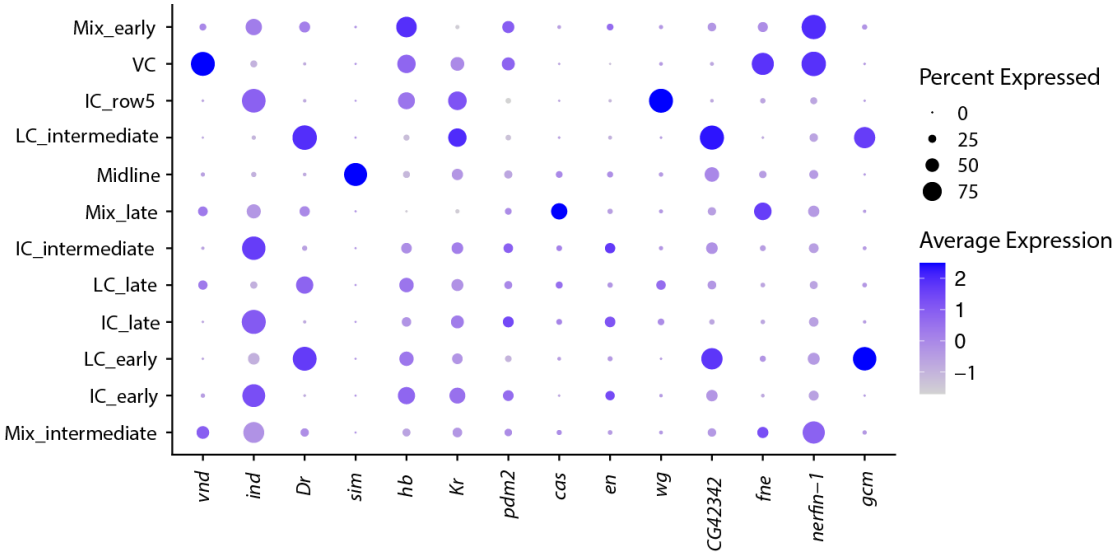
Supplementary Figure S 9 Enriched genes in the IC are not exclusive to its domain of the neuroectoderm

Dot-Plots showing the top enriched genes in the IC population. There is no specific expression of markers in the IC the genes are also expressed in the adjacent domains, VC and LC.



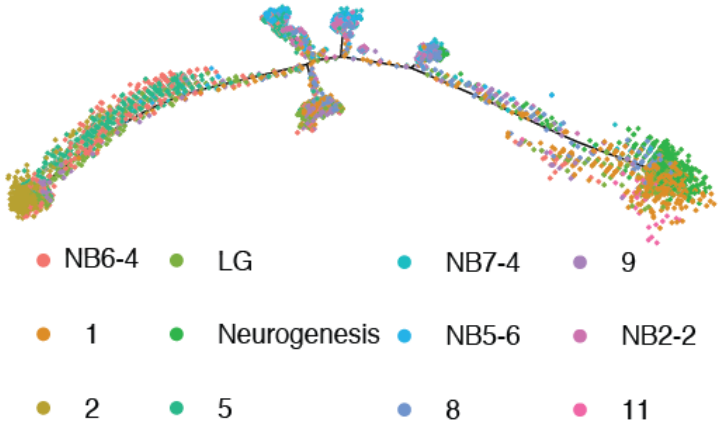
Supplementary Figure S 10 Expression of later precursor markers in the DV populations

(**A-D**) Expression of later markers of glia (A/C) and neuronal (B/D) precursor show expression in the later timepoint. Glia markers (A/C) are expressed in the LC domain exclusively but the neuronal marker *nerfin-1* (B) is expressed where the column identities merge, indicating a cellular fate identity overwriting a spatial one.



