From Dept. of Medical Biochemistry and Biophysics
Karolinska Institutet, Stockholm, Sweden

NAVIGATING THE OLIGODENDROGLIAL CHROMATIN LANDSCAPE USING HIGH-THROUGHPUT SINGLE-CELL EPIGENOMICS

Mukund Kabbe

Karolinska Institutet

Stockholm 2024

Published by Karolinska Institutet.
Printed by Universitetsservice US-AB, 2024

ISBN 978-91-8017-354-4
Cover illustration: *Extracting epigenetic information from the chromatin of individual neural cells.*
Artwork by Amagoia Agirre

Navigating The Oligodendroglial Chromatin Landscape Using High-Throughput Single-Cell Epigenomics

Thesis for Doctoral Degree (Ph.D.)

By

Mukund Kabbe

The thesis will be defended in public at **08:30 on Friday, May 24th 2024**
at **Eva & Georg Klein Hall**, Biomedicum, Solnavägen 9, Stockholm

Principal Supervisor:
Prof. Gonçalo Castelo-Branco
Karolinska Institutet
Department of Medical Biochemistry and Biophysics
Division of Molecular Neurobiology

Co-supervisor(s):
Dr. Marek Bartosovic
Stockholm University
Department of Biochemistry & Biophysics

Assoc. Prof. Simon Elsässer
Karolinska Institutet
Department of Medical Biochemistry and Biophysics
Division of Genome Biology

Opponent:
Prof. Bing Ren
University of California, San Diego
Department of Cellular and Molecular Medicine

Examination Board:
Assoc. Prof. Claudio Cantu
Linköping University
Department of Biomedical and Clinical Sciences

Prof. Rickard Sandberg
Karolinska Institutet
Department of Cell and Molecular Biology

Prof. Johan Jakobsson
Lund University
Department of Experimental Medical Science

To my parents

"The further you are from shore, the deeper the water gets."

- Venetian proverb

Popular science summary of the thesis

DNA is the genetic blueprint of every cell, housed within the nucleus. This blueprint acts as an instruction manual, dictating the code that generates every cell type in the body. Epigenetic mechanisms, also called gene regulatory mechanisms, control how this manual is interpreted and ensure that each cell type only has access to the instructions relevant to its function. Studying these mechanisms can help inform us about how cell types are specified, how they differ from each other, and importantly, how things go awry in disease. The work contained in this thesis is focused on navigating the many layers of the gene regulatory landscape in a specific cell lineage called oligodendroglia.

The oligodendroglial lineage, found in the brain and spinal cord, is composed of oligodendrocytes, and their progenitor population, oligodendrocyte progenitor cells (OPCs). Oligodendrocytes are responsible for producing a fatty membrane called myelin, which wraps tightly around neurons to form the myelin sheath. The myelin sheath provides insulation to the neuron, which significantly increases the speed of the electrical signal (called the action potential) passing through the neuron.

Multiple sclerosis (MS) is an inflammatory disease where the body's immune system attacks oligodendrocytes and myelin. The loss of the insulating myelin sheath disrupts the transmission of the action potential and the communication between neurons which leads to motor impairments in patients with the disease. Long thought to simply be victims of the attack, oligodendroglia were recently shown to be active responders and modulators of the inflammatory cascade, through the expression of several immune-associated genes.

In **Paper I**, we sought to understand the regulatory mechanisms through which disease-associated immune genes are activated in oligodendroglia in disease, using a mouse model of MS. In **Paper III**, we switched our focus to the human CNS and investigated regional differences in the regulatory landscape of the oligodendroglial lineage. The second broad aim of this thesis was to develop new tools to further study gene regulation. These efforts are captured in **Paper II** and **Paper IV**.

Abstract

The central nervous system (CNS), which includes the brain and spinal cord, is a remarkably complex system composed of billions of cells controlling everything ranging from basic metabolic functions to higher executive processes. This is achieved through the coordinated actions between neurons and glia. Oligodendrocytes, a specific glial population, are the myelinating cells of the CNS and wrap neuronal axons with the myelin sheath to facilitate the action potential and coordinate neuronal circuits. Recent years have shown that these cells, along with their progenitor population – OPCs – are more transcriptionally and functionally diverse than was previously believed. The epigenome acts as a function that converts the conserved genetic code to the diverse phenotypes that are observed, and so studying the epigenome can provide insights into how these states arise. The rise of high-throughput single-cell sequencing technologies has enabled the profiling of complex tissues and unveiled cellular diversity in the epigenome and transcriptome.

This thesis aims to broaden our understanding of gene regulation in the oligodendroglial lineage in both healthy and diseased contexts using a range of single-cell epigenomic methods.

In **Paper I** we used single-nucleus ATAC-seq to investigate the chromatin landscape of oligodendroglia in experimental autoimmune encephalomyelitis (EAE), a mouse model of Multiple sclerosis. We sought to understand the regulatory mechanisms that underlie the immune-like transcriptomic states that oligodendroglia exhibit at peak EAE. We found that healthy oligodendroglia exhibit primed chromatin at a subset of immune genes. We then show that these genes are activated in EAE, through a coordinated action involving the Polycomb Repressive Complex 2, and alterations in the histone landscape.

In **Paper II**, we developed scCUT&Tag, a new single-cell technology that enables the profiling of histone modifications and transcription factors (TFs). We applied the method to the juvenile mouse brain targeting the histone modifications H3K27me3, H3K27ac, H3K4me3, and H3K36me3, as well as the chromatin-associated proteins OLIG2 and RAD21. We use scCUT&Tag to delineate regulatory principles such as promoter bivalency, H3K4me3 spreading and promoter-enhancer interactions.

In **Paper III**, we used snATAC-seq and nanoCUT&Tag to profile adult human CNS tissue and capture the chromatin accessibility, H3K27me3 and H3K27ac histone landscape in different neural cell types. We unveiled a primed chromatin signature at the development-associated HOX genes in spinal cord-derived oligodendroglia (OLG). Using Micro-C to profile the chromatin architecture, we found that iPS-derived human OPCs exhibit a HOX architecture that is compatible both with the primed chromatin state seen in the adult OLGs and with high-grade pontine gliomas. Our results suggest that spinal cord-derived adult OLGs retain epigenetic memory of these genes, which may enable them to promptly transcribe these genes in regenerative contexts but may also make them susceptible to gliomagenesis.

In **Paper IV**, we developed nanoCTAR (pronounced "nano-star") a 4-in-1 multimodal single-cell technology. We used nanoCTAR to simultaneously capture accessible chromatin, two histone modifications and the transcriptome in single cells from the developing mouse brain. We showcase nanoCTAR as a versatile, easy-to-implement, cost-effective method for high-complexity single-cell profiling.

List of scientific papers

I. Epigenomic priming of immune genes implicates oligodendroglia in multiple sclerosis susceptibility

Mandy Meijer*, Eneritz Agirre*, **Mukund Kabbe**, Cassandra A. van Tuijn, Abeer Heskol, Chao Zheng, Ana Mendanha Falcao, Marek Bartosovic, Leslie Kirby, Daniela Calini, Michael R. Johnson, M. Ryan Corces, Thomas J. Montine, Xingqi Chen, Howard Y. Chang, Dheeraj Malhotra, Gonçalo Castelo-Branco

Neuron *2022*

II. Single-cell CUT&Tag profiles histone modifications and transcription factors in complex tissues

Marek Bartosovic, **Mukund Kabbe**, Gonçalo Castelo-Branco

Nature Biotechnology *2021*

III. Single-nuclei histone modification profiling of the adult human central nervous system unveils epigenetic memory of developmental programs

Mukund Kabbe, Eneritz Agirre, Karl E. Carlström, Fabio Baldivia Pohl, Nicolas Ruffin, David van Bruggen, Mandy Meijer, Luise A. Seeker, Nadine Bestard-Cuche, Alex R. Lederer, Jilin Zhang, Virpi Ahola, Steven A. Goldman, Marek Bartosovic, Maja Jagodic, Anna Williams, Gonçalo Castelo-Branco

bioRxiv *2024*

IV. Multimodal single-cell epigenome and transcriptome co-profiling

Mukund Kabbe*, Mattia Zaghi*, Oluwatoba Ajani, Naomi Rijk, Marek Bartosovic, Gonçalo Castelo-Branco

Manuscript *2024*

** Equal Contribution*

Scientific papers not included in this thesis

Developmental landscape of human forebrain at a single-cell level identifies early waves of oligodendrogenesis

David van Bruggen, Fabio Pohl, Christoffer Mattson Langseth, Petra Kukanja, Hower Lee, Alejandro Mossi Albiach, **Mukund Kabbe**, Mandy Meijer, Sten Linnarsson, Markus M Hilscher, Mats Nilsson, Erik Sundström, Gonçalo Castelo-Branco

Developmental Cell *2022*

Brain Matters: unveiling the distinct contributions of region, age, and sex to glia diversity and CNS function

Luise A. Seeker, Nadine Bestard-Cuche, Sarah Jäkel, Nina-Lydia Kazakou, Sunniva M. K. Bøstrand, Laura J. Wagstaff, Justyna Cholewa-Waclaw, Alastair M. Kilpatrick, David Van Bruggen, **Mukund Kabbe**, Fabio Baldivia Pohl, Zahra Moslehi, Neil C. Henderson, Catalina A. Vallejos, Gioele La Manno, Gonçalo Castelo-Branco, Anna Williams

Acta Neuropathologica Communications *2023*

Distinct transcriptomic and epigenomic responses of mature oligodendrocytes during disease progression in a mouse model of multiple sclerosis

Chao Zheng*, Bastien Hervé*, Mandy Meijer, Leslie Ann Rubio Rodríguez-Kirby, André Ortlieb Guerreiro Cacais, Petra Kukanja, **Mukund Kabbe**, Tomas Olsson, Eneritz Agirre, Gonçalo Castelo-Branco

bioRxiv *2023*

Contents

List of abbreviations

A-P	Anterior-Posterior
ABC	Activity-by-contact model
AST	Astrocyte
ATAC-seq	Assay for transposase-accessible chromatin with sequencing
ATP	Adenosine triphosphate
BBB	Blood-brain barrier
BMP	Bone Morphogenetic protein
c-Dom	Centromeric domain
CBEX	Cerebellar excitatory neuron
CBINH	Cerebellar inhibitory neuron
CCA	Canonical correlation analysis
cCRE	Candidate cis-regulatory element
CDC42EP1	CDC42 Effector Protein 1
CFA	Complete Freud's adjuvant
ChIP	Chromatin immunoprecipitation
CNS	Central nervous system
COP	Committed OPC
CRISPR	Clustered regularly interspaced short palindromic repeats
CTCF	CCCTC-binding factor
CUT&RUN	Cleavage under target and release using nuclease
CUT&Tag	Cleavage under target and tagmentation
CXEX	Cortical excitatory neuron
CXINH	Cortical inhibitory neuron
D-V	Dorsal-Ventral
dCas9	Dead Cas9
DHS	DNase Hypersensitive
DMG	Diffuse midline glioma
DNA	Deoxyribonucleic acid
E	Embryonic day
EAE	Experimental autoimmune encephalomyelitis
EEL	Enhanced Electric FISH
EGA	European Genome-phenome Archive
ENCODE	Encyclopaedia of DNA elements
ENDO	Endothelial cell
EZH2	Enhancer of Zeste homolog 2
EZHIP	Enhancer of Zeste homolog inhibitory protein
FACS	Fluorescence-activated cell sorting
FAIR	Findability, Accessibility, Interoperability, and Reuse
FANTOM	Functional annotation of the mammalian genome

FGF	Fibroblast Growth Factor
FISH	Fluorescent in situ hybridization
GM	Gray matter
GRN	Gene regulatory network
GWAS	Genome-wide association study
H3K27M	Lysine 27 to Methionine mutation on H3
H3K27me3	Trimethylation of Lysine 27 on H3
HAT	Histone acetyltransferase
HC	Heterochromatin
HCA	Human cell atlas
HDAC	Histone deacetylase
HGG	High-grade glioma
HMM	Hidden Markov Model
HOX	Homeobox
hPTM	Histone post-translational modification
IFN-g	Interferon gamma
iPS-hOPC	iPS-derived human OPC
iPSC	Induced pluripotent stem cell
ISS	In situ Sequencing
KMT2A/B	Lysine-methyltransferase 2A/B
kNN	k-Nearest Neighbours
LGE	Lateral ganglionic eminence
LIANTI	Linear amplification via transposon insertion
LLM	Large language models
lncRNA	Long non-coding RNA
LSI	Latent semantic indexing
ME	Mosaic end
MERFISH	Multiplexed error-robust FISH
mESC	Mouse embryonic stem cell
MFOL	Myelin-forming oligodendrocyte
MGE	Medial ganglionic eminence
MHC	Major histocompatibility complex
mi-R	microRNA
MIGL	Microglia
MNase	Micrococcal nuclease
MOG	Myelin oligodendrocyte glycoprotein
MOL	Mature oligodendrocyte
mRNA	Messenger RNA
MS	Multiple sclerosis
nanoCTAR	nanoCUT&Tag + ATAC + RNA
nanoTn5	Nanobody-Tn5 fusion protein
NFOL	Newly formed oligodendrocyte

NGS	Next-Gen sequencing
NLP	Natural language processing
NPC	Neural progenitor cells
OL	Oligodendrocyte
OLG	Oligodendroglia
OPC	Oligodendrocyte progenitor cell
pA-MNase	Protein A-MNase fusion protein
pA-Tn5	Protein A-Tn5 fusion protein
PCA	Principal component analysis
PCR	Polymerase chain reaction
PCW	Post-conception week
PDGFRA	Platelet-derived growth factor receptor alpha
PFA-EP	Posterior fossa group A ependymomas
PLAC-seq	Proximity Ligation-assisted ChIP
PNS	Peripheral nervous system
PPMS	Primary progressive multiple sclerosis
pR1N	partial Read 1 Nextera
PRC2	Polycomb repressive complex 2
RA	Retinoic Acid
RNA	Ribonucleic acid
RRMS	Relapsing-remitting multiple sclerosis
RT	Reverse transcription
sc	Single-cell
SHH	Sonic hedgehog
sn	Single-nucleus
SNP	Single-nucleotide polymorphism
SOX10	SRY-Box transcription factor 10
SPMS	Secondary progressive multiple sclerosis
SVD	Singular vector decomposition
t-Dom	Telomeric domain
TAD	Topologically associating domain
TF	Transcription factor
TF-IDF	Term frequency-inverse document frequency
TSS	Transcription start site
VA	Valproic acid
WM	White matter
ZPA	Zone of polarizing activity
ZRS	ZPA regulatory sequence

1 Oligodendrocytes and the CNS

1.1 CNS development

The vertebrate central nervous system (CNS) is a complex organ system composed of the brain and spinal cord. It begins to form around the fourth week of gestation in humans, from the outermost germ layer of the developing embryo – the ectoderm. A specialised column of ectodermal cells along the dorsal midline of the embryo responds to inductive signals from neighbouring cells to form the neuroectoderm, which thickens to form the neural plate (Kandel, 2013). Through a process termed neurulation, the thickened neural plate invaginates and hollows out centrally to form a tube-like structure which closes on the dorsal surface to produce the fully enclosed neural tube. The cells of the neural tube will form the CNS, while the neuroectoderm cells at the neural plate border form the neural crest which will give rise to the cells of the peripheral nervous system (PNS). On both sides of the neural tube, the cells of the mesoderm form structures called somites which will subsequently give rise to the dermal skin, muscle, and bone (Yusuf & Brand-Saberi, 2006). The ventromedial mesoderm forms the notochord which secretes morphogens for patterning the CNS (Kandel, 2013). A series of rapid, non-uniform cell divisions along the length of the neural tube gives rise to specialised structures that will form distinct parts of the brain and spinal cord. The anterior (rostral) neural tube divides into three vesicles called the prosencephalon, mesencephalon, and rhombencephalon which will form the forebrain, midbrain, and hindbrain respectively, while the posterior (caudal) neural tube forms the spinal cord (Ishikawa et al., 2012).

1.2 Anterior-posterior and dorsal-ventral CNS patterning

The formation of distinct CNS structures is enabled through a complex pattern of morphogen signalling and transcription factor expression. The primary drivers of dorsoventral (DV) patterning in the spinal cord are Sonic Hedgehog (SHH), Bone Morphogenetic Proteins (BMP) and Wnt proteins. The cells of the floor plate and notochord which run along the ventral midline secrete SHH (Ericson et al., 1996), while those of the roof plate along the dorsal midline secrete BMP and Wnt proteins. The secretion of these signalling proteins from fixed sources creates a natural DV gradient with SHH concentration highest and lowest at the ventral

surface and dorsal surface respectively, with the opposite occurring for BMP and Wnt. The combinatorial action of these opposing morphogen gradients triggers different intracellular signalling cascades which lead to the expression of distinct transcription factors (TFs) (Alvarez-Medina et al., 2008; Bond et al., 2012; Cayuso & Martí, 2005). The spatial specificity of TF expression divides the DV axis into transient progenitor domains, which give rise to specific cell types.

The anterior-posterior (AP) axis, or rostrocaudal, axis is patterned through the action of several proteins, including the well-studied fibroblast growth factor (FGF), retinoic acid (RA) and family of homeobox proteins (HOX). While FGF plays a more prominent role in the anterior patterning of the telencephalon, RA acts in posteriorizing the neural tube (del Corral & Storey, 2004). The HOX family of transcription factors also play a prominent role in the developing hindbrain and spinal cord and is a prominent feature in **Paper III** of this thesis, and so will be elucidated on further.

1.3 HOX expression in development

The HOX genes belong to a family of homeodomain-containing TFs and are evolutionarily conserved across all bilaterian organisms (Garcia-Fernàndez, 2005). They were first discovered as master genes that regulated body segmentation in *Drosophila melanogaster* (Lewis, 1978). Since then, they have been deeply studied by evolutionary and developmental biologists due to their essential roles in organising the major body axis during development (Krumlauf, 1994). In most vertebrates, there are 39 HOX genes, each belonging to one of 13 paralogue groups. Individual paralogues are grouped in four clusters – HOXA, HOXB, HOXC, and HOXD – and are located on different chromosomes, though not all paralogues are present in all clusters (**Figure 1**). Each of the genes contains a core 60 amino acid-encoding DNA sequence, which forms the DNA-binding homeodomain of the translated protein.

During embryonic development, the expression of the HOX genes helps coordinate the AP patterning of the hindbrain and spinal cord (Philippidou & Dasen, 2013), and the pattern of expression is dictated by the physical location of the gene in the cluster, the location of the cell along the neural tube, and the period of development. This spatiotemporally locked pattern is termed Hox collinearity. Genes at the 3' ends of each HOX cluster are expressed earlier in development and in more anterior regions, while the genes at the 5' end are expressed later and in more posterior regions. The earliest HOX genes are expressed in the hindbrain,

where they segment and specify the transient rhombomeric domains (Lumsden, 2004) and the latest HOX genes are expressed in the lumbar regions of the spinal cord.

The precision of this HOX timer is controlled through several mechanisms including chromatin architecture, histone tail modifications and CTCF binding (Narendra et al., 2016; Noordermeer et al., 2011). Prior to their expression, the HOX genes exist in a closed and insulated domain and are coated with the repressive H3K27me3 mark. The activation of these genes is associated with the sequential opening of the chromatin, deposition of the active H3K27ac and H3K4me3 marks and the association of CTCF-binding sites with regulatory domains flanking the clusters (Noordermeer et al., 2011; Rekaik et al., 2023). The process of chromatin opening unfolds like a zipper, gradually opening from the 3' to the 5' end of the gene cluster. This exposes the genes slated for expression while still leaving the more 3' located genes accessible (Rekaik et al., 2023).

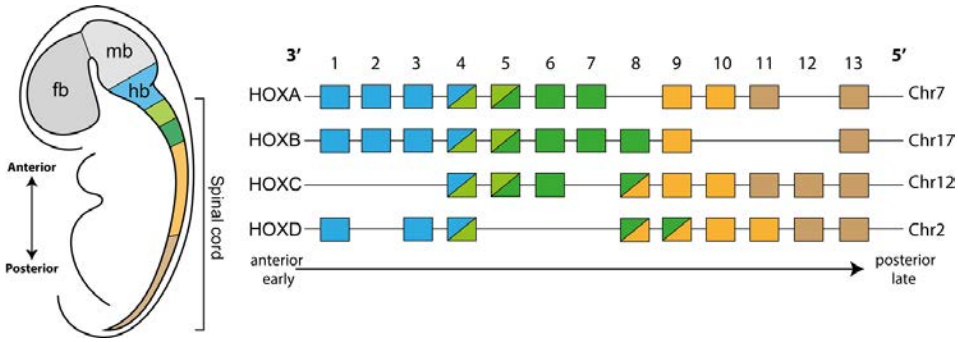

***Figure 1**: Spatiotemporal pattern of HOX activation in development. Left – schematic of developing brain and spinal cord showing forebrain (fb), midbrain (mb), hindbrain (hb) and broad segmentation of the spinal cord. Right – HOX gene arrangement in the four clusters, coloured by region of expression. 3' genes in clusters are expressed more anterior and early, while 5' genes are expressed more posterior and later in development.*

1.4 Oligodendrocyte biology

1.4.1 Overview of the lineage

Oligodendrocytes (OLs) are the resident myelinating cells of the CNS. They produce a lipid-rich myelin membrane which wraps around neuronal axons,

forming the myelin sheath. The myelin sheath provides electrical insulation to the axon, which increases the conduction velocity and enables saltatory conduction of the action potential down the length of the axon (Frankenhaeuser, 1952; Stämpfli, 1954). Unlike Schwann cells, the myelinating cells of the PNS which contact and wrap a single axon, a single oligodendrocyte can contact and ensheath multiple axons. The length and thickness of the myelin sheath are adaptive and can dynamically respond to neuronal activity to coordinate and fine-tune circuits (Gibson et al., 2014; Sherman & Brophy, 2005). In addition to providing structural support, oligodendrocytes also metabolically support axons by shuttling lactate through monocarboxylate transporters located at the myelin-axon interface (Fünfschilling et al., 2012). The lactate provided by the oligodendrocyte transporters is critical for neuronal survival and can serve as an alternative energy source when glucose availability is low (Brown et al., 2001; Tekkök et al., 2005).

The discovery of oligodendrocytes dates to the 1920s when the Spanish neuroscientist Pio del Rio Hortega identified them in mouse brain tissue sections using his silver carbonate staining method (del Río Hortega, 1922). Del Rio Hortega categorised OLs into four distinct types based on the morphological structures he observed. Type I OLs exhibited small cell bodies with numerous fine processes and were present in both white matter (WM) and grey matter (GM). Type II OLs had fewer, thicker processes and were exclusive to the WM; Type III possessed large cell bodies with thick processes and myelin sheaths in the WM, while Type IV OLs displayed an elongated morphology with bipolar processes extending towards large axons in WM of the brainstem (del Río Hortega, 1922; Pérez-Cerdá et al., 2015) .

OLs arise from a *Pdgfra* expressing progenitor population – oligodendrocyte progenitor cells (OPCs) – which unlike the WM-dominated OLs, are uniformly distributed across the CNS. The path from OPCs to OLs is not a binary step; single-cell transcriptomic studies have revealed that the differentiation trajectory is studded with many intermediate transcriptional states including committed OPCs (COPs), newly formed OLs (NFOLs), and myelin-forming OLs (MFOLs), before branching out into 6 classes of mature OLs (MOLs) (Marques et al., 2016). The transcriptional states seen in MOLs, and the morphological heterogeneity identified by del Rio Hortega have hinted at potential functional differences in the lineage, though this is still not completely understood. The border between the MOL sub-populations is blurry and not well defined and may reflect transient

transcriptional states between which the MOLs switch based on context. However, some regional, functional, and age-associated differences within the MOL populations have been observed. The MOL2 population is enriched in the spinal cord, relative to the brain, and more specifically within the WM tracts of the spinal cord. MOL5/6 on the other hand, show an increase in their abundance across the CNS with age, and within the spinal cord are preferentially enriched in GM. MOL2 also appear to be more susceptible to spinal cord injury (Floriddia et al., 2020).

1.4.2 Oligodendrocytes in development

Oligodendrogenesis occurs at distinct periods during embryonic CNS development. In the mouse brain, OPCs arise from neuroepithelial cells in three embryonic waves. The first telencephalic wave occurs at E12.5 around the ventricular zone of the medial ganglionic eminence (MGE) and is characterised by the expression of NKX2.1. After production, these ventral OPCs migrate radially outwards and dorsally to populate the entire cortex. The second wave of OPC generation occurs more dorsally at E15.5 around the lateral ganglionic eminence (LGE) and is marked by the expression of GSX2. The third population, expressing EMX1, arises before birth from the cortical subventricular zone (Kessaris et al., 2006). The third population largely supplants the first two and is the main OPC population found in the adult mouse brain (Richardson et al., 2006) (**Figure 2**, left).

A similar pattern of development occurs in the developing mouse spinal cord but with only two waves. The first SHH-dependent population (NKX2.1 positive) arises in the ventral pMN domain at E12.5, followed by a second SHH-independent wave in the dorsal dP3-dP5 domains around E15 (Cai et al., 2005; Fogarty et al., 2005; Kessaris et al., 2006; Pringle et al., 1996; Richardson et al., 2006; Vallstedt et al., 2005). The ventrally-derived population accounts for the majority, comprising 85% of the final spinal cord OPC pool, whereas the dorsal population constitutes the remaining 15% (Richardson et al., 2006) (**Figure 2**, right).

In the mouse brain and spinal cord, the OPCs from each wave transcriptionally converge after birth, and this homogenous pool begins to differentiate into oligodendrocytes around post-natal day 7 (p7) into the 6 classes of MOLs, as described earlier (Marques et al., 2016, 2018).

