Inaugural dissertation

for

obtaining the doctoral degree

of the

Combined Faculty of Mathematics, Engineering and Natural Sciences

of the

Ruprecht – Karls – University

Heidelberg

Presented by

M. Sc. Lukas Peter Maria Kremer

born in: Aachen, Germany

Oral examination: 04. 12. 2023

Investigating the Molecular Basis of Adult Neurogenesis:

Single-Cell Multi-Omics Unveils DNA Methylation as a Key Regulator of Astrocyte Stemness

Referees: Prof. Dr. Ana Martin-Villalba

Jun.-Prof. Dr. Simon Anders

Summary

The ventricular-subventricular zone (vSVZ) of adult mammalian brains harbors specialized astrocytes, called adult neural stem cells (NSCs), which are capable of generating both neurons and glial cells. In contrast, common parenchymal astrocytes perform a wide range of structural, metabolic, homeostatic and neurosupportive functions. Despite these distinct functions, studies employing immunostaining and single-cell RNA-sequencing have demonstrated that NSCs and astrocytes largely express the same set of genes, which raises the question how stem cell function is molecularly encoded. To address this question, I analyzed a single-cell triple-omic data set that contains information on gene expression, chromatin accessibility and DNA methylation for hundreds of cells of the adult NSC lineage, as well as common parenchymal astrocytes.

To enable this analysis, I developed "`scbs`", a Python software for the analysis of single-cell methylation data. I devised and implemented two major improvements over the current state of the art analysis workflow: First, instead of segmenting the genome into fixed intervals, I devised an approach to scan the entire genome for informative regions called variably methylated regions (VMRs). Second, instead of simply averaging methylation values within these tiles, I devised a more robust measure of DNA methylation.

By making use of these new methods, I demonstrated that adult NSCs possess a unique DNA methylation profile that is not found in common parenchymal astrocytes. This NSC methylome is characterized by hypomethylation of genes required for neurogenesis. I propose that this feature contributes to the neurogenic capabilities of NSCs by enabling the transcriptional activation of these genes. In contrast, common parenchymal astrocytes are locked in their current astrocyte fate by DNA methylation. To test this hypothesis, I analyzed single-cell multi-omic data from mice that were subjected to ischemia, as ischemia is known to induce a neurogenic response in common parenchymal astrocytes and NSCs. My analysis suggests that this gain of neurogenic capabilities is accompanied by gain of an NSC methylome, which supports the idea that a specific DNA methylome is required for stem cell function. Overall, my results demonstrate that DNA methylation is dynamic even in adult tissues and not just in embryonic development, and furthermore unveils DNA methylation as a crucial factor that constrains or enables alternative cellular fates.

Zusammenfassung

Die ventrikuläre-subventrikuläre Zone (vSVZ) des Gehirns erwachsener Säugetiere beherbergt spezialisierte Astrozyten, sogenannte adulte neurale Stammzellen (NSCs), die sowohl Neuronen als auch Gliazellen hervorbringen können. Im Gegensatz dazu erfüllen gewöhnliche parenchymale Astrozyten ein breites Spektrum struktureller, metabolischer, homöostatischer und unterstützender Funktionen. Trotz dieser unterschiedlichen Funktionen haben Studien mit Immunfärbungen und single-cell RNA-sequencing gezeigt, dass NSCs und Astrozyten größtenteils denselben Gensatz exprimieren, was die Frage aufwirft, wie Stammzellfunktion auf molekularer Ebene kodiert wird. Um diese Frage zu beantworten, habe ich einen single-cell multi-omics-Datensatz analysiert, der Informationen zur Genexpression, Chromatinzugänglichkeit und DNA-Methylierung hunderter NSCs, NSC-Tochterzellen, sowie gewöhnlicher parenchymaler Astrozyten enthält.

Um diese Analyse zu ermöglichen, habe ich "scbs" entwickelt, eine Python-Software zur Analyse von Methylierungsdaten einzelner Zellen. Ich habe zwei wesentliche Verbesserungen gegenüber dem üblicherweise verwendeten Analyse-Workflow entwickelt und implementiert: Anstatt das Genom in feste Intervalle zu unterteilen, habe ich zunächst einen Ansatz entwickelt, um das gesamte Genom nach informativen Regionen zu durchsuchen, die ich als variabel methylierte Regionen (VMRs) bezeichne. Zweitens habe ich, anstatt einfach die Methylierungswerte innerhalb dieser Regionen zu mitteln, ein robusteres Maß für die DNA-Methylierung entwickelt.

Mithilfe dieser neuen Methoden habe ich gezeigt, dass adulte NSCs ein einzigartiges DNA-Methylierungsprofil besitzen, das in gewöhnlichen parenchymalen Astrozyten nicht zu finden ist. Dieses NSC-Methylom ist durch eine Hypomethylierung der für die Neurogenese erforderlichen Gene gekennzeichnet. Ich stelle die These auf, dass diese Eigenschaft zu den neurogenen Fähigkeiten von NSCs beiträgt, indem sie die transkriptionelle Aktivierung dieser Gene ermöglicht. Im Gegensatz dazu sind gewöhnliche parenchymale Astrozyten durch DNA-Methylierung an ihre aktuelle Zellidentität gebunden. Um diese Hypothese zu testen, analysierte ich single-cell multi-omics-Daten von Mäusen, die einer Ischämie ausgesetzt waren, da Ischämie eine neurogene Reaktion in gewöhnlichen parenchymalen Astrozyten und NSCs auslöst. Meine Analyse legt nahe, dass Astrozyten welche Stammzellfunktionen erwerben ihre DNA-Methylierung verändern, was meine Annahme unterstützt, dass ein spezifisches DNA-Methylom für die Stammzellfunktion erforderlich ist. Insgesamt zeigen meine Ergebnisse, dass die DNA-Methylierung auch in Geweben erwachsener Säugetiere und nicht nur in der Embryonalentwicklung dynamisch ist, und enthüllen darüber hinaus die DNA-Methylierung als einen entscheidenden Faktor, der bestimmte Zellschicksale einschränkt oder ermöglicht.

Contents

List of Figures

List of Abbreviations

5caC	5-carboxylcytosine
5fC	5-formylcytosine
5hmC	5-hydroxymethylcytosine
5mC	5-methylcytosine
A, C, G, T	adenine, cytosine, guanine, thymine
aNSC	active neural stem cell
ATAC-seq	assay for transposase-accessible chromatin using sequencing
ATP	adenosine triphosphate
BER	base excision repair
bp	base pair
BrdU	5-bromo-3′-deoxyuridine
ChIP-seq	chromatin immunoprecipitation followed by sequencing
cKO	conditional gene knockout
CNS	central nervous system
CpG	CG-dinucleotide (5′—C—phosphate—G—3′)
CRISPR	clustered regularly interspaced short palindromic repeats
CSR	compressed sparse row
DMR	differentially methylated region
DNA	deoxyribonucleic acid
DNMT	DNA methyltransferase
dpi	days post ischemia
E	embryonic day (e. g. E10)
EGF	epidermal growth factor
Eph	ephrin receptors
FACS	fluorescence-activated cell sorting
GO	gene ontology
GpC	GC-dinucleotide (5′—G—phosphate—C—3′)
H3K9me2	di-methylation at the 9^{th} lysine of histone H3
IFNAGRKO	interferon-α-receptor and interferon-γ-receptor knockout
iPSC	induced pluripotent stem cell
kb	kilobase pair
KO	gene knockout
LMR	lowly methylated region
MAB-seq	methylase-assisted bisulfite sequencing
MBD	methyl-CpG-binding domain
Mb	megabase pair
mRNA	messenger ribonucleic acid
MS	multiple sclerosis
NCoR1/2	nuclear receptor co-repressor 1/2
NEIL	Nei-like
NOMe-seq	nucleosome occupancy and methylation sequencing
NSC	adult neural stem cell

OPC	oligodendrocyte precursor/progenitor cell
PC	principal component
PCA	principal component analysis
PCR	polymerase chain reaction
poly(dT)	polydeoxythymine
qNSC	quiescent neural stem cell
qPCR	quantitative polymerase chain reaction
RAM	random-access memory
RNA	ribonucleic acid
RNA-seq	RNA sequencing
scATAC-seq	single-cell ATAC sequencing
scBS	single-cell bisulfite sequencing
scbs	my software for the analysis of single-cell bisulfite sequencing data
scNMT-seq	single-cell nucleosome, methylation and transcription sequencing
scRNA-seq	single-cell RNA sequencing
scTAM-seq	single-cell targeted analysis of the methylome
snhmC-seq	single-nucleus 5hmC sequencing
TAP	transit-amplifying progenitor
TCF/LEF	T cell factor/lymphoid enhancer factor family
TDG	thymine DNA glycosylase
TF	transcription factor
TFBS	transcription factor binding site
TSS	transcription start site
UMAP	uniform manifold approximation and projection
UMI	unique molecular identifier
VAR	variably accessible region
VMR	variably methylated region
vSVZ	ventricular-subventricular zone
WGBS	whole-genome bisulfite sequencing
YFP	yellow fluorescent protein

List of Gene Symbols

Gene Symbol	Protein Name	Alternative Name(s) (notable ones in **bold**)
Aldoc	Fructose-bisphosphate aldolase C	*Aldo3*, *Scrg2*
Apoe	Apolipoprotein E	Apo-E
Aqp4	Aquaporin-4	
Ascl1	Achaete-scute homolog 1	***Mash1***, *Ash1*, mASH-1
Atp1b1	Sodium/potassium-transporting ATPase subunit beta-1	*Atp4b*
Cas9	CRISPR-associated endonuclease Cas9	
Cd9	CD9 antigen	
Clu	Clusterin	
Cspg4	Chondroitin sulfate proteoglycan 4	**NG2**
Dcx	Neuronal migration protein doublecortin	
Dlx2	Homeobox protein DLX-2	*Tes1*
Dnmt1	DNA (cytosine-5)-methyltransferase 1	*Met-1*, MmuI, MCMT
Dnmt3a	DNA (cytosine-5)-methyltransferase 3A	
Dnmt3b	DNA (cytosine-5)-methyltransferase 3B	MmuIIIB
Dnmt3c	DNA (cytosine-5)-methyltransferase 3C	
Dnmt3l	DNA (cytosine-5)-methyltransferase 3-like	
Efnb1	Ephrin-B1	*Epl2*, *Eplg2*, *Lerk2*, *Stra1*
Efnb2	Ephrin-B2	*Elf2*, *Epl5*, *Eplg5*, *Htkl*, *Lerk5*
Egfr	Epidermal growth factor receptor	
Emx2	Homeobox protein EMX2	Empty spiracles homolog 2
Epha4	Ephrin type-A receptor 4	*Sek*, *Sek1*
Fabp7	Fatty acid-binding protein, brain	**BLBP**, B-FABP
Gfap	Glial fibrillary acidic protein	
Glul	Glutamine synthetase	*Glns*
Ifngr1	Interferon gamma receptor 1	*Ifngr*
Kcnj10	ATP-sensitive inward rectifier potassium channel 10	Kir4.1
Lcat	Phosphatidylcholine-sterol acyltransferase	
Mag	Myelin-associated glycoprotein	Siglec-4a
Mbp	Myelin basic protein	Myelin A1 protein
Mecp2	Methyl-CpG-binding protein 2	
Mki67	Proliferation marker protein Ki-67	
Mn1	Transcriptional activator MN1	
Mobp	Myelin-associated oligodendrocyte basic protein	
Mog	Myelin-oligodendrocyte glycoprotein	
Ncam1	Neural cell adhesion molecule 1,	CD56, *Ncam*
Nes	Nestin	
Neurog2	Neurogenin-2	***Ngn2***, *Ath4a*, *Atoh4*
Nfia	Nuclear factor 1 A-type	NF1-A, NF-I/A, NFI-A
Nfic	Nuclear factor 1 C-type	NF1-C, NF-I/C, NFI-C
Nfix	Nuclear factor 1 X-type	NF1-X, NF-I/X, NFI-X
Notum	Palmitoleoyl-protein carboxylesterase	

	NOTUM	
Nr2e1	Nuclear receptor subfamily 2 group E member 1	**TLX**, Tll, mTll, Protein tailless homolog
Nrxn1	Neurexin-1-beta	
Nr4a2	Nuclear receptor subfamily 4 group A member 2	***Nurr1***
Olig1	Oligodendrocyte transcription factor 1	Oligo1
Olig2	Oligodendrocyte transcription factor 2	Oligo2
Pax8	Paired box protein Pax-8	
Plp1	Myelin proteolipid protein	PLP, Lipophilin
Prom1	Prominin-1	**CD133**
Rbpj	Recombining binding protein suppressor of hairless	*Igkjrb1, Igkrsbp, Rbpsuh*, RBP-J kappa
S100b	Protein S100-B	S-100 protein beta chain
Sfrp1	Secreted frizzled-related protein 1	
Slc1a2	Excitatory amino acid transporter 2	*Eaat2, Glt1*
Slc1a3	Excitatory amino acid transporter 1	**GLAST**, *Eaat1, Gmt1*
Slc41a2	Solute carrier family 41 member 2	
Sox10	Transcription factor SOX-10	SOX-21, SOX-M
Tcf4	Transcription factor 4	*Itf2, Sef2*, ME2, SL3-3, SEF-2
Tcf12	Transcription factor 12	*Alf1, Meb*, ME1, HTF4
Tet1	Methylcytosine dioxygenase TET1	*Kiaa1676, Cxxc6*
Tet2	Methylcytosine dioxygenase TET2	*Kiaa1546*
Tnfrsf19	Tumor necrosis factor receptor superfamily member 19	***Troy***, *Taj*
Trdmt1	tRNA aspartic acid methyltransferase 1	***Dnmt2***, *Met2*
Uhrf1	E3 ubiquitin-protein ligase UHRF1	*Np95*
Vim	Vimentin	
Wnt3a	Protein Wnt-3a	
Zfp503	Zinc finger protein 503	*Nolz1, Zfp503*

Chapter 1

Introduction

The brain is a remarkable organ that hosts our consciousness, enables us to experience the world and allows us to form critical thoughts. To understand how such intricate functions can emerge from an organ, and how brain functions compromised upon injury can be restored, it is necessary to elucidate the brains complexity in all its facets. History has shown that breakthroughs in neuroscience were often enabled by the development of new laboratory methods. For instance, Santiago Ramón y Cajal's pioneering work on the structure of the nervous system was made possible by Camillo Golgi's method to stain single neurons with potassium dichromate and silver nitrate. While this work undoubtedly paved the way for modern neuroscience, Ramón y Cajal and other authorities in the field also assumed that neurogenesis, the production of new neurons, occurs only during development and not in adulthood. This view became one of the most pervasive dogmas in neuroscience that persisted for over a hundred years (reviewed in Gross, 2000). Again, novel methods such as labeling of dividing cells with ^{3}H-Thymidine and later 5-bromo-3'-deoxyuridine (BrdU) paved the way for a series of studies that ultimately led to a paradigm shift in the field, and the acceptance that neurogenesis occurs in adult brains (e.g. Altman, 1962; Kaplan and Hinds, 1977; Goldman and Nottebohm, 1983; Kuhn et al., 1996).

The examples above demonstrate that discoveries in biology are often facilitated by the development of new methods. I started my doctoral research during the rise of single-cell RNA-sequencing (scRNA-seq) technologies, which made it possible to characterize gene expression of brain cells at unprecedented resolution. Shortly after the rise of scRNA-seq, single-cell sequencing protocols were adapted to capture other molecular features, including epigenetic features such as DNA methylation. After a brief introduction, the first major section of this thesis will delineate how I analyzed such data of this new type and explain new computational methods which I developed to enable this analysis. In the second major section, I will then focus on my findings in adult neurogenesis of the ventricular-subventricular zone, and explain why I propose that DNA methylation underlies stemness of adult neural stem cells.

1.1 Adult neurogenesis

It is now established that mammalian brains generate new neurons not only in development, but also in adulthood. Here I will briefly review the topic of adult neurogenesis with a focus on rodents, particularly the house mouse *Mus musculus*, which serves as a well-studied model organism. As the foundation for adult neurogenesis is laid in embryonic development, I will first delineate neurogenesis of the developing cerebral cortex.

1.1.1 Developmental origins of adult neural stem cells

As for all cells of our bodies, the origin of adult neural stem cells, the cells responsible for neurogenesis in adult brains, can be traced back to one of the three germ layers that emerge during gastrulation of the early embryo. One of these germ layers, the ectoderm, harbors multipotent stem cells called neuroepithelial cells (also known as neuroectodermal cells). Neuroepithelial cells have the ability to self-renew, i. e. to produce more stem cells, via symmetric cell division, i. e. a cell division that yields two identical daughter cells (Huttner and Brand, 1997; Shenghui et al., 2009). This leads to formation of the neural tube, the embryonic precursor of the central nervous system (CNS). Between mouse gestational day 9 and 10 (E9 – 10), neuroepithelial cells convert into **radial glial cells** – the main source of neural progenitor cells during CNS development that give rise to a variety of neurons, as well as glial cells including astrocytes, oligodendrocytes and ependymal cells (Fig. 1.1, Malatesta et al., 2003; Spassky et al., 2005).

During the conversion to radial glial cells, neuroepithelial cells not only lose some characteristics of epithelial cells, such as the expression of tight junction components (Aaku-Saraste et al., 1996) but also acquire features of astroglial cells (reviewed in Kriegstein and Alvarez-Buylla, 2009; Malatesta et al., 2008). These astroglial features include expression of the glutamate transporter GLAST (encoded by the gene *Slc1a3*), brain lipid-binding protein BLBP (*Fabp7*), expression of several intermediate filaments such as nestin (*Nes*) and vimentin (*Vim*), but also the establishment of glycogen storage granules. As depicted in Fig. 1.1, radial glial cells are elongated cells with apical-basal polarity. Their somata reside in the embryonic ventricular zone, which allows for contact with the developing ventricular system on their apical end. In contrast, the basal end of radial glial cells contains an elongated primary cilium that extends to the pial surface, i. e. the outer layer of the developing cerebral cortex (Kriegstein and Alvarez-Buylla, 2009; Malatesta et al., 2008).

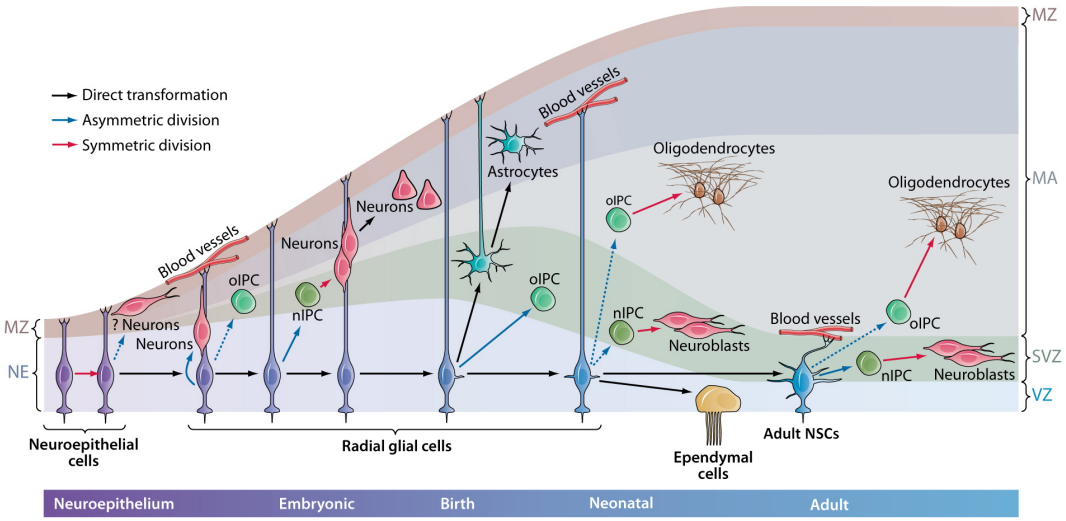

FIGURE 1.1: The origin of adult neural stem cells in embryonic development. During early development, neuroepithelial cells self-renew and ultimately give rise to radial glial cells. These cells have a bipolar morphology (apical: down, basal: up) and contact both the ventricle and blood vessels. Radial glia first give rise to a large number of neurons, and later give rise to glial cells including oligodendrocytes. Both neurons and oligodendrocytes are mostly generated indirectly through actively dividing intermediate progenitor cells (neurogenic nIPCs and oligodendrocytic oIPCs). As the brain thickens, radial glial cells elongate and serve as a scaffold for migrating neurons. Toward the end of neurodevelopment, most radial glia detach from the apical side and convert into astrocytes. A small number of radial glia retains apical contact and stem cell function, however. These cells are known as adult neural stem cells (NSCs) and are capable of generating both neurons (via nIPCs known as transit-amplifying progenitors) and glial cells throughout adulthood. Solid arrows are supported by experimental evidence; dashed arrows were hypothesized by Kriegstein and Alvarez-Buylla (2009). MA, mantle; MZ, marginal zone; NE, neuroepithelium; SVZ, subventricular zone; VZ, ventricular zone. Figure and caption adapted from Kriegstein and Alvarez-Buylla (2009).

As embryonic development progresses, radial glial cells populate the brain with neurons. In the developing cortex, it is believed that neurogenesis occurs by means of asymmetric divisions that give rise to one radial glial cell and one daughter cell. This daughter cell can either be a neuron or, more commonly, the daughter cell represents an intermediate progenitor state that further increases neuronal output by undergoing one or more rounds of division. These transit-amplifying cells, termed intermediate progenitor cells or basal progenitor cells, are often restricted to the generation of a specific type of progeny (e. g. oligodendrocytes or neurons; Fig. 1.1). This fate-restriction is determined by spatial and temporal cues. For instance, different sub-regions of the developing ventricular zone give rise to different neuronal subtypes that express distinct sets of transcription factors (Guillemot, 2005; Greig et al., 2013; Kriegstein and Alvarez-Buylla, 2009). Another factor that affects progenitor fate-restriction is time: While the early phases of brain development are dedicated to the production of neurons, the focus shifts to the generation of glial cells around embryonic day 18 (E18), in the so-called neurogenic-to-gliogenic switch (Miller and Gauthier, 2007; Malatesta et al., 2008).

Toward the end of embryonic development, most radial glial cells transform into **astrocytes** (Noctor et al., 2008). During this transformation, radial glial cells lose

their characteristic bipolar morphology, as well as their contact with the ventricle and pial surface. These cells then migrate toward the cortical plate, assume the characteristic "star-shaped" astrocyte morphology, and perform various neuroprotective, metabolic and homeostatic tasks in the brain parenchyma (see section 1.2).

Importantly though, some radial glial cells persist throughout adulthood. These cells retain their bipolar morphology, their contacts with the vasculature and the ventricle, and most importantly their ability to act as neural stem cells (Merkle et al., 2004). These adult neural stem cells (NSCs) persist in specialized niches of the adult brain (Merkle et al., 2004; Fuentealba et al., 2015; Furutachi et al., 2015). The two major stem cell niches of adult vertebrate brains are the subgranular zone that is part of the dentate gyrus of the hippocampus (Abbott and Nigussie, 2020; Abdissa et al., 2020) and the ventricular-subventricular zone on the walls of the lateral ventricles (Doetsch et al., 1999). More recently, it has also been suggested that adult neurogenesis might occur in other brain areas including the hypothalamus, cortex and amygdala (reviewed in Jurkowski et al., 2020).

1.1.2 The ventricular-subventricular zone

The ventricular-subventricular zone (vSVZ), also known as subependymal zone (e. g. Kazanis et al., 2017) or simply adult subventricular zone, is one of the two major neural stem cell niches of mammals (Fig. 1.2). Mammalian brains feature interconnected cavities filled with cerebrospinal fluid. The vSVZ corresponds to the inner lining of the two largest of these cavities, the lateral ventricles, and harbors several thousand NSCs (also known as type B(1) cells or SVZ astrocytes, see section 1.1.4). In adult mice, vSVZ NSCs generate large amounts of neuroblasts (young neurons) which migrate along the so-called rostral migratory stream toward the olfactory bulb where they integrate into the existing neural circuitry. Albeit to a lesser extent, NSCs also generate glial cells. This includes the generation of astrocytes (Sohn et al., 2015, see section 1.2) and oligodendrocytes (Menn et al., 2006; Gonzalez-Perez and Alvarez-Buylla, 2011; Ortega et al., 2013), a type of glial cell that insulates neuronal axons in a process called myelination (see section 1.3).

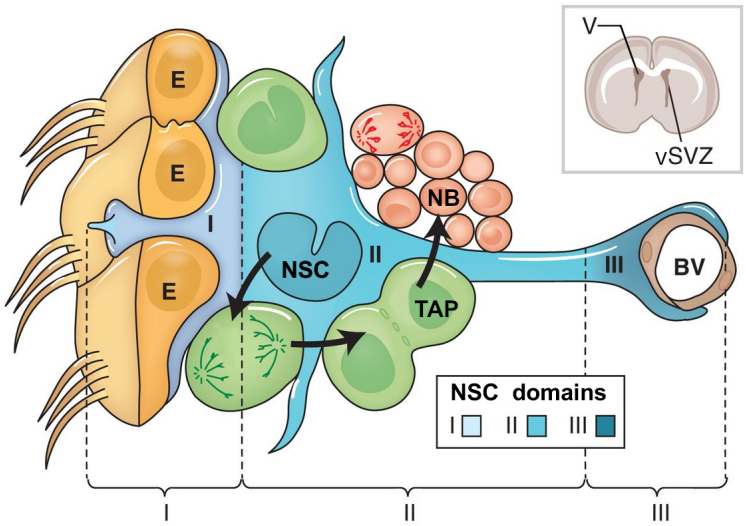

FIGURE 1.2: Adult NSCs and their progeny in the vSVZ.
Top right: Coronal section of the adult mouse brain. Specialized astrocytes called adult NSCs reside at the inner lining of both lateral ventricles (V), in the ventricular-subventricular zone (vSVZ). Adult NSCs (a single NSC, divided into three domains, is depicted in blue) are capable of self-renewal and also give rise to the rapidly dividing transit-amplifying progenitors (TAPs, green). TAPs give rise to neuroblasts (red), which migrate along the rostral migratory stream toward the olfactory bulb. The apical NSC domain (I) contains an apical process that harbors a primary cilium and makes contact with the ventricle. The apical process is surrounded by ependymal cells (E) in a characteristic pinwheel structure. The intermediate NSC domain (II) contains the cell body which makes contact with neuroblasts, TAPs and other NSCs. The basal domain (III) comprises a process that contacts blood vessels (BV, brown). Figure and caption adapted from Lim and Alvarez-Buylla (2016).

The vSVZ can be divided into four microdomains: the lateral (also called septal), dorsal, medial and ventral vSVZ. Lineage-tracing experiments have shown that NSCs located in different subdomains express different sets of transcription factors and are biased toward the production of different neuronal sub-types (López-Juárez et al., 2013; Merkle et al., 2014; Chaker et al., 2016). Similarly, NSCs located in the dorsal and dorso-lateral vSVZ are more likely to generate oligodendrocytes compared to NSCs from other vSVZ microdomains (Ortega et al., 2013; Azim et al., 2016; Mizrak et al., 2019).

The inner lining of the lateral ventricles is formed by a layer of ciliated glia cells called ependymal cells (Jiménez et al., 2014). These cells form a tight barrier that displaces the soma of NSCs from direct contact with the lateral ventricle. Nonetheless, NSCs maintain contact to the ventricle and cerebrospinal fluid by extending a cellular process on their apical end. Within the vSVZ, NSCs and ependymal cells are arranged in a manner that resembles a pinwheel, where the apical processes of several NSCs are surrounded by a rosette of ependymal cells (Mirzadeh et al., 2008).

1.1.3 The neurogenic adult neural stem cell lineage

NSCs of the murine vSVZ retain labels incorporated during cell division for multiple weeks and are able to replenish the stem cell pool after ablation of cycling cells

(Morshead et al., 1994; Doetsch et al., 1999; Kalamakis et al., 2019). These findings suggest that the majority of NSCs is in a quiescent state. **Quiescent NSCs (qNSCs)** are out of the cell cycle and maintain a low metabolic rate (Wang et al., 2011). In addition to qNSCs, a small subset of NSCs is in an active state characterized by the expression of genes involved in the cell cycle and protein biosynthesis, and by the expression of epidermal growth factor receptor (EGFR) (Codega et al., 2014; Llorens-Bobadilla et al., 2015; Zywitza et al., 2018). Both qNSCs and **active NSCs (aNSCs)** co-exist in the vSVZ and even in the same microenvironment (i. e., pinwheel structure, Codega et al., 2014). The existence of two separate active and quiescent somatic stem cell populations was also described in other somatic stem cell niches including hair follicles, crypts of the intestinal epithelium, and bone marrow (Li and Clevers, 2010). This partition is thought to contribute to regenerative capabilities by providing a dedicated population of cells that can quickly respond when necessary, for instance upon injury. In contrast, a dedicated quiescent population is less likely to accumulate potentially tumorigenic mutations during DNA replication and serves as a backup pool to replenish active stem cells (Li and Clevers, 2010).

Once activated, NSCs give rise to the rapidly dividing **transit-amplifying progenitors (TAPs)**, also known as transit-amplifying cells, transit-amplifying precursors or type C cells. Much like intermediate progenitor cells in development (section 1.1.1), these cells serve to amplify the number of progeny and are thought to be fate-restricted. In the murine vSVZ, TAPs undergo approximately three rounds of symmetric divisions and then give rise to neuroblasts (Ponti et al., 2013). **Neuroblasts**, also called type A cells, are immature neurons that migrate a considerable distance toward the olfactory bulb. In order to migrate, neuroblasts essentially migrate along each other and form a network of chains that ultimately converges in one large stream of migrating neuroblasts called the rostral migratory stream (Altman, 1969; Lois and Alvarez-Buylla, 1994). Neuroblasts are actively dividing cells. During their chain migration to the olfactory bulb, neuroblasts undergo approximately one to two cell divisions (Ponti et al., 2013). Upon arrival at the olfactory bulb, neuroblasts disperse radially, differentiate into functional interneurons and integrate into the neural circuitry of the olfactory bulb (Lim and Alvarez-Buylla, 2016).

1.1.4 Adult neural stem cells are specialized astrocytes

As adult NSCs are direct descendants of radial glial cells (see section 1.1.1), they retain many characteristics of these cells. Among these are features of astroglial cells, including ultrastructural qualities observed in electron microscopy images, as well as protein expression of astrocyte marker genes such as GFAP and GLAST visualized by immunostaining (Doetsch et al., 1997). For these reasons, NSCs are considered specialized astrocytes (Doetsch et al., 1999; Kriegstein and Alvarez-Buylla, 2009; Schneider et al., 2019; Cebrian-Silla et al., 2021). Similar to other types of astrocytes (Díaz-Castro et al., 2023), NSCs also contact the vasculature by extending cellular

processes with specialized endfeet to nearby blood vessels (Mirzadeh et al., 2008). However, in contrast to non-neurogenic astrocytes which are typically multipolar, NSCs retain the unique apical-basal polarity of radial glial cells. Unlike the multiciliated ependymal cells, these vSVZ astrocytes are monociliated and express GLAST as well as Prominin-1 (*Prom1*, CD133, Mirzadeh et al., 2008), which is also found in the cilia of radial glial cells (Mirzadeh et al., 2008; Beckervordersandforth et al., 2010; Llorens-Bobadilla et al., 2015). While the basal process of NSCs contacts nearby blood vessels, the apical process of NSCs is surrounded by ependymal cells in the aforementioned pinwheel structure and extends toward the lateral ventricle. This apical process harbors the primary cilium of NSCs, contacts the cerebrospinal fluid and represents a unique feature of NSCs that is not found in common parenchymal astrocytes (Mirzadeh et al., 2008). Since the cerebrospinal fluid contains many crucial signaling molecules known to regulate adult neurogenesis, such as Wnt ligands and bone morphogenic proteins, it is believed that this connection to the ventricle provides regulatory cues that are then integrated and acted upon by NSCs, in addition to signals provided by adjacent cells and the vasculature (Lehtinen et al., 2011).

Heterogeneity of vSVZ astrocytes

Reports of astrocyte heterogeneity in the vSVZ date back to 1997, when Doetsch et al. first described the astrocyte identity of NSCs. In this publication, Doetsch et al. identified two types of vSVZ astrocytes with distinct morphologies, which they labeled **B1** and **B2**. While it is now commonly accepted that type B1 cells are *bona fide* adult NSCs, little is known about the function of type B2 cells, which are rarely discussed in the literature and, if so, sometimes with conflicting information. Most review articles on adult neurogenesis omit B2 cells entirely and focus solely on (B1) NSCs (e. g. Lim and Alvarez-Buylla, 2016; Bond et al., 2015). Articles that mention B2 cells typically describe them as astrocytes located below the ependyma and closer to the adjacent parenchyma of the striatum, which contact blood vessels but not the lateral ventricle (Ihrie and Álvarez-Buylla, 2011; Pastrana et al., 2011). These cells possess a multipolar morphology with bushy, branched processes similar to astrocytes found in other regions of the brain (Mirzadeh et al., 2008). Obernier et al. (2018) demonstrated that these B2 cells are generated by NSCs, and furthermore speculated that they might act as a reserve population of NSCs as they were previously shown to proliferate (Doetsch et al., 1997; Capilla-Gonzalez et al., 2014).

In addition to NSCs and B2 astrocytes, the vSVZ also contains other astrocytes that are often referred to with catch-all terms such as "niche astrocytes" or "parenchymal astrocytes" (Pastrana et al., 2011) (although the term niche astrocytes has also been used to refer to B2 astrocytes specifically in Platel and Bordey, 2016). Parenchymal vSVZ astrocytes do not possess the bipolar NSC morphology and are generally assumed to be non-neurogenic astrocytes with supporting homeostatic function

(Garcia et al., 2004; Pastrana et al., 2011). Although vague, I will employ the term "parenchymal astrocytes" in this thesis to refer to all astrocytes that are primarily concerned with tissue-homeostatic function and not NSC function, to maintain congruence with previous literature. To summarize, at least three sub-types of astrocytes were reported in the vSVZ: NSCs (quiescent and active), type B2 astrocytes, and parenchymal astrocytes. Whether the latter two sub-types contribute to neurogenesis is currently poorly understood.

The search for NSC markers

The astrocyte cell identity of NSCs, combined with the occurence of other types of astrocytes in the vSVZ, poses a significant challenge to researchers trying to study NSCs in isolation (Ihrie and Álvarez-Buylla, 2011; Pastrana et al., 2011). While astrocyte markers such as GFAP or GLAST allow for isolation of NSCs by means of e. g. fluorescence-activated cell sorting (FACS), this strategy inevitably leads to the isolation of both NSCs and common parenchymal astrocytes. For this reason, considerable efforts were devoted to the identification of NSC-specific marker genes which would allow for isolation of NSCs while precluding contamination of the sample with other astrocytes. Several such markers have been suggested. These include, for instance, Prominin-1 (Mirzadeh et al., 2008), Tlx (Tailless, *Nr2e1*, Shi et al., 2004; Liu et al., 2008), Nestin (*Nes*, Yamaguchi et al., 2000), LeX (LewisX, CD15, SSEA-1, Capela and Temple, 2002) and Troy (*Tnfrsf19*, Basak et al., 2018). While these markers serve as useful tools for the isolation and visualization of NSCs, there are also caveats to their use. For instance, Prominin-1 is dynamically regulated and thus not constantly expressed in all NSCs, and this marker is also expressed by other vSVZ cells including oligodendrocytes and ependymal cells (Beckervordersandforth et al., 2010; Mirzadeh et al., 2008). Similar caveats apply to other markers, and thus it is generally advisable to combine multiple markers (Beckervordersandforth et al., 2010; Codega et al., 2014; Llorens-Bobadilla et al., 2015).

1.1.5 Adult neurogenesis through the lens of scRNA-seq

The rise of scRNA-seq (see section 1.5) has enabled the simultaneous quantification of gene expression profiles of thousands of cells. One of the major advantages of scRNA-seq is that even heterogeneous samples consisting of multiple cell types can be processed and interpreted, as the single-cell resolution allows users to break down the sample into cell types *in silico*. For this reason, scRNA-seq raised hopes as a means to discover rare cell types or sub-populations, and to identify new marker genes. Thus, the adult neurogenesis community was quick to adapt this new technology, resulting in several publications that utilized scRNA-seq to dissect cellular heterogeneity of the vSVZ (e. g., Llorens-Bobadilla et al., 2015; Zywitza et al., 2018; Kalamakis et al., 2019; Mizrak et al., 2019; Kremer et al., 2021; Cebrian-Silla et al., 2021; Carvajal Ibañez et al., 2023).

Single-cell sequencing offered a new perspective on adult neurogenesis and enabled the field to consider the entire transcriptome instead of just select marker genes. However, this advancement also brought a new challenge, namely to reconcile existing cell type and cell state definitions with the new data sets obtained by scRNA-seq. As discussed later in section 1.5.4 (see also Amezquita et al., 2020; Heumos et al., 2023), a typical scRNA-seq analysis utilizes dimensionality reduction for data visualization, as well as a clustering algorithm to define groups of data points (i. e., cells). The cell groups thus obtained by clustering ("clusters") are then manually labeled by assessing gene expression of the contained cells, for instance by inspecting the expression of known marker genes. While this approach is often described as *unbiased*, since groups of cells are automatically determined by the clustering algorithm and not by the analyst, the labels attached to each cluster are still ultimately the result of expert judgement. Furthermore, there is no guarantee that the obtained clusters correspond to cell populations previously described in the literature. As a result, scRNA-seq data sets derived from vSVZ samples have been interpreted in different ways.

Overall, studies employing scRNA-seq to study adult neurogenesis in the murine vSVZ obtained largely consistent results across different laboratories. Sufficiently large data sets revealed a continuous cell trajectory, characterized by expression of a wide range of astrocyte marker genes on one end of the trajectory and expression of neuroblast marker genes on the other end (Zywitza et al., 2018; Kalamakis et al., 2019; Cebrian-Silla et al., 2021; Kremer et al., 2021). Thus, scRNA-seq confirmed the established view that NSCs are specialized astrocytes. However, dimensionality reduction plots also revealed signs of astrocyte heterogeneity in the vSVZ, as indicated by detection of three (or more, depending on clustering parameters and cell number) visually separated groups of astrocytes. Dimensionality reduction and pseudotemporal ordering of cells suggests that these three clusters might represent three consecutive cell states of the NSC lineage. It is agreed upon that the third of these clusters corresponds to aNSCs, due to high expression of aNSC markers such as *Egfr* and expression of genes required for mitosis. Accordingly, the community also agrees that the cluster "upstream" of these aNSCs consists of qNSCs.

Different labels have been attached to the first cluster at the putative start of the cell trajectory, however. Across data sets, cells in this cluster show very high expression of astrocyte markers such as *S100b* and *Aqp4* (Zywitza et al., 2018; Kalamakis et al., 2019; Cebrian-Silla et al., 2021; Carvajal Ibañez et al., 2023). For this reason, many authors assumed that this cluster corresponds to common non-neurogenic **parenchymal astrocytes** and even excluded these cells from further analysis (e. g. Cebrian-Silla et al., 2021; Zywitza et al., 2018). According to this interpretation, NSCs occur in two activation states, aNSC and qNSC. Although not explicitly stated, the main rationale for this interpretation is likely that common parenchymal astrocytes, which are generally assumed to be non-neurogenic, were previously reported to reside in the vSVZ (Garcia et al., 2004; Pastrana et al., 2011, section 1.1.4), and that

these cells should also be captured by scRNA-seq. Although it is difficult to distinguish common parenchymal astrocytes from NSCs based on gene expression due to shared expression of astrocyte marker genes, Cebrian-Silla et al. and Zywitza et al. likely reasoned that the astrocyte cluster with the largest distance to e. g. aNSCs and TAPs must correspond to common parenchymal astrocytes.

In contrast, others have interpreted the initial astrocyte cluster as a separate NSC activation state (e. g. Kalamakis et al., 2019; Carvajal Ibañez et al., 2023). In this interpretation, NSCs occur in three possible states: a deeply quiescent or **dormant state** ("qNSC1"), a primed-quiescent state ("qNSC2"), and the aNSC state. An argument in favor of this interpretation is that cells in this initial astrocyte cluster show many differences in gene expression compared to astrocytes isolated from the striatum and cortex, such as expression of the cell surface glycoprotein CD9 (*Cd9*, Llorens-Bobadilla et al., 2015; Kalamakis et al., 2019). Furthermore, Kalamakis et al. (2019) demonstrated that the two initial astrocyte/NSC clusters share a highly similar transcriptome, comparable with the similarity of qNSC2 cells and aNSCs, suggesting that they might be part of the same continuous lineage.

To summarize, while scRNA-seq of the vSVZ yields similar results across different laboratories, these data sets have been interpreted in different ways. Specifically, scRNA-seq of the vSVZ reveals two distinct, albeit similar, groups of quiescent astrocytes, and it is currently unclear whether these correspond to two sub-states of qNSCs, or to qNSCs and common non-neurogenic parenchymal astrocytes. In addition to this ambiguity, it is currently entirely unclear which single-cell transcriptomes correspond to type B2 cells (section 1.1.4), as no study to date pinpointed this cell population in an scRNA-seq data set.

1.2 NSCs and other astrocytes in health and disease

As reviewed earlier, only a small fraction of astrocytes contributes to neurogenesis in adult mammalian brains. Nonetheless, the remaining common parenchymal astrocytes are essential for brain function, which I will briefly review in this section.

1.2.1 Common parenchymal astrocytes

Astrocytes are glial cells that perform a wide range of structural, metabolic, homeostatic and neurosupportive functions in adult and developing brains (reviewed in Sofroniew and Vinters, 2010; Sloan and Barres, 2014; Linnerbauer and Rothhammer, 2020). The name *astro*cyte derives from their characteristic star-shaped morphology, which is the result of their many cellular processes. Using these processes, astrocytes make contact with synapses, dendrites, axons, blood vessels and even other astrocytes (Fig. 1.3).

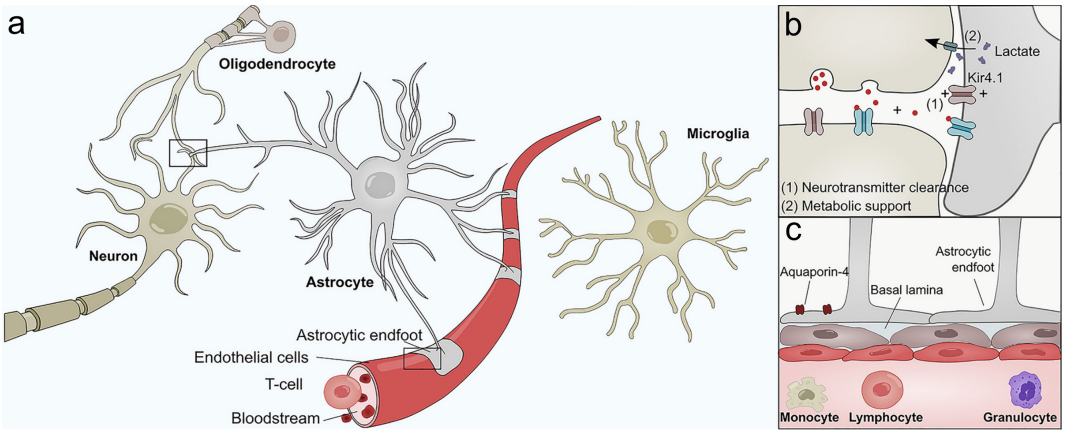

FIGURE 1.3: Function of common parenchymal astrocytes in homeostatic conditions. **(a)** Astrocytes interact with neurons, oligodendrocytes, microglia, and cells of the blood brain barrier during steady state conditions. **(b)** Astrocytes modulate synaptic transmission by providing metabolic support and clearing neurotransmitters, using transmembrane proteins such as the inward rectifying K^+ channel Kir4.1 (*Kcnj10*). **(c)** Astrocytic endfeet contact the vasculature of the brain and contribute to function of the blood brain barrier. These endfeet express high levels of the water channel Aquaporin-4 (*Aqp4*) and interact with pericytes of the basal lamina of the brain. Figure and caption adapted from Linnerbauer and Rothhammer (2020).

Astrocytes tile the entire CNS in a continuous, structured network in which they connect to each other via gap junctions. These gap junctions allow for exchange of small molecules and ions such as Ca^{2+} and K^+ (Bennett et al., 2003). Although astrocytes do not propagate action potentials like neurons, this interconnected network of astrocytes allows for the transmission of signals, for instance in the form of Ca^{2+} waves (Bazargani and Attwell, 2016; Rasmussen et al., 2023).

The main function of astrocytes, however, is to support neurons in their function. Astrocytes achieve this by utilizing their many connections with other cells to distribute energy metabolites, water, sterols, ions and neurotransmitters in the brain. To achieve this feat, astrocytes express an array of **transmembrane proteins**, which allow for transport and diffusion of ions and molecules. Many of these proteins, such as the water channel Aquaporin-4 (*Aqp4*) and the amino acid transporter GLAST (*Slc1a3*) are commonly used as astrocyte marker genes. Via their connections with the vasculature, astrocytes take up molecules such as water and glucose, which they then distribute among adjacent neurons. Furthermore, astrocytes act as a local fuel reserve by storing glucose in the form of glycogen granules. Once required, for instance during low levels of blood sugar, the stored glycogen is broken down to lactate, which astrocytes then transfer to adjacent neurons as a source of energy (Sofroniew and Vinters, 2010; Sloan and Barres, 2014). Like energy metabolites, astrocytes also provide neurons with cholesterol, which they synthesize and then traffic to neurons and glia using apolipoprotein E (Apo-E, *Apoe*, Wang et al., 2021).

In addition to their crucial functions in maintaining brain homeostasis and supporting neurons, astrocytes also actively modulate the transmission of nervous impulses. In the long term, astrocytes achieve this by controlling the formation

and elimination of synapses (Stevens et al., 2007; Sloan and Barres, 2014) and by releasing growth factors and related molecules (Sofroniew and Vinters, 2010). In the short term, astrocytes modulate synapse function by releasing and taking up neurotransmitters, amino acids and other synaptically active molecules (reviewed in Lalo et al., 2021). But the ability of astrocytes to take up neurotransmitters is not only important for synaptic function, it is also crucial to prevent accumulation of these molecules in the brain. In a similar manner, astrocytes also employ ion channels and transporters to regulate ion concentration and pH of the extracellular space. Furthermore, owing to their ability to form tight junctions, secrete growth factors and transport fluids, astrocytes contribute critically to the formation and maintenance of the blood brain barrier (Cabezas et al., 2014). In summary, astrocytes are crucial regulators of synapses, blood flow, fluid homeostasis, energy metabolism and cholesterol metabolism in the brain (Sofroniew and Vinters, 2010). While my summary of astrocyte function describes the main functions performed by a typical astrocyte, it is important to note that astrocyte morphology, function and gene expression is heterogeneous and differs both on a local level and across different regions of the CNS (Haim and Rowitch, 2017; Chai et al., 2017; Matias et al., 2019).

1.2.2 Reactive astrocytes

In addition to their crucial functions in maintaining brain homeostasis of the healthy brain, astrocytes are also known for their ability to sense and respond to brain pathologies such as injury, infection and ischemia. Although this response was described almost a century ago (del Río Hortega and Penfield, 1927), to date there is still substantial disagreement and uncertainty about its extent and the functional implications. As a result, various definitions, molecular hallmarks and marker genes associated with this phenomenon are described in the literature, some of which are conflicting or only apply in specific scenarios such as disruption of the blood brain barrier. To clarify these ambiguities and agree on a common definition, several leading figures in the field published a consensus statement in 2021, where they define the phenomenon as follows:

> "We suggest [the term] 'reactive astrogliosis' to define the process whereby, in response to pathology, astrocytes engage in molecularly defined programs involving changes in transcriptional regulation, as well as biochemical, morphological, metabolic, and physiological remodeling, which ultimately result in gain of new function(s) or loss or upregulation of homeostatic ones."
>
> (Escartin et al. 2021)

The functional role of such **reactive astrocytes** is complex and still an active field of research. Perhaps the most well-studied aspect of reactive astrogliosis is glial scar formation, a process in which reactive astrocytes contribute to the formation of a physical barrier upon injury to the CNS. While this glial scar seals the site of injury

and thus protects against further infection and cellular damage, gliosis was for many decades considered detrimental to CNS recovery as it is a hallmark of many CNS injuries and impedes axonal re-growth (e. g. Liuzzi and Lasek, 1987). More recently though, some studies painted a more nuanced view on reactive astrocytes, revealing detrimental but also protective functions, such as secretion of anti-inflammatory and protective compounds (Faulkner et al., 2004; Linnerbauer and Rothhammer, 2020). Nonetheless, recent experiments on the spiny mouse *Acomys cahirinus* demonstrated that this species is capable of fast and efficient CNS regeneration after spinal cord injury in the absence of glial scar formation, which suggests that glial scarring indeed impedes CNS recovery (Nogueira-Rodrigues et al., 2022).