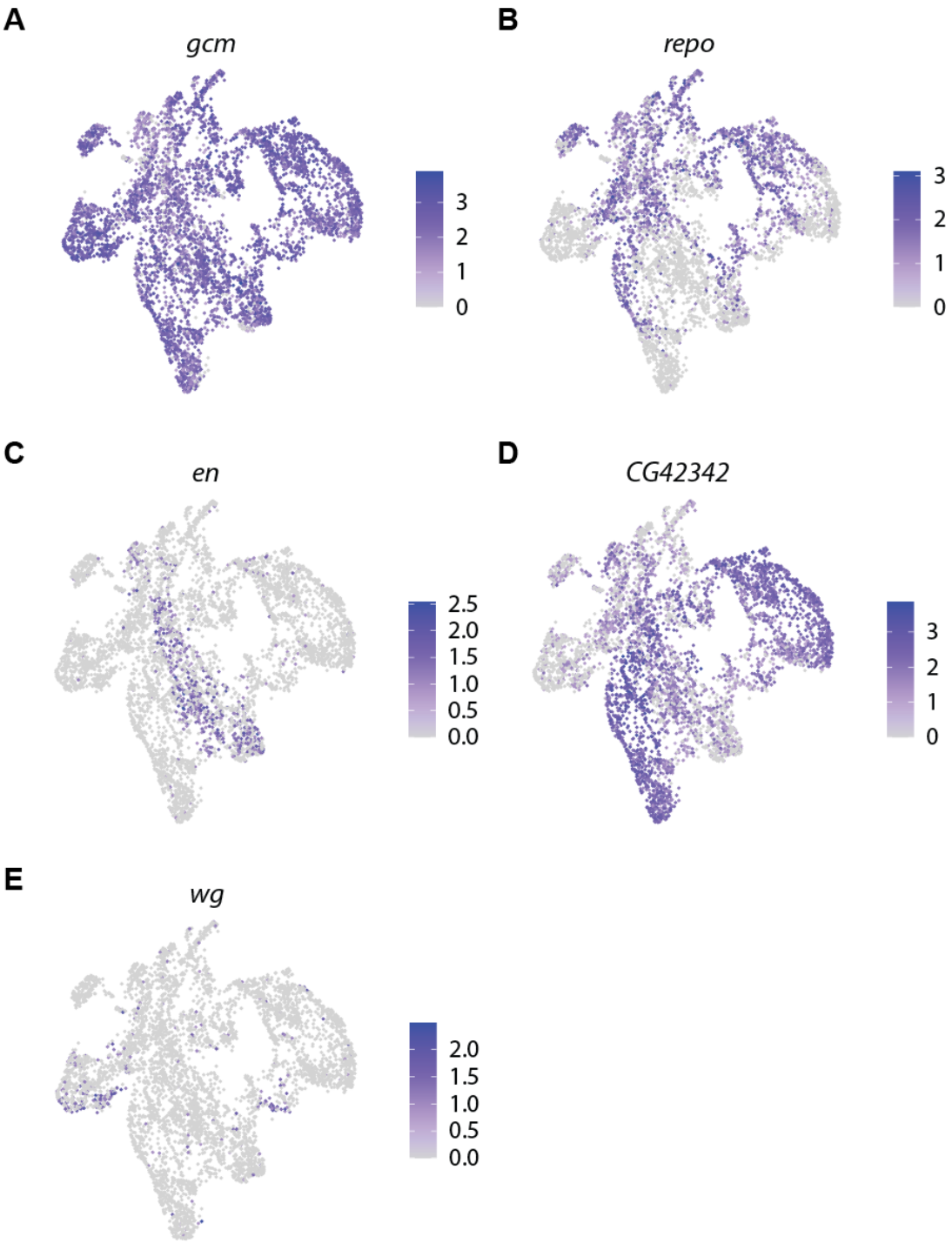
Supplementary Figure S 11 Cluster assignment for dorsoventral populations

Dot-plot of genes used for cluster assignment of the dorsoventral populations. Distinct clusters were identified for the four dorsoventral populations, Midline, ventral column (VC), intermediate column (IC) and lateral column (LC); clusters where the dorsoventral populations do not separate are labeled as "Mix". Further division of the column cluster were made based on expression temporal transcription factors (*hb* and *Kr* – early; *pdm2* – intermediate, *pdm2* and *cas* – late).



