From the Department of Medical Biochemistry and Biophysics
Karolinska Institutet, Stockholm, Sweden

CELL STATES ALONG OLIGODENDROCYTE DEVELOPMENT AND DISEASE

David van Bruggen

Karolinska Institutet

Stockholm 2021

Published by Karolinska Institutet.
Printed by Universitetsservice US-AB, 2021

ISBN 978-91-8016-103-9
Cover illustration: Network depicting developing brain cells

CELL STATES ALONG OLIGODENDROCYTE DEVELOPMENT AND DISEASE

THESIS FOR DOCTORAL DEGREE (Ph.D.)

By

David van Bruggen

The thesis will be defended in public at **Eva & Georg Klein Hall**, Solnavägen 9,
Friday January 15, 14:00

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

Co-supervisor(s):
Prof. Sten Linnarsson
Karolinska Institutet
Department of Medical Biochemistry and Biophysics
Division of Molecular Neurobiology

Asscociate Prof. Carsten Daub
Karolinska Institutet
Department of Biosciences and Nutrition

Opponent:
Associate Prof. Rahul Satija
New York University
Center for Genomics and Systems Biology
New York Genome Center

Examination Board:
Associate Prof. Eva Hedlund
Karolinska Institutet
Department of Cell and Molecular Biology

Prof. Thomas Perlmann
Karolinska Institutet
Department of Cell and Molecular Biology

Prof. Ragnhildur Thóra Káradóttir
University of Cambridge
University of Iceland
Department of Veterinary Medicine

"There is grandeur in this view of life,
with its several powers, having been originally breathed into a few forms or into one; and that,
whilst this planet has gone cycling on according to the fixed law of gravity,
from so simple a beginning endless forms most beautiful and most wonderful have been,
and are being, evolved."

Charles Darwin - On the Origin of Species by Means of Natural Selection, 1859

ABSTRACT

The brain, one of the most complex organs in the body, where an immense diversity of cell states emerges from simple structure, where function arises from sets of regulatory principles and pattern persist where individual cells do not. Revealing the regulatory underpinnings of the brain, from unspecified cell states to diversity, is paramount for achieving a thorough understanding of the development process and generating insight into the disease states of the brain. This thesis is an exploration into how canonical regulatory factors and elements, such as transcription factors and genes, lock a regulatory system in a multi-outcome network with limited possible states.

The work in this thesis focuses on the oligodendrocyte lineage, a glial cell known for it's supportive role in the central nervous system, where it facilitates electrical transmission through the enscheathment of axons. Oligodendrocytes (OLs) lie at the heart of multiple sclerosis (MS), a disease where an immune response is mounted against myelin. As a response, oligodendrocyte precursor cells (OPCs) move towards lesions and remyelinate axons, however, this mechanism fails in later stages of the disease. Thus, an understanding to how OPCs develop is vital to amelioration of the altered oligodendrocyte population.

In **Paper I** we reveal a previously underestimated heterogeneity within the oligodendrocyte lineage in mouse. We show that OL maturation is an ongoing process, albeit, decreasing in frequency with age. Furthermore, complex wheel training in mice revealed that the OLs respond to this challenge through an increase in differentiation.
Paper II investigates the cellular response in the experimental autoimmune encephalo-myelitis (EAE) disease mouse model of MS, where we find a tailored response by the resident OL population, changed from its normal transcriptional program, expressing a spectrum of genes related to survival, immunological stimulation, phagocytosis, and active differentiation. Furthermore, we provide evidence that OLs can elicit responses from T cells.
In **Paper III** we explore the different waves of OPC generation in the developing mouse brain at embryonic day 13.5 and postnatal day 7. We show that recently Pdgfra expressing cells at the E13.5 time point exhibit a multitude of patterning genes, and we show the emergence of a possible OPC progenitor through the inclusion of a bridging E17.5 time point population. This pre-OPC population is biased towards expressing glial and OL lineage specifying genes such as Olig1, Olig2, Ptprz1, and Bcan. Furthermore, lineage tracing of OPC developmental waves, shows no transcriptional differences, leading us to conclude that OPCs are generally naïve to the time or region of specification.
In **Paper IV** we show that we are able to detect OPC formation in the developing human forebrain. We detect OPCs at the earliest sampled time point post conception week 8. We attempt to recover the path of OPC formation, and investigate the regulatory dynamics in the specification of OPCs.