Several new insights in reactive astrocyte biology were enabled by transcriptome profiling techniques including quantitative polymerase chain reaction (qPCR), microarrays and later RNA-sequencing (RNA-seq), which demonstrated that reactive astrocytes are characterized by increased expression of distinct gene sets that greatly depend on the pathological condition (Zamanian et al., 2012; Liddelow et al., 2017; Escartin et al., 2021). Specifically, Zamanian et al. quantified gene expression changes observed after neuroinflammation, induced by systemic lipopolysaccharide injection, and gene expression changes observed after ischemia, i. e. restricted blood supply to the brain, in this case induced by occluding the middle cerebral artery. The authors found that these two conditions induced the expression of two distinct, albeit slightly overlapping, sets of genes. In a follow-up study, Liddelow et al. (2017) labeled these two sub-states of reactive astrocytes **A1** and **A2**. Based on gene expression, the authors proposed that A2 reactive astrocytes, which were observed after ischemia, may have a neuroprotective function. A2 reactive astrocytes, in contrast, showed a decrease in many homeostatic and neuroprotective functions and instead became neurotoxic *in vitro*. This A2 reactive state was shown to be induced by microglia, which in turn were activated by neuroinflammation (Liddelow et al., 2017).

In the following years, many biomedical studies adopted the A1 / A2 nomenclature to classify reactive astrocytes observed after various CNS pathologies. However, as Escartin et al. (2021) note, this binary classification into "good" and "bad" reactive astrocytes is overly simplistic as it does not capture the wide and diverse range of reactive astrocyte states. Instead, the authors suggest to move beyond the A1 / A2 nomenclature and instead characterize astrocyte states on a case-by-case basis, using functional and molecular assays.

1.2.3 Neurogenic response to injury

Reactive astrogliosis demonstrates that astrocytes possess unique abilities to sense and respond to brain injury. In non-mammalian vertebrates such as salamanders (Joven and Simon, 2018) and zebrafish (Kyritsis et al., 2012; Diotel et al., 2020), injury-induced glial response mechanisms allow for regeneration of the brain, including

the generation of new neurons and re-growth of lost brain tissue. Since mammalian astrocytes are also able to sense and respond to injury, and since a small subset of astrocytes (i. e., NSCs) generates neurons even in adult mammalian brains, it is tempting to speculate that all astrocytes may possess a latent neurogenic potential that is merely blocked in most conditions. It is possible that this putative block of neurogenic capabilities is introduced in development, as common parenchymal astrocytes isolated from young animals before postnatal day 11 can give rise to neurospheres *in vitro* (Laywell et al., 2000).

Parenchymal astrocytes of the striatum and cortex

Several studies in the past decade aimed to unlock the neurogenic potential of common parenchymal astrocytes located outside adult germinal niches. In 2014, Magnusson et al. showed that ischemia caused by transient occlusion of the middle cerebral artery is sufficient to induce a neurogenic program in astrocytes of the striatum, a brain area located adjacent to the vSVZ that is not considered neurogenic in homeostatic conditions. Specifically, these striatal astrocytes first acquired expression of the proneural transcription factor Ascl1 (*Ascl1*) and the proliferation marker Ki67 (*Mki67*), and later downregulated the astrocyte marker S100-B (*S100b*) while acquiring expression of the neuronal migration protein doublecortin (*Dcx*) and polysialylated neural cell adhesion molecule (PSA–NCAM, *Ncam1*). By inhibiting Notch signaling in healthy mice, as well as artificially increasing Notch signaling in ischemic mice, Magnusson et al. (2014) demonstrated that decreased Notch signaling is both required and sufficient to induce this neurogenic response in striatal astrocytes. In the following year, two further studies conducted by other laboratories confirmed the observation that striatal astrocytes differentiate into neurons after ischemia (Duan et al., 2015) and after inducing a brain lesion by injecting quinolinic acid (Nato et al., 2015).

In a follow-up study, Magnusson et al. (2020) utilized scRNA-seq to demonstrate that such neurogenic striatal astrocytes, induced by depletion of Notch signaling, proceed along a differentiation lineage that resembles the vSVZ NSC lineage. A direct comparison with the vSVZ NSC lineage showed that striatal astrocytes are transcriptionally similar to qNSC1 cells and thus locate upstream of the canonical qNSC2 $\rightarrow$ neuroblast lineage. Once activated though, striatal astrocytes proceed along a differentiation trajectory that closely recapitulates cell states previously described in the vSVZ NSC lineage (section 1.1.3), including cells resembling aNSCs, TAPs and neuroblasts. During this transition, striatal astrocytes gradually lose their parenchymal astrocyte morphology but do not assume the characteristic bipolar vSVZ NSC morphology. Magnusson et al. (2020) furthermore showed that some striatal astrocytes became stalled before entering the TAP-like state, and that this could be resolved by exposure to epidermal growth factor (EGF) (Fig. 1.4).

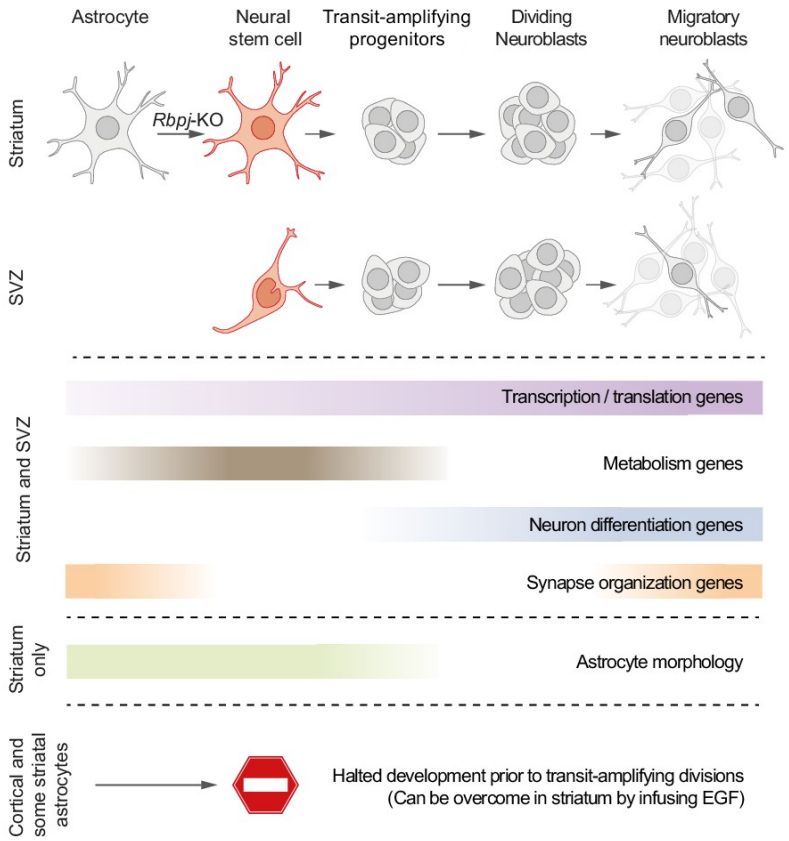

FIGURE 1.4: Activation of a neurogenic program in striatal astrocytes.
This schematic summarizes the main findings of Magnusson et al. (2020), who propose that striatal astrocytes can activate a neurogenic program upon *Rbpj* knockout (KO). Overall, this neurogenic program is highly similar to that of adult vSVZ NSCs. However, neurogenic striatal astrocytes partially retain their distinct astrocyte morphology and show differential expression of few individual genes (see Magnusson et al., 2020, for details). This neurogenic response is also seen in astrocytes of the somatosensory cortex, although these cells halt their development before entering the transit-amplifying stage. This stall is also observed in some neurogenic striatal astrocytes and can be overcome by exposure to epidermal growth factor (EGF). Figure and caption modified from Magnusson et al. (2020).

To assess whether astrocytes located in other brain areas have similar neurogenic potential, Magnusson et al. (2020) also isolated Notch-depleted astrocytes from the somatosensory cortex. While a modest number of cortical astrocytes activated the putative neurogenic program, all of these cells stalled before entering the TAP state. A previous publication from the same laboratory demonstrated that this stall can be overcome when Notch depletion is combined with stab wound injury, leading to the generation of neuroblasts and even interneurons (Zamboni et al., 2020). Similarly, Mattugini et al. (2019) combined stab wound injury with overexpression of two neurogenic factors, *Neurog2* and *Nr4a2* (also known as *Ngn2* and *Nurr1*), to induce neurogenesis in astrocytes of the cerebral cortex. Overall, these studies confirm that astrocytes in various brain areas possess a hidden neurogenic potential. In the striatum, injury alone is sufficient to generate neuroblasts, suggesting that striatal astrocytes are particularly suited for such experiments. Their ability to generate neuroblasts upon injury might also suggest that striatal astrocytes are more similar

to vSVZ NSCs than other astrocytes such as cortical astrocytes, which require more manipulation to enter the neurogenic state.

vSVZ NSCs

Even though NSCs of the vSVZ already generate neurons under homeostatic conditions, there is substantial evidence that these astrocytes, too, can be activated by injury. Such evidence is found in the scRNA-seq data of Llorens-Bobadilla et al. (2015), who isolated vSVZ cells from both healthy mice and mice that were subjected to transient bilateral common carotid artery occlusion. Compared to healthy controls, the authors observed a clear depletion of qNSC1 cells two days after ischemia and reasoned that this shift corresponds to an activation of these cells, i. e., a transition from the dormant qNSC1 state to the primed-quiescent qNSC2 state. Similarly, Delgado et al. (2021) showed that inducing demyelination by lysolecithin injection causes activation of vSVZ NSCs and generation of new glial cells. Based on gene expression, Llorens-Bobadilla et al. (2015) speculated that the injury-induced response of NSCs might be driven by **interferons**, i. e. cytokines which modulate the innate and adaptive immune response upon injury and infection (reviewed in Mazewski et al., 2020). In line with this hypothesis, knockout of interferon-γ-receptor 1 (*Ifngr1*) greatly impeded the injury-induced depletion of qNSC1 cells (Llorens-Bobadilla et al., 2015).

More recent studies further explored the regulation of vSVZ NSCs by interferons, particularly in the context of aging (Kalamakis et al., 2019; Dulken et al., 2019). Specifically, two concurrent studies found differences in the proportion of NSC cell states between young and old mice, and attributed these differences eiter to secretion of interferon-γ by T cells that infiltrate the vSVZ (Dulken et al., 2019) or to inflammation of other resident vSVZ cells (Kalamakis et al., 2019). Of note, recent findings suggest that interferons do not only control NSC function in injured or old brains, but also in healthy brains of all ages, revealing interferons as a crucial regulator of NSC function (Carvajal Ibañez et al., 2023). In summary, adult NSCs of the vSVZ sense injury and respond by activation, similar to common parenchymal astrocytes. In the case of NSCs, this response is mediated by interferon signaling.

1.3 Oligodendrocytes

The three major types of glial cells in the CNS are astrocytes (see section 1.2), oligodendrocytes (reviewed in Baumann and Pham-Dinh, 2001; Bradl and Lassmann, 2010; Kuhn et al., 2019) and microglia (reviewed in Prinz et al., 2019). Unlike microglia, which descend from early erythroid myeloid progenitors during early embryonic development (Prinz et al., 2019), both astrocytes and oligodendrocytes descent from radial glial cells (Kriegstein and Alvarez-Buylla, 2009). As adult NSCs

represent a population of radial glial cells that is set to quiescence around E14 (Furutachi et al., 2015), they also retain the ability to generate oligodendrocytes in addition to their major neurogenic function (Ortega et al., 2013; Delgado et al., 2021).

Like astrocytes, oligodendrocytes are supportive cells of the CNS that are indispensable for proper neuronal function in mammals. Oligodendrocytes accomplish this by forming and maintaining **myelin sheaths** around neuronal axons, which act as an insulation for nerve impulses. Myelin sheaths are distributed in regular intervals along the length of an axon, with regular gaps called nodes of Ranvier. This arrangement allows for fast signal transduction, as the low capacitance of ensheathed axon segments allows action potentials to quickly "jump" from one node of Ranvier to the next in a process called saltatory conduction (from Latin *saltus*, "jump") (Huxley and Stämpfli, 1949). A single oligodendrocyte is able to myelinate up to fifty axonal segments (Baumann and Pham-Dinh, 2001). In contrast to Schwann cells, which fulfill a similar role in the peripheral nervous system, oligodendrocytes are able to myelinate the axons of several nearby neurons (Salzer, 2015). Similar to astrocytes, albeit to a lesser extent, oligodendrocytes also support neurons metabolically by transferring energy in the form of lactate (Bercury and Macklin, 2015).

The myelin sheath itself consists of the oligodendrocyte's plasma membrane, which tightly wraps around the axon in multiple layers (Kuhn et al., 2019). In addition to the lipids that make up the plasma membrane, the myelin sheath also contains large quantities of proteins such as myelin basic protein (*Mbp*), which are required for formation and stabilization of the myelin sheath, and might also enhance insulation (Boggs, 2006). Several genes that encode for proteins expressed in the myelin sheath are used as oligodendrocyte marker genes. This includes, for instance, *Mbp*, the transmembrane protein myelin proteolipid protein (*Plp1*), myelin-oligodendrocyte glycoprotein (*Mog*) and myelin associated glycoprotein (*Mag*), but also transcription factors that determine oligodendrocyte cell identity such as *Sox10* and *Olig2* (Pozniak et al., 2010; Kuhn et al., 2019).

1.3.1 Loss and restoration of myelin

Myelination is a tightly regulated and selective process that furthermore requires vast amounts of lipids and proteins. It has been estimated that a mature oligodendrocyte can support a plasma membrane that weighs up to 100 times the weight of its cell body (Bradl and Lassmann, 2010). To form and maintain these myelin sheaths, oligodendrocytes have high metabolic rates and are thus subjected to toxic byproducts of metabolism such as reactive oxygen species. As a result, oligodendrocytes are vulnerable to CNS conditions such as inflammation, oxidative stress and injury, which can lead to death of oligodendrocytes and demyelination (McTigue and Tripathi, 2008). To recover from such conditions, adult mammalian brains possess the ability to restore myelin sheaths to exposed axons in a process called

remyelination. But remyelination is not only important for regeneration after damage, as low rates of oligodendrocyte death are normal even in healthy adult brains. Furthermore, adaptive myelination is important for neuronal plasticity: Learning a new motor skill such as juggling (or more realistically in the case of rodents, running in a wheel with irregularly spaced rugs) requires remyelination (McKenzie et al., 2014). Remyelination is thought to involve the generation of new oligodendrocytes by a dedicated pool of quiescent oligodendrocyte precursor/progenitor cells (OPCs) that is distributed throughout the brain (Bradl and Lassmann, 2010; McKenzie et al., 2014; Buchanan et al., 2023). However, note that radiocarbon dating of human white matter oligodendrocytes revealed only limited oligodendrocytes turnover despite high rates of myelin exchange (Yeung et al., 2014). Similarly to OPCs, oligodendrocytes generated by vSVZ NSCs might also contribute to remyelination (Menn et al., 2006; Nait-Oumesmar et al., 2007; Delgado et al., 2021).

Due to the relative vulnerability of oligodendrocytes and their essential role in brain function, demyelination is a hallmark of many CNS disorders. The most common demyelinating disease is **multiple sclerosis** (MS), a chronic inflammatory, neurodegenerative disorder that affects the CNS and is accompanied by characteristic demyelinating lesions (reviewed in Filippi et al., 2018; Rodríguez Murúa et al., 2022). Risk factors for MS include environmental factors such as Epstein Barr Virus infection and smoking, but also genetic factors. Several reports also suggest that epigenetic marks such as DNA methylation are altered in MS, although the role of these alterations is currently poorly understood (Huynh et al., 2014; Chomyk et al., 2017; Webb and Guerau-de Arellano, 2017). Similarly, the etiology of MS is currently unclear and still an active field of research. To date, the most common model is that MS is caused by infiltration of autoreactive immune cells into the CNS, which leads to demyelination and ultimately neurodegeneration (Filippi et al., 2018; Rodríguez Murúa et al., 2022).

1.4 DNA methylation

DNA, the polymer that carries the genetic instructions required for development and function of our bodies, encodes this information in a sequence consisting of four nucleotides: cytosine (C), guanine (G), adenine (A) and thymine (T). In 1948 however, long before DNA was identified as the carrier of genetic information, Hotchkiss made a curious discovery when he subjected hydrolyzed calf thymus DNA to paper chromatography: Instead of the expected four bands, one for each nucleotide, he observed a faint fifth band which he labeled "epi-cytosine" (Fig. 1.5).

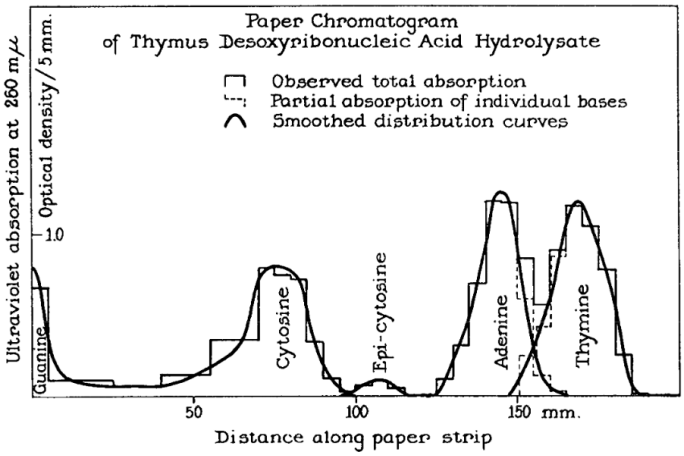

FIGURE 1.5: The first evidence for DNA methylation in mammalian genomes.
Guanine, cytosine, adenine, thymine and "epi-cytosine" (5-methylcytosine), as detected by paper chromatography in 1948. Figure taken from Hotchkiss (1948).

As methylated cytosine (5-methylcytosine, 5mC) was previously detected in nucleic acids isolated from *Mycobacterium tuberculosis* (Johnson and Coghill, 1925), Hotchkiss speculated that the fifth band in his chromatogram might correspond to 5mC. Today, we know that this suggestion was correct, as the genomes of many prokaryotes, plants, fungi, insects and mammals were shown to contain 5mC (Mattei et al., 2022).

In mammals, DNA methylation occurs predominantly in a specific sequence context, namely at the cytosine of 5′–**C**–**p**hosphate–**G**–3′ (CpG) dinucleotides (Sinsheimer et al., 1954). Although rare exceptions exist, CpG methylation is typically symmetric, which means that both the cytosine on the plus strand, as well as the cytosine complementary to guanine on the minus strand, are methylated (Xu and Corces, 2018). The genomic distribution of these CpG dinucleotides shows at least two statistical anomalies: First, CpG dinucleotides occur less frequently than would be expected by random chance. This phenomenon called CpG depletion is the result of the high mutation rate of 5mC, which is subject to spontaneous deamination. The result of this mutation is a thymine base that is improperly matched with a guanine. DNA repair mechanisms oftentimes replace this mismatch with an A=T base pair instead of the original C≡G, which leads to depletion of CpG dinucleotides over evolutionary timescales (Bird, 1980). Second, CpG dinucleotides are not uniformly

distributed across the genome. Instead, many CpG sites occur in clusters known as CpG islands. Most of these CpG-rich genomic regions are 300 to 3000 base pairs (bp) wide and lowly methylated (Bird, 1986; Fatemi et al., 2005). In adult mammals, however, most CpG sites located outside CpG islands are methylated. The total fraction of methylated CpG sites differs between cell types and tissues. For instance, a recent study which quantified CpG methylation in individual cells of various murine brain cell types reported average CpG methylation levels ranging from 61.6% to 88.8% (Liu et al., 2022).

Although less frequently, cytosines in a non-CpG context may also be methylated. Compared to CpG methylation non-CpG methylation is poorly understood, but recent evidence suggests it plays a role in neuronal diversity and development (Price et al., 2019; Jeong et al., 2021; Liu et al., 2022). Unless explicitly stated otherwise, all following mentions of DNA methylation, or simply methylation, refer to CpG methylation.

1.4.1 Molecular mechanisms of (de)methylation

Mammalian cells are equipped with an array of enzymes which allow them to add and remove methyl groups from cytosines in their genome. Early research on this molecular methylation machinery focused mostly on a family of enzymes called DNA methyltransferases (DNMTs), which catalyze the transfer of a methyl group from the universal methyl donor S-adenosylmethionine to C5 of the cytosine ring (reviewed in Lyko, 2018). The genome of *M. musculus* contains six members of the DNMT gene family: *Dnmt1*, *Trdmt1*, *Dnmt3a*, *Dnmt3b*, *Dnmt3c* and *Dnmt3l*.

DNMT1 is the **maintenance methyltransferase** of mammalian genomes, which means that it ensures that DNA methylation patterns are maintained during cell division. This is necessary, as the nascent strands synthesized during DNA replication are initially unmethylated. To restore the original symmetric DNA methylation pattern, DNMT1 interacts with other factors such as UHRF1 to bind hemi-methylated, i. e., asymmetrically methylated, CpG dinucleotides and methylates the unmethylated cytosine (Hirasawa et al., 2008; Xu and Corces, 2018).

In contrast to the maintenance methyltransferase DNMT1, ***de novo* methyltransferases** are able to introduce DNA methylation at CpG dinucleotides that were previously unmethylated, thus giving them the ability to establish new epigenetic patterns. In mice, this function is carried out by DNMT3A and DNMT3B (Okano et al., 1999), as well as the recently discovered, rodent-specific DNMT3C (Barau et al., 2016). The activity of these methyltransferases is well-studied during early embryonic development, where DNA methylation is first erased from the genome and then re-established *de novo* (reviewed in Lee et al., 2014; Chen and Zhang, 2020). Knockout of *Dnmt3a*, *Dnmt3b* or *Dnmt1* leads to premature death of developing mice at 4 weeks post natal, E13.5 and E9.5, respectively (Clark et al., 2022).

Other members of the DNMT gene family include DNMT3L and TRDMT1. DNMT3L is a truncated homolog of DNMT3A/B that lost methyltransferase activity, but nonetheless plays an important role in *de novo* methylation as it directs the activity of DNMT3A/B using histone modification marks as a guide (Ooi et al., 2007; Lee et al., 2014). The enzyme formerly known as DNMT2 was recently re-named to TRDMT1, as Goll et al. (2006) demonstrated that it methylates tRNA and not DNA.

Active and passive demethylation

The removal of DNA methylation can be achieved in two different ways. The first of these mechanisms, called **passive demethylation**, is simply the absence of maintenance methylation during cell division, resulting in a relatively indiscriminate global decrease of DNA methylation. This mode of demethylation naturally occurs in primordial germ cells, which lose about 90% of their global DNA methylation between E6.5 and E13.5 (Lee et al., 2014). Mechanistically, passive demethylation at this developmental time point is likely the result of multiple factors, including exclusion of DNMT1's interaction partner UHRF1 from the nucleus, downregulation of *Uhrf1*, loss of H3K9me2, as well as downregulation of the *de novo* methyltransferases (Lee et al., 2014). In addition to naturally occurring passive demethylation, unspecific global demethylation can also be induced artificially using DNMT1 inhibitors such as 5-azacytidine (Jones et al., 2016).

In contrast to passive demethylation, **active demethylation** is the targeted removal of DNA methylation from specific cytosines in the genome. The enzymes responsible for active demethylation were unknown for many decades (Mattei et al., 2022). Only recently, a computational sequence-similarity search and follow-up experiments unveiled the Ten-Eleven Translocation (TET) family as the enzymes responsible for active demethylation in mammals (Tahiliani et al., 2009). TET enzymes achieve this feat in a stepwise series of oxidations which results in oxidized methylcytosine intermediates including 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) (Fig. 1.6, An et al., 2017). 5fC and 5caC are then replaced by unmethylated cytosines by the base excision repair, which completes the active demethylation process. However, as DNTM1 and UHRF1 are unable to recognize oxidized intermediates, TET enzymes also facilitate replication-dependent demethylation (An et al., 2017).

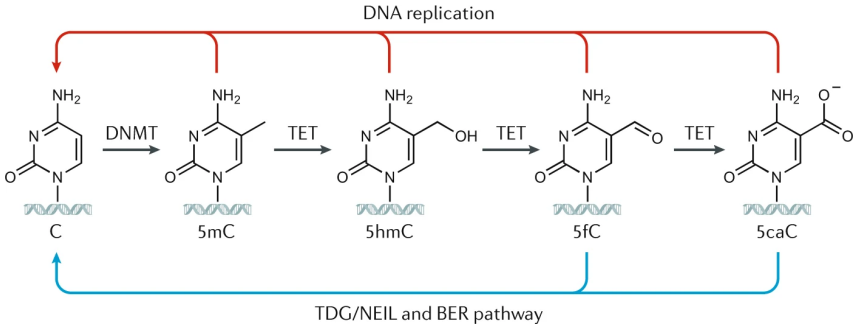

FIGURE 1.6: Function of DNMT and TET enzymes in passive and active DNA demethylation.
DNA methyltransferases (DNMTs) catalyze the methylation of cytosine (C), yielding 5-methylcytosine (5mC). This reaction can be reverted by TET (ten-eleven translocation) enzymes, which catalyze a series of oxidations that produce 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC). 5fC and 5hmC are the target of the thymine DNA glycosylase (TDG)/Nei-like (NEIL) and base excision repair (BER) machineries, which replace them with unmethylated cytosines (active demethylation, blue arrows). Furthermore, 5mC may be depleted in the absence of maintenance methylation during DNA replication (passive demethylation, red arrows). Figure and caption taken from Parry et al. (2021).

1.4.2 Methylation and gene regulation

Since its discovery almost a century ago, DNA methylation has been implicated in a vast range of biological processes including gene regulation, development, imprinting, genome organization, cell fate decisions, cancer, and aging (reviewed in Mattei et al., 2022; Parry et al., 2021; Lyko, 2018; Moore et al., 2013; Lee et al., 2014; Jones, 2012). In this section I will focus on gene regulation and cell fate decisions and briefly review how these two topics are linked to DNA methylation, as these aspects are most relevant for my research.

Methylation as a repressive epigenetic mark

Historically, DNA methylation has been viewed as a repressive epigenetic mark. A number of findings support this view. For instance, artificially induced methylation of individual promoters was shown to decrease gene expression and genome-wide studies found that DNA methylation generally anticorrelates with transcription (Mattei et al., 2022). In the last decades, several studies also offered mechanistic explanations on how DNA methylation can lead to lower gene expression. Methylation readers such as methyl-CpG binding protein 2 (MeCP2, *Mecp2*), for instance, might bind to methylated CpG dinucleotides and recruit the nuclear receptor corepressor 1/2 (NCoR1/2) complex, which promotes the formation of repressive chromatin structures as well as histone deacetylation (Tillotson and Bird, 2020). Recent evidence also suggests that many transcription factors are sensitive to DNA methylation, i. e., they bind preferentially to sequence motifs containing 5mC or unmethylated cytosines (Yin et al., 2017; Kaluscha et al., 2022). Surprisingly, effects of nearby DNA methylation on transcription factor binding were even described for transcription factors whose binding motif does not contain a CpG dinucleotide

(You et al., 2011). Interactions between DNA methylation and transcription factor binding are bidirectional though, as some transcription factors are also able to affect DNA methylation (Tchieu et al., 2019; Donaghey et al., 2018; Reizel et al., 2018).

While the repressive properties of DNA methylation are still acknowledged today, more recent research also revealed more nuances to this relationship as they distinguished between methylation of promoters, gene bodies, enhancers and other genomic features (Moore et al., 2013; Lee et al., 2014). It is long known that most mammalian proximal promoters are lowly methylated as the majority of promoters, and almost all promoters of housekeeping genes, contain CpG islands which are generally hypomethylated (Bird, 1986; Gardiner-Garden and Frommer, 1987; Saxonov et al., 2006). Accordingly, since most promoters are lowly methylated, an unmethylated promoter alone is not sufficient to enable gene expression. Promoter hypermethylation, however, is associated with silencing of gene expression (Saxonov et al., 2006; Rollins et al., 2006; Mattei et al., 2022).

Surprisingly though, several studies suggest that promoter hypermethylation is often introduced after a gene is already silenced, indicating that promoter methylation is not the cause, but rather a consequence of gene silencing. Bird (2002) thus proposed that promoter methylation serves to silence genes "irrevocably", i. e. to achieve stable, long-term silencing of gene expression. For this reason, modern targeted gene silencing tools using engineered dCas9 [1] fusion proteins often utilize methyltransferase domains (Nakamura et al., 2021). It is worth noting, however, that there is still substantial disagreement and uncertainty on the role of promoter methylation in gene regulation after many decades of research on this topic. Only recently, Ford et al. (2017) conducted a screening in which they artificially induced methylation in thousands of promoters. Surprisingly, the authors reported that this often did not result in gene repression; a claim that was swiftly rebutted in a follow-up pre-print in which Ford et al.'s data was re-analyzed by Korthauer and Irizarry (2018), which again prompted a follow-up screening with additional controls (de Mendoza et al., 2022).

Gene body methylation

Despite the aforementioned view that DNA methylation represses gene expression, several studies found that the gene bodies of expressed genes tend to be methylated (Lister et al., 2009; Jones, 2012). Several explanations for this surprising positive correlation between gene expression and gene body methylation were proposed. For instance, gene body methylation might serve to silence cryptic intragenic promoters (Tran et al., 2005; Neri et al., 2017) or regulate splicing (Laurent et al., 2010; Shukla et al., 2011). Despite frequent reports of a positive correlation between gene body methylation and gene expression, some studies also obtained seemingly conflicting

[1] dCas9: endonuclease deficient ("dead") CRISPR (clustered regularly interspaced short palindromic repeats)-associated protein Cas9.

results when they found a negative correlation between methylation of the first exon (Brenet et al., 2011), the first intron (Anastasiadi et al., 2018; Schlosberg et al., 2017) or immediately downstream of the transcription start site (TSS) (Hovestadt et al., 2014). These findings suggest that gene body methylation near the 5′ end of the gene has a repressive effect on gene expression, much like promoter methylation. For this reason, it might be wise to divide the gene body into a 5′ domain (where methylation is typically repressive) and a 3′ domain (where methylation is often positively correlated with gene expression) when quantifying gene body methylation (Brenet et al., 2011; Moore et al., 2013).

Cell type-specific methylation patterns at cis-regulatory elements

Driven by advances in epigenome sequencing and other technologies, research in the last decade has deepened our understanding of epigenetic differences between tissues and cell types. Genome-wide epigenome profiling techniques such as **whole-genome bisulfite-sequencing (WGBS)** and chromatin immunoprecipitation followed by sequencing (ChIP-seq) made it possible to assess epigenetic marks, as well as their interactions, for the entire genome. An important conclusion derived from such studies is that different cell types and lineages possess unique methylation profiles, characterized by specific **lowly methylated regions (LMRs)** (Lister et al., 2009; Stadler et al., 2011; Lee et al., 2014; Argelaguet et al., 2019). As a result, cell types and tissues are readily distinguished by assessing DNA methylation at relatively few genomic loci (Ziller et al., 2013; Lee et al., 2014). Many of these LMRs appear to be distal gene regulatory elements, as they bind cell type-specific transcription factors and are enriched for enhancer chromatin marks (Stadler et al., 2011; Argelaguet et al., 2019) or repressive sequence motifs characteristic of silencers (Cedar et al., 2022). Methylation at these distal regulatory sites might thus provide a basic framework supporting the expression of genes required in specific tissues or cell types (Ramsköld et al., 2009; Lee et al., 2014), while possibly also silencing other genes which are not currently required (Edrei et al., 2021). How exactly methylation at these distal sites might regulate gene expression is incompletely understood, but it is evident that DNA methylation interacts with other gene regulatory components such as histone modifications and transcription factors (Fig. 1.7) (Noh et al., 2015; Angeloni and Bogdanovic, 2019; Cedar et al., 2022).

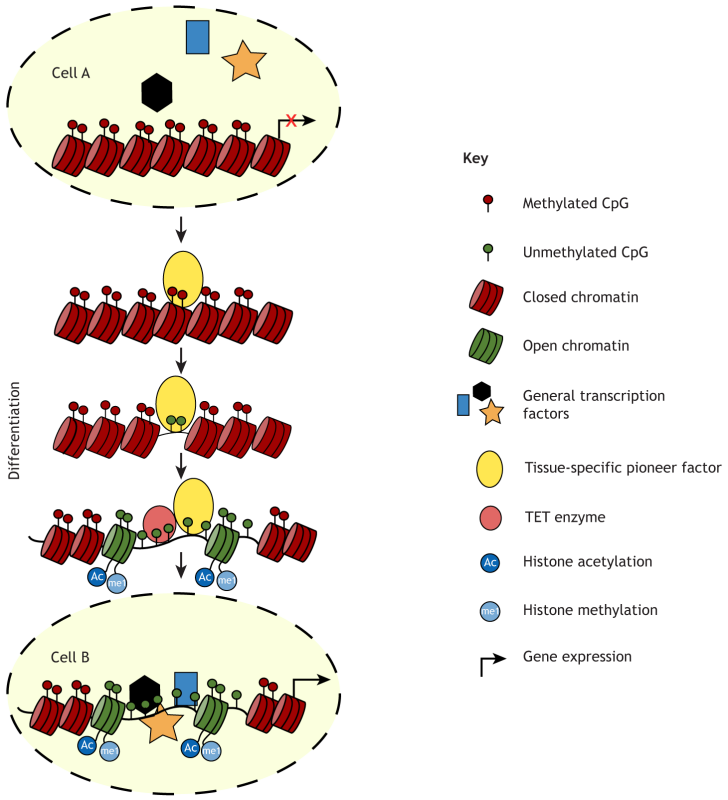

FIGURE 1.7: Model of the influence of DNA methylation on regulatory elements during cell differentiation.
This schematic depicts a model proposed by Cedar et al. (2022), who argue that demethylation of regulatory elements may promote cell type-specific gene expression. In cell A (top), these regulatory elements are methylated, packed in closed chromatin and inaccessible to transcription factors. During development, pioneer transcription factors bind sequence motifs at regulatory elements, cause demethylation (for instance by recruiting TET enzymes), and promote local chromatin accessibility. This process subsequently leads to a broad hypomethylated region of increased chromatin accessibility, decreased nucleosome occupancy and specific histone modifications. This broad region then remains stably open and is thus accessible to transcription factors that are ubiquitously expressed in various cell types and tissues. Ultimately, these regions enable cell-type specific gene expression in cell B (bottom), even in the absence of pioneer transcription factors which were initially required. Figure and caption modified from Cedar et al. (2022).

1.4.3 Effects on cell differentiation and fate decisions

In stark contrast to somatic tissues, which possess global methylation levels between 60% and 80%, embryonic stem cells (ESCs) possess surprisingly low methylation levels between 5% and 30%. This feature of ESCs is the result of the global epigenetic reprogramming that occurs as part of early embryonic development, which includes two major waves of demethylation (reviewed in Lee et al., 2014; Chen and Zhang, 2020; Cedar et al., 2022). A similar phenomenon is observed during artificially induced cellular reprogramming: When induced pluripotent stem cells (iPSCs) are generated from somatic cells, this is also accompanied by a global decrease in DNA methylation, which makes it tempting to speculate that loss of DNA methylation is required for pluripotency and stem cell function (Lee et al., 2014).

Recently, advances in single-cell epigenomics and methylation reporter systems allowed for a more detailed view on epigenome dynamics in early development. Using a newly developed single-cell triple-omics protocol (Clark et al., 2018, see section 1.5.3 for details), the laboratory of Wolf Reik showed that gastrulation is accompanied by drastic methylation changes at lineage-specific regulatory elements. Specifically, Argelaguet et al. (2019) showed that lineage specification into one of the three primordial germ layers requires TET-mediated active demethylation at lineage-specific regulatory elements. In a follow-up study, Clark et al. (2022) furthermore demonstrated that DNMTs are required to silence alternative cell fates and exit pluripotency during early organogenesis. Similarly, knockout of TET enzymes prevented the generation of primitive erythrocytes, strengthening the view that active demethylation contributes to lineage decisions (Clark et al., 2022).

Curiously, research of the last decade demonstrated that turnover of DNA methylation during cell differentiation occurs in a highly dynamic, oscillatory fashion (reviewed in Parry et al., 2021). This dynamic is the result of simultaneous activity of both *de novo* methyltransferases and TET enzymes at the same distal regulatory sites. Parry et al. hypothesize that this stochastic turnover increases the diversity of stem cell populations, which may lead to transient lineage biases in a sub-set of the stem cell pool. Amplification of these lineage biases might then lead to establishment of stable lineages with distinct gene regulatory networks.

Dynamic DNA methylation in adult tissues

While most research on DNA methylation dynamics in cellular differentiation processes has focused on the developing embryo, some recent studies also inquired whether methylation changes occur in adult mammals. To this end, two recent studies disrupted the DNA methylation machinery in the hematopoietic stem cell lineage, which resulted in skewed lineage proportions (Izzo et al., 2020; Ostrander et al., 2020). While knockout of *Dnmt3a* favored erythroid over myelomonocytic cells, knockout of *Tet2* showed the opposite effect, suggesting that both active demethylation and *de novo* methylation are crucial for hematopoietic lineage decisions. Along the same line, Schönung et al. (2023) very recently employed methylation microarrays to reveal a progressive establishment of DNA methylation patterns at cis-regulatory elements during murine hematopoiesis. A small number of studies also reported dynamic DNA methylation in other adult tissues. For instance, Reizel et al. (2015) observed specific DNA methylation changes in the livers of adult male mice upon reaching sexual maturity, while Dos Santos et al. (2015) observed such changes in the mammary gland upon pregnancy. Overall, research of the last decade challenged the classical view that DNA methylation is a static epigenetic mark.

1.5 Single-cell sequencing

Sequencing techniques such as RNA-seq and WGBS have revolutionized modern biology by providing a means to "zoom out" from individual genes or loci, enabling us to study the entirety of all genes, transcripts or loci at once. In the last two decades, so-called *omics* techniques based on next-generation sequencing (reviewed in Hu et al., 2021) were a driving force behind discoveries in many different fields including adult NSC biology (section 1.1) and the study of DNA methylation (section 1.4). To date, most of these protocols are **bulk sequencing** protocols, which means they are designed to characterize or quantify nucleic acids in a given sample. Most biological samples consist of a mixture of different cells, and in bulk sequencing the obtained measurements correspond to the population average. A common application of, for instance, bulk RNA-seq is to compare gene expression between samples subject to different treatments or conditions. This allows for the discovery of differentially expressed genes which might be associated with a given treatment or condition.

As many samples consist of a heterogeneous mixture of cells, however, a fundamental limitation of bulk RNA-seq is that any observed differences cannot be attributed to a specific cell population. For instance, RNA-seq of healthy and inflamed tissue samples might lead to the identification of genes upregulated in inflammation, but it cannot inform us whether these genes are equally upregulated in all cell types of the tissue, or merely in resident immune cells. Similar limitations apply to other bulk sequencing protocols such as WGBS, which allows us to observe genome-wide DNA methylation profiles for a sample, but does not inform us about potential methylation differences between cell types within that sample. To address these shortcomings, many bulk sequencing protocols were subsequently modified to allow for sequencing of individual cells (reviewed in Kashima et al., 2020; Wen and Tang, 2022).

In this section, I will briefly review the **single-cell sequencing** methods employed in the course of my doctoral research. Furthermore, I will elaborate on computational methods used to analyze single-cell RNA-seq data, as they inspired the methods I developed for the analysis of single-cell methylome data (section 2).

1.5.1 Single-cell RNA-sequencing

scRNA-seq protocols employ basic principles from bulk RNA-seq (reviewed in Van den Berge et al., 2019) and modify them to achieve cell-wise quantification of mRNA transcripts. A key principle of scRNA-seq is to separate single cells into individual reaction compartments [2]. Inside each compartment, distinct, compartment-specific molecular tags are then added to each transcript, which allows for cell-wise

[2] Although not yet widely adopted, some recent methods based on combinatorial indexing no longer follow this principle, see for instance Datlinger et al. (2021) or, for a brief summary of such methods, Baysoy et al. (2023) and Fig. 3 therein.

demultiplexing of transcripts at the data processing stage. Currently, the most popular scRNA-seq protocols achieve this either using microwell plates or microfluidic devices (Heumos et al., 2023; Ding et al., 2020).

Plate-based scRNA-seq protocols (Picelli et al., 2013; Hashimshony et al., 2016; Hagemann-Jensen et al., 2020) require the deposition of single cells into individual microwells using manual pipetting, FACS, or liquid handling robots. In these protocols, transcript tagging is typically achieved by the use of commercially available index primers that allow users to pool multiple libraries and sequence them together (Picelli et al., 2013). Compared to other methods, plate-based scRNA-seq is relatively low-throughput since the number of cells is bounded by the number of wells on the plate. Obtaining large data sets comprising thousands of cells thus requires the use of multiple plates and sequencing runs, which is labor-intensive and can lead to the introduction of additional technical noise and batch effects (Heumos et al., 2023).

Droplet-based scRNA-seq methods (Macosko et al., 2015; Zheng et al., 2017) use microfluidics to encapsulate single cells in small droplets containing reagents and one barcoded bead each. Each bead is coated with numerous primers that typically consist of

- a poly(dT) sequence used to capture the polyadenylated transcripts
- a barcode sequence used to tag all transcripts in the droplet with the same unique sequence ("cell barcode" or simply "barcode")
- an additional barcode used to tag all transcripts in the droplet with a different unique sequence (unique molecular identifier, UMI)
- a PCR handle.

The cell barcode is used to identify transcripts that were encapsulated in the same droplet, i. e., ideally transcripts that originated from one single cell, while UMIs are used to quantify the original mRNA molecules instead of duplicates introduced in subsequent PCR (Islam et al., 2014; Macosko et al., 2015).

Historically, UMIs were a unique advantage of droplet-based methods, but note that recent plate-based approaches also incorporate UMIs (Hashimshony et al., 2016; Hagemann-Jensen et al., 2020). Contemporary advantages of droplet-based methods include their increased throughput, as well as the commercial availability of ready-to-use kits provided by companies such as 10X Genomics. In contrast to plate-based methods, droplet-based methods suffer from a lower overall sensitivity (although this can be partially explained by lower recommended sequencing depths) and a higher rate of doublets, i. e., two cells that were erroneously assigned one cell barcode (Ding et al., 2020). Furthermore, unlike many plate-based protocols, droplet-based scRNA-seq typically does not achieve full-length transcript coverage as transcripts are usually sequenced from the 3′ end (Macosko et al., 2015).

1.5.2 Single-cell epigenome sequencing

The rise of scRNA-seq methods demonstrated the immense utility of distinguishing individual cells in omics assays. Naturally, many epigenomics methods were also adapted to enable single-cell resolution. The most popular and mature single-cell epigenomics method to date is **single-cell ATAC-seq** (**a**ssay for **t**ransposase-**a**ccessible **c**hromatin using **seq**uencing, Buenrostro et al., 2015). Similar to bulk ATAC-seq (Buenrostro et al., 2013), this method detects open chromatin using a mutated, hyperactive Tn5 transposase which inserts sequencing primers at accessible genomic regions.

Single-cell methylome sequencing

The development of sequencing methods that allow for genome-wide quantification of DNA methylation at single-cell resolution is an ongoing endeavor that started one decade ago when the first such protocols were reported (Guo et al., 2013; Smallwood et al., 2014; Luo et al., 2017; Mulqueen et al., 2018; Luo et al., 2018; Shareef et al., 2021; Nichols et al., 2022; Chatterton et al., 2023). Like bulk WGBS, **single-cell bisulfite sequencing (scBS)** protocols are based on a technique called bisulfite sequencing, whereby genomic DNA is treated with sodium bisulfite (Frommer et al., 1992). While 5mC (and oxidized methylcytosine intermediates such as 5hmC, Huang et al., 2010) is protected from the mutagenic effects of bisulfite, unmethylated cytosine is converted to uracil, which is read as thymine in subsequent PCR and sequencing steps. As a result, unmethylated cytosines are marked by C–to–T transitions, which are readily detected when sequencing reads are mapped to a reference genome (Krueger and Andrews, 2011).

While scBS allows us to assess DNA methylation at single-base and single-cell resolution, there are also downsides to this approach. First, bisulfite conversion is unable to distinguish 5mC from oxidized methylcytosine intermediates such as 5hmC (Huang et al., 2010). Second, sodium bisulfite causes DNA degradation (Grunau et al., 2001). Third, unlike scRNA-seq which focuses on transcripts, scBS aims to quantify a genome-wide property, which means that a large number of sequencing reads are required to achieve modest sequencing coverage. As a result of the latter two points, scBS is currently expensive and the obtained data sets are very sparse, which introduces additional challenges during downstream analysis (section 1.5.4).

1.5.3 Single-cell multi-omics

Very recently, single-cell sequencing methods with different molecular readouts were combined to enable simultaneous quantification of gene expression and, for instance, epigenomic features of the same cells (reviewed in Baysoy et al., 2023). The journal Nature Methods selected these single-cell multi-omics approaches as

the Method of the Year 2019 (Teichmann and Efremova, 2020), as they hold great promise for the investigation of gene regulatory networks (Zhu et al., 2020), stem cell biology (Kucinski and Gottgens, 2020), immunology (Nathan et al., 2019) and neurobiology (Huang and Paul, 2019).

Compared to plain scRNA-seq, single-cell multi-omics has several key advantages. Combining gene expression with other molecular readouts such as chromatin accessibility, histone modifications or DNA methylation might allow for more accurate identification of cell types and cell states while also shedding light on the molecular features that underpin these. Furthermore, some cellular features such as stem cell fate bias are only partially detectable in the transcriptome and might be encoded in epigenetic features instead (Weinreb et al., 2020). While these questions might also be explored with parallel scRNA-seq and single-cell epigenomics experiments, obtaining multiple molecular readouts from each cell has the unique advantage that these modalities can be directly related, for instance to identify enhancers and infer gene regulatory networks (Bravo González-Blas et al., 2023). Furthermore, the joint analysis of parallel "single-omic" data sets requires elaborate methods to integrate data across modalities and enable transfer of cell type labels (e. g. Stuart et al., 2019), which may introduce errors or artifacts. Finally, a unique advantage of single-cell multi-omics is that the scRNA-seq portion of the data can be used to guide the analysis and interpretation of other under-examined molecular features. For instance, the transcriptome layer might be used to relate new multi-omics data sets to existing scRNA-seq data sets, or to annotate cell types based on the expression of marker genes. Both of these tasks would pose a significant challenge for the analysis of e. g. single-omic scBS data sets.

scNMT-seq

Single-cell Nucleosome, Methylation and Transcription sequencing (scNMT-seq) is a recently developed single-cell multi-omics protocol that allows for simultaneous quantification of gene expression, DNA methylation and nucleosome occupancy (Argelaguet et al., 2019, Fig. 1.8). In essence, scNMT-seq is a combination of Nucleosome Occupancy and Methylation sequencing (NOMe-seq, Kelly et al., 2012) with the popular scRNA-seq protocol Smart-seq2 (Picelli et al., 2013). **NOMe-seq** is a bulk sequencing protocol based on WGBS that allows for the simultaneous quantification of DNA methylation and nucleosome occupancy (which is tightly linked to chromatin accessibility, Tsompana and Buck, 2014; Nordström et al., 2019). This is achieved using the viral methyltransferase M.CviPI, which preferentially targets GpC dinucleotides in accessible, nucleosome-depleted genomic regions (Kelly et al., 2012). Following bisulfite conversion, methylated cytosines can then be detected at the data processing stage. Depending on the sequence context, methylated cytosines are either interpreted as endogenous methylation marks (CpG sites), or as nucleosome-depleted, accessible genomic loci (GpC sites). scNMT-seq was

previously used to study epigenetic dynamics in early embryonic development (Argelaguet et al., 2019; Clark et al., 2022, see section 1.4.3).

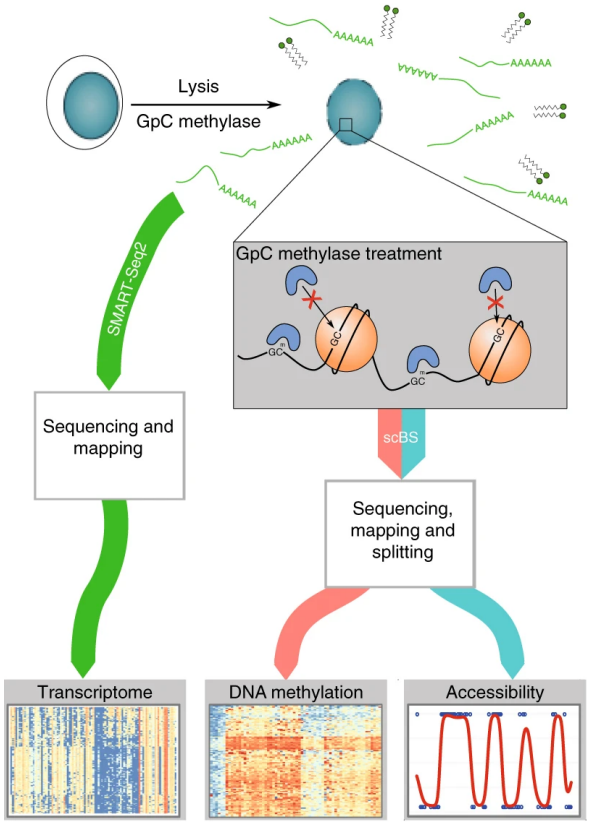

FIGURE 1.8: Schematic of the scNMT-seq protocol.
Individual cells are lysed and accessible chromatin is labeled using an enzyme that preferentially methylates GpC-dinucleotides at accessible chromatin. DNA and mRNA are then separated using magnetic beads. mRNA is sequenced using the Smart-seq2 protocol (Picelli et al., 2013), while DNA is sequenced with the scBS protocol of Clark et al. (2017). Depending on the sequencing context in which DNA methylation is observed (CpG or GpC), DNA methylation is either interpreted as endogenous methylation or as chromatin accessibility in the downstream bioinformatic analysis. Figure and caption adapted from Clark et al. (2018).