Studying the pattern of OPC formation in human foetal development is far more challenging, however, a similar pattern of temporally distinct waves appears to be

conserved (Jakovcevski & Zecevic, 2005; Mo & Zecevic, 2009; Rakic & Zecevic, 2003). OPCs were originally detectable at the 16th post-conception week (PCW), which corresponds to a later time point than the first embryonic wave in mice. This population was detected in larger numbers by PCW 22. Further, an EGFR-expressing population of pre-OPCs was detected at PCW 20-24. However, as OPCs were detected at PCW 16, this population likely corresponds to a second wave. The quest for the initiation of the first wave remained unsolved, until recently: Single-cell transcriptomic and spatial imaging of the human embryonic forebrain in the early first trimester revealed a population of pre-OPCs and OPCs arising as early as PCW 6-8 (Braun et al., 2023; van Bruggen et al., 2022), and a similar OPC production was seen in the first trimester of the developing human spinal cord (Rayon et al., 2021).

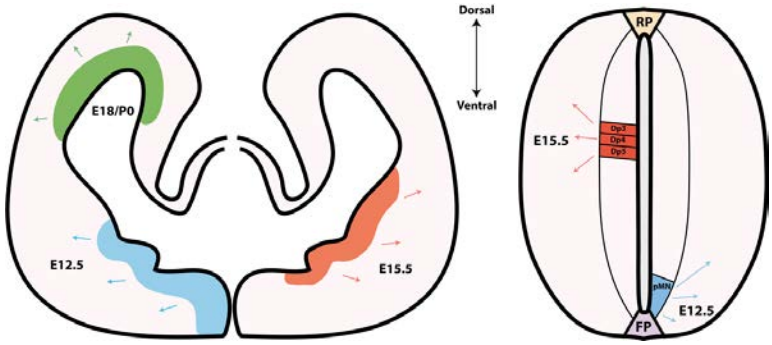

***Figure 2**: OPC specification in development. The first (blue), second (red), and third (green) waves of embryonic OPC specification in the brain (left) and spinal cord (right), showing the time window, region of specification and direction of migration. RP – roof plate. FP – floor plate.*

1.5 Demyelinating disease

1.5.1 Multiple Sclerosis

Multiple sclerosis (MS) is a chronic, inflammatory, multi-factorial autoimmune disease of the CNS, affecting over 2.3 million people globally (Thompson et al., 2018). Though the exact cause and aetiology of MS remains unknown, genetic risk factors, environment and lifestyle have all been linked to the occurrence of the disease. The geographical distribution of MS reveals a growing prevalence with increasing distance from the equator, hinting at environmental factors such as

sunlight exposure and vitamin D deficiency. From a genetic perspective, genome-wide association studies (GWAS) have identified over 200 MS risk loci in the genome, mostly associated with immune pathways, leukocyte activation, and antigen presentation (International Multiple Sclerosis Genetics Consortium, 2019; Ramagopalan et al., 2010).

MS is characterized by the presence of inflammatory, demyelinating lesions, predominantly in the WM of the brain and spinal cord. A targeted autoimmune attack mounted on the oligodendroglial lineage leads to the destruction of the myelin sheath (demyelination). This loss of electrical insulation impairs the neuron's conduction velocity which manifests as motor symptoms in patients. Neurons initially attempt to resolve this loss by redistributing Na^+ ion channels around the demyelinated segment to aid the non-saltatory conduction (England et al., 1990; Smith et al., 1982). However, the loss of myelin-dependent trophic support and the exposure of the axons to reactive oxygen species (ROS) in the inflammatory microenvironment leads to further neuronal damage.

In the early stages of the disease, the demyelinating attack is countered with a robust remyelination response, involving OPCs which are mobilized to lesion areas where they differentiate into oligodendrocytes and remyelinate the denuded axons (Franklin & Ffrench-Constant, 2008). This initial remyelination effort resolves the clinical symptoms in patients and is called the remission phase. However, repeated inflammatory attacks lead to a relapse into the demyelinated state, which is then met again with a remyelination effort. The presentation of MS displaying this back-and-forth pattern of injury and repair is known as relapsing-remitting MS (RRMS). As the disease progresses, however, the efficiency of the remyelination effort weakens and eventually fails altogether, which leads to a progressive worsening of symptoms. This second phase of continuous degeneration is called secondary progressive MS (SPMS). However, approximately 15% of patients develop the disease directly in the progressive state, known as primary progressive MS (PPMS) (Miller & Leary, 2007). Unfortunately, while several immunomodulatory therapies are available, MS remains largely incurable, and patients experience progressive degeneration.

Two opposing models have been proposed to explain the initiation of MS. The "outside-in" model posits that a dysregulation of the immune system in the periphery is the primary cause that drives the immune cells into the CNS and attacks the myelin. The "inside-out" model on the other hand posits a CNS-

intrinsic process or myelin-associated degeneration or injury that leads to the influx of the immune cells and subsequent inflammatory cascade (Titus et al., 2020).

1.5.2 Experimental Autoimmune Encephalomyelitis (EAE)

MS is a complex disease, and the pathophysiology is difficult to study in humans. To aid in this, several animal models of MS have been developed, which aim to recapitulate different aspects of the disease. One of the most common mouse models of MS is experimental autoimmune encephalomyelitis (EAE), which supports the "outside-in" model of disease and helps investigate the immune cascade following demyelinating injury (Baxter, 2007).

EAE is induced in mice through an immunization protocol targeting myelin-associated proteins, such as myelin oligodendrocyte glycoprotein (MOG). There are several options for inducing EAE - either by using a peptide from myelin oligodendrocyte glycoprotein (MOG), the whole MOG protein, or through the adoptive transfer of MOG-specific T cells, or the use of a different myelin-associated protein. The immunization used by our group involves an initial administration with an emulsion of the MOG_{35-55} peptide in complete Freud's adjuvant (CFA), which acts as an immune booster. This is complemented with a pertussis toxin injection to permeabilize the blood-brain barrier (BBB). The EAE pathology is primarily seen in the spinal cord, and symptoms develop from the base of the tail upwards – starting with limp tail and progressing up to limb paralysis.

1.5.3 Oligodendroglia in disease

In the traditional view of MS, oligodendrocytes were perceived as passive victims within the inflammatory cascade, while immune cells were regarded as the primary perpetrators. In addition, the role of OPCs was viewed as being confined to processes such as migration, proliferation, and differentiation at lesion sites. However, the discovery of immunomodulatory functions within the OLG lineage has challenged this view (Kirby & Castelo-Branco, 2021).

Using single-cell transcriptomics, our group has shown that the OLG lineage presents an altered disease-specific transcriptomic profile in EAE, not found in control mice. A subset of EAE-specific OPCs upregulated genes associated with major histocompatibility complexes I and II (MHC-I, MHC-II) and antigen presentation pathways (Falcão et al., 2018). The expression of the MHC-I and

MHC-II genes was shown to lead to a functional protein and enabled OPCs to cross-present antigens to T-cells, a process previously thought to be exclusive to professional antigen-presenting cells, such as dendritic cells (Falcão et al., 2018; Kirby et al., 2019). That OPCs could express MHC genes in response to the cytokine, interferon-gamma (IFN-g) was first seen in the 1980s (Calder et al., 1988; Suzumura et al., 1986; Wong et al., 1984), however, the functional role was unclear at the time. A different sub-set of OPCs showed expression of genes associated with immunosuppression, which suggested that these cells may play a role in decreasing inflammation (Falcão et al., 2018). However, whether the immune-OPCs are beneficial or detrimental to the disease progression remains unclear, and further investigation is necessary.

The transition of OPCs to an inflammatory, immune-like state also inhibits their differentiation potential, likely hampering their ability to participate in the remyelination process (Falcão et al., 2018; Kirby et al., 2019). However, recent findings in post-mortem human MS tissue have challenged the notion that remyelination is solely reliant on OPC differentiation: two independent studies used orthogonal approaches to show that existing oligodendrocytes can also contribute to remyelination, a finding that has reshaped our understanding of MS pathology (Jäkel et al., 2019; Yeung et al., 2019).

Yeung et.al. used a unique approach utilising radioactive ^{14}C birth-dating to estimate the age of cells within MS tissue. The approach was based on the context of elevated atmospheric ^{14}C levels that followed nuclear testing conducted in the early 1900s. As the atmospheric carbon entered the food chain through plants and subsequently into humans, it would be incorporated into cells during DNA replication. By leveraging the subsequent and steady decline of atmospheric ^{14}C over time, the age of a cell could be gauged based on the amount of radioactive material found within it (Yeung et al., 2014). Using this approach, the investigators found that a subset of patients had newly generated oligodendrocytes in the remyelinated lesions, indicating they were derived from OPC differentiation. However, many patients also presented remyelinated lesions with much older oligodendrocytes, that could not have been derived from OPCs after the demyelination event, suggesting that the previously existing oligodendrocytes were remyelinating the axons (Yeung et al., 2019).

A parallel study from Jäkel et.al. used single nucleus transcriptomics in adult MS autopsy material to show that the distribution of OPC and MOL populations in MS

was skewed compared to Control material. One population of stable, myelinating MOLs was almost absent in MS tissue, and another population with immune gene expression was almost exclusive to the disease. Further, the MOLs in the active lesions of MS displayed increased expression of genes associated with a myelination program, suggesting they were involved in the remyelination of these lesions. Remyelinated lesions on the other hand showed a depletion of OPCs and intermediate oligodendrocyte populations (Jäkel et al., 2019).

Collectively, these studies have unveiled novel insights into MS pathology and ushered in a new era of investigation into the roles that oligodendroglia play in the contexts of neuroimmunology and remyelination.

2 Epigenetic regulation of gene expression

2.1 Overview of the non-coding genome

As referenced in the abstract of this thesis, cellular diversity can be captured at the level of function, morphology, protein content and transcriptome. Despite this diversity, the genomic DNA from which mRNA is derived (and thereafter protein), remains largely identical across all somatic cells within an organism. For centuries, scientists have been fascinated with understanding how the complex human body arises from this static genetic code. While it was understood that DNA contained genes which encoded functional proteins, a puzzling phenomenon emerged regarding the ratio of protein-coding to non-coding regions within the genome (Hahn & Wray, 2002). Despite the human genome consisting of over 3.1 billion nucleotides, the protein-coding fraction accounts for a mere 1-2%. The remaining 98% was initially dismissed as "junk DNA" with no biological purpose. However, considering another paradox where the amount of genetic material seemed not to correlate with organismal complexity, scientists began to entertain the idea that perhaps the non-coding fraction may indeed harbour valuable information (Choi et al., 2020). Subsequently, we have discovered that various classes of regulatory elements within the non-coding genome are indeed critical in governing and shaping gene function (Shen et al., 2012; Thurman et al., 2012). Large-scale research initiatives such as ENCODE (Encyclopaedia of DNA elements), FANTOM (Functional annotation of the mammalian genome), ROADMAP Epigenomics Program, and the 4D Nucleome Program among others have contributed significantly to advancing our understanding of the role of the non-coding genome, and the regulatory principles governing gene function both in health and disease (Bernstein et al., 2010; de Hoon et al., 2015; Dekker et al., 2017; Dunham et al., 2012).

2.2 The epigenome and cellular diversity

Understanding the complex mechanism through which the static genome is parsed to give rise to the diverse array of phenotypes seen across all cells falls within the realm of epigenetics. Epigenetic mechanisms operate across multiple hierarchical levels, encompassing macro-environmental cues to micro-environmental signalling, including intracellular and intranuclear pathways. These

mechanisms extend down to chromatin restructuring and reorganisation, orchestrating gene expression and influencing cellular function. This information can also be stably inherited to ensure that daughter cells (from symmetric cell division) will generate the same cell type. At the core of epigenetics lies the concept that all cells in an organism are descendants of a single fertilized egg (zygote). This totipotent zygote divides to produce pluripotent stem cells, which further differentiate to generate all specialized cell types in an organism – all without any changes in the genome (Berger et al., 2009; Bird, 2007; Trerotola et al., 2015).

In 1957, the British geneticist Conrad Waddington, introduced a metaphor known as Waddington's landscape, which illustrated the epigenetic landscape and the process of cellular specialization. The metaphor places the totipotent zygote at the top of a hilly landscape. As the cell rolls down the hill, it traverses and enters different epigenetic valleys, which restrict it to specific cellular lineages and sub-lineages (Waddington, 2014). The landscape highlights how cells commit to lineages by making fate choices, restricting their epigenetic potential while enhancing functional specialization. This landscape, while useful to visualise epigenetic potential, has its limitations. A feature of the landscape is the uni-directional flow of information, implying that once a cell commits to a lineage, it cannot reverse its fate or move up to a higher potential state. However, the discovery of the Yamanaka re-programming factors in 2006 challenged this notion, by revealing that the coordinated expression of a four-transcription factor cocktail – OCT4, SOX2, KLF4, and c-MYC – could reprogram the expression profile of a cell to mimic a more pluripotent state, referred to as an induced pluripotent stem cell (iPSC) (Takahashi et al., 2007; Takahashi & Yamanaka, 2006). In the landscape analogy, this would be akin to rolling a cell back up the hill to an earlier point in time, making it possible to take on alternative fates (Hochedlinger & Plath, 2009).

2.3 Regulatory elements

The regulatory epigenome is incredibly complex and has multiple cis-acting and trans-acting elements that act in a synchronised manner to elicit transcriptional responses. Promoters and enhancers are key cis-regulatory elements (CRE) involved in initiating and modulating gene expression. Promoter elements are short DNA sequences typically proximal to the transcription start site (TSS) of a gene and serve as initiation hubs for RNA polymerase and transcription factors,

along with other members of the transcription complex to assemble before transcription (Deng & Roberts, 2005). Enhancers are another class of CREs, which are located distally from the promoter, and enhance the expression of the target gene. The first enhancer element was identified in 1982 by Banerji and colleagues, who found that the insertion of a short 72-bp sequence from the simian virus 40 (SV40) genome strongly upregulated the expression of the β-globin gene in HeLa cells (Banerji et al., 1981). The enhanced expression was independent of the location of the sequence relative to the β-globin promoter. This observation highlights two key features of genomic enhancers: first, they can act remotely and can be located very far from the target gene, and second, the function of the enhancer is orientation-agnostic (Schoenfelder & Fraser, 2019). Since then, many enhancers have been identified and well-studied. Enhancers contain a short DNA sequence or motif that is recognised by transcription factors (TF), a class of trans-acting elements. TFs contain evolutionarily conserved DNA binding domains that allow them to bind their cognate recognition motif at enhancers. TF binding triggers a cascade of events including chromatin remodelling, recruitment of the mediator complex, assembly of the transcriptional complex and eventually transcription. (Bergman et al., 2022; Rebeiz & Tsiantis, 2017) TF-enhancer interactions are a fascinating area of investigation. TFs often exhibit cell-type specificity, and with their specific recognition motifs, enhancers also display cell-type specificity. Enhancers can also present tissue specificity, adding another complex layer to gene regulation. However, for TFs to bind, they must have access to the DNA, which can be occluded by the presence of nucleosomes, a large nucleoprotein complex and the fundamental unit of chromatin (explained in detail in section **2.4**). This leads to another feature of active enhancers, which is their likelihood to be situated within open, nucleosome-free regions of the genome. However, a special class of TFs, called pioneer TFs, are uniquely capable of binding compact, nucleosome-bound DNA and initiating chromatin remodelling. Pioneer TFs play an important role in fate commitment and can drive the expression of lineage-specifying transcriptional programs (Iwafuchi-Doi & Zaret, 2014; Meers et al., 2019; Soufi et al., 2015).

Apart from enhancing gene expression, some elements are also involved in repressing transcription or insulating from enhancer activity. One classic example is the CCCTC-binding factor, CTCF. CTCF is a well-characterized, highly conserved DNA-binding protein involved in organising the 3D chromatin architecture (Lobanenkov et al., 1990; Ong & Corces, 2014; Phillips & Corces,

2009). Broadly, CTCF can organise the genome in a manner that can insulate gene promoters from enhancers, or even promote and facilitate the interaction. The specific mechanism through which CTCF achieves this is explained in section **2.6.1**.

In addition to these elements, several other classes also contribute to the regulation of gene expression. One example is the large family of non-coding RNAs, which are highly diverse both in structure and function and participate in processes ranging from post-transcriptional regulation to chromatin remodelling and serving as molecular scaffolds (Dykes & Emanueli, 2017). Another regulatory mechanism is DNA methylation, a chromatin-associated modification with key roles in gene silencing (Attwood et al., 2002).

While there are several layers of epigenetic control, now, I will focus deeper on three key pillars of intra-nuclear chromatin-associated epigenetic regulation.

2.4 Chromatin accessibility

The nucleosome, a nucleoprotein complex of DNA (~147bp) wrapped around a histone octamer, is the fundamental unit of chromatin. The histone octamer is composed of two copies each of four positively-charged histone proteins – H2A, H2B, H3, and H4. The wrapping of negatively charged DNA around the positively-charge histone proteins allows for increased compaction of the DNA. Given the central role of genomic DNA in cells, this has the added benefit of protecting the DNA from chromatin-binding proteins and nucleases. DNA predominantly exists in a tightly bundled state called heterochromatin (Kaplan et al., 2008; Kornberg & Thomas, 1974; Olins & Olins, 1974). However, gene expression, the gene and its corresponding regulatory elements need to be unwrapped (into euchromatin) and made accessible to allow for the binding of different transcription-associated proteins including TFs and RNA polymerase (Klemm et al., 2019; Shlyueva et al., 2014). The selectivity in opening the chromatin and its corresponding regulatory value is reflected in the observation that accessible chromatin comprises only 2% of the whole genome and yet encapsulates over 90% of TF-binding events (Thurman et al., 2012). Consequently, the accessible fraction of chromatin in a cell reflects the epigenetic blueprint underlying a particular cell state.

Referring to chromatin as “accessible” or “open” may incorrectly imply exposed segments of the double-stranded DNA. However, DNA is almost ubiquitously associated with nucleosomes. Rather, the extent of accessibility simply refers to the degree to which other chromatin-binding proteins can physically contact the

chromatin and reflect a more decondensed state of nucleosomal packaging. As such, chromatin accessibility is very dynamic, and histones compete with other proteins such as TFs for occupancy on the DNA (Krebs et al., 2017; Poirier et al., 2008). Nucleosomes, when evicted from the chromatin, are quickly replaced by other chromatin-binding proteins (Brahma & Henikoff, 2024). This binding precludes the nucleosome from binding at that region, and due to the larger footprint that the nucleosome occupies relative to transcription factors or other proteins such as polymerases, the nucleosome-depleted region can be treated as accessible.

Nucleosome depletion also leads to increased susceptibility to enzymatic degradation. This can be leveraged by controlled usage of the DNase I enzyme. This was first leveraged in the 1970s when researchers used the DNase I enzyme to identify vulnerable regions, called DNase Hypersensitive sites (DHS). Surveying these sites revealed interesting patterns. As expected, gene promoters were enriched, likely due to the increased occupancy with components of the transcriptional machinery (Keene et al., 1981; McGhee et al., 1981). Within DHS sites, some loci were bound by transcription factors, occluding DNase activity at the binding site, creating a "cold spot" or a "footprint" in the signal. Surveying these hypersensitive sites has enabled the identification of a range of candidate CREs (cCREs).

2.4.1 DNase-seq

As mentioned in the previous section, measuring DNase hypersensitivity was one of the earliest methods used to characterize accessible chromatin. DNase-seq couples the DNase I digestion assay to high-throughput sequencing for genome-wide analysis (Boyle et al., 2008; Hesselberth et al., 2009). The technique involves extracting permeabilized nuclei and treating with DNase I. The enzymatic treatment introduces double-stranded breaks at accessible regions of the chromatin, which are then ligated to adaptors, amplified, and sequenced. Despite its effectiveness, the protocol requires a great degree of optimization of input cell numbers and titration of enzyme concentration. Nonetheless, this method has been used extensively for profiling DHS regions in a wide range of cell types and for a long time was considered the gold standard (Boyle et al., 2008; Thurman et al., 2012).

2.4.2 MNase-seq

As the name suggests, MNase-seq uses the micrococcal nuclease (MNase) enzyme to digest accessible chromatin in permeabilized cells. However, unlike the endonuclease DNase, MNase has both endonuclease and exonuclease activity. It uses the endonuclease activity to create a double-stranded cut, but then in a "pacman-like" manner uses the exonuclease function to digest away chromatin until a nucleosome or TF-protected region is encountered. As such, the fragments sequenced using MNase-seq correspond to nucleosome-bound regions or TF-bound regions, which yields an inverted signal compared to DNase-seq (Schones et al., 2008).

2.4.3 ATAC-seq

The Assay for Tn5-Accessible Chromatin (ATAC) is a relative newcomer in the accessibility toolbox but has become one of the most widely used and popular techniques for profiling accessible chromatin (Buenrostro & Greenleaf, 2013). The method employs the hyperactive cut-and-paste Tn5 transposase enzyme, whose DNA transposition activity has been well characterized (Goryshin & Reznikoff, 1998; R. C. Johnson & Reznikoff, 1983; Reznikoff, 2003). When loaded with sequencing adaptors and incubated with permeabilized cells or nuclei, Tn5 inserts its payload into regions of accessible DNA (Buenrostro & Greenleaf, 2013; Picelli et al., 2014). The fragments of chromatin which are bookended by the inserted adaptors are then amplified using primers containing the P5 and P7 Illumina sequencing handles, which allows for quick generation of sequencing-ready libraries. In this thesis, ATAC-seq is the sole method of choice for profiling chromatin accessibility and is used extensively in **Papers I**, **III**, and **IV**, and warrants a deeper analysis of the transposition mechanism and the challenges associated with ATAC-seq.

2.4.4 Tn5-mediated transposition

Tn5 transposase has a high affinity for double-stranded DNA (dsDNA) and is catalytically active in a dimerized state (R. C. Johnson & Reznikoff, 1983). For sequencing applications, Tn5 is loaded with oligonucleotide cassettes that are application-specific (Adey, 2021), however, the presence of a minimal sequence corresponding to a conserved mosaic end (ME) is required for the Tn5 to recognize and bind to the DNA cassette (R. C. Johnson & Reznikoff, 1983). For ATAC-seq, the Tn5 is loaded with two distinct oligonucleotide cassettes: ME-

A/rev and ME-B/rev. Each contains the common and short ME-rev sequence (non-transfer strand) annealed to a longer ME-A or ME-B sequence (transfer strand), which extends as a 5′ overhang. Upon contact with accessible DNA, the loaded Tn5 makes a 9bp-staggered double-stranded cut, and covalently ligates the transfer strands to the native DNA, at each of the two cut-sites. The DNA fragment encapsulated between two distinct tagmentation events is amplified via PCR, after which it can be sequenced (**Figure 3**).

While the fragment between two tagmentation events forms the basic unit of an ATAC-seq library, the two cut-sites bookending the fragment are the true location of the accessible chromatin. Consequently, the fragment itself may span nucleosomes or other chromatin-interacting proteins. This results in the characteristic nucleosomal banding pattern in the fragment size distribution of the final library. Fragment sizes are roughly quantized by different multiples of nucleosomes. Furthermore, in the absence of nucleosomes, free DNA is more prone to extensive tagmentation, leading to an increased density of cut-sites, and therefore resulting in shorter sub-nucleosomal fragments (**Figure 4**).

Like DNase-seq, ATAC-seq signals are enriched around promoters, enhancers, and the TSS of active genes, due to the increased transcriptional activity and binding of various elements of the transcriptional machinery. As such, a common quality control measure for ATAC-seq libraries is to check the aggregate signal enrichment at all TSS relative to distal upstream and downstream loci. A high-quality ATAC-seq library will be reflected in robust TSS enrichment scores.

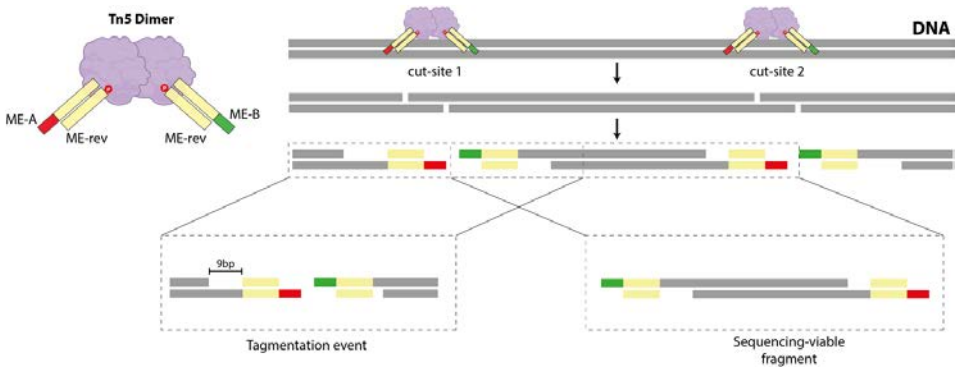

***Figure 3**: Visualizing Tn5-mediated transposition. Left - Tn5 dimer loaded with ME-A/rev and ME-B/rev oligonucleotide cassettes. Non-transfer strand bears a 5′ phosphate group. Right –Tn5 binds dsDNA and generates a 9bp-staggered double-stranded break. Adaptors are inserted by covalent ligation of the 3′ transfer strand to the free 5′ ends at the cut-site. Two separate tagmentation events generate an amplifiable fragment for sequencing. Created with BioRender.com.*

2.4.5 Challenges in ATAC-seq

For simplicity, the challenges in working with ATAC-seq data can be distributed into three broad buckets:

i) Biological sparsity

While transcriptomic assays benefit from a rich target pool of hundreds to thousands of mRNA transcripts per gene, chromatin assays are constrained to two copies of any given locus on the DNA (in a diploid organism). This biological constraint presents a reduced sample space for profiling and contributes to the inherent sparsity of ATAC-seq data, but also other chromatin-associated profiling methods.

ii) Technology sensitivity

In addition to the limited targets in the chromatin, the sensitivity of the capture mechanism can compound the issue of sparse data. As shown in **Figure 4**, two separate tagmentation events are required for the generation of an amplifiable fragment. However, to be sequenced, the fragment must be flanked with distinct ME-A and ME-B adaptors (A-B or B-A configuration). Only this configuration will allow the correct addition of the universal P5 and P7 Illumina handles, needed for sequencing.