LIST OF SCIENTIFIC PAPERS

I. *Sueli Marques*, Amit Zeisel*, Simone Codeluppi,* ***David van Bruggen****, Ana Mendanha Falcão, Lin Xiao, Huiliang Li, Martin Häring, Hannah Hochgerner, Roman A Romanov, Daniel Gyllborg, Ana B Muñoz-Manchado, Gioele La Manno, Peter Lönnerberg, Elisa M Floriddia, Fatemah Rezayee, Patrik Ernfors, Ernest Arenas, Jens Hjerling-Leffler, Tibor Harkany, William D Richardson, Sten Linnarsson, Gonçalo Castelo-Branco*
Oligodendrocyte heterogeneity in the mouse juvenile and adult central nervous system
Science *2016, Vol. 352 Issue 6291, Page 1326-1329*

II. *Ana Mendanha Falcão*,* ***David van Bruggen****, Sueli Marques, Mandy Meijer, Sarah Jäkel, Eneritz Agirre, Samudyata, Elisa M. Floriddia, Darya Vanichkina, Charles ffrench-Constant, Anna Williams, André Ortlieb Guerreiro-Cacais, Gonçalo Castelo-Branco*
Disease-specific oligodendrocyte lineage cells arise in multiple sclerosis
Nature Medicine *2018, Vol. 24, Page 1837-1844*

III. *Sueli Marques*,* ***David van Bruggen****, Darya Pavlovna Vanichkina*, Elisa Mariagrazia Floriddia, Hermany Munguba, Leif Väremo, Stefania Giacomello, Ana Mendanha Falcão, Mandy Meijer, Åsa Kristina Björklund, Jens Hjerling-Leffler, Ryan James Taft, Gonçalo Castelo-Branco*
Transcriptional Convergence of Oligodendrocyte Lineage Progenitors during Development
Developmental Cell *2018, Vol. 46, Page 504-517*

IV. ***David van Bruggen****, Erik Sundström, Gonçalo Castelo-Branco*
Single cell transcriptomics reveals cell-states along human neural developmental landscape
Manuscript

Publications not included in this thesis

*Elisa M Floriddia, Tânia Lourenço, Shupei Zhang, **David van Bruggen**, Markus M Hilscher, Petra Kukanja, João P Gonçalves Dos Santos, Müge Altınkök, Chika Yokota, Enric Llorens-Bobadilla, Sara B Mulinyawe, Mário Grãos, Lu O Sun, Jonas Frisén, Mats Nilsson, Gonçalo Castelo-Branco*
Distinct oligodendrocyte populations have spatial preference and different responses to spinal cord injury
***Nature communications** 2020 November 17*

*Sueli Marques, **David van Bruggen**, Gonçalo Castelo-Branco*
Single-Cell RNA Sequencing of Oligodendrocyte Lineage Cells from the Mouse Central Nervous System
***Oligodendrocytes – Methods in Molecular Biology** 2019 March 01*

*Sarah Jäkel, Eneritz Agirre, Ana Mendanha Falcão, **David Van Bruggen**, Ka Wai Lee, Irene Knuesel, Dheeraj Malhotra, Anna Williams, Gonçalo Castelo-Branco*
Altered human oligodendrocyte heterogeneity in multiple sclerosis
***Nature** 2019 Januari 23*

*Gioele La Manno, Ruslan Soldatov, Amit Zeisel, Emelie Braun, Hannah Hochgerner, Viktor Petukhov, Katja Lidschreiber, Maria E Kastriti, Peter Lönnerberg, Alessandro Furlan, Jean Fan, Lars E Borm, Zehua Liu, **David van Bruggen**, Jimin Guo, Xiaoling He, Roger Barker, Erik Sundström, Gonçalo Castelo-Branco, Patrick Cramer, Igor Adameyko, Sten Linnarsson, Peter V Kharchenko*
RNA velocity of single cells
***Nature** 2018 August 18*