A limitation of scNMT-seq is that cytosines in the G–C–G context have to be discarded from the analysis, since it is not possible to distinguish endogenous CpG methylation from artificially introduced GpC methylation at these sites (Clark et al., 2018). Similarly, as the viral methyltransferase possesses off-target activity at CC-dinucleotides (Kelly et al., 2012), Argelaguet et al. furthermore discarded cytosines in the C–C–G context. Thus, the quantification of chromatin accessibility comes at the cost of ca. 48% of CpG sites, which exacerbates the sparsity of the obtained single-cell methylome data.

In contrast to count-based methods such scATAC-seq, scNMT-seq encodes chromatin accessibility through bisulfite conversion. Compared to scATAC-seq, which cannot distinguish stochastic lack of sequencing coverage from inaccessible chromatin, this offers a unique advantage. The downside of this approach, however,

is that scNMT-seq aims for genome-wide coverage, while scATAC-seq merely sequences accessible chromatin. This results in a higher number of required sequencing reads, which in turn gives rise to higher sequencing costs or increased sparsity. Finally, a major limitation of scNMT-seq is its low throughput and the lack of UMIs in the scRNA-seq portion of the protocol. Both of these limitations were addressed by Cerrizuela et al. (2022) during the course of this work.

1.5.4 Computational analysis of single-cell omics data

The popularization of new sequencing protocols necessitates the development of computational methods that are tailored specifically to the analysis of these new data sets. For this reason, the maturation of scRNA-seq technologies was accompanied by the emergence of workflows to analyze the obtained data sets (reviewed in Amezquita et al., 2020; Heumos et al., 2023), as well as user-friendly software packages such as Seurat (Stuart et al., 2019) and ScanPy (Wolf et al., 2018) that implemented these workflows. A similar trend is currently observed in the scATAC-seq field, which also gave rise to custom software and algorithms (e. g. Pliner et al., 2018; Stuart et al., 2021; Bravo González-Blas et al., 2023).

Analysis of single-cell RNA-sequencing data

Similar to bulk RNA-seq, the initial scRNA-seq data processing steps consist of mapping the sequencing reads to a reference genome or transcriptome, followed by quantification of reads at the gene level. Depending on the protocol used, additional steps might be required to resolve cell barcodes and UMIs (reviewed in Heumos et al., 2023). These additional processing steps are implemented in scRNA-seq tools such as Cell Ranger (Zheng et al., 2017) or zUMIs (Parekh et al., 2018). Ultimately, the main goal of initial data processing is to obtain a cell × gene **count matrix** [3], i. e. a matrix that denotes, for every gene in every cell, how many sequencing reads (or UMIs) were assigned to that gene (Fig. 1.9). This matrix summarizes the obtained gene expression data and serves as input for downstream exploratory data analysis. A characteristic feature of scRNA-seq count matrices is that they are typically very sparse, i. e. most of the count values are zero. This phenomenon is likely the result of multiple factors including the low mRNA content of single cells, the stochasticity of gene expression, as well as random sampling of mRNA molecules during mRNA capture and sequencing (Svensson, 2020; Qiu, 2020).

[3] Sometimes also written as gene × cell. Also note that microfluidic droplets or microwells might accidentally contain multiple cells, or no cells at all. Thus, not all rows in this count matrix are guaranteed to correspond to a single cell.

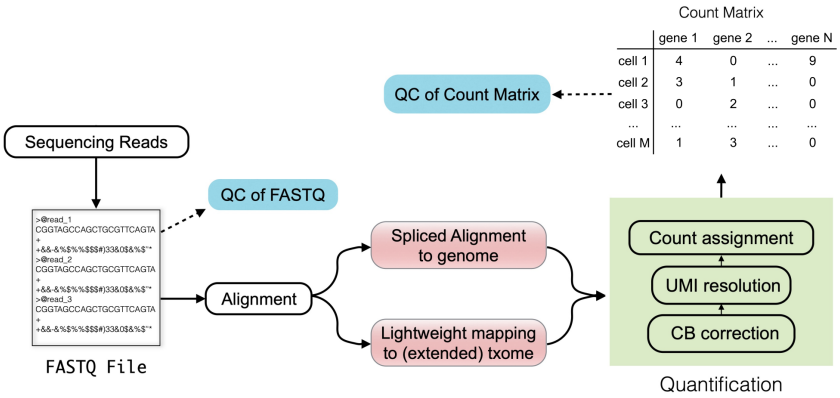

FIGURE 1.9: Initial data processing steps of typical droplet-based scRNA-seq. Sequencing reads in FASTQ format are first subjected to quality control (QC) and, if the quality is acceptable, read 2 of each read pair is aligned to the genome or transcriptome (txome) of the sequenced species. Read 1 of each read pair is then inspected for two barcode sequences: The cell barcode (CB) and the unique molecular identified (UMI). Allowing for few mismatches due to potential sequencing errors, each read pair is then assigned a CB and a UMI. Finally, the obtained data is summarized in a cell × gene count matrix. This matrix reports how many unique molecules, i. e., reads after UMI-based deduplication, were aligned to each gene. The count matrix is furthermore subjected to QC, which includes the removal of UMIs with very few observed molecules (i. e., empty droplets or low-quality cells). Figure adapted from Heumos et al. (2023).

Once the raw count matrix is obtained, it is subjected to quality filtering steps such as the detection and subsequent removal of low-quality cells, doublets and empty droplets or wells (Heumos et al., 2023; Amezquita et al., 2020). The rows of a count matrix, corresponding to the cells, are then "log-normalized". This means they are first normalized to account for differences in library size between cells, and subsequently transformed to reduce heteroskedasticity, i. e., the circumstance that the counts of highly expressed genes have a higher variance than those of lowly expressed genes. This transformation typically involves logarithmization after adding a small numeric value, to prevent taking the logarithm of zero, although other transformations have been proposed (see Ahlmann-Eltze and Huber, 2023, for a systematic comparison of different transformations).

A key requirement for a successful scRNA-seq analysis is the ability to assign cell types and cell states. Not all genes are equally useful for this task, however. Examples of uninformative genes might include, for instance, housekeeping genes with near-constant expression across all sampled cells, as well as lowly-expressed genes that were only observed in a very few cells. Thus, count matrices are subjected to feature selection. In its most basic form, this involves the use of two thresholds: a threshold on mean gene expression, to discard lowly expressed genes, and a threshold on the variance of the log-normalized counts, to select so-called **highly variable genes** (Amezquita et al., 2020). The reasoning behind variance thresholding is to exclude uninformative genes with near-constant expression across cell types, while selecting those genes that have higher expression in some cells compared to others, as these might be useful to distinguish cell states.

One of the main reasons for the abovementioned pre-processing and quality filtering steps is to obtain a matrix that is suitable as input for principal component analysis (PCA), a popular dimensionality reduction technique that is used to reduce the cell × gene matrix to a cell × PC (principal component) matrix that has fewer columns but retains most of the information. This procedure reduces noise as uninformative PCs are typically discarded, while also decreasing subsequent compute time as the reduced matrix is much smaller than the full count matrix. Furthermore, PCA serves as a starting point for data exploration, as PCs reveal groups of genes with correlated gene expression, which often correspond to biological processes (Wagner et al., 2016).

The reduced cell × PC matrix is then used as input for dedicated data visualization algorithms such as uniform manifold approximation and projection (UMAP, McInnes et al., 2018), which place cells on a two-dimensional plane while attempting to place cells with similar transcriptomes close to each other. To annotate cell types, the most common approach is to subject cells to a clustering algorithm (e. g. Leiden clustering, Traag et al., 2019) to discover groups of cells with similar transcriptomes. These clusters of cells are then inspected for marker genes and manually annotated (see e. g. Pliner et al., 2019; Frauhammer and Anders, 2022, for alternatives to this approach). Common downstream analysis methods include pseudotemporal analysis, i. e. ordering the cells along a two-dimensional differentiation trajectory (Haghverdi et al., 2016; Qiu et al., 2017; Street et al., 2018), the identification of differentially expressed genes between groups of cells (Soneson and Robinson, 2018), as well as the integration of two or more distinct data sets to allow for their joint analysis while reducing the influence of batch effects (Haghverdi et al., 2018; Stuart et al., 2019).

Analysis of single-cell methylome data

Due to the wealth of existing scRNA-seq methods, a common strategy is to adapt these methods for scBS data. However, there are fundamental differences between scRNA-seq and scBS data sets that hinder this transfer of methods:

- scRNA-seq quantifies discrete features (transcripts or genes), while scBS quantifies DNA methylation of the entire genome.
- scRNA-seq measures transcript abundances and thus yields count data, while scBS measures the methylation state of individual cytosines, yielding binary data.
- In contrast to scRNA-seq, where a lack of sequencing reads (a zero in the count matrix) indicates low gene expression, scBS data contains true missing values: The methylation status of cytosines that are not covered by a sequencing read in a given cell is simply unknown.

Raw scBS data can be represented by a cell × cytosine matrix where each value is either 1 for methylated cytosines, 0 for unmethylated cytosines, or NA (not available) for cytosines with unknown methylation status due to lack of sequencing coverage.

Using this matrix for data exploration is challenging as the vast majority of values is NA and since most data exploration algorithms are not designed to work with binary data.

To bypass these challenges and ultimately make scBS data more similar to scRNA-seq data, most scBS studies preprocess scBS data by first segmenting ("**tiling**") the genome into intervals of a fixed width, often 100 kb, and then averaging DNA methylation in each interval for each cell (Luo et al., 2017; Liu et al., 2022) (Fig. 1.10). Other studies forego genome tiling in favor of a more targeted approach to define genomic intervals of interest. For instance, Argelaguet et al. (2019) anticipated methylation change at regulatory elements and thus performed ChIP-seq on histone modifications to define a set of genomic intervals. Other studies employ a hybrid strategy combining both regulatory annotations and genomic tiles (Farlik et al., 2015; Chatterton et al., 2023).

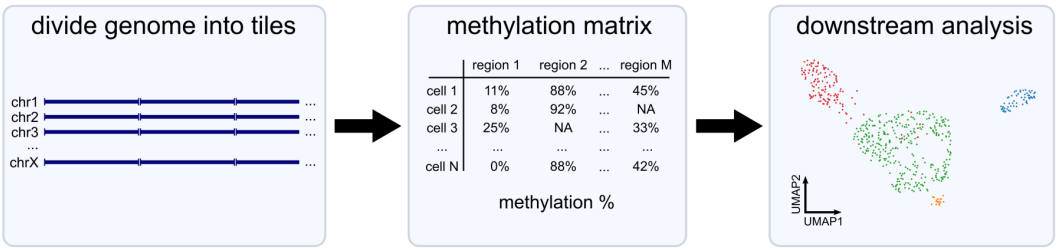

FIGURE 1.10: The current default strategy for analyzing scBS data.
In a first step, the entire genome is segmented into fixed intervals of a constant length such as 100 kb. Next, CpG methylation is quantified and averaged within each of these tiles for every cell. The obtained methylation fractions, often represented as a percentage, are then represented by a cell × tile methylation matrix. Since sequencing coverage is typically very low for individual cells, some intervals may not contain a single sequencing read, which means that the methylation matrix may contain missing values (NA). The obtained matrix may subsequently be subjected to downstream analysis methods such as dimensionality reduction (e. g., PCA, UMAP) or clustering. Figure adapted from Kremer et al. (2022a).

This tiling and averaging approach addresses all three aforementioned challenges with scBS data: Tiling the genome defines a set of features, similar to the genes which are quantified in scRNA-seq. Averaging within these tiles yields continuous values, often represented as percentages, which are more suitable as input for downstream analysis algorithms than binary values. Lastly, tiling reduces the number of missing values, as sufficiently large genomic tiles are likely to contain at least one sequencing read and NA values are ignored when averaging. Thus, tiling and averaging offers a pragmatic solution to some of the challenges posed by scBS data. Nonetheless, as I will demonstrate in the next chapter, this approach also leads to dilution of signal and increased levels of noise. To address these issues, I devised an alternative approach to analyze scBS data, which I will present in chapter 2.

1.6 Aims of this study

Adult NSCs are specialized astrocytes that possess neurogenic capabilities (section 1.1.4, Doetsch et al., 1999; Kriegstein and Alvarez-Buylla, 2009; Schneider et al., 2019). Despite this unique function, the transcriptome of NSCs closely resembles that of common parenchymal astrocytes, which raises the question which molecular features underlie NSC stemness. In this work, I discovered that adult NSCs possess a unique DNA methylome that is not observed in common parenchymal astrocytes. I propose that it is this unique epigenetic makeup that permits stem cell function by enabling the expression of genes required for neurogenesis.

In detail, the objectives of this study were to:

I. Characterize epigenetic change along the adult NSC lineage.
As outlined in section 1.1.5, the use of single-cell transcriptomics allows for a detailed characterization of changes in gene expression that occur along the adult NSC lineage. However, whether progression along the NSC lineage is likewise accompanied by epigenetic changes is currently unknown. To explore this possibility, I analyzed a large single-cell triple-omic data set which contains information on gene expression, chromatin accessibility, and DNA methylation. This data set comprises NSC-lineage cells, oligodendrocytes, and striatal astrocytes. More specifically, I aimed to quantify at which stages of the NSC lineage epigenetic changes are introduced, whether these changes are enriched for specific transcription factor binding site (TFBS) motifs, and how these changes relate to gene expression.

II. Develop computational methods that improve the analysis of single-cell methylome data.
Analyzing the abovementioned single-cell triple-omic data set requires computational methods designed to address the unique challenges posed by scBS data (section 1.5.4). Due to the lack of software designed for this purpose, I aimed to develop and implement several algorithms specifically designed for scBS data. I demonstrated that my novel methods yield better results than existing approaches and furthermore released these methods to the scientific community as a Python software package called `scbs`.

III. Characterize DNA methylation differences between adult NSCs and common parenchymal astrocytes.
My exploratory analysis of single-cell multi-omics data revealed striking differences in DNA methylation between adult NSCs and common parenchymal astrocytes. To assess the implications of these differences, I aimed to functionally characterize the genes which occur in close proximity to differentially methylated regions. By relating gene expression dynamics in the NSC lineage to methylation dynamics, I then proposed a model that aims to explain how DNA methylation might contribute to stemness of NSCs (section 4.2.2).

IV. Assess whether the DNA methylome is remodeled during injury-induced neurogenesis.

My model predicts that a specific epigenetic makeup is required for neurogenesis. This NSC methylome is not observed in common parenchymal astrocytes of the striatum. However, previous studies (e. g. Magnusson et al., 2014) demonstrated that striatal astrocytes are able to initiate a neurogenic program under specific circumstances such as injury. My model would thus predict that striatal astrocytes undergo epigenome remodeling upon injury, which is why I next aimed to characterize the DNA methylomes of striatal astrocytes isolated from mouse brains that were subjected to ischemic injury. I was previously reported that vSVZ NSCs, too, become activated in response to injury, and that this response depends on interferon signaling (Llorens-Bobadilla et al., 2015, section 1.2.3). Thus, I set out to answer whether this response might be accompanied by changes in DNA methylation, and whether these potential changes are hampered in interferon-deficient mice.

Results and Discussion

Chapter 2

Computational methods for single-cell methylation data

2.1 Improvements to standard single-cell methylation analysis

Recent years have brought about new sequencing techniques that enabled quantification of DNA methylation at single-cell resolution. However, computational methods to analyze single-cell methylation data are still in their infancy. While scRNA-seq data is typically analyzed with the help of user-friendly tool kits such as ScanPy (Wolf et al., 2018) or Seurat (Stuart et al., 2019), studies employing scBS report the use of custom algorithms and scripts. As a result, analyzing scBS data is challenging, laborious, and often not reproducible.

To address these issues, I developed `scbs`, a software that facilitates storage, exploration and analysis of scBS data. These methods are briefly described in Kremer et al. (2022a). In the following section, I will describe the methods that I developed and implemented in `scbs` in more detail.

2.1.1 Storage of single-cell methylation data

A common strategy in scBS data analysis is to obtain a cell × feature methylation matrix, analogous to the count matrix in scRNA-seq. This makes it possible to utilize methods that have proven useful in the more mature scRNA-seq field, such as dimensionality reduction or clustering of cells. However, due to the many differences between scRNA-seq and scBS data (see section 1.5.4), obtaining this methylation matrix is not straightforward.

To construct a methylation matrix, it is necessary to quantify DNA methylation in a given set of genomic intervals in all cells that were sequenced. This requires efficient storage and access of scBS data. However, due to the lack of bisulfite-aware read mappers specifically designed for single-cell methylation data, one currently has to

resort to mappers designed for bulk WGBS [1]. Bulk WGBS mappers such as Bismark (Krueger and Andrews, 2011) or MethylPy (Schultz et al., 2015) typically report methylation calls in a tabular format, where each row represents a single cytosine (or CpG site). This file is generated for each bulk sample. In scBS, where there are potentially hundreds or thousands of single-cell samples, this procedure thus results in a large number of tabular files. To save disk space and, most importantly, enable fast access to methylation values of a given genomic interval, I first devised an approach to efficiently store scBS data.

To achieve this, I first had to represent the obtained methylation data as numerical values. In the case of a single cytosine base, the methylation status can simply be represented by a binary value, i. e. 1 to denote a methylated cytosine and 0 to denote an unmethylated cytosine. However, this simplicity does not fully apply to a single CpG site, which contains two cytosines in total due to base pairing, which may differ in their methylation status. Similarly, at the single-cell level, the methylation status of a CpG site may differ between alleles. However, both of these sources of heterogeneity are rare (Arand et al., 2015). Furthermore, since scBS currently results in very low read coverage per cell, it is very rare to obtain more than one read per CpG site, which means that this heterogeneity cannot usually be detected. In the rare case where multiple conflicting reads are mapped to a CpG site, I thus used a simple majority vote to obtain a binary methylation value that summarizes the methylation status of that CpG. In the even rarer case where exactly half of the reads suggested methylation, I discarded the CpG site from the data set for that cell.

Applying this approach to all CpG sites of all sequenced cells yields binary values that can be represented by a matrix where rows represent CpG sites and columns represent cells. However, due to the low sequencing coverage of scBS data, most CpG sites are not covered by a sequencing read in the majority of cells, which results in many missing values in this matrix. To some extent, this resembles scRNA-seq count matrices, which are typically sparse, i. e. contain mostly zeroes. Most scRNA-seq tools make use of this property by using compressed sparse matrix formats, which greatly reduce the disk space required to store sparse matrices (Wolf et al., 2018; Stuart et al., 2019). This compression is mainly achieved by storing only non-zero values, while zeroes are not explicitly stored. Instead, all matrix positions that were not stored are simply assumed to be zero.

To compress scBS data in a similar manner, I deviated from the abovementioned binary representation of the data (1, 0 and missing values), and instead represented methylated CpG sites as 1, unmethylated CpG sites as -1, and missing values as 0. Coding missing values as 0 and storing the matrix in a compressed sparse matrix format means that missing values are no longer explicitly stored, which decreases disk space requirements and makes it feasible to hold a typical scBS data

[1] Although note that this problem was addressed in a very recent publication by Fischer and Schulz (2023), who implemented a dedicated scBS mapper.

set in random-access memory (RAM). Since a common use-case of scBS data is to assess DNA methylation in genomic intervals, I used the compressed sparse row (CSR) format (Tinney and Walker, 1967; Buluç et al., 2009) that is optimized for row-wise access. Furthermore, since scBS data is usually processed chromosome by chromosome, I stored the data in separate matrices for each chromosome, which further lowers memory requirements.

To additionally make row-wise access to genomic sites more straightforward, I constructed the matrix in such a way that each row corresponds to a single genomic base pair, instead of a single CpG site. Genomic intervals are usually denoted by providing the base pair coordinates, and this adjustment means that e. g. the methylation state of CpG sites in the interval `1:5000-6000` can be retrieved by simply accessing rows 5000 to 6000 in the matrix of chromosome 1. While this adjustment increases the matrix dimensions and thus the required disk and memory space, this increase is modest since I only stored methylation values located in CpG dinucleotides. All non-CpG rows are thus populated with zeroes, which are not explicitly stored.

2.1.2 Quantification of methylation in genomic intervals

To obtain a cell × feature methylation matrix, one has to choose a measure of DNA methylation for a given genomic interval. The most obvious choice, commonly used in bulk WGBS, is to simply use the mean methylation of all CpG sites contained in the interval. However, due to the low sequencing coverage currently achievable for single cells, this measure of DNA methylation is problematic since intervals of intermediate size (ca. 1-10 kb) often contain just a single sequencing read, or no read at all. This means that region-wise methylation averages are not directly comparable across cells, because the average may greatly depend on the exact location of the sequencing read within the interval. Fig. 2.1 depicts an example of a genomic interval that is quantified in three single-cell samples.

In this example, two of the three cells obtained a sequencing read in the interval of interest. However, since these reads are not long enough to cover the entire interval, the methylation status of some CpG sites is unknown. In a naive calculation of the mean methylation, these un-observed sites would simply be ignored. The resulting values of 22% and 77% would indicate that methylation in this interval is very different between the two measured cells. However, upon closer inspection of the individual CpG sites, it becomes clear that there is no evidence for methylation differences between the two cells, since the CpG sites that were observed in both cells are in agreement. Thus, the observed methylation difference might simply be the result of stochastic variation of the read position.

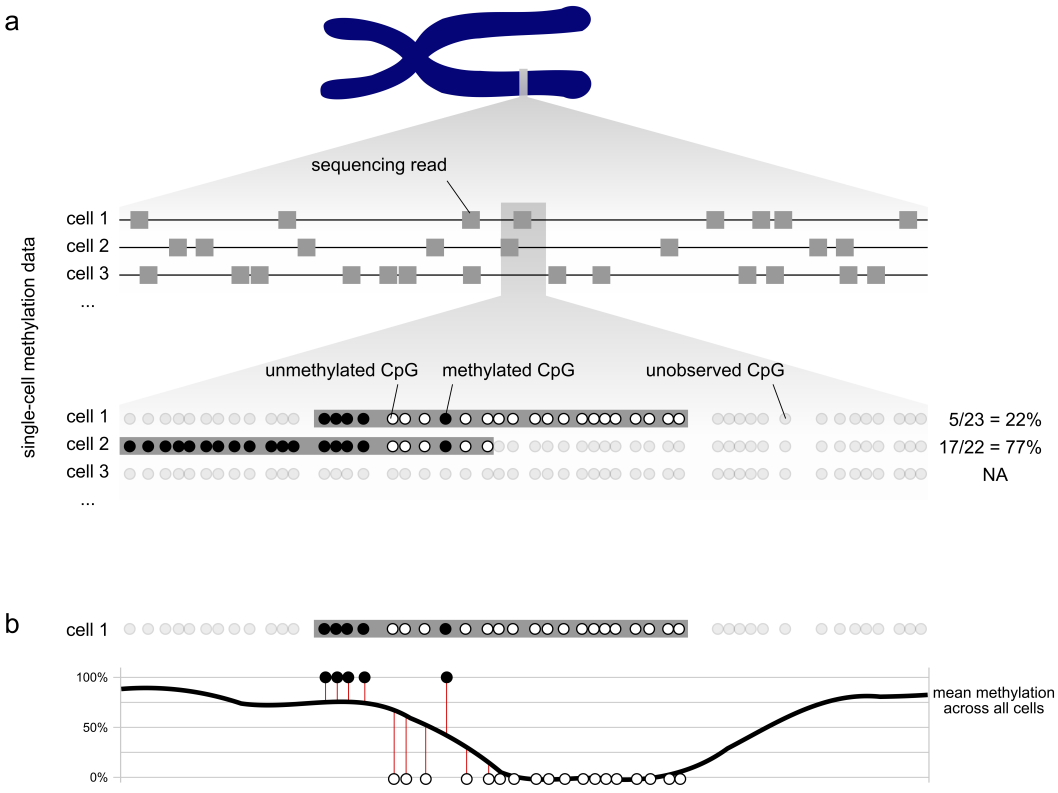

FIGURE 2.1: Improved quantification of DNA methylation in genomic intervals.
(a) I depicted a genomic interval along a chromosome, for which DNA methylation is to be quantified. Two cells cover different parts of this interval with one sequencing read each. If I simply count the fraction of methylated CpG sites for each cell, ignoring missing values, I obtain very different values for the two cells. **(b)** To obtain the thick black average methylation curve, I first average each CpG site's methylation over all cells and apply smoothing. I then calculate a relative methylation value, the shrunken mean of the residuals, as illustrated here for cell 1 from **a**. This methylation value is the shrunken mean of the cell's signed residuals to the smoothed average methylation curve, i. e. the lengths of the vertical red lines are averaged, counting the residuals as positive values for methylated CpGs and negative for unmethylated CpGs. Figure and caption modified from Kremer et al. (2022a).

To combat this source of noise, I devised a quantification method that takes into account the expected methylation value of each observed CpG site. In the depicted example, the ensemble average of all cells ("pseudobulk") suggests that the center of the genomic interval is generally lowly methylated, while the flanks are more highly methylated. It is thus not surprising that the mean methylation of cell 1 is lower, since its read mapped to the center of the region. To reduce the influence of read position, for each CpG, I subtract the expected methylation value from the binary methylation value that was measured. The resulting CpG-wise residuals, indicated by the red vertical lines in Fig. 2.1b, are then averaged to obtain an estimate of methylation in the interval.

Another source of noise is the sequencing coverage itself, i. e. the number of read-covered CpG sites for which the methylation status was measured. It is clear that methylation estimates are unreliable in the case where only few CpG sites were measured in a given cell. For example, the average methylation of cells with only

one observed CpG site in a given interval will always be 0% or 100%. To prevent these extreme, and thus unlikely, estimates for low-coverage intervals, I furthermore employ shrinkage when calculating the mean of the residuals by adding a pseudo-count to the denominator. This has almost no effect on intervals with high coverage, but pushes the mean towards zero, i. e. the ensemble average, for low-coverage intervals.

2.1.3 Detecting variably methylated regions

Previous studies typically quantify DNA methylation in subsequent genomic bins of e. g. 100 kb (see section 1.5.4 and Fig. 1.10). However, this strategy is unlikely to be optimal since large genomic bins may span both informative and uninformative genomic regions, in which case the informative signal may be diluted. Thus, I devised and implemented an approach to scan scBS data of the entire genome for informative genomic regions (Fig. 2.2).

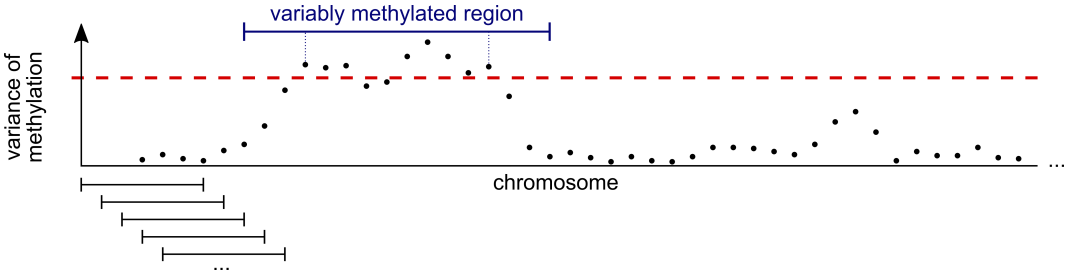

FIGURE 2.2: Finding variably methylated regions (VMRs).
I first divide each chromosome into overlapping windows (intervals at the bottom left). I then calculate the methylation value of each interval (the shrunken mean of the residuals as described in section 2.1.2 and Fig. 2.1b). I detect variably methylated regions (VMRs) by calculating the variance of these methylation values for each window and setting a variance threshold (dashed line). I then merge all intersecting or adjacent windows above this treshold. The obtained merged intervals are labeled VMRs. Figure and caption modified from Kremer et al. (2022a).

Whether a region is informative is determined based on the variance of the cell-wise methylation estimates, calculated as described in section 2.1.2. A high variance denotes that the region is heterogeneously methylated, i. e. some cells are highly methylated while others are lowly methylated, which means that the region is potentially useful to distinguish cell types or other groups of cells. A similar strategy is used in scRNA-seq, where genes with high variance of gene expression, termed highly variable genes, are selected for PCA (see section 1.5.4 and Amezquita et al., 2020).

To scan the entire genome while allowing for flexible start and end positions of detected regions, I first divide each chromosome into overlapping windows of a fixed width (`--bandwidth` parameter, default: 2 kb). These overlapping windows are shifted by a fixed step size (`--stepsize` parameter, default: 10 bp). Once all variances of the largest chromosome are calculated, they are used to determine a

variance threshold so that a quantile of the variances is greater than this threshold (`--var-threshold` parameter, default: 0.02). This threshold is then applied to all chromosomes. Overlapping windows above the threshold are subsequently merged, yielding the variably methylated regions (VMRs). This merging procedure is able to detect VMRs of variable size, since each VMR consists of one or more windows.

2.2 scbs, a software for the analysis of scBS data

To make my newly developed scBS methods available to the scientific community, I implemented them in a Python package called `scbs`. Many algorithms in this package are computationally expensive since they perform operations on the entire genome of potentially thousands of cells. To achieve fast execution times, I avoided conversion to dense matrices and instead implemented custom algorithms that use the sparse matrix format described in 2.1.1. Additionally, I used Numba (Lam et al., 2015) to translate crucial inner-loop code sections to fast machine code, and parallelized computationally expensive steps such as the sliding window approach described in 2.1.3.

2.2.1 List of subcommands implemented in the scbs package

Once installed from the Python package index, the methods implemented in the `scbs` package are accessible via a command line interface that consists of the following subcommands (Fig. 2.3):

`scbs` displays the version number and briefly explains all available subcommands.

`scbs prepare` collects and stores scBS data in a compressed, sparse matrix format as described in section 2.1.1.

`scbs filter` discards low-quality cells according to user-specified thresholds.

`scbs smooth` aggregates scBS data from all cells in the data set and calculates their smoothed ensemble average along the whole genome.

`scbs scan` scans the whole genome for VMRs as described earlier in section 2.1.3.

`scbs matrix` generates a methylation matrix, reporting both the methylation average as well as the shrunken mean of residuals as described in section 2.1.2.

`scbs profile` calculates the average methylation of all cells around a set of user-specified genomic intervals as showcased later in Fig. 3.3.

`scbs diff` detects differentially methylated regions (DMRs) between two user-defined groups of cells. This method was developed and implemented by Martina Braun under my supervision and is described in Braun (2023).

The command line interface offers a brief explanation of all required input files, generated output files, and parameters of these subcommands. The source code is hosted at GitHub [2]. This repository also contains more documentation, as well as a tutorial that showcases all subcommands on a small example data set.

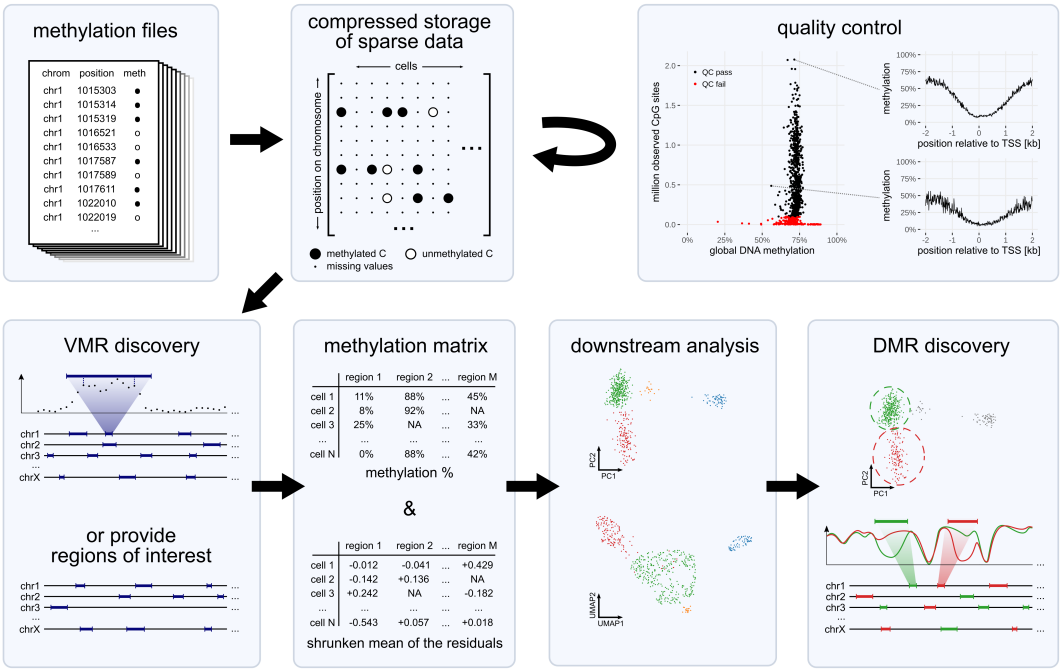

FIGURE 2.3: Overview of the functionalities implemented in the scbs package.
`scbs prepare` parses methylation files produced by common bisulfite sequencing mappers and stores their contents in a compressed format optimized for efficient access to genomic intervals. `scbs` also produces cell-wise summary statistics and quality plots (here: average methylation around the transcription start site) that are used to detect low-quality cells. These cells can be discarded with `scbs filter`. To obtain a methylation matrix, similar to the count matrices used in scRNA-seq, the user must first decide in which genomic intervals methylation is to be quantified. The user may either provide genomic regions of *a priori* interest, or they may choose to discover VMRs (variably methylated regions) in the data with `scbs scan`. The resulting methylation matrix can then be used for downstream analysis such as cell clustering and dimensionality reduction. Figure and caption taken from Kremer et al. (2022a).

2.3 Application and benchmarks

To demonstrate the value of the methods implemented in `scbs`, I analyzed two data sets with `scbs`: The first data set is our own single-cell triple-omic data set that comprises single-cell methylomes, transcriptomes, and chromatin accessibility data of 1566 cells isolated from murine forebrains (Kremer et al., 2022b). Note that this data set is discussed and interpreted in much greater detail in chapter 3; here I merely use it as a means to assess the performance of my analysis methods. Detailed author attributions on the generation of this data set are reported in External Contributions. The second data set, previously published by Luo et al. (2017), consists of 3377 single-cell CpG methylomes from the murine frontal cortex, comprising 16 neuronal sub-types.

[2] https://github.com/LKremer/scbs

2.3.1 scbs improves identification of cell types

I first tested whether my newly developed methods improve the ability to distinguish cell types and cell states (Fig. 2.4). To this end, I obtained cell type/cell state labels based on the single-cell transcriptomic portion of our data, using Leiden clustering (Traag et al., 2019) and manual annotation of the obtained clusters (Fig. 2.4a).

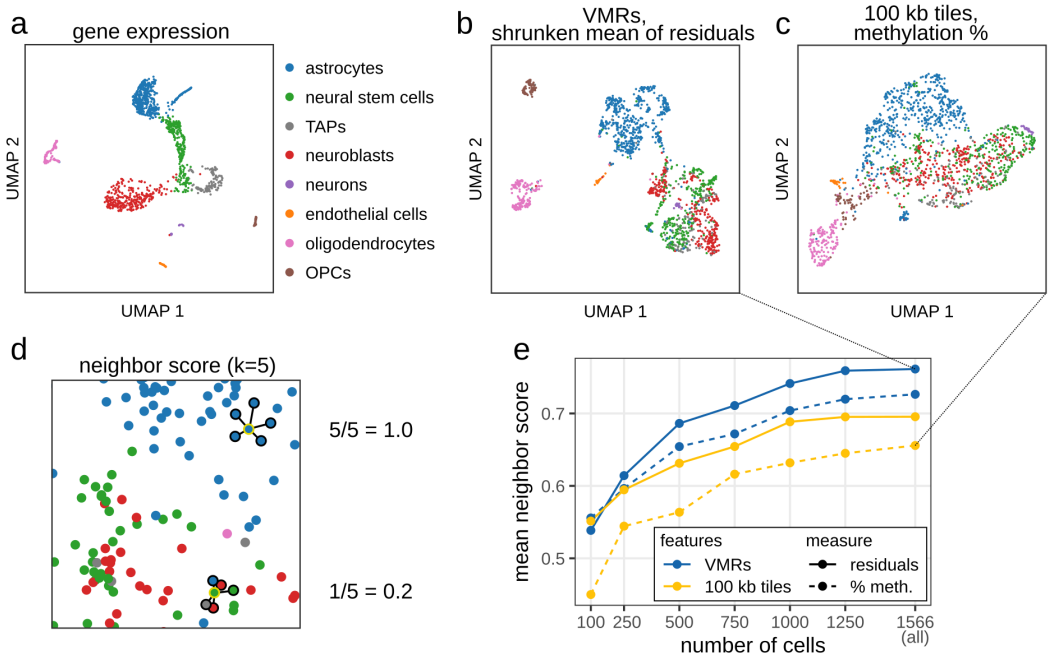

FIGURE 2.4: Benchmark of my methods on single-cell multi-omic data of murine forebrain cells.
(a) Cell labels based on clustering of single-cell transcriptomes of our own single-cell triple-omic data set (reported in Kremer et al. (2022b) and discussed later in section 3). **(b, c)** Exemplary UMAPs obtained when analyzing the data set with my proposed methods **(b)** or conventional methods based on genome tiling **(c)**. **(d)** Illustration of the neighbor score. Note that distances are measured in 15-dimensional PC space, not in the two dimensional UMAP space. **(e)** Mean neighbor score obtained after analyzing single-cell methylomes with different combinations of methods. I subjeted either 100 kb genomic tiles or VMRs to PCA and UMAP. I pre-processed tiles or VMRs as proposed earlier in this section (shrunken mean of residuals subjected to PCA) or using conventional methods (PCA on methylation percentages, lightly imputed as done in Luo et al., 2017). I sub-sampled the full 1566-cell data set to simulate smaller data sets (x axis). A higher neighbor score implies better separation of cell types inferred from single-cell transcriptomes of the same cells **(a)**. This figure is based on a previous iteration of this analysis, reported in Kremer et al. (2022a).

I considered the cell labels derived from the transcriptomic data as ground truth and tested whether I am able to distinguish the same groups of cells based on their methylomes. To do this, I subjected the corresponding single-cell methylomes to four variants of the scBS analysis workflow described earlier: I defined genomic intervals in two ways and quantified methylation in these intervals in two ways, then always subjected the resulting matrix to PCA. To define the intervals, I either simply tiled the genome into 100 kb bins as is commonly done (e. g. Luo et al., 2017), or I used my VMR detection approach. To quantify methylation in these intervals, I either

used the current default strategy (averaging to obtain methylation percentages as depicted in Fig. 1.10), or I used my new strategy of calculating the shrunken mean of the residuals as depicted in Fig. 2.1.

Visual inspection of the resulting UMAPs revealed that my proposed combination of methods results in more clearly separated cell types, compared to a UMAP obtained with default analysis methods (Fig. 2.4b,c). While all cells form a continuous point cloud when using default methods, my improvements led to a clear separation of oligodendrocytes, OPCs and endothelial cells. Furthermore, even cellular sub-states of cells in the continuous neural stem cell lineage were partially separated.

To quantify this performance gain in a more rigorous manner, I used a score that quantifies whether cells were placed, in 15-dimensional PC space, in a neighborhood comprising cells of the same cell type ("neighbor score", see Fig. 2.4d). Since scBS protocols are costly and labor-intensive, I first asked how the neighbor score depends on the total number of cells that were sequenced. To simulate smaller data sets, I sub-sampled the 1566-cell data set into smaller data sets, analyzed them again with the four workflow variants, and calculated the mean neighbor score for each. My results, depicted in Fig. 2.4e, demonstrate that the forebrain cell clusters are more cleanly separated when using my proposed methods. This performance gain was observed also in smaller data sets, although cell cluster separation generally becomes more difficult in those. Both the use of VMRs instead of genomic tiles, as well as the use of the shrunken mean of the residuals over methylation averages improved cell cluster separation.

2.3.2 scbs outperforms default methods on an external data set

I repeated this benchmark on the data set published by Luo et al. (2017), which consists of 3377 single-cell CpG methylomes from the murine frontal cortex, comprising 16 neuronal sub-types. The authors annotated cell types based on non-CpG methylation instead of CpG-methylation. Using these labels as ground truth, I re-analyzed the full and sub-sampled CpG methylation data with the four methods combinations and found that my suggested methods consistently outperformed default methods (Fig. 2.5). Note that the benchmarks depicted in Fig. 2.4 and Fig. 2.5 are inspired by a previous iteration of this analysis conducted by Leonie Küchenhoff under my supervision. This previous analysis, which I modified and repeated here, is reported in Kremer et al. (2022a).

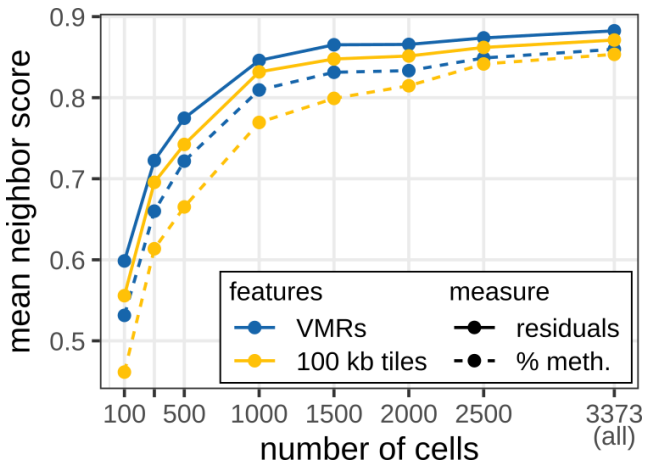

FIGURE 2.5: Benchmark of my methods on single-cell methylomes of neural sub-types of the murine frontal cortex.
I benchmarked my methods on single-cell methylomes from Luo et al. (2017) as described in Fig. 2.4. The cell type labels which act as ground truth were reported by Luo et al. and are based on non-CpG methylation and not CpG methylation.

2.3.3 scbs performs robustly across a wide range of VMR detection parameters

Lastly, I assessed whether my proposed workflow requires fine-tuning of parameters (Fig. 2.6). To this end, I re-analyzed both data sets, as well as sub-samples of the data with different VMR detection parameters, namely the width of the sliding window in bp, (set with option `--bandwidth` in the `scbs` software, default 2000 bp) and the variance threshold above which windows are merged to VMRs (`--var-threshold`, default 0.02). This parameter sweep showed that my workflow gives good results over a wide range of parameter values. For the data of Luo et al. (2017), results are nearly independent of the parameters (Fig. 2.6b). In our own more challenging data set (Fig. 2.6a), cell types were less cleanly separated when I selected very large bandwidths, or very strict variance thresholds. However, very small bandwidths or very lenient thresholds resulted in a much higher number of VMRs and thus long computing times. Overall, however, my default parameter combination provided good results and fast compute times in both data sets.

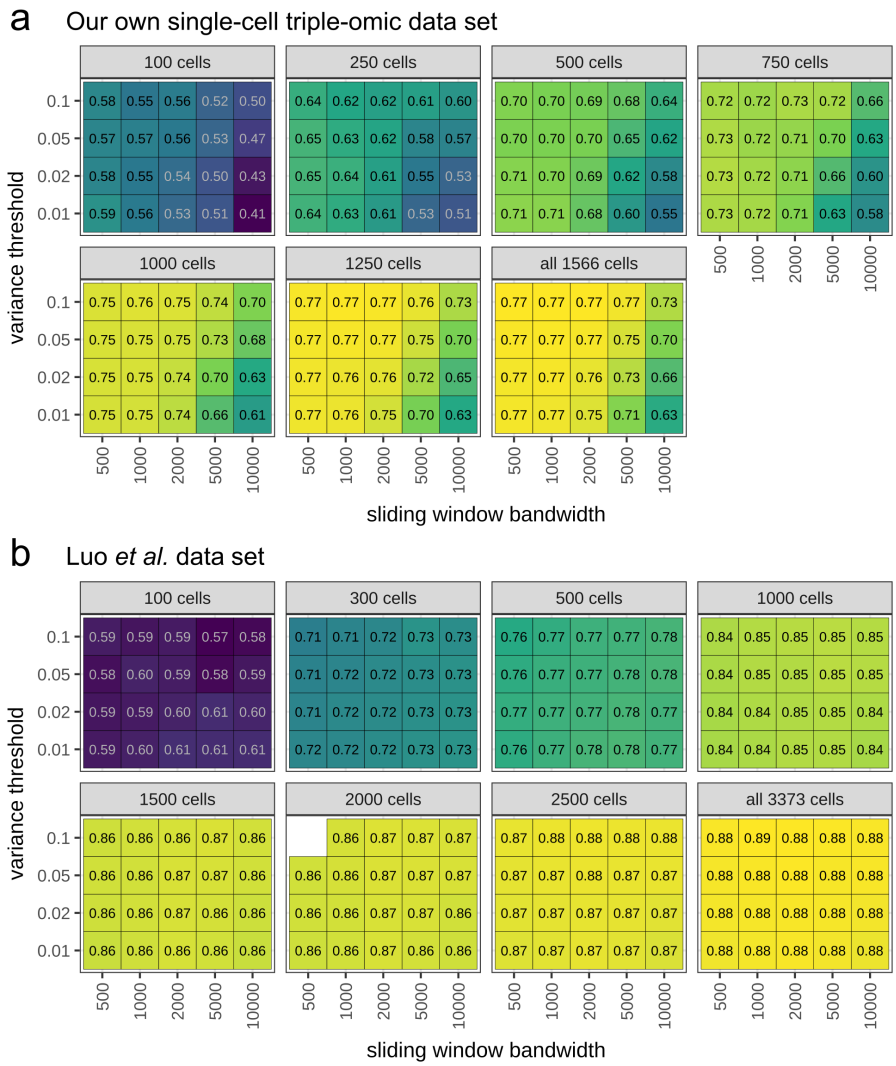

FIGURE 2.6: Effect of VMR detection parameters on separation of cell types.
Mean neighbor scores obtained after analyzing single-cell methylomes from Kremer et al. (2022b) **(a)** or Luo et al. (2017) **(b)** with my proposed methods. I detected VMRs with `scbs scan` using various sliding window bandwidths or variance thresholds, and on sub-samples of the full data sets. Note that one parameter combination did not yield any results since the resulting methylation matrix did not fit into RAM.

Chapter 3

Characterization of methylation dynamics in the adult vSVZ

Single-cell sequencing enabled characterization of molecular changes in the adult NSC lineage at unprecedented resolution. Due to the rise and commercial availability of scRNA-seq methods specifically (Ding et al., 2020), to date most single-cell studies on vSVZ neurogenesis focus on gene expression (e. g., Llorens-Bobadilla et al., 2015; Zywitza et al., 2018; Kalamakis et al., 2019; Cebrian-Silla et al., 2021). As a result, transcriptional changes occurring along the NSC-to-neuroblast differentiation trajectory are well-characterized (see section 1.1.5). However, knowledge on other molecular features such as potential changes in epigenetic marks, is currently lagging behind.

3.1 Single-cell multi-omics of the healthy vSVZ NSC lineage

To characterize epigenetic changes in the adult NSC lineage, I teamed up with Dr. Santiago Cerrizuela, a postdoctoral researcher in the Martin-Villalba group. Together with Prof. Martin-Villalba and Prof. Anders, we agreed to employ scNMT-seq (Clark et al., 2018, summarized in section 1.5.3) to simultaneously quantify DNA methylation, chromatin accessibility, and gene expression at single-cell resolution. All data presented in this section were generated in experiments either performed by Dr. Cerrizuela, or by assistants under his supervision (see External Contributions for details).

Dr. Cerrizuela and assistants dissected the vSVZ and olfactory bulb of several adult male mice to isolate cells from various points of the adult NSC lineage (Fig. 3.1a). Detailed information on the mice that they used in these experiments is available in the supplementary material of Kremer et al. (2022b). Furthermore, Dr. Cerrizuela and assistants isolated GLAST$^+$ astrocytes from the striatum, in order to contrast molecular features of these common astrocytes with those of adult NSCs, the specialized astrocytes residing in the vSVZ.