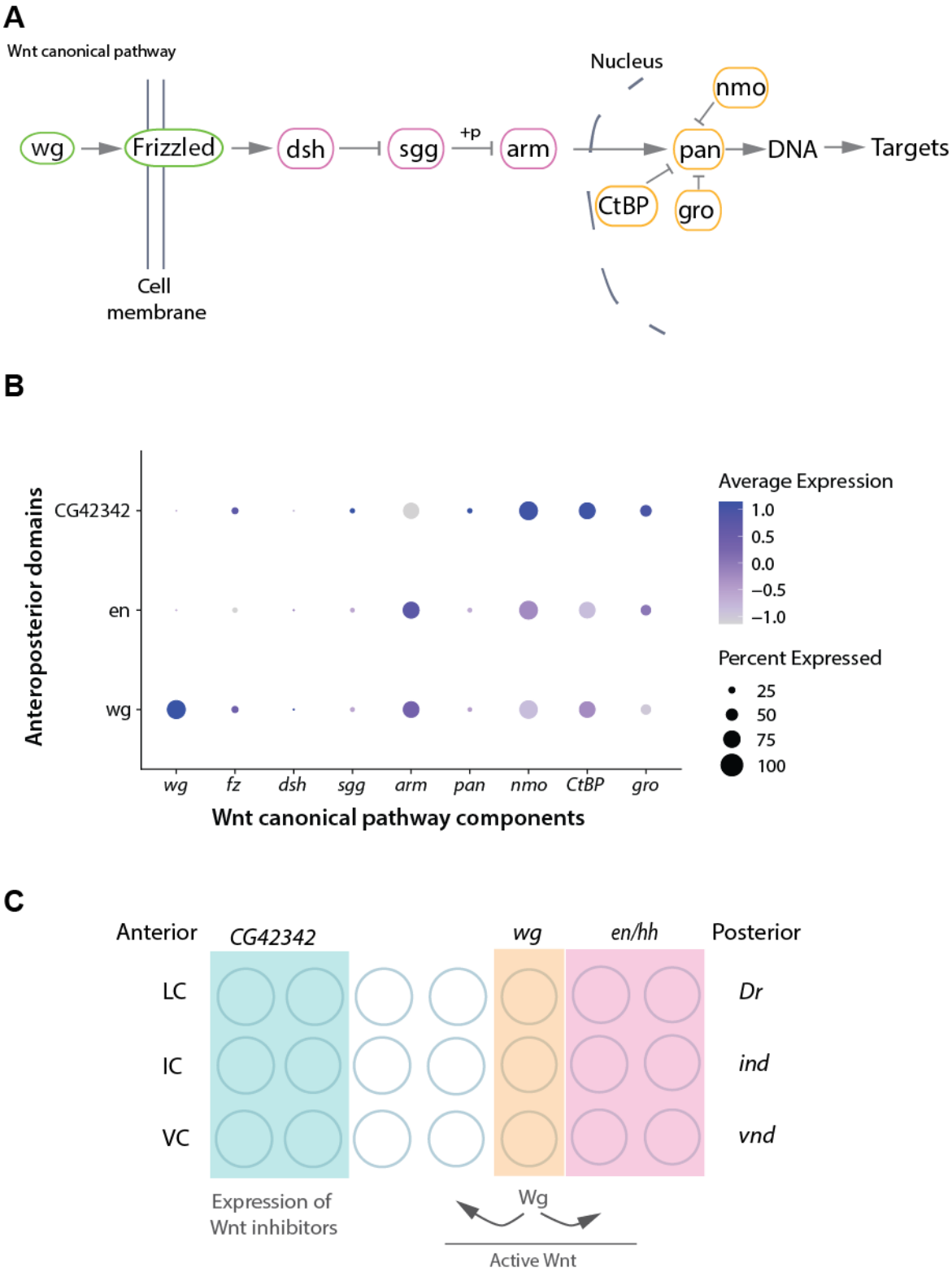
Supplementary Figure S 12 Glia population pseudotime

Pseudotime trajectory plot of the glia population labelled by cluster (see Figure 7-13).