Tn5 loading is a semi-stochastic process, involving incubations of the Tn5 with an equimolar mixture of ME-A/rev (A) and ME-B/rev (B) oligonucleotide cassettes, such that each unit in the dimer will bind to one cassette. This means that a loaded Tn5 dimer can exist in four possible adapter configurations: A/A, A/B, B/A, and B/B. After tagmentation, this leads to 16 possible fragment configurations, with the sequencing-compatible A-B and B-A flanked fragments expected with 50% frequency (**Figure 4**). The incompatibility of the remaining 50% corresponds to lost information, which further exacerbates the sparsity issue. Innovative strategies have been developed to circumvent this issue, such as in nanoCUT&Tag which employs linear amplification and dual tagmentation (explained in section **3.4.3**) (Bartosovic & Castelo-Branco, 2023), LIANTI which uses in vitro transcription to increase fragment yield (Chen et al., 2017), and s3-ATAC which uses complex molecular manipulation to generate 100% sequencing viability (Mulqueen et al., 2021).

iii) Feature selection

The third challenge arises during downstream data analysis, following sequencing. In transcriptomics, the sequenced reads can be mapped straightforwardly to a discrete feature set, corresponding to the genes in the organism, as mRNA is a direct transcript of the gene. Accessible chromatin, on the other hand, can encompass several classes of functional elements (Dunham et al., 2012; Thurman et al., 2012), which are scattered throughout the genome. One approach to tackle this is to aggregate the sequencing reads and identify regions of the genome with a significantly higher pileup of reads relative to neighbouring regions. These high signal-to-noise regions, known as peaks, are used as the feature set for building a count matrix (Y. Zhang et al., 2008). While this is powerful and a commonly used approach, one drawback is the possibility of masking a weaker but bona fide signal adjacent to a strong peak. Therefore, another common approach is to build a count matrix using fixed-width genomic bins. This ensures that the entire genome is scanned for reads, before filtering out empty bins or bins with low counts and is explored in section **3.5.1**.

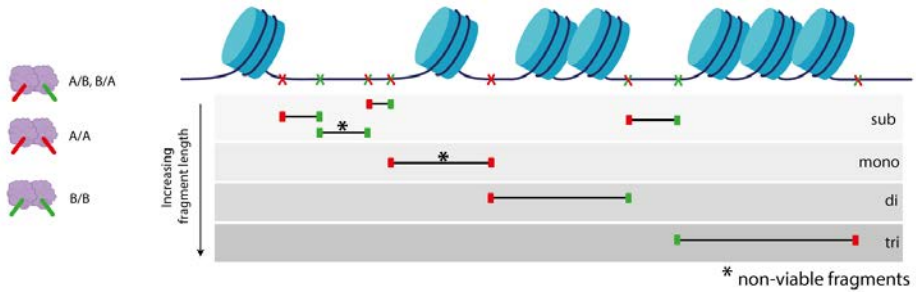

***Figure 4**: Viability and size distribution of fragments. Left – Four possible Tn5 configurations after adapter loading. Right – Tagmentation events on chromatin are shown as colour-coded crosses on the DNA and correspond to the Tn5 configuration at that cut-site. Fragments generated based on the inserted adaptor orientations. Shorter fragments are enriched in nucleosome-depleted regions while longer fragments are distributed according to nucleosomal periodicity. Fragments flanked by same end adaptors are non-viable for sequencing and are indicated with an asterisk. Nucleosomes are depicted as blue circles. Created with BioRender.com.*

2.5 Histone-tail modifications

Profiling chromatin accessibility is an important technique for identifying cCREs for all the reasons stated earlier. However, accessibility is only one layer of epigenetic regulation. To get a more granular understanding of the epigenome, we also need to look at another layer of the blueprint – histone tail modifications.

The histone proteins in a nucleosome have long N-terminal residues that emanate out from the core, known as the "histone tail". Residues on the histone tail can bear a wide range of post-translational modifications (PTMs), which serve as signalling beacons for proteins such as transcription factors, chromatin remodelling complexes and other chromatin-interacting proteins. Consequently, these modifications functionally annotate the local chromatin at a molecular level and form what is known as the histone code (Allis & Jenuwein, 2016; Kouzarides, 2007; Strahl & Allis, 2000).

The genetic code is virtually universal with a clear link between a DNA sequence, the transcribed mRNA sequence, and the ensuing amino acid sequence. The histone code on the other hand is far more complex and analogue, with histone PTMs investing the chromatin with functional identity, which can alter in the context of adjacent modifications. The precision of the code is exemplified by the fact that the addition or removal of a single methyl group (three hydrogens attached to a carbon, -CH3) to the same amino acid residue (H3K4me2 versus H3K4me3) or added to the same amino acid located elsewhere in the tail (H3K9me3 versus H3K4me3), can alter the functional state of the chromatin, and alter the downstream response. Hundreds of histone PTMs have been identified which can be broadly categorized into the following six groups based on the chemical groups involved – methylation, acetylation, ubiquitylation, sumoylation, citrullination and phosphorylation, although the number of identified histone PTMs continues to increase. These groups have often distinct but also overlapping roles affecting DNA replication, transcription, DNA repair, remodelling, and condensation. While all histones can be modified, most modifications found have been associated with the H3 histone protein (Bannister & Kouzarides, 2011; Kouzarides, 2007). Adding to the complexity, histone modifications are not static. While some may endure to represent a more stable state, many are dynamic and changing in response to ongoing intracellular signalling and cellular demands. Numerous studies have contributed to partially deciphering the functional role of different classes of histone PTMs. Methylation can be an active or repressive mark,

depending on the context: Trimethylation of lysine 4 on histone H3 (H3K4me3) is an active mark and is associated with actively transcribed promoters (Santos-Rosa et al., 2002). However, when the same histone H3 bears trimethylation of lysine 9 (H3K9me3) or lysine 27 (H3K27me3), it represents inactive heterochromatin. Heterochromatin (HC) can also be defined at two levels – facultative HC and constitutive HC. While both groups are transcriptionally silent, facultative HC exhibits a higher potential to transition to an active state and can mark primed or poised regions of chromatin, whereas constitutive HC represents more stably repressed regions of the genome. This distinction is also captured in the histone code, where H3K27me3 typically marks facultative HC and H3K9me3 marks constitutive HC (Craig, 2005; Talbert & Henikoff, 2006; Trojer & Reinberg, 2007).

Indeed, the H3K27 residue can also display differential methylation, and while tri-methylation represents facultative HC (H3K27me3), mono-methylation (H3K27me1) is associated with transcriptional activation along with H3K36me3, and di-methylation (H3K27me2) is associated with enhancer-specific repression (Ferrari et al., 2014).

Acetylation on the other hand is an active chromatin mark, and H3K27ac is found to mark active enhancers and promoters (Gräff & Tsai, 2013).

The deposition and removal of different histone PTMs is carried out by conserved proteins and protein complexes. Histone acetyltransferases (HATs) are a family of proteins that add the acetyl group, whereas histone deacetylases (HDACs) remove the acetyl group. As such, HDAC activity is associated with gene repression, and the inhibition of HDACs with valproic acid (VA) can lead to hyperacetylation which has been shown to activate enhancers and even trigger cell differentiation (Göttlicher et al., 2001).

This highlights the complex nature of this code and how these enzymes constantly scan and interact with chromatin to maintain an ultra-fine balance and regulate gene expression.

2.5.1 Importance of targeting the histone code

As mentioned at the start of this section, profiling chromatin accessibility is a powerful approach, and can serve as an important tool to identify cCREs. The development of several computational methods has aided this effort. Co-accessibility analysis, which relies on the increased likelihood of regulatory

elements and their target genes displaying a shared accessibility state can be used to identify regulatory interactions and build gene regulatory networks (GRNs) (Pliner et al., 2018). TF footprinting algorithms rely on the dip in accessibility seen within open chromatin peaks, corresponding to the TF binding event to identify putative regulatory TFs in different cell types (Boyle et al., 2008; Hesselberth et al., 2009). Similarly, the chromVar algorithm scans accessible peaks for TF motifs and identifies motifs with increased accessibility which can be used as a proxy for TF activity (Schep et al., 2017).

The reliance of chromatin accessibility on computational inference to study gene regulation makes it an effective first-pass approach, as it can pinpoint *which* regions of the genome are important for the regulatory blueprint. However, it may fall short in providing a definitive understanding of *why* they are important.

This is exemplified when we tie in what is known about the histone code. Tools such as Hidden Markov Models (HMM) can identify patterns in chromatin accessibility to segment the genome into discrete "states" corresponding to putative function (Ernst & Kellis, 2010; Thurman et al., 2012). H3K4me3 and H3K27ac are both active chromatin marks, found at active promoters or active promoters and enhancers respectively, and co-localize with accessible chromatin. Consequently, the ATAC signal could encapsulate regions bearing both marks. However, profiling with ATAC-seq would provide limited resolution, relative to profiling the individual marks themselves. Similarly, H3K27me3, as a marker of facultative HC, can be located at primed chromatin regions, which are typically accessible, and is commonly seen in embryonic development and in pluripotent stem cells (Bernstein et al., 2006). However, ATAC-seq would not be able to distinguish active from primed chromatin, which again highlights the difficulty of teasing out functional annotations from accessibility alone.

Hence, to gain a more comprehensive understanding of the regulatory epigenome and "sharpen" the image of the blueprint snapshot, it is imperative to have robust methods capable of profiling histone marks.

2.5.2 ChIP-seq

Chromatin Immuno-Precipitation with sequencing (ChIP-seq) has long been the gold standard for profiling chromatin-associated proteins and modifications (D. S. Johnson et al., 2007). It involves formaldehyde-based fixation to chemically freeze the state of the cell, followed by sonication to shear and fragment the DNA. The

fragmented DNA is targeted with an antibody against the protein or modification of interest and pulled down, producing an enriched pool of protein-bound DNA. After reverse cross-linking, these fragments can then be constructed into a library and sequenced.

While ChIP-seq is powerful and robust, it suffers from some drawbacks that make the method less than ideal. The high number of input cells required can make it difficult to profile rare cell populations or from limited material. The protocol also takes several days to generate sequencing-ready libraries, with many steps requiring careful optimisation. ChIP-seq data is also inherently noisy due to the immunoprecipitation, making it difficult to extract bona fide signals from the noisy background.

2.5.3 CUT&RUN

CUT&RUN (Cleavage under target and release using nuclease) is a relatively recent development which uses a protein-A micrococcal nuclease fusion protein (pA-MNase) to precisely cut out the region of the DNA bound by the protein of interest (Skene & Henikoff, 2017). The protocol is straightforward and involves sequentially incubating a primary antibody, secondary antibody and the pA-MNase with permeabilized nuclei. The assembled complex is activated using calcium which induces the MNase to cut around its tethered site. MNase has both exonuclease and endonuclease activity which allows it to first create a double-stranded cut and then digest away the chromatin up to the precise site of protein binding. These excised fragments are released, amplified into a library construct, and sequenced. Unlike ChIP-seq, CUT&RUN can be performed on fewer cells, making it attractive for profiling low input samples (Hainer et al., 2019; Patty & Hainer, 2021). The protocol is also simpler and can be completed within a day, and the high specificity of the cutting generates higher signal-to-noise in the data. However, both ChIP-seq and CUT&RUN require a separate library construction step for generating the final sequencing-ready libraries, which can be cumbersome.

2.5.4 CUT&Tag

A combination of CUT&RUN and ATAC-seq, CUT&Tag uses a protein-A Tn5 fusion protein (pA-Tn5) to perform antibody-tethered tagmentation (Kaya-Okur et al., 2019). The initial steps are similar to CUT&RUN, however, it relies on magnesium to activate the Tn5 transposase and insert sequencing adaptors around the bound

site. A major advantage of CUT&Tag over the previous two methods is the in-protocol library generation step. By loading the Tn5 with the appropriate Nextera-sequence bearing adaptors, full-length sequencing-ready libraries are generated immediately after PCR amplification. This significantly shortens the overall protocol, making it more robust and easier to adopt, and also makes it amenable to single-cell applications (see section **3.4**).

However, specific applications may make CUT&RUN more attractive than CUT&Tag. The higher salt concentrations in CUT&Tag buffers restrict the mobility of tethered Tn5, which allows for a more targeted, precise tagmentation around the binding site. Simultaneously, the higher concentration also destabilizes protein-DNA binding, increasing the likelihood of transcription factors being evicted from the chromatin. This issue can be circumvented using CUT&RUN which uses lower salt concentrations. Therefore, for profiling sparse TFs, CUT&RUN may be more attractive, whereas, for histone modifications which are more abundant in the genome, CUT&Tag is better.

2.6 The 3-D chromatin architecture

2.6.1 Higher-order structures

The complex genome of higher-order organisms features regulatory elements that can be situated several hundred kilobases away from their cognate target genes in linear 2D space (Schoenfelder & Fraser, 2019). However, the nucleus is not a 2D entity, and the chromatin contained within is also packed hierarchically in volumetric space. At the lowest level, the double-stranded DNA helix is wrapped around histones to form nucleosomes. Strings of nucleosomes form chromatin fibres which are then organised into more complex structures such as loops (Rowley & Corces, 2018).

According to the loop extrusion model, the CTCF protein and the Cohesin complex work together to extrude out and fold the chromatin into loop-like structures (Kim et al., 2019). CTCF contains a DNA-binding domain, which recognizes and binds to DNA at the well-characterized CTCF motif. The assembled Cohesin complex, envelops chromatin like a ring, and driven by ATP, progresses along the chromatin. This movement involves extracting the intervening chromatin on both sides, akin to pulling a thread through a ring. Loop extrusion continues until the Cohesin complex encounters CTCF proteins bound at convergent motif orientations (converging CTCF). The loop anchors are therefore dictated by CTCF binding, and

disruption of CTCF motif orientations can disrupt loop formation and structure (Hakimi et al., 2002; Kim et al., 2019; Sun et al., 2023).

Multiple sub-loops can be organized within a single larger loop which are more stable. These stable domains, called topologically associating domains (TADs) correspond to linear blocks of the genome that display increased contact frequency relative to regions outside the domain (Dixon et al., 2012). On genomic contact maps, TADs appear as sharp triangle-like structures. The base of the triangle corresponds to the block of the genome that is contained within the domain, while the apex of the triangle reflects the interaction between the outer anchors of the domain. TADs can span from few hundred kilobases up to megabase resolution (Dixon et al., 2012).

At even higher levels, the genome is organised into large compartments, which appear as a checkerboard-like pattern on genomic contact maps. These compartments, arbitrarily named A and B, appear as the first principal component after decomposing the genome-wide contact matrix, and reflect the highest order of chromatin organization (Lieberman-Aiden et al., 2009). Like TADs, regions within compartments exhibit increased contact frequency. These compartments are not static; different genomic regions can switch between compartments. Loci within the A Compartment are enriched for more activating marks and display increased correlation with active gene expression, while B Compartment loci tend to be more compact, exhibiting higher contact frequencies at the same genomic distance relative to the A compartment. Hence, the A and B compartments correspond to the active and inactive regions of the genome respectively (Lieberman-Aiden et al., 2009).

2.6.2 Role of TADs in regulatory interactions

The loop extrusion model of chromatin conformation provides an explanation for long-range enhancer-promoter interactions, and for the "promoter-skipping" phenomenon, wherein enhancers can bypass neighbouring, proximal promoters. Furthermore, promoters and their regulatory elements are often contained within the same TAD to facilitate increased contact frequency for gene expression. However, the degree to which TAD integrity contributes to maintaining precise enhancer-promoter contact is poorly understood. Studies investigating the regulation of the Sonic hedgehog (*Shh*) promoter via the 850kb distal ZRS element have made a compelling case for the role of TADs in facilitating enhancer-promoter interactions (Lettice et al., 2003; Symmons et al., 2016). Symmons et.al.

elegantly showed that when both the promoter and enhancer were embedded within the same TAD, gene expression was robust, and was independent of the distance between the two elements *within* the TAD. However, disruption of the TAD structure led to weaker expression, which could to an extent be rescued by rearranging the loci to be more proximal in linear space (Symmons et al., 2016). The function of TADs in creating neighbourhoods for increased regulatory contact however cannot explain the regulatory control of genes by enhancers located in different TADs, which has also been shown (Javierre et al., 2016).

2.6.3 Hi-C

Hi-C is a proximity ligation-based technology enabling unbiased profiling of genome-wide contacts (Lieberman-Aiden et al., 2009). Chromatin contacts are fixed in place by formaldehyde-mediated crosslinking. The cross-linked nuclei are digested with restriction endonucleases, and the free ends are ligated together, amplified, and sequenced. When aligning the reads to the genome, paired-end reads which appear to originate from distal regions of the genome, called chimeric reads, are considered as valid read pairs. These chimeric reads represent two distal regions of the genome that were in proximity at the time of fixation and may represent genuine chromatin interactions. However, one limitation of this method is its global contact capture, which includes both genuine and spurious non-significant contacts. Therefore, a substantial number of input cells are required to decipher the true population-level signal. Nonetheless, Hi-C remains a popular method and is responsible for the discovery of genomic compartments and TADs and advanced our understanding of chromatin architecture (Dixon et al., 2012).

2.6.4 Micro-C

Using restriction endonucleases to fragment the chromatin, as in Hi-C, presents a resolution limitation, as the fragmentation only occurs at loci containing the recognition sequence for the enzyme. At a 100% cutting efficiency, a 6-cutter enzyme such as *HindIII* would theoretically be expected to cut every 4096bp (4^6), providing at best a 4kb resolution. However, the recognition sequences are not uniformly distributed, nor is digestion perfectly efficient, which leads to a lower resolution (>4kb). Even with the use of a combination of restriction enzymes, the resolution is limited to the kilobase scale.

Micro-C is a technique similar to Hi-C, but replaces the restriction enzymes with an MNase enzyme to digest the chromatin (Hsieh et al., 2015). As seen earlier, the

dual endo- and exonuclease enzyme can digest away DNA until it encounters a nucleosome. The presence of nucleosome-sized fragments theoretically enables nucleosome-level resolution. Although obtaining this high-resolution is still a function of the sequencing depth, it has been used to study folding principles in greater detail, revealing new insights into cohesin-CTCF dynamics (Krietenstein et al., 2020).

2.6.5 HiChIP & PLAC-seq

Analysing genome-wide contacts can provide valuable insights, however, for certain applications, it may be desirable to specifically capture genome-wide contacts mediated by a protein of interest. The proximity-ligation methods HiChIP and PLAC-seq enable this specific profiling (Fang et al., 2016; Mumbach et al., 2016). These methods combine Hi-C and immunoprecipitation to pull down ligated genomic fragments bearing the protein of interest. This approach is particularly useful for identifying enhancer-promoter contacts. Active enhancers are bound by H3K27ac, and targeting this modification can help identify target promoters for different enhancers. Another advantage of these methods is the reduced sequencing depth required, as only a subset of the Hi-C contacts are captured.

3 Single-cell epigenomics

3.1 Single-cell sequencing

Over the last two decades, the rise of Next-Gen Sequencing (NGS) methods has facilitated the profiling of diverse cell types and tissues. These methods have found applications in transcriptomics, epigenomics and whole-genome sequencing. Bulk sequencing whole tissue samples or cells provides a population-level snapshot of the state of the sample. This is fine for homogeneous samples, but for complex tissue samples, the signal can be dominated by the most abundant cell types, while smaller, rarer populations get masked. One workaround is to purify cell populations using sorting techniques such as fluorescent-activated cell/nuclei sorting (FACS/FANS). This approach however relies on *a priori* knowledge of surface or intra-cellular proteins that can be targeted during sorting. Single-cell sequencing, as the name suggests, captures molecular information at a cellular resolution. The granularity achieved through this circumvents the masking problem associated with bulk sequencing studies and has been key to identifying new cell types in complex tissues as well as subtle yet distinct differences within cell types (Linnarsson & Teichmann, 2016).

The first single-cell study using NGS was performed in 2009 and profiled the transcriptome of a mouse blastomere at the four-cell stage (Tang et al., 2009). This small but incredibly important pilot study in four cells set in motion an avalanche of innovations that have led to an explosion of single-cell studies in many organisms and complex tissue types encompassing several millions of cells in a single study.

3.1.1 Achieving cellular resolution

NGS methods involve capturing information in a biological medium and transferring it to an *ex-situ* dsDNA strand which can then be read on sequencing instruments. This workflow presents several engineering challenges, starting with extracting the biological information using molecular tools. However, the key component is annotating this information using an identifier, or a barcode. These barcodes are then used to link the sequencing data with the original sample information. Barcodes are typically introduced during strand synthesis steps in a PCR or reverse transcription (RT) workflow and consist of a short 8-16bp

nucleotide sequence. At the bulk level, barcodes can be introduced using a sample-specific barcode in the workflow. However, for single-cell applications, this is a non-trivial task as each cell requires its own unique barcode. This requires cells to first be partitioned into separate reaction chambers where their barcodes can be introduced, after which they can be re-grouped for the remainder of the workflow.

3.1.2 Partitioning methods

Partitioning methods can broadly be grouped into plate-based partitioning and droplet-based partitioning, both of which involve creating some form of barrier between individual cells (Baran-Gale et al., 2018). Plate-based partitioning methods create a physical barrier by sorting single cells into individual wells of a multi-well plate. Droplet-based methods employ a liquid barrier typically by encapsulating cells into nanolitre-sized microdroplets. Plate-based methods are limited by the number of wells, and while some like the Takara ICELL8 can have up to 5184 nano wells in a single plate, these typically have lower throughput compared to droplet-based methods which can partition hundreds of thousands of cells using microfluidic devices. Microfluidics instruments are often quite expensive, however, though the cost may be offset by the lower per-cell cost achieved from the increased throughput. Nonetheless, innovations such as split-pool barcoding strategies can circumvent this limitation in plate-based methods through multiple rounds of partial barcoding (Cao et al., 2017; Macosko et al., 2015). Plate-based methods typically also have improved sensitivity with higher complexity in the final sequencing libraries (Cusanovich et al., 2015). Hybrid technologies that combine plate-based and droplet-based methods are also becoming popular and can enable ultra-high throughput profiling (Datlinger et al., 2021).

3.2 Challenges in single-cell data

The improved granularity of single-cell data also comes with considerable challenges. The smaller sample size, represented by individual cells rather than thousands as in a bulk experiment, results in a more constrained sampling pool for information capture. Naturally, the single-cell signal is sparser compared to bulk-level samples. This persists in all single-cell methods but is exacerbated in epigenomic methods which target the DNA, where only two alleles per cell are putative targets in a diploid organism.

Further, the same granularity that enables the unmasking of signals from rare cell populations also has the downside that weaker signals from individual cells cannot be rescued or strengthened by the aggregate signal of the population. Therefore, single-cell data is more sensitive to dropouts or inadequate information capture. Furthermore, the absence of a signal is less informative than the presence of a signal since the absence could be due to the true absence or due to signal dropouts. This is less relevant in bulk studies, where dropouts from a few cells may not affect the overall signal in the sample.

3.3 The ideal experiment

Several of the drawbacks mentioned earlier could be addressed in experiment and technology design. To explore this further, we can consider a hypothetical framework for an ideal experiment, which involves complete cell capture, zero barcode collisions, and infinite sensitivity.

3.3.1 Cell capture

A perfect experiment would ensure that every cell in the starting sample – cell culture or tissue – is captured and profiled in the protocol. However, most single-cell experiments only profile a subset of the cells in a population. Usually, this is due to technical limitations of the method itself, with even the highest of throughputs being limited to a couple of thousand cells per channel. Although a single sample could be split across multiple parallel runs, it remains statistically unlikely that every cell would be captured. Another source of cell loss is during tissue dissociation, which despite extensive optimization cannot ensure a perfect dissociation where all cells are retained intact within the single-cell suspension. This is important as it reflects that these experiments only provide a sampling of the true population. Furthermore, the choice of sample dissociation protocol has also been shown to elicit dissociation-specific changes in cells, which is unrelated to the underlying biology, but will show up in the data, skewing the interpretation (Marsh et al., 2022).

3.3.2 Barcode collisions

Barcode collisions, also called doublets or multiplets, arise when a single barcode is shared between multiple cells, leading to an inability to distinguish the respective cells from each other in the data. This one-to-many mapping of barcodes to cells is not an uncommon feature of single-cell experiments, arising

again due to technical limitations. Accidental sorting of two cells into the same well (plate-based) or multiple cells being encapsulated into the same droplet (droplet-based) will lead to those cells being indexed with the same barcode. Commercially available droplet-based methods tackle this by creating barcode and cell droplet emulsions at limiting dilutions, such that majority of droplets are empty and do not contain cells. This increases the likelihood that the droplets that do contain cells will be limited to a single cell. However, this also leads to incredibly low efficiency where only 1% of the operational capacity is used to capture cells, which also ties into the earlier point about sub-optimal cell capture (Datlinger et al., 2021). This trade-off however aims to minimize the likelihood of barcode collisions. There are also several computational methods for identifying putative collisions in the sequenced data, ranging from simple approaches such as selecting cells with abnormally high number of reads (outliers) to more sophisticated approaches such as computationally mixing cells and generating synthetic collisions that can then be used to build a model for identifying cells that mimic the simulated doublets (Wolock et al., 2019). The ideal experiment would ensure a perfect one-to-one mapping of barcodes to cells.

3.3.3 Sensitivity

The ability to discern differences in cell populations and cell states in a single-cell library is dependent on the complexity or richness of the library at the single-cell level. This sensitivity refers to the ability of a technology to detect and capture molecular information even at minimal concentrations. Technology sensitivity is largely a function of the chemistry being used to capture the molecular information (Svensson et al., 2018). Generally, droplet-based methods have lower sensitivity than plate-based methods (Wang et al., 2021). However, this trade-off is balanced by the higher throughput offered in droplet-based methods. Single-cell transcriptomics methods capturing mRNA molecules have a large target pool that can be captured, however, genes with ultra-low expression may have transcripts at concentrations too low to be detected, which can bias the interpretation of the data. Therefore, the ideal experiment would have perfect 100% sensitivity capturing all molecular information and providing an unbiased view of the state of the cell.