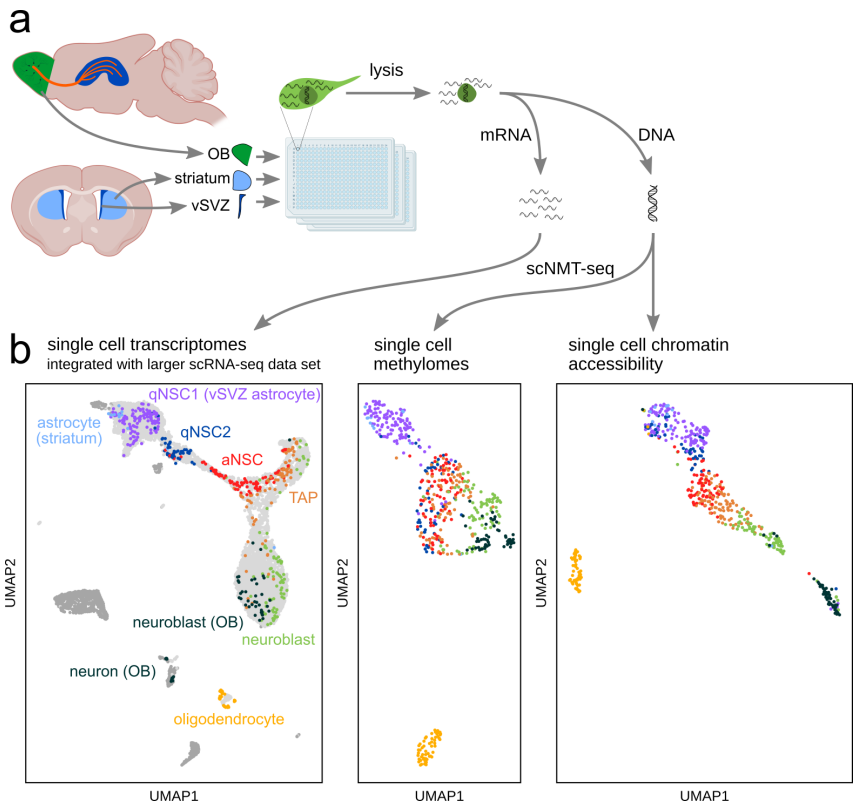

FIGURE 3.1: Single-cell triple-omics of the adult NSC lineage.
(a) Schematic of the workflow performed by Dr. Santiago Cerrizuela and colleagues to obtain scNMT-seq data from three brain regions. **(b)** UMAPs that I generated for each molecular layer of this data set. For the transcriptomics data, I integrated the cells with the larger data set of Carvajal Ibañez et al. (2023), which consists of cells from the vSVZ (light gray) and olfactory bulb (OB, dark gray). Figure and caption modified from Kremer et al. (2022b).

3.1.1 Quality assessment of scNMT-seq data

To assess the quality of the resulting scNMT-seq data set, I computed several quality metrics for each cell, including the number of detected genes and the number of observed CpG sites (Fig. 3.2). Furthermore, I used `scbs profile` to visualize DNA methylation and chromatin accessibility around transcription start sites (TSSs), which are known to be lowly methylated and highly accessible compared to their surroundings (Fig. 3.3). After filtering off-target cells and cells with unsatisfactory quality metrics or TSS profiles (see Extended Methods M.2), I obtained a triple-omic data set consisting of 540 high-quality cells obtained from healthy wild type mice (Fig. 3.1b).

On average, I detected expression of 5188 genes (Fig. 3.2). This quality metric greatly exceeds those typically achieved by droplet-based scRNA-seq protocols, possibly since they employed the recently developed Smart-seq3 (Hagemann-Jensen et al., 2020) and since they sequenced more deeply than usual. The average single-cell epigenome contained information on 678 186 CpG sites, which corresponds to ca. 5% of CpG sites in the mouse genome, a value that is expected for scNMT-seq (Clark et al., 2018).

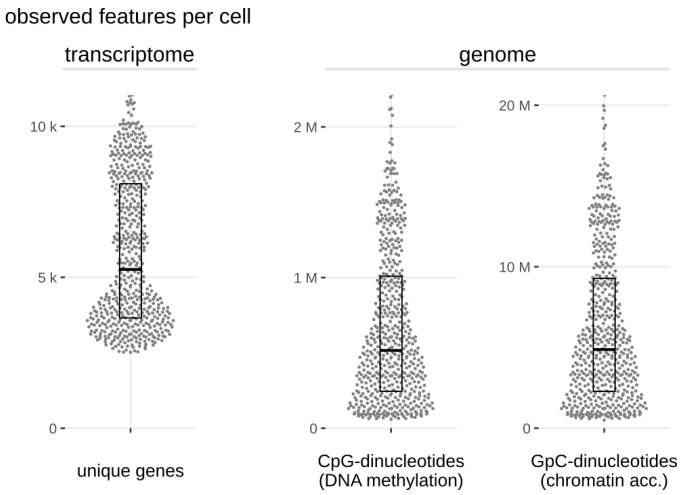

FIGURE 3.2: Quality metrics of cells retained after quality filtering.
scNMT-seq quality metrics which I computed for all cells that passed quality filtering. "Unique genes" refers to the number of genes with at least one detected sequencing read per cell. "CpG/GpC-dinucleotides" refers to the number of methylation sites with sequencing coverage and thus known methylation status. Boxes denote the interquartile range; center lines denote the median. Figure and caption taken from Kremer et al. (2022b).

Usage of `scbs profile` demonstrated that the vast majority of cells showed low DNA methylation and high chromatin accessibility at TSSs as expected (section 1.4.2). An example TSS profile of a single high-quality neuroblast is depicted in Fig. 3.3. To further probe the resolution and quality of our single-cell epigenomes, I also visualized methylation and accessibility around CTCF-binding sites. The resulting graphs revealed oscillations of both methylation and chromatin accessibility, which most likely correspond to regularly spaced depositions of nucleosomes as previously described (Teif et al., 2014). Furthermore, CTCF-binding sites show a distinct accessibility peak with a sharp dip in its center. The latter dip may correspond to chromatin that is rendered inaccessible due to deposition of the CTCF protein itself. Plots such as the ones depicted in Fig. 3.3 are typically used to assess the quality of bulk epigenomic data. The level of detail visible in this data, at the single-cell level, demonstrates that this epigenomic data is of high quality.

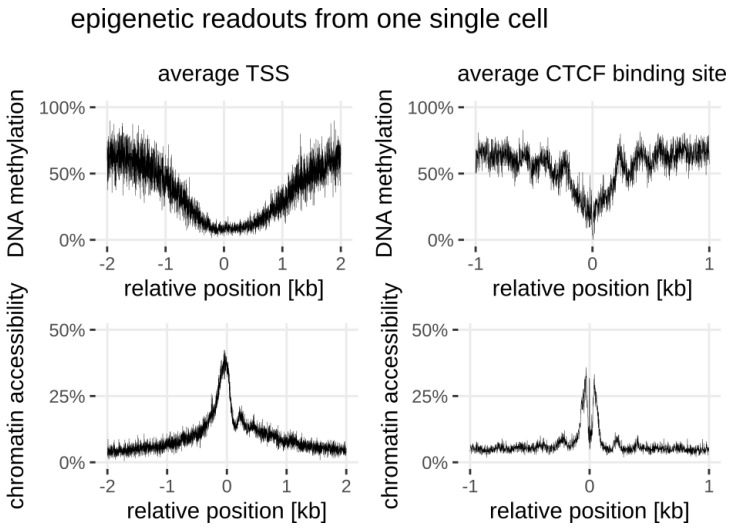

FIGURE 3.3: Transcription start sites are lowly methylated and highly accessible. Average methylation and chromatin accessibility levels around transcription start sites (TSSs) and CTCF-binding sites of a single neuroblast. Figure and caption taken from Kremer et al. (2022b).

3.1.2 Annotation and exploration of scNMT-seq data

By integrating the transcriptomic portion of our data with a larger scRNA-seq data set of the vSVZ and olfactory bulb (Carvajal Ibañez et al., 2023), I confirmed that we had captured cells of the NSC lineage. In order to annotate cell types and cell states, I first ordered cells according to their position in the NSC differentiation trajectory. To do this, I combined MOFA+ (Argelaguet et al., 2020) with slingshot (Street et al., 2018) to compute pseudotime values that incorporate information from all three molecular layers. I then assessed gene expression of known lineage marker genes and transcription factors (TFs). The observed expression dynamics are in agreement with previous data sets of the adult NSC lineage (Llorens-Bobadilla et al., 2015; Kalamakis et al., 2019) and allowed me to assign cell types and cell states (Fig. 3.4a-b).

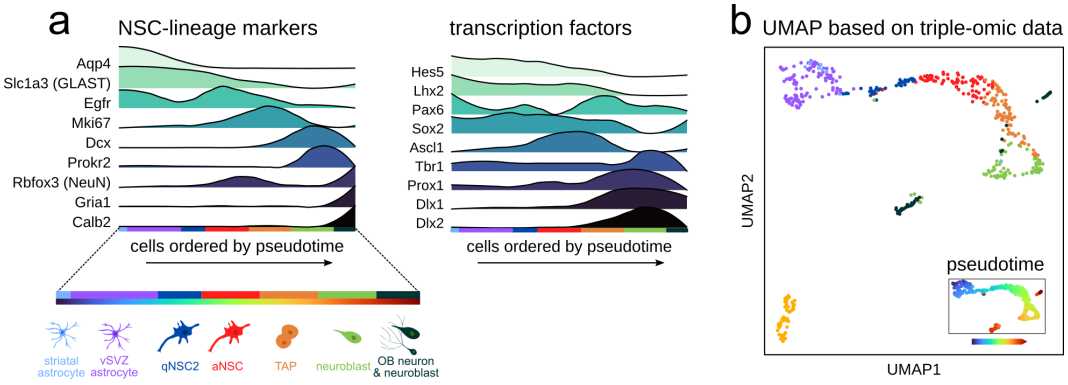

FIGURE 3.4: Expression of marker genes and transcription factors in pseudotime. **(a)** Gene expression of key marker genes and transcription factors along pseudotime. **(b)** UMAP and pseudotime assignments (small inset plot) based on omics data from all three molecular layers. I colored cells by their inferred cell type or cell state, based on expression of marker genes in **a**. Figure and caption taken from Kremer et al. (2022b).

To enable analysis of the DNA methylation and chromatin accessibility data sets, I used `scbs` to employ the newly developed computational methods described in chapter 2. In brief, I first stored both epigenomic data sets and then used `scbs scan` to identify VMRs in the CpG data set and variably accessible regions (VARs) in the GpC data set. I then quantified CpG methylation at VMRs and GpC methylation (chromatin accessibility) at VARs, and separately performed PCA and UMAP on each data set. This enabled two-dimensional representation of the two epigenomic data sets (Fig. 3.1b).

The resulting UMAPs visually resemble the UMAP of the scRNA-seq data, indicating that cell types and cell states of the vSVZ and olfactory bulb can be distinguished not only by gene expression, but also by epigenetic information. In contrast to scRNA-seq, however, the DNA methylation and chromatin accessibility UMAPs did not show a separation of cycling cells (late aNSCs, TAPs, early neuroblasts) from the rest of the NSC lineage. This may indicate that cycling cells can only be detected by gene expression, and not by epigenomic methods. A possible explanation is that expression of cell cycle-related genes may not be not regulated by changes in DNA methylation and chromatin accessibility.

Furthermore, the UMAP of single-cell methylomes displayed a striking separation of striatal astrocytes and qNSC1 cells on one hand, and qNSC2 cells, aNSCs, TAPs, and neuroblasts on the other hand. This observation suggests the existence of strong epigenetic differences between these two groups of cells. In all three UMAPs, oligodendrocytes are clearly separated from the remaining cells, suggesting that they possess gene expression and epigenomic profiles distinct from the neurogenic vSVZ lineage.

3.1.3 VMR methylation anticorrelates with gene expression

One advantage of single-cell multi-omic data is that relationships between the measured omics-layers can be readily studied since multiple readouts are obtained from each cell. This makes it possible to identify candidate genomic regions that may be involved in regulating gene expression. For instance, if a gene is strongly expressed in some cells but not in others, one may speculate that expression of this gene is regulated by methylation of its promoter. This hypothesis can be tested in scNMT-seq data by correlating gene expression with promoter methylation, i. e. by checking whether those cells that are strongly expressing the gene also possess lower methylation of its promoter. To systematically identify genes regulated by promoter methylation in the adult brain, I thus correlated, for each gene, promoter methylation with gene expression across all cells (Fig. 3.5a).

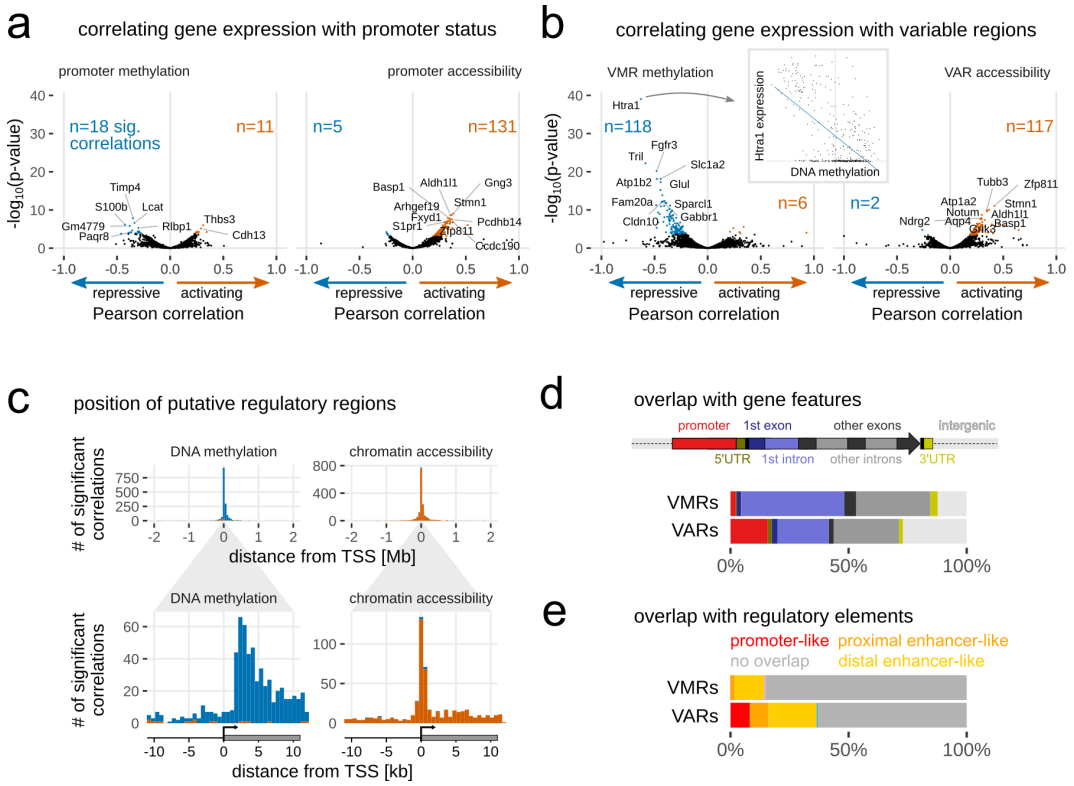

FIGURE 3.5: Correlating epigenetic features with gene expression.
(a) Correlation of promoter methylation (left) and promoter accessibility (right) with gene expression. Negative correlations (blue) indicate a repressive effect of methylation/accessibility while positive correlations (orange) indicate activation. n: number of significant correlations after Benjamini-Hochberg adjustment. **(b)** Correlation of VMR (variably methylated region) methylation and VAR (variably accessible region) accessibility with expression of the nearest gene. **(c)** Distance histogram of all significant correlations between gene expression and all VMRs and VARs within 2 Mb of the TSS. Negative correlations are orange, positive correlations are blue. The bottom panel is a zoom-in of the top panel. **(d, e)** Overlap of significantly correlating VMRs and VARs with gene features **(d)** and candidate cis-regulatory elements from ENCODE **(e)**. Figure and caption modified from Kremer et al. (2022b).

The obtained correlation coefficients are overall modest and only a small number of genes shows a significant correlation between promoter methylation and gene expression. This data demonstrates that the gene expression differences observed between cell types and cell states of the vSVZ and olfactory bulb are not the result of differences in promoter methylation. In contrast, correlating chromatin accessibility of promoters with gene expression resulted in a much higher number of significant correlations, indicating that promoter accessibility is one mechanism that affects gene expression. Most of these correlations were positive, in agreement with the established view that accessible promoters are more easily accessed by TFs and thus more readily transcribed (Felsenfeld et al., 1996; Pliner et al., 2018).

Since I only identified few significantly correlating promoters, I next asked whether there might be other genomic regions whose methylation status is associated with gene expression. To test whether VMRs and VARs identified with `scbs scan` are more strongly associated with gene expression, I next correlated their methylation and chromatin accessibility estimates with gene expression. This correlation analysis

requires matching of VMRs (or VARs) with genes. However, in contrast to the promoter-gene-matches depicted in Fig. 3.5a, it is not immediately obvious how VMRs and genes should be matched, since VMRs might occur in intergenic regions, or since one gene may intersect multiple VMRs. To nonetheless enable a correlation analysis similar to Fig. 3.5a, I thus matched every gene with the VMR closest to its TSS.

Correlating VMR methylation with expression of the nearest gene resulted in more significant correlations and more extreme correlation coefficients (Fig. 3.5b), indicating that `scbs scan` may be able to identify regulatory elements that are more predictive of gene expression than promoter methylation. In contrast, correlating VARs with gene expression yielded results similar to those obtained when correlating promoter accessibility. This suggests that, in the case of chromatin accessibility data, there is little or no additional benefit of quantifying VARs instead of promoters. Of note, most significant VMR-to-gene correlations were negative, while most VAR-to-gene correlations were positive. This observation is in line with the common notion that DNA methylation is repressive (e. g. Mattei et al., 2022, also see section 1.4.2) while accessible chromatin promotes gene expression (Felsenfeld et al., 1996; Pliner et al., 2018).

3.1.4 Methylation of the first intron may contribute to gene regulation

To investigate where putative regulatory regions are located relative to the position of their potential target genes, I disregarded VMR-to-gene matches and instead correlated expression of each gene with *all* VMRs and VARs within 2 Mb of the TSS (Fig. 3.5c). This revealed that most significant correlations were located in the center of the $\pm$2 Mb region, close to the TSS. Zooming into this region by limiting the view to $\pm$10 kb showed that most significant correlations do not occur at the TSS itself, but rather downstream of the TSS, with a peak at approximately +2 kb. In contrast, most significantly correlating VARs are located at the TSS. This finding confirms that the current scATAC-seq practice of using TSS accessibility to predict gene expression (Pliner et al., 2018) is substantiated.

The abundance of relevant VMRs just downstream of the TSS suggests that most of these VMRs are located in the gene body. By intersecting VMRs with gene features, I showed that almost half of them are located in the first intron (Fig. 3.5d). This suggests that methylation of the first intron may contribute to gene regulation. In comparison to VMRs, VARs are more frequently located in promoters, in agreement with their aggregation near the TSS (Fig. 3.5c). Finally, to understand whether VMRs and VARs correspond to known regulatory elements, I intersected their genomic coordinates with those of candidate cis-regulatory elements (cCREs) reported by the ENCODE consortium (Moore et al., 2020) (Fig. 3.5e). Most VMRs and VARs did not intersect with known cCREs, however. This either means that they correspond to regulatory elements that are not annotated by ENCODE, or that methylation at these

sites affects gene expression in a manner that is independent of classical promoters or enhancers. Another possible explanation is that the observed VMRs are merely correlating with gene expression, but not causally influencing gene expression.

3.1.5 Epigenetic changes in pseudotime

Next, I utilized the newly generated triple-omic data set to assess whether epigenetic changes occur along the NSC differentiation lineage. For reference, I first visualized gene expression changes along the lineage using a correlation heatmap (Fig. 3.6). This visualization illustrates that the transcriptomes of striatal astrocytes, qNSC1 cells, and qNSC2 cells are very similar, in agreement with the notion that adult NSCs are specialized astrocytes that possess an astrocyte transcriptome (Doetsch et al., 1999, section 1.1.4). This transcriptomic state changes upon NSC activation, i. e. the transition between qNSC2 and aNSC. Overall, my visualizations of transcriptomic changes along the NSC lineage (Fig. 3.6, Fig. 3.4) are in agreement with previous scRNA-seq studies of the vSVZ (Kalamakis et al., 2019, section 1.1.5).

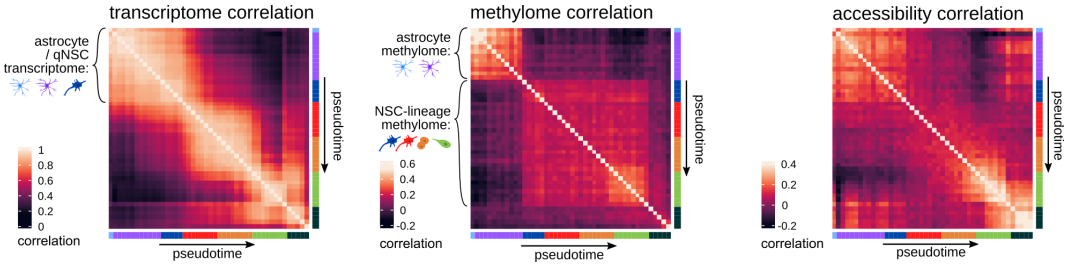

FIGURE 3.6: Correlation matrices of gene expression, DNA methylation and chromatin accessibility.
I ordered and binned cells according to pseudotime, with each bin containing 10 cells on average. Bright squares indicate blocks of cells that are transcriptionally (or epigenetically) homogeneous. Transitions between these blocks indicate molecular changes that occur upon NSC differentiation. Figure and caption modified from Kremer et al. (2022b).

Next, I used the same approach to visualize DNA methylation changes along the lineage. This resulted in a very different pattern, characterized by two large blocks that represent two groups of cells with similar methylomes. In stark contrast to the transcriptome data, here the first block consists of just striatal astrocytes and qNSC1 cells, indicating that these two populations possess a highly similar methylome. Despite their transcriptomic similarity to qNSC1 cells, qNSC2 cells are found in the second methylome block that also contains aNSCs, TAPs and neuroblasts. For these reason, I propose that qNSC1 cells correspond to common parenchymal astrocytes of the vSVZ, while qNSC2 cells are *bona fide* vSVZ qNSCs that regularly contribute to neurogenesis (see sections 3.2, 4.2.2 and 4.3 for a deeper exploration of this idea). Thus, I will henceforth refer to qNSC1 cells as vSVZ astrocytes.

Shortly before the transition from TAPs to neuroblasts, a third bright square is visible, suggesting that additional methylation change occurs at this point in pseudotime (explored in greater detail in section 3.1.6).

Lastly, I also generated a correlation heatmap based on the chromatin accessibility data. Compared to the transcriptome or methylome heatmaps, the transitions between blocks of correlating cells are more blurry in the accessibility heatmap. This would suggest that accessibility changes in the lineage occur more gradually than the relatively sharp changes observed in the two other modalities. However, there are also more noticeable dark bands of weakly-correlating pseudotime bins, suggesting that the accessibility data is overall more noisy. Importantly, the first block of correlating cells comprises striatal astrocytes, as well as qNSC1 and qNSC2 cells. The clear distinction between qNSC1 and qNSC2 that I previously observed in the methylome data is not visible.

3.1.6 Two waves of methylation change occur in the adult NSC lineage

Due to the intriguing pattern of methylome changes observed in Fig. 3.6, I decided to quantify methylation and demethylation events in the NSC lineage. For each VMR, I thus fitted a step function to the methylation values as a function of pseudotime. I then determined those VMRs that clearly show a rapid increase or decrease of methylation and simultaneously inferred the most likely change point in pseudotime (see Extended Methods M.7 for details). In agreement with the patterns observed in Fig. 3.6, a histogram of these change points (Fig. 3.7a) shows that most methylation or demethylation events occur in one of three waves.

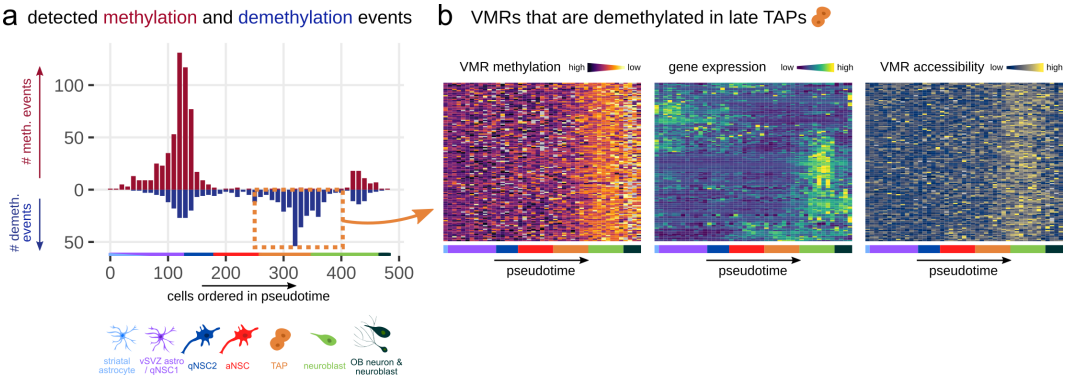

FIGURE 3.7: DNA methylation changes in pseudotime.
(a) Histogram of inferred pseudotime points at which VMRs become methylated (red, up) or demethylated (blue, down). **(b)** Expression and epigenetic state of genes intersecting with VMRs that become demethylated in late TAPs (dashed orange box in **a**). I ordered the rows by hierarchical clustering on gene expression values. Note that most of these genes are expressed in neuroblasts. Figure and caption modified from Kremer et al. (2022b).

The first wave of methylation change corresponds to the transition between qNSC1 and qNSC2. It is characterized by both methylation and demethylation events, although methylation events were observed more frequently. The second wave occurs at the late TAP stage and consists almost exclusively of demethylation events. Lastly, towards the end of pseudotime in late neuroblasts, both methylation and demethylation events are observed, which may correspond to the implementation of neuron-specific methylomes as previously described (e. g. Luo et al., 2017). It is

worth noting, however, that inferred (de)methylation events of this third wave are likely less reliable, since their prediction depends on the methylomes of very few olfactory bulb neurons. These neurons furthermore showed below average quality metrics, and the methylation changes of this third wave are not clearly visible in Fig. 3.6. For these reasons, I decided to focus on the two other (de)methylation waves.

I will first discuss epigenomic changes occurring in the second wave, at the late TAP stage. Visualization of VMRs that change at this point in pseudotime demonstrated that their methylation indeed decreases at the late TAP stage (Fig. 3.7b). To estimate whether these methylation changes are accompanied by changes in gene expression, I furthermore visualized the expression of genes intersecting with these VMRs. Ordering these genes by hierarchical clustering showed that roughly half of these intersecting genes become transcriptionally active at the neuroblast stage, i. e. shortly after the demethylation wave (Fig. 3.7b). This suggests that demethylation may pave the way for the expression of genes required at the neuroblast stage.

However, it is worth noting that not all genes follow this expression pattern. Among the intersecting genes are also several genes that are predominantly expressed in qNSC1 cells (vSVZ astrocytes), or in aNSCs and later stages. A possible explanation is that the heuristic method I used to match genes and VMRs resulted in incorrect matches. For instance, some VMRs might not affect expression of the nearest gene and instead regulate more distant genes, or they may not influence gene expression at all. Furthermore, there may also be VMRs where high methylation does not repress and instead enhances gene expression.

Lastly, visualization of chromatin accessibility (Fig. 3.7b) revealed that demethylation of the identified VMRs coincides with a wave of increased accessibility of these VMRs. Unlike demethylation, however, this increase in accessibility is merely transient and no longer detected in neuroblasts and neurons of the olfactory bulb. This transient increase in VMR accessibility might enable binding of TFs and TET enzymes required for demethylation.

3.1.7 VMRs are enriched for cell type-specific transcription factor motifs

Next, I aimed to identify TFs that might be responsible for methylation differences between the various cell populations in the data set. To do so, I performed a motif enrichment analysis to find transcription factor binding site (TFBS) motifs that frequently occur in VMRs. Specifically, I used HOMER (Heinz et al., 2010) to scan the DNA sequences of VMRs for sequence motifs from the JASPAR2022 (Castro-Mondragon et al., 2022) TFBS motif data base. I used the previously computed PC vectors based on VMR methylation to group VMRs into three groups (Fig. 3.8a):

- VMRs with high PC1 scores, indicative of low methylation in oligodendrocytes
- VMRs with low PC2 scores, indicative of low methylation in astrocytes

- VMRs with high PC2 scores, indicative of low methylation in the neurogenic lineage.

To narrow down the list of TF candidates, I furthermore considered the average expression of each TF, highlighting only those TFs that are expressed as indicated by the scRNA-seq portion of the data.

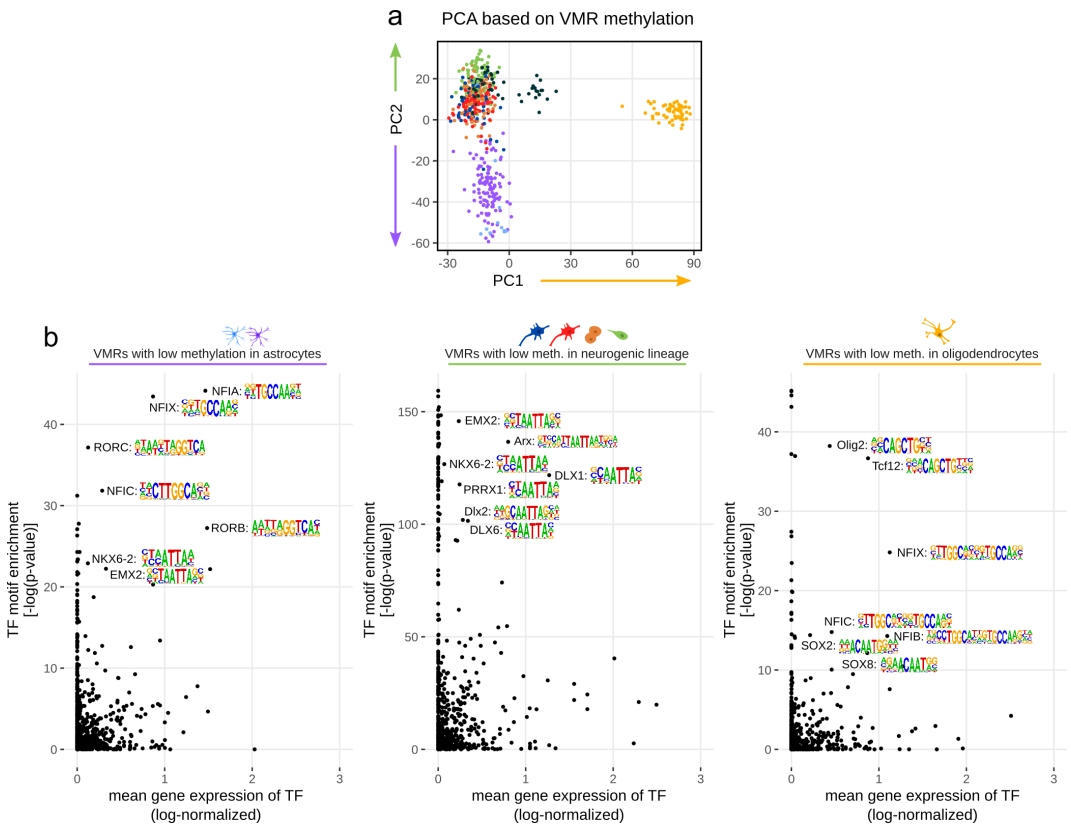

FIGURE 3.8: Transcription factor motifs enriched in variably methylated regions. Transcription factor (TF) binding site motif enrichment of variably methylated regions (VMRs) with low methylation in astrocytes, neuroblasts or oligodendrocytes. **(a)** PCA of single-cell methylomes, based on VMR methylation values. Since PC1 and PC2 separate oligodendrocytes and astrocytes, respectively, from the other cells, I used these PCs to identify the genomic regions (i.e., VMRs) used for motif enrichment. **(b)** Scatter plots of transcription factors, showing the TF motif's enrichment p-values (reported by HOMER, Heinz et al., 2010) on the y axis, and the TF mean gene expression in the respective cell population on the x axis. Figure and caption modified from Kremer et al. (2022b).

Motifs associated with low methylation in astrocytes / qNSC1 cells

Motif enrichment of each VMR group revealed several enriched TFBS motifs, many of which were enriched only in one of the three VMR groups (Fig. 3.8b). VMRs lowly methylated in qNSC1 cells (vSVZ astrocytes) are enriched for the motifs of several nuclear factors, including NFIA, NFIX and NFIC. While motif enrichment is no definite evidence that these TFs are involved in astrocyte-specific demethylation, NFIA is used to convert human iPSC-derived NSCs to astrocytes, and in doing so induces demethylation at GFAP, an astrocyte-specific intermediary filament that is commonly used as an astrocyte marker (Tchieu et al., 2019). Together with the

results from my TFBS enrichment analysis, this suggest that NFIA may promote astrocyte cell identity by inducing demethlyation near genes required for astrocyte function.

In contrast, another nuclear factor whose motif is enriched in qNSC1 cells, NFIX, was previously linked to NSC function: Martynoga et al. (2013) previously demonstrated that NFIX promotes NSC quiescence, but does not simultaneously induce the expression of astrocyte markers such as S100-B. A follow-up study also showed that NFIX suppresses oligodendrogenesis in adult vSVZ NSCs (Zhou et al., 2015), which raises the question why I found the NFIX motif to be enriched in astrocyte-specific regions. First, since all NFI proteins bind to the same DNA recognition motif (Gronostajski, 2000), with only slight variations in the composition of surrounding base pairs, my analysis cannot distinguish between individual members of this gene family. It is thus possible that e. g. low methylation of NFIA motifs led to enrichment of NFIX motifs, purely due to their motif similarity and not because NFIX is responsible for astrocyte-specific demethylation. Alternatively, it is also possible that vSVZ astrocytes (qNSC1) share some epigenomic properties with NSCs, and that these depend on NFIX. A first step towards entangling the role of nuclear factors in NSCs and astrocytes would be to characterize their expression and the methylation status of their TBFS in greater detail, and to assess whether the NFI motifs that are demethylated in vSVZ astrocytes are similarly demethylated in NSCs.

Motifs associated with low methylation in the neurogenic lineage

VMRs lowly methylated in the neurogenic lineage, i. e., in cells at the qNSC2 stage up until neuroblasts, are enriched for the TFBS motifs of several neurogenic TFs (Fig. 3.8b). Among these are EMX2, a homeobox TF that controls proliferation of adult vSVZ NSCs (Galli et al., 2002; Gangemi et al., 2001), and Dlx2. The role of Dlx2 in forebrain development is well-characterized, where it regulates interneuron migration (Anderson et al., 1997) and promotes generation of interneurons over oligodendrocyte progenitor cells (Petryniak et al., 2007). It was also shown that Dlx2 is crucial for the generation of olfactory bulb interneurons from the lateral vSVZ (Brill et al., 2008).

Motifs associated with low methylation in oligodendrocytes

Lastly, inspection of TFBS enriched in oligodendrocyte-specific LMRs revealed Olig2 and Tcf12 as the top hits (Fig. 3.8b). Olig2 is a well-studied master regulator of oligodendrocyte cell identity (Zhang et al., 2022; Zhou et al., 2000). Tcf12 is a member of the TCF/LEF (T cell factor/lymphoid enhancer factor family) TF family, which regulates expression of Wnt target genes (Behrens et al., 1996). It was recently shown that Wnt ligands, specifically Wnt3a, stimulate oligodendrogenesis of adult NSCs in the dorsal vSVZ (Ortega et al., 2013). The authors also demonstrated that Tcf4, another member of the TCF/LEF family, is required for this. Whether Tcf12

plays a similar role in adult oligodendrogenesis, and whether this process involves demethylation near Tcf12-binding sites, remains to be determined.

Overall, many of the TFs identified in this enrichment analysis correspond to cell-type specific functions, or even act as master regulators of cell fates or lineage decisions. Nonetheless, there are also caveats to the presented analysis. First, my approach cannot inform us about the causality that led to decreased methylation at the identified TFBS motifs: Are the identified TFs pioneer TFs that cause changes in methylation and chromatin accessibility (Tchieu et al., 2019; Donaghey et al., 2018; Reizel et al., 2018)? Or is binding of the identified TFs sensitive to DNA methylation (Kaluscha et al., 2022; Hernandez-Corchado and Najafabadi, 2022), which would mean that binding of these TFs might be regulated by changes in DNA methylation? Or are the identified motifs not functionally linked to changes in DNA methylation, but merely a byproduct of demethylation near gene sets that are regulated by certain TFs? Experimental validation of individual TFs is required to distinguish between these options. Second, my analysis is limited to TFs with known motifs listed in the JASPAR2022 database, and several motifs correspond to human TFs (indicated by their capitalization). Last, since most TFs are part of a larger protein family with very similar motifs, it is often not possible to know which exact TF binds to the identified motifs. This limitation is clearly visible in the motif enrichment of VMRs with low methylation in the neurogenic lineage (Fig. 3.8b), which identified several members of the homeobox TF family that all share a variant of the TAATTA-motif. To alleviate this issue, I disregarded TFs that are not expressed according to the transcriptomic data, but in many cases multiple members of the same family are expressed. Despite these limitations, the TFs identified here are promising candidates for TFs that might regulate cell fate by a mechanism that involves changes in DNA methylation.

3.1.8 Epigenetic changes at marker genes

In the preceding analyses, I was able to distinguish cell types based on DNA methylation alone and found that VMRs were enriched for TFBS motifs linked to cell type-specific functions. Thus, I hypothesized that DNA methylation regulates the expression of genes specifically expressed in certain cell types or cell states, as it was previously proposed for cell differentiation in embryonic development (Cedar et al., 2022, Fig. 1.7). To assess whether this is the case, I first selected genes that contain a VMR in their gene body and then used the single-cell transcriptomic data to identify 100 VMR-containing marker genes for each observed cell type/state.I then visualized the average expression of these markers, as well as the average methylation and accessibility of the intersecting VMRs and promoters, in pseudotime (Fig. 3.9).

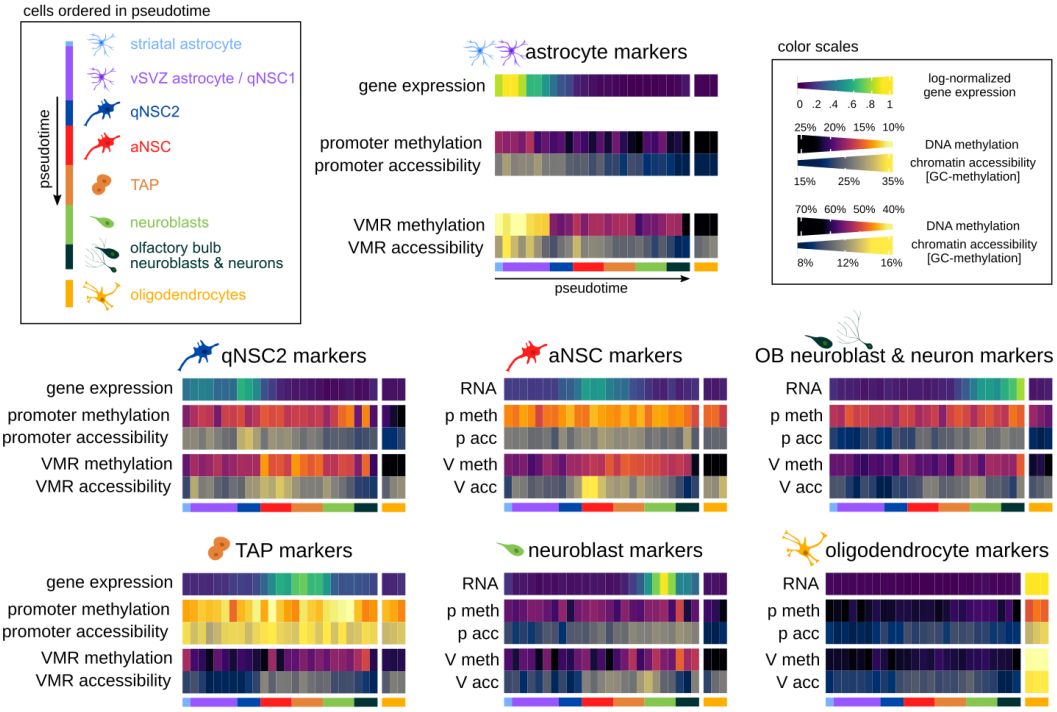

FIGURE 3.9: Epigenetic changes at cell type-specific genes.
Expression and epigenetic status of genes expressed at different states of the NSC lineage (top legend). The top heatmap row indicates the average expression of 100 marker genes which I identified in a differential expression analysis of the single-cell transcriptomes. Middle and bottom rows indicate the average DNA methylation (meth) and chromatin accessibility (acc) at the 100 promoters (p) of these markers, and at 100 VMRs (V) overlapping marker gene bodies. Figure and caption modified from Kremer et al. (2022b).

This analysis revealed several patterns. First, despite explicitly searching for genes that are expressed only in qNSC1 cells and striatal astrocytes, the identified set of 100 astrocyte marker genes was also lowly expressed in qNSC2 cells. This once again highlights the challenge of finding reliable marker genes to distinguish common parenchymal astrocytes from NSCs (reviewed e. g. in Lim and Alvarez-Buylla, 2016, section 1.1.4), due to their transcriptomic similarity (Llorens-Bobadilla et al., 2015; Zywitza et al., 2018; Cebrian-Silla et al., 2021). In stark contrast, methylation at VMRs intersecting these genes showed a very distinct cut that clearly distinguishes vSVZ astrocytes (qNSC1 cells) from qNSC2 cells. This clear separation was not visible in any of the other epigenetic features, which displayed more gradual, continuous change over pseudotime. Overall, these observations suggest that VMRs at astrocyte marker genes are specifically demethylated in common parenchymal astrocytes of the striatum and vSVZ, but not in qNSC2 cells or later stages of the NSC lineage. Nonetheless, despite the high methylation of these VMRs, qNSC2 cells retain some expression of the identified astrocyte marker genes.

The analysis of qNSC2- and aNSC-specific marker genes was overall more challenging. The reason for this is that, as discussed above, qNSC1 and qNSC2 cells are transcriptomically similar, and thus I was unable to identify strong markers that are exclusively expressed in qNSC2 cells. Since I additionally restricted the analysis

to markers that intersect VMRs and only few VMRs arise at e. g. the aNSC stage (Fig. 3.7a), my approach was unable to identify 100 very specific marker genes for these stages. As a result, the average marker gene did not display a very specific expression peak at the qNSC2 or aNSC stage and instead was also lowly expressed at other stages. Epigenetic readouts at VMRs and promoters near these marker genes overall appeared to be noisy and often in disagreement with changes in gene expression. For instance, VMRs at aNSC markers were more lowly methylated at differentiation stages past aNSCs. Due to the unclear patterns observed at aNSC- and qNSC2-markers, possibly due the dearth of VMRs specific to these stages, this analysis did not allow me to draw strong conclusions. Similarly, promoters and VMRs near neuroblast markers and markers of olfactory bulb neuroblasts/neurons showed only weak patterns. In general, VMRs and promoters of both marker gene sets tend to become more accessible and less methylated in pseudotime. These changes occur gradually in pseudotime and might correspond to a gradual transition from an astrocytic to a neuronal cell identity.

Inspection of TAP markers, however, showed a very characteristic pattern distinct from that of other marker genes: In contrast to other markers, promoters of these genes were both lowly methylated and highly accessible. Curiously, this promoter state was not specific to the TAP stage, but also clearly observed at all other stages of the NSC lineage and even in oligodendrocytes. A reason for this might be that many TAP marker genes are involved in the cell cycle (e. g. Llorens-Bobadilla et al., 2015). This means that many identified TAP markers do not perform a cell type-specific function *per se*, but rather a basic cellular function, i. e. cell division, that might also be active on other cycling cells and not only TAPs. For this reason, the epigenetic makeup of TAP markers might resemble that of housekeeping genes, i. e. constitutively expressed genes involved in core cellular functions. It is long known that promoter methylation differs strongly between housekeeping genes and cell type-specific genes. Promoters of housekeeping genes have a higher CpG density and are generally more lowly methylated, while CpG sites at promoters of tissue-specific genes are more variably methylated and occur at a lower density (Bird, 1986; Gardiner-Garden and Frommer, 1987; Larsen et al., 1992; Saxonov et al., 2006; Zhu et al., 2008). Since the cell cycle is not a tissue-specific or cell-type specific function, I thus speculate that cell cycle genes share epigenetic properties, including constitutively low promoter methylation, with housekeeping genes. Indeed, a recent study in which genes were clustered based on their CpG distribution described that both housekeeping genes and genes related to cell division were found in the same gene cluster, characterized by high density and low methylation of CpG sites (Tian et al., 2022).

3.2 A unique methylome distinguishes NSCs from other astrocytes

All my previously described analyses suggested that there are strong methylation differences between qNSC1 cells and qNSC2 cells: Visualization of single-cell methylomes with PCA (Fig. 3.8a) and UMAP (Fig. 3.1b) resulted in a clear separation of qNSC1 cells from the remaining cells. In agreement with this, visualization (Fig. 3.6) and quantification (Fig. 3.7a) of methylation changes suggested that many VMRs are indeed differentially methylated between qNSC1 and qNSC2 cells. Furthermore, astrocytes isolated from the striatum grouped with qNSC1 cells in these analyses, suggesting that the methylome of qNSC1 cells resembles that of common parenchymal astrocytes. For these reasons, I proposed that qNSC1 cells correspond to common parenchymal astrocytes of the vSVZ.

If this is the case, the strong methylation differences that I detected between qNSC1 cells and qNSC2 cells might support different functions: either functions of a common parenchymal astrocyte, or stem cell function in the case of qNSC2. To test this hypothesis, I next aimed to identify and functionally characterize the genomic regions that make up the observed methylome difference.

To identify differentially methylated regions (DMRs) between qNSC1 cells and cells of the neurogenic lineage (qNSC2 $\rightarrow$ neuroblast), I considered the set of previously identified VMRs and tested each VMR for differential methylation between the two cell groups. To this end, I performed a Wilcoxon rank sum test between the two groups for each VMR and corrected for multiple testing with the Benjamini-Hochberg method. Of note, this analysis predates development of our improved DMR detection approach (Braun, 2023), which is why I resorted to *post hoc* testing of VMRs. This approach identified 1 372 genomic regions that were lowly methylated in qNSC1 cells, which I labeled astrocyte lowly-methylated regions (**astrocyte LMRs**), as well as 621 regions that were lowly methylated in the neurogenic lineage (**NSC LMRs**) (Fig. 3.10a). In this volcano plot, I furthermore colored each VMR by the $\log_2$-fold change in expression of the nearest gene, to assess whether methylation differences go hand to in hand with differences in gene expression. This visualization revealed that low VMR methylation is often accompanied by increased expression of the nearest gene, but this tendency was not evident for all VMRs.

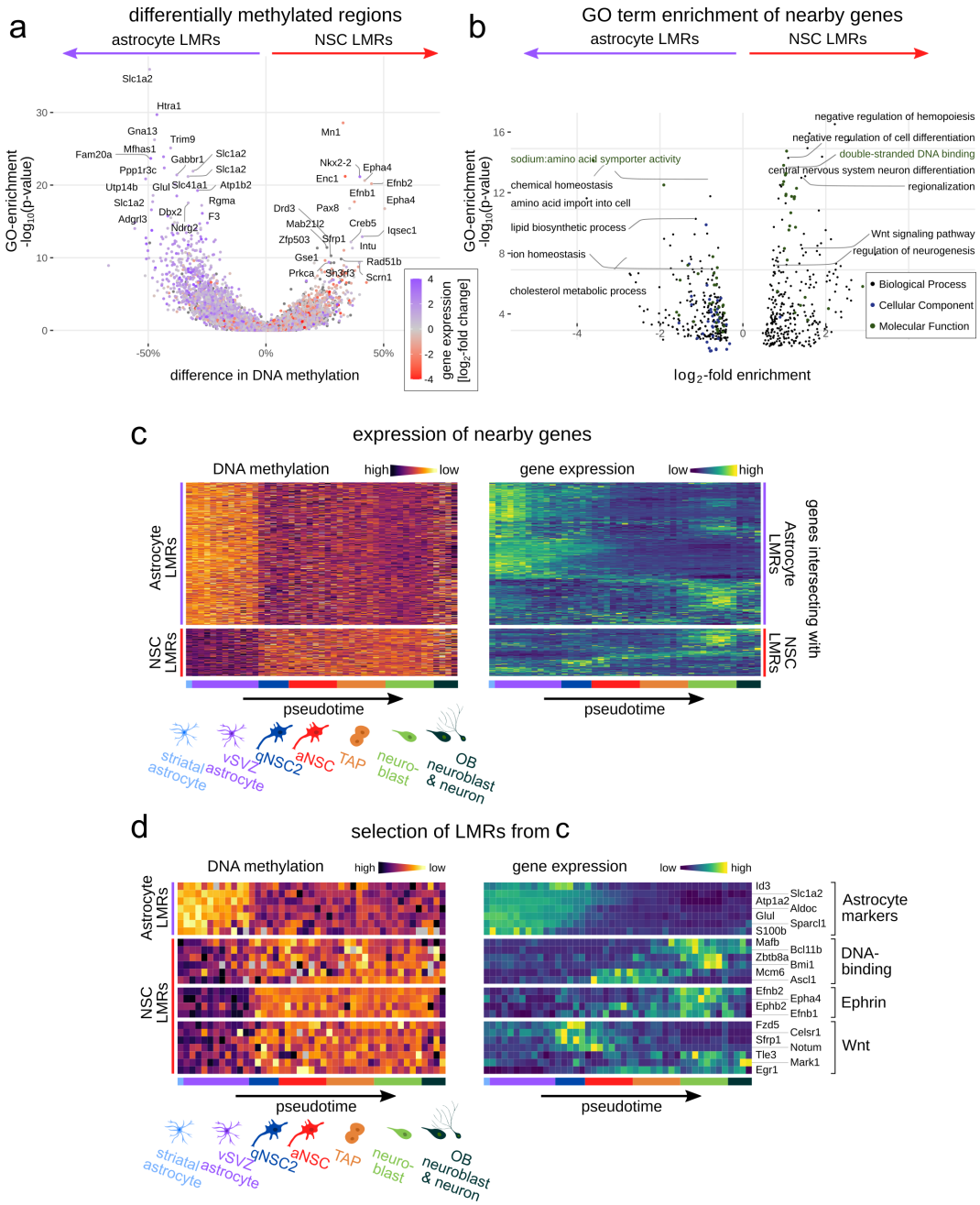

FIGURE 3.10: NSCs possess a pro-neurogenic methylome that clearly distinguishes them from other astrocytes.
(a) Volcano plot of lowly methylated regions (LMRs), which I identified by testing for differential methylation between vSVZ astrocytes (qNSC1) and NSC-lineage cells (qNSC2 → neuroblast). **(b)** Gene ontology (GO) term enrichment of genes near astrocyte LMRs and NSC LMRs identified in **a**. **(c)** Heatmap of LMR methylation (left) and expression of the nearest gene (right) along pseudotime. Rows are ordered by hierarchical clustering on gene expression values. **(d)** Selection of LMRs and genes from **c**. Note the clear separation of NSCs and other astrocytes (striatal astrocytes, vSVZ astrocytes = qNSC1) in the methylation data. Figure and caption taken from Kremer et al. (2022b).