***David van Bruggen**, Eneritz Agirre, Gonçalo Castelo-Branco*
Single-cell transcriptomic analysis of oligodendrocyte lineage cells
***Current opinion in neurobiology** 2017 December 01*

CONTENTS

LIST OF ABBREVIATIONS

CCA	cannonical correlation analysis
CGE	caudal ganglionic eminence
CNS	central nervous system
EAE	experimental autoimmune encephalo-myelitis mouse model
EMT	epithelial to mesenchymal transitioning
KNN-graph	*K*-Nearest neighbor graph
LGE	lateral ganglionic eminence
MGE	medial ganglionic eminence
MS	multiple sclerosis
OL	oligodendrocyte
OPC	oligodendrocyte precursor cell
pMN	progenitor of motor neuron domain
PPMS	primary progressive MS
Pseudotime	inferred ordering of cells through a biological process
RNAseq	Massively parallel RNA sequencing
RRMS	relapsing-remitting MS
Seq	sequencing
SPMS	progressive MS
SVZ	sub ventricular zone
UMI	unique molecular identifier
VZ	ventricular zone

1 CELL TYPES AND THE EMERGENCE OF NEURAL STRUCTURES

Understanding of the regulatory and developmental patterns underlying current neural structures is best sought through the lens of its evolutionary past, by which complicated transcription factor patterns can be observed in the context of their first emergence.

The first known structures resembling a rudimentary nerve net can be traced back to a common ancestor probably shared by cnidarians; a group pertaining to jellyfish, corals, sea anemones, and other soft bodied polyps [1,2]. Cell-type diversification events and the evolution of certain types of gene-modules can be combined to chart the emergence of neuronal cell-types during evolution using comparative genomics approaches enabled by the emergence of ctenophore, placozoa, and cnidarian annotated genomes [3–6], allowing for molecular resolution phylogenetics.

1.1 DEFINITION OF CELL TYPES

The definition of cell-types according to Arendt, Musser, et al. 2016 is described as, *"a set of cells in an organism that change in evolution together, partially independent of other cells, and are evolutionarily more closely related to each other than to other cells"*. Thus, cell-types are seen as individual evolutionary units, which stem from ancient ancestor cell-types, in which the first core regulatory units evolved that enabled the first cell-types to gain specific functions, and subsequently a gain of diversification [7]. The proper definition and elucidation of cell-types is ever more important with current efforts underway, generating mouse and human cell atlases, making use of novel techniques that make available molecular resolution data in individual cells [8].

1.1.1 A brief primer on present day cell types in the CNS

Early studies by Ramón y Cajal revealed a great diversity and morphological distinctions of cells in the central nervous system (CNS). The simplest distinction is made by grouping the cells in the category of nerve cells and supporting cells. Nerve cells communicate over large distances, through structures such as axons, which are long projections of cells enabling the neurons to form circuits throughout the brain. Hundreds of different neuronal cell types exist, with amazingly intricate morphologies. Some neurons stay in a layer in the brain, others project outward into other layers. Most neurons in the cortex project into others layers, generating intricate morphologies and cortical layer specific cell types. Aside from neurons, the supporting cells in the brain, glia, include oligodendrocytes, astrocytes, and microglia.

Microglia are immune-competent cells widely spread in the CNS, where they assist in maintainance, immune-surveilance and phagocytosis. However they do not originate from the brain during development, instead they migrate from the yolk-sac progenitors in a brief time-window during development.

Astrocytes are the most abundant glial cell in the CNS. Astrocytes have many responses to a plethora of signals in the brain, but the most prominent functions of the astrocytes are the maintenance of homeostasis of the environment, supportive participation in synapse chemistry, and blood brain barrier maintainance.