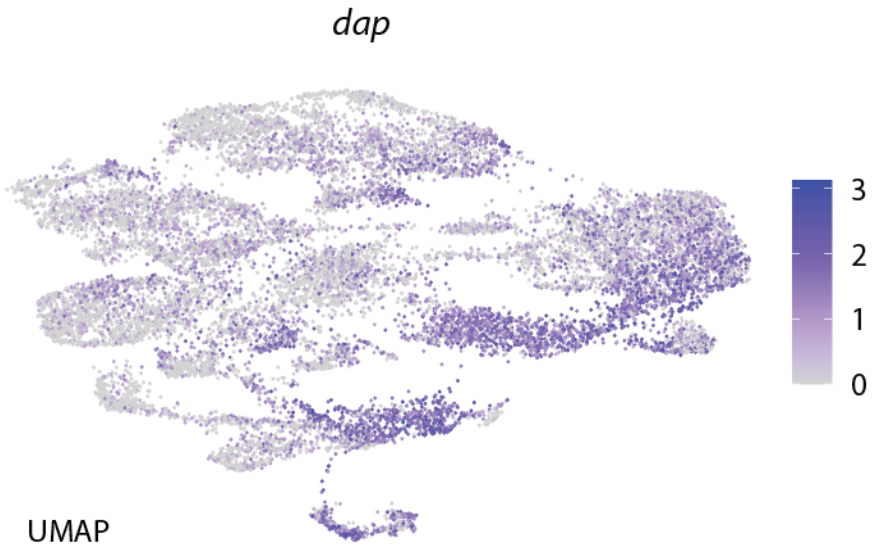
Supplementary Figure S 13 Glia and AP axis markers in the glia population

(**A**) Expression of *gcm* indicates that all cells exhibit at least partial glial identity. (**B**) *repo* expression in a subset of the cells indicates differential induction of Gcm targets in distinct cell populations. (**C-E**) Glia neuroblast emerge from distinct AP domains. (**C**) *en* domain allows for the identification of neuroblasts 6-4 and 7-4. (**D**) The *CG42342* domain from where most of the glia neuroblasts emerge (LG, NB1-1, NB2-2, NB2-5). (**E**) *wg* domain from where one neuroblast that gives rise to glia emerges (NB5-6).



Supplementary Figure S 14 Regulation of Wnt pathway in the embryonic segment

(**A**) Schematic of the canonical Wnt pathway in *Drosophila* showing the major components. (**B**) Dot-plot showing the expression of the Wnt pathway components in the different anteroposterior domains. (**C**) Schematic of the segment and its anteroposterior domains as well as where the Wnt pathway is active or inhibited.



Supplementary Figure S 15 Expression of *dacapo* (cell cycle inhibitor) in the neuroblasts (dpn-positive cells).

Feature Plot showing *dacapo* expression in the clusters (cell populations) where the regulon E2f1 is not expressed (see Figure 7-14), showing that expression of cell cycle inhibitors downregulates the expression of cell cycle genes.

11.3 Supplementary Tables

Supplementary Table S 1 Sequencing sample information

General information on all the sequencing samples generated in this work.

Sample	FACS	Number of reads (millions)	Number of UMI (millions)	Final number of cells
whole_embryo	No	442	98	7694
cells_fresh_rep1	No	46	40	6636
cells_fresh_rep2	No	45	39	
cells_fix_rep1	No	48	41	2631
cells_fix_rep2	No	46	41	
nuclei_fresh_rep1	No	50	42	1579
nuclei_fresh_rep2	No	47	40	
nuclei_fix_rep1	No	45	39	2997
nuclei_fix_rep2	No	55	44	
nbs_tp1_rep1	Sorted for wor (Alexa 555)	151	101	3250
nbs_tp1_rep2	Sorted for wor (Alexa 555)	132	115	3971
nbs_tp2_rep1	Sorted for wor (Alexa 555)	151	128	5108
nbs_tp2_rep2	Sorted for wor (Alexa 555)	125	101	4646
nbs_tp3_rep1	Sorted for wor (Alexa 555)	125	100	2242
nbs_tp3_rep2	Sorted for wor (Alexa 555)	126	102	6459
nbs_tp4_rep1	Sorted for wor (Alexa 555)	148	135	6428
nbs_tp4_rep2	Sorted for wor (Alexa 555)	125	102	6674
nbs_tp5_rep1	Sorted for wor (Alexa 555)	113	103	6255