3.4 Targeting the single cell epigenome

Shortly after the rise and explosion of single-cell transcriptomics, the focus shifted to single-cell epigenomics. Broadly, the most popular methods have focused on chromatin and chromatin-protein interactions, DNA methylation and 3D genome architecture. Here, only the first two will be covered.

3.4.1 Single-cell ATAC-seq

Single-cell ATAC-seq was developed shortly after its bulk counterpart and profiled 254 cells using an integrated fluidics circuit (IFC) for the compartmentalisation (Buenrostro et al., 2015). A plate-based combinatorial barcoding approach soon followed, capturing single-cell chromatin profiles in the nematode *C. elegans*, and adult mouse tissue (Cao et al., 2017; Cusanovich et al., 2015, 2018). The release of the commercial scATAC-seq droplet-based platform by 10x Genomics in 2019 enabled higher cell throughputs per experiment. Now, single-cell chromatin accessibility profiling has been widely implemented and applied in various settings, to characterize chromatin profiles in complex tissues as well as whole organisms (Domcke et al., 2020; Mannens et al., 2023; Markenscoff-Papadimitriou et al., 2020; Satpathy et al., 2019; Zhang et al., 2021; Zu et al., 2023).

3.4.2 Single-cell CUT&Tag

As CUT&Tag is basically antibody-tethered ATAC-seq, it lends itself well to single-cell applications. This was highlighted in the original CUT&Tag study, where the investigators used the Takara iCELL8 nano-dispensing system to sort tagmented nuclei into a multi-well plate, followed by introducing well-specific barcodes through PCR (Kaya-Okur et al., 2019). This was later expanded on in multi-CUT&Tag, which could target multiple histone PTMs, but also on a plate-based platform (Gopalan et al., 2021). In an effort to increase throughput, we then developed scCUT&Tag, by optimising the bulk CUT&Tag protocol to be compatible with the commercial 10x Genomics scATAC-seq platform (Bartosovic et al., 2021). This study is captured in **Paper II** and is explained in detail in **Chapter 5**. In parallel, another study by Wu et.al. used a similar approach to profile H3K27me3 in individual cells of human brain tumours (Wu et al., 2021).

3.4.3 nanoCUT&Tag

nanoCUT&Tag was developed as a successor to scCUT&Tag (Bartosovic & Castelo-Branco, 2022), and featured three prominent changes that shortened the protocol, improved read metrics, and enabled multiplexed targeting. These changes are discussed in detail here, as these principles were leveraged to build the method detailed in **Paper IV**.

i) Nanobody-Tn5 fusion protein

Secondary nanobodies are single-domain antibodies that retain the antigen binding site but with a smaller footprint than a traditional antibody (Pleiner et al., 2017). Bartosovic et. al. fused secondary nanobodies with the Tn5 protein to produce nanoTn5 fusion proteins which replaced the need for using separate secondary antibodies and a pA-Tn5 protein, as in scCUT&Tag. Further, the nanoTn5 and primary antibody could be added simultaneously to permeabilized nuclei to allow binding. In addition to shortening the length of the protocol, another benefit was the reduced cell number requirement at the start of the protocol.

ii) Deterministic adapter loading

As highlighted in section **2.4.5**, the semi-stochastic nature of Tn5 loading leads to a loss of 50% of the fragments. This is due to the requirement for the appropriate flanking sequence combination (ME-A and ME-B). Here, Bartosovic et.al. loaded the nanoTn5 with a ME-A construct only, ensuring that all tagmentation events resulted in the incorporation of a ME-A fragment. Following droplet encapsulation, the fragments were then amplified, incorporating the cell barcode on the P5 end. However, this amplification is linear and not exponential, due to the missing ME-B cassette which would have served as the template for the reverse primer. This was then followed by a second tagmentation step using a wild-type Tn5 protein, loaded with the ME-B cassette only. This tagmentation event introduced the ME-B sequence randomly into the linearly amplified DNA fragments, creating viable fragments that were now flanked with both ME-A and ME-B.

The deterministic loading followed by a two-step tagmentation improved the sensitivity of the protocol and captured more fragments per cell relative to the original scCUT&Tag protocol (Bartosovic & Castelo-Branco, 2022).

iii) Antibody barcoding

Finally, by using mouse-specific and rabbit-specific nanoTn5 proteins, it became possible to combine this with rabbit and mouse primary antibodies, thereby enabling the targeting of two epitopes simultaneously. The addition of an 8bp antibody barcode in the adapters loaded on the species-specific nanoTn5 proteins allowed the demultiplexing of the modalities in downstream data analysis. Further, by adding a similar antibody barcode to a wildtype Tn5, it was also possible to target accessible chromatin (ATAC) before antibody-nanoTn5 binding, which enabled the simultaneous capture of 3 modalities in the same cell.

3.5 Data analysis approaches

The single-cell epigenomics field is relatively young, but the rapid pace of technology development has necessitated the development of computational methods for analysing the complex data that is generated. Here, rather than getting into the intricate details of every step of the analysis pipeline, I will describe some of the key steps and considerations involved in processing high-dimensional single-cell data.

3.5.1 Building the count matrix

After the first step of aligning the sequencing data to a reference genome and calling cells using the unique barcode, the next step is to build a count matrix for analysis.

The count matrix will typically have individual cells as rows and features as columns, or vice versa. As alluded to in section **2.4.5**, feature selection for genomic information is an important parameter that can influence the resulting analysis. While both peaks and genomic bins can be used, single-cell experiments involving complex tissue may benefit more from using the latter, simply due to the diversity of putative cell types in the sample. The robustness of the peaks called over the whole dataset will likely be skewed depending on the frequency distribution of cell types in the data.

Using genomic bins allows for unbiased sampling across the entire genome, after which empty bins and low-coverage bins can be removed. If using bins, another point to consider is the bin size. While using smaller width bins increases the granularity of the features, it naturally leads to a lower read count per bin. In addition, it is also important to consider the computational requirements when

working with large matrices resulting from narrow bins. For instance, covering the genome using a bin width of 2000bp, would generate approximately 1.5 million bins – an incredibly high number, and potentially resource intensive particularly if the dataset is of high quality with deep sequencing depth and high complexity. The bin width can also be chosen based on the type of dataset. snATAC-seq data may benefit from smaller bins due to the nature of regulatory elements being captured in accessible chromatin – promoters, enhancers, CTCF-bound sites etc. – which all have relatively small footprints. However, in scCUT&Tag and nanoCUT&Tag, the width can be selected depending on the likely distribution pattern of the histone PTM being profiled. H3K36me3, an active mark associated with active transcription, is typically deposited across the gene body. The repressive H3K9me3 and H3K27me3 marks which typically mark constitutive and facultative heterochromatin respectively, also span multiple nucleosomes and have broader signal. These three modifications would likely benefit from broader bins ranging from 10kb to 50kb, whereas narrower bins would be better for other marks such as H3K27ac and H3K4me3, marking active enhancers and promoters respectively.

3.5.2 Normalization and dimensionality reduction

Data normalization is an important step to account for differences and variability in read depth, as well as sparsity. In addition, the dimensionality of these datasets also poses a challenge. Each feature in the count matrix informs about one molecular attribute and contributes to one dimension of data space for the analysis. While adding more features can seem more informative (for instance – having information about 10 attributes is more informative than just 2 attributes), the addition of each feature exponentially increases the data space, creating significant computational challenges during analysis. One approach to simplify this is to use the manifold assumption, which states that single-cell states span a low-dimensional manifold, rather than high-dimensional ambient space (Moon et al., 2018). Dimensionality reduction approaches, therefore, aim to transform the high-dimensional feature matrix into a lower-dimension matrix, while preserving the true variation present in the dataset, using methods such as principal component analysis (PCA) and singular vector decomposition (SVD). This reduction approach reduces the data space in which the data lies and eases the computational complexity, while still capturing most of the underlying biological differences.

In single-cell epigenomics, a common approach used for normalization is latent semantic indexing (LSI), which first performs term frequency-inverse document frequency (TF-IDF) for normalization followed by SVD for dimensionality reduction (Aizawa, 2003; Cusanovich et al., 2018). This approach is rooted in the field of natural language processing (NLP), where it is used to extract information and relationships between words (terms) in a corpus of text documents. This approach is especially useful when dealing with sparse data, for which single-cell epigenomics fits the bill. In essence, TF-IDF normalization alters the count matrix by assigning weights to each item in the matrix depending on the frequency of a feature in a cell (term frequency) and its global frequency across cells (document frequency). This aims to dampen the feature signal shared across multiple cells (less informative), and amplify the rare, less common signal (more informative). Following this with SVD then produces a new lower-dimensional matrix representing the "latent space" (Cusanovich et al., 2018).

3.5.3 Batch correction and clustering

Dimensionality reduction methods aim to capture the major sources of variation, but it is important to distinguish undesirable technical variation from the desired biological variation. Technical variation can arise due to many factors, including random fluctuations and unintentional biases introduced during the experimental workflow due to operator bias or batch-to-batch variation with reagents and sequencing. Collectively, these are called batch effects and can skew data analysis and interpretation. While measures can be taken to minimise batch effects in the upstream experimental workflow through proper planning, sample randomization and minimizing experimental and sequencing batches, complete elimination is not always feasible. This limitation is evident during the integration of external datasets in downstream analysis, where experimental factors are beyond our control, and can introduce substantial variation (Stuart et al., 2019). Many batch correction algorithms have been developed to assist in the removal of this unwanted technical variation such as Harmony, Scanorama, ComBat, and CCA, among others (Butler et al., 2018; Hie et al., 2019; W. E. Johnson et al., 2007; Korsunsky et al., 2019; Stuart et al., 2019). The choice of the algorithm also depends on the application as different methods prioritize different aspects of the correction, with some performing an aggressive technical correction, at the cost of losing some biological variation, while others are optimized at preserving the biological variation with a softer technical correction (Luecken et al., 2022).

After batch correction, a cell-to-cell connectivity graph based on a k-nearest neighbours (kNN) approach is constructed over which clustering is performed. Clustering, or community detection, aims to find cell communities with similar feature profiles in the latent space. At the highest level, clusters reflect broad cell types captured in the dataset, but at higher resolutions can reveal finer differences and heterogeneity. A useful approach to identify fine clusters, is to first perform broad level clustering, and then either repeat the clustering on each broad cluster, or re-run the LSI using variable features found in the broad clusters, to tease out finer distinctions (Granja et al., 2021).

3.6 A note on multimodal methods

Single-cell technologies have revolutionised our understanding of molecular biology and revealed incredible insights into tissue and cell biology. Just as scRNA-seq experiments have revealed transcriptomic heterogeneity, epigenomic experiments have revealed molecular mechanisms and regulatory pathways potentially underlying these diverse states. However, profiling individual modalities has a major shortcoming: cellular function is remarkably complex and dynamic, and processes occur at multiple levels. Each single-cell application targeting a specific modality provides only a partial perspective, akin to illuminating an object in a dark room with a flashlight from a fixed angle. The resulting shadow on the opposite wall represents the data generated and offers only one perspective of the biological truth. Several computational methods have been developed which can integrate datasets from different modalities, and while useful, the limitation remains that the datasets were generated from different experiments. To gain a more comprehensive understanding, profiling additional modalities in the same cell is essential, by shining multiple lights within the same room on the same object. multimodal technologies aim to do just this by capturing multiple modalities within the same cell, and not different cells, providing a more accurate snapshot of the underlying biology.

Multimodal profiling is far more challenging than single-modality profiling, as multiple molecular moieties, with different biochemical properties, need to be preserved and tagged with the cell barcode. Following this, the modalities also need to be split and processed separately. Despite this, the past few years have seen a surge in the development of these methods capable of targeting and capturing multiple sets of features within a cell. **Paper IV** of this thesis also

attempts to do the same by simultaneously profiling 4 orthogonal modalities within the same cell, to increase our understanding of cellular function.

4 Research Aims

This thesis has two overarching objectives:

1. To develop new single-cell technologies for profiling the epigenome and transcriptome (**Papers II** and **IV**)
2. To apply single-cell methods to characterize oligodendroglial chromatin states (**Papers I** and **III**)

Specifically, these are addressed in each paper's research aims:

Paper I

- Profile accessible chromatin of oligodendroglia in a mouse model of MS
- Identify regulatory mechanisms controlling the immune-like transcriptional state in these cells
- Characterize the interplay between specific transcription factors and histone modifications in activating immune gene expression

Paper II

- Develop a method for profiling histone post-translational modifications and transcription factors at single-cell resolution
- Delineate regulatory principles such as promoter bivalency, H3K4me3 spreading and promoter-enhancer interactions in different neural cells

Paper III

- Characterize the accessible chromatin and histone PTM landscape in adult human neural cells
- Identify regional differences in adult human oligodendroglia
- Deconvolute regulatory mechanisms for HOX gene repression in spinal-cord derived oligodendroglia

Paper IV

- Develop a new multi-modal single-cell technology capable of simultaneously capturing multiple layers of the epigenome and the transcriptome

5 Results

5.1 Paper I - Epigenomic priming of immune genes implicates oligodendroglia in multiple sclerosis susceptibility

In the inflammatory environments of MS and EAE, OPCs are recruited to the site of lesions to differentiate and remyelinate denuded axons (Franklin & Ffrench-Constant, 2008). We and others have previously shown that oligodendroglial (OLG) lineage cells actively participate and modulate the inflammatory microenvironment by upregulating several immune-associated genes including proteins of the MHC-I and MHC-II antigen presentation complex (Falcão et al., 2018; Kirby et al., 2019). These transcriptional changes and functional activity opened a new avenue into understanding the role of these cells in disease.

The aim of this study was to characterize the regulatory landscape of OLG lineage cells in EAE, to better understand how these immune programs are transcriptionally activated. We first performed snATAC-seq in OLG-enriched cell populations from spinal cords of CFA-Control and EAE mice, using droplet-based (10x Genomics) and plate-based (pi-ATAC) platforms.

In the EAE-derived OLGs, we found several peaks proximal to genes associated with immune pathways, suggesting the chromatin changes were in line with the previously published transcriptomic changes. Interestingly, we found that only 11.48% of differentially accessible peaks in EAE-OLGs were associated with promoters while 33.65% and 49.02% were intergenic and intronic elements, which suggested that changes in enhancer utilization may drive the transcriptomic changes that had been observed. In line with this, we found that several genes in the MHC-I and MHC-II pathways such as *Tap1*, *Psmb8*, *Psmb9*, *Tap2, Nlrc5, H2-Aa,* and *H2-Ab1* displayed promoter accessibility in both CFA-Control and EAE-derived OLGs, though accessibility was higher in disease, indicating promoter priming. We then performed multiOme ATAC+RNA co-profiling in CFA-Control and EAE and confirmed that accessibility of the MHC-I and MHC-II genes were not associated with expression, unlike in EAE, where promoter accessibility correlated with their expression.

We had previously shown that IFN-g activates immune gene expression in OPCs in vitro. Further, since several IFN-g response genes were activated in EAE, we performed ATAC-seq and RNA-seq in IFN-g treated primary mouse OPCs to check if chromatin accessibility changes underlie the activation of these genes in vitro. Overall, we found 867 genes with altered expression but only 47 chromatin sites with differential accessibility in IFN-g treated OPCs. Like we saw in EAE, most chromatin alterations were found at intergenic and intronic regions. We also observed that a large subset of upregulated immune response genes did not display any alteration in chromatin accessibility. Of these genes, some displayed increased accessibility at regions close to the promoter suggesting enhancer activation. Collectively, our in vitro data indicated that OPCs display a primed chromatin state at several immune gene promoters.

Enhancer sequences are bound by TFs which is associated with their activation. Given that most chromatin accessibility changes in EAE were observed at distal enhancer elements, we used chromVar to search for changes in TF motif accessibility in EAE-OLGs. Among others, we found several TFs with immunoregulatory functions such as KLF4, KLF13, FOS, JUNB, BACH1, BACH2, IRF1, IRF2, STAT1, STAT2 and STAT3 with increased motif accessibility in EAE-OLGs. Given the roles of BACH1 and STAT1 in regulating inflammatory macrophage differentiation and IFN-g signal transduction, we performed siRNA-mediated knockdown of *Stat1* and *Bach1* separately in IFN-g treated primary OPCs, followed by bulk RNA-seq. *Bach1* knockdown led to increased expression of the MHC-I genes *H2-Q4* and *H2-Q7* and the MHC-II gene *Cd74*, which suggested that BACH1 may be a negative regulator of these immune genes. In contrast, S*tat1* knockdown decreased the expression of several immune-related genes including *H2-Ab1*, *Cd74*, *Il12rb1*, *Cd274*, *Irf1*, *Irf8*, *Nlrc5*, *Tap1*, and *Psmb8*, suggesting STAT1 may positively regulate these genes upon IFN-g treatment in OPCs. This was further corroborated by performing CUT&Tag targeting STAT1 in IFN-g treated OPCs, which showed increased STAT1 binding at these upregulated genes.

While enhancer activation plays a role in target gene expression, it is not the only layer regulating this. To get a more holistic view of the regulatory landscape, we performed CUT&RUN targeting the histone modifications H3K27ac (active enhancers), H3K27me3 (inactive/primed promoters), H3K4me3 (active promoters) as well as the insulator protein, CTCF. We also performed HiChIP against H3K27ac to evaluate promoter-enhancer interactions. Using the Activity-by-contact (ABC) model, we integrated the ATAC, H3K27ac and HiChIP data to

identify enhancers and their target promoters. We observed increased H3K27ac deposition at enhancers around MHC-I and MHC-II genes upon IFN-g treatment. Interestingly, many of the immune genes which were activated upon treatment also presented low H3K4me3 and high H3K27me3 deposition in Control, suggesting these were bivalent promoters. Upon IFN-g treatment, H3K27me3 was removed and H3K4me3 was increased.

The Enhancer of Zeste homolog 2 (EZH2) protein is the component of the Polycomb Repressive Complex 2 (PRC2) responsible for the deposition of H3K27me3. Hence, we then tested if depletion of H3K27me3 through pharmacological inhibition of EZH2 could increase the expression of immune genes. Following 96h treatment with an EZH2 inhibitor, we treated the cells with IFN-g and found increased expression of several immune genes, indicating that EZH2-mediated H3K27me3 deposition maintains the genes in a primed state.

We also observed similar priming of immune genes (*TAP1*, *TAP2*, *PSMB8*, *PSMB9*) in non-immune neural cells in the human brain. Finally, we found that a subset of MS-associated SNPs overlapped with regulatory regions in OLGs, suggesting that MS susceptibility may be linked to alterations in the regulatory activity of these distal enhancers.

5.2 Paper II - Single-cell CUT&Tag profiles histone modifications and transcription factors in complex tissues

Profiling chromatin modifications in individual cells can deconvolute regulatory principles underlying different cell states, however most methods that were available were limited by sensitivity and throughput. In this study, we developed scCUT&Tag by combining the 10x Genomics scATAC-seq platform with CUT&Tag to profile histone modifications and transcription factors in thousands of single cells with improved sensitivity over existing methods. We showcased the potential of scCUT&Tag by applying it to the juvenile mouse brain targeting a range of histone modifications, the transcription factor OLIG2 and the cohesin component RAD21.

CUT&Tag is similar to ATAC-seq, with the main distinction being that the tagmentation step is antibody-tethered, in the former. The original bulk CUT&Tag protocol used concanavalin-A coated magnetic beads to bind cells and nuclei, and a magnet to hold the bead-cell complex during wash steps. As these magnetic beads would be incompatible with the 10x Genomics microfluidics

platform, we opted for centrifugation-based wash steps in place of the magnet-based separation. These steps were optimized by adding BSA to the buffers to minimize nuclei clumping induced through repeated centrifugation steps. Following tagmentation, we integrated the CUT&Tag protocol with the 10x Genomics scATAC-seq (v1 and v1.1) kits for single-cell encapsulation up till library completion. We first tested the protocol by targeting H3K27me3 in a mixture of three mouse cell lines and were able to deconvolute all cell lines in the sequenced data.

We then applied scCUT&Tag to the juvenile mouse brain targeting the histone modifications H3K27me3, H3K27ac, H3K4me3 and H3K36me3. We used peaks proximal to marker gene populations to initially annotate the broad neural cell types. In the active marks of H3K4me3, H3K27ac and H3K36me3, we found a strong marker gene signal for each cell type and an inverse signal for the repressive mark H3K27me3. Comparing the data quality with other published technologies revealed that scCUT&Tag performed at par or better for signal specificity and captured more cells per experiment.

Single-cell transcriptomic atlases serve as excellent references for annotating cell types. We extracted the top 100 differentially expressed genes for different neural populations in a published mouse atlas and aggregated the H3K4me3 signal around the promoter of these gene sets to generate a per-cell metagene score. This RNA semi-agnostic method also worked well to identify different cell types, which we validated by integrating the H3K4me3 dataset with the transcriptomic atlas using the standard approach - canonical correlation analysis (CCA).

H3K4me3 marks active promoters, however the breadth of the H3K4me3 mark has been linked to transcriptional activity of the gene. In line with this, we observed that the H3K4me3 breadth at marker gene promoters was greater compared to all genes. To see how the mark changed upon OPC differentiation, we first ranked the cells of the OLG lineage in a pseudotime trajectory and observed the H3K4me3 signal breadth from OPCs to MOLs at the top 100 MOL promoters. We observed a steady increase in the width of the signal along the pseudotime trajectory, which was in line with the expected increase in expression of these genes. These results indicate that the scCUT&Tag data faithfully recapitulate histone PTM dynamics at the cellular level.

We also tested scCUT&Tag against the transcription factor OLIG2 and the cohesin complex subunit RAD21. While the resolution of the data was lower compared to

the histone modifications, we could still identify two clusters in the OLIG2 dataset corresponding to an OLG population and a non-OLG population. Performing a de novo motif search on the peak sets in both datasets identified motif sequences that were similar to the CTCF-binding motif (for RAD21) and other published OLIG2 motifs (for OLIG2).

Finally, we tested the quality of H3K27ac peaks called in the OLG populations by feeding them into the ABC model to predict putative promoter-enhancer interactions. As an orthogonal measurement, we performed HiChIP against H3K27ac in cultured primary mouse OPCs and validated the ABC-inferred loops using a signal pileup analysis on the HiChIP dataset. This revealed strong signal enrichment at the inferred loops, highlighting the ability to use scCUT&Tag data for predicting cell type-specific regulatory interactions.

5.3 Paper III - Single-nuclei histone modification profiling of the adult human central nervous system unveils epigenetic memory of developmental programs

The cells of the human central nervous system display marked transcriptional heterogeneity. However, the epigenetic landscape underlying this diversity is not well understood. In this study, we sought to chart the chromatin accessibility and histone modification landscape of different neural populations by performing snATAC-seq and nanoCUT&Tag on archival post-mortem tissue samples from different regions of the adult human CNS.

Using snATAC-seq, we could identify the broad neural populations. Interestingly, while the neurons clearly separated out based on their electrophysiological profile and tissue of origin, the glial cell populations displayed a far more homogenous structure. We then performed nanoCUT&Tag, co-profiling H3K27ac and H3K27me3 in the same cell, with metrics comparable to that of our previously published juvenile mouse dataset. As H3K27ac marks active enhancers and overlaps with regions of accessible chromatin, we could integrate the snATAC-seq data with the H3K27ac data to annotate the cells and use the shared barcode between the histone modalities to annotate the H3K27me3 data. Our nanoCUT&Tag dataset is the first joint multi-modal single-cell histone PTM dataset of the adult human CNS.

Using co-accessibility analysis, we found an OLG-specific link between the *SOX10* promoter and a distal upstream enhancer that had not been previously associated with the *SOX10* gene. We found a conserved binding motif for the transcription factor, TFAP2A at the core of the enhancer peak, which has been shown to regulate the expression of *SOX10*, albeit via a different enhancer. Furthermore, within MOLs the enhancer was also co-accessible with the *CDC42EP1* gene, which is associated with myelin sheath compaction. These results highlighted the utility of our dataset for finding putative regulatory interactions within neural populations.

We also used the H3K27ac and ATAC datasets to build a core regulatory transcription factor network, where we found several HOX proteins within the OPC network displaying strong predicted regulatory activity. These TFs were also identified within the MOL regulatory network, but with lower predicted activity. As the HOX genes are expressed in the developing spinal cord, we then checked if there was regional specificity between the motor cortex and cervical spinal cord. Differential accessibility analysis revealed increased HOX chromatin accessibility within spinal cord-derived OLGs. Despite this, spinal cord OLGs did not display robust HOX expression, ascertained by performing multiOme ATAC+RNA co-profiling and by checking an adult CNS transcriptomic atlas, which suggested epigenetic memory of these genes in the adult OLGs. Genome browser tracks further revealed increased H3K27ac and ATAC signal and reduced H3K27me3 signal around 3′ HOX-A and HOX-D genes. This signal distribution was inverted at the 5′ end of the clusters. Though H3K27me3 at the 3′ end was less than that at the 5′ end, it was still greater than the signal outside the cluster, indicating that it may be sufficient to prevent robust transcription.

HOX activation during development is also regulated by the 3D chromatin architecture around the clusters. Given the epigenetic memory of the chromatin at HOX loci in OLGs, we then investigated if the 3D architecture plays a role in maintaining the development-like state. We performed Micro-C in iPS-derived human OPCs (hOPCs) and observed the development-associated c-Dom and t-Dom TAD structures at the HOX-A and HOX-D clusters. These TADs contain enhancers that regulate the expression of the HOX genes in development. Interestingly, we observed increased activation (ATAC + H3K27ac) signal at the enhancers in the adult spinal cord OLGs, suggesting these genes may also be primed for expression through 3D contact with the distal enhancers.

HOX activation is also seen in glioblastoma, and recent studies have shown that the molecular architecture around the HOX genes in midline high-grade gliomas (HGG) faithfully recapitulates the locus of origin, captured by the spatiotemporal address of the cell of origin. Of the HOX genes activated in H3K27M mutation-bearing pontine HGGs, several presented primed promoters with high ATAC and H3K27ac signal in our non-diseased spinal cord-derived OLGs. Further, the 3D chromatin architecture in pontine gliomas resembled a similar pattern to our Micro-C dataset. Collectively, these results indicated that epigenetic memory around HOX genes is maintained well into adulthood, but repressed via H3K27me3, and that the disruption of H3K27me3 deposition in H3K27M-bearing HGGs may de-repress these genes, driving their expression.