Mature oligodendrocytes function as the myelinating cell in the CNS. Ensheathing axons and promoting fast salutatory signal propagation essential for proper function of the CNS. Many functions of oligodendrocytes remain to be discovered, from the presence of non-myelinating oligodendrocytes, to the nature of the oligodendrocyte precursor cell (OPC). OPC function, aside from generating oligodendrocyte lineage cells, is far from clear. Recent research suggests OPC involvement in direct communication with neurons, as well as a broader role in progenitor function.

1.1.2 Cell types as a collection of gene modules

Gene-modules or core regulatory units, expressed in present day synapses are most likely originating from a common ancestor of choanoflagellates and metazoans which express certain proteins that are associated to synaptic function [9]. Several studies [2,10–13] suggest that the post-synaptic module first evolved as a primitive chemosensory module that could perhaps be sensitive to amino acid detection like glutamate, which later evolved into a bigger signaling role such as in the post-synaptic function, and in a similar manner ionotropic glutamate receptor (GABA) families existed in the common metazoan ancestor, right before the emergence of neurons as a cell type through most likely diversification of chemosensory cells for which Neuroligin and Neurexin seem to have been a key event [1,2]. Furthermore, sponges are known to release “neurotransmitters” from cells lining their channels including glutamate, coordinating contractions along muscle cells in the body wall, yet are lacking a clear cell type that could be considered a neuron [2]. The fact that the neuronal lineage constitutes one of the most diversified cell-lineages of any organism [14], allows speculation that patterning factors must have evolved as a consequence of the diversification event in the neuronal lineage, right after the appearance of neurons in the last common ancestor of cnidarians, such as anemones and jellyfish, and this is indeed observed [15].

1.2 THE EVOLUTION OF THE NERVOUS SYSTEM

Primitive nerve nets as those found in cnidarians seem to be developmentally subdivided into regions [16]. Recent comparative analyses suggest a neurodevelopmental similarity in molecular and spatial arrangements between annelids (segmented worms) and vertebrates, interestingly recent molecular data also shows that this similarity extents to the cnidarian sea anemone, leading to the hypothesis that this regionalization already existed in a common ancestor of cnidarians, such as jellyfish, and bilaterians, such as the mammalia [17].

The similarities of the ectodermal tissue (one of the three germinal layers), and more specifically the neuro-ectoderm (which gives rise to neurons), is proposed to stem from a common ancestor possessing a crude nerve net [3,17]. The phylogenetic analysis indicates shared transcription factors that control the regionalization of the developing neural tissue which in the common ancestor gave rise to two specialized parts of the nervous system, one on the apical pole, and the other on the blastoporal developmental region as analysed and postulated by [15]. Comparisons between phyla in different developmental stages can then shed light on the early evolution of the nervous system. The common ancestor of cnidarians and bilaterians must have had a gastrula-like basic organization, because it has been shown that cnidarian and bilaterian similarities include WNT-β-catenin signaling interacting and activating achaete-scute (ASCL) families, basic helix-loop-helix (bHLH) proneural genes, Notch signaling, as well as SOXB2 [18,19] which has similarities with SOX2 in humans [20] indicating homology to bilaterian neurogenic ectoderm.

The before mentioned neuron diversification event in evolution has as a consequence that specific patterning factors must have evolved to support regionalization and specification of neurons. Cnidarians and bilaterians express homologous transcription factors, and morphogens conserved in evolution, Sine oculis 3 (SIX3), Forkhead box A (FOXA), Sonic Hedgehog (SHH), Brachyury, Forkhead box B (FOXB), NK2.2, and LIM homeobox (LMX), are restricted to the most-medial region, PAX is expressed more laterally, overlapping with NK6 and Msh homeobox (MSX). The most peripheral or lateral region expresses Iroquois homeobox (IRX) and Distal-less homeobox (DLX), as mentioned and compared in Arendt et al., 2016 (Fig 1a-c), and related to mammalian dorso-ventral patterning. Additionally, Arendt et al., 2016 proposes homology between the cnidarian gastric pouches and bilaterian somites, which interestingly are both expressing genes homologous to the HOX family and GBX transcription factors (Fig 1d-f). Furthermore, conserved WNT signalling of genes such as WntA, Wnt1, Wnt7, and Wnt4, and their orthologes WNT1, WNT7, WNT4, and WNT2 in

combination with genes such as SHH, might provide a glimpse into what ancient neural celltypes might have been present in the common ancestor [15]. Thus, it comes to no surprise that these conserved factors play a major role in the formation of the neural crest, which lies at the basis of vertebrate and mammalian central nervous system (CNS) development.