nbs_tp5_rep2	Sorted for wor (Alexa 555)	101	77	7169
nbs_tp6_rep1	Sorted for wor (Alexa 555)	125	102	3949
nbs_tp6_rep2	Sorted for wor (Alexa 555)	125	101	7356

Supplementary Table S 2 Table of all oligos used in this work

Table shows oligo name, sequence and its purpose.

Name	Sequence	Description
cas_fwd_ish	AGTCCCAATATGACCTCCACG	Specific for *cas* cDNA for ISH probe.
cas_rev_t7_ish	NNNNTAATACGACTCACTATAGG GCGGCTGGCAGTTGGATTTTC	Specific for *cas* cDNA for ISH probe with T7 sequence.
gcm_fwd_ish	AGCCAGTGTGCATCTGAGAC	Specific for *gcm* cDNA for ISH probe.
gcm_rev_t7_ish	NNNNTAATACGACTCACTATAGG GTACGAGCTCCCTTGCTTGAC	Specific for *gcm* cDNA for ISH probe with T7 sequence.
nerfin1_fwd_ish	ACACCGATACCGATTCCAGC	Specific for *nerfin-1* cDNA for ISH probe.
nerfin1_rev_t7_ish	NNNNTAATACGACTCACTATAGG GTCCACAATCTGAAGCCGTCC	Specific for *nerfin-1* cDNA for ISH probe with T7 sequence.
dpn_fwd_ish	TATGGAGAAACGTCGCCGAG	Specific for *dpn* cDNA for ISH probe.
dpn_rev_t7_ish	NNNNTAATACGACTCACTATAGG GATGGCGCCAGCTAGGAATTT	Specific for *dpn* cDNA for ISH probe with T7 sequence.
Obp99a_fwd_ish	GTTTTCGTTGCCATCTGCGT	Specific for *Obp99a* cDNA for ISH probe.
Obp99a_rev_t7_ish	NNNNTAATACGACTCACTATAGG GTCCCCACTGAATCGAGAAGGA	Specific for *Obp99a* cDNA for ISH probe with T7 sequence.
opa_fwd_ish	CTCCTCGGATCGGAAGAAGC	Specific for *opa* cDNA for ISH probe.
opa_rev_t7_ish	NNNNTAATACGACTCACTATAGG GTGGGAGGGGATGACGAGAAT	Specific for *opa* cDNA for ISH probe with T7 sequence.
en _fwd_ish	GGGCAGGTGACAAGGCTAAG	Specific for *en* cDNA for ISH probe.
en_rev_t7_ish	NNNNTAATACGACTCACTATAGG GCATCCGAGGACTCCATTCCG	Specific for *en* cDNA for ISH probe with T7 sequence.
wg _fwd_ish	TTCACTCCTCCGCTCGAAAC	Specific for *wg* cDNA for ISH probe.
wg_rev_t7_ish	NNNNTAATACGACTCACTATAGG GGGCGAAGGCTCCAGATAGAC	Specific for *wg* cDNA for ISH probe with T7 sequence.
CG42342_fwd_ish	GCTGGGCTATGCGGAGATAG	Specific for *CG42342* cDNA for ISH probe.
CG42342_rev_t7_ish	NNNNTAATACGACTCACTATAGG GTTTTCAAGCGGCTTTCCTGC	Specific for *CG42342* cDNA for ISH probe with T7 sequence.
CG12496_fwd_ish	AGATTCCCGTGCACGATCAG	Specific for *CG12496* cDNA for ISH probe.
CG12496_rev_t7_ish	NNNNTAATACGACTCACTATAGG GGCTCGCTTTCTCGTCTCCAT	Specific for *CG12496* cDNA for ISH probe with T7 sequence.
slp2_fwd_ish	GAGCAACGGACGTGAAATCG	Specific for *slp2* cDNA for ISH probe.
slp2_rev_t7_ish	NNNNTAATACGACTCACTATAGG GCTACGGTGACGGGTTTGTGA	Specific for *slp2* cDNA for ISH probe with T7 sequence.
inv_fwd_ish	CTATCGACCGTCGGCATTCA	Specific for *inv* cDNA for ISH probe.