3.2.1 Astrocyte LMRs and NSC LMRs support distinct cellular functions

To assess the functional implications of NSC LMRs and astrocyte LMRs, I inspected genes which overlap these putative regulatory regions. Among the genes located

near astrocyte LMRs were many genes associated with classical astrocyte functions. This includes genes involved in amino acid metabolism and transport such as the excitatory amino acid transporter *Slc1a2* and the glutamine synthetase *Glul*, as well as other ion transporters such as the magnesium transporter *Slc41a2* and the ATP-dependent sodium/potassium transporter *Atp1b1*. Low methylation near these genes might contribute to the release and uptake of neurotransmitters by astrocytes and to their role in maintaining ion homeostasis. Similarly, low methylation near *Lcat*, a gene involved in cholesterol metabolism, might support astrocytic production of cholesterol. Gene ontology (GO) term enrichment of VMRs, using the GREAT tool (McLean et al., 2010), indeed showed an enrichment of several GO terms related to ion transport, lipid/cholesterol metabolism, and ion homeostasis and transport (Fig. 3.10b).

Genes near NSC LMRs included several genes involved in Eph/ephrin signaling (*Epha4*, *Efnb1*, *Efnb2*), which regulates both neuroblast migration and NSC quiescence (Holmberg et al., 2005; Jiao et al., 2008; Nomura et al., 2010; Ricard et al., 2006) (Fig. 3.10a). Enriched GO terms such as "negative regulation of cell differentiation", "central nervous system neuron differentiation" and "regulation of neurogenesis" furthermore suggest that low methylation at NSC LMRs may contribute to the neurogenic function of NSCs (Fig. 3.10b).

Curiously, the top enriched GO term was "negative regulation of hemopoiesis". A possible explanation is that other somatic stem cells, including hematopoietic stem cells, employ a similar set of genes to regulate cell differentiation. Indeed, a closer inspection of the gene sets associated with the identified GO terms showed that 110 of the 111 genes associated with "negative regulation of hemopoiesis" are also associated with the more general GO term "negative regulation of cell differentiation", which is the second most enriched GO term in my analysis. I suspect that many of these genes are not only involved in hematopoietic stem cell function, but also important for other somatic stem cells such as adult NSCs. The functional annotation of genes involved in hematopoiesis might simply be more accurate than that of other somatic stem cells, since hematopoiesis is a well-studied process.

Other enriched GO terms include "double-stranded DNA binding" and "Wnt signaling pathway". The former GO term most likely corresponds to the presence of hypomethylated VMRs near TFs such as *Nkx2-2*, *Pax8*, *Mn1* and *Zfp503* (Fig. 3.10a). These TFs may contribute to NSC-specific functions. In the case of *Zfp503*, it was recently shown that this gene encodes for a transcriptional repressor that targets the astrocyte marker gene *Gfap* and promotes neurogenesis over astrogliogenesis in the developing mouse cerebral cortex. Low methylation near this gene might thus contribute to the transition from an astrocyte cell identity to the neuroblast cell identity in NSC-to-neuroblast differentiation. Similarly, the presence of LMRs near Wnt signaling genes might contribute to the regulation of adult neurogenesis by

Wnt signaling as reported for the vSVZ (Qu et al., 2010) and dentate gyrus (Lie et al., 2005).

3.2.2 Genes required for neurogenesis are hypomethylated in NSCs

The observed DNA methylation differences and associated GO terms seem to confirm the notion that DNA methylation underpins either astrocyte function in the case of qNSC1 cells, or the generation of neuroblasts in the case of qNSC2 cells. However, previous studies (e. g. Kalamakis et al., 2019) and my own analysis (Fig. 3.6) showed that qNSC1 and qNSC2 cells have similar transcriptomes. Thus, it seems unlikely that all identified DMRs are tied to gene expression differences between qNSC1 cells and qNSC2 cells. But is there any link between DNA methylation differences and gene expression? To investigate, I determined the closest gene for each significant LMR and visualized LMR methylation alongside gene expression in pseudotime (Fig. 3.10c). Additionally, to illustrate some exemplary gene-to-LMR matches, I picked some of the most significant LMRs and grouped them into four categories (Fig. 3.10d).

Hierarchical clustering of the log-normalized gene expression values showed that most genes near hypomethylated VMRs are indeed more strongly expressed in the respective population. For instance, approximately two thirds of the genes near astrocyte LMRs are more strongly expressed in early pseudotime, i. e. in astrocytes and qNSCs. However, unlike the methylation state of LMRs, gene expression of nearby genes does not show a clear change point between qNSC1 cells and qNSC2 cells, as also previously observed in Fig. 3.6. Instead, many genes near astrocyte LMRs appear to be gradually fading out in pseudotime (Fig. 3.10c). Among these genes are several astrocyte marker genes such as *S100b* and *Aldoc*, which I highlighted in Fig. 3.10d. As was already reported long ago (e. g. in Doetsch et al., 1999), these astrocyte markers are not only expressed in striatal astrocytes and qNSC1 cells, but also in qNSC2 cells, which means that they are not suitable to distinguish common parenchymal astrocytes from NSCs. However, LMRs near these astrocyte markers are clearly methylated at the qNSC2 state and at later pseudotime points of the neurogenic lineage, while the same LMRs are hypomethylated in striatal astrocytes and qNSC1 cells. Surprisingly, this might mean that astrocyte marker genes are suitable to distinguish common parenchymal astrocytes from NSCs, but only if one considers their LMR methylation instead of their expression.

Why astrocyte marker genes are expressed in qNSC1 cells despite the high methylation of nearby astrocyte LMRs is unclear. One possible explanation is that cells are constantly transitioning between the qNSC1 and qNSC2 states, and that astrocyte marker transcripts are retained and thus detectable in qNSC2 cells that recently transitioned from the qNSC1 state. However, since I observed moderate expression of most astrocyte markers in qNSC2 cells, this scenario would require significant turnover between the qNSC1 and qNSC2 populations and/or very long persistence

of mRNA molecules. Another possibility is that methylation at the identified LMRs does not silence gene expression but rather modulates it in another way. For instance, one might speculate that hypermethylation of astrocyte LMRs does not influence the expression of astrocyte marker genes directly. Instead, methylation at astrocyte LMRs may simply be the first step toward silencing nearby astrocyte markers in the course of NSC-to-neuroblast differentiation. A possible mechanism is that methylated CpG dinucleotides in astrocyte LMRs enable binding of methyl-CpG-binding domain (MBD) proteins such as MeCP2. During NSC activation, MBD family proteins might bind astrocyte LMRs, leading to changes in chromatin conformation and histone modifications, and subsequently to silencing of astrocyte marker genes. Clemens et al. (2020) recently uncovered a similar mechanism in the murine cerebral cortex, where intragenic enhancers of certain genes are first methylated during early postnatal development. Later in adulthood, methylated cytosines in these enhancers are then bound by MeCP2, which leads to repression of these genes.

However, not all genes near astrocyte LMRs are expressed early in pseudotime. Roughly one third of these genes, illustrated in Fig. 3.10c, shows a more diffuse pattern that is not anticorrelated with LMR methylation. Many of these genes appear to peak in their expression at the neuroblast stage. I suspect that some of these genes are mismatched, which means that the displayed gene is simply not regulated by the assigned LMR, since matching LMRs and genes by proximity is a simple heuristic that likely results in some erroneous matches. Another option is that DNA methylation at some of these LMRs is not repressive, but is rather positively correlated with gene expression as previously reported for e. g. gene body methylation (Tran et al., 2005; Neri et al., 2017, section 1.4.2).

Genes located near NSC LMRs are expressed at various time points of the adult NSC lineage (Fig. 3.10c). Approximately half of these genes are expressed in neuroblasts, which suggests that the observed methylation differences are related to the expression of genes required at the neuroblast stage. In Fig. 3.10d, I manually selected some of the most significant NSC LMRs and grouped them by their function. Again, the progression of gene expression along pseudotime paints a slightly different picture than the observed changes in DNA methylation: While the methylation state of NSC LMRs clearly differs between qNSC1 cells and striatal astrocytes on the one hand and qNSC2 cells and later stages on the other hand, gene expression appears to be much more dynamic and often restricted to specific cell states.

Expression of the previously discussed genes involved in Eph/ephrin signaling peaks at the neuroblast stage and a similar trend was also observed for several genes encoding DNA-binding proteins and for genes involved in Wnt signaling. Again, the data indicate that DNA methylation does not simply mirror gene expression. Instead, the relationship between DNA methylation and gene expression appears to be more complex and temporally uncoupled. Since VMRs near many genes that become transcriptionally active in neuroblasts are already lowly methylated much

earlier, at the qNSC2 stage, I suspect that low methylation alone is not sufficient to enable transcription. Instead, low methylation at NSC LMRs might be a prerequisite for gene regulatory mechanisms that cause gene expression at a later stage. This would mean that qNSC2 cells have the epigenetic license to express neuroblast genes, while qNSC1 cells are locked in their astrocyte fate by methylation at NSC LMRs, which may prevent the expression of genes required for neurogenesis. For a more detailed formulation of this model, please refer to section 4.2.2. Overall, my analysis suggests that the observed methylation differences promote neurogenesis in qNSC2 cells and later stages, and common parenchymal astrocyte functions in qNSC1 cells and striatal astrocytes.

3.3 Ischemia induces methylation changes in NSCs and other astrocytes

Based on my analysis of our scNMT-seq data of the murine NSC lineage, I proposed that the two methylome states that I discovered support stem cell function and parenchymal astrocyte function, respectively (section 3.2). However, testing this hypothesis is challenging. To assess whether the NSC methylome, characterized by hypomethylated NSC LMRs and hypermethylated astrocyte LMRs, is required for the generation of neuroblasts, one would ideally endow striatal astrocytes with a NSC methylome and observe whether this leads to the generation of neuroblasts. Alternatively, one could also induce methylation of NSC LMRs in cells of the vSVZ, to assess whether this impairs neurogenesis. However, methods to modify DNA methylation are still in their infancy. While recent advances in CRISPR/Cas9-based technologies enable editing of DNA methylation at precisely defined loci (reviewed in Nakamura et al., 2021; Nuñez et al., 2021), these methods are only suitable to target few loci simultaneously. However, my analysis identified at least 2 043 astrocyte LMRs and NSC LMRs, which means that employing epigenome modifiers to convert one of the methylome states into the other is not feasible.

Thus, we devised a different experiment to determine whether the previously described NSC methylome is essential for neurogenesis. As reviewed in section 1.2.3, common parenchymal astrocytes of the striatum can acquire NSC properties and generate neuroblasts under certain conditions such as ischemic injury. If my hypothesis is true and a specific DNA methylation profile is required for NSC function, this acquisition of neurogenic properties by striatal astrocytes must be accompanied by a transition from the astrocyte methylome to the NSC methylome.

To test this hypothesis, Dr. Cerrizuela and colleagues subjected male 2 month old mice to transient global brain ischemia for 22 minutes. They then isolated GLAST$^+$ cells from the striatum and vSVZ of mice 2 days post ischemic injury (dpi) and 21 dpi and subjected them to scNMT-seq (Fig. 3.11). To ensure that any potential neuroblasts or NSC-like cells in the striatum are not simply vSVZ NSCs that migrated toward

the site of ischemic injury, they furthermore used a transgenic mouse line that allows for labeling of vSVZ NSCs via the tamoxifen-inducible Cre recombinase system. In these mice, injection of tamoxifen leads to expression of yellow fluorescent protein (YFP) in cells expressing the NSC marker gene *Nr2e1*, commonly known as tailless (Tlx).

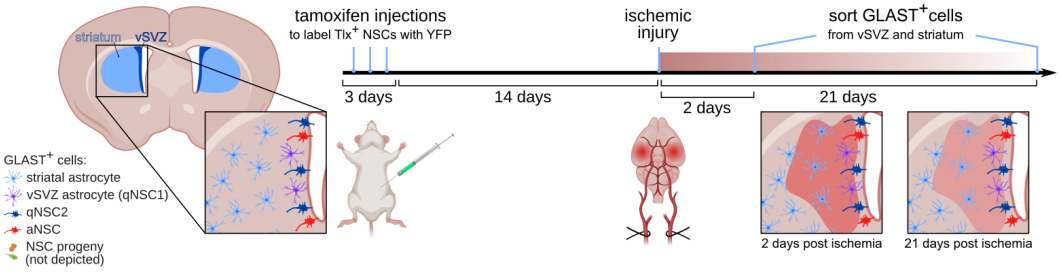

FIGURE 3.11: Experiment to assess the effects of ischemia on GLAST$^+$ cells of the vSVZ and striatum.
This experiment was carried out by Dr. Santiago Cerrizuela and colleagues (see Kremer et al., 2022b, for details). I analyzed the resulting data. GLAST$^+$ cells comprise vSVZ NSCs, vSVZ astrocytes (qNSC1) and striatal astrocytes. Both tissues were sequenced at two time points: 2 dpi (days post ischemia) and 21 dpi. Tamoxifen was injected to label *Nr2e1*$^+$ (commonly known as Tlx) vSVZ NSCs via Cre-inducible YFP (yellow fluorescent protein) expression to detect potential NSCs that may have migrated to the striatum. Figure and caption modified from Kremer et al. (2022b).

3.3.1 Ischemia induces a neurogenic response in the vSVZ and striatum

To assess the effects of ischemia on astrocytes and NSCs of the vSVZ and striatum, I processed single-cell transcriptomes and methylomes as previously described for cells from naive mice. I then integrated the single-cell transcriptomes of scNMT-seq cells from naive mice and post-ischemic mice with the larger 10X Genomics scRNA-seq data set from Carvajal Ibañez et al. (2023) (Fig. 3.12).

This visualization revealed dramatic changes in gene expression two days after ischemia in both the vSVZ and the striatum. In the vSVZ of naive mice, the transcriptomes of GLAST$^+$ cells are distributed along a continuous trajectory that starts with vSVZ astrocytes (qNSC1 cells) and ends with neuroblasts. In contrast, the naive striatum harbors a homogeneous population of striatal astrocytes that closely resemble qNSC1 cells of the vSVZ. Two days after ischemic injury, the distribution of cells in both the vSVZ and striatum shifts drastically: The transcriptomic space previously occupied by vSVZ astrocytes and striatal astrocytes contains fewer cells two days after ischemia, suggesting that ischemia led to the depletion of this cell population. Another striking difference in cells isolated two days after ischemia is the presence of a new GLAST$^+$ cell population in transcriptomic space that was not previously occupied in cells from naive mice. This ischemia-specific cell population, located outside transcriptomic space occupied by cells of the naive NSC lineage, was observed in both the vSVZ and the striatum. The observed depletion of striatal and vSVZ astrocytes and the concomitant emergence of another cell population might be the result of ischemia-induced gene expression changes in astrocytes. This

would mean that the ischemia-specific cell population located outside the naive NSC lineage consists of striatal astrocytes and vSVZ astrocytes that underwent expression changes in response to ischemia.

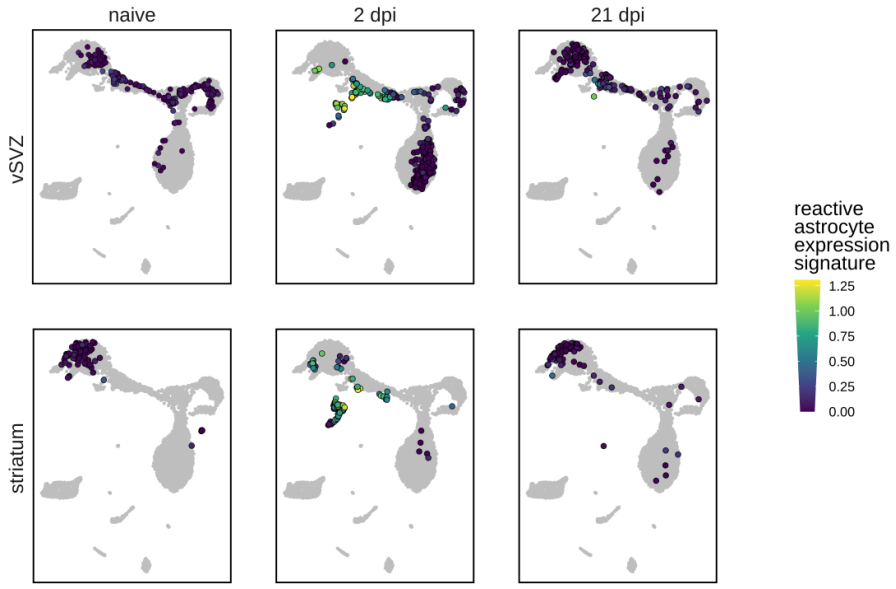

FIGURE 3.12: Injury-induced changes in gene expression.
Comparison of the cell states 2 dpi (days post ischemia) and 21 dpi with the cell states in naive mice. I colored the cells by their mean log-normalized expression of reactive astrocyte marker genes from Liddelow et al. (2017). Figure and caption taken from Kremer et al. (2022b).

Reactive astrogliosis

It is known that astrocytes respond to brain injury by entering a reactive state that is characterized by changes in morphology, gene expression and function (Escartin et al., 2021, section 1.2.2). To test whether the newly emerged cell population corresponds to reactive astrocytes, I obtained a list of reactive astrocyte marker genes from Liddelow et al. (2017). In this publication, the authors described two sub-populations of reactive astrocytes: A1 reactive astrocytes, to which the authors attributed neurotoxic properties, and A2 reactive astrocytes, which the authors described as neuroprotective. I obtained their marker gene list to assess whether the novel cell population I observed two days after ischemia consists of reactive astrocytes. To do so, I colored each cell by its average log-normalized pan-reactive astrocyte marker gene expression (Fig. 3.12). This revealed that both cells in the early NSC lineage, as well as cells in the ischemia-specific transcriptomic space express genes that were previously ascribed to reactive astrogliosis.

Visualizing expression of marker genes for reactive astrocyte sub-types, I showed that post-ischemic cells in the early neurogenic lineage express markers of A1 reactive astrocytes, while the cells positioned outside the naive lineage are more similar to the A2 reactive astrocyte sub-type (Fig. 3.13a-b). However, it is worth noting that not all marker genes reported by Liddelow et al. (2017) follow this pattern and several of these marker genes have low gene expression values in our

data set (Fig. 3.13c). This is to be expected, since both the binary classification into A1 and A2 reactive astrocytes, as well as the presence of marker genes from Liddelow et al. (2017) is not universally applicable to all tissues and conditions (Escartin et al., 2021).

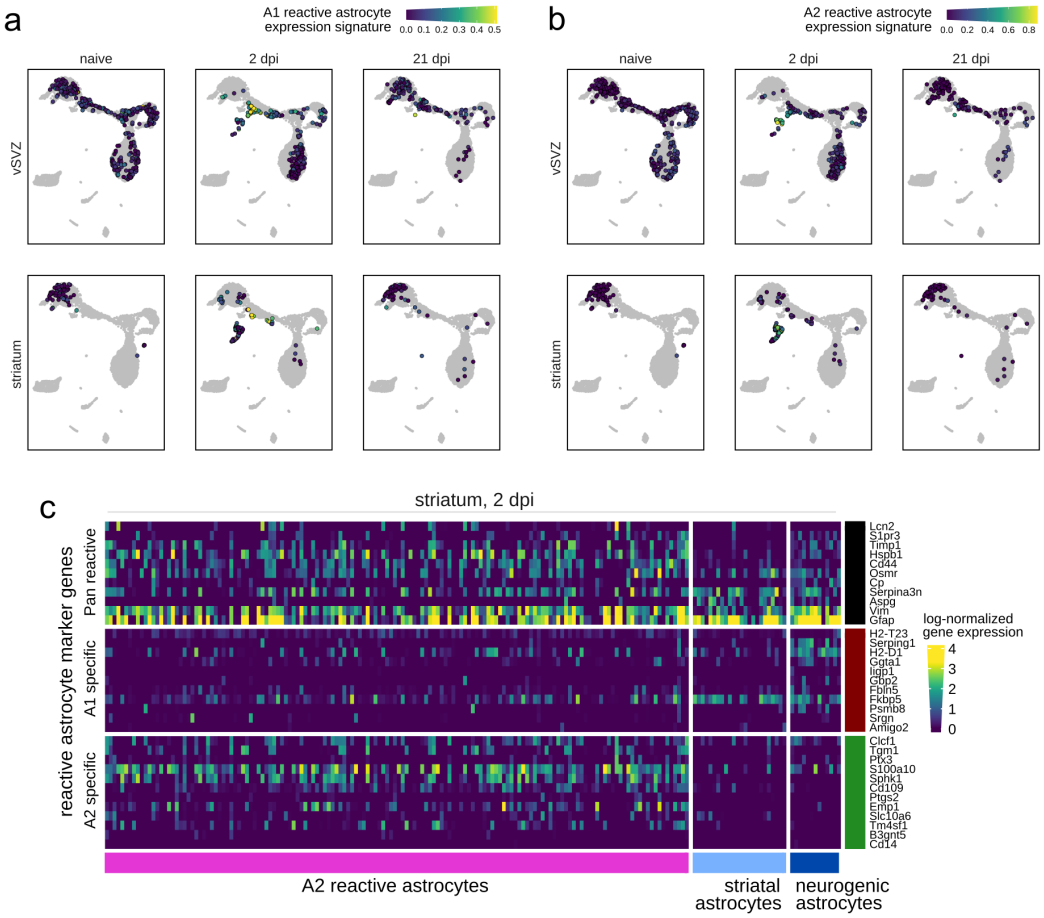

FIGURE 3.13: Expression of reactive astrocyte marker genes.
(a, b) Mean log-normalized expression of marker genes for the two sub-types of reactive astrocytes (**a**: A1, **b**: A2) listed in Liddelow et al. (2017). **(c)** Expression of reactive astrocyte marker genes in the striatum 2 dpi (days post ischemia). Figure and caption taken from Kremer et al. (2022b).

Nonetheless, the presence of two distinct cell populations expressing reactive astrocyte marker genes in our data suggests a heterogeneous response of astrocytes to ischemia that roughly resembles the binary classification into A1 and A2 reactive astrocytes. Already 21 days after ischemia, reactive astrocytes are no longer detected and striatal astrocyte and vSVZ astrocyte populations are replenished (Fig. 3.12).

Exit from the astrocyte state

The observation that one of the injury-responsive astrocyte sub-populations is located in transcriptomic space that is occupied by qNSC2 cells in naive mice suggests that these cells might contribute to neurogenesis. In the striatum, this putative shift from striatal astrocytes toward transcriptomic states previously observed in the

naive vSVZ lineage is especially striking: While the naive striatum contains a homogeneous population of striatal astrocytes, the striatum 2 dpi contains several cells resembling vSVZ NSCs, TAPs and neuroblasts (Fig. 3.12). Overall, this suggests that ischemia caused activation of a neurogenic program in striatal astrocytes as previously described by Magnusson et al. (2014) and others. 21 days after ischemia, the striatal astrocyte population is restored, although several neuroblast-like and NSC-like cells are still observed.

In the vSVZ 2 dpi, I also identified a greater number of neuroblasts than in the naive vSVZ. This suggests that depletion of the vSVZ astrocyte population is not only accompanied by reactive astrogliosis, but potentially also by an increase in the generation of neuroblasts. Taken together, the depletion of vSVZ astrocytes and simultaneous increase in the number of neuroblasts suggests that vSVZ astrocytes/NSCs are activated upon ischemia, as previously reported by Llorens-Bobadilla et al. (2015).

3.3.2 Post-ischemic astrocytes acquire an NSC methylome

My analysis of scRNA-seq transcriptomes of cells isolated from post-ischemic mice suggested that ischemia induced gain of neurogenic capabilities in common parenchymal astrocytes of the striatum. To assess whether this gain of neurogenic capabilities is accompanied by remodeling of the DNA methylome, I once again quantified DNA methylation at those NSC LMRs and astrocyte LMRs that I had previously discovered in data from naive mice. Most strikingly, striatal astrocytes that have entered the neurogenic lineage 2 dpi are characterized by low methylation of NSC LMRs and high methylation of astrocyte LMRs (Fig. 3.14, Fig. 3.15). This suggests that ischemia-induced gain of neurogenic capabilities causes large-scale, genome-wide remodeling of the astrocyte methylome into an NSC-like methylome that supports neurogenesis and stem cell function. Unlike those cells that seemingly entered the neurogenic lineage, reactive astrocytes outside the neurogenic lineage (resembling A2 reactive astrocytes) do not have an NSC-like methylome, which once again demonstrates that not all astrocytes activate a neurogenic program in response to ischemia, and that epigenome remodeling is not necessarily linked to reactive astrogliosis.

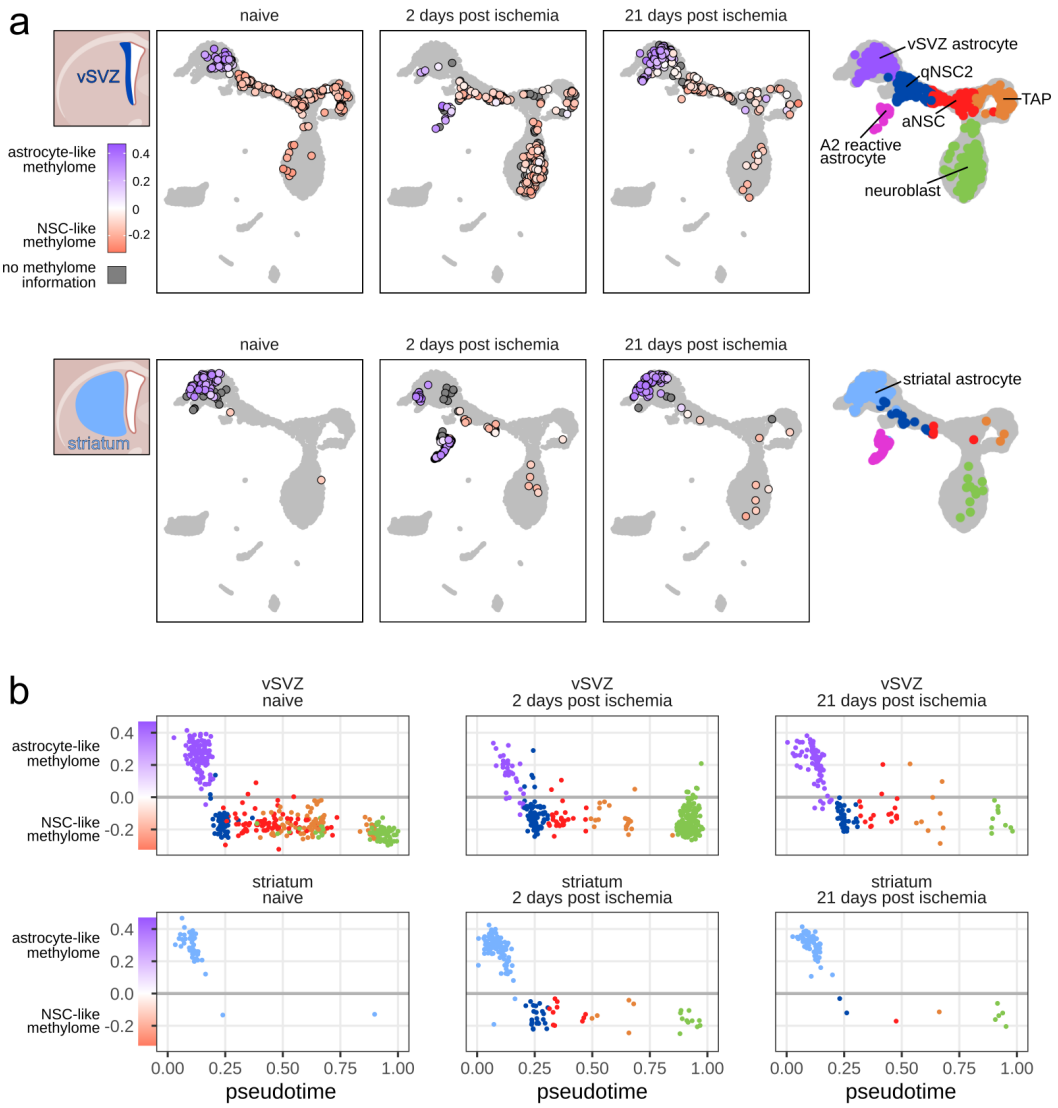

FIGURE 3.14: Injury-induced changes in DNA methylation.
(a) Comparison of transcriptional and epigenetic cell states 2 dpi and 21 dpi with the cell states in naive mice. I colored each cell according to its average methylation of NSC LMRs and astrocyte LMRs from Fig. 3.10. The depicted score is the difference between the average methylation of all astrocyte LMRs and the average methylation of all NSC LMRs (also see Fig. 3.15). **(b)** Alternative visualization of the data in **a**. Here, I represented the transcriptome state of each cell by a single pseudotime value instead of UMAP coordinates. Figure and caption modified from Kremer et al. (2022b).

Both the vSVZ astrocyte population and the striatal astrocyte population is restored 21 dpi. These replenished astrocytes possess the previously described astrocyte methylome, which suggests that injury-induced methylome remodeling is rapidly reversed. However, it is important to note that our data set does not allow us to draw definite conclusions about the origins of these cells. While it seems likely that these 21 dpi astrocytes correspond to cells that returned from the neurogenic lineage back to the astrocyte state, there are also other possible explanations: First, it is possible that these restored astrocytes are astrocytes that migrated to the vSVZ and striatum from neighboring brain areas. Second, these cells might simply correspond

to astrocytes that did not respond to ischemia. While almost all astrocytes in the 2 dpi sample showed a strong response to ischemia, is might be that the individual mouse sampled for the 21 dpi data point showed a weaker response. The complete lack of a response to ischemia in this individual seems unlikely, however, since I detected several cells in the 21 dpi striatum that epigenetically and transcriptomically resemble cells from the vSVZ NSC lineage (Fig. 3.14). Nonetheless, follow-up experiments should aim to replicate my results on a larger number of mice.

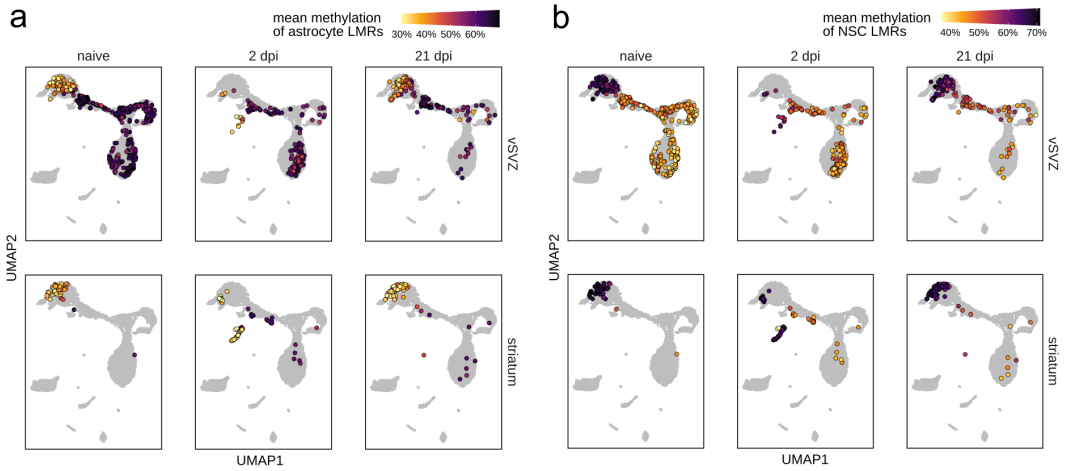

FIGURE 3.15: Methylation of astrocyte LMRs and NSC LMRs.
Average methylation of astrocyte LMRs **(a)** and NSC LMRs **(b)** from Fig. 3.10. Figure and caption modified from Kremer et al. (2022b).

3.3.3 Lack of interferon signaling hinders the neurogenic response to ischemia

Llorens-Bobadilla et al. (2015) previously demonstrated that interferon signaling is required for the concerted activation of vSVZ astrocytes in ischemia. But does impaired interferon signaling also prevent remodeling of the DNA methylome? To address this question, Dr. Cerrizuela and colleagues obtained mice lacking both the interferon-γ-receptor and the interferon-α-receptor (IFNAGRKO mice, Huang et al., 1993; Müller et al., 1994). Again, they subjected these mice to transient global brain ischemia and, two days after ischemia, isolated GLAST$^+$ cells for scNMT-seq.

My analysis of the transcriptomes of these cells revealed that, compared to cells from wild type 2 dpi, IFNAGRKO mice had a more limited response to ischemia, characterized by a higher retention of cells in the astrocyte state (Fig. 3.16). Importantly, my quantification of DNA methylation demonstrated that these cells also retained an astrocyte methylome with low methylation of astrocyte LMRs and high methylation of NSC LMRs. IFNAGRKO samples furthermore did not display the gene expression signature of A1-reactive astrocytes (Fig. 3.17). In contrast, the A2-reactive gene expression signature was clearly present also in IFNAGRKO cells, suggesting that interferon signaling may be required for the upregulation of some genes involved in reactive astrogliosis upon ischemia, but not for others. In my previous analysis

of wild type 2 dpi cells, I observed that astrocytes 2 dpi, which likely entered the neurogenic lineage, express A1-reactive markers but not A2-reactive markers which are expressed by cells outside the neurogenic lineage. As IFNAGRKO impairs both the neurogenic response to ischemia and the upregulation of A1-reactive astrocyte marker genes, it is tempting to speculate that both of these responses are related and depend on interferon signaling.

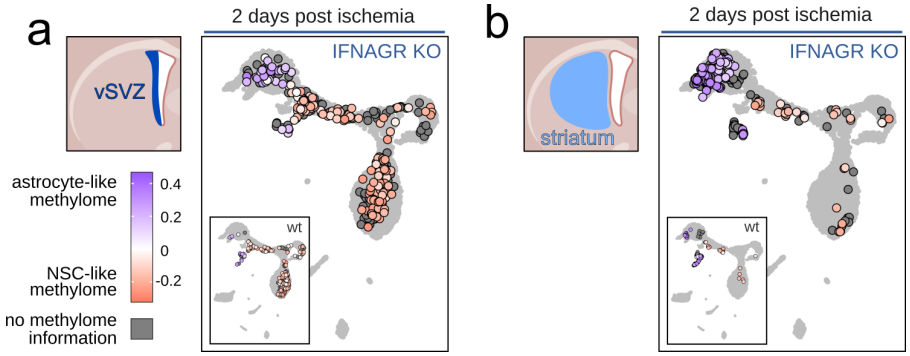

FIGURE 3.16: Lack of interferon signaling impairs the generation of neurogenic astrocytes by ischemia.
GLAST$^+$ cells isolated from the vSVZ **(a)** and striatum **(b)** two days after ischemia, depicted as in Fig. 3.14. For reference, the inset depicts GLAST$^+$ cells isolated from wild type mice 2 dpi (from Fig. 3.14). A control of cells isolated from naive IFNAGRKO mice is shown in Fig. 3.17. Figure and caption modified from Kremer et al. (2022b).

Surprisingly, even though the ischemia-induced depletion of astrocytes was inhibited in IFNAGRKO, I still detected an increased number of neuroblasts in the IFNAGRKO vSVZ 2 dpi sample. This increase is clearly visible when comparing the sample to both the naive wild type vSVZ and the naive IFNAGRKO control vSVZ (Fig. 3.17). While additional independent statistical replicates are required to derive strong conclusions from this observation, a possible interpretation is that ischemia increases the production of vSVZ neuroblasts even in the absence of interferon signaling. But how is increased neurogenesis achieved, if it is not the result of increased astrocyte-to-qNSC2 transition? Carvajal Ibañez et al. (2023) recently showed that IFNAGRKO affects both NSC activation and self-renewal, and that the overall effect on the rate of neural progenitor (TAPs, neuroblasts) production strongly depends on age. In young mice, as I have studied them here, their model predicts that IFNAGRKO leads to a strong decrease in activation, a modest decrease in self-renewal, and decreased production of neural progenitors. As I observed high neuroblast production and retention of astrocytes upon ischemia in IFNAGRKO, it is possible that ischemia leads to increased self-renewal despite absence of interferon signaling.

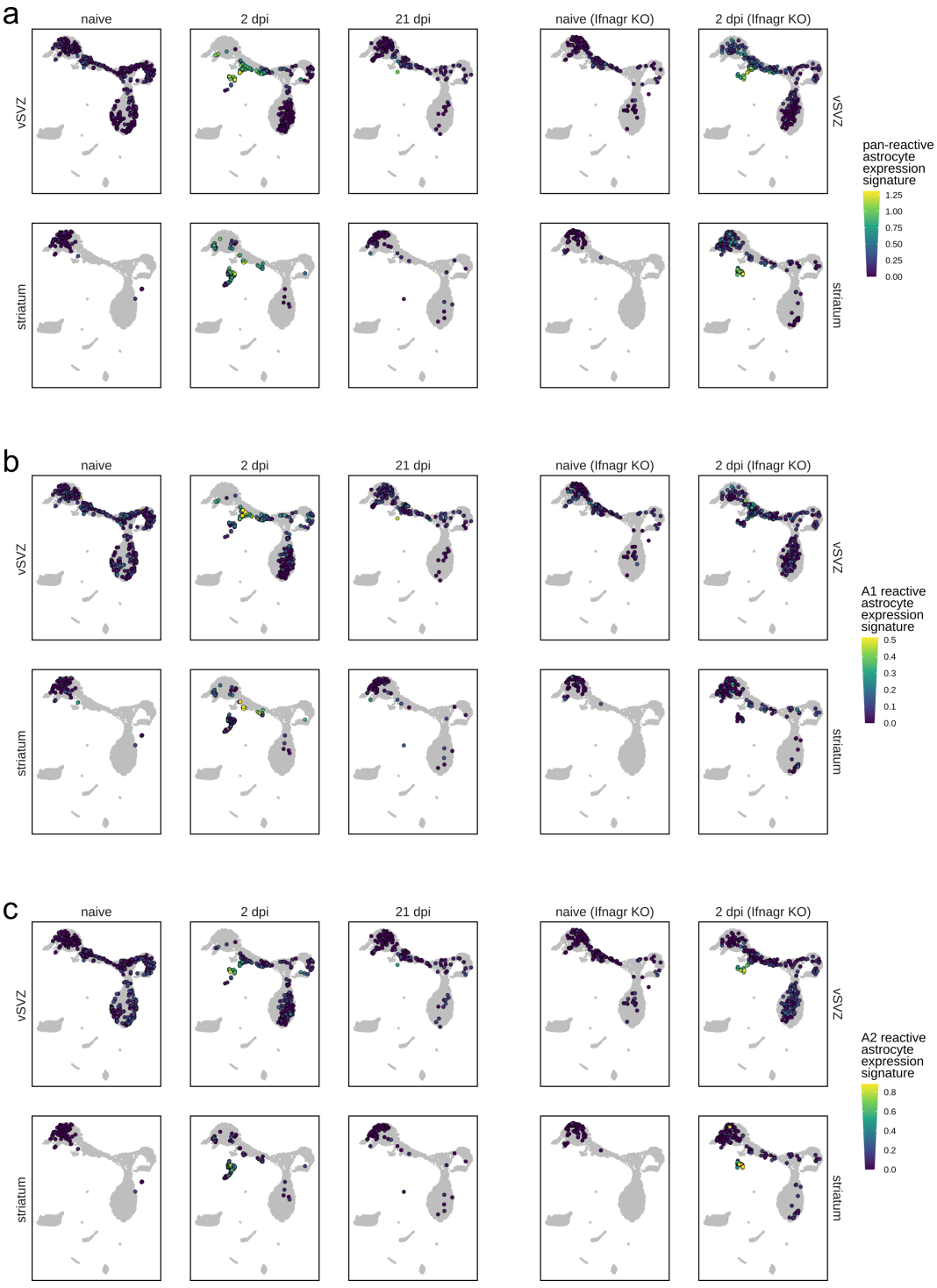

FIGURE 3.17: Comparison of reactive astrocyte marker gene expression between wild type and IFNAGRKO.
Mean log-normalized gene expression of three reactive astrocyte marker gene sets in wild type (left) and IFNAGRKO (right). To aid the comparison between wild type and IFNAGRKO, the wild type panels from Fig. 3.12 and Fig. 3.13a-b are repeated here. **(a)** Pan-reactive astrocyte markers. **(b)** A1 reactive astrocyte markers. **(c)** A2 reactive astrocyte markers. Figure and caption modified from Kremer et al. (2022b); reactive astrocyte marker gene sets were obtained from Liddelow et al. (2017).

To summarize, my analysis suggests that interferon signaling might be an upstream regulator of several responses to ischemia including astrocyte-to-qNSC2 transition, the associated methylome remodeling, and the upregulation of A1 reactive astrocyte markers.

3.3.4 Ischemia induces additional methylation changes

In my previous analysis of single-cell methylomes isolated from naive and post-ischemic mice, I quantified DNA methylation specifically at astrocyte LMRs and NSC LMRs. This analysis indicated that ischemia induces epigenetic changes at these genomic intervals that are consistent with a transition from the astrocyte state to the neurogenic stem cell fate. However, since I restricted this methylome analysis to LMRs which I had discovered in healthy mice, it was unable to reveal any additional methylome changes that might occur outside of astrocyte/NSC LMRs upon ischemia. To investigate whether ischemia also induces other methylation changes, I utilized the single-cell DMR detection approach implemented by Martina Braun (2023) under my supervision. To this end, I selected qNSC2 cells, aNSCs and TAPs of the vSVZ. I then used `scbs diff` to scan the entire genome of these cells for potential DMRs between 2 dpi and naive samples. This approach revealed several hypomethylated and hypermethylated **ischemia DMRs** (Fig. 3.18). Although GO term enrichment analysis did not yield any significant hits, possible due to the modest number of identified DMRs, I recognized genes involved in astrocyte function (*Apoe*, *Clu*) and synapse function (*Nrxn1*) among the nearby genes (Fig. 3.18a).

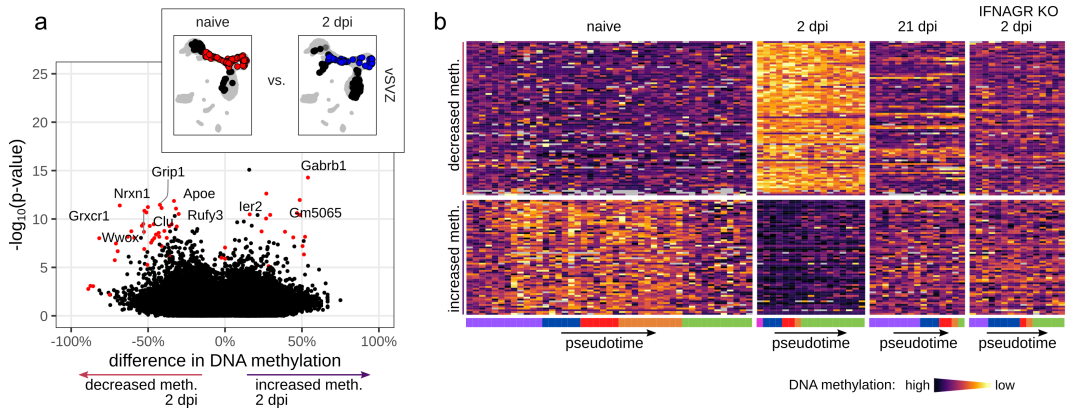

FIGURE 3.18: Methylation changes two days after ischemia.
(a) Volcano plot of DMRs which I detected 2 dpi (red: adjusted p-value < 0.05). Inset: NSC-lineage cells 2 dpi (blue; qNSC2 cells, aNSCs and TAPs) were tested against NSC-lineage cells from naive mice (red). Only DMRs that intersect a gene body are depicted. **(b)** Heatmap of DMR methylation of cells isolated from naive mice and post-ischemic mice, as well as 2 dpi IFNAGRKO mice. All DMRs with an adjusted p-value < 0.05 are depicted. Figure and caption modified from Kremer et al. (2022b).

To also assess the methylation status of the identified DMRs at 21 dpi and in IFNAGRKO 2 dpi, I visualized these DMRs in a heatmap (Fig. 3.18b). This visualization confirmed that the identified genomic regions are indeed differentially methylated

between cells isolated from naive mice, and cells isolated 2 dpi. This difference in methylation was almost entirely absent in cells from the 21 dpi sample. Nonetheless, a small number of DMRs that I found to be hypomethylated 2 dpi were also hypomethylated 21 dpi, as indicated by the bright bands in the upper 21 dpi section of the heatmap. This might suggest that these DMRs retain their low methylation status for at least three weeks.

Next, I also interrogated the 21 dpi sample for ischemia DMRs, using the same approach described earlier. Unlike the 2 dpi comparison, this analysis detected mostly DMRs that were hypomethylated 21 dpi (Fig. 3.19a).

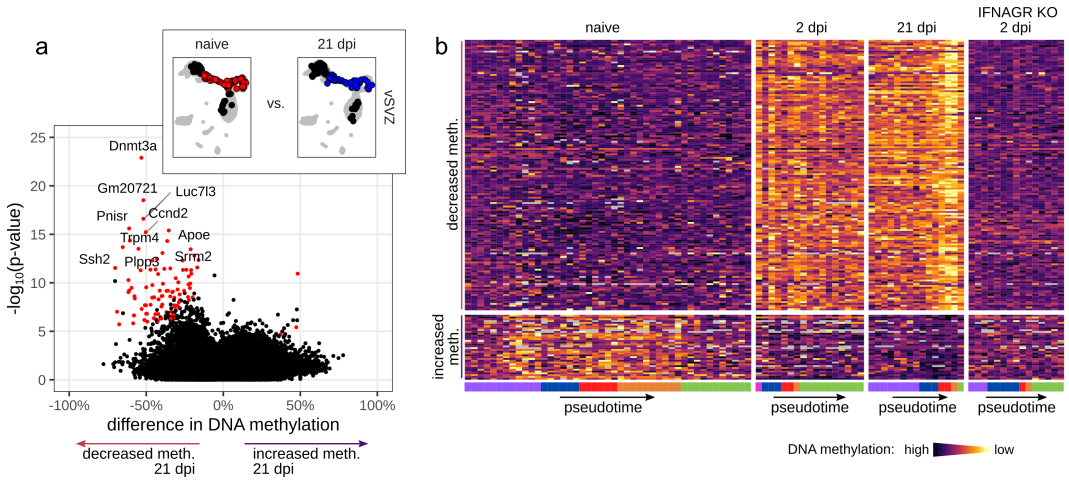

FIGURE 3.19: Methylation changes 21 days after ischemia.
(a) Volcano plot of DMRs which I detected 21 dpi (red: adjusted p-value < 0.05). Inset: NSC-lineage cells 21 dpi (blue) were tested against NSC-lineage cells from naive mice (red). Only DMRs that intersect a gene body are depicted. **(b)** Heatmap of DMR methylation of cells isolated from naive mice and post-ischemic mice, as well as 21 dpi IFNAGRKO mice. All DMRs with an adjusted p-value < 0.05 are depicted. Figure and caption modified from Kremer et al. (2022b).