5.4 Paper IV - 4-in-1 multimodal single-cell epigenome and transcriptome co-profiling

Multimodal single-cell technologies have improved our understanding of cellular processes. By capturing multiple orthogonal components within the same cell, these technologies enable direct comparisons, rather than necessitating correlations and integrations of datasets from different cells and experiments. While many multimodal approaches are available, some combinations remain more difficult to profile than others – such as chromatin proteins and the transcriptome. In this study, we present nanoCTAR (pronounced: nano-star), a 4-in-1 approach that simultaneously profiles chromatin accessibility, two histone modifications and the transcriptome in thousands of single cells.

The multiOme platform uses a ligation reaction to introduce the cell barcode within the microdroplets, whereas the scATAC-seq platform does so using strand extension in PCR. Hence, we designed a new adapter cassette for the Tn5 which allows for ligation and incorporates a DNA modality barcode. The longer protocol also presented a challenge for maintaining RNA integrity, and so we added RNase inhibitors to all buffers and performed all steps (except tagmentation) on ice or at 4°C. The shared cell barcode within individual droplets links the DNA modalities to the RNA, and the DNA barcode on the adaptor cassettes allows deconvolution of each DNA modality. The custom adapters also necessitated custom read setup and sequencing primers.

We tested nanoCTAR on freshly dissociated embryonic mouse brains at E14.5, a time point characterised by dynamic cell lineage transitions, targeting the histone

modifications H3K27ac and H3K27me3. We sequenced to a depth of 50 million reads each for the epigenomic library (ATAC + H3K27ac + H3K27me3) and the transcriptomic library (RNA) and resolved 9820 cells shared across all four modalities. The genomic distribution of the different DNA modalities faithfully reflected the expected patterns, with high TSS enrichment scores for H3K27ac and ATAC and low for H3K27me3. The RNA dataset captured ~4000 RNA molecules and ~2000 genes per cell, with minimal mitochondrial genes.

We then integrated the RNA dataset with a published mouse brain atlas to identify the cell types and observed coherence between all four modalities at cell type-specific genes. Collectively, these results showcase the potential of nanoCTAR for co-profiling four orthogonal modalities in the same cell, even with shallow sequencing.

6 Conclusions & Perspectives

The work presented in this thesis initially aimed to further our understanding of the epigenetic landscape and gene regulatory mechanisms, particularly within the OLG lineage and the broader central nervous system. **Paper I** and **Paper III** represent endeavours to explore these aspects in mouse and human contexts, respectively. Along the journey, the aim expanded to include technology development for epigenetic profiling that could help navigate the dense and complex layers of the epigenome. This second endeavour is captured in both **Paper II** and **Paper IV**, but the latter in particular aimed to push the boundaries of what was possible to profile within individual cells.

6.1 On oligodendroglial biology

The great themes in OLG research have broadly ranged from understanding the spatiotemporal events underlying OPC production during development, the molecular cues driving OPC differentiation, microenvironmental stimuli triggering myelin production in MOLs, myelination dynamics and demyelination-associated pathologies. The single-cell era has certainly helped us march faster and further down these research avenues but has also revealed new paths ripe for exploration.

In multiple sclerosis, OLGs were historically seen as passive victims of the immune cell-mediated inflammatory cascade. However, there is growing evidence that refutes this and shows that OLGs are indeed capable of interacting with and potentially modulating their microenvironment. MHC-II gene expression in OPCs and antigen presentation functionality certainly make them culpable, but it is unclear whether this aggravates or ameliorates the immune cascade. In **Paper I**, we showed that OLGs display promoter priming at several immune genes that are activated in the context of EAE. We also proposed a de-repression model of activation of these genes mediated by the EZH2 subunit of the PRC2 complex. While chromatin accessibility was observed at single-cell resolution in EAE, the histone landscape was characterized in bulk assays in cultured OPCs. This prompts the question – do all OPCs in EAE exhibit promoter priming with H3K27me3, or is it confined to a subset? While the bulk assays cannot answer this, scCUT&Tag (**Paper II**) and its successor nanoCUT&Tag could provide the granularity to address this. We know that OPCs presenting the immune-like state have a diminished capacity for differentiating into MOLs (Falcão et al., 2018; Kirby

et al., 2019). Conversely, OPCs play a well-established role in migrating to lesions and differentiating to MOLs (Franklin & Ffrench-Constant, 2008). These opposing points can be reconciled by the observation that only a subset of EAE-derived OPCs display the immune-like phenotype, encompassing both antigen-presenting and immunosuppressive characteristics. Consequently, we hypothesised that OPCs may encounter a fate choice – to contribute to remyelination efforts or transition to an immune-OPC state. However, the temporal aspect of the choice remains unresolved – is the fate choice triggered by the inflammatory context, or is it predetermined, and if so, when? The nanoCUT&Tag experiments again may help address this, and I suspect that a subset of OPCs are predisposed to initiate the immune transition in EAE. The first line of support for this is the observation that while OPCs tile the entire CNS, only a fraction differentiates into MOLs. A parsimonious explanation for this is that OPCs have alternative roles in physiological contexts. Indeed, several lines of evidence provide more concrete support for this.

A study in the zebrafish spinal cord by Marisca et.al. identified a sub-group of OPCs exhibiting increased calcium signalling, but a limited capacity for differentiation. This sub-group acts almost like sentinels, surveying and interacting with the neural circuit and in response to signalling cues, can divide to give rise to a different sub-group of OPCs with increased potential for differentiating into MOLs (Marisca et al., 2020). Furthermore, OPCs also express ion channels enabling them to be electrically responsive, and as they age, become functionally heterogenous in their ion channel density, possibly affecting their differentiation potential (Spitzer et al., 2019). Finally, OPCs in the mouse brain are also involved in fine-tuning neuronal circuits either by regulating axon growth and axon phagocytosis (Buchanan et al., 2022; Xiao et al., 2022). This capacity for phagocytosis aligns well with the observations of antigen uptake for presentation seen in OPCs (Falcão et al., 2018; Kirby et al., 2019). It would be interesting to investigate whether similar pathways and mechanisms are used in both settings.

Our H3K27ac-directed HiChIP data in IFN-g treated OPCs showed increased looping between MHC-II genes and upstream and downstream enhancers, in line with these genes being expressed upon IFN-g treatment. Interestingly, we also observed looping interactions in control OPCs, though the signal was weaker (Meijer et al., 2022). Although cells were exposed to IFN-g for six hours in the experiment, preliminary data showed an increase in MHC-II gene expression within 15 minutes of IFN-g exposure. This rapid activation suggests that a subset of OPCs

may already possess a primed chromatin architecture in the absence of inflammatory signals, which could allow for rapid activation of these genes when needed. A caveat here is that the HiChIP experiment was conducted in a bulk population. Therefore, while it is difficult to definitively conclude that a subset of cells display this primed 3D architecture in control conditions, it is plausible that the presence of a weak HiChIP signal in Control that increased upon IFN-g exposure reflects an increase in the proportion of cells displaying the promoter-enhancer looping.

However, another intriguing question arises – could there be an alternative advantage for maintaining the chromatin architecture in the first place? We know the histone landscape is primed, so why prime the architecture too?

This question primes us well for discussing **Paper III**, where we observed epigenetic memory of the development-associated HOX genes in adult spinal cord-derived OLGs. HOX gene expression orchestrates anterior-posterior patterning in the developing neural tube, which explains the signal being observed in the neuroectoderm-derived OPC, MOL, and AST populations, but not the yolk-sac-derived MIGL population (Alliot et al., 1999; Krumlauf, 1994). We did not find spinal cord-derived neurons in our dataset, likely due to the difficulty in isolating these hypoxia-sensitive cells in the tissue. We also observed a decoupling of HOX accessibility from robust gene expression assessed by multiOme analysis. Nonetheless, proving a negative signal, i.e. absence of HOX transcripts, is challenging, and could also be due to the sensitivity of the technology or technical dropouts. Another challenge in asserting the attenuated HOX expression in the adult spinal cord OLGs is that HOX genes can indeed be expressed postnatally in different cell lineages. A classic example of this is in skin fibroblasts, where the cells maintain transcriptional memory of HOX gene expression from development (Rinn et al., 2006, 2007). This memory is robustly maintained in young and old fibroblasts, indicating the HOX program may serve as an internal map for remembering positional identity. Bulk transcriptomic datasets have also provided insights into HOX expression in the spinal cord (GTEx Consortium, 2020). However, the bulk nature of these tissue-level datasets makes it difficult to discern specific cell type contributions to the signal. While it has been shown that neurons can and do maintain HOX gene expression into adulthood, less is understood about expression patterns in the OLG lineage (Briscoe & Wilkinson, 2004; Hutlet et al., 2016). We and others have previously demonstrated that patterning genes, including the HOX genes, are expressed by OPCs in the

developing mouse spinal cord, but are subsequently downregulated after birth (Marques et al., 2018; Zeisel et al., 2018). Our multiOme analysis in the adult spinal cord corroborates this finding, showing attenuated HOX expression relative to other genes. Similarly, a transcriptomic atlas by Siletti et.al. also revealed that the majority of HOX genes were not expressed in adult spinal cord-derived OPCs or MOLs (Siletti et al., 2023). Even the most prominently detected HOX genes – *HOXB1*, *HOXD1*, and *HOXD3* – were found in less than 15% of cells and at residual levels. In contrast, the TF-encoding *SOX10*, *OLIG1*, and *OLIG2* genes, showed higher expression levels and were detected in nearly all cells (though the high fraction of detection is likely due to their crucial role in the lineage). While it may be argued that HOX genes are challenging to detect in droplet-based assays, such as those used in our multiOme analysis and the transcriptomic atlas, another study using the same droplet-based platform and focusing on the embryonic human spinal cord, clearly captured HOX genes both at higher levels and in up to 60% of cells (Rayon et al., 2021). As HOX expression in the developmental dataset would be expected, it serves as a positive control and suggests that technical limitations are unlikely to be the issue. Nonetheless, using orthogonal approaches such as probe and imaging-based in-situ Sequencing (ISS), EEL, and MERFISH, could help ascertain the true levels of HOX expression in these cells (Borm et al., 2023; K. H. Chen et al., 2015; Fang et al., 2022; Ke et al., 2013).

The finding that a small proportion of OLGs may express low levels of HOX genes is supported by our observation in the nanoCUT&Tag data. The co-profiling of H3K27me3 and H3K27ac in the same cell lets us rank the cells according to one modality and look at the signal in the other modality for that same cell. Doing so revealed that OLGs could segregate into two sub-groups based on the H3K27me3 deposition at the 3′ end of the cluster. While H3K27me3 was reduced at the 3′ end of the cluster relative to the 5′ end in all cells, one sub-group presented further reduction of H3K27me3, while the H3K27ac signal remained the same in all cells. These findings suggest that this sub-group, with an increased H3K27ac-to-H3K27me3 ratio, may be the source of the residual HOX gene expression that we and others observed (Seeker et al., 2023). However, to assert this definitively would require an experiment where the histone landscape is co-profiled with the transcriptome – achievable through nanoCTAR (**Paper IV**) or with the Droplet Paired-Tag method (Xie et al., 2023; Zhu et al., 2021).

Another area of investigation is the link (if present) between the maintenance of epigenetic memory at HOX genes and their expression in gliomas (Jessa et al.,

2022). Our data would indicate that the basal level of H3K27me3 at 3′ HOX genes is sufficient to maintain silencing of these genes, in the presence of H3K27ac and chromatin accessibility. Furthermore, diffuse midline gliomas (DMGs) bearing the lysine-to-methionine mutation of histone H3 (H3K27M), exhibit global disruption of PRC2 activity (Bender et al., 2013; P. W. Lewis et al., 2013; Venneti et al., 2013). PRC2-mediated H3K27me3 deposition at HOX genes has been well characterized and shown to maintain HOX gene silencing in Drosophila (Coleman & Struhl, 2017). Therefore, HOX expression may be a consequence of the molecular state of the chromatin coupled to the mutation. However, non-H3K27M-bearing cancers such as posterior fossa group A ependymomas (PFA-EP), also express HOX genes. PFA-EP tumours express the enhancer of Zestes homolog inhibitory protein, EZHIP, which inhibits PRC2 activity through direct interaction with the EZH2 subunit (Jain et al., 2019, 2020). Therefore, the convergent mechanisms of H3K27M-bearing and non-bearing gliomas would indicate that PRC2 disruption potentially drives their expression. Perturbing PRC2 activity in vitro could test this hypothesis. In **Paper II**, we used a pharmacological EZH2 inhibitor to investigate the effect of H3K27me3 depletion on the primed MHC-II gene expression in primary mouse OPCs. Although our focus was to observe the effect of EZH2 inhibition in Control and IFN-g treatments, interestingly, several HOX genes were upregulated in both treatments, indicating it was an effect of the EZH2 inhibition. Further, the top GO terms for genes upregulated upon EZH2 inhibition included "pattern specification process", "embryonic morphogenesis", and "appendage development", all of which included the HOX genes in the gene set (**Paper II** – Supplementary Table 7). These experiments were performed in mouse OPCs, but it would be pertinent to perform similar experiments in the iPS-derived hOPCs too. A downside of these experiments is that global and non-specific EZH2 inhibition makes it difficult to decipher if HOX expression is a primary or secondary effect. Here, using a targeted inhibition approach, as shown by Levy et.al., employing an EED-binder (EB) fused to dCas9 (EB-dCas9) could allow for locus-specific EZH2 inhibition. The EB protein outcompetes EZH2 for binding to EED, another subunit of the PRC2, thereby mimicking the inhibition of EZH2 while providing a dCas9 and guide RNA-driven locus specificity (Levy et al., 2022).

The focus on perturbing PRC2 to elicit HOX expression is logical given its role at this locus, however, other complexes may also be involved. KMT2A/B is a multi-protein complex responsible for laying down H3K4me3 at active and bivalent genes (Cenik & Shilatifard, 2021). Recently, Sparbier et. al. found that Menin, a sub-

unit of KMT2A/B, specifically played a role in directing the complex towards active promoters. Inhibiting Menin released the complex from the active genes, shifting the localization in favour of bivalent genes. This led to an increase in H3K4me3 at the bivalent promoters, leading to the de-repression of these genes, phenocopying effects seen in leukaemia (Sparbier et al., 2023). H3K4me3 deposition at HOX loci has been associated with the expression of these genes in the developing mouse tail bud (Soshnikova & Duboule, 2009), and though we have not yet characterized H3K4me3 in human OPCs, these findings suggest it is worth investigating. This also motivates the experiment inhibiting Menin alone, or in combination with EZH2, to see the effect on HOX expression in these cells.

These perturbation experiments, though powerful, focus on the molecular landscape, but we also ought to look at the 3D chromatin conformation. HOX expression is driven by enhancer interactions in the flanking centromeric (c-Dom) and telomeric domains (t-Dom). A recent study by Deforzh et.al. showed that cortical astrocytes exhibit 3D chromatin interactions between the HOX-D embedded microRNA, miR-10b, and a t-Dom embedded enhancer-associated lncRNA, LINC01116. mi-R10b expression is undetectable in the healthy brain but is expressed in over 90% of high-grade gliomas (HGG), thereby showing strong disease association. The study found that activation of miR-10b or the LINC011116 drove HOXD gene expression and induced a glioma-like hyperproliferative phenotype in cultured astrocytes. Further, knockdown of either disrupted the 3D chromatin contact and TAD structure and led to glioma cell death (Deforzh et al., 2022). Our iPS-derived hOPCs also display 3D contact between miR-10b and LINC01116. Preliminary CRISPR activation experiments targeting miR-10b and LINC01116 using dCas9-p300 in the hOPCs are suggestive of a similar transcriptional response, with an increase in HOXD gene expression, though whether the hyperproliferative phenotype is recapitulated remains to be tested.

These observations hint that primed chromatin architecture and 3D contact seem to be a precursor for the expression of these genes and formed the basis for the question posed earlier when discussing the MHC-II promoter-enhancer priming in (potentially) a subset of OPCs, assessed by HiChIP. Chromatin looping is driven by interactions between CTCF insulator proteins and the ring-like Cohesin complex. DNA loop extrusion through the cohesin ring is an ATP-driven process, and although looping can occur independently of ATP, the rates are much slower (Kim et al., 2019). Looping stops upon encountering CTCF anchors bound to the chromatin. Therefore, if maintaining a looped structure is ATP-independent, it is

plausible that cells may find it energetically advantageous to pre-establish 3D contacts before circumstances warrant gene expression. Indeed, a similar phenomenon has been observed in memory encoding in neurons. Memory consolidation was found to reorganise chromatin architecture, favouring enhancer-promoter priming (Marco et al., 2020). Logically, this would make sense – maintain 3D contact and then modify the local histone landscape to drive expression. The implication isn't that all 3D contacts are pre-established, but perhaps only those which are associated with quick-response genes and primed genes. This would enable rapid gene activation when needed, as the contact has been established. That 3D interaction lies upstream of gene activation is not a novel concept, and the discovery of TADs supports this (Dixon et al., 2012). TADs provide an insulated neighbourhood for increased contact between CREs and promoters, as discussed in **Chapter 2**. In an elegant study, Bonev et.al. used Hi-C to capture the global changes in chromatin architecture in mouse neural cells from the embryonic stem (ES) cell stage to neural progenitors (NPCs) to cortical neurons. A prominent feature was the massive re-wiring that occurred predominantly from the ES to NPC stage, indicating that perhaps less rewiring is needed for the neural lineage cell specification (Bonev et al., 2017).

The energetics hypothesis also has an interesting corollary – could observing 3D interactions *alone* inform about putative future states? The findings of **Paper I**, where we saw the primed chromatin interactions between MHC-II genes and their enhancers in the absence of expression, suggest that they could. The observations of **Paper III**, where chromatin priming in adult spinal cord OLGs was in line with the 3D contacts seen in the iPS-derived hOPCs, also suggest the same, although we would need to observe the chromatin architecture within the same adult tissue to be sure. Nonetheless, the histone priming, coupled with the hOPC 3D interactions and the miR-10b – LINC01116 functional studies all support the idea of using 3D contact information to identify putative transcriptional programs.

6.2 On single-cell methods

Single-cell methods have significantly advanced our understanding of cellular processes, a theme that has been explored throughout this thesis. At its core, the development of these methods revolves around three key phases: capture, read, and interpret. The reading phase involves library sequencing, while interpretation relies on demultiplexing data, and using computational software and algorithms to extract insights. While most of the focus (deservedly so) is spent in the

interpretation phase, either for handling noisy sparse data, technical biases, or through developing powerful computational methods for analysis, it is the first capture phase that I find particularly challenging and exciting. The capture phase is an exercise in creativity and engineering where we need to leverage the known properties of different biomolecules to tag, selectively enrich, extract, and store the information in a DNA format. Multimodal methods, including nanoCTAR in **Paper IV**, do the same but with the added challenge of preserving all modalities during the capture process.

Looking forward, it will be interesting to see how these methods are used to generate further insights. All the papers in this thesis have focused quite extensively on chromatin accessibility and the histone landscape, and on their role in identifying CREs. The argument for profiling the histone landscape is the increased granularity that it provides relative to chromatin accessibility. However, the reliance on antibodies for profiling the histone landscape and other chromatin-associated proteins such as TFs, is limited by the availability of high-quality antibodies. In this regard, chromatin accessibility has an advantage. However, while it has been increasingly clear that relying on sequence alone may not be sufficient to understand regulatory dynamics, several recent efforts have attempted to reverse-engineer sequence information to predict regulatory elements. This is done either through the design of synthetic enhancers (Minnoye et al., 2020; Taskiran et al., 2024), or more recently, using deep learning-based approaches to decipher meaning from sequence alone (Geng et al., 2022; Mannens et al., 2023; Piecyk et al., 2022; H. Zhang et al., 2019).

As analysis methods advance and reveal further insights, they will undoubtedly lead to more questions and highlight the limitations of current methods. This will underscore the need to keep revisiting and refining the initial capture phase of methods development to address these challenges and further our understanding of biological systems.

In conclusion, this thesis has leveraged both established and newly-developed single-cell methods to uncover novel regulatory mechanisms in oligodendroglia, both in healthy and diseased settings. These discoveries have also highlighted previously unexplored roles for this lineage, expanding our understanding of its potential functions in various contexts.

7 Ethical Considerations

The work carried out in this thesis involved working with both mouse (**Papers I, II, IV**) and human tissue (**Paper III**).

7.1 Mouse experiments

All mouse experiments were planned thoroughly beforehand in line with the PREPARE and the Three R guidelines, to both minimize the number of experiments and animals being used. Mouse handling and tissue collection procedures were done following the local animal protection legislation, and only after the approval of all ethical permits.

The mouse model of MS used in **Paper I** – EAE – induces demyelination in the mouse CNS, leading to paralysis of the tail and limbs. To reduce the suffering caused by weakness, a diet gel and extra ground food were provided to the mice. In addition, grids were placed in the cage to make access to food and water easier, though saline injections were administered if signs of dehydration were observed. These interventions aim to reduce the suffering but unfortunately cannot eliminate it. During the experiment, all mice were checked on daily (or twice daily after symptom onset) and an open dialogue was maintained with animal technicians and veterinarians.

7.2 Human experiments

In **Paper III**, we used adult human post-mortem tissue brain and spinal cord of donors without pathology sourced from the Edinburgh Brain Bank. Consent was obtained from the family of all donors prior to tissue collection. Donor identity was represented solely through a pseudonymized alphanumeric code. While broader metadata regarding donor sex, age, post-mortem interval and primary cause of death was shared with us, information such as name, address, ethnicity, religion, socioeconomic status, date of death, sexual orientation or any other sensitive personal information that could identify the donor or the family was not available to us.

The snATAC-seq and nanoCUT&Tag experiments performed on the human tissue generated DNA libraries that were sequenced using paired-end sequencing. While the DNA fragments in the libraries contain anywhere between 150 bp - 2000 bp

of genomic information from the donors, a maximum of 50 bp from each end of the library was sequenced. The intervening information was not read directly from the genomic fragments but was mapped to the reference genome during the paired-end read alignment step. Nonetheless, the captured 50bp constitutes sensitive information, and as such was only stored and processed on secure servers approved for the storage of human data. In keeping with the Findability, Accessibility, Interoperability, and Reuse (FAIR) principles for data sharing and accessibility, we deposited the processed data (count matrices, browser tracks etc) on servers such as the UCSC Genome Browser and the CZI CellxGene platform. These data formats cannot be linked back to the raw fastq files containing the sequence information. The raw sequencing data was deposited into the European Genome-Phenome Archive (EGA), a secure storage repository for sensitive data, with restricted data access. Access to the EGA-secured data by external applicants is only granted after we, as the data access controllers, manually approve the motivated application.

7.3 On the role of technology

Animal research is a contentious topic but holds promise in driving forward our understanding of human health and disease. Nonetheless, it is important to recognize that conducting this kind of research needs to take into account the pros and cons and should be at the forefront during study design.

In addition to minimising animal usage, we also have an ethical responsibility to maximise the information extracted from each sacrificed animal. Though often overlooked, multimodal technologies hold promises in this regard. By capturing more modalities in a single assay, we move towards maximising the information leached from the starting sample. This has the potential to lower the number of experiments overall, and the dependence on multiple samples, particularly if each sample corresponds to the sacrifice of an animal.

Finally, computational advancements, including the recent strides in large language models (LLMs), have significantly expanded our capabilities in virtually all domains. With these LLMs now making their way into biological research (Cui et al., 2024; Hou & Ji, 2024), there is general excitement about the insights they may yield when applied to the cumulative research corpus that has been collected over the decades. Undoubtedly, these powerful models will accelerate the pace of research, but hopefully, the discoveries they bring will also reduce the reliance on animal research.

8 Acknowledgements

They say it takes a village…

This journey has been long, challenging, and incredibly rewarding. I am grateful for the many wonderful people I have met along the way, each of whom has contributed to making this experience more enjoyable.

First, and foremost, to **Gonçalo**! Thank you for taking me under your wing 6 years ago as a master's student. Based on the lab environment you had created and your leadership style, I immediately knew I wanted to stay and do my PhD here – thank you for later giving me that opportunity. You are a brilliant scientist, and I have learned so much from you, but it can be summarized in something I call the N+1 philosophy - whenever I would present to you, even if I had thought of N possibilities to explain the data, you had this amazing ability to think for a moment, find the core issue, and provide a $(N+1)^{th}$ perspective that I hadn't thought of. Over the years, in chasing N+1 myself, I learned from you how to think more deeply and critically about every aspect of science and research, and for that I am grateful. Thank you for being a caring mentor, an inspiring leader, and for giving me the room and opportunities to explore, learn, and grow into an independent scientist. (P.S. even now at the end of the PhD, you still manage to find N+1, though I'd like to think I make you work for it!)

Thank you, **Carlos Rovira**, for our conversation back in 2017. You pointed me towards Stockholm and to Gonçalo's lab, and for that, I will be forever grateful.

Marek, thank you for being a great co-supervisor, mentor and epigenomic trailblazer! You were proof that it was possible to excel at the bench and in analysis, and encouraged and supported me when I wanted to learn bioinformatics. Thanks for all the sharp and critical feedback over the years, informal brainstorming sessions and for being a great sounding board to bounce crazy ideas off.

Thank you, **Simon**, for being my co-supervisor, and for the discussions about science but also for the advice on entrepreneurial paths.

Thank you, **Bing Ren**, for being my opponent – I look forward to our discussion at the defence! Thank you, **Claudio Cantu**, **Rickard Sandberg**, and **Johan Jakobsson** for agreeing to be on my examination board.