Name	Sequence	Description
inv_rev_t7_ish	NNNNTAATACGACTCACTATAGG GAGCAACAAAAACGGCGTGTA	Specific for *inv* cDNA for ISH probe with T7 sequence.
rdx_fwd_ish	GATTGGATTCGCCATTCGCC	Specific for *rdx* cDNA for ISH probe.
rdx_rev_t7_ish	NNNNTAATACGACTCACTATAGG GGGCTCGCTATGCTTTTGCTC	Specific for *rdx* cDNA for ISH probe with T7 sequence.
drl_fwd_ish	GCCTAATCGGAGTCTCGGC	Specific for *drl* cDNA for ISH probe.
drl _rev_t7_ish	NNNNTAATACGACTCACTATAGG GTAGACGTTTTCAGGCCGCTT	Specific for *drl* cDNA for ISH probe with T7 sequence.
Ilp4_fwd_ish	TTTTATGCCCGGTGAGAGCA	Specific for *Ilp4* cDNA for ISH probe.
Ilp4_rev_t7_ish	NNNNTAATACGACTCACTATAGG GCTGCTGCTGCTTGGAGAGTAA	Specific for *Ilp4* cDNA for ISH probe with T7 sequence.
CG10479_fwd_ish	GGGTCGCCTGTTCCAATACA	Specific for *CG10479* cDNA for ISH probe.
CG10479_rev_t7_ish	NNNNTAATACGACTCACTATAGG GCTCAGGATGCCGAGTAGCTG	Specific for *CG10479* cDNA for ISH probe with T7 sequence.
wor_fwd_ish	CGTGTGGCGAAAGCAAATGA	Specific for *wor* cDNA for ISH probe.
wor_rev_t7_ish	NNNNTAATACGACTCACTATAGG GATCCTCCAGGCTGGCTTCTA	Specific for *wor* cDNA for ISH probe with T7 sequence.
sim_fwd_ish	TAGTCACTCATTCGCTCGCC	Specific for *sim* cDNA for ISH probe.
sim_rev_t7_ish	NNNNTAATACGACTCACTATAGG GCCGGAGCAGTGTATCACCTTA	Specific for *sim* cDNA for ISH probe with T7 sequence.
hb_fwd_ish	GCCACGAAACGCCGTCTA	Specific for *hb* cDNA for ISH probe.
hb_rev_t7_ish	NNNNTAATACGACTCACTATAGG GTCGATTCGAATTCGCTTTCAAC	Specific for *hb* cDNA for ISH probe with T7 sequence.
kr_fwd_ish	TGCTTCAAGACGCACAAACG	Specific for *Kr* cDNA for ISH probe.
kr_rev_t7_ish	NNNNTAATACGACTCACTATAGG GAATGCATGTTTAGAGCGCCG	Specific for *Kr* cDNA for ISH probe with T7 sequence.
CG3036_fwd_ish	ATTGTCGAGTGCAAGAGCCG	Specific for *CG3036* cDNA for ISH probe.
CG3036_rev_t7_ish	NNNNTAATACGACTCACTATAGG GAGTGCATCGGTGTACAGTGG	Specific for *CG3036* cDNA for ISH probe with T7 sequence.
CG8407_fwd_ish	GCCATCGTTAGTTCGTGGCT	Specific for *CG8407* cDNA for ISH probe.
CG8407_rev_t7_ish	NNNNTAATACGACTCACTATAGG GAAGTCACATGGTTGCCAGGG	Specific for *CG8407* cDNA for ISH probe with T7 sequence.
vnd_fwd_ish	AATGTGTAGAGTGCGCTGCT	Specific for *vnd* cDNA for ISH probe.
vnd_rev_t7_ish	NNNNTAATACGACTCACTATAGG GTTGGTGAACAGGACTCGTCG	Specific for *vnd* cDNA for ISH probe with T7 sequence.
Con_fwd_ish	CGATAATCGCATTGAGCGCAT	Specific for *Con* cDNA for ISH probe.
Con_rev_t7_ish	NNNNTAATACGACTCACTATAGG GTGCTTGCTCCTTATGCTCGG	Specific for *Con* cDNA for ISH probe with T7 sequence.
CG8353_fwd_ish	CTGTAAACCCACCCCTCGTT	Specific for *CG8353* cDNA for ISH probe.
CG8353_rev_t7_ish	NNNNTAATACGACTCACTATAGG GGTGAAAACTGCTCGGCAACA	Specific for *CG8353* cDNA for ISH probe with T7 sequence.
hkb_fwd_ish	CCAGGACAGTGAAGCAGGAG	Specific for *hkb* cDNA for ISH probe.