Visualizing the DMRs in a heatmap showed that the hypomethylated DMRs detected 21 dpi were also hypomethylated in the 2 dpi sample, albeit to a lesser extent (Fig. 3.19b). Even though I detected these DMRs 21 dpi, my observation that they are also demethylated 2 dpi suggests that they were demethylated as an immediate response to ischemia in the course of at most two days. What might be the reason why ischemia DMRs detected 21 dpi are shared by both time points, while most ischemia DMRs detected 2 dpi are observed at this time point exclusively? Since DMRs detected at the later time point are also hypomethylated earlier, it appears that the DMR detection approach simply missed most of these DMRs in the 2 dpi sample. This might be due to lower statistical power, i. e. lower coverage or cell count at this time point, or simply since the difference in methylation is lower at 2 dpi, as is also visible in (Fig. 3.19b). But why are not all 2 dpi ischemia DMRs also shared by the 21 dpi sample? It might be that these DMRs are only transiently demethylated upon ischemia and return to their former methylation state within

three weeks. Furthermore, some observed ischemia DMRs might be the result of epigenetic variation of individual mice due to limited sample size.

Methylation at *Dnmt3a*

Surprisingly, the top DMR detected 21 dpi intersects with the gene *Dnmt3a*, which encodes the methyltransferase responsible for *de novo* methylation of DNA (Fig. 3.19b, also see section 1.4.1). To investigate this specific DMR in greater detail, I visualized its methylation status for single cells (Fig. 3.20). This visualization confirmed that the DMR `12:3895478-3915578` near *Dnmt3a* was indeed lowly methylated specifically at 21 dpi, but not at other time points.

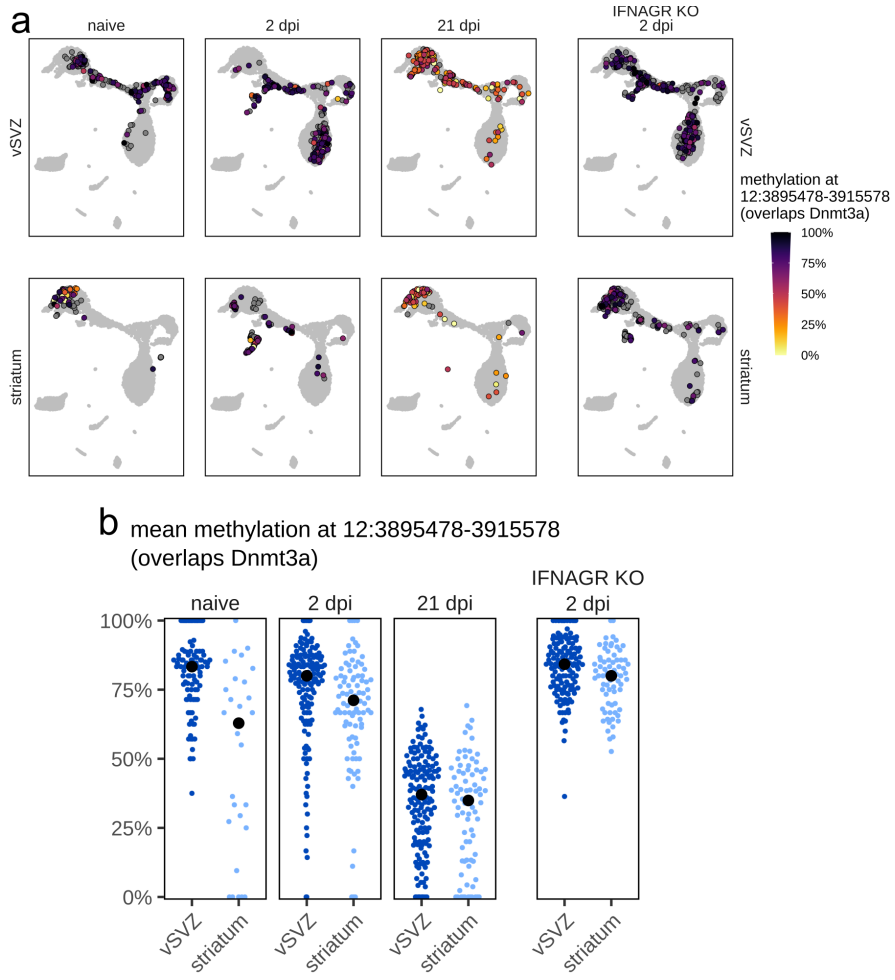

FIGURE 3.20: Methylation at *Dnmt3a* is decreased 21 days after ischemia.
(a) Mean methylation of the top DMR from Fig. 3.19a, which overlaps the gene *Dnmt3a*. Cells with no sequencing coverage at this DMR are shown in dark gray. I integrated the single cell transcriptome data with a larger scRNA-seq data set from Carvajal Ibañez et al. (2023) (small light gray background cells). **(b)** Bee swarm plot of data from **a**. Small colored points represent single cells, large black points denote the median. Figure and caption taken from Kremer et al. (2022b).

What could be the functional role of demethylation at *Dnmt3a* upon ischemia? My analysis of naive and post-ischemic scNMT-seq samples suggests that astrocytes

undergo epigenetic remodeling, including the demethylation of astrocyte LMRs, upon ischemia. I also observed that astrocytes with a regular astrocyte methylome are replenished already 21 days after ischemia. Similarly, as shown in Fig. 3.18b, it appears that several DMRs that acquire low methylation upon ischemia return to their original hypermethylated state after 21 days. If this dynamic is true, it requires a mechanism by which DNA methylation is restored at e. g. astrocyte LMRs, most likely by DNTM3A. I speculate that demethylation at *Dnmt3a* might be part of a negative feedback loop that occurs upon ischemia: Ischemia might lead to hypomethylation near *Dnmt3a*, which in turn might cause increased expression of this gene. Increased DNMT3A levels might then facilitate re-methylation of loci that were previously demethylated upon ischemia, such as astrocyte LMRs. While this model might help explain how ischemia-induced demethylation is reversed already 21 dpi, I was unable to find evidence for increased gene expression of *Dnmt3a* in the single-cell transcriptomics data. Considering that ischemia-induced changes in DNA methylation are already reverted 21 dpi, it might be that the proposed increase in *Dnmt3a* has already ceased at this time point.

Both the ischemia DMRs observed 2 dpi, as well as those observed 21 dpi, were not differentially methylated in cells isolated from post-ischemic IFNAGRKO mice (Fig. 3.18b, Fig. 3.19b). This suggests that interferon signaling is not only an upstream regulator of the neurogenic response to ischemia, but might also be required for the establishment of ischemia DMRs.

To summarize, my analysis confirmed that ischemia leads to activation of a neurogenic program in astrocytes of the vSVZ and striatum. Excitingly, this transition from a common parenchymal astrocyte to a specialized stem cell astrocyte is accompanied not only by changes in gene expression, but also by genome-wide remodeling of the DNA methylome. Methylome remodeling occurs predominantly at genomic sites that are already differentially methylated between common parenchymal astrocytes and NSCs in the naive brain. This suggests that the LMRs I identified in cells isolated from naive brains are indeed linked to neurogenesis. In addition, I identified a small number of ischemia-specific DMRs that are not present in the naive brain, one of which might facilitate the return of cells to the parenchymal astrocyte state after ischemia. Knockout of the interferon α- and γ-receptors impaired both gene expression changes and methylome remodeling, which unveils interferons as potential upstream regulators of this response to ischemia.

3.4 Active demethylation in oligodendrocytes

The single-cell triple-omic data set of the naive vSVZ comprised a large number of oligodendrocytes. In my analysis of single-cell methylomes I noticed that oligodendrocytes readily separated from other cell populations in PCA (Fig. 3.8), even when using basic analysis methods (Fig. 2.4c). Based on this observation, I speculated that oligodendrocytes may possess a unique methylome with distinct characteristics. Thus, I set out to identify genomic regions that are lowly methylated in oligodendrocytes specifically.

3.4.1 Genes required for myelination are hypomethylated in oligodendrocytes

Since this analysis predates the development of `scbs diff` (Braun, 2023), I used the same approach that I used to detect astrocyte LMRs and NSC LMRs (see section 3.2): I considered the set of previously identified VMRs and tested them for differential methylation between oligodendrocytes and other vSVZ cells. I detected a large number of DMRs, many of which showed an average methylation difference of $\pm 50\%$ (Fig. 3.21a). I suspected that these DMRs might underpin the expression of genes that are highly expressed in oligodendrocytes. To investigate, I utilized the scRNA-seq layer of the data and computed the $\log_2$-fold change in expression between oligodendrocytes and other vSVZ cells for each gene. I then restricted the analysis to VMRs that intersect a gene and colored each VMR by the $\log_2$-fold change in gene expression.

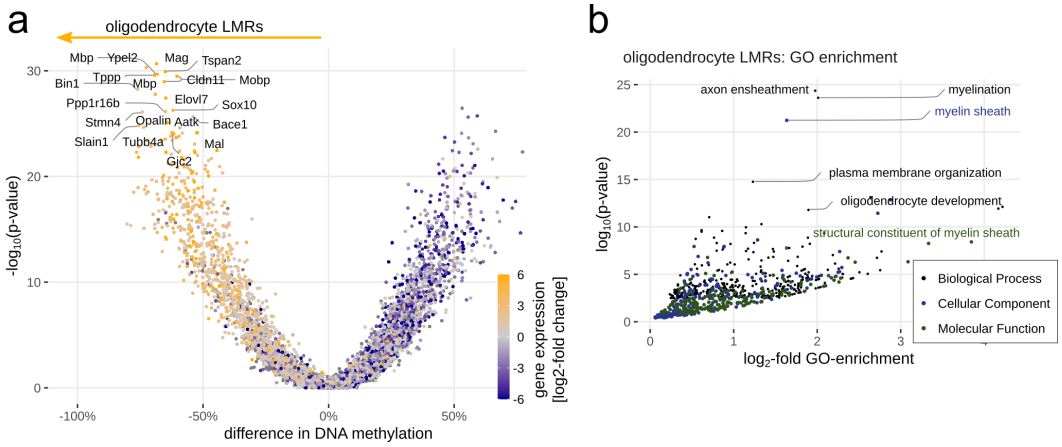

FIGURE 3.21: Identification of oligodendrocyte LMRs.
(a) Volcano plot of VMRs that intersect gene bodies, tested for differential methylation between oligodendrocytes and all other cells (from naive wild type mice only). VMRs are colored by the $\log_2$-fold change in gene expression between oligodendrocytes and other cells. **(b)** Gene ontology (GO) term enrichment of genes near oligodendrocyte LMRs identified in **a**. Each point corresponds to one GO term and is colored by its sub-ontology. Figure and caption modified from Kremer et al. (2022b).

Among the intersecting genes were several commonly used oligodendrocyte marker genes such as *Mbp*, *Mobp*, *Mag* and *Sox10*. This visualization also revealed a clear link

between differential methylation and differential gene expression: Almost all VMRs that were lowly methylated in oligodendrocytes also intersected with a gene that had higher expression in oligodendrocytes, in line with the common notion (Mattei et al., 2022, section 1.4.2) that DNA methylation acts as a repressive epigenetic mark (Fig. 3.21a). Only few VMRs appeared to deviate from this relationship, which is why I decided to focus on genomic regions that are *lowly* methylated in oligodendrocytes (**oligodendrocyte LMRs**).

Visual inspection of Fig. 3.21a suggested that oligodendrocyte LMRs occur near genes that are highly expressed in oligodendrocytes. To gain insight into the function of these genes, I used GREAT (McLean et al., 2010) for region-based GO term enrichment analysis. Among the top enriched GO terms were several terms linked to the primary function of oligodendrocytes, namely formation of the myelin sheath around axons (Fig. 3.21b). Taken together, these results indicate that oligodendrocyte LMRs support oligodendrocyte function by enhancing the expression of genes required for myelination.

3.4.2 Oligodendrocytes show traces of active demethylation

Next, I wondered by which molecular mechanism the cell type / cell state-specific LMRs I identified in oligodendrocytes, astrocytes and NSCs are introduced. As summarized in section 1.4.1, DNA methylation can be removed either passively during DNA replication, or actively via TET enzymes. TET enzymes remove DNA methylation by catalyzing a targeted, stepwise series of oxidations (Tahiliani et al., 2009, section 1.4.1). As illustrated in Fig. 3.22a, this series of oxidations contains three intermediate stages that represent partially oxidized derivatives of 5-methylcytosine. Two of these intermediates, 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC), can be detected by methylase-assisted bisulfite sequencing (MAB-seq), a technique based on bulk WGBS (Wu et al., 2014). To assess whether active demethylation via TET enzymes occurs in cells of the adult vSVZ, Dr. Sascha Dehler and Dr. Santiago Cerrizuela had performed MAB-seq on several cell populations of the vSVZ, isolated by FACS. Of these samples, I analyzed the following three cell populations:

- $O4^+$ oligodendrocytes
- PSA-$NCAM^+$ neuroblasts
- $GLAST^+/CD9^{high}$ vSVZ astrocytes and NSCs.

This data set comprises two replicates per cell population, with each replicate consisting of twelve pooled mice.

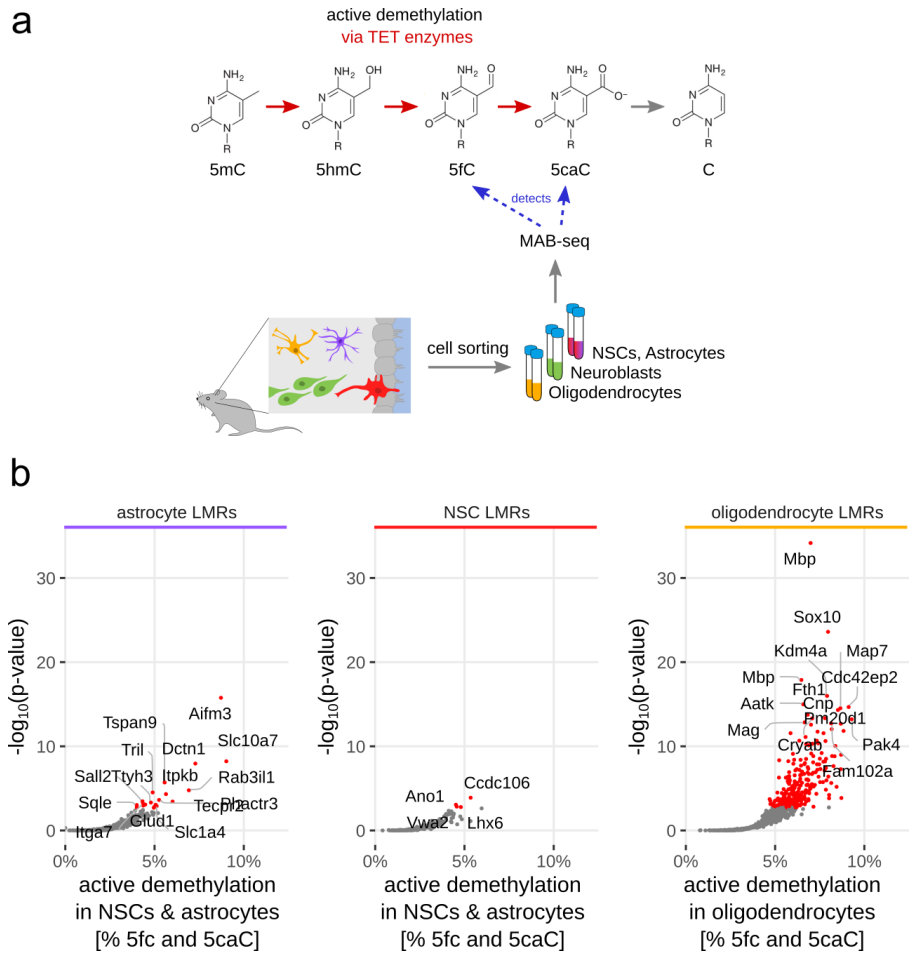

FIGURE 3.22: Using MAB-seq to quantify active DNA demethylation.
(a) Schematic of TET-mediated active demethylation (top) and the experimental workflow to detect active demethylation with MAB-seq (bottom). This experiment was performed by Dr. Sascha Dehler and Dr. Santiago Cerrizuela. I analyzed the resulting data. 5mC: 5-methylcytosine; 5hmc: 5-hydroxymethylcytosine; 5fC: 5-formylcytosine; 5caC: 5-carboxylcytosine. **(b)** Quantification of active demethylation in LMRs of specific cell types and cell states. I previously identified the genome coordinates of LMRs (lowly methylated regions) in scNMT-seq data (Fig. 3.10a and Fig. 3.21a). I assessed these coordinates in the MAB-seq data to quantify active demethylation at these LMRs. Figure and caption modified from Kremer et al. (2022b).

I utilized this MAB-seq data set to examine astrocyte LMRs, NSC LMRs and oligodendrocyte LMRs for traces of active demethylation, i. e. 5fC and 5caC. Specifically, I used a binomial generalized linear model with a logit-link function, testing the alternative hypothesis that the average 5fC and 5caC level is higher than the global average level of the samples. I then used the Benjamini-Hochberg method to adjust p-values for multiple testing. My analysis identified several LMRs with strong evidence for traces of active demethylation (Fig. 3.22b).

My analysis showed that only a modest number of astrocyte LMRs and NSC LMRs contain molecular traces of TET-mediated demethylation. However, it is important to note that MAB-seq can likely only detect recent demethylation events, since 5fC and 5caC may only persist for a limited time. Thus, I cannot exclude the possibility that astrocyte LMRs and NSC LMRs are the result of past TET activity that is no

longer detectable. Nonetheless, the quantity of 5fC/5caC-containing astrocyte and NSC LMRs may allow me to speculate about the rate at which cells transition between the vSVZ astrocyte (qNSC1) state and the qNSC2 state under homeostatic conditions *in vivo*. Assuming that this transition is accompanied by TET-mediated methylome remodeling, the modest number of 5fC/5caC-containing LMRs suggests that the majority of cells in the bulk MAB-seq samples did not recently transition between the qNSC1 and qNSC2 states. However, the detection of *some* actively demethylated astrocyte and NSC LMRs may still provide evidence for the transition between these two states. To summarize, my analysis of MAB-seq samples of NSCs and vSVZ astrocytes indicates low TET activity at astrocyte LMRs and NSC LMRs, which suggests that cells do transition between the vSVZ astrocyte and qNSC2 states, albeit at a low rate. I expect that samples isolated from mice that were recently subjected to ischemia would contain more traces of active demethylation, particularly at NSC LMRs, which become demethylated upon ischemia (Fig. 3.14).

In stark contrast to astrocyte LMRs and NSC LMRs, my analysis identified hundreds of oligodendrocyte LMRs with high 5fC/5caC content (Fig. 3.22b). This finding indicates high activity of TET enzymes in oligodendrocytes, at those regions which my previous scNMT-seq analysis revealed to be hypomethylated specifically in oligodendrocytes (Fig. 3.21).

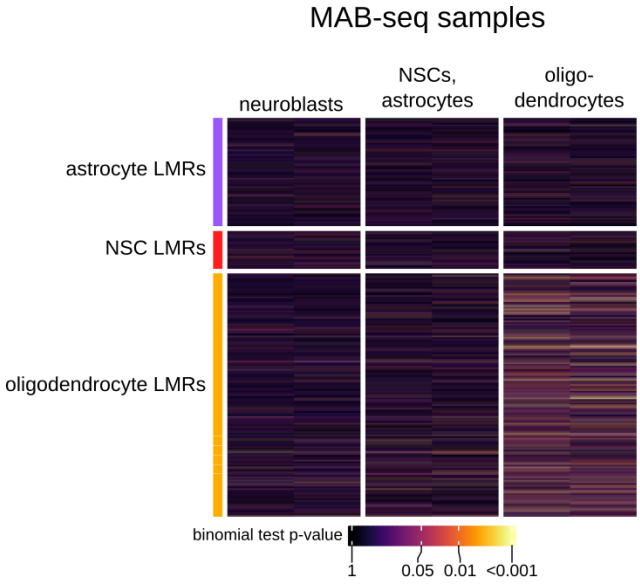

FIGURE 3.23: Active demethylation at three sets of LMRs.
Each column represents one MAB-seq replicate consisting of 12 pooled mice. For visualization only, I used a binomial test for each individual replicate here, instead of the binomial generalized linear model which incorporates data from both replicates (Fig. 3.22b).

Next, I asked whether the high levels of 5fC/5caC occur specifically at these oligodendrocyte LMRs, or whether oligodendrocytes generally possess higher 5fC/5caC levels at regions with variable methylation. To assess this, I quantified 5fC and 5caC levels in all three sets of LMRs, tested for increased 5fC/5caC levels in all three cell

populations, and visualized the resulting p-values in a heatmap (Fig. 3.23). This visualization clearly showed that only oligodendrocyte LMRs, but not astrocyte or NSC LMRs, show traces of active demethylation.

Strikingly, the top hit among those LMRs with significant evidence for active demethylation was located near *Mbp*, the gene encoding myelin basic protein. This protein is the second most abundant constituent of the myelin sheath and plays a major role in myelin sheath stabilization and formation Boggs (2006). To illustrate oligodendrocyte-specific demethylation and traces of active demethylation near *Mbp*, I visualized both the scNMT-seq methylome data and the MAB-seq data at this genomic site (Fig. 3.24). To do so, I aggregated the single-cell methylomes into pseudobulk data, based on cell type or cell state, and then displayed the average DNA methylation level of each CpG site per cell population. To visualize active demethylation, I displayed for each CpG site the proportion of reads with evidence for active demethylation (T instead of C, indicative of bisulfite-converted 5fC and 5caC).

FIGURE 3.24: *Mbp* is lowly methylated in oligodendrocytes and shows evidence for active demethylation by TET enzymes.
DNA methylation in pseudobulk scNMT-seq cell populations (top) and molecular traces of active demethylation (bottom) near the gene encoding myelin basic protein (*Mbp*). Figure and caption modified from Kremer et al. (2022b).

This visualization clearly shows that *Mbp* is demethylated specifically in oligodendrocytes, but not in other cell populations, and that the VMR detection approach that I implemented in `scbs scan` successfully captured this genomic site. Furthermore, hypomethylation near *Mbp* coincides with accumulation of several CpG sites with high MAB-seq signal. This indicates that low methylation near *Mbp* is the result

of recent TET activity in the oligodendrocyte populations that were sampled for MAB-seq.

In summary, my results suggests that oligodendrocytes employ TET enzymes to decrease DNA methylation near genes involved in myelination. The strong signal indicates recent TET activity in a large fraction of the oligodendrocytes that were sampled for MAB-seq. Why might this TET activity be required? First, some of the sampled oligodendrocytes might be immature, or they might have matured only recently. I suspect that demethylation of genes required for myelination might be a crucial aspect of oligodendrocyte maturation, since expression of these genes is required for the function of mature oligodendrocytes. Second, it is also possible that TET enzymes are not only active during oligodendrocyte maturation, but also in mature oligodendrocytes. At this stage, oligodendrocytes might use TET enzymes to modulate and fine-tune the expression of e. g. myelin components.

In line with my observations, a concurrent study by Zhang et al. (2021) found solid evidence that TET-mediated demethylation is indeed required for function and maturation of oligodendrocytes. In this study, the authors first utilized antibody-based 5hmC immunoprecipitation combined with high-throughput sequencing to identify hydroxymethylated loci in oligodendrocytes isolated from the murine cortex at postnatal day 6. Similar to my analysis of MAB-seq data, Zhang et al. (2021) identified these remains of active demethylation near genes required for myelination, such as *Mbp* and *Mag*, but also near genes associated with differentiation and maturation of oligodendrocytes and OPCs. By performing experiments on mice where *Tet1* was conditionally knocked out (cKO) in *Olig1*$^+$ oligodendrocytes and *Cspg4*$^+$ (commonly known as NG2) OPCs, the authors confirmed that *Tet1* is required for OPC and oligodendrocyte function. *Tet1* cKO mice showed impaired OPC differentiation and myelination especially at young ages, as well as behavioral defects. However, the authors also demonstrated that *Tet1* is required for the function of mature oligodendrocytes in adult brains: When inducing demyelinated lesions by injecting lysolecithin into the corpus callosum, oligodendrocytes of control mice swiftly remyelinated exposed axons, but this response was greatly impaired in *Tet1* cKO mice. Taken together, my findings and the findings reported by Zhang et al. (2021) reveal that active demethylation by TET enzymes is crucial for normal functioning of oligodendrocytes, including their maturation and myelination capabilities.

3.4.3 Active demethylation may be impaired in multiple sclerosis

My findings suggest that oligodendrocytes of adult mammalian brains employ TET-mediated active demethylation to regulate the expression genes involved in myelination. Thus, I wondered whether this mechanism might be compromised in demyelinating diseases such as multiple sclerosis (MS). To link the murine MAB-seq data set to potential epigenetic aberrations in human MS patients, I obtained data

from Huynh et al. (2014). In this study, the authors profiled CpG methylation of pathology-free regions isolated from human MS (and control) brains using Illumina 450K methylation arrays. The authors reported thousands of genomic regions that were differentially methylated in MS (**MS DMRs**). I speculated that some of these MS DMRs might be the target of TET enzymes in healthy brains, and that this active demethylation might be compromised in MS.

To test this hypothesis, I first set out to identify regions in the mouse genome that are orthologous to human MS DMRs, using the liftOver [1] tool. I then quantified traces of active demethylation at these MS DMR orthologs in the MAB-seq samples and tested whether the signal exceeds background 5fC/5caC levels (Fig. 3.25a) as done earlier for LMRs (see Fig. 3.22b). In total, 30.8% of the 315 murine MS DMR orthologs with sufficient sequencing coverage showed significant evidence for active demethylation, suggesting that many MS DMRs are indeed subject to active demethylation in healthy brains (Fig. 3.25b).

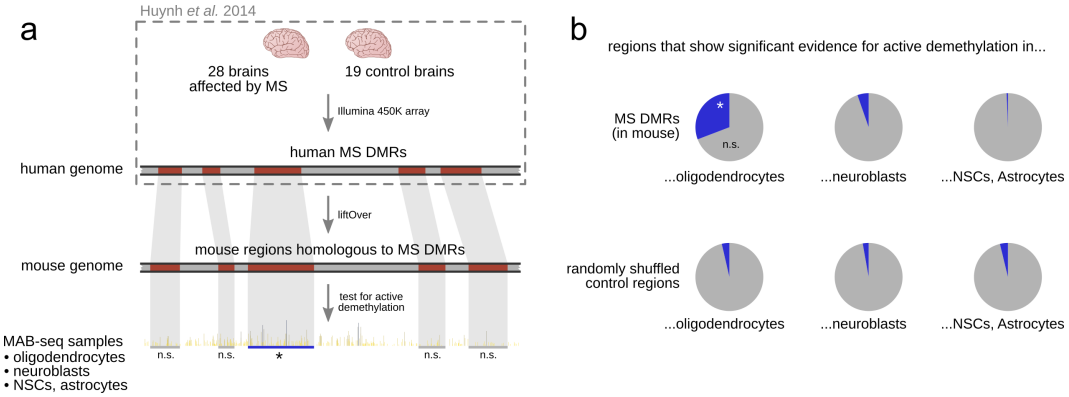

FIGURE 3.25: Active demethylation at regions orthologous to multiple sclerosis (MS) DMRs. **(a)** Quantification of active demethylation in genomic regions orthologous to MS DMRs reported by Huynh et al. (2014), using the MAB-seq data. **(b)** 30.8% of MS DMRs show significant evidence for active demethylation in murine oligodendrocytes, in contrast to other cell types (top) or random control intervals of the same size distribution (bottom). Figure and caption taken from Kremer et al. (2022b).

To ensure that the large proportion of actively demethylated regions did not arise due to random chance, I performed two control analyses: First, I also quantified active demethylation at murine MS DMR orthologs in MAB-seq samples that do not consist of oligodendrocytes, namely neuroblast samples and NSC/astrocyte samples. Both of these cell populations contained a much smaller proportion of regions with significant evidence for active demethylation, indicating that MS DMRs are targeted by TET enzymes only in oligodendrocytes, but not in other cell types. Second, I also generated a control set of random genomic regions. To do this, I used `bedtools shuffle` (Quinlan and Hall, 2010) to generate a randomly placed set of genomic intervals that matches the size distribution of MS DMR orthologs. Only a small fraction of these control regions showed significant evidence for active

[1] https://genome.ucsc.edu/cgi-bin/hgLiftOver

demethylation, which demonstrates that oligodendrocytes deposit 5fC/5caC at MS DMR orthologs specifically (Fig. 3.25b).

To summarize, my findings suggest that oligodendrocytes of adult murine brains employ active demethylation via TET enzymes to regulate myelination. Specifically, I propose that active demethylation near genes encoding myelin components such as *Mbp*, as well as other genes required for myelination, is used to regulate the expression of these genes. This mechanism might be employed in mature oligodendrocytes to fine-tune the expression of myelin components, or whenever a burst of myelin production is acutely required. The latter option might apply, for instance, when pre-myelinating oligodendrocytes mature and begin to ensheath an axon, or upon remyelination after injury. In line with this interpretation, Zhang et al. (2021) recently demonstrated that *Tet1* is required for both oligodendrocyte maturation and remyelination. Furthermore, I showed that genomic regions linked to MS are targets of active demethylation in murine oligodendrocytes, suggesting that MS might involve misregulation of myelination genes due to improper TET activity.

Chapter 4

Conclusions and open questions

4.1 Advances in single-cell methylome sequencing

Research of the last decade provided us with new sequencing protocols which allow for genome-wide quantification of DNA methylation at single cell resolution (section 1.5.2). Similarly to the early days of single-cell transcriptomics, most research in the single-cell methylome field is still focused on the development of new methods (e. g. Nichols et al., 2022; Chatterton et al., 2023) or the creation of large reference data resources (Luo et al., 2017; Liu et al., 2022). I am confident that the field will soon enter a phase where the newly developed methods are utilized to uncover new biology that was previously obscured by a lack of single-cell resolution. The first examples include the observation that commitment to the mesoderm and endoderm, but not to the ectoderm cell fate requires TET-mediated epigenome remodeling of enhancers (Argelaguet et al., 2019), as well as my finding that adult NSCs possess a pro-neurogenic methylome distinct from that of other astrocytes (Kremer et al., 2022b, section 3).

Despite these successful applications, the single-cell methylome field is still held back by two major limitations which prevent the widespread application of scBS. First, scBS data are very sparse, which hampers their analysis and interpretation. Second, the generation of scBS data is associated with prohibitive costs. High sequencing costs not only restrict such experiments to well-funded laboratories only, but also encourage shallow sequencing which furthermore increases sparsity in return. At the heart of these issues lie two fundamental limitations of scBS: First, scBS aims to achieve whole-genome sequencing coverage, which necessitates deep sequencing to achieve acceptable coverage. In contrast, scRNA-seq (e. g. Macosko et al., 2015) and scATAC-seq (Buenrostro et al., 2015) only quantify transcripts and accessible chromatin, respectively, which means that much fewer sequencing reads are required to generate informative data sets. Second, scBS protocols struggle with DNA degradation as an inevitable but undesirable side-effect of bisulfite conversion (Grunau et al., 2001), which furthermore exacerbates sparsity.

In the long run, bisulfite-free approaches for genome-wide methylome profiling are a promising alternative to current protocols. Such enzyme-based protocols

were recently described in Liu et al. (2019), Vaisvila et al. (2021) and Wang et al. (2023). But even with currently available approaches based on bisulfite conversion, it is possible to reduce sequencing costs and sparsity. To achieve this, I propose to sacrifice whole-genome coverage in favor of higher sequencing coverage at informative loci when possible. This can be achieved, for instance, by enriching for reads that are rich in CpG sites (Guo et al., 2015) or by enriching for reads which correspond to gene bodies, regulatory elements or other regions of interest. Naturally, a downside of such approaches is that any informative methylation pattern that might occur outside the targeted regions will inevitably be missed. For this reason, it might be useful to combine both genome-wide and targeted approaches: In a pilot experiment, the tissue or cell lineage of interest may be profiled with whole-genome scBS, which allows for the discovery of informative genomic regions (i. e. DMRs or VMRs identified with `scbs`). These costly experiments might be carried out by large consortia or well-funded laboratories which aim to generate methylome cell atlases, i. e. large reference data sets. Follow-up experiments may then be carried out with more cost-effective, targeted approaches such as the recently developed scTAM-seq (single-cell targeted analysis of the methylome), which allows for profiling of up to 650 selected CpG sites in thousands of cells and achieves over 90% sequencing coverage (Bianchi et al., 2022). This strategy not only reduces sequencing costs, but also facilitates the study of informative regions at a higher resolution in a greater number of cells.

4.1.1 Analysis of single-cell methylation data

As outlined in section 1.5.4, the interpretation of scBS experiments is currently very challenging since scBS data is extremely sparse and fundamentally different from scRNA-seq data. Furthermore, there is a lack of dedicated software for its analysis. I am confident that my software `scbs` (Kremer et al., 2022a) addresses this bottleneck as it is not only easy to use, but also improves results and facilitates the discovery of putative regulatory genomic sites (section 2.3). Thus, my contributions have significantly eased the processing, analysis and interpretation of scBS data while also providing more accurate results.

Despite these significant advances, scBS data analysis methods are still in their infancy and further contributions are needed to elevate scBS data analysis methods to the more mature level of scRNA-seq or scATAC-seq methods. First, most scRNA-seq tools are designed for use in an interactive computing environment. In contrast, since `scbs` comprises many computationally demanding data processing steps, its features are currently only accessible from the command line. To facilitate interactive data exploration and visualization, for instance to inspect genomic loci such as DMRs, a future release of `scbs` might contain methods to support its use in interactive computing environments.

Second, there is an urgent need for the development of scBS analysis methods designed to handle data sets that comprise multiple replicates, i. e. data obtained from several individuals. Specifically, scBS currently lacks methods to account for technical batch effects. Furthermore, just like many methods for differential gene expression analysis in scRNA-seq (Nguyen et al., 2023), the DMR detection approach implemented in `scbs diff` currently considers single cells as experimental units (Braun, 2023). In the same vein, the sliding window approach implemented in `scbs scan` likely captures some genomic intervals that have a high variance of methylation due to technical batch effects or differences between individuals, and not due to differences between cell states. Future methods for single-cell DMR detection should aim to incorporate cell-wise sample labels into differential methylation testing, to ensure that reported DMRs occur robustly in all sampled individuals.

Finally, due to the ever-increasing size of single-cell sequencing data sets, it is important to ensure that scBS analysis methods are computationally efficient and scalable. To ensure this, I used Numba (Lam et al., 2015) to parallelize and accelerate crucial inner-loop code, and furthermore implemented an option to parse large data sets in chunks. To uncover any additional bottlenecks in the analysis of large data sets, `scbs` should be benchmarked on real or simulated scBS data comprising tens of thousands of cells.

To summarize, both sequencing protocols and data analysis methods for single-cell methylome analysis are still in their infancy. My work outlined in this thesis not only offers new tools for the analysis of such data, but also provides a good example of how such data can be used to uncover new biology.

4.2 Emerging roles of DNA methylation

4.2.1 DNA methylation is dynamic in adult tissues

Historically, DNA methylation has been viewed as a static epigenetic mark that is faithfully copied in subsequent cell divisions (Mattei et al., 2022). Only in the past decade, this static view was overturned by novel insights including the discovery of TET enzymes (Tahiliani et al., 2009), the identification of tissue-specific methylation differences using WGBS (Ziller et al., 2013), as well as the in-depth characterization of methylation dynamics in embryonic development (Argelaguet et al., 2019; Parry et al., 2021, also see section 1.4).

My observations not only provide additional evidence for this emerging view that DNA methylation is dynamic, but furthermore demonstrate that this concept may also apply to adult tissues and not just embryonic development: First, my identification of methylation and demethylation events in the healthy adult NSC lineage strongly suggests that adult NSC differentiation is accompanied by dynamic changes in DNA methylation (section 3.1.6). Second, my analysis of scNMT-seq data

derived from post-ischemic mice suggests that some cells of adult tissues undergo methylome remodeling in response to environmental stimuli such as ischemia (section 3.3). Third, my analysis of MAB-seq data suggests that dynamic changes in DNA methylation might be required for myelination in adult brains (section 3.4, Fig. 4.1), an assertion that was recently confirmed by Zhang et al. (2021).

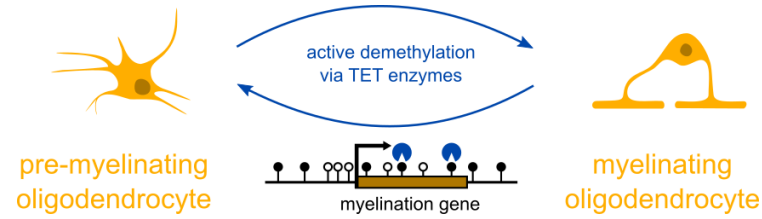

FIGURE 4.1: Summary and interpretation of my characterization of oligodendrocyte methylomes.
Oligodendrocytes possess unique DNA methylomes, characterized by low methylation levels near genes required for myelination. My analysis of MAB-seq data revealed that these sites are the target of TET enzymes. I propose that this active demethylation is an important regulator of myelination. Figure taken from Kremer et al. (2022b).

Taken together, these findings demonstrate that DNA methylation is dynamic in various cell types and functional contexts of the adult brain. I suspect that future studies will uncover similar methylation dynamics in other adult tissues.

4.2.2 A permissive methylome may contribute to stemness of NSCs

My results outlined in chapter 3 demonstrate that NSCs possess a unique DNA methylome, which sets them apart from other astrocytes of the vSVZ and striatum and likely contributes to their neurogenic capabilities. Furthermore, my analyses suggest that ischemia causes astrocytes of the vSVZ and striatum to enter the neurogenic lineage, and that this transition is accompanied by methylome remodeling (Fig. 4.2).

How exactly might the unique methylome of NSCs contribute to stemness? Already in 1995, Adrian Bird speculated that one function of DNA methylation might be to reduce transcriptional noise by acting as a global repressor that decreases transcription of most of the genome. An extension of this concept was later formulated to describe embryonic development, where it was reasoned that DNA methylation at regulatory elements is used to repress genes that are only required in specific cell types or tissues (Meissner et al., 2008; Ziller et al., 2013; Suzuki et al., 2017; Reizel et al., 2018; Argelaguet et al., 2019; Edrei et al., 2021; Cedar et al., 2022).

In line with this notion, I view the DNA methylome as a repressive force that imposes constraints on the transcriptome space that a given cell can occupy. Thus, a restrictive methylome might lock a cell into a specific cellular fate by silencing those genes that would be required to (de)differentiate to an alternative fate. In contrast, a permissive methylome may promote the acquisition of new cell fates by enabling the expression of genes required for such transitions. Permissive methylomes might thus be characterized by hypomethylation near genes required for cell differentiation,

and possibly also by hypomethylation near genes that are functionally important for the terminally differentiated endpoint.

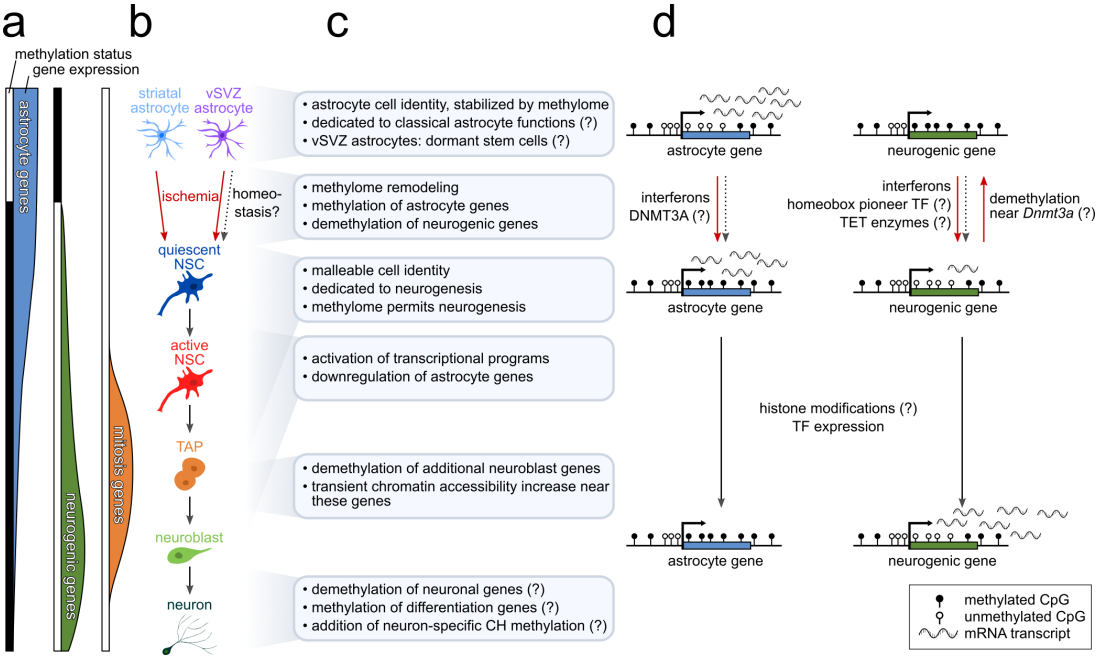

FIGURE 4.2: Summary and interpretation of my main findings.
(a) Schematic illustration of changes in gene expression and DNA methylation in the NSC lineage (depicted in **b**). Genes required for astrocyte function are expressed in both astrocytes and qNSCs and gradually fade out. In contrast, VMRs near these genes are methylated (black) in the entire NSC lineage, but hypomethylated (white) in astrocytes including vSVZ astrocytes (qNSC1). Genes required for neurogenesis show an inverse pattern, characterized by VMR methylation in astrocytes and gradual upregulation at various time points of NSC lineage progression. Mitotic genes are induced regardless of their epigenetic state, as they are constitutively demethylated. **(b)** Schematic depiction of the NSC lineage including lineage transitions that occur under homeostatic conditions (gray arrows), lineage transitions which I hypothesize (gray dotted arrow), and lineage transitions which I observed upon ischemia (red arrows). Only forward transitions (i. e., neurogenic transitions) are depicted. vSVZ astrocytes and qNSC cells correspond to the qNSC1 and qNSC2 states described in Kalamakis et al. (2019), respectively. **(c)** Annotation and interpretation of epigenetic change which I observed in the NSC lineage. Speculative statements are marked with question marks. **(d)** Schematic depiction of astrocyte / neurogenic genes and their methylation and expression status at three points of the NSC lineage. I frequently observed methylation change downstream of the promoter region, in the first intron. Putative regulatory factors of expression and methylation change are denoted by text. Arrows colored as in **b**, with one backward arrow to signify a potential negative feedback loop by demethylation near *Dnmt3a*. Some figure elements were taken from Kremer et al. (2022b).

In the case of adult NSCs, I suggest that their DNA methylome does not primarily promote stemness by directly affecting gene expression. This is evidenced by my observation that DNA methylation does not mirror gene expression, as seen very clearly in Fig. 3.10c-d. Rather, hypomethylation of NSC LMRs might resemble an open door that permits NSCs to drift along the NSC lineage trajectory, toward the neuroblast state. In contrast, from the perspective of a common parenchymal astrocyte, the door that leads to the neuroblast state is locked by DNA methylation, and thus parenchymal astrocytes are stably locked in their fate (Fig. 4.3).

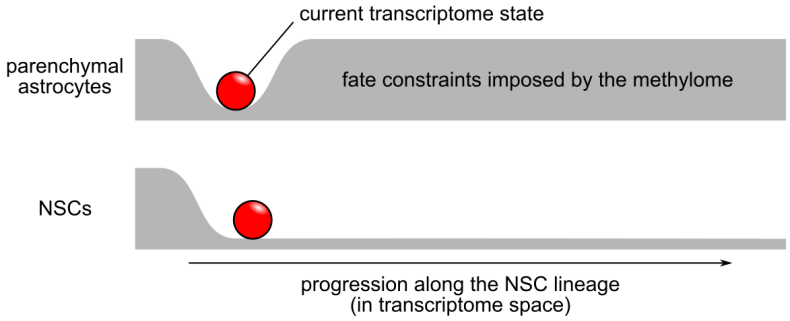

FIGURE 4.3: The methylome fate-locks common parenchymal astrocytes but permits differentiation of NSCs.
I propose that DNA methylation locks common parenchymal astrocytes, including striatal astrocytes and vSVZ astrocytes (qNSC1) in their astrocyte fate by repressing genes required for neurogenesis. NSCs, however, possess a permissive methylome that permits their progression along the neurogenic lineage. This methylome is characterized by the absence of repressive DNA methylation near neurogenic genes, which means that NSCs possess the epigenetic license to express these genes.

As reviewed by Stricker and Götz (2018), evidence to support the idea that DNA methylation guides neural fate decisions is currently very limited, although the authors acknowledge that DNA methylation might be important for differentiation and neuronal maturation. While I did not investigate whether methylation might predict alternative fate decisions, my results show convincing evidence that specific methylation patterns are required for neurogenesis. Furthermore, my analysis of single-cell methylomes obtained from post-ischemic mice (Fig. 3.14) demonstrates that injury-induced exit from the common parenchymal astrocyte fate is accompanied by methylome remodeling, which provides new evidence for the idea that DNA methylation is an important regulator of cell fate. To irrevocably demonstrate whether DNA methylation plays a causal role in neurogenesis, future studies should aim to manipulate DNA methylation using targeted epigenome modifiers (Nakamura et al., 2021; Nuñez et al., 2021), or aim to prevent injury-induced epigenome remodeling using conditional knockout or inhibition of the DNA methylation machinery. A criticism voiced by Stricker and Götz is that previous studies found only poor correlation of DNA methylation and gene expression. I am confident that my model depicted in Fig. 4.3 offers a convincing explanation why this might be the case, as DNA methylation may not acutely control gene expression, but may rather permit the expression of genes required at future cell states.

4.3 Common parenchymal astrocytes may act as a backup stem cell pool

After decades of research on astrocyte sub-populations of the adult vSVZ, astrocyte heterogeneity of this region is still incompletely understood (section 1.1.4). One of the challenges in this field is to reconcile classical astrocyte sub-state definitions, derived from imaging and immunostainings, with those cellular sub-states determined with

scRNA-seq (section 1.1.5). Due to the uncertainties associated with manual labeling of scRNA-seq clusters, the most basal (i. e., most astrocyte-like) scRNA-seq cluster was either interpreted as parenchymal non-neurogenic astrocytes (e. g. Cebrian-Silla et al., 2021, who labeled this cluster "astrocytes" and discarded it from further analysis) or as a population of deeply dormant NSCs (e. g. Kalamakis et al., 2019, who labeled this cluster "qNSC1"). However, it is challenging to find support for either of these proposed functions using scRNA-seq data alone, as transcriptomes of this cluster are highly similar to those of qNSC2 cells.

In the work presented here, I was finally able to identify strong molecular differences between qNSC1 cells (vSVZ astrocytes) and qNSC2 cells: As clearly visible in Fig. 3.6, qNSC1 transcriptomes are highly similar to qNSC2 transcriptomes, while DNA methylation at VMRs is not correlated at all between these two populations. Classical astrocyte marker genes, including various transmembrane proteins required for parenchymal astrocyte function, are only hypomethylated in qNSC1 cells (Fig. 3.10). For this reason, I propose that qNSC1 cells are primarily dedicated to parenchymal astrocyte functions and tissue homeostasis of the vSVZ, which is why I labeled them "vSVZ astrocytes". In contrast, I propose that qNSC2 cells represent the *bona fide* qNSCs that regularly contribute to neurogenesis. Nonetheless, my results also demonstrated that the parenchymal astrocyte fate of qNSC1 cells is not set in stone, as these cells underwent methylome remodeling and became neurogenic upon ischemia. Thus, vSVZ astrocytes (qNSC1) act as a deeply quiescent backup NSC population that is recruited in times of need, for instance upon ischemic injury. Why might astrocytes, and not other resident brain cells, evolve to act as a backup stem cell pool? It might be that astrocytes are particularly suited for this purpose due to their innate ability to sense and act upon a diverse range of stimuli, including injury (Sofroniew and Vinters, 2010; Sloan and Barres, 2014; Linnerbauer and Rothhammer, 2020, section 1.2).

It remains to be investigated whether recruitment of qNSC1 cells to the neurogenic lineage also occurs in homeostasis. A promising strategy to test this hypothesis is the use of clonal lineage-tracing combined with scRNA-seq (reviewed in Kester and Van Oudenaarden, 2018; Chen et al., 2022), or even single-cell multi-omics, to reconstruct lineage relationships among cells of the vSVZ. This approach might unveil some qNSC1 and qNSC2 cells that share a recent common ancestor. A downside of this approach is that evidence for shared ancestry does not inform us about the directionality of the lineage relationships, i. e. it might not be possible to distinguish qNSC1 $\rightarrow$ qNSC2 from qNSC2 $\rightarrow$ qNSC1 transitions. For this reason, it might be worthwhile to supplement clonal lineage tracing experiments with data obtained from molecular recorders, which record transcriptional activity over a defined period of time and thus make it possible to gain insight on past transcriptional activity of cells (Masuyama et al., 2022; Boers et al., 2023). Such data might reveal qNSC2 cells which had the transcriptome of a qNSC1 cell in the past. Interpreting these data

might be challenging, however, since qNSC1 and qNSC2 cells share highly similar transcriptomes.