To the stellar cast that make up the GCB lab –

Mandy, my first supervisor in the lab, I have learned a lot from you! Thank you for all the fun conversations, and discussions about epigenetics, for teaching me a range of methods and techniques in my first year, and of course, your crystal-clear instructions (first step - we wear our lab coats!). **Petra**, my academic twin, it's crazy to think how far we've come in the past 6 years! I am grateful for our friendship and know that I can come talk to you about life, career, or anything that will end up with me saying "oh Petra" at least once. You're an amazing scientist and a wonderful person. I will miss your awkward humour, our after-works (we should bring back the three whiskeyteers with Fabio!), and the mutual understanding of not talking on the bus in the mornings! **Fabio**, you are an incredible friend and one of the kindest people I know. Thank you for your support, for always being down for a fika chat, and for reminding me to take breaks, so I don't lose myself in the stress. Thank you for introducing me to true Brazilian culture, including the barbecues (though I will not miss the chicken hearts), and of course the beloved feijoada (even though I prefer the vegan version). Sorry for the pain from the recent workout, but we'll go again, right?! **Chao**, thanks for always being willing to help, and somehow always having the answer to any administrative questions! Thanks for keeping my reflexes on point when you randomly pop up and surprise me at my desk, and for helping me finish the stock of Fisherman's Friend mints. **Eneritz**, the stoic guru! Thank you for all your help and patient guidance with bioinformatics over the years, and for the chat in Berlin about the importance of balance in life. **Leslie**, thank you for your kindness and your incredible intuition, with timely words of encouragement and support when they were needed the most. Thanks for all the laughs, fun conversations, and for (unwittingly) showing how to take tipping to an absurd level! **Yonglong**, thanks for being a great office mate, deep talks on protocol development, and for the contagious laugh when you saw my Tn5 illustration. I will forever be impressed by your ability to summon the fires of hell and stuff it inside innocuous items like cucumbers. **Tony**, thanks for the humorous conversations, and for all your work in keeping things running smoothly in the lab – it is not an easy job, and thanks for teaching me the only way to correctly eat a taco. **Karl**, I have learned a lot about resilience from you. Thank you for all the discussions, kind words of support, and for all the help with experiments. **Bastien**, thanks for bringing a new level of order to data storage and code tracking in the lab. Your competitive spirit is legendary, and I won't understand how you managed to swim faster than David and I could kayak.

Neemat, for introducing me to Nanaimo bars, eh! Thanks for showing me how to season maple syrup with waffles (yes, you read that correctly). **Mattia Z.**, you are possibly one of the fastest learners I've ever met! Thank you for taking on the challenges of multi-modal method development with me. **Ting**, thanks for reviving the social scene at MolNeuro! I look forward to joining the whiskey nights after submitting this thesis. **Federico**, thanks for the guitar serenade that memorable evening in Visby, and the occasional chat. **Charles**, snowboarder first, scientist second? ;) Thanks for all the cheerful conversations and for introducing me to the best fish and chips in Stockholm! **Özge**, we didn't overlap much, but I'm sure you will do great in your PhD! **Naomi** and **Toba**, it was wonderful mentoring you both and seeing your progress through your thesis projects. Thank you for your effort and tenacity in what was a rather challenging project. I look forward to seeing the heights you scale in your respective PhDs. **Esmee**, **Fleur**, and **Mattia G.**, thanks for the fun lunch conversations, the encouragement to go jump on trampolines, and like Ting's whisky nights, I hope to join your game evenings soon. **David**, the forever unfazed! Thanks for introducing me to Scanpy before Seurat. **Alessandro**, thanks for your sharp and intense questions which taught me to a lot think more critically about data. **Ana F.**, for your kindness, encouragement and for being the grand-mentor! Thank you, **Florian**, **Liepa**, **Milda**, and **Natali** for the fun times as master students.

Huge thanks to all the MolNeuro PIs – **Sten**, **Ulrika**, **Jens**, **Ernest**, **Per**, **Onur**, and **Patrik**, for making this the amazing place it is. Many have said this before, but MolNeuro really is a unique place in KI. Having world-class science with titans in the field, and combining it with a vibrant, collaborative, and friendly atmosphere made this place almost magical.

Camiel, my calmest friend, thanks for countless conversations, after-works, dinners, discussions about ATAC analysis, and of course the shared interest in F1 (I apologize for my crazy side coming out during Monza 2021!). I will miss our lunch chats and seeing your Bialetti French press flip me off! **Emilia**, thanks for being a fun office mate, for understanding "hoodie mode", and for the long chats about the most random things ranging from research, post-leg day soreness, PhD stresses and woes, to the intricacies of sleep tracking. **Lars**, for the fun conversations, groovy music, and inspiring, jaw-dropping EEL visuals in your presentations. **Alejandro**, for the nerdy chats about longevity, NMN, and NR. **Shigeaki**, for showing me the mouse brain in 3D VR, that was super cool! **Ka Wai** thanks for the pre-covid dinners, your tasty desserts, and shared love for peanut

butter. **Lijuan**, for all the help with 10x kits! **Dmitry**, for enlightening me about the non-trivial issues with nuclei isolation from monkey neurons. **Simone C.**, for the fun conversations, and encouragement to learn bioinformatics. **Jokubas**, for the deadpan humour. **Peter** for all the help with Monod. **Wei** for the chats about post-PhD careers. **Yizhou**, for the discussions and questions about projects. **Kasra,** for the sick music in the lab, and the occasional F1 chats. **Roland**, for all your work in keeping MolNeuro running smoothly and the fun Friday emails. **Mingdong**, for the chats, when we crossed each other on weekends. **Elin V.**, **Elin G.**, **Natalie**, **Hayley**, **Marla**, and **Doris**, for the lunch conversations. Thank you, **Fatima**, **Naoya**, **Dagmara**, **Dongoh**, **Marina**, **Ziwei**, **Guochang**, **Elisa**, **Sam**, **Alessandra**, **Chiara**, **Ivana**, **Jussi**, **Jie**, **Carmen**, **Lauri**, **Shanzheng**, **Emelie**, **Lisbeth**, **Lisa**, and so many more at **MolNeuro** for all the great memories!

Thank you to **Elisabet E.** at NGI for all the help with sequencing and for bearing with my urgent orders, **KI Career Services** for arranging external internships, and **Christopher U.** for mentoring me at SEB, where I gained valuable perspective, the **Medbioinfo program** for the courses, the network, and the fantastic retreats. Thanks, **CZI**, for funding my project, the global network, and also for flying me business class to California!

To my friends in Sweden – thank you to **Imran** & **Kätlin** for all the inevitably deep yet fun conversations whenever we meet. Thank you, **Mariana**, **Helene**, **Pablo**, **Ana**, **Tanja**, **Lukas**, and **Savreen** for the fun times and dinners. To my friends around the world –**Arun**, **Aditya** (Jupudi!), **Niharika** (Zing!), **Prakhar**, **Sudharshinee**, **Utkarsh**, and **Vasu** thanks for the support and encouragement despite my periodic absence. **Aneesh**, thanks for being a good friend, for staking out at the airport as we waited for my luggage, showing me around San Francisco, and no, I didn't manage to genetically engineer a four-leaf clover in my PhD, sorry!

Thank you to my high school teacher **Gargi S.**, you sparked my excitement about molecular biology, setting me on the path that led me to where I am today.

And now a heartfelt thank you to my entire family, who have been incredibly supportive and loving throughout.

To my dear in-laws, **Amma** & **Papa Shenoy**, thank you for welcoming me into the family and all your support!

To my sister **Apeksha**, the more creative sibling - Thank you for feeding me with cakes and pastries, and possibly the best cookies I've ever had. You are incredibly

talented, hard-working and most importantly, kind. With so much to offer the world, I look forward to seeing where life takes you in the future.

Mummy and **Papa**, this thesis is dedicated to you both. I will be forever indebted to you for everything you have done for me and am deeply grateful for all your support, sacrifice, and love. Mummy, you raised me to believe that no dream was too big and no goal was too lofty. Your constant encouragement and support gave me the confidence to pursue my passions with joy. Papa, you are the embodiment of discipline and support, and showed through example, the importance of work ethic, integrity, and dedication, in the pursuit of success. With these values instilled by both of you from a young age, I had the courage and conviction to chase my dreams.

Finally, to dearest **Aditi**, my wonderful wife and best friend, words cannot express how grateful I am to have you as my partner in this journey. Sharing the experience of doing our PhDs together is something I always cherish. We have celebrated, consoled, and navigated the many ups and downs of this crazy ride side by side, and I wouldn't have had it any other way. You are my rock, and your endless support and love have always given me the strength to keep moving forward. Thank you for everything! I look forward to whatever life has in store for us next and tackling it head-on by your side.

9 References

Adey, A. C. (2021). Tagmentation-based single-cell genomics. *Genome Research*, *31*(10), 1693–1705. https://doi.org/10.1101/GR.275223.121

Aizawa, A. (2003). An information-theoretic perspective of tf–idf measures q. *Information Processing and Management*.

Alliot, F., Godin, I., & Pessac, B. (1999). Microglia derive from progenitors, originating from the yolk sac, and which proliferate in the brain. *Brain Research. Developmental Brain Research*, *117*(2), 145–152. https://doi.org/10.1016/s0165-3806(99)00113-3

Allis, C. D., & Jenuwein, T. (2016). The molecular hallmarks of epigenetic control. *Nature Reviews Genetics*, *17*(8), 487–500. https://doi.org/10.1038/nrg.2016.59

Alvarez-Medina, R., Cayuso, J., Okubo, T., Takada, S., & Martí, E. (2008). Wnt canonical pathway restricts graded Shh/Gli patterning activity through the regulation of Gli3 expression. *Development*, *135*(2), 237–247. https://doi.org/10.1242/dev.012054

Attwood, J. T., Yung, R. L., & Richardson, B. C. (2002). DNA methylation and the regulation of gene transcription. *Cellular and Molecular Life Sciences CMLS*, *59*(2), 241–257. https://doi.org/10.1007/s00018-002-8420-z

Banerji, J., Rusconi, S., & Schaffner, W. (1981). Expression of a β-globin gene is enhanced by remote SV40 DNA sequences. *Cell*, *27*(2), 299–308. https://doi.org/10.1016/0092-8674(81)90413-X

Bannister, A. J., & Kouzarides, T. (2011). Regulation of chromatin by histone modifications. *Cell Research*, *21*(3), 381–395. https://doi.org/10.1038/cr.2011.22

Baran-Gale, J., Chandra, T., & Kirschner, K. (2018). Experimental design for single-cell RNA sequencing. *Briefings in Functional Genomics*, *17*(4), 233–239. https://doi.org/10.1093/BFGP/ELX035

Bartosovic, M., & Castelo-Branco, G. (2022). Multimodal chromatin profiling using nanobody-based single-cell CUT&Tag. *bioRxiv*, 2022.03.08.483459-2022.03.08.483459. https://doi.org/10.1101/2022.03.08.483459

Bartosovic, M., & Castelo-Branco, G. (2023). Multimodal chromatin profiling using nanobody-based single-cell CUT&Tag. *Nature Biotechnology*, *41*(6), 794–805. https://doi.org/10.1038/s41587-022-01535-4

Bartosovic, M., Kabbe, M., & Castelo-Branco, G. (2021). Single-cell CUT&Tag profiles histone modifications and transcription factors in complex tissues. *Nature Biotechnology*, *39*(7), 825–835. https://doi.org/10.1038/S41587-021-00869-9

Baxter, A. G. (2007). The origin and application of experimental autoimmune encephalomyelitis. *Nature Reviews Immunology*, *7*(11), 904–912. https://doi.org/10.1038/nri2190

Bender, S., Tang, Y., Lindroth, A. M., Hovestadt, V., Jones, D. T. W., Kool, M., Zapatka, M., Northcott, P. A., Sturm, D., Wang, W., Radlwimmer, B., Højfeldt, J. W., Truffaux, N., Castel, D., Schubert, S., Ryzhova, M., Seker-Cin, H., Gronych, J., Johann, P. D., …

Pfister, S. M. (2013). Reduced H3K27me3 and DNA hypomethylation are major drivers of gene expression in K27M mutant pediatric high-grade gliomas. *Cancer Cell*, *24*(5), 660–672. https://doi.org/10.1016/j.ccr.2013.10.006

Berger, S. L., Kouzarides, T., Shiekhattar, R., & Shilatifard, A. (2009). An operational definition of epigenetics. *Genes & Development*, *23*(7), 781–783. https://doi.org/10.1101/GAD.1787609

Bergman, D. T., Jones, T. R., Liu, V., Ray, J., Jagoda, E., Siraj, L., Kang, H. Y., Nasser, J., Kane, M., Rios, A., Nguyen, T. H., Grossman, S. R., Fulco, C. P., Lander, E. S., & Engreitz, J. M. (2022). Compatibility rules of human enhancer and promoter sequences. *Nature*, *607*(7917), Article 7917. https://doi.org/10.1038/s41586-022-04877-w

Bernstein, B. E., Mikkelsen, T. S., Xie, X., Kamal, M., Huebert, D. J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K., Jaenisch, R., Wagschal, A., Feil, R., Schreiber, S. L., & Lander, E. S. (2006). A Bivalent Chromatin Structure Marks Key Developmental Genes in Embryonic Stem Cells. *Cell*, *125*(2), 315–326. https://doi.org/10.1016/j.cell.2006.02.041

Bernstein, B. E., Stamatoyannopoulos, J. A., Costello, J. F., Ren, B., Milosavljevic, A., Meissner, A., Kellis, M., Marra, M. A., Beaudet, A. L., Ecker, J. R., Farnham, P. J., Hirst, M., Lander, E. S., Mikkelsen, T. S., & Thomson, J. A. (2010). The NIH Roadmap Epigenomics Mapping Consortium. *Nature Biotechnology*, *28*(10), 1045–1048. https://doi.org/10.1038/nbt1010-1045

Bird, A. (2007). Perceptions of epigenetics. *Nature 2007 447:7143*, *447*(7143), 396–398. https://doi.org/10.1038/nature05913

Bond, A. M., Bhalala, O. G., & Kessler, J. A. (2012). The dynamic role of bone morphogenetic proteins in neural stem cell fate and maturation. *Developmental Neurobiology*, *72*(7), 1068–1084. https://doi.org/10.1002/dneu.22022

Bonev, B., Mendelson Cohen, N., Szabo, Q., Fritsch, L., Papadopoulos, G. L., Lubling, Y., Xu, X., Lv, X., Hugnot, J.-P., Tanay, A., & Cavalli, G. (2017). Multiscale 3D Genome Rewiring during Mouse Neural Development. *Cell*, *171*(3), 557–572.e24. https://doi.org/10.1016/j.cell.2017.09.043

Borm, L. E., Mossi Albiach, A., Mannens, C. C. A., Janusauskas, J., Özgün, C., Fernández-García, D., Hodge, R., Castillo, F., Hedin, C. R. H., Villablanca, E. J., Uhlén, P., Lein, E. S., Codeluppi, S., & Linnarsson, S. (2023). Scalable in situ single-cell profiling by electrophoretic capture of mRNA using EEL FISH. *Nature Biotechnology*, *41*(2), 222–231. https://doi.org/10.1038/s41587-022-01455-3

Boyle, A. P., Davis, S., Shulha, H. P., Meltzer, P., Margulies, E. H., Weng, Z., Furey, T. S., & Crawford, G. E. (2008). High-Resolution Mapping and Characterization of Open Chromatin across the Genome. *Cell*, *132*(2), 311–311. https://doi.org/10.1016/J.CELL.2007.12.014

Brahma, S., & Henikoff, S. (2024). The BAF chromatin remodeler synergizes with RNA polymerase II and transcription factors to evict nucleosomes. *Nature Genetics*, *56*(1), 100–111. https://doi.org/10.1038/s41588-023-01603-8

Braun, E., Danan-Gotthold, M., Borm, L. E., Lee, K. W., Vinsland, E., Lönnerberg, P., Hu, L., Li, X., He, X., Andrusivova, Z., Lundeberg, J., Barker, R. A., Arenas, E., Sundström, E., & Linnarsson, S. (2023). Comprehensive cell atlas of the first-trimester developing

human brain | Science. *Science*, *382*(6667). https://doi.org/10.1126/science.adf1226

Briscoe, J., & Wilkinson, D. G. (2004). Establishing neuronal circuitry: Hox genes make the connection. *Genes & Development*, *18*(14), 1643–1648. https://doi.org/10.1101/gad.1227004

Brown, A. M., Wender, R., & Ransom, B. R. (2001). Metabolic substrates other than glucose support axon function in central white matter. *Journal of Neuroscience Research*, *66*(5), 839–843. https://doi.org/10.1002/jnr.10081

Buchanan, J., Elabbady, L., Collman, F., Jorstad, N. L., Bakken, T. E., Ott, C., Glatzer, J., Bleckert, A. A., Bodor, A. L., Brittain, D., Bumbarger, D. J., Mahalingam, G., Seshamani, S., Schneider-Mizell, C., Takeno, M. M., Torres, R., Yin, W., Hodge, R. D., Castro, M., … da Costa, N. M. (2022). Oligodendrocyte precursor cells ingest axons in the mouse neocortex. *Proceedings of the National Academy of Sciences of the United States of America*, *119*(48), e2202580119. https://doi.org/10.1073/pnas.2202580119

Buenrostro, J. D., & Greenleaf, W. J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nature Methods*, *12*. https://doi.org/10.1038/nmeth.2688

Buenrostro, J. D., Wu, B., Litzenburger, U. M., Ruff, D., Gonzales, M. L., Snyder, M. P., Chang, H. Y., & Greenleaf, W. J. (2015). Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature*, *523*(7561), 486–490. https://doi.org/10.1038/nature14590

Butler, A., Hoffman, P., Smibert, P., Papalexi, E., & Satija, R. (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nature Biotechnology*, *36*(5), 411–420. https://doi.org/10.1038/nbt.4096

Cai, J., Qi, Y., Hu, X., Tan, M., Liu, Z., Zhang, J., Li, Q., Sander, M., & Qiu, M. (2005). Generation of oligodendrocyte precursor cells from mouse dorsal spinal cord independent of Nkx6 regulation and Shh signaling. *Neuron*, *45*(1), 41–53. https://doi.org/10.1016/j.neuron.2004.12.028

Calder, V. L., Wolswijk, G., & Noble, M. (1988). The differentiation of O-2A progenitor cells into oligodendrocytes is associated with a loss of inducibility of Ia antigens. *European Journal of Immunology*, *18*(8), 1195–1201. https://doi.org/10.1002/eji.1830180808

Cao, J., Packer, J. S., Ramani, V., Cusanovich, D. A., Huynh, C., Daza, R., Qiu, X., Lee, C., Furlan, S. N., Steemers, F. J., Adey, A., Waterston, R. H., Trapnell, C., & Shendure, J. (2017). Comprehensive single-cell transcriptional profiling of a multicellular organism. *Science (New York, N.Y.)*, *357*(6352), 661–667. https://doi.org/10.1126/SCIENCE.AAM8940

Cayuso, J., & Martí, E. (2005). Morphogens in motion: Growth control of the neural tube. *Journal of Neurobiology*, *64*(4), 376–387. https://doi.org/10.1002/neu.20169

Cenik, B. K., & Shilatifard, A. (2021). COMPASS and SWI/SNF complexes in development and disease. *Nature Reviews Genetics*, *22*(1), 38–58. https://doi.org/10.1038/s41576-020-0278-0

Chen, C., Xing, D., Tan, L., Li, H., Zhou, G., Huang, L., & Xie, X. S. (2017). Single-cell whole-genome analyses by Linear Amplification via Transposon Insertion (LIANTI). *Science*, *356*(6334), 189–194. https://doi.org/10.1126/science.aak9787

Chen, K. H., Boettiger, A. N., Moffitt, J. R., Wang, S., & Zhuang, X. (2015). Spatially resolved, highly multiplexed RNA profiling in single cells. *Science*, *348*(6233), aaa6090. https://doi.org/10.1126/science.aaa6090

Choi, I.-Y., Kwon, E.-C., & Kim, N.-S. (2020). The C- and G-value paradox with polyploidy, repeatomes, introns, phenomes and cell economy. *Genes & Genomics*, *42*(7), 699–714. https://doi.org/10.1007/s13258-020-00941-9

Coleman, R. T., & Struhl, G. (2017). Causal role for inheritance of H3K27me3 in maintaining the OFF state of a Drosophila HOX gene. *Science (New York, N.Y.)*, *356*(6333), eaai8236. https://doi.org/10.1126/science.aai8236

Craig, J. M. (2005). Heterochromatin—Many flavours, common themes. *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology*, *27*(1), 17–28. https://doi.org/10.1002/bies.20145

Cui, H., Wang, C., Maan, H., Pang, K., Luo, F., Duan, N., & Wang, B. (2024). scGPT: Toward building a foundation model for single-cell multi-omics using generative AI. *Nature Methods*, 1–11. https://doi.org/10.1038/s41592-024-02201-0

Cusanovich, D. A., Daza, R., Adey, A., Pliner, H. A., Christiansen, L., Gunderson, K. L., Steemers, F. J., Trapnell, C., & Shendure, J. (2015). Multiplex single-cell profiling of chromatin accessibility by combinatorial cellular indexing. *Science*, *348*(6237), 910–914. https://doi.org/10.1126/SCIENCE.AAB1601/SUPPL_FILE/CUSANOVICH-SM.PDF

Cusanovich, D. A., Hill, A. J., Aghamirzaie, D., Daza, R. M., Pliner, H. A., Berletch, J. B., Filippova, G. N., Huang, X., Christiansen, L., DeWitt, W. S., Lee, C., Regalado, S. G., Read, D. F., Steemers, F. J., Disteche, C. M., Trapnell, C., & Shendure, J. (2018). A Single-Cell Atlas of In Vivo Mammalian Chromatin Accessibility. *Cell*, *174*(5), 1309-1324.e18. https://doi.org/10.1016/j.cell.2018.06.052

Datlinger, P., Rendeiro, A. F., Boenke, T., Senekowitsch, M., Krausgruber, T., Barreca, D., & Bock, C. (2021). Ultra-high-throughput single-cell RNA sequencing and perturbation screening with combinatorial fluidic indexing. *Nature Methods 2021 18:6*, *18*(6), 635–642. https://doi.org/10.1038/s41592-021-01153-z

de Hoon, M., Shin, J. W., & Carninci, P. (2015). Paradigm shifts in genomics through the FANTOM projects. *Mammalian Genome*, *26*(9), 391–402. https://doi.org/10.1007/s00335-015-9593-8

Deforzh, E., Uhlmann, E. J., Das, E., Galitsyna, A., Arora, R., Saravanan, H., Rabinovsky, R., Wirawan, A. D., Teplyuk, N. M., El Fatimy, R., Perumalla, S., Jairam, A., Wei, Z., Mirny, L., & Krichevsky, A. M. (2022). Promoter and enhancer RNAs regulate chromatin reorganization and activation of miR-10b/HOXD locus, and neoplastic transformation in glioma. *Molecular Cell*, *82*(10), 1894-1908.e5. https://doi.org/10.1016/j.molcel.2022.03.018

Dekker, J., Belmont, A. S., Guttman, M., Leshyk, V. O., Lis, J. T., Lomvardas, S., Mirny, L. A., O'Shea, C. C., Park, P. J., Ren, B., Politz, J. C. R., Shendure, J., & Zhong, S. (2017). The

4D nucleome project. *Nature*, *549*(7671), 219–226. https://doi.org/10.1038/nature23884

del Corral, R. D., & Storey, K. G. (2004). Opposing FGF and retinoid pathways: A signalling switch that controls differentiation and patterning onset in the extending vertebrate body axis. *BioEssays*, *26*(8), 857–869. https://doi.org/10.1002/bies.20080

del Río Hortega, P. (1922). *?` Son homologables la glía de escasas radiaciones y la célula de Schwan?* Junta para Ampliación de Estudios e Investigaciones Científicas.

Deng, W., & Roberts, S. G. E. (2005). A core promoter element downstream of the TATA box that is recognized by TFIIB. *Genes & Development*, *19*(20), 2418–2423. https://doi.org/10.1101/gad.342405

Dixon, J. R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J. S., & Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*, *485*(7398), Article 7398. https://doi.org/10.1038/nature11082

Domcke, S., Hill, A. J., Daza, R. M., Cao, J., O'Day, D. R., Pliner, H. A., Aldinger, K. A., Pokholok, D., Zhang, F., Milbank, J. H., Zager, M. A., Glass, I. A., Steemers, F. J., Doherty, D., Trapnell, C., Cusanovich, D. A., & Shendure, J. (2020). A human cell atlas of fetal chromatin accessibility. *Science*, *370*(6518), eaba7612. https://doi.org/10.1126/science.aba7612

Dunham, I., Kundaje, A., Aldred, S. F., Collins, P. J., Davis, C. A., Doyle, F., Epstein, C. B., Frietze, S., Harrow, J., Kaul, R., Khatun, J., Lajoie, B. R., Landt, S. G., Lee, B.-K., Pauli, F., Rosenbloom, K. R., Sabo, P., Safi, A., Sanyal, A., … HudsonAlpha Institute, C., UC Irvine, Stanford group (data production and analysis). (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature*, *489*(7414), 57–74. https://doi.org/10.1038/nature11247

Dykes, I. M., & Emanueli, C. (2017). Transcriptional and Post-Transcriptional Gene Regulation by Long Non-Coding RNA. *Genomics, Proteomics & Bioinformatics*, *15*(3), 177–186. https://doi.org/10.1016/j.gpb.2016.12.005

England, J. D., Gamboni, F., Levinson, S. R., & Finger, T. E. (1990). Changed distribution of sodium channels along demyelinated axons. *Proceedings of the National Academy of Sciences of the United States of America*, *87*(17), 6777–6780. https://doi.org/10.1073/pnas.87.17.6777

Ericson, J., Morton, S., Kawakami, A., Roelink, H., & Jessell, T. M. (1996). Two critical periods of Sonic Hedgehog signaling required for the specification of motor neuron identity. *Cell*, *87*(4), 661–673. https://doi.org/10.1016/s0092-8674(00)81386-0

Ernst, J., & Kellis, M. (2010). Discovery and characterization of chromatin states for systematic annotation of the human genome. *Nature Biotechnology 2010 28:8*, *28*(8), 817–825. https://doi.org/10.1038/nbt.1662

Falcão, A. M., Bruggen, D. V., Marques, S., Meijer, M., Jäkel, S., Agirre, E., Floriddia, E. M., Vanichkina, D. P., Williams, A., Guerreiro-cacais, A. O., & Castelo-branco, G. (2018). Disease-specific oligodendrocyte lineage cells arise in multiple sclerosis. *Nature Medicine*. https://doi.org/10.1038/s41591-018-0236-y

Fang, R., Xia, C., Close, J. L., Zhang, M., He, J., Huang, Z., Halpern, A. R., Long, B., Miller, J. A., Lein, E. S., & Zhuang, X. (2022). Conservation and divergence of cortical cell

organization in human and mouse revealed by MERFISH. *Science*, *377*(6601), 56–62. https://doi.org/10.1126/science.abm1741

Fang, R., Yu, M., Li, G., Chee, S., Liu, T., Schmitt, A. D., & Ren, B. (2016). Mapping of long-range chromatin interactions by proximity ligation-assisted ChIP-seq. *Cell Research*, *26*(12), 1345–1348. https://doi.org/10.1038/cr.2016.137

Ferrari, K. J., Scelfo, A., Jammula, S. G., Cuomo, A., Barozzi, I., Stützer, A., Fischle, W., Bonaldi, T., & Pasini, D. (2014). Polycomb-Dependent H3K27me1 and H3K27me2 Regulate Active Transcription and Enhancer Fidelity. *Molecular Cell*, *53*(1), 49–62. https://doi.org/10.1016/J.MOLCEL.2013.10.030/ATTACHMENT/8393BA87-CDCE-424D-BE5D-381A67DE8C56/MMC3.XLS

Floriddia, E. M., Lourenço, T., Zhang, S., van Bruggen, D., Hilscher, M. M., Kukanja, P., Gonçalves dos Santos, J. P., Altınkök, M., Yokota, C., Llorens-Bobadilla, E., Mulinyawe, S. B., Grãos, M., Sun, L. O., Frisén, J., Nilsson, M., & Castelo-Branco, G. (2020). Distinct oligodendrocyte populations have spatial preference and different responses to spinal cord injury. *Nature Communications 2020 11:1*, *11*(1), 1–15. https://doi.org/10.1038/s41467-020-19453-x

Fogarty, M., Richardson, W. D., & Kessaris, N. (2005). A subset of oligodendrocytes generated from radial glia in the dorsal spinal cord. *Development (Cambridge, England)*, *132*(8), 1951–1959. https://doi.org/10.1242/dev.01777

Frankenhaeuser, B. (1952). Saltatory conduction in myelinated nerve fibres. *The Journal of Physiology*, *118*(1), 107–112.