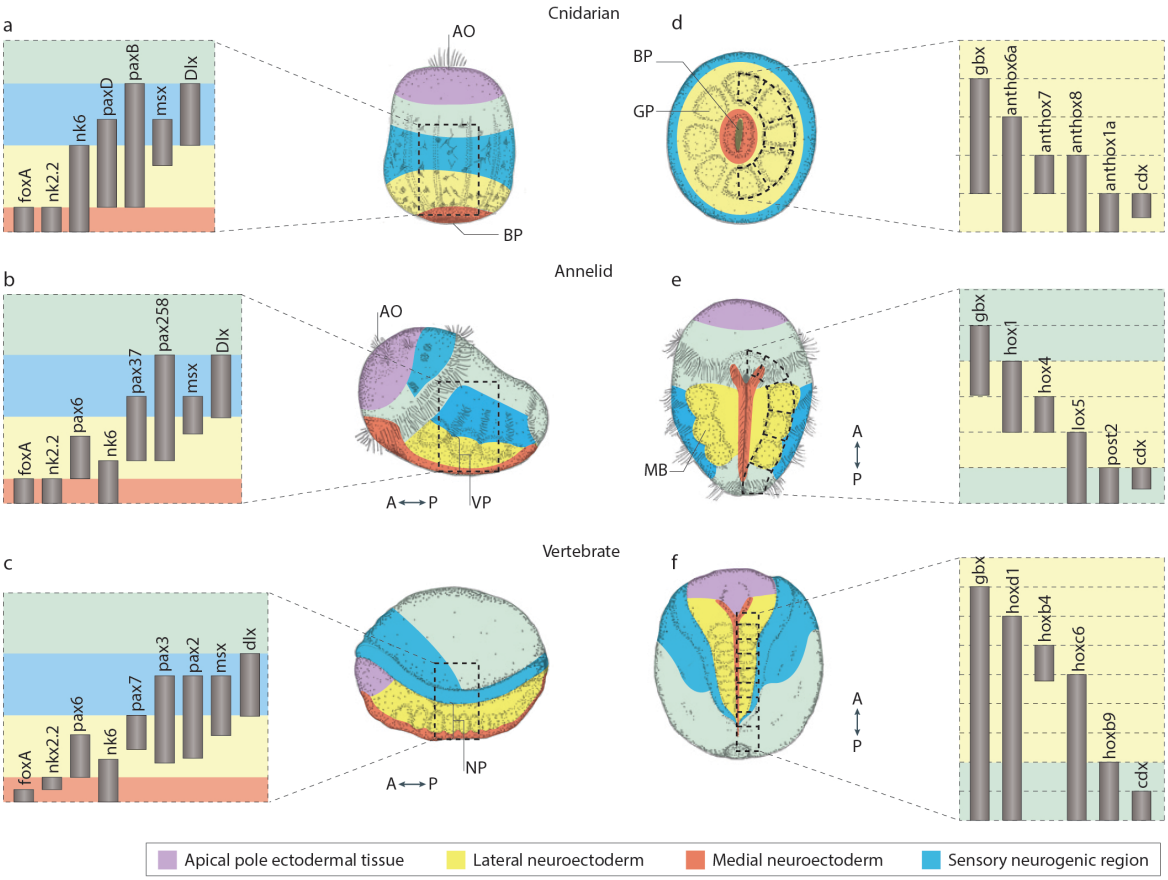

Figure 1 Cnidarians and bilaterians express homologous transcription factors conserved in evolution. Adapted from Arendt et al., 2016 with permission.