To avoid the abovementioned ambiguity, I suggest to also consider the methylome when inferring past cell states. A promising strategy to achieve this is to detect past epigenome remodeling events at single-cell resolution, for instance using the recently described single-nucleus 5hmC sequencing (snhmC-seq), which allows for the detection of 5hmC, i. e. the detection of recent TET activity (Fabyanic et al., 2021). However, the scNMT-seq data set discussed in this thesis might already contain some evidence for homeostatic qNSC1 $\rightarrow$ qNSC2 transitions: My methylome analysis of cells isolated from the post-ischemic vSVZ revealed a small number of cells with an astrocyte methylome but an NSC transcriptome (Fig. 3.14b), suggesting that these cells recently transitioned from the qNSC1 to the qNSC2 state but that methylome remodeling had not occurred yet. Albeit at a minuscule quantity, I also observed such putative transitioning cells in the naive vSVZ. While a much greater sample size is required to draw definitive conclusions, this observation may provide some limited evidence that qNSC1 $\rightarrow$ qNSC2 transitions occur in homeostasis, albeit at a lower frequency compared to ischemia.

Further evidence for the latent stem cell potential of qNSC1 cells is the observation that GLAST$^+$/Prominin-1$^+$ cells, as isolated by FACS, populate both the qNSC1 and qNSC2 scRNA-seq clusters (Llorens-Bobadilla et al., 2015). As the first two scNMT-seq plates of our scNMT-seq data set also contain GLAST$^+$/Prominin-1$^+$ cells isolated by FACS, I was able to confirm this observation (Kremer et al., 2022b, Supplementary Table 1). The presence of Prominin-1, a protein found in the cilia of radial glial cells and adult NSCs (Mirzadeh et al., 2008; Beckervordersandforth et al., 2010), suggests that qNSC1 cells possess the characteristic ciliated apical process also found in qNSC2 cells, a feature which would set them apart from common parenchymal astrocytes of other brain regions.

My investigation of single-cell triple-omic data of the vSVZ made it possible to attach the common parenchymal astrocyte label to specific cells with a higher confidence, as methylation data provides an additional, more informative molecular layer for this task. Nonetheless, to date some astrocyte sub-populations previously described in the vSVZ, such as type B2 cells (Doetsch et al., 1997), were not yet pinpointed in scRNA-seq data sets of the vSVZ. I anticipate that advances in spatial transcriptomics (reviewed in Cheng et al., 2023) will ultimately make it possible to throw a bridge between classical imaging-based nomenclature of vSVZ cell populations and those that were recently used to label scRNA-seq data sets.

Appendix

Chapter M

Extended methods

As part of my doctoral research, I developed, implemented and applied novel computational methods for the analysis of scBS data. For a detailed breakdown of these methods see chapter 2, as well as Kremer et al. (2022a) for mathematical details. For details on laboratory experiments (all of which were performed by Dr. Santiago Cerrizuela and colleagues, see External Contributions), as well as details on the mice used for these experiments, see the methods section and supplementary material of Kremer et al. (2022b). In this extended methods section, I will provide additional details on my analysis of scNMT-seq data and MAB-seq data.

M.1 Processing of single-cell multi-omic sequencing data

I processed single-cell triple-omic sequencing data as previously described in Cerrizuela et al. (2022). Specifically, I processed the transcriptomic portion of the sequencing data with the zUMIs pipeline version 2.9.4f, which uses STAR version 2.7.3a (Dobin et al., 2013) internally for read mapping. I mapped to the mouse genome build GRCm38/mm10, using the *Mus musculus* genome annotation of Ensembl Release 102 (Martin et al., 2023).

To process genomic sequencing reads, I first trimmed these reads in paired-end mode with Trim Galore [1] and then mapped them to the mouse genome (GRCm38) using the single-end, non-directional mode of Bismark version 0.22.3 (Krueger and Andrews, 2011). I then used the `deduplicate_bismark` script, included in Bismark software, to filter PCR duplicates. After merging single-end alignments per cell, I used the `coverage2cytosine` script with the `--nome-seq` option to generate reports separate reports for CpG and GpC methylation. I used these reports as input for downstream analysis with `scbs`.

M.2 Quality filtering of single-cell multi-omic data

To assess data quality, I computed three quality metrics for each cell: The number of observed genes, the number of observed

[1] https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/

GpC sites (Fig. 3.2). Both GpC and CpG methylation data was processed with `scbs` version 0.3.2. To ensure that only non-empty wells and high quality cells were part of the analysis, I discarded all cells with less than 2500 observed genes and/or less than 50 000 observed CpG sites. Setting a separate threshold for the number of observed GpC sites was not necessary, since this value is highly correlated with the number of observed CpG sites as both metrics depend on sequencing coverage.

To furthermore assess the quality of the epigenomic data, I used `scbs profile` to visualize the average DNA methylation and chromatin accessibility around TSSs for each cell. An example TSS profile of a single high-quality neuroblast is depicted in Fig. 3.3. Since TSSs are known to be more accessible and less methylated than their surroundings, I manually discarded a small number of cells that clearly violated these criteria. Note that I later also identified off-target cells such as ependymal cells using our single-cell transcriptomic data. These cells were then discarded from all three parallel (triple-omic) data sets. For a complete list of all cells that passed quality filtering, as well as additional metadata on each cell, see the supplementary material of Kremer et al. (2022b).

M.3 Transcriptome integration

To process scRNA-seq data, I used Seurat version 4.1.0 (Stuart et al., 2019). I used Seurat to integrate the single-cell triple-omic data with a larger scRNA-seq data set from Carvajal Ibañez et al. (2023), only using the wild type cells from this source. Specifically, I used the default Seurat normalization workflow for both data sets, identified 3000 highly variable genes, and then used Seurat's `FindIntegrationAnchors` and `IntegrateData` functions with 30 dimensions. I then used the integrated data for PCA on 30 PCs, followed by UMAP.

M.4 Dimensionality reduction

To process the quality-filtered epigenomic data, I first used `scbs smooth` with a bandwidth of 1000 bp for CpG data and a bandwidth of 500 bp for GpC data. I then used `scbs scan` with default options for VMR detection and VAR detection. In the case of CpG data I used the default bandwidth of 2000 bp; for GpC data I used 1000 bp. To obtain matrices for dimensionality reduction, I used `scbs matrix` to quantify CpG and GpC and methylation at VMRs, VARs and promoters (defined as ±1000 bp intervals around TSSs). For CpG data, I used the shrunken means of the residuals of VMRs, using only VMRs that were observed in at least 20% of cells. For GpC data, I used the shrunken means of the residuals of both VARs and promoters as features, using only features observed in at least 40% of cells. As artificially induced GpC methylation showed technical variation between cells, I centered the rows (cells) of the GpC matrix. I next centered the columns (features) of both matrices and separately subjected them to a modified implementation of PCA that handles

missing values (proposed by Simon Anders and described in Kremer et al., 2022a). For visualization, I used UMAP (McInnes et al., 2018) on the top 15 PCs, excluding PC 5 in the case of CpG data as it was strongly associated with data quality.

M.5 Pseudotime and cell type annotation

To order cells isolated from naive mice along the NSC differentiation trajectory using information from all three molecular layers, I combined MOFA+ version 1.6.0 (Argelaguet et al., 2020) with slingshot version 2.4.0 (Street et al., 2018). Specifically, I used the CpG and GpC matrices described in the previous sections, as well as the scRNA-seq count matrix of 3000 highly variable genes (transformed with SCTransform, Hafemeister and Satija, 2019) as input for MOFA+. Using MOFA+, I reduced this data to 15 dimensions called factors. I excluded factors 4 and 11 as they correlated with cell quality and then used Leiden clustering (Traag et al., 2019) on the remaining factors. I then subjected the factors to pseudotime inference with slingshot, defining the cluster corresponding to astrocytes as start point and the cluster corresponding to neurons as end point. To annotate cell types, I first visualized the expression of well-known NSC lineage markers in pseudotime (Fig. 3.4a). I then manually segmented the pseudotime values to derive groups of cells which correspond to NSC cell states described in the literature. For pseudotime analyses which also include cells isolated from post-ischemic mice, I repeated the identification of Leiden clusters and pseudotime inference with slingshot in integrated PCA transcriptome space.

M.6 Correlating gene expression with epigenetic features

I used the R function `cor.test` to separately test for significant Pearson correlation between CpG and GpC methylation values of nearby genomic intervals including promoters, VMRs and VARs. I used log-normalized expression values and the shrunken means of the residuals in this correlation analysis, considering only epigenomic features which had sequencing coverage in at least five cells. To correlate VMRs and VARs with gene expression, I correlated their methylation values with the expression of the nearest gene. Nearby genes were determined with bedtools version 2.27.1 using the command `bedtools closest -D 'b' -a VMRs.bed -b genes.bed`. When multiple VMRs/VARs intersected the same gene, I matched only the largest VMR/VAR with that gene. I corrected the p-values obtained with `cor.test` for multiple testing with the Benjamini-Hochberg method.

To quantify how frequently putative regulatory VMRs or VARs intersect with various genomic features, I selected only those features that showed a significant (adjusted p-value < 0.05) correlation with gene expression. I then used ChIPseeker version 1.32.0 (Yu et al., 2015) to quantify overlaps with gene features, using the options `tssRegion=c(-1000, 1000)` and `overlap="all"`. To quantify intersection

of VMRs and VARs with putative regulatory elements, I obtained mouse genome coordinates of candidate cis-regulatory elements (cCREs) [2] annotated by the ENCODE consortium (Moore et al., 2020). I quantified overlaps with these features with the `mergeByOverlaps` function of the R package GenomicRanges version 1.48.0 (Lawrence et al., 2013).

M.7 Epigenetic changes in pseudotime

To assess epigenetic changes occurring along the NSC differentiation trajectory, I first binned cells in pseudotime. In this binning, I ensured that each bin contains 10 (or close to 10) cells and that each bin contains only cells of the same cellular state and tissue. I used this binning for all heatmaps of naive wild type data. Log-normalized gene expression values of the top 3000 highly variable genes, as well as methylation and accessibility values of VMRs and VARs, were averaged per bin. To visualize changes of these three molecular layers in pseudotime, I computed and visualized the Pearson correlation of all bins with the R package ComplexHeatmap (Gu et al., 2016).

To furthermore quantify methylation and demethylation events in the NSC lineage, I focused on VMRs with sequencing coverage in at least 100 NSC lineage cells. I fit a step function to the methylation values (shrunken mean of the residuals) as a function of pseudotime, separately for each of these VMRs. This step function comprises the parameter s, which represents a change point in pseudotime, as well as two constant values corresponding to the mean methylation values before and after s. To find the most likely value for s for each VMR, I minimized the sum of squared residuals over these parameters. If this procedure resulted in a 15% lower sum of squares than a constant fit without a step, I counted the VMR as a (de)methylation event that occurs at pseudotime point s. I visualized these values for s in a histogram in Fig. 3.7a.

M.8 Epigenetic changes near marker genes

To identify representative marker genes for all cellular states and cell types, I used the two-sided Wilcoxon rank sum test. Specifically, I tested the log-normalized gene expression values in the cell group of interest against the expression values of all other cells. I restricted the resulting gene list to genes with a Benjamini-Hochberg adjusted p-value < 0.05 and to genes that intersect a VMR. Of these, I defined the top 100 most overexpressed genes, according to $\log_2$-fold change, as marker genes. For each marker gene, I quantified CpG methylation of the intersecting VMR, as well as the TSS±1000 bp promoter region of the gene. Log-normalized expression values and cell-wise methylation percentages were averaged per pseudotime bin and visualized

[2] downloaded on 2021-08-03 from the registry of cCREs V3 at `https://screen.encodeproject.org/`

in Fig. 3.9. I repeated the same procedure with VARs instead of VMRs, and GpC methylation instead of CpG, to assess dynamics of chromatin accessibility. For a complete list of the identified markers genes, as well as the genomic coordinates of overlapping VMRs, see the supplementary material of Kremer et al. (2022b).

M.9 Enrichment of transcription factor binding site motifs

As the PCA of VMR methylation values separated oligodendrocytes from other cells on PC 1, I selected the 5000 VMRs with the highest loading of PC 1 to identify TFBS motifs that are lowly methylated specifically in oligodendrocytes. To perform motif enrichment, I used HOMER version 4.4 (Heinz et al., 2010) in conjunction with the motif database JASPAR2022 (Castro-Mondragon et al., 2022) using the following command:

```
findMotifsGenome.pl VMRs.bed mm10r output/ -len 5,6,7,8,9,10,11,12 -size given -mcheck
↪  JASPAR2022.db -mknown JASPAR2022.db -bg JASPAR2022.db
```

The same strategy was used to identify TFBS motifs enriched in VMRs which are hypomethylated in other cell states. These VMRs were identified using PC 2, as PC 2 separates astrocytes/qNSC1 cells from the rest of the NSC lineage. Accordingly, NSC lineage VMRs were identified by selecting the top 5000 VMRs with the highest PC 2 loading, while astrocyte VMRs were identified by selecting the bottom 5000.

M.10 Detection of LMRs and GO term enrichment

To identify genomic regions that are lowly methylated in one group of cells compared to all other sampled cells (lowly methylated regions; LMRs), I considered the set of previously identified VMRs and restricted the analysis to cells isolated from naive mice only. For each VMR, I used Wilcoxon's two-sided rank sum test to test for differences in DNA methylation, as measured by the shrunken means of the residuals, between cells in the group and all other cells in the quality-filtered data set. I assigned the term "LMR" to VMRs that were more lowly methylated in the cell group of interest and had a Benjamini-Hochberg adjusted p-value below 0.05. I limited all LMR heatmaps and volcano plots to those LMRs whose genome coordinates intersect with a gene body, to facilitate interpretation of LMRs. For a complete list of the genomic coordinates of astrocyte LMRs and NSC LMRs, see the supplementary material of Kremer et al. (2022b).

I used the GREAT tool (McLean et al., 2010) with the option "basal plus extension", specifying a constitutive 20 kb downstream regulatory domain and up to 1000 kb maximum extension, to identify GO terms associated with cell type-specific LMR sets. Note that this LMR detection approach precedes the development of a more

appropriate method to detect DMRs in single-cell methylation data (Braun, 2023), which I employed in a later analysis described in the next section.

M.11 Detection of post-ischemic methylation and expression change

I acquired three sets of reactive astrocyte marker genes from (Liddelow et al., 2017). These three sets (pan-reactive, A1 and A2) were used to derive reactive astrocyte gene expression signatures by averaging the log-normalized gene expression values of the respective markers for each cell.

To visualize methylation differences in UMAPs, I calculated a cell-wise methylation score that denotes whether a cell has an NSC-like or an astrocyte-like methylome. This methylation score is simply the difference between the mean methylation of all astrocyte LMRs and the mean methylation of all NSC LMRs, using the raw methylation fraction (percentage) and not the shrunken means of residuals to aid interpretability.

To detect DMRs induced by ischemia, I restricted the analysis to qNSC2 cells, aNSCs and TAPs of the vSVZ. Using `scbs diff` (Braun, 2023), I tested cells isolated from mice 2 dpi against cells isolated from naive mice. I used the same procedure to test cells isolated 21 dpi against cells from naive mice:

```
scbs diff --bandwidth 2000 --stepsize 100 --threshold 0.05 --min-cells 6 data_dir/
↪  group_labels.txt DMRs.bed
```

M.12 Analysis of MAB-seq data

I obtained bulk MAB-seq data generated by Dr. Sascha Dehler, Dr. Santiago Cerrizuela and Dr. Dieter Weichenhan. Of this data set, I used four bulk MAB-seq samples: Two MAB-seq replicates of $O4^{+}$ oligodendrocytes and two MAB-seq replicates of $GLAST^{+}/CD9^{high}$ vSVZ astrocytes and NSCs. I first trimmed MAB-seq reads with Trim Galore version 0.4.5 in paired-end mode. I mapped the trimmed reads to the mouse genome build GRCm38 with bwa-meth version 0.2.2 (Pedersen et al., 2014) using default options. I then discarded all reads with a mapping quality below 30 and/or with an unmapped mate. PCR duplicate reads were discarded with the `MarkDuplicates` script version 2.18.14, which is part of Picard tools [3]. To quantify methylation I used MethylDackel [4] version 0.3.0. Common mouse single nucleotide polymorphisms were filtered with BS-SNPer (Gao et al., 2015) with options `--minhetfreq 0.1 --minhomfreq 0.85 --minquali 20 --mincov 10 --mapvalue 30`.

[3] https://broadinstitute.github.io/picard/
[4] https://github.com/dpryan79/MethylDackel

To answer the question whether LMRs which I had previously detected in scNMT-seq data are the target of active demethylation by TET enzymes, I used `bedtools intersect` version 2.27.1 to extract only those CpG sites from the methylation reports generated by MethylDackel that intersect with an LMR. I then restricted the analysis to only those LMRs with a mean sequencing coverage greater than 5 in both MAB-seq replicates. These LMRs were separately tested for evidence of active demethylation in the two astrocyte/NSC replicates, and in the two oligodendrocyte replicates. Specifically, to make use of the two bulk MAB-seq replicates per cell population, I used a binomial generalized linear model (glm) with a logit-link function to test the alternative hypothesis that the average methylation in a given LMR is lower than the background methylation level. I used the mean genome-wide methylation level of the two samples as the background methylation level. I adjusted the resulting p-values for multiple testing with the Benjamini-Hochberg method.

To test MS DMRs for evidence of active demethylation, I obtained genome coordinates of MS DMRs from the supplementary materials of Huynh et al. (2014). I used the liftOver tool [5] with the option `-minMatch=0.1` to first convert the human genome (hg18) coordinates to homologous mouse genome coordinates (mm9). I then converted the mm9 coordinates to mm10 coordinates using liftOver with default options. The obtained regions correspond to regions in the mouse genome that are homologous to MS DMRs reported by Huynh et al. (2014). I tested these regions for evidence of active demethylation using the binomial glm approach described above.

[5] `https://genome.ucsc.edu/cgi-bin/hgLiftOver`

Acknowledgements / Danksagung

First and foremost, I would like to thank my PhD supervisors Ana Martin-Villalba and Simon Anders for their outstanding mentorship and guidance throughout the past years. I am immensely happy that I moved to Heidelberg because your collaboration provided an exceptionally rare, fruitful environment that allowed me to work on both exciting biology and exciting computational methods.

Thanks to you, Ana, for your immense creativity, your exceptional supervision, and your continuous encouragement. Even with your busy schedule, you were always very approachable and found the time to talk with me. I am deeply grateful for everything you taught me, and for your enthusiasm for my results which really motivated me to strive for greatness. I would also like to thank you for the many amazing opportunities you provided to me when you sent me to conferences to speak on your behalf.

Thanks to you, Simon, for your excellent guidance and for the many hours of detailed explanations and brainstorming sessions. These sessions were intense but taught me a whole new way of thinking and gave me the confidence and tools to not only use computational methods, but also to design them. Thank you for this invaluable knowledge, but also for helping me keep my sanity in a project with so many possible directions. With your calm and rational suggestions, you helped me sharpen my focus.

I am also grateful to my TAC members Kyung-Min Noh and Matthias Schlesner, who provided invaluable feedback in a supportive and encouraging environment and offered to join my PhD defense committee. Along the same line, I would also like to thank Jan Lohmann for joining my defense committee on a rather short notice.

I am immensely indebted to all my colleagues who generated this amazing, one-of-a-kind data set which I had the pleasure of analyzing here. Most importantly, I would like to thank you, Santi, for taking on the massive challenge of mastering the obscure scNMT-seq protocol, for your diligence and immense scientific output, and for your infectious cheerfulness. Without your and Ana's initiative to go for this protocol and perform new experiments, I would probably still be analyzing bulk WGBS data. And I also want to thank you, Hadil, for stepping into this project at such a challenging phase of your personal life, and for delivering results that are absolutely critical for the paper revisions. I was also blessed with two excellent students, Martina Braun and Leonie Küchenhoff, who surpassed all of my expectations and made several crucial contributions to `scbs`. Lastly, I want to thank all of the diligent contributors who helped generate the scNMT-seq and MAB-seq data sets, including Mohammed, Tobias, Jannes, Sascha, Aylin, as well as our collaborators Dieter Weichenhan and Christoph Plass.

Finally, I want to stop talking about research and focus on all my colleagues, friends and family that supported me in the past years. First, I want to extend a heartfelt "thank you" to all members of the Martin-Villalba and Anders labs, who made me feel right at home from the very first day when I moved to Heidelberg. Thanks for the countless suggestions and for the supportive and accepting work environment, but also for all the fun we had in the past years. I will always remember the fun games and activities inside and outside the lab, the traditional Friday 5 PM "meetings", and all the friendships that formed during my PhD.

Last but not least, I would like to thank my family.
Vielen Dank, Mama und Papa, dafür dass ihr mich bei jeder meiner Entscheidungen unterstützt habt. Dafür, dass ihr mir mein Studium einschließlich der Zeit in Sheffield ermöglicht habt und dafür, dass ihr mich immer ermutigt habt meine Träume zu verwirklichen – selbst wenn es darum ging, Aachen zu verlassen um ein Fach zu studieren, mit dem ihr beide wenig anfangen könnt. Danke auch an dich, Kira, für die vielen schönen gemeinsamen Erinnerungen und deine Gastfreundschaft in Berlin.

Und zu guter Letzt: Vielen Dank auch an dich, Caro, für deine Liebe, deine Unterstützung und dafür, dass du mir nun schon seit fast acht Jahren mein Leben versüßt.

External Contributions

Dr. Santiago Cerrizuela performed all scNMT-seq laboratory experiments (animal handling, tissue dissection, FACS, library preparation etc.) and thus generated all multi-omic data sets reported in this work. In these experiments, Dr. Cerrizuela was supported by Mohammad Eid Al Shukairi, Tobias Ellinger, Jannes Straub and Aylin Korkmaz. Mohammad Eid Al Shukairi and Dr. Cerrizuela conducted ischemia experiments.

MAB-seq laboratory experiments were performed by Dr. Sascha Dehler and Dr. Cerrizuela and supervised by Dr. Dieter Weichenhan and Prof. Christoph Plass.

Martina Braun implemented an approach to detect DMRs in single-cell methylation data under my supervision (`scbs diff`; Braun, 2023). Leonie Küchenhoff performed initial benchmarks of the `scbs` software under my supervision, some of which inspired my own benchmarks reported in section 2.3.

I performed all computational analyses, interpreted results, created all figures except those in the introduction, and developed the `scbs` software under supervision of Prof. Ana Martin-Villalba and Jun.-Prof. Simon Anders. My results presented in this thesis were also published in Kremer et al. (2022a) and Kremer et al. (2022b).

Bibliography

Aaku-Saraste, E., Hellwig, A., and Huttner, W. B. (1996). Loss of occludin and functional tight junctions, but not ZO-1, during neural tube closure-remodeling of the neuroepithelium prior to neurogenesis. *Developmental Biology*, 180(2):664–679.

Abbott, L. C. and Nigussie, F. (2020). Adult neurogenesis in the mammalian dentate gyrus. *Anatomia, Histologia, Embryologia*, 49(1):3–16.

Abdissa, D., Hamba, N., and Gerbi, A. (2020). Review article on adult neurogenesis in humans. *Translational Research in Anatomy*, 20:100074.

Ahlmann-Eltze, C. and Huber, W. (2023). Comparison of transformations for single-cell RNA-seq data. *Nature Methods*, pages 1–8.

Altman, J. (1962). Are new neurons formed in the brains of adult mammals? *Science*, 135(3509):1127–1128.

Altman, J. (1969). Autoradiographic and histological studies of postnatal neurogenesis. IV. cell proliferation and migration in the anterior forebrain, with special reference to persisting neurogenesis in the olfactory bulb. *Journal of Comparative Neurology*, 137(4):433–457.

Amezquita, R. A., Lun, A. T., Becht, E., Carey, V. J., Carpp, L. N., Geistlinger, L., Marini, F., Rue-Albrecht, K., Risso, D., Soneson, C., Waldron, L., Hervé, P., Smith, M. L., Huber, W., Morgan, M., Gottardo, R., and Hicks, S. C. (2020). Orchestrating single-cell analysis with Bioconductor. *Nature Methods*, 17(2):137–145 and accompanying online book accessed from https://bioconductor.org/books/release/OSCA/ on 21.07.2023.

An, J., Rao, A., and Ko, M. (2017). TET family dioxygenases and DNA demethylation in stem cells and cancers. *Experimental & Molecular Medicine*, 49(4):e323–e323.

Anastasiadi, D., Esteve-Codina, A., and Piferrer, F. (2018). Consistent inverse correlation between DNA methylation of the first intron and gene expression across tissues and species. *Epigenetics & Chromatin*, 11(1):1–17.

Anderson, S., Eisenstat, D., Shi, L., and Rubenstein, J. (1997). Interneuron migration from basal forebrain to neocortex: dependence on Dlx genes. *Science*, 278(5337):474–476.

Angeloni, A. and Bogdanovic, O. (2019). Enhancer DNA methylation: implications for gene regulation. *Essays in Biochemistry*, 63(6):707–715.

Arand, J., Wossidlo, M., Lepikhov, K., Peat, J. R., Reik, W., and Walter, J. (2015). Selective impairment of methylation maintenance is the major cause of DNA methylation reprogramming in the early embryo. *Epigenetics & Chromatin*, 8(1):1–14.

Argelaguet, R., Arnol, D., Bredikhin, D., Deloro, Y., Velten, B., Marioni, J. C., and Stegle, O. (2020). MOFA+: a statistical framework for comprehensive integration of multi-modal single-cell data. *Genome Biology*, 21(1):1–17.

Argelaguet, R., Clark, S. J., Mohammed, H., Stapel, L. C., Krueger, C., Kapourani, C.-A., Imaz-Rosshandler, I., Lohoff, T., Xiang, Y., Hanna, C. W., et al. (2019). Multi-omics profiling of mouse gastrulation at single-cell resolution. *Nature*, 576(7787):487–491.

Azim, K., Berninger, B., and Raineteau, O. (2016). Mosaic subventricular origins of forebrain oligodendrogenesis. *Frontiers in Neuroscience*, 10:107.

Barau, J., Teissandier, A., Zamudio, N., Roy, S., Nalesso, V., Hérault, Y., Guillou, F., and Bourc'his, D. (2016). The DNA methyltransferase DNMT3C protects male germ cells from transposon activity. *Science*, 354(6314):909–912.

Basak, O., Krieger, T. G., Muraro, M. J., Wiebrands, K., Stange, D. E., Frias-Aldeguer, J., Rivron, N. C., van de Wetering, M., van Es, J. H., van Oudenaarden, A., et al. (2018). Troy^{+} brain stem cells cycle through quiescence and regulate their number by sensing niche occupancy. *Proceedings of the National Academy of Sciences*, 115(4):E610–E619.

Baumann, N. and Pham-Dinh, D. (2001). Biology of oligodendrocyte and myelin in the mammalian central nervous system. *Physiological Reviews*, 81(2):871–927.

Baysoy, A., Bai, Z., Satija, R., and Fan, R. (2023). The technological landscape and applications of single-cell multi-omics. *Nature Reviews Molecular Cell Biology*, pages 1–19.

Bazargani, N. and Attwell, D. (2016). Astrocyte calcium signaling: the third wave. *Nature Neuroscience*, 19(2):182–189.

Beckervordersandforth, R., Tripathi, P., Ninkovic, J., Bayam, E., Lepier, A., Stempfhuber, B., Kirchhoff, F., Hirrlinger, J., Haslinger, A., Lie, D. C., et al. (2010). In vivo fate mapping and expression analysis reveals molecular hallmarks of prospectively isolated adult neural stem cells. *Cell Stem Cell*, 7(6):744–758.

Behrens, J., Von Kries, J. P., Kühl, M., Bruhn, L., Wedlich, D., Grosschedl, R., and Birchmeier, W. (1996). Functional interaction of β-catenin with the transcription factor LEF-1. *Nature*, 382(6592):638–642.

Bennett, M. V., Contreras, J. E., Bukauskas, F. F., and Sáez, J. C. (2003). New roles for astrocytes: gap junction hemichannels have something to communicate. *Trends in Neurosciences*, 26(11):610–617.

Bercury, K. K. and Macklin, W. B. (2015). Dynamics and mechanisms of CNS myelination. *Developmental Cell*, 32(4):447–458.

Bianchi, A., Scherer, M., Zaurin, R., Quililan, K., Velten, L., and Beekman, R. (2022). scTAM-seq enables targeted high-confidence analysis of DNA methylation in single cells. *Genome Biology*, 23(1):1–17.

Bird, A. (2002). DNA methylation patterns and epigenetic memory. *Genes & Development*, 16(1):6–21.

Bird, A. P. (1980). DNA methylation and the frequency of CpG in animal DNA. *Nucleic Acids Research*, 8(7):1499–1504.

Bird, A. P. (1986). CpG-rich islands and the function of DNA methylation. *Nature*, 321(6067):209–213.

Bird, A. P. (1995). Gene number, noise reduction and biological complexity. *Trends in Genetics*, 11(3):94–100.

Boers, R., Boers, J., Tan, B., van Leeuwen, M. E., Wassenaar, E., Sanchez, E. G., Sleddens, E., Tenhagen, Y., Mulugeta, E., Laven, J., et al. (2023). Retrospective analysis of enhancer activity and transcriptome history. *Nature Biotechnology*, pages 1–11.

Boggs, J. (2006). Myelin basic protein: a multifunctional protein. *Cellular and Molecular Life Sciences CMLS*, 63(17):1945–1961.

Bond, A. M., Ming, G.-l., and Song, H. (2015). Adult mammalian neural stem cells and neurogenesis: five decades later. *Cell Stem Cell*, 17(4):385–395.

Bradl, M. and Lassmann, H. (2010). Oligodendrocytes: biology and pathology. *Acta Neuropathologica*, 119:37–53.

Braun, M. (2023). Uncovering mCH accumulation in migratory neuroblasts preceeding neuronal differentiation in the adult brain and DMR detection in scBS-seq data by scbs diff. Master's thesis, Universität Heidelberg, Heidelberg, Germany.

Bravo González-Blas, C., De Winter, S., Hulselmans, G., Hecker, N., Matetovici, I., Christiaens, V., Poovathingal, S., Wouters, J., Aibar, S., and Aerts, S. (2023). SCENIC+: single-cell multiomic inference of enhancers and gene regulatory networks. *Nature Methods*, pages 1–13.

Brenet, F., Moh, M., Funk, P., Feierstein, E., Viale, A. J., Socci, N. D., and Scandura, J. M. (2011). DNA methylation of the first exon is tightly linked to transcriptional silencing. *PloS one*, 6(1):e14524.

Brill, M. S., Snapyan, M., Wohlfrom, H., Ninkovic, J., Jawerka, M., Mastick, G. S., Ashery-Padan, R., Saghatelyan, A., Berninger, B., and Götz, M. (2008). A Dlx2-and Pax6-dependent transcriptional code for periglomerular neuron specification in the adult olfactory bulb. *Journal of Neuroscience*, 28(25):6439–6452.

Buchanan, J., da Costa, N. M., and Cheadle, L. (2023). Emerging roles of oligodendrocyte precursor cells in neural circuit development and remodeling. *Trends in Neurosciences*.

Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y., and Greenleaf, W. J. (2013). Transposition of native chromatin for multimodal regulatory analysis and personal epigenomics. *Nature Methods*, 10(12):1213.

Buenrostro, J. D., Wu, B., Litzenburger, U. M., Ruff, D., Gonzales, M. L., Snyder, M. P., Chang, H. Y., and Greenleaf, W. J. (2015). Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature*, 523(7561):486–490.

Buluç, A., Fineman, J. T., Frigo, M., Gilbert, J. R., and Leiserson, C. E. (2009). Parallel sparse matrix-vector and matrix-transpose-vector multiplication using compressed sparse blocks. In *Proceedings of the twenty-first annual symposium on Parallelism in algorithms and architectures*, pages 233–244.

Cabezas, R., Ávila, M., Gonzalez, J., El-Bachá, R. S., Báez, E., García-Segura, L. M., Jurado Coronel, J. C., Capani, F., Cardona-Gomez, G. P., and Barreto, G. E. (2014). Astrocytic

modulation of blood brain barrier: perspectives on Parkinson's disease. *Frontiers in Cellular Neuroscience*, 8:211.

Capela, A. and Temple, S. (2002). LeX/ssea-1 is expressed by adult mouse CNS stem cells, identifying them as nonependymal. *Neuron*, 35(5):865–875.

Capilla-Gonzalez, V., Cebrian-Silla, A., Guerrero-Cazares, H., Garcia-Verdugo, J. M., and Quiñones-Hinojosa, A. (2014). Age-related changes in astrocytic and ependymal cells of the subventricular zone. *Glia*, 62(5):790–803.

Carvajal Ibañez, D., Skabkin, M., Hooli, J., Cerrizuela, S., Göpferich, M., Jolly, A., Volk, K., Zumwinkel, M., Bertolini, M., Figlia, G., et al. (2023). Interferon regulates neural stem cell function at all ages by orchestrating mTOR and cell cycle. *EMBO Molecular Medicine*, page e16434.

Castro-Mondragon, J. A., Riudavets-Puig, R., Rauluseviciute, I., Berhanu Lemma, R., Turchi, L., Blanc-Mathieu, R., Lucas, J., Boddie, P., Khan, A., Manosalva Pérez, N., et al. (2022). JASPAR 2022: the 9^{th} release of the open-access database of transcription factor binding profiles. *Nucleic Acids Research*, 50(D1):D165–D173.

Cebrian-Silla, A., Nascimento, M. A., Redmond, S. A., Mansky, B., Wu, D., Obernier, K., Rodriguez, R. R., Gonzalez-Granero, S., García-Verdugo, J. M., Lim, D. A., et al. (2021). Single-cell analysis of the ventricular-subventricular zone reveals signatures of dorsal and ventral adult neurogenesis. *Elife*, 10:e67436.

Cedar, H., Sabag, O., and Reizel, Y. (2022). The role of DNA methylation in genome-wide gene regulation during development. *Development*, 149(2):dev200118.

Cerrizuela, S., Kaya, O., Kremer, L. P. M., Sarvari, A., Ellinger, T., Straub, J., Brunken, J., Sanz-Morejón, A., Korkmaz, A., and Martin-Villalba, A. (2022). High-throughput scNMT protocol for multiomics profiling of single cells from mouse brain and pancreatic organoids. *Star Protocols*.

Chai, H., Diaz-Castro, B., Shigetomi, E., Monte, E., Octeau, J. C., Yu, X., Cohn, W., Rajendran, P. S., Vondriska, T. M., Whitelegge, J. P., et al. (2017). Neural circuit-specialized astrocytes: transcriptomic, proteomic, morphological, and functional evidence. *Neuron*, 95(3):531–549.

Chaker, Z., Codega, P., and Doetsch, F. (2016). A mosaic world: puzzles revealed by adult neural stem cell heterogeneity. *Wiley Interdisciplinary Reviews: Developmental Biology*, 5(6):640–658.

Chatterton, Z., Lamichhane, P., Ahmadi Rastegar, D., Fitzpatrick, L., Lebhar, H., Marquis, C., Halliday, G., and Kwok, J. B. (2023). Single-cell DNA methylation sequencing by combinatorial indexing and enzymatic DNA methylation conversion. *Cell & Bioscience*, 13(1):2.

Chen, C., Liao, Y., and Peng, G. (2022). Connecting past and present: single-cell lineage tracing. *Protein & Cell*, 13(11):790–807.

Chen, Z. and Zhang, Y. (2020). Role of mammalian DNA methyltransferases in development. *Annual Review of Biochemistry*, 89:135–158.

Cheng, M., Jiang, Y., Xu, J., Mentis, A.-F. A., Wang, S., Zheng, H., Sahu, S. K., Liu, L., and Xu, X. (2023). Spatially resolved transcriptomics: a comprehensive review of their technological advances, applications, and challenges. *Journal of Genetics and Genomics*.

Chomyk, A. M., Volsko, C., Tripathi, A., Deckard, S. A., Trapp, B. D., Fox, R. J., and Dutta, R. (2017). DNA methylation in demyelinated multiple sclerosis hippocampus. *Scientific Reports*, 7(1):8696.

Clark, S. J., Argelaguet, R., Kapourani, C.-A., Stubbs, T. M., Lee, H. J., Alda-Catalinas, C., Krueger, F., Sanguinetti, G., Kelsey, G., Marioni, J. C., et al. (2018). scNMT-seq enables joint profiling of chromatin accessibility DNA methylation and transcription in single cells. *Nature Communications*, 9(1):1–9.

Clark, S. J., Argelaguet, R., Lohoff, T., Krueger, F., Drage, D., Göttgens, B., Marioni, J. C., Nichols, J., and Reik, W. (2022). Single-cell multi-omics profiling links dynamic DNA methylation to cell fate decisions during mouse early organogenesis. *Genome Biology*, 23(1):1–20.

Clark, S. J., Smallwood, S. A., Lee, H. J., Krueger, F., Reik, W., and Kelsey, G. (2017). Genome-wide base-resolution mapping of DNA methylation in single cells using single-cell bisulfite sequencing (scBS-seq). *Nature Protocols*, 12(3):534–547.

Clemens, A. W., Wu, D. Y., Moore, J. R., Christian, D. L., Zhao, G., and Gabel, H. W. (2020). MeCP2 represses enhancers through chromosome topology-associated DNA methylation. *Molecular Cell*, 77(2):279–293.

Codega, P., Silva-Vargas, V., Paul, A., Maldonado-Soto, A. R., DeLeo, A. M., Pastrana, E., and Doetsch, F. (2014). Prospective identification and purification of quiescent adult neural stem cells from their in vivo niche. *Neuron*, 82(3):545–559.

Datlinger, P., Rendeiro, A. F., Boenke, T., Senekowitsch, M., Krausgruber, T., Barreca, D., and Bock, C. (2021). Ultra-high-throughput single-cell RNA sequencing and perturbation screening with combinatorial fluidic indexing. *Nature Methods*, 18(6):635–642.

de Mendoza, A., Nguyen, T. V., Ford, E., Poppe, D., Buckberry, S., Pflueger, J., Grimmer, M. R., Stolzenburg, S., Bogdanovic, O., Oshlack, A., et al. (2022). Large-scale manipulation of promoter DNA methylation reveals context-specific transcriptional responses and stability. *Genome Biology*, 23(1):163.

del Río Hortega, P. and Penfield, W. (1927). *Cerebral cicatrix: the reaction of neuroglia and microglia to brain wounds*. Junta para la Ampliación de Estudios e Investigaciones Científicas.

Delgado, A. C., Maldonado-Soto, A. R., Silva-Vargas, V., Mizrak, D., von Känel, T., Tan, K. R., Paul, A., Madar, A., Cuervo, H., Kitajewski, J., et al. (2021). Release of stem cells from quiescence reveals gliogenic domains in the adult mouse brain. *Science*, 372(6547):1205–1209.

Díaz-Castro, B., Robel, S., and Mishra, A. (2023). Astrocyte endfeet in brain function and pathology: Open questions. *Annual Review of Neuroscience*, 46.

Ding, J., Adiconis, X., Simmons, S. K., Kowalczyk, M. S., Hession, C. C., Marjanovic, N. D., Hughes, T. K., Wadsworth, M. H., Burks, T., Nguyen, L. T., et al. (2020). Systematic comparison of single-cell and single-nucleus RNA-sequencing methods. *Nature Biotechnology*, 38(6):737–746.

Diotel, N., Lübke, L., Strähle, U., and Rastegar, S. (2020). Common and distinct features of adult neurogenesis and regeneration in the telencephalon of zebrafish and mammals. *Frontiers in Neuroscience*, 14:957.

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, 29(1):15–21.

Doetsch, F., Caille, I., Lim, D. A., García-Verdugo, J. M., and Alvarez-Buylla, A. (1999). Subventricular zone astrocytes are neural stem cells in the adult mammalian brain. *Cell*, 97(6):703–716.

Doetsch, F., Garcıa-Verdugo, J. M., and Alvarez-Buylla, A. (1997). Cellular composition and three-dimensional organization of the subventricular germinal zone in the adult mammalian brain. *Journal of Neuroscience*, 17(13):5046–5061.

Donaghey, J., Thakurela, S., Charlton, J., Chen, J. S., Smith, Z. D., Gu, H., Pop, R., Clement, K., Stamenova, E. K., Karnik, R., et al. (2018). Genetic determinants and epigenetic effects of pioneer-factor occupancy. *Nature Genetics*, 50(2):250–258.

Dos Santos, C. O., Dolzhenko, E., Hodges, E., Smith, A. D., and Hannon, G. J. (2015). An epigenetic memory of pregnancy in the mouse mammary gland. *Cell Reports*, 11(7):1102–1109.

Duan, C.-L., Liu, C.-W., Shen, S.-W., Yu, Z., Mo, J.-L., Chen, X.-H., and Sun, F.-Y. (2015). Striatal astrocytes transdifferentiate into functional mature neurons following ischemic brain injury. *Glia*, 63(9):1660–1670.

Dulken, B. W., Buckley, M. T., Navarro Negredo, P., Saligrama, N., Cayrol, R., Leeman, D. S., George, B. M., Boutet, S. C., Hebestreit, K., Pluvinage, J. V., et al. (2019). Single-cell analysis reveals T cell infiltration in old neurogenic niches. *Nature*, 571(7764):205–210.

Edrei, Y., Levy, R., Marom, A., Radlwimmer, B., and Hellman, A. (2021). Methylation-mediated retuning on the enhancer-to-silencer activity scale of networked regulatory elements guides driver-gene misregulation. *bioRχiv*, pages 2021–03.

Escartin, C., Galea, E., Lakatos, A., O'Callaghan, J. P., Petzold, G. C., Serrano-Pozo, A., Steinhäuser, C., Volterra, A., Carmignoto, G., Agarwal, A., et al. (2021). Reactive astrocyte nomenclature, definitions, and future directions. *Nature Neuroscience*, 24(3):312–325.

Fabyanic, E. B., Hu, P., Qiu, Q., Wang, T., Berríos, K. N., Flournoy, J., Connolly, D. R., Zhou, Z., Kohil, R. M., and Wu, H. (2021). Quantitative single cell 5hmC sequencing reveals non-canonical gene regulation by non-CG hydroxymethylation. *bioRχiv*, pages 2021–03.

Farlik, M., Sheffield, N. C., Nuzzo, A., Datlinger, P., Schönegger, A., Klughammer, J., and Bock, C. (2015). Single-cell DNA methylome sequencing and bioinformatic inference of epigenomic cell-state dynamics. *Cell Reports*, 10(8):1386–1397.

Fatemi, M., Pao, M. M., Jeong, S., Gal-Yam, E. N., Egger, G., Weisenberger, D. J., and Jones, P. A. (2005). Footprinting of mammalian promoters: use of a CpG DNA methyltransferase revealing nucleosome positions at a single molecule level. *Nucleic Acids Research*, 33(20):e176–e176.

Faulkner, J. R., Herrmann, J. E., Woo, M. J., Tansey, K. E., Doan, N. B., and Sofroniew, M. V. (2004). Reactive astrocytes protect tissue and preserve function after spinal cord injury. *Journal of Neuroscience*, 24(9):2143–2155.

Felsenfeld, G., Boyes, J., Chung, J., Clark, D., and Studitsky, V. (1996). Chromatin structure and gene expression. *Proceedings of the National Academy of Sciences*, 93(18):9384–9388.

Filippi, M., Bar-Or, A., Piehl, F., Preziosa, P., Solari, A., Vukusic, S., and Rocca, M. (2018). Multiple sclerosis. *Nature Reviews Disease Primers*.

Fischer, J. and Schulz, M. H. (2023). Efficiently quantifying DNA methylation for bulk-and single-cell bisulfite data. *bioRχiv*, pages 2023–01.

Ford, E., Grimmer, M. R., Stolzenburg, S., Bogdanovic, O., Mendoza, A. d., Farnham, P. J., Blancafort, P., and Lister, R. (2017). Frequent lack of repressive capacity of promoter DNA methylation identified through genome-wide epigenomic manipulation. *bioRχiv*, page 170506.

Frauhammer, F. and Anders, S. (2022). cellpypes: Cell type pipes for R. available at https://zenodo.org/record/6555728.

Frommer, M., McDonald, L. E., Millar, D. S., Collis, C. M., Watt, F., Grigg, G. W., Molloy, P. L., and Paul, C. L. (1992). A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proceedings of the National Academy of Sciences*, 89(5):1827–1831.

Fuentealba, L. C., Rompani, S. B., Parraguez, J. I., Obernier, K., Romero, R., Cepko, C. L., and Alvarez-Buylla, A. (2015). Embryonic origin of postnatal neural stem cells. *Cell*, 161(7):1644–1655.

Furutachi, S., Miya, H., Watanabe, T., Kawai, H., Yamasaki, N., Harada, Y., Imayoshi, I., Nelson, M., Nakayama, K. I., Hirabayashi, Y., et al. (2015). Slowly dividing neural progenitors are an embryonic origin of adult neural stem cells. *Nature Neuroscience*, 18(5):657–665.

Galli, R., Fiocco, R., De Filippis, L., Muzio, L., Gritti, A., Mercurio, S., Broccoli, V., Pellegrini, M., Mallamaci, A., and Vescovi, A. L. (2002). Emx2 regulates the proliferation of stem cells of the adult mammalian central nervous system. *Development*.

Gangemi, R. M., Daga, A., Marubbi, D., Rosatto, N., Capra, M. C., and Corte, G. (2001). Emx2 in adult neural precursor cells. *Mechanisms of Development*, 109(2):323–329.

Gao, S., Zou, D., Mao, L., Liu, H., Song, P., Chen, Y., Zhao, S., Gao, C., Li, X., Gao, Z., et al. (2015). BS-SNPer: SNP calling in bisulfite-seq data. *Bioinformatics*, 31(24):4006–4008.

Garcia, A. D. R., Doan, N. B., Imura, T., Bush, T. G., and Sofroniew, M. V. (2004). GFAP-expressing progenitors are the principal source of constitutive neurogenesis in adult mouse forebrain. *Nature Neuroscience*, 7(11):1233–1241.

Gardiner-Garden, M. and Frommer, M. (1987). CpG islands in vertebrate genomes. *Journal of Molecular Biology*, 196(2):261–282.

Goldman, S. A. and Nottebohm, F. (1983). Neuronal production, migration, and differentiation in a vocal control nucleus of the adult female canary brain. *Proceedings of the National Academy of Sciences*, 80(8):2390–2394.

Goll, M. G., Kirpekar, F., Maggert, K. A., Yoder, J. A., Hsieh, C.-L., Zhang, X., Golic, K. G., Jacobsen, S. E., and Bestor, T. H. (2006). Methylation of $tRNA^{Asp}$ by the DNA methyltransferase homolog Dnmt2. *Science*, 311(5759):395–398.

Gonzalez-Perez, O. and Alvarez-Buylla, A. (2011). Oligodendrogenesis in the subventricular zone and the role of epidermal growth factor. *Brain Research Reviews*, 67(1-2):147–156.

Greig, L. C., Woodworth, M. B., Galazo, M. J., Padmanabhan, H., and Macklis, J. D. (2013). Molecular logic of neocortical projection neuron specification, development and diversity. *Nature Reviews Neuroscience*, 14(11):755–769.

Gronostajski, R. M. (2000). Roles of the NFI/CTF gene family in transcription and development. *Gene*, 249(1-2):31–45.

Gross, C. G. (2000). Neurogenesis in the adult brain: death of a dogma. *Nature Reviews Neuroscience*, 1(1):67–73.

Grunau, C., Clark, S., and Rosenthal, A. (2001). Bisulfite genomic sequencing: systematic investigation of critical experimental parameters. *Nucleic Acids Research*, 29(13):e65–e65.

Gu, Z., Eils, R., and Schlesner, M. (2016). Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics*, 32(18):2847–2849.

Guillemot, F. (2005). Cellular and molecular control of neurogenesis in the mammalian telencephalon. *Current Opinion in Cell Biology*, 17(6):639–647.

Guo, H., Zhu, P., Guo, F., Li, X., Wu, X., Fan, X., Wen, L., and Tang, F. (2015). Profiling DNA methylome landscapes of mammalian cells with single-cell reduced-representation bisulfite sequencing. *Nature Protocols*, 10(5):645–659.

Guo, H., Zhu, P., Wu, X., Li, X., Wen, L., and Tang, F. (2013). Single-cell methylome landscapes of mouse embryonic stem cells and early embryos analyzed using reduced representation bisulfite sequencing. *Genome Research*, 23(12):2126–2135.

Hafemeister, C. and Satija, R. (2019). Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biology*, 20(1):1–15.

Hagemann-Jensen, M., Ziegenhain, C., Chen, P., Ramsköld, D., Hendriks, G.-J., Larsson, A. J., Faridani, O. R., and Sandberg, R. (2020). Single-cell RNA counting at allele and isoform resolution using Smart-seq3. *Nature Biotechnology*, 38(6):708–714.