Franklin, R. J. M., & Ffrench-Constant, C. (2008). Remyelination in the CNS: From biology to therapy. *Nature Reviews. Neuroscience*, *9*(11), 839–855. https://doi.org/10.1038/nrn2480

Fünfschilling, U., Supplie, L. M., Mahad, D., Boretius, S., Saab, A. S., Edgar, J., Brinkmann, B. G., Kassmann, C. M., Tzvetanova, I. D., Möbius, W., Diaz, F., Meijer, D., Suter, U., Hamprecht, B., Sereda, M. W., Moraes, C. T., Frahm, J., Goebbels, S., & Nave, K. A. (2012). Glycolytic oligodendrocytes maintain myelin and long-term axonal integrity. *Nature 2012 485:7399*, *485*(7399), 517–521. https://doi.org/10.1038/nature11007

Garcia-Fernàndez, J. (2005). The genesis and evolution of homeobox gene clusters. *Nature Reviews Genetics*, *6*(12), 881–892. https://doi.org/10.1038/nrg1723

Geng, Q., Yang, R., & Zhang, L. (2022). A deep learning framework for enhancer prediction using word embedding and sequence generation. *Biophysical Chemistry*, *286*, 106822. https://doi.org/10.1016/j.bpc.2022.106822

Gibson, E. M., Purger, D., Mount, C. W., Goldstein, A. K., Lin, G. L., Wood, L. S., Inema, I., Miller, S. E., Bieri, G., Zuchero, J. B., Barres, B. A., Woo, P. J., Vogel, H., & Monje, M. (2014). Neuronal activity promotes oligodendrogenesis and adaptive myelination in the mammalian brain. *Science*, *344*(6183). https://doi.org/10.1126/SCIENCE.1252304/SUPPL_FILE/PAPV2.PDF

Gopalan, S., Wang, Y., Harper, N. W., Garber, M., & Fazzio, T. G. (2021). Simultaneous profiling of multiple chromatin proteins in the same cells. *Molecular Cell*, *81*(22), 4736-4746.e5. https://doi.org/10.1016/j.molcel.2021.09.019

Goryshin, I. Y., & Reznikoff, W. S. (1998). Tn5 in vitro transposition. *The Journal of Biological Chemistry*, *273*(13), 7367–7374. https://doi.org/10.1074/JBC.273.13.7367

Göttlicher, M., Minucci, S., Zhu, P., Krämer, O. H., Schimpf, A., Giavara, S., Sleeman, J. P., Lo Coco, F., Nervi, C., Pelicci, P. G., & Heinzel, T. (2001). Valproic acid defines a novel class of HDAC inhibitors inducing differentiation of transformed cells. *The EMBO Journal*, *20*(24), 6969–6978. https://doi.org/10.1093/emboj/20.24.6969

Gräff, J., & Tsai, L. H. (2013). Histone acetylation: Molecular mnemonics on the chromatin. *Nature Reviews Neuroscience 2013 14:2*, *14*(2), 97–111. https://doi.org/10.1038/nrn3427

Granja, J. M., Corces, M. R., Pierce, S. E., Bagdatli, S. T., Choudhry, H., Chang, H. Y., & Greenleaf, W. J. (2021). ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. *Nature Genetics*, *53*(3), Article 3. https://doi.org/10.1038/s41588-021-00790-6

GTEx Consortium. (2020). The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science (New York, N.Y.)*, *369*(6509), 1318–1330. https://doi.org/10.1126/science.aaz1776

Hahn, M., & Wray, G. (2002). The g-value paradox. *Evolution & Development*, *4*, 73–75. https://doi.org/10.1046/j.1525-142X.2002.01069.x

Hainer, S. J., Bošković, A., McCannell, K. N., Rando, O. J., & Fazzio, T. G. (2019). Profiling of Pluripotency Factors in Single Cells and Early Embryos. *Cell*, *177*(5), 1319-1329.e11. https://doi.org/10.1016/j.cell.2019.03.014

Hakimi, M.-A., Bochar, D. A., Schmiesing, J. A., Dong, Y., Barak, O. G., Speicher, D. W., Yokomori, K., & Shiekhattar, R. (2002). A chromatin remodelling complex that loads cohesin onto human chromosomes. *Nature*, *418*(6901), 994–998. https://doi.org/10.1038/nature01024

Hesselberth, J. R., Chen, X., Zhang, Z., Sabo, P. J., Sandstrom, R., Reynolds, A. P., Thurman, R. E., Neph, S., Kuehn, M. S., Noble, W. S., Fields, S., & Stamatoyannopoulos, J. A. (2009). Global mapping of protein-DNA interactions in vivo by digital genomic footprinting. *Nature Methods 2009 6:4*, *6*(4), 283–289. https://doi.org/10.1038/nmeth.1313

Hie, B., Bryson, B., & Berger, B. (2019). Efficient integration of heterogeneous single-cell transcriptomes using Scanorama. *Nature Biotechnology*, *37*(6), 685–691. https://doi.org/10.1038/s41587-019-0113-3

Hochedlinger, K., & Plath, K. (2009). Epigenetic reprogramming and induced pluripotency. *Development (Cambridge, England)*, *136*(4), 509–523. https://doi.org/10.1242/DEV.020867

Hou, W., & Ji, Z. (2024). Assessing GPT-4 for cell type annotation in single-cell RNA-seq analysis. *Nature Methods*, 1–4. https://doi.org/10.1038/s41592-024-02235-4

Hsieh, T.-H. S., Weiner, A., Lajoie, B., Dekker, J., Friedman, N., & Rando, O. J. (2015). Mapping Nucleosome Resolution Chromosome Folding in Yeast by Micro-C. *Cell*, *162*(1), 108–119. https://doi.org/10.1016/j.cell.2015.05.048

Hutlet, B., Theys, N., Coste, C., Ahn, M.-T., Doshishti-Agolli, K., Lizen, B., & Gofflot, F. (2016). Systematic expression analysis of Hox genes at adulthood reveals novel patterns

in the central nervous system. *Brain Structure and Function*, *221*(3), 1223–1243. https://doi.org/10.1007/s00429-014-0965-8

International Multiple Sclerosis Genetics Consortium. (2019). Multiple sclerosis genomic map implicates peripheral immune cells and microglia in susceptibility. *Science (New York, N.Y.)*, *365*(6460), eaav7188. https://doi.org/10.1126/science.aav7188

Ishikawa, Y., Yamamoto, N., Yoshimoto, M., & Ito, H. (2012). The Primary Brain Vesicles Revisited: Are the Three Primary Vesicles (Forebrain/Midbrain/Hindbrain) Universal in Vertebrates. *Brain Behavior and Evolution*, *79*(2), 75–83. https://doi.org/10.1159/000334842

Iwafuchi-Doi, M., & Zaret, K. S. (2014). Pioneer transcription factors in cell reprogramming. *Genes & Development*, *28*(24), 2679–2692. https://doi.org/10.1101/gad.253443.114

Jain, S. U., Do, T. J., Lund, P. J., Rashoff, A. Q., Diehl, K. L., Cieslik, M., Bajic, A., Juretic, N., Deshmukh, S., Venneti, S., Muir, T. W., Garcia, B. A., Jabado, N., & Lewis, P. W. (2019). PFA ependymoma-associated protein EZHIP inhibits PRC2 activity through a H3 K27M-like mechanism. *Nature Communications*, *10*(1), 2146. https://doi.org/10.1038/s41467-019-09981-6

Jain, S. U., Rashoff, A. Q., Krabbenhoft, S. D., Hoelper, D., Do, T. J., Gibson, T. J., Lundgren, S. M., Bondra, E. R., Deshmukh, S., Harutyunyan, A. S., Juretic, N., Jabado, N., Harrison, M. M., & Lewis, P. W. (2020). H3 K27M and EZHIP Impede H3K27-Methylation Spreading by Inhibiting Allosterically Stimulated PRC2. *Molecular Cell*, *80*(4), 726-735.e7. https://doi.org/10.1016/j.molcel.2020.09.028

Jäkel, S., Agirre, E., Mendanha Falcão, A., van Bruggen, D., Lee, K. W., Knuesel, I., Malhotra, D., ffrench-Constant, C., Williams, A., & Castelo-Branco, G. (2019). Altered human oligodendrocyte heterogeneity in multiple sclerosis. *Nature 2019 566:7745*, *566*(7745), 543–547. https://doi.org/10.1038/s41586-019-0903-2

Jakovcevski, I., & Zecevic, N. (2005). Sequence of oligodendrocyte development in the human fetal telencephalon. *Glia*, *49*(4), 480–491. https://doi.org/10.1002/GLIA.20134

Javierre, B. M., Burren, O. S., Wilder, S. P., Kreuzhuber, R., Hill, S. M., Sewitz, S., Cairns, J., Wingett, S. W., Várnai, C., Thiecke, M. J., Burden, F., Farrow, S., Cutler, A. J., Rehnström, K., Downes, K., Grassi, L., Kostadima, M., Freire-Pritchett, P., Wang, F., … Fraser, P. (2016). Lineage-Specific Genome Architecture Links Enhancers and Non-coding Disease Variants to Target Gene Promoters. *Cell*, *167*(5), 1369-1384.e19. https://doi.org/10.1016/j.cell.2016.09.037

Jessa, S., Mohammadnia, A., Harutyunyan, A. S., Hulswit, M., Varadharajan, S., Lakkis, H., Kabir, N., Bashardanesh, Z., Hébert, S., Faury, D., Vladoiu, M. C., Worme, S., Coutelier, M., Krug, B., Faria Andrade, A., Pathania, M., Bajic, A., Weil, A. G., Ellezam, B., … Kleinman, C. L. (2022). K27M in canonical and noncanonical H3 variants occurs in distinct oligodendroglial cell lineages in brain midline gliomas. *Nature Genetics*, *54*(12), 1865–1880. https://doi.org/10.1038/s41588-022-01205-w

Johnson, D. S., Mortazavi, A., Myers, R. M., & Wold, B. (2007). Genome-wide mapping of in vivo protein-DNA interactions. *Science*, *316*(5830), 1497–1502. https://doi.org/10.1126/SCIENCE.1141319/SUPPL_FILE/JOHNSON.SOM-5-30.PDF

Johnson, R. C., & Reznikoff, W. S. (1983). DNA sequences at the ends of transposon Tn5 required for transposition. *Nature*, *304*(5923), 280–282. https://doi.org/10.1038/304280A0

Johnson, W. E., Li, C., & Rabinovic, A. (2007). Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics*, *8*(1), 118–127. https://doi.org/10.1093/biostatistics/kxj037

Kandel, E. R. (2013). *Principles of Neural Science, Fifth Edition*. McGraw Hill Professional.

Kaplan, N., Moore, I. K., Fondufe-Mittendorf, Y., Gossett, A. J., Tillo, D., Field, Y., LeProust, E. M., Hughes, T. R., Lieb, J. D., Widom, J., & Segal, E. (2008). The DNA-encoded nucleosome organization of a eukaryotic genome. *Nature 2008 458:7236*, *458*(7236), 362–366. https://doi.org/10.1038/nature07667

Kaya-Okur, H. S., Wu, S. J., Codomo, C. A., Pledger, E. S., Bryson, T. D., Henikoff, J. G., Ahmad, K., & Henikoff, S. (2019). CUT&Tag for efficient epigenomic profiling of small samples and single cells. *Nature Communications 2019 10:1*, *10*(1), 1–10. https://doi.org/10.1038/s41467-019-09982-5

Ke, R., Mignardi, M., Pacureanu, A., Svedlund, J., Botling, J., Wählby, C., & Nilsson, M. (2013). In situ sequencing for RNA analysis in preserved tissue and cells. *Nature Methods*, *10*(9), 857–860. https://doi.org/10.1038/nmeth.2563

Keene, M. A., Corces, V., Lowenhaupt, K., & Elgin, S. C. (1981). DNase I hypersensitive sites in Drosophila chromatin occur at the 5′ ends of regions of transcription. *Proceedings of the National Academy of Sciences*, *78*(1), 143–146. https://doi.org/10.1073/pnas.78.1.143

Kessaris, N., Fogarty, M., Iannarelli, P., Grist, M., Wegner, M., & Richardson, W. D. (2006). Competing waves of oligodendrocytes in the forebrain and postnatal elimination of an embryonic lineage. *Nature Neuroscience*, *9*(2), 173–179. https://doi.org/10.1038/nn1620

Kim, Y., Shi, Z., Zhang, H., Finkelstein, I. J., & Yu, H. (2019). Human cohesin compacts DNA by loop extrusion. *Science*, *366*(6471), 1345–1349. https://doi.org/10.1126/science.aaz4475

Kirby, L., & Castelo-Branco, G. (2021). Crossing boundaries: Interplay between the immune system and oligodendrocyte lineage cells. *Seminars in Cell & Developmental Biology*, *116*, 45–52. https://doi.org/10.1016/j.semcdb.2020.10.013

Kirby, L., Jin, J., Cardona, J. G., Smith, M. D., Martin, K. A., Wang, J., Strasburger, H., Herbst, L., Alexis, M., Karnell, J., Davidson, T., Dutta, R., Goverman, J., Bergles, D., & Calabresi, P. A. (2019). Oligodendrocyte precursor cells present antigen and are cytotoxic targets in inflammatory demyelination. *Nature Communications*, *10*(1), 1–20. https://doi.org/10.1038/s41467-019-11638-3

Klemm, S. L., Shipony, Z., & Greenleaf, W. J. (2019). Chromatin accessibility and the regulatory epigenome. *Nature Reviews Genetics 2018 20:4*, *20*(4), 207–220. https://doi.org/10.1038/s41576-018-0089-8

Kornberg, R. D., & Thomas, J. O. (1974). Chromatin structure: Oligomers of the histones. *Science*, *184*(4139), 865–868. https://doi.org/10.1126/SCIENCE.184.4139.865/ASSET/1613F602-AE4C-4658-B3BF-F78C936D31B1/ASSETS/SCIENCE.184.4139.865.FP.PNG

Korsunsky, I., Millard, N., Fan, J., Slowikowski, K., Zhang, F., Wei, K., Baglaenko, Y., Brenner, M., Loh, P., & Raychaudhuri, S. (2019). Fast, sensitive and accurate integration of single-cell data with Harmony. *Nature Methods*, *16*(12), Article 12. https://doi.org/10.1038/s41592-019-0619-0

Kouzarides, T. (2007). Chromatin modifications and their function. *Cell*, *128*(4), 693–705. https://doi.org/10.1016/j.cell.2007.02.005

Krebs, A. R., Imanci, D., Hoerner, L., Gaidatzis, D., Burger, L., & Schübeler, D. (2017). Genome-wide Single-Molecule Footprinting Reveals High RNA Polymerase II Turnover at Paused Promoters. *Molecular Cell*, *67*(3), 411-422.e4. https://doi.org/10.1016/j.molcel.2017.06.027

Krietenstein, N., Abraham, S., Venev, S. V., Abdennur, N., Gibcus, J., Hsieh, T.-H. S., Parsi, K. M., Yang, L., Maehr, R., Mirny, L. A., Dekker, J., & Rando, O. J. (2020). Ultrastructural Details of Mammalian Chromosome Architecture. *Molecular Cell*, *78*(3), 554-565.e7. https://doi.org/10.1016/j.molcel.2020.03.003

Krumlauf, R. (1994). *Hox* genes in vertebrate development. *Cell*, *78*(2), 191–201. https://doi.org/10.1016/0092-8674(94)90290-9

Lettice, L. A., Heaney, S. J. H., Purdie, L. A., Li, L., de Beer, P., Oostra, B. A., Goode, D., Elgar, G., Hill, R. E., & de Graaff, E. (2003). A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Human Molecular Genetics*, *12*(14), 1725–1735. https://doi.org/10.1093/hmg/ddg180

Levy, S., Somasundaram, L., Raj, I. X., Ic-Mex, D., Phal, A., Schmidt, S., Ng, W. I., Mar, D., Decarreau, J., Moss, N., Alghadeer, A., Honkanen, H., Sarthy, J., Vitanza, N., Hawkins, R. D., Mathieu, J., Wang, Y., Baker, D., Bomsztyk, K., & Ruohola-Baker, H. (2022). dCas9 fusion to computer-designed PRC2 inhibitor reveals functional TATA box in distal promoter region. *Cell Reports*, *38*(9), 110457. https://doi.org/10.1016/j.celrep.2022.110457

Lewis, E. B. (1978). A gene complex controlling segmentation in Drosophila. *Nature*, *276*(5688), 565–570. https://doi.org/10.1038/276565a0

Lewis, P. W., Müller, M. M., Koletsky, M. S., Cordero, F., Lin, S., Banaszynski, L. A., Garcia, B. A., Muir, T. W., Becher, O. J., & Allis, C. D. (2013). Inhibition of PRC2 activity by a gain-of-function H3 mutation found in pediatric glioblastoma. *Science (New York, N.Y.)*, *340*(6134), 857–861. https://doi.org/10.1126/science.1232245

Lieberman-Aiden, E., van Berkum, N. L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B. R., Sabo, P. J., Dorschner, M. O., Sandstrom, R., Bernstein, B., Bender, M. A., Groudine, M., Gnirke, A., Stamatoyannopoulos, J., Mirny, L. A., Lander, E. S., & Dekker, J. (2009). Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. *Science*, *326*(5950), 289–293. https://doi.org/10.1126/science.1181369

Linnarsson, S., & Teichmann, S. A. (2016). Single-cell genomics: Coming of age. *Genome Biology*, *17*(1), 1–3. https://doi.org/10.1186/S13059-016-0960-X/METRICS

Lobanenkov, V. V., Nicolas, R. H., Adler, V. V., Paterson, H., Klenova, E. M., Polotskaja, A. V., & Goodwin, G. H. (1990). A novel sequence-specific DNA binding protein which

interacts with three regularly spaced direct repeats of the CCCTC-motif in the 5′-flanking sequence of the chicken c-myc gene. *Oncogene*, *5*(12), 1743–1753.

Luecken, M. D., Büttner, M., Chaichoompu, K., Danese, A., Interlandi, M., Mueller, M. F., Strobl, D. C., Zappia, L., Dugas, M., Colomé-Tatché, M., & Theis, F. J. (2022). Benchmarking atlas-level data integration in single-cell genomics. *Nature Methods*, *19*(1), 41–50. https://doi.org/10.1038/s41592-021-01336-8

Lumsden, A. (2004). Segmentation and compartition in the early avian hindbrain. *Mechanisms of Development*, *121*(9), 1081–1088. https://doi.org/10.1016/j.mod.2004.04.018

Macosko, E. Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, A. R., Kamitaki, N., Martersteck, E. M., Trombetta, J. J., Weitz, D. A., Sanes, J. R., Shalek, A. K., Regev, A., & McCarroll, S. A. (2015). Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell*, *161*(5), 1202–1214. https://doi.org/10.1016/j.cell.2015.05.002

Mannens, C. C. A., Hu, L., Lönnerberg, P., Schipper, M., Reagor, C., Li, X., He, X., Barker, R. A., Sundström, E., Posthuma, D., & Linnarsson, S. (2023). *Dynamics of chromatin accessibility during human first-trimester neurodevelopment* (p. 2023.08.18.553878). bioRxiv. https://doi.org/10.1101/2023.08.18.553878

Marco, A., Meharena, H. S., Dileep, V., Raju, R. M., Davila-Velderrain, J., Zhang, A. L., Adaikkan, C., Young, J. Z., Gao, F., Kellis, M., & Tsai, L.-H. (2020). Mapping the epigenomic and transcriptomic interplay during memory formation and recall in the hippocampal engram ensemble. *Nature Neuroscience*, *23*(12), 1606–1617. https://doi.org/10.1038/s41593-020-00717-0

Marisca, R., Hoche, T., Agirre, E., Hoodless, L. J., Barkey, W., Auer, F., Castelo-Branco, G., & Czopka, T. (2020). Functionally distinct subgroups of oligodendrocyte precursor cells integrate neural activity and execute myelin formation. *Nature Neuroscience*, *23*(3), 363–374. https://doi.org/10.1038/S41593-019-0581-2

Markenscoff-Papadimitriou, E., Whalen, S., Przytycki, P., Thomas, R., Binyameen, F., Nowakowski, T. J., Kriegstein, A. R., Sanders, S. J., State, M. W., Pollard, K. S., & Rubenstein, J. L. (2020). A Chromatin Accessibility Atlas of the Developing Human Telencephalon. *Cell*, *182*(3), 754-769.e18. https://doi.org/10.1016/j.cell.2020.06.002

Marques, S., Bruggen, D. van, Vanichkina, D. P., Floriddia, E. M., Munguba, H., Väremo, L., Giacomello, S., Falcão, A. M., Meijer, M., Björklund, Å. K., Hjerling-Leffler, J., Taft, R. J., & Castelo-Branco, G. (2018). Transcriptional Convergence of Oligodendrocyte Lineage Progenitors during Development. *Developmental Cell*, *46*(4), 504-517.e7. https://doi.org/10.1016/j.devcel.2018.07.005

Marques, S., Zeisel, A., Codeluppi, S., Van Bruggen, D., Mendanha Falcão, A., Xiao, L., Li, H., Häring, M., Hochgerner, H., Romanov, R. A., Gyllborg, D., Muñoz Manchado, A., La Manno, G., Lönnerberg, P., Floriddia, E. M., Rezayee, F., Ernfors, P., Arenas, E., Hjerling-leffler, J., … Castelo-Branco, G. (2016). Oligodendrocyte heterogeneity in the mouse juvenile and adult central nervous system. *Science*, *352*(6291), 1326–1329. https://doi.org/10.1126/science.aaf6463

Marsh, S. E., Walker, A. J., Kamath, T., Dissing-Olesen, L., Hammond, T. R., de Soysa, T. Y., Young, A. M. H., Murphy, S., Abdulraouf, A., Nadaf, N., Dufort, C., Walker, A. C., Lucca,

L. E., Kozareva, V., Vanderburg, C., Hong, S., Bulstrode, H., Hutchinson, P. J., Gaffney, D. J., … Stevens, B. (2022). Dissection of artifactual and confounding glial signatures by single-cell sequencing of mouse and human brain. *Nature Neuroscience*, *25*(3), 306–316. https://doi.org/10.1038/s41593-022-01022-8

McGhee, J. D., Wood, W. I., Dolan, M., Engel, J. D., & Felsenfeld, G. (1981). A 200 base pair region at the 5′ end of the chicken adult β-globin gene is accessible to nuclease digestion. *Cell*, *27*(1), 45–55. https://doi.org/10.1016/0092-8674(81)90359-7

Meers, M. P., Janssens, D. H., & Henikoff, S. (2019). Pioneer Factor-Nucleosome Binding Events during Differentiation Are Motif Encoded. *Molecular Cell*, *75*(3), 562-575.e5. https://doi.org/10.1016/j.molcel.2019.05.025

Meijer, M., Agirre, E., Kabbe, M., van Tuijn, C. A., Heskol, A., Zheng, C., Mendanha Falcão, A., Bartosovic, M., Kirby, L., Calini, D., Johnson, M. R., Corces, M. R., Montine, T. J., Chen, X., Chang, H. Y., Malhotra, D., & Castelo-Branco, G. (2022). Epigenomic priming of immune genes implicates oligodendroglia in multiple sclerosis susceptibility. *Neuron*. https://doi.org/10.1016/J.NEURON.2021.12.034

Miller, D. H., & Leary, S. M. (2007). Primary-progressive multiple sclerosis. *The Lancet Neurology*, *6*(10), 903–912.