Name	Sequence	Description
hkb_rev_t7_ish	NNNNTAATACGACTCACTATAGG GGATTGGGTTTGGTGAGTGCG	Specific for *hkb* cDNA for ISH probe with T7 sequence.
gsb_fwd_ish	GCGGATATCCCTTTCAAGGACAA	Specific for *gsb* cDNA for ISH probe.
gsb_rev_t7_ish	NNNNTAATACGACTCACTATAGG GGGAACAAGCAGGGATCGTCT	Specific for *gsb* cDNA for ISH probe with T7 sequence.
sprt_fwd_ish	CCGTGTAATCCGAACCCGAT	Specific for *sprt* cDNA for ISH probe.
sprt_rev_t7_ish	NNNNTAATACGACTCACTATAGG GGCGACGATCCAGTCCATTGT	Specific for *sprt* cDNA for ISH probe with T7 sequence.
CG6218_fwd_ish	ATCAAAGAGTCGCTGAGCCC	Specific for *CG6218* cDNA for ISH probe.
CG6218_rev_t7_ish	NNNNTAATACGACTCACTATAGG GCAGCGGATTTCGTCAAAGGA	Specific for *CG6218* cDNA for ISH probe with T7 sequence.
Ptx1_fwd_ish	CGCAAAGATGGCAATCGCAG	Specific for *Ptx1* cDNA for ISH probe.
Ptx1_rev_t7_ish	NNNNTAATACGACTCACTATAGG GTTAATTGCTGATTTGGGGGCG	Specific for *Ptx1* cDNA for ISH probe with T7 sequence.
CG2865_fwd_ish	GAATCTGCCCAGAACCAGCG	Specific for *CG2865* cDNA for ISH probe.
CG2865_rev_t7_ish	NNNNTAATACGACTCACTATAGG GCCAATCCGCCCGATCATCTT	Specific for *CG2865* cDNA for ISH probe with T7 sequence.

Supplementary Table S 3 Cluster markers for timepoint 1

Top 25 differentially expressed genes between the clusters of timepoint 1 (see Figure 7-4).