1.3 A BRIEF OVERVIEW OF DORSO-VENTRAL PATTERNING IN THE MAMMALIAN NEURAL CREST AND NEURAL TUBE, SPECIFYING OLIGODENDROCYTE PRECURSOR DOMAINS

The neural crest, derived from neuro-ectoderm, is unique to vertebrate development. The formation of the neural crest hinges on the specification of a neural plate border during gastrulation. This process starts with neural specification regulated by an interplay of WNT inhibitors, FGFs, and BMP signals, forming a gradient in the neural plate [21]. WNT and BMPs are expressed from lateral sides of the developing embryo, whereas inhibitors are expressed from medial regions such as the neural plate. These established morphogens are then responsible for the formation of a gradient containing a region or zone of intermediate WNT and BMP activity, from where the future neural plate border cells arise [22].

The specification of the neural plate border brings about a new wave of expression and patterning factors, which is not completely homogenous throughout the neural plate border. These neural plate border-specifiers involve factors such as Tfap2, Msx1, Zic1, Gbx2, Pax3/7, Dlx5/6, Gata2/3, Foxi1/2, and Hes1/4 [23–25], and are in crosstalk through feedback loops, ultimately leading to the establishment of a stable patterning across the neural plate border. The non-uniformity is a key factor in subsequent regionalization and roles in later specification. After neural plate border specification the neural crest specifiers start being established, factors such as FoxD3, Ets1, and Snai1/2, are responsible for the establishment of the premigratory neural crest [25]. Further regionalization is established by the formation of ectodermal placodes on the lateral sides of the neural crest border, and give rise to sensory neurons and other cells of the peripheral nervous system. The ectodermal placodes are established by interactions of Dlx5/6, Gata2/3, Foxi1/3, from which key placodal regulators start to be expressed such as Six1, Eya1/2 and Irx1 [26,27]. The neural crest then undergoes epithelial to mesenchymal transitioning (EMT) regulated or initiated by Snai1/2, and WNT signaling, amongst others [28,29]. Several Sox family factors, such as Sox8/9/10 and Sox5/6 are implicated in differentiation and initiation of the migratory neural crest into several fates such as chondrocytes, melanocytes, and neurons [25,30]. The neural crest develops into many different cell-types of the PNS, completed through the process of neurulation, of which a key event is the generation of the neural tube [31].

Neural tube formation is achieved through a complicated interplay of a tightly regulated network of dorso-ventral patterning genes that are both repressive and activating leading to sharply defined boundaries as well as fusion of the neural plate in a process called “neurulation”, in ways that are still poorly understood. Compared to neural crest formation, far less is known about neural tube formation and the complicated process of closure and fusion of the neural tube at the dorsal side. It is now known that dorso-ventral patterning is established through gradients of signaling molecules secreted from different structures during development, leading to the expression of a tight knit network of local patterning factors along the dorsal-ventral axis that produce neural progenitor domains through mutual repressive and activating activities. Many of these factors are homologous to cnidarian regionalization, such as expression of factors such as Lhx, Pax3/6, Ascl1 (Mash), Ngn1/2, Dbx1/2, Irx3, and Nkx6.1/2.2 [15,32]. Subsequently, it is possible to perturb domain formation by repressing expression of some of these factors resulting in expansion or ablation of certain progenitor domains. Neural tube formation is highly dependent on mesodermal structures starting with the notochord. This structure is located in the central midline of the embryo with

the developing neural tube more dorsally and the notochord more ventrally located [33]. Additionally, mesoderm is also present on the lateral sides of the forming neural tube, which will segment into somites. The notochord induces floorplate formation, primarily by inhibition of proliferation producing a thin layer of cells that become the floorplate [34]. Evidence exists that sonic hedgehog (Shh), expressed from the notochord can aid in inhibition of cell cycle kinases. Conversely, the lateral mesoderm segmenting into somites, produces an FGF gradient stimulating cell proliferation and thereby thickening the lateral sides of the neural tube [35]. The combined gradients of Shh from the floorplate, FGF laterally, and Bmp4 from the roofplate provide a stratified regionalization on which further structure can be generated.