Haghverdi, L., Büttner, M., Wolf, F. A., Buettner, F., and Theis, F. J. (2016). Diffusion pseudotime robustly reconstructs lineage branching. *Nature Methods*, 13(10):845–848.

Haghverdi, L., Lun, A. T., Morgan, M. D., and Marioni, J. C. (2018). Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. *Nature Biotechnology*, 36(5):421–427.

Haim, L. B. and Rowitch, D. H. (2017). Functional diversity of astrocytes in neural circuit regulation. *Nature Reviews Neuroscience*, 18(1):31–41.

Hashimshony, T., Senderovich, N., Avital, G., Klochendler, A., De Leeuw, Y., Anavy, L., Gennert, D., Li, S., Livak, K. J., Rozenblatt-Rosen, O., et al. (2016). CEL-Seq2: sensitive highly-multiplexed single-cell RNA-seq. *Genome Biology*, 17(1):1–7.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y. C., Laslo, P., Cheng, J. X., Murre, C., Singh, H., and Glass, C. K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Molecular Cell*, 38(4):576–589.

Hernandez-Corchado, A. and Najafabadi, H. S. (2022). Toward a base-resolution panorama of the in vivo impact of cytosine methylation on transcription factor binding. *Genome Biology*, 23(1):151.

Heumos, L., Schaar, A. C., Lance, C., Litinetskaya, A., Drost, F., Zappia, L., Lücken, M. D., Strobl, D. C., Henao, J., Curion, F., et al. (2023). Best practices for single-cell analysis across modalities. *Nature Reviews Genetics*, pages 1–23 and accompanying online book accessed from https://sc–best–practices.org on 18.07.2023.

Hirasawa, R., Chiba, H., Kaneda, M., Tajima, S., Li, E., Jaenisch, R., and Sasaki, H. (2008). Maternal and zygotic Dnmt1 are necessary and sufficient for the maintenance of DNA methylation imprints during preimplantation development. *Genes & Development*, 22(12):1607–1616.

Holmberg, J., Armulik, A., Senti, K.-A., Edoff, K., Spalding, K., Momma, S., Cassidy, R., Flanagan, J. G., and Frisén, J. (2005). Ephrin-A2 reverse signaling negatively regulates neural progenitor proliferation and neurogenesis. *Genes & Development*, 19(4):462–471.

Hotchkiss, R. D. (1948). The quantitative separation of purines, pyrimidines, and nucleosides by paper chromatography. *Journal of Biological Chemistry*, 175(1):315–332.

Hovestadt, V., Jones, D. T., Picelli, S., Wang, W., Kool, M., Northcott, P. A., Sultan, M., Stachurski, K., Ryzhova, M., Warnatz, H.-J., et al. (2014). Decoding the regulatory landscape of medulloblastoma using DNA methylation sequencing. *Nature*, 510(7506):537–541.

Hu, T., Chitnis, N., Monos, D., and Dinh, A. (2021). Next-generation sequencing technologies: An overview. *Human Immunology*, 82(11):801–811.

Huang, S., Hendriks, W., Althage, A., Hemmi, S., Bluethmann, H., Kamijo, R., Vilček, J., Zinkernagel, R. M., and Aguet, M. (1993). Immune response in mice that lack the interferon-γ receptor. *Science*, 259(5102):1742–1745.

Huang, Y., Pastor, W. A., Shen, Y., Tahiliani, M., Liu, D. R., and Rao, A. (2010). The behaviour of 5-hydroxymethylcytosine in bisulfite sequencing. *PloS one*, 5(1):e8888.

Huang, Z. J. and Paul, A. (2019). The diversity of GABAergic neurons and neural communication elements. *Nature Reviews Neuroscience*, 20(9):563–572.

Huttner, W. B. and Brand, M. (1997). Asymmetric division and polarity of neuroepithelial cells. *Current Opinion in Neurobiology*, 7(1):29–39.

Huxley, A. and Stämpfli, R. (1949). Evidence for saltatory conduction in peripheral myelinated nerve fibres. *The Journal of Physiology*, 108(3):315.

Huynh, J. L., Garg, P., Thin, T. H., Yoo, S., Dutta, R., Trapp, B. D., Haroutunian, V., Zhu, J., Donovan, M. J., Sharp, A. J., et al. (2014). Epigenome-wide differences in pathology-free regions of multiple sclerosis–affected brains. *Nature Neuroscience*, 17(1):121–130.

Ihrie, R. A. and Álvarez-Buylla, A. (2011). Lake-front property: a unique germinal niche by the lateral ventricles of the adult brain. *Neuron*, 70(4):674–686.

Islam, S., Zeisel, A., Joost, S., La Manno, G., Zajac, P., Kasper, M., Lönnerberg, P., and Linnarsson, S. (2014). Quantitative single-cell RNA-seq with unique molecular identifiers. *Nature Methods*, 11(2):163–166.

Izzo, F., Lee, S. C., Poran, A., Chaligne, R., Gaiti, F., Gross, B., Murali, R. R., Deochand, S. D., Ang, C., Jones, P. W., et al. (2020). DNA methylation disruption reshapes the hematopoietic differentiation landscape. *Nature Genetics*, 52(4):378–387.

Jeong, H., Mendizabal, I., Berto, S., Chatterjee, P., Layman, T., Usui, N., Toriumi, K., Douglas, C., Singh, D., Huh, I., et al. (2021). Evolution of DNA methylation in the human brain. *Nature Communications*, 12(1):2021.

Jiao, J.-w., Feldheim, D. A., and Chen, D. F. (2008). Ephrins as negative regulators of adult neurogenesis in diverse regions of the central nervous system. *Proceedings of the National Academy of Sciences*, 105(25):8778–8783.

Jiménez, A. J., Domínguez-Pinos, M.-D., Guerra, M. M., Fernández-Llebrez, P., and Pérez-Fígares, J.-M. (2014). Structure and function of the ependymal barrier and diseases associated with ependyma disruption. *Tissue Barriers*, 2(1):e28426.

Johnson, T. B. and Coghill, R. D. (1925). Researches on pyrimidines. C111. the discovery of 5-methyl-cytosine in tuberculinic acid, the nucleic acid of the tubercle bacillus. *Journal of the American Chemical Society*, 47(11):2838–2844.

Jones, P. A. (2012). Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nature Reviews Genetics*, 13(7):484–492.

Jones, P. A., Issa, J.-P. J., and Baylin, S. (2016). Targeting the cancer epigenome for therapy. *Nature Reviews Genetics*, 17(10):630–641.

Joven, A. and Simon, A. (2018). Homeostatic and regenerative neurogenesis in salamanders. *Progress in Neurobiology*, 170:81–98.

Jurkowski, M. P., Bettio, L., K. Woo, E., Patten, A., Yau, S.-Y., and Gil-Mohapel, J. (2020). Beyond the hippocampus and the SVZ: adult neurogenesis throughout the brain. *Frontiers in Cellular Neuroscience*, 14:576444.

Kalamakis, G., Brüne, D., Ravichandran, S., Bolz, J., Fan, W., Ziebell, F., Stiehl, T., Catalá-Martinez, F., Kupke, J., Zhao, S., et al. (2019). Quiescence modulates stem cell maintenance and regenerative capacity in the aging brain. *Cell*, 176(6):1407–1419.

Kaluscha, S., Domcke, S., Wirbelauer, C., Stadler, M. B., Durdu, S., Burger, L., and Schübeler, D. (2022). Evidence that direct inhibition of transcription factor binding is the prevailing mode of gene and repeat repression by DNA methylation. *Nature Genetics*, pages 1–12.

Kaplan, M. S. and Hinds, J. W. (1977). Neurogenesis in the adult rat: electron microscopic analysis of light radioautographs. *Science*, 197(4308):1092–1094.

Kashima, Y., Sakamoto, Y., Kaneko, K., Seki, M., Suzuki, Y., and Suzuki, A. (2020). Single-cell sequencing techniques from individual to multiomics analyses. *Experimental & Molecular Medicine*, 52(9):1419–1427.

Kazanis, I., Evans, K. A., Andreopoulou, E., Dimitriou, C., Koutsakis, C., Karadottir, R. T., and Franklin, R. J. (2017). Subependymal zone-derived oligodendroblasts respond to

focal demyelination but fail to generate myelin in young and aged mice. *Stem Cell Reports*, 8(3):685–700.

Kelly, T. K., Liu, Y., Lay, F. D., Liang, G., Berman, B. P., and Jones, P. A. (2012). Genome-wide mapping of nucleosome positioning and DNA methylation within individual DNA molecules. *Genome Research*, 22(12):2497–2506.

Kester, L. and Van Oudenaarden, A. (2018). Single-cell transcriptomics meets lineage tracing. *Cell Stem Cell*, 23(2):166–179.

Korthauer, K. and Irizarry, R. A. (2018). Genome-wide repressive capacity of promoter DNA methylation is revealed through epigenomic manipulation. *bioRχiv*, page 381145.

Kremer, L. P., Küchenhoff, L., Cerrizuela, S., Martin-Villalba, A., and Anders, S. (2022a). Analyzing single-cell bisulfite sequencing data with scbs. *bioRχiv*.

Kremer, L. P. M., Cerrizuela, S., Al Shukairi, M. E., Ellinger, T., Straub, J., Dehler, S., Korkmaz, A., Weichenhan, D., Plass, C., Anders, S., and Martin-Villalba, A. (2022b). Single-cell triple-omics uncovers DNA methylation as key feature of stemness in the healthy and ischemic adult brain. *bioRχiv*.

Kremer, L. P. M., Cerrizuela, S., Dehler, S., Stiehl, T., Weinmann, J., Abendroth, H., Kleber, S., Laure, A., El Andari, J., Anders, S., Marciniak-Czochra, A., Grimm, D., and Martin-Villalba, A. (2021). High throughput screening of novel AAV capsids identifies variants for transduction of adult NSCs within the subventricular zone. *Molecular Therapy - Methods & Clinical Development*, 23:33–50.

Kriegstein, A. and Alvarez-Buylla, A. (2009). The glial nature of embryonic and adult neural stem cells. *Annual Review of Neuroscience*, 32:149–184.

Krueger, F. and Andrews, S. R. (2011). Bismark: a flexible aligner and methylation caller for bisulfite-seq applications. *Bioinformatics*, 27(11):1571–1572.

Kucinski, I. and Gottgens, B. (2020). Advancing stem cell research through multimodal single-cell analysis. *Cold Spring Harbor Perspectives in Biology*, 12(7):a035725.

Kuhn, H. G., Dickinson-Anson, H., and Gage, F. H. (1996). Neurogenesis in the dentate gyrus of the adult rat: age-related decrease of neuronal progenitor proliferation. *Journal of Neuroscience*, 16(6):2027–2033.

Kuhn, S., Gritti, L., Crooks, D., and Dombrowski, Y. (2019). Oligodendrocytes in development, myelin generation and beyond. *Cells*, 8(11):1424.

Kyritsis, N., Kizil, C., Zocher, S., Kroehne, V., Kaslin, J., Freudenreich, D., Iltzsche, A., and Brand, M. (2012). Acute inflammation initiates the regenerative response in the adult zebrafish brain. *Science*, 338(6112):1353–1356.

Lalo, U., Koh, W., Lee, C. J., and Pankratov, Y. (2021). The tripartite glutamatergic synapse. *Neuropharmacology*, 199:108758.

Lam, S. K., Pitrou, A., and Seibert, S. (2015). Numba: A LLVM-based Python JIT compiler. In *Proceedings of the Second Workshop on the LLVM Compiler Infrastructure in HPC*, pages 1–6.

Larsen, F., Gundersen, G., Lopez, R., and Prydz, H. (1992). CpG islands as gene markers in the human genome. *Genomics*, 13(4):1095–1107.

Laurent, L., Wong, E., Li, G., Huynh, T., Tsirigos, A., Ong, C. T., Low, H. M., Sung, K. W. K., Rigoutsos, I., Loring, J., et al. (2010). Dynamic changes in the human methylome during differentiation. *Genome Research*, 20(3):320–331.

Lawrence, M., Huber, W., Pages, H., Aboyoun, P., Carlson, M., Gentleman, R., Morgan, M. T., and Carey, V. J. (2013). Software for computing and annotating genomic ranges. *PLoS Computational Biology*, 9(8):e1003118.

Laywell, E. D., Rakic, P., Kukekov, V. G., Holland, E. C., and Steindler, D. A. (2000). Identification of a multipotent astrocytic stem cell in the immature and adult mouse brain. *Proceedings of the National Academy of Sciences*, 97(25):13883–13888.

Lee, H. J., Hore, T. A., and Reik, W. (2014). Reprogramming the methylome: erasing memory and creating diversity. *Cell Stem Cell*, 14(6):710–719.

Lehtinen, M. K., Zappaterra, M. W., Chen, X., Yang, Y. J., Hill, A. D., Lun, M., Maynard, T., Gonzalez, D., Kim, S., Ye, P., et al. (2011). The cerebrospinal fluid provides a proliferative niche for neural progenitor cells. *Neuron*, 69(5):893–905.

Li, L. and Clevers, H. (2010). Coexistence of quiescent and active adult stem cells in mammals. *Science*, 327(5965):542–545.

Liddelow, S. A., Guttenplan, K. A., Clarke, L. E., Bennett, F. C., Bohlen, C. J., Schirmer, L., Bennett, M. L., Münch, A. E., Chung, W.-S., Peterson, T. C., et al. (2017). Neurotoxic reactive astrocytes are induced by activated microglia. *Nature*, 541(7638):481–487.

Lie, D.-C., Colamarino, S. A., Song, H.-J., Désiré, L., Mira, H., Consiglio, A., Lein, E. S., Jessberger, S., Lansford, H., Dearie, A. R., et al. (2005). Wnt signalling regulates adult hippocampal neurogenesis. *Nature*, 437(7063):1370–1375.

Lim, D. A. and Alvarez-Buylla, A. (2016). The adult ventricular–subventricular zone (V-SVZ) and olfactory bulb (OB) neurogenesis. *Cold Spring Harbor Perspectives in Biology*, 8(5):a018820.

Linnerbauer, M. and Rothhammer, V. (2020). Protective functions of reactive astrocytes following central nervous system insult. *Frontiers in Immunology*, 11:573256.

Lister, R., Pelizzola, M., Dowen, R. H., Hawkins, R. D., Hon, G., Tonti-Filippini, J., Nery, J. R., Lee, L., Ye, Z., Ngo, Q.-M., et al. (2009). Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature*, 462(7271):315–322.

Liu, H., Zeng, Q., Zhou, J., Bartlett, A., Wang, B., Berube, P., Tian, W., Kenworthy, M., Altshul, J., Nery, J. R., et al. (2022). Single-cell DNA methylome and 3D multi-omic atlas of the adult mouse brain. *bioRχiv*.

Liu, H.-K., Belz, T., Bock, D., Takacs, A., Wu, H., Lichter, P., Chai, M., and Schütz, G. (2008). The nuclear receptor tailless is required for neurogenesis in the adult subventricular zone. *Genes & Development*, 22(18):2473–2478.

Liu, Y., Siejka-Zielińska, P., Velikova, G., Bi, Y., Yuan, F., Tomkova, M., Bai, C., Chen, L., Schuster-Böckler, B., and Song, C.-X. (2019). Bisulfite-free direct detection of 5-methylcytosine and 5-hydroxymethylcytosine at base resolution. *Nature biotechnology*, 37(4):424–429.

Liuzzi, F. J. and Lasek, R. J. (1987). Astrocytes block axonal regeneration in mammals by activating the physiological stop pathway. *Science*, 237(4815):642–645.

Llorens-Bobadilla, E., Zhao, S., Baser, A., Saiz-Castro, G., Zwadlo, K., and Martin-Villalba, A. (2015). Single-cell transcriptomics reveals a population of dormant neural stem cells that become activated upon brain injury. *Cell Stem Cell*, 17(3):329–340.

Lois, C. and Alvarez-Buylla, A. (1994). Long-distance neuronal migration in the adult mammalian brain. *Science*, 264(5162):1145–1148.

López-Juárez, A., Howard, J., Ullom, K., Howard, L., Grande, A., Pardo, A., Waclaw, R., Sun, Y.-Y., Yang, D., Kuan, C.-Y., et al. (2013). Gsx2 controls region-specific activation of neural stem cells and injury-induced neurogenesis in the adult subventricular zone. *Genes & Development*, 27(11):1272–1287.

Luo, C., Keown, C. L., Kurihara, L., Zhou, J., He, Y., Li, J., Castanon, R., Lucero, J., Nery, J. R., Sandoval, J. P., et al. (2017). Single-cell methylomes identify neuronal subtypes and regulatory elements in mammalian cortex. *Science*, 357(6351):600–604.

Luo, C., Rivkin, A., Zhou, J., Sandoval, J. P., Kurihara, L., Lucero, J., Castanon, R., Nery, J. R., Pinto-Duarte, A., Bui, B., et al. (2018). Robust single-cell DNA methylome profiling with snmC-seq2. *Nature Communications*, 9(1):3824.

Lyko, F. (2018). The DNA methyltransferase family: a versatile toolkit for epigenetic regulation. *Nature Reviews Genetics*, 19(2):81–92.

Macosko, E. Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, A. R., Kamitaki, N., Martersteck, E. M., et al. (2015). Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell*, 161(5):1202–1214.

Magnusson, J. P., Göritz, C., Tatarishvili, J., Dias, D. O., Smith, E. M., Lindvall, O., Kokaia, Z., and Frisén, J. (2014). A latent neurogenic program in astrocytes regulated by Notch signaling in the mouse. *Science*, 346(6206):237–241.

Magnusson, J. P., Zamboni, M., Santopolo, G., Mold, J. E., Barrientos-Somarribas, M., Talavera-Lopez, C., Andersson, B., and Frisén, J. (2020). Activation of a neural stem cell transcriptional program in parenchymal astrocytes. *Elife*, 9:e59733.

Malatesta, P., Appolloni, I., and Calzolari, F. (2008). Radial glia and neural stem cells. *Cell and Tissue Research*, 331(1):165–178.

Malatesta, P., Hack, M. A., Hartfuss, E., Kettenmann, H., Klinkert, W., Kirchhoff, F., and Götz, M. (2003). Neuronal or glial progeny: regional differences in radial glia fate. *Neuron*, 37(5):751–764.

Martin, F. J., Amode, M. R., Aneja, A., Austine-Orimoloye, O., Azov, A. G., Barnes, I., Becker, A., Bennett, R., Berry, A., Bhai, J., et al. (2023). Ensembl 2023. *Nucleic Acids Research*, 51(D1):D933–D941.

Martynoga, B., Mateo, J. L., Zhou, B., Andersen, J., Achimastou, A., Urbán, N., van den Berg, D., Georgopoulou, D., Hadjur, S., Wittbrodt, J., et al. (2013). Epigenomic enhancer annotation reveals a key role for NFIX in neural stem cell quiescence. *Genes & Development*, 27(16):1769–1786.

Masuyama, N., Konno, N., and Yachie, N. (2022). Molecular recorders to track cellular events. *Science*, 377(6605):469–470.

Matias, I., Morgado, J., and Gomes, F. C. A. (2019). Astrocyte heterogeneity: impact to brain aging and disease. *Frontiers in Aging Neuroscience*, 11:59.

Mattei, A. L., Bailly, N., and Meissner, A. (2022). DNA methylation: a historical perspective. *Trends in Genetics*.

Mattugini, N., Bocchi, R., Scheuss, V., Russo, G. L., Torper, O., Lao, C. L., and Götz, M. (2019). Inducing different neuronal subtypes from astrocytes in the injured mouse cerebral cortex. *Neuron*, 103(6):1086–1095.

Mazewski, C., Perez, R. E., Fish, E. N., and Platanias, L. C. (2020). Type I interferon (IFN)-regulated activation of canonical and non-canonical signaling pathways. *Frontiers in Immunology*, 11:606456.

McInnes, L., Healy, J., and Melville, J. (2018). Umap: Uniform manifold approximation and projection for dimension reduction. *arχiv preprint. arχiv:1802.03426*.

McKenzie, I. A., Ohayon, D., Li, H., Paes de Faria, J., Emery, B., Tohyama, K., and Richardson, W. D. (2014). Motor skill learning requires active central myelination. *Science*, 346(6207):318–322.

McLean, C. Y., Bristor, D., Hiller, M., Clarke, S. L., Schaar, B. T., Lowe, C. B., Wenger, A. M., and Bejerano, G. (2010). GREAT improves functional interpretation of cis-regulatory regions. *Nature Biotechnology*, 28(5):495–501.

McTigue, D. M. and Tripathi, R. B. (2008). The life, death, and replacement of oligodendrocytes in the adult CNS. *Journal of Neurochemistry*, 107(1):1–19.

Meissner, A., Mikkelsen, T. S., Gu, H., Wernig, M., Hanna, J., Sivachenko, A., Zhang, X., Bernstein, B. E., Nusbaum, C., Jaffe, D. B., et al. (2008). Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature*, 454(7205):766–770.

Menn, B., Garcia-Verdugo, J. M., Yaschine, C., Gonzalez-Perez, O., Rowitch, D., and Alvarez-Buylla, A. (2006). Origin of oligodendrocytes in the subventricular zone of the adult brain. *Journal of Neuroscience*, 26(30):7907–7918.

Merkle, F. T., Fuentealba, L. C., Sanders, T. A., Magno, L., Kessaris, N., and Alvarez-Buylla, A. (2014). Adult neural stem cells in distinct microdomains generate previously unknown interneuron types. *Nature Neuroscience*, 17(2):207–214.

Merkle, F. T., Tramontin, A. D., García-Verdugo, J. M., and Alvarez-Buylla, A. (2004). Radial glia give rise to adult neural stem cells in the subventricular zone. *Proceedings of the National Academy of Sciences*, 101(50):17528–17532.

Miller, F. D. and Gauthier, A. S. (2007). Timing is everything: making neurons versus glia in the developing cortex. *Neuron*, 54(3):357–369.

Mirzadeh, Z., Merkle, F. T., Soriano-Navarro, M., Garcia-Verdugo, J. M., and Alvarez-Buylla, A. (2008). Neural stem cells confer unique pinwheel architecture to the ventricular surface in neurogenic regions of the adult brain. *Cell Stem Cell*, 3(3):265–278.

Mizrak, D., Levitin, H. M., Delgado, A. C., Crotet, V., Yuan, J., Chaker, Z., Silva-Vargas, V., Sims, P. A., and Doetsch, F. (2019). Single-cell analysis of regional differences in adult V-SVZ neural stem cell lineages. *Cell Reports*, 26(2):394–406.

Moore, J. E., Purcaro, M. J., Pratt, H. E., Epstein, C. B., Shoresh, N., Adrian, J., Kawli, T., Davis, C. A., Dobin, A., Kaul, R., et al. (2020). Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature*, 583(7818):699–710.

Moore, L. D., Le, T., and Fan, G. (2013). DNA methylation and its basic function. *Neuropsychopharmacology*, 38(1):23–38.

Morshead, C. M., Reynolds, B. A., Craig, C. G., McBurney, M. W., Staines, W. A., Morassutti, D., Weiss, S., and van der Kooy, D. (1994). Neural stem cells in the adult mammalian forebrain: a relatively quiescent subpopulation of subependymal cells. *Neuron*, 13(5):1071–1082.

Müller, U., Steinhoff, U., Reis, L. F., Hemmi, S., Pavlovic, J., Zinkernagel, R. M., and Aguet, M. (1994). Functional role of type I and type II interferons in antiviral defense. *Science*, 264(5167):1918–1921.

Mulqueen, R. M., Pokholok, D., Norberg, S. J., Torkenczy, K. A., Fields, A. J., Sun, D., Sinnamon, J. R., Shendure, J., Trapnell, C., O'Roak, B. J., et al. (2018). Highly scalable generation of DNA methylation profiles in single cells. *Nature Biotechnology*, 36(5):428–431.

Nait-Oumesmar, B., Picard-Riera, N., Kerninon, C., Decker, L., Seilhean, D., Höglinger, G. U., Hirsch, E. C., Reynolds, R., and Baron-Van Evercooren, A. (2007). Activation of the subventricular zone in multiple sclerosis: evidence for early glial progenitors. *Proceedings of the National Academy of Sciences*, 104(11):4694–4699.

Nakamura, M., Gao, Y., Dominguez, A. A., and Qi, L. S. (2021). CRISPR technologies for precise epigenome editing. *Nature Cell Biology*, 23(1):11–22.

Nathan, A., Baglaenko, Y., Fonseka, C. Y., Beynor, J. I., and Raychaudhuri, S. (2019). Multimodal single-cell approaches shed light on T cell heterogeneity. *Current Opinion in Immunology*, 61:17–25.

Nato, G., Caramello, A., Trova, S., Avataneo, V., Rolando, C., Taylor, V., Buffo, A., Peretto, P., and Luzzati, F. (2015). Striatal astrocytes produce neuroblasts in an excitotoxic model of Huntington's disease. *Development*, 142(5):840–845.

Neri, F., Rapelli, S., Krepelova, A., Incarnato, D., Parlato, C., Basile, G., Maldotti, M., Anselmi, F., and Oliviero, S. (2017). Intragenic DNA methylation prevents spurious transcription initiation. *Nature*, 543(7643):72–77.

Nguyen, H. C., Baik, B., Yoon, S., Park, T., and Nam, D. (2023). Benchmarking integration of single-cell differential expression. *Nature Communications*, 14(1):1570.

Nichols, R. V., O'Connell, B. L., Mulqueen, R. M., Thomas, J., Woodfin, A. R., Acharya, S., Mandel, G., Pokholok, D., Steemers, F. J., and Adey, A. C. (2022). High-throughput robust single-cell DNA methylation profiling with sciMETv2. *Nature Communications*, 13(1):7627.

Noctor, S. C., Martínez-Cerdeño, V., and Kriegstein, A. R. (2008). Distinct behaviors of neural stem and progenitor cells underlie cortical neurogenesis. *Journal of Comparative Neurology*, 508(1):28–44.

Nogueira-Rodrigues, J., Leite, S. C., Pinto-Costa, R., Sousa, S. C., Luz, L. L., Sintra, M. A., Oliveira, R., Monteiro, A. C., Pinheiro, G. G., Vitorino, M., et al. (2022). Rewired glycosylation activity promotes scarless regeneration and functional recovery in spiny mice after complete spinal cord transection. *Developmental Cell*, 57(4):440–450.

Noh, K.-M., Wang, H., Kim, H. R., Wenderski, W., Fang, F., Li, C. H., Dewell, S., Hughes, S. H., Melnick, A. M., Patel, D. J., et al. (2015). Engineering of a histone-recognition domain in Dnmt3a alters the epigenetic landscape and phenotypic features of mouse ESCs. *Molecular Cell*, 59(1):89–103.

Nomura, T., Göritz, C., Catchpole, T., Henkemeyer, M., and Frisén, J. (2010). EphB signaling controls lineage plasticity of adult neural stem cell niche cells. *Cell Stem Cell*, 7(6):730–743.

Nordström, K. J., Schmidt, F., Gasparoni, N., Salhab, A., Gasparoni, G., Kattler, K., Müller, F., Ebert, P., Costa, I. G., consortium, D., et al. (2019). Unique and assay specific features of NOMe-, ATAC- and DNase I-seq data. *Nucleic Acids Research*, 47(20):10580–10596.

Nuñez, J. K., Chen, J., Pommier, G. C., Cogan, J. Z., Replogle, J. M., Adriaens, C., Ramadoss, G. N., Shi, Q., Hung, K. L., Samelson, A. J., et al. (2021). Genome-wide programmable transcriptional memory by CRISPR-based epigenome editing. *Cell*, 184(9):2503–2519.

Obernier, K., Cebrian-Silla, A., Thomson, M., Parraguez, J. I., Anderson, R., Guinto, C., Rodriguez, J. R., Garcia-Verdugo, J.-M., and Alvarez-Buylla, A. (2018). Adult neurogenesis is sustained by symmetric self-renewal and differentiation. *Cell Stem Cell*, 22(2):221–234.

Okano, M., Bell, D. W., Haber, D. A., and Li, E. (1999). DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell*, 99(3):247–257.

Ooi, S. K., Qiu, C., Bernstein, E., Li, K., Jia, D., Yang, Z., Erdjument-Bromage, H., Tempst, P., Lin, S.-P., Allis, C. D., et al. (2007). DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. *Nature*, 448(7154):714–717.

Ortega, F., Gascón, S., Masserdotti, G., Deshpande, A., Simon, C., Fischer, J., Dimou, L., Chichung Lie, D., Schroeder, T., and Berninger, B. (2013). Oligodendrogliogenic and neurogenic adult subependymal zone neural stem cells constitute distinct lineages and exhibit differential responsiveness to Wnt signalling. *Nature Cell Biology*, 15(6):602–613.

Ostrander, E. L., Kramer, A. C., Mallaney, C., Celik, H., Koh, W. K., Fairchild, J., Haussler, E., Zhang, C. R., and Challen, G. A. (2020). Divergent effects of Dnmt3a and Tet2 mutations on hematopoietic progenitor cell fitness. *Stem Cell Reports*, 14(4):551–560.

Parekh, S., Ziegenhain, C., Vieth, B., Enard, W., and Hellmann, I. (2018). zUMIs - a fast and flexible pipeline to process RNA sequencing data with UMIs. *Gigascience*.

Parry, A., Rulands, S., and Reik, W. (2021). Active turnover of DNA methylation during cell fate decisions. *Nature Reviews Genetics*, 22(1):59–66.

Pastrana, E., Silva-Vargas, V., and Doetsch, F. (2011). Eyes wide open: a critical review of sphere-formation as an assay for stem cells. *Cell Stem Cell*, 8(5):486–498.

Pedersen, B. S., Eyring, K., De, S., Yang, I. V., and Schwartz, D. A. (2014). Fast and accurate alignment of long bisulfite-seq reads. *arχiv preprint. arχiv:1401.1129*.

Petryniak, M. A., Potter, G. B., Rowitch, D. H., and Rubenstein, J. L. (2007). Dlx1 and Dlx2 control neuronal versus oligodendroglial cell fate acquisition in the developing forebrain. *Neuron*, 55(3):417–433.

Picelli, S., Björklund, Å. K., Faridani, O. R., Sagasser, S., Winberg, G., and Sandberg, R. (2013). Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nature Methods*, 10(11):1096–1098.

Platel, J. and Bordey, A. (2016). The multifaceted subventricular zone astrocyte: From a metabolic and pro-neurogenic role to acting as a neural stem cell. *Neuroscience*, 323:20–28.

Pliner, H. A., Packer, J. S., McFaline-Figueroa, J. L., Cusanovich, D. A., Daza, R. M., Aghamirzaie, D., Srivatsan, S., Qiu, X., Jackson, D., Minkina, A., et al. (2018). Cicero predicts cis-regulatory DNA interactions from single-cell chromatin accessibility data. *Molecular Cell*, 71(5):858–871.

Pliner, H. A., Shendure, J., and Trapnell, C. (2019). Supervised classification enables rapid annotation of cell atlases. *Nature Methods*, 16(10):983–986.

Ponti, G., Obernier, K., and Alvarez-Buylla, A. (2013). Lineage progression from stem cells to new neurons in the adult brain ventricular-subventricular zone. *Cell Cycle*, 12(11):1649–1650.

Pozniak, C. D., Langseth, A. J., Dijkgraaf, G. J., Choe, Y., Werb, Z., and Pleasure, S. J. (2010). Sox10 directs neural stem cells toward the oligodendrocyte lineage by decreasing suppressor of fused expression. *Proceedings of the National Academy of Sciences*, 107(50):21795–21800.

Price, A. J., Collado-Torres, L., Ivanov, N. A., Xia, W., Burke, E. E., Shin, J. H., Tao, R., Ma, L., Jia, Y., Hyde, T. M., et al. (2019). Divergent neuronal DNA methylation patterns across human cortical development reveal critical periods and a unique role of CpH methylation. *Genome Biology*, 20:1–20.

Prinz, M., Jung, S., and Priller, J. (2019). Microglia biology: one century of evolving concepts. *Cell*, 179(2):292–311.

Qiu, P. (2020). Embracing the dropouts in single-cell RNA-seq analysis. *Nature Communications*, 11(1):1169.

Qiu, X., Mao, Q., Tang, Y., Wang, L., Chawla, R., Pliner, H. A., and Trapnell, C. (2017). Reversed graph embedding resolves complex single-cell trajectories. *Nature Methods*, 14(10):979–982.

Qu, Q., Sun, G., Li, W., Yang, S., Ye, P., Zhao, C., Yu, R. T., Gage, F. H., Evans, R. M., and Shi, Y. (2010). Orphan nuclear receptor TLX activates Wnt/β-catenin signalling to stimulate neural stem cell proliferation and self-renewal. *Nature Cell Biology*, 12(1):31–40.

Quinlan, A. R. and Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26(6):841–842.

Ramsköld, D., Wang, E. T., Burge, C. B., and Sandberg, R. (2009). An abundance of ubiquitously expressed genes revealed by tissue transcriptome sequence data. *PLoS Computational Biology*, 5(12):e1000598.

Rasmussen, R. N., Asiminas, A., Carlsen, E. M. M., Kjaerby, C., and Smith, N. A. (2023). Astrocytes: integrators of arousal state and sensory context. *Trends in Neurosciences*.

Reizel, Y., Sabag, O., Skversky, Y., Spiro, A., Steinberg, B., Bernstein, D., Wang, A., Kieckhaefer, J., Li, C., Pikarsky, E., et al. (2018). Postnatal DNA demethylation and its role in tissue maturation. *Nature Communications*, 9(1):1–11.

Reizel, Y., Spiro, A., Sabag, O., Skversky, Y., Hecht, M., Keshet, I., Berman, B. P., and Cedar, H. (2015). Gender-specific postnatal demethylation and establishment of epigenetic memory. *Genes & Development*, 29(9):923–933.

Ricard, J., Salinas, J., Garcia, L., and Liebl, D. J. (2006). EphrinB3 regulates cell proliferation and survival in adult neurogenesis. *Molecular and Cellular Neuroscience*, 31(4):713–722.

Rodríguez Murúa, S., Farez, M. F., and Quintana, F. J. (2022). The immune response in multiple sclerosis. *Annual Review of Pathology: Mechanisms of Disease*, 17:121–139.

Rollins, R. A., Haghighi, F., Edwards, J. R., Das, R., Zhang, M. Q., Ju, J., and Bestor, T. H. (2006). Large-scale structure of genomic methylation patterns. *Genome Research*, 16(2):157–163.

Salzer, J. L. (2015). Schwann cell myelination. *Cold Spring Harbor Perspectives in Biology*, 7(8):a020529.

Saxonov, S., Berg, P., and Brutlag, D. L. (2006). A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *Proceedings of the National Academy of Sciences*, 103(5):1412–1417.

Schlosberg, C. E., VanderKraats, N. D., and Edwards, J. R. (2017). Modeling complex patterns of differential DNA methylation that associate with gene expression changes. *Nucleic Acids Research*, 45(9):5100–5111.

Schneider, J., Karpf, J., and Beckervordersandforth, R. (2019). Role of astrocytes in the neurogenic niches. *Astrocytes: Methods and Protocols*, pages 19–33.

Schönung, M., Hartmann, M., Krämer, S., Stäble, S., Hakobyan, M., Kleinert, E., Aurich, T., Cobanoglu, D., Heidel, F. H., Fröhling, S., et al. (2023). Dynamic DNA methylation reveals novel cis-regulatory elements in mouse hematopoiesis. *Experimental Hematology*, 117:24–42.

Schultz, M. D., He, Y., Whitaker, J. W., Hariharan, M., Mukamel, E. A., Leung, D., Rajagopal, N., Nery, J. R., Urich, M. A., Chen, H., et al. (2015). Human body epigenome maps reveal noncanonical DNA methylation variation. *Nature*, 523(7559):212–216.

Shareef, S. J., Bevill, S. M., Raman, A. T., Aryee, M. J., van Galen, P., Hovestadt, V., and Bernstein, B. E. (2021). Extended-representation bisulfite sequencing of gene regulatory elements in multiplexed samples and single cells. *Nature Biotechnology*, 39(9):1086–1094.

Shenghui, H., Nakada, D., and Morrison, S. J. (2009). Mechanisms of stem cell self-renewal. *Annual Review of Cell and Developmental Biology*, 25:377–406.

Shi, Y., Chichung Lie, D., Taupin, P., Nakashima, K., Ray, J., Yu, R. T., Gage, F. H., and Evans, R. M. (2004). Expression and function of orphan nuclear receptor TLX in adult neural stem cells. *Nature*, 427(6969):78–83.

Shukla, S., Kavak, E., Gregory, M., Imashimizu, M., Shutinoski, B., Kashlev, M., Oberdoerffer, P., Sandberg, R., and Oberdoerffer, S. (2011). CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. *Nature*, 479(7371):74–79.

Sinsheimer, R. L., Koerner, J. F., Vadla, J., and Lunan, K. (1954). The action of pancreatic desoxyribonuclease. I. isolation of mono-and dinucleotides. *Journal of Biological Chemistry*, 208(1):445–459.

Sloan, S. A. and Barres, B. A. (2014). Mechanisms of astrocyte development and their contributions to neurodevelopmental disorders. *Current Opinion in Neurobiology*, 27:75–81.

Smallwood, S. A., Lee, H. J., Angermueller, C., Krueger, F., Saadeh, H., Peat, J., Andrews, S. R., Stegle, O., Reik, W., and Kelsey, G. (2014). Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity. *Nature Methods*, 11(8):817–820.

Sofroniew, M. V. and Vinters, H. V. (2010). Astrocytes: biology and pathology. *Acta Neuropathologica*, 119:7–35.

Sohn, J., Orosco, L., Guo, F., Chung, S.-H., Bannerman, P., Ko, E. M., Zarbalis, K., Deng, W., and Pleasure, D. (2015). The subventricular zone continues to generate corpus callosum and rostral migratory stream astroglia in normal adult mice. *Journal of Neuroscience*, 35(9):3756–3763.

Soneson, C. and Robinson, M. D. (2018). Bias, robustness and scalability in single-cell differential expression analysis. *Nature Methods*, 15(4):255–261.

Spassky, N., Merkle, F. T., Flames, N., Tramontin, A. D., García-Verdugo, J. M., and Alvarez-Buylla, A. (2005). Adult ependymal cells are postmitotic and are derived from radial glial cells during embryogenesis. *Journal of Neuroscience*, 25(1):10–18.

Stadler, M. B., Murr, R., Burger, L., Ivanek, R., Lienert, F., Schöler, A., Nimwegen, E. v., Wirbelauer, C., Oakeley, E. J., Gaidatzis, D., et al. (2011). DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature*, 480(7378):490–495.

Stevens, B., Allen, N. J., Vazquez, L. E., Howell, G. R., Christopherson, K. S., Nouri, N., Micheva, K. D., Mehalow, A. K., Huberman, A. D., Stafford, B., et al. (2007). The classical complement cascade mediates CNS synapse elimination. *Cell*, 131(6):1164–1178.

Street, K., Risso, D., Fletcher, R. B., Das, D., Ngai, J., Yosef, N., Purdom, E., and Dudoit, S. (2018). Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *BMC genomics*, 19(1):1–16.

Stricker, S. H. and Götz, M. (2018). DNA-methylation: master or slave of neural fate decisions? *Frontiers in Neuroscience*, 12:5.

Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck III, W. M., Hao, Y., Stoeckius, M., Smibert, P., and Satija, R. (2019). Comprehensive integration of single-cell data. *Cell*, 177(7):1888–1902.

Stuart, T., Srivastava, A., Madad, S., Lareau, C. A., and Satija, R. (2021). Single-cell chromatin state analysis with Signac. *Nature Methods*, 18(11):1333–1341.

Suzuki, T., Maeda, S., Furuhata, E., Shimizu, Y., Nishimura, H., Kishima, M., and Suzuki, H. (2017). A screening system to identify transcription factors that induce binding site-directed DNA demethylation. *Epigenetics & Chromatin*, 10(1):1–14.

Svensson, V. (2020). Droplet scRNA-seq is not zero-inflated. *Nature Biotechnology*, 38(2):147–150.

Tahiliani, M., Koh, K. P., Shen, Y., Pastor, W. A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L. M., Liu, D. R., Aravind, L., et al. (2009). Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science*, 324(5929):930–935.

Tchieu, J., Calder, E. L., Guttikonda, S. R., Gutzwiller, E. M., Aromolaran, K. A., Steinbeck, J. A., Goldstein, P. A., and Studer, L. (2019). NFIA is a gliogenic switch enabling rapid derivation of functional human astrocytes from pluripotent stem cells. *Nature Biotechnology*, 37(3):267–275.

Teichmann, S. and Efremova, M. (2020). Method of the Year 2019: single-cell multimodal omics. *Nature Methods*, 17(1):2020.

Teif, V. B., Beshnova, D. A., Vainshtein, Y., Marth, C., Mallm, J.-P., Höfer, T., and Rippe, K. (2014). Nucleosome repositioning links DNA (de)methylation and differential CTCF binding during stem cell development. *Genome Research*, 24(8):1285–1295.

Tian, H., He, Y., Xue, Y., and Gao, Y. Q. (2022). Expression regulation of genes is linked to their CpG density distributions around transcription start sites. *Life Science Alliance*, 5(9).

Tillotson, R. and Bird, A. (2020). The molecular basis of MeCP2 function in the brain. *Journal of Molecular Biology*, 432(6):1602–1623.

Tinney, W. F. and Walker, J. W. (1967). Direct solutions of sparse network equations by optimally ordered triangular factorization. *Proceedings of the IEEE*, 55(11):1801–1809.

Traag, V. A., Waltman, L., and Van Eck, N. J. (2019). From Louvain to Leiden: guaranteeing well-connected communities. *Scientific Reports*, 9(1):1–12.

Tran, R. K., Henikoff, J. G., Zilberman, D., Ditt, R. F., Jacobsen, S. E., and Henikoff, S. (2005). DNA methylation profiling identifies CG methylation clusters in arabidopsis genes. *Current Biology*, 15(2):154–159.

Tsompana, M. and Buck, M. J. (2014). Chromatin accessibility: a window into the genome. *Epigenetics & Chromatin*, 7(1):1–16.

Vaisvila, R., Ponnaluri, V. C., Sun, Z., Langhorst, B. W., Saleh, L., Guan, S., Dai, N., Campbell, M. A., Sexton, B. S., Marks, K., et al. (2021). Enzymatic methyl sequencing detects DNA methylation at single-base resolution from picograms of DNA. *Genome Research*, 31(7):1280–1289.

Van den Berge, K., Hembach, K. M., Soneson, C., Tiberi, S., Clement, L., Love, M. I., Patro, R., and Robinson, M. D. (2019). RNA sequencing data: hitchhiker's guide to expression analysis. *Annual Review of Biomedical Data Science*, 2:139–173.

Wagner, A., Regev, A., and Yosef, N. (2016). Revealing the vectors of cellular identity with single-cell genomics. *Nature Biotechnology*, 34(11):1145–1160.

Wang, H., Kulas, J. A., Wang, C., Holtzman, D. M., Ferris, H. A., and Hansen, S. B. (2021). Regulation of beta-amyloid production in neurons by astrocyte-derived cholesterol. *Proceedings of the National Academy of Sciences*, 118(33):e2102191118.

Wang, T., Fowler, J. M., Liu, L., Loo, C. E., Luo, M., Schutsky, E. K., Berríos, K. N., DeNizio, J. E., Dvorak, A., Downey, N., et al. (2023). Direct enzymatic sequencing of 5-methylcytosine at single-base resolution. *Nature Chemical Biology*, pages 1–9.

Wang, Y.-Z., Plane, J. M., Jiang, P., Zhou, C. J., and Deng, W. (2011). Concise review: quiescent and active states of endogenous adult neural stem cells: identification and characterization. *Stem cells*, 29(6):907–912.

Webb, L. M. and Guerau-de Arellano, M. (2017). Emerging role for methylation in multiple sclerosis: beyond DNA. *Trends in Molecular Medicine*, 23(6):546–562.

Weinreb, C., Rodriguez-Fraticelli, A., Camargo, F. D., and Klein, A. M. (2020). Lineage tracing on transcriptional landscapes links state to fate during differentiation. *Science*, 367(6479):eaaw3381.

Wen, L. and Tang, F. (2022). Recent advances in single-cell sequencing technologies. *Precision Clinical Medicine*, 5(1):pbac002.

Wolf, F. A., Angerer, P., and Theis, F. J. (2018). SCANPY: large-scale single-cell gene expression data analysis. *Genome Biology*, 19(1):1–5.

Wu, H., Wu, X., Shen, L., and Zhang, Y. (2014). Single-base resolution analysis of active DNA demethylation using methylase-assisted bisulfite sequencing. *Nature Biotechnology*, 32(12):1231–1240.

Xu, C. and Corces, V. G. (2018). Nascent DNA methylome mapping reveals inheritance of hemimethylation at CTCF/cohesin sites. *Science*, 359(6380):1166–1170.

Yamaguchi, M., Saito, H., Suzuki, M., and Mori, K. (2000). Visualization of neurogenesis in the central nervous system using nestin promoter-GFP transgenic mice. *Neuroreport*, 11(9):1991–1996.

Yeung, M. S., Zdunek, S., Bergmann, O., Bernard, S., Salehpour, M., Alkass, K., Perl, S., Tisdale, J., Possnert, G., Brundin, L., et al. (2014). Dynamics of oligodendrocyte generation and myelination in the human brain. *Cell*, 159(4):766–774.

Yin, Y., Morgunova, E., Jolma, A., Kaasinen, E., Sahu, B., Khund-Sayeed, S., Das, P. K., Kivioja, T., Dave, K., Zhong, F., et al. (2017). Impact of cytosine methylation on DNA binding specificities of human transcription factors. *Science*, 356(6337):eaaj2239.

You, J. S., Kelly, T. K., De Carvalho, D. D., Taberlay, P. C., Liang, G., and Jones, P. A. (2011). OCT4 establishes and maintains nucleosome-depleted regions that provide additional layers of epigenetic regulation of its target genes. *Proceedings of the National Academy of Sciences*, 108(35):14497–14502.

Yu, G., Wang, L.-G., and He, Q.-Y. (2015). ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics*, 31(14):2382–2383.

Zamanian, J. L., Xu, L., Foo, L. C., Nouri, N., Zhou, L., Giffard, R. G., and Barres, B. A. (2012). Genomic analysis of reactive astrogliosis. *Journal of Neuroscience*, 32(18):6391–6410.

Zamboni, M., Llorens-Bobadilla, E., Magnusson, J. P., and Frisén, J. (2020). A widespread neurogenic potential of neocortical astrocytes is induced by injury. *Cell Stem Cell*, 27(4):605–617.

Zhang, K., Chen, S., Yang, Q., Guo, S., Chen, Q., Liu, Z., Li, L., Jiang, M., Li, H., Hu, J., et al. (2022). The oligodendrocyte transcription factor 2 OLIG2 regulates transcriptional repression during myelinogenesis in rodents. *Nature Communications*, 13(1):1–13.

Zhang, M., Wang, J., Zhang, K., Lu, G., Liu, Y., Ren, K., Wang, W., Xin, D., Xu, L., Mao, H., et al. (2021). Ten-eleven translocation 1 mediated-DNA hydroxymethylation is required for myelination and remyelination in the mouse brain. *Nature Communications*, 12(1):1–21.

Zheng, G. X., Terry, J. M., Belgrader, P., Ryvkin, P., Bent, Z. W., Wilson, R., Ziraldo, S. B., Wheeler, T. D., McDermott, G. P., Zhu, J., et al. (2017). Massively parallel digital transcriptional profiling of single cells. *Nature Communications*, 8(1):14049.

Zhou, B., Osinski, J. M., Mateo, J. L., Martynoga, B., Sim, F. J., Campbell, C. E., Guillemot, F., Piper, M., and Gronostajski, R. M. (2015). Loss of NFIX transcription factor biases postnatal neural stem/progenitor cells toward oligodendrogenesis. *Stem Cells and Development*, 24(18):2114–2126.

Zhou, Q., Wang, S., and Anderson, D. J. (2000). Identification of a novel family of oligodendrocyte lineage-specific basic helix–loop–helix transcription factors. *Neuron*, 25(2):331–343.

Zhu, C., Preissl, S., and Ren, B. (2020). Single-cell multimodal omics: the power of many. *Nature Methods*, 17(1):11–14.

Zhu, J., He, F., Hu, S., and Yu, J. (2008). On the nature of human housekeeping genes. *Trends in Genetics*, 24(10):481–484.

Ziller, M. J., Gu, H., Müller, F., Donaghey, J., Tsai, L. T.-Y., Kohlbacher, O., De Jager, P. L., Rosen, E. D., Bennett, D. A., Bernstein, B. E., et al. (2013). Charting a dynamic DNA methylation landscape of the human genome. *Nature*, 500(7463):477–481.

Zywitza, V., Misios, A., Bunatyan, L., Willnow, T. E., and Rajewsky, N. (2018). Single-cell transcriptomics characterizes cell types in the subventricular zone and uncovers molecular defects impairing adult neurogenesis. *Cell Reports*, 25(9):2457–2469.