Cluster	Markers	Cluster	Markers	Cluster	Markers
NBs (High dpn; red cluster)	Obp99a	NBs (High dpn, green)	CG10035	Glia Precursors	gcm
	Dl		eEF1beta		repo
	MRE16		eEF5		NimC4
	Calr		eEF1alpha1		Argk
	RNASEK		ATPsynCF6		CR30009
	Ama		roh		Glycogenin
	COX7A		ND-MLRQ		CG6218
	sna		ATPsynO		shep
	CG42394		ATPsyndelta		cib
	sala		ATPsynE		Dr
	CG15646		UQCR-11		CG3036
	UQCR-14		eEF2		zfh1
	eEF1alpha1		ATPsynB		CG42342
	Nph		CG13427		Act5C
	eIF4A		ATPsynF		Mes2
	COX6B		COX4		Cyp1
	ImpL2		eEF1gamma		tap
	ATPsynbeta		ND-15		drl
	Thor		Cys		Trx-2
	rdx		COX5B		Tapdelta
	Sirup		ATPsynG		CG11267
	l(1)sc		COX5A		crok
	eEF5		RNASEK		dap
	CG11825		COX6B		eEF1delta
			eIF4A		Ama
neuroblasts (High dpn; olive green)	RNASEK	NBs (High dpn; lilac)	sala	Midline/Brain	CG9650
	ATPsynF		mt:ATPase6		CR45388
	ATPsynB		eEF1beta		cas
	ATPsynE		eEF2		CR30009
	roh		Cys		CG13920
	ND-SGDH		CG13427		CG18619
	eEF5		CG34224		vvl
	COX6B		COX5B		ATPsyn-beta
	ND-15		ND-MLRQ		HmgD
	COX4		ATPsynE		eIF-4a

Cluster	Markers	Cluster	Markers	Cluster	Markers
	UQCR-Q		eEF5		HmgZ
	ND-MLRQ		eEF1gamma		CG32230
	COX5B		MRE16		SmG
	Phf5a		ATPsynCF6		His2Av
	ATPsynbeta		COX8		CG3560
	SNRPG		ATPsynG		geminin
	ATPsynG		CG10035		CoVa
	UQCR-14		eIF4A		CG7911
	UQCR-11		UQCR-14		l(2)06225
	Nph		roh		CR34335
	COX5A		eEF1alpha1		eIF-5A
	eEF2		COX5A		CG9548
	eIF4A		Calr		EF2
	eEF1alpha1		COX4		CoVb
	eEF1beta		COX6B		Ef1alpha48D
NBs (High dpn; dark green)	Ef1alpha48D	NPs	nerfin-1		
	eIF-5A		lin-28		
	Ef1beta		insb		
	Ucrh		elav		
	CG32230		Oli		
	EF2		Imp		
	eIF-4a		tap		
	l(2)06225		lola		
	CG30415		ftz		
	CoVb		dap		
	CG40127		CG8353		
	CG3321		chrb		
	CoVIb		VepD		
	Crc		l(3)neo38		
	CoVa		MRE16		
	CoVIII		Calr		
	CG7911		eIF3k		
	CG3560		kek5		
	Ef1gamma		eIF1		
	ATPsyn-beta		REPTOR-BP		
	CoIV		Nrg		
	CG7580				
	CoVIIc				
	SmG				
	CR34335				

Supplementary Table S 4 Markers for the anteroposterior populations

Top 25 differentially expressed genes between the anteroposterior domains identified (see Figure 7-8).

Cluster	Markers	Cluster	Markers	Cluster	Markers
wg	wg	CG42342	CG42342	en	en
	gsb		gcm		inv
	Wnt4		drl		ind
	slp2		cv-c		Sp1
	slp1		NimC4		Tollo
	gsb-n		CG3036		D
	CG12496		shep		Lac
	18w		cib		Ten-a
	PsGEF		Act5C		CG5059
	Dtg		Argk		rdx
	E2f1		trn		nerfin-1
	ImpL3		Fas3		

	sca		repo	
	CycE		CG6218	
	Rcc1		tap	
	insv		Dr	
	CycB3		luna	
	klu		CrebA	
	CG1943		CR30009	
	CG34224		CG14903	
	dpn		zfh1	
	wech		knrl	
	CG10035		CG11267	
	Dl		side	
	Ubx		emc	

Supplementary Table S 5 Genes enriched at different points in pseudotime

Lists of genes found enriched in the beginning, middle and end of pseudotime (q.value $<1x10^{-4}$). Available upon request.