1.4 ESTABLISHMENT OF RADIAL GLIA, AND THE GLIAL LINEAGE

Around mouse embryonic day 9-10.5 the established neuro-epithelium of ectodermal origin begins to form a regionalized layer of cells with an emerging glial phenotype called the radial glial cells [36]. These bi-polar cells retain contact with both pial and ventricular surfaces and are restricted to the ventricular zone, where they begin to express a set of unique markers such as GLAST (Slc1a3), TNC, and BLBP (Fabp7), as well as Nestin, Vimentin, and in some cases GFAP [36–38]. This transformation of the neuro-epithelium to radial glial fate accompanies some structural changes such as the emergence of adherens junctions, necessary for maintaining the integrity and to bind the apical domains of the radial glial cells together forming the ventricular zone (VZ) [39–41].

The VZ is a thin layer of cells lining the ventricular surface from which the radial glial cells produce progenitor cells through mainly asymmetrical divisions, producing all neuronal lineage cells in the brain. Similarly to the neural tube from which they stem, radial glial cells are subject to the patterning factors present in the regions from which they have emerged and generate distinct neural progenitors in specified regions of the neuro-epithelial sheet derived ventricular surface. Radial glia maintain apical-basal polarity similar to the neuroepithelium from which they are derived, project a pial process towards the cortical plate, and feature adherens junctions that maintain an organized basal organization on the apical side forming the cortical lamina which is important for maintaining VZ integrity. The evolutionary conserved genes Numb and Numbl have a role in maintaining the adherens junctions and are essential for maintaining cell polarity [42].