Minnoye, L., Taskiran, I. I., Mauduit, D., Fazio, M., Van Aerschot, L., Hulselmans, G., Christiaens, V., Makhzami, S., Seltenhammer, M., Karras, P., Primot, A., Cadieu, E., van Rooijen, E., Marine, J.-C., Egidy, G., Ghanem, G.-E., Zon, L., Wouters, J., & Aerts, S. (2020). Cross-species analysis of enhancer logic using deep learning. *Genome Research*, *30*(12), 1815–1834. https://doi.org/10.1101/gr.260844.120

Mo, Z., & Zecevic, N. (2009). Human fetal Radial glia cells generate oligodendrocytes in vitro. *Glia*, *57*(5), 490–490. https://doi.org/10.1002/GLIA.20775

Moon, K. R., Stanley, J. S., Burkhardt, D., van Dijk, D., Wolf, G., & Krishnaswamy, S. (2018). Manifold learning-based methods for analyzing single-cell RNA-sequencing data. *Current Opinion in Systems Biology*, *7*, 36–46. https://doi.org/10.1016/j.coisb.2017.12.008

Mulqueen, R. M., Pokholok, D., O'Connell, B. L., Thornton, C. A., Zhang, F., O'Roak, B. J., Link, J., Yardımcı, G. G., Sears, R. C., Steemers, F. J., & Adey, A. C. (2021). High-content single-cell combinatorial indexing. *Nature Biotechnology*, *39*(12), 1574–1580. https://doi.org/10.1038/S41587-021-00962-Z

Mumbach, M. R., Rubin, A. J., Flynn, R. A., Dai, C., Khavari, P. A., Greenleaf, W. J., & Chang, H. Y. (2016). HiChIP: Efficient and sensitive analysis of protein-directed genome architecture. *Nature Methods*, *13*(11), 919–922. https://doi.org/10.1038/nmeth.3999

Narendra, V., Bulajić, M., Dekker, J., Mazzoni, E. O., & Reinberg, D. (2016). CTCF-mediated topological boundaries during development foster appropriate gene regulation. *Genes & Development*, *30*(24), 2657–2662. https://doi.org/10.1101/gad.288324.116

Noordermeer, D., Leleu, M., Splinter, E., Rougemont, J., De Laat, W., & Duboule, D. (2011). The dynamic architecture of Hox gene clusters. *Science (New York, N.Y.)*, *334*(6053), 222–225. https://doi.org/10.1126/science.1207194

Olins, A. L., & Olins, D. E. (1974). Spheroid Chromatin Units (ν Bodies). *Science*, *183*(4122), 330–332. https://doi.org/10.1126/SCIENCE.183.4122.330

Ong, C.-T., & Corces, V. G. (2014). CTCF: An architectural protein bridging genome topology and function. *Nature Reviews Genetics*, *15*(4), 234–246. https://doi.org/10.1038/nrg3663

Patty, B. J., & Hainer, S. J. (2021). Transcription factor chromatin profiling genome-wide using uliCUT&RUN in single cells and individual blastocysts. *Nature Protocols*, *16*(5), Article 5. https://doi.org/10.1038/s41596-021-00516-2

Pérez-Cerdá, F., Sánchez-Gómez, M. V., & Matute, C. (2015). Pío del Río Hortega and the discovery of the oligodendrocytes. *Frontiers in Neuroanatomy*, *9*. https://doi.org/10.3389/fnana.2015.00092

Philippidou, P., & Dasen, J. S. (2013). Hox genes: Choreographers in neural development, architects of circuit organization. *Neuron*, *80*(1), 12–34. https://doi.org/10.1016/j.neuron.2013.09.020

Phillips, J. E., & Corces, V. G. (2009). CTCF: Master weaver of the genome. *Cell*, *137*(7), 1194–1211. https://doi.org/10.1016/j.cell.2009.06.001

Picelli, S., Björklund, Å. K., Reinius, B., Sagasser, S., Winberg, G., & Sandberg, R. (2014). Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. *Genome Research*, *24*(12), 2033–2040. https://doi.org/10.1101/GR.177881.114

Piecyk, R. S., Schlegel, L., & Johannes, F. (2022). Predicting 3D chromatin interactions from DNA sequence using Deep Learning. *Computational and Structural Biotechnology Journal*, *20*, 3439–3448. https://doi.org/10.1016/j.csbj.2022.06.047

Pleiner, T., Bates, M., & Görlich, D. (2017). A toolbox of anti–mouse and anti–rabbit IgG secondary nanobodies. *Journal of Cell Biology*, *217*(3), 1143–1154. https://doi.org/10.1083/jcb.201709115

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., Adey, A. C., Steemers, F. J., Shendure, J., & Trapnell, C. (2018). Cicero Predicts cis-Regulatory DNA Interactions from Single-Cell Chromatin Accessibility Data. *Molecular Cell*, *71*(5), 858-871.e8. https://doi.org/10.1016/J.MOLCEL.2018.06.044

Poirier, M. G., Bussiek, M., Langowski, J., & Widom, J. (2008). Spontaneous Access to DNA Target Sites in Folded Chromatin Fibers. *Journal of Molecular Biology*, *379*(4), 772–786. https://doi.org/10.1016/j.jmb.2008.04.025

Pringle, N. P., Yu, W. P., Guthrie, S., Roelink, H., Lumsden, A., Peterson, A. C., & Richardson, W. D. (1996). Determination of neuroepithelial cell fate: Induction of the oligodendrocyte lineage by ventral midline cells and sonic hedgehog. *Developmental Biology*, *177*(1), 30–42. https://doi.org/10.1006/dbio.1996.0142

Rakic, S., & Zecevic, N. (2003). Early oligodendrocyte progenitor cells in the human fetal telencephalon. *Glia*, *41*(2), 117–127. https://doi.org/10.1002/GLIA.10140

Ramagopalan, S. V., Dobson, R., Meier, U. C., & Giovannoni, G. (2010). Multiple sclerosis: Risk factors, prodromes, and potential causal pathways. *The Lancet Neurology*, *9*(7), 727–739. https://doi.org/10.1016/S1474-4422(10)70094-6

Rayon, T., Maizels, R. J., Barrington, C., & Briscoe, J. (2021). Single-cell transcriptome profiling of the human developing spinal cord reveals a conserved genetic

programme with human-specific features. *Development (Cambridge, England)*, *148*(15). https://doi.org/10.1242/DEV.199711

Rebeiz, M., & Tsiantis, M. (2017). Enhancer evolution and the origins of morphological novelty. *Current Opinion in Genetics & Development*, *45*, 115–123. https://doi.org/10.1016/j.gde.2017.04.006

Rekaik, H., Lopez-Delisle, L., Hintermann, A., Mascrez, B., Bochaton, C., Mayran, A., & Duboule, D. (2023). Sequential and directional insulation by conserved CTCF sites underlies the Hox timer in stembryos. *Nature Genetics*, *55*(7), Article 7. https://doi.org/10.1038/s41588-023-01426-7

Reznikoff, W. S. (2003). Tn5 as a model for understanding DNA transposition. *Molecular Microbiology*, *47*(5), 1199–1206. https://doi.org/10.1046/J.1365-2958.2003.03382.X

Richardson, W. D., Kessaris, N., & Pringle, N. (2006). Oligodendrocyte wars. *Nature Reviews Neuroscience*, *7*(1), 11–18. https://doi.org/10.1038/nrn1826

Rinn, J. L., Bondre, C., Gladstone, H. B., Brown, P. O., & Chang, H. Y. (2006). Anatomic demarcation by positional variation in fibroblast gene expression programs. *PLoS Genetics*, *2*(7), e119. https://doi.org/10.1371/journal.pgen.0020119

Rinn, J. L., Kertesz, M., Wang, J. K., Squazzo, S. L., Xu, X., Brugmann, S. A., Goodnough, L. H., Helms, J. A., Farnham, P. J., Segal, E., & Chang, H. Y. (2007). Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell*, *129*(7), 1311–1323. https://doi.org/10.1016/j.cell.2007.05.022

Rowley, M. J., & Corces, V. G. (2018). Organizational principles of 3D genome architecture. *Nature Reviews Genetics*, *19*(12), Article 12. https://doi.org/10.1038/s41576-018-0060-8

Santos-Rosa, H., Schneider, R., Bannister, A. J., Sherriff, J., Bernstein, B. E., Emre, N. C. T., Schreiber, S. L., Mellor, J., & Kouzarides, T. (2002). Active genes are tri-methylated at K4 of histone H3. *Nature*, *419*(6905), 407–411. https://doi.org/10.1038/NATURE01080

Satpathy, A. T., Granja, J. M., Yost, K. E., Qi, Y., Meschi, F., McDermott, G. P., Olsen, B. N., Mumbach, M. R., Pierce, S. E., Corces, M. R., Shah, P., Bell, J. C., Jhutty, D., Nemec, C. M., Wang, J., Wang, L., Yin, Y., Giresi, P. G., Chang, A. L. S., … Chang, H. Y. (2019). Massively parallel single-cell chromatin landscapes of human immune cell development and intratumoral T cell exhaustion. *Nature Biotechnology*, *37*(8), 925–936. https://doi.org/10.1038/s41587-019-0206-z

Schep, A. N., Wu, B., Buenrostro, J. D., & Greenleaf, W. J. (2017). chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. *Nature Methods 2017 14:10*, *14*(10), 975–978. https://doi.org/10.1038/nmeth.4401

Schoenfelder, S., & Fraser, P. (2019). Long-range enhancer–promoter contacts in gene expression control. *Nature Reviews Genetics 2019 20:8*, *20*(8), 437–455. https://doi.org/10.1038/s41576-019-0128-0

Schones, D. E., Cui, K., Cuddapah, S., Roh, T. Y., Barski, A., Wang, Z., Wei, G., & Zhao, K. (2008). Dynamic regulation of nucleosome positioning in the human genome. *Cell*, *132*(5), 887–898. https://doi.org/10.1016/J.CELL.2008.02.022

Seeker, L. A., Bestard-Cuche, N., Jäkel, S., Kazakou, N.-L., Bøstrand, S. M. K., Wagstaff, L. J., Cholewa-Waclaw, J., Kilpatrick, A. M., Van Bruggen, D., Kabbe, M., Baldivia Pohl, F., Moslehi, Z., Henderson, N. C., Vallejos, C. A., La Manno, G., Castelo-Branco, G., & Williams, A. (2023). Brain matters: Unveiling the distinct contributions of region, age, and sex to glia diversity and CNS function. *Acta Neuropathologica Communications*, *11*(1), 84. https://doi.org/10.1186/s40478-023-01568-z

Shen, Y., Yue, F., McCleary, D. F., Ye, Z., Edsall, L., Kuan, S., Wagner, U., Dixon, J., Lee, L., Lobanenkov, V. V., & Ren, B. (2012). A map of the cis-regulatory sequences in the mouse genome. *Nature*, *488*(7409), 116–120. https://doi.org/10.1038/nature11243

Sherman, D. L., & Brophy, P. J. (2005). Mechanisms of axon ensheathment and myelin growth. *Nature Reviews Neuroscience*, *6*(9), 683–690. https://doi.org/10.1038/nrn1743

Shlyueva, D., Stampfel, G., & Stark, A. (2014). Transcriptional enhancers: From properties to genome-wide predictions. *Nature Reviews Genetics 2014 15:4*, *15*(4), 272–286. https://doi.org/10.1038/nrg3682

Siletti, K., Hodge, R., Mossi Albiach, A., Lee, K. W., Ding, S.-L., Hu, L., Lönnerberg, P., Bakken, T., Casper, T., Clark, M., Dee, N., Gloe, J., Hirschstein, D., Shapovalova, N. V., Keene, C. D., Nyhus, J., Tung, H., Yanny, A. M., Arenas, E., … Linnarsson, S. (2023). Transcriptomic diversity of cell types across the adult human brain. *Science (New York, N.Y.)*, *382*(6667), eadd7046. https://doi.org/10.1126/science.add7046

Skene, P. J., & Henikoff, S. (2017). An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *eLife*, *6*. https://doi.org/10.7554/ELIFE.21856

Smith, K. J., Bostock, H., & Hall, S. M. (1982). Saltatory conduction precedes remyelination in axons demyelinated with lysophosphatidyl choline. *Journal of the Neurological Sciences*, *54*(1), 13–31. https://doi.org/10.1016/0022-510x(82)90215-5

Soshnikova, N., & Duboule, D. (2009). Epigenetic Temporal Control of Mouse Hox Genes in Vivo. *Science*, *324*(5932), 1320–1323. https://doi.org/10.1126/science.1171468

Soufi, A., Garcia, M. F., Jaroszewicz, A., Osman, N., Pellegrini, M., & Zaret, K. S. (2015). Pioneer transcription factors target partial DNA motifs on nucleosomes to initiate reprogramming. *Cell*, *161*(3), 555–568.

Sparbier, C. E., Gillespie, A., Gomez, J., Kumari, N., Motazedian, A., Chan, K. L., Bell, C. C., Gilan, O., Chan, Y.-C., Popp, S., Gough, D. J., Eckersley-Maslin, M. A., Dawson, S.-J., Lehner, P. J., Sutherland, K. D., Ernst, P., McGeehan, G. M., Lam, E. Y. N., Burr, M. L., & Dawson, M. A. (2023). Targeting Menin disrupts the KMT2A/B and polycomb balance to paradoxically activate bivalent genes. *Nature Cell Biology*, *25*(2), 258–272. https://doi.org/10.1038/s41556-022-01056-x

Spitzer, S. O., Sitnikov, S., Kamen, Y., Evans, K. A., Kronenberg-Versteeg, D., Dietmann, S., de Faria, O., Agathou, S., & Káradóttir, R. T. (2019). Oligodendrocyte Progenitor Cells Become Regionally Diverse and Heterogeneous with Age. *Neuron*, *101*(3), 459-471.e5. https://doi.org/10.1016/J.NEURON.2018.12.020

Stämpfli, R. (1954). Saltatory Conduction in Nerve. *Physiological Reviews*, *34*(1), 101–112. https://doi.org/10.1152/physrev.1954.34.1.101

Strahl, B. D., & Allis, C. D. (2000). The language of covalent histone modifications. *Nature*, *403*(6765), 41–45. https://doi.org/10.1038/47412

Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W. M., Hao, Y., Stoeckius, M., Smibert, P., & Satija, R. (2019). Comprehensive Integration of Single-Cell Data. *Cell*, *177*(7), 1888-1902.e21. https://doi.org/10.1016/j.cell.2019.05.031

Sun, Y., Xu, X., Zhao, W., Zhang, Y., Chen, K., Li, Y., Wang, X., Zhang, M., Xue, B., Yu, W., Hou, Y., Wang, C., Xie, W., Li, C., Kong, D., Wang, S., & Sun, Y. (2023). RAD21 is the core subunit of the cohesin complex involved in directing genome organization. *Genome Biology*, *24*(1), 155. https://doi.org/10.1186/s13059-023-02982-1

Suzumura, A., Silberberg, D. H., & Lisak, R. P. (1986). The expression of MHC antigens on oligodendrocytes: Induction of polymorphic H-2 expression by lymphokines. *Journal of Neuroimmunology*, *11*(3), 179–190. https://doi.org/10.1016/0165-5728(86)90002-0

Svensson, V., Vento-Tormo, R., & Teichmann, S. A. (2018). Exponential scaling of single-cell RNA-seq in the past decade. *Nature Protocols*, *13*(4), 599–604. https://doi.org/10.1038/nprot.2017.149

Symmons, O., Pan, L., Remeseiro, S., Aktas, T., Klein, F., Huber, W., & Spitz, F. (2016). The Shh Topological Domain Facilitates the Action of Remote Enhancers by Reducing the Effects of Genomic Distances. *Developmental Cell*, *39*(5), 529–543. https://doi.org/10.1016/J.DEVCEL.2016.10.015

Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., & Yamanaka, S. (2007). Induction of Pluripotent Stem Cells from Adult Human Fibroblasts by Defined Factors. *Cell*, *131*(5), 861–872. https://doi.org/10.1016/J.CELL.2007.11.019

Takahashi, K., & Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell*, *126*(4), 663–676. https://doi.org/10.1016/J.CELL.2006.07.024

Talbert, P. B., & Henikoff, S. (2006). Spreading of silent chromatin: Inaction at a distance. *Nature Reviews Genetics*, *7*(10), 793–803. https://doi.org/10.1038/nrg1920

Tang, F., Barbacioru, C., Wang, Y., Nordman, E., Lee, C., Xu, N., Wang, X., Bodeau, J., Tuch, B. B., Siddiqui, A., Lao, K., & Surani, M. A. (2009). mRNA-Seq whole-transcriptome analysis of a single cell. *Nature Methods*, *6*(5), 377–382. https://doi.org/10.1038/nmeth.1315

Taskiran, I. I., Spanier, K. I., Dickmänken, H., Kempynck, N., Pančíková, A., Ekşi, E. C., Hulselmans, G., Ismail, J. N., Theunis, K., Vandepoel, R., Christiaens, V., Mauduit, D., & Aerts, S. (2024). Cell-type-directed design of synthetic enhancers. *Nature*, *626*(7997), 212–220. https://doi.org/10.1038/s41586-023-06936-2

Tekkök, S. B., Brown, A. M., Westenbroek, R., Pellerin, L., & Ransom, B. R. (2005). Transfer of glycogen-derived lactate from astrocytes to axons via specific monocarboxylate transporters supports mouse optic nerve activity. *Journal of Neuroscience Research*, *81*(5), 644–652. https://doi.org/10.1002/jnr.20573

Thompson, A. J., Banwell, B. L., Barkhof, F., Carroll, W. M., Coetzee, T., Comi, G., Correale, J., Fazekas, F., Filippi, M., Freedman, M. S., Fujihara, K., Galetta, S. L., Hartung, H. P., Kappos, L., Lublin, F. D., Marrie, R. A., Miller, A. E., Miller, D. H., Montalban, X., … Cohen, J. A. (2018). Diagnosis of multiple sclerosis: 2017 revisions of the McDonald

criteria. *The Lancet Neurology*, *17*(2), 162–173. https://doi.org/10.1016/S1474-4422(17)30470-2

Thurman, R. E., Rynes, E., Humbert, R., Vierstra, J., Maurano, M. T., Haugen, E., Sheffield, N. C., Stergachis, A. B., Wang, H., Vernot, B., Garg, K., John, S., Sandstrom, R., Bates, D., Boatman, L., Canfield, T. K., Diegel, M., Dunn, D., Ebersol, A. K., … Stamatoyannopoulos, J. A. (2012). The accessible chromatin landscape of the human genome. *Nature 2012 489:7414*, *489*(7414), 75–82. https://doi.org/10.1038/nature11232

Titus, H. E., Chen, Y., Podojil, J. R., Robinson, A. P., Balabanov, R., Popko, B., & Miller, S. D. (2020). Pre-clinical and Clinical Implications of "Inside-Out" vs. "Outside-In" Paradigms in Multiple Sclerosis Etiopathogenesis. *Frontiers in Cellular Neuroscience*, *14*. https://doi.org/10.3389/fncel.2020.599717

Trerotola, M., Relli, V., Simeone, P., & Alberti, S. (2015). Epigenetic inheritance and the missing heritability. *Human Genomics*, *9*(1). https://doi.org/10.1186/S40246-015-0041-3

Trojer, P., & Reinberg, D. (2007). Facultative Heterochromatin: Is There a Distinctive Molecular Signature? *Molecular Cell*, *28*(1), 1–13. https://doi.org/10.1016/j.molcel.2007.09.011

Vallstedt, A., Klos, J. M., & Ericson, J. (2005). Multiple dorsoventral origins of oligodendrocyte generation in the spinal cord and hindbrain. *Neuron*, *45*(1), 55–67. https://doi.org/10.1016/j.neuron.2004.12.026

van Bruggen, D., Pohl, F., Langseth, C. M., Kukanja, P., Lee, H., Albiach, A. M., Kabbe, M., Meijer, M., Linnarsson, S., Hilscher, M. M., Nilsson, M., Sundström, E., & Castelo-Branco, G. (2022). Developmental landscape of human forebrain at a single-cell level identifies early waves of oligodendrogenesis. *Developmental Cell*, *57*(11), 1421-1436.e5. https://doi.org/10.1016/j.devcel.2022.04.016

Venneti, S., Garimella, M. T., Sullivan, L. M., Martinez, D., Huse, J. T., Heguy, A., Santi, M., Thompson, C. B., & Judkins, A. R. (2013). Evaluation of histone 3 lysine 27 trimethylation (H3K27me3) and enhancer of Zest 2 (EZH2) in pediatric glial and glioneuronal tumors shows decreased H3K27me3 in H3F3A K27M mutant glioblastomas. *Brain Pathology (Zurich, Switzerland)*, *23*(5), 558–564. https://doi.org/10.1111/bpa.12042

Waddington, C. H. (2014). The strategy of the genes: A discussion of some aspects of theoretical biology. *The Strategy of the Genes: A Discussion of Some Aspects of Theoretical Biology*, 1–262. https://doi.org/10.4324/9781315765471

Wang, X., He, Y., Zhang, Q., Ren, X., & Zhang, Z. (2021). Direct Comparative Analyses of 10X Genomics Chromium and Smart-seq2. *Genomics, Proteomics & Bioinformatics*, *19*(2), 253–266. https://doi.org/10.1016/j.gpb.2020.02.005

Wolock, S. L., Lopez, R., & Klein, A. M. (2019). Scrublet: Computational Identification of Cell Doublets in Single-Cell Transcriptomic Data. *Cell Systems*, *8*(4), 281-291.e9. https://doi.org/10.1016/j.cels.2018.11.005

Wong, G. H. W., Bartlett, P. F., Clark-Lewis, I., Battye, F., & Schrader, J. W. (1984). Inducible expression of H–2 and Ia antigens on brain cells. *Nature*, *310*(5979), 688–691. https://doi.org/10.1038/310688a0

Wu, S. J., Furlan, S. N., Mihalas, A. B., Kaya-Okur, H. S., Feroze, A. H., Emerson, S. N., Zheng, Y., Carson, K., Cimino, P. J., Keene, C. D., Sarthy, J. F., Gottardo, R., Ahmad, K., Henikoff, S., & Patel, A. P. (2021). Single-cell CUT&Tag analysis of chromatin modifications in differentiation and tumor progression. *Nature Biotechnology*, *39*(7), 819–824. https://doi.org/10.1038/S41587-021-00865-Z

Xiao, Y., Petrucco, L., Hoodless, L. J., Portugues, R., & Czopka, T. (2022). Oligodendrocyte precursor cells sculpt the visual system by regulating axonal remodeling. *Nature Neuroscience*, *25*(3), 280–284. https://doi.org/10.1038/s41593-022-01023-7

Xie, Y., Zhu, C., Wang, Z., Tastemel, M., Chang, L., Li, Y. E., & Ren, B. (2023). Droplet-based single-cell joint profiling of histone modifications and transcriptomes. *Nature Structural & Molecular Biology*, *30*(10), 1428–1433. https://doi.org/10.1038/s41594-023-01060-1

Yeung, M. S. Y., Djelloul, M., Steiner, E., Bernard, S., Salehpour, M., Possnert, G., Brundin, L., & Frisén, J. (2019). Dynamics of oligodendrocyte generation in multiple sclerosis. *Nature*, *566*(7745), 538–542. https://doi.org/10.1038/S41586-018-0842-3

Yeung, M. S. Y., Zdunek, S., Bergmann, O., Bernard, S., Salehpour, M., Alkass, K., Perl, S., Tisdale, J., Possnert, G., Brundin, L., Druid, H., & Frisén, J. (2014). Dynamics of Oligodendrocyte Generation and Myelination in the Human Brain. *Cell*, *159*(4), 766–774. https://doi.org/10.1016/j.cell.2014.10.011

Yusuf, F., & Brand-Saberi, B. (2006). The eventful somite: Patterning, fate determination and cell division in the somite. *Anatomy and Embryology*, *211*(1), 21–30. https://doi.org/10.1007/s00429-006-0119-8

Zeisel, A., Hochgerner, H., Lönnerberg, P., Johnsson, A., Memic, F., van der Zwan, J., Häring, M., Braun, E., Borm, L. E., La Manno, G., Codeluppi, S., Furlan, A., Lee, K., Skene, N., Harris, K. D., Hjerling-Leffler, J., Arenas, E., Ernfors, P., Marklund, U., & Linnarsson, S. (2018). Molecular Architecture of the Mouse Nervous System. *Cell*, *174*(4), 999-1014.e22. https://doi.org/10.1016/j.cell.2018.06.021

Zhang, H., Hung, C.-L., Liu, M., Hu, X., & Lin, Y.-Y. (2019). NCNet: Deep Learning Network Models for Predicting Function of Non-coding DNA. *Frontiers in Genetics*, *10*. https://doi.org/10.3389/fgene.2019.00432

Zhang, K., Hocker, J. D., Miller, M., Hou, X., Chiou, J., Poirion, O. B., Qiu, Y., Li, Y. E., Gaulton, K. J., Wang, A., Preissl, S., & Ren, B. (2021). A single-cell atlas of chromatin accessibility in the human genome. *Cell*, *184*(24), 5985-6001.e19. https://doi.org/10.1016/j.cell.2021.10.024

Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., Nussbaum, C., Myers, R. M., Brown, M., Li, W., & Shirley, X. S. (2008). Model-based analysis of ChIP-Seq (MACS). *Genome Biology*, *9*(9), 1–9. https://doi.org/10.1186/GB-2008-9-9-R137/FIGURES/3

Zhu, C., Zhang, Y., Li, Y. E., Lucero, J., Behrens, M. M., & Ren, B. (2021). Joint profiling of histone modifications and transcriptome in single cells from mouse brain. *Nature Methods 2021 18:3*, *18*(3), 283–292. https://doi.org/10.1038/s41592-021-01060-3

Zu, S., Li, Y. E., Wang, K., Armand, E. J., Mamde, S., Amaral, M. L., Wang, Y., Chu, A., Xie, Y., Miller, M., Xu, J., Wang, Z., Zhang, K., Jia, B., Hou, X., Lin, L., Yang, Q., Lee, S., Li, B., …

Ren, B. (2023). Single-cell analysis of chromatin accessibility in the adult mouse brain. *Nature*, *624*(7991), 378–389. https://doi.org/10.1038/s41586-023-06824-9