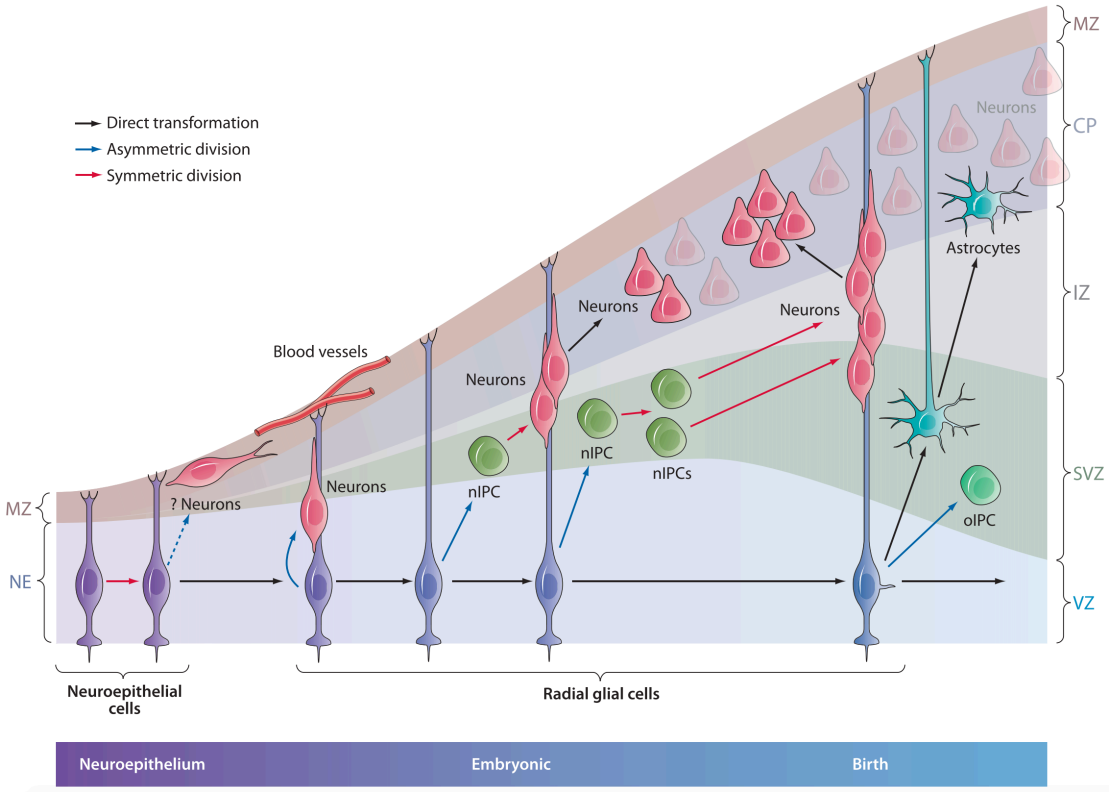

Figure 2 Neurogenesis during cortical development. CP, cortical plate; IZ, intermediate zone; MZ, marginal zone; nIPC, neurogenic intermediate progenitor cell; oIPC, oligodendrocyte generating intermediate progenitor cell; RG, radial glia; SVZ, subventricular zone; VZ, ventricular zone. Adapted from *Kriegstein, Alvarez-Buylla, 2009* with permission.

Radial glial proliferation and progenitor differentiation occurs during cortical neurogenesis, where RG cells proliferate and differentiate in an asymmetrical manner maintaining their apico-basal polarity. Recently, studies have shown that RG cells produce daughter cells with restricted lineage potential called transit amplifying cells, however RG cells can also directly produce neurons, both are generated through asymmetric division events [43]. During neurogenesis, the newly formed daughter cell does not inherit the pial process in most cases but migrates along the pial fiber of the RG cell towards the cortical plate and the upper layers, while the RG cell can divide again. The intermediate progenitors move from the VZ to the emerging sub-ventricular zone (SVZ) and have the potential to self renew and produce a specific sequence of neuronal subtypes, subtype progenitors or glial progenitor cells depending on the regional identity determined by the dorso-ventral patterning of the RG cell population, thus, a single RG cell population produces a sequence of restricted progenitors in a spatiotemporal dependent manner [43]. A recent study found a remarkable similarity in the VZ and SVZ of different developmental regions, indicating similar maturation trajectories between different developing regions of the brain diversify post-mitotically [44].

1.5 TELENCEPHALON FORMATION AND DORSO-VENTRAL REGIONALIZATION

Guided by patterning factors along the anterior and posterior axis, the anterior part of the neuro-epithelium is destined to become the brain, initiated by the formation of the telencephalon. Similar to the neural tube formation, several cell extrinsic factors such as FGF, SHH, WNT and BMP families, play a role in the regionalization of the brain. However, the regionalization is further established by transcription factors, analogous to the formation of the neural tube, a tight-knit network of factors establish zones and transient structures during development through mutual repression and activation combinations establishing sharp boundaries of neural progenitor populations in the developing telencephalon. This network of regionalization mainly consists of transcription factors, encoded by genes such as Foxg1, Gli3, Pax6, Lhx2, Gsx2, Nkx2.1, and Emx2 [45–48].

Telencephalon formation is initiated by expression of the transcription factor Foxg1 [45,46]. Initiation of Foxg1 expression is established through mutual signaling of FGFs, Wnts, Shh, and BMPs in a complicated series of repression and activation events generating a niche that induces Foxg1 expression in the developing anterior end of the anterior neural ridge, followed by subdivisions into domains expressing Nkx2.1, Gsx2, Pax6, and Emx2 [49]. Before complete closure of the neural tube, a broad domain organization already exists dividing the developing neuroepithelium into a dorsal and a ventral domain, generating glutamatergic and GABAergic neurons respectively [50]. Establishment of dorsal and ventral domains is achieved through patterning of Gli3 in combination with Shh, similarly to the spinal cord. Initially Gli3 is expressed from all cells in the developing telencephalon, but is then restricted by Shh, expressed from the ventral domain of the closed neuroepithelium, and is needed for proper expression initiation of further patterning factors such as Dlx2, Gsx2, and Nkx2.1 [51]. Additionally, development of proper ventral progenitor domains also depends on expression of Foxg1 and FGF signalling, but also for dorsal development in addition to Gli3, where both are needed for the development of the dorsal domain [52].

Consolidation of pattering in the early neural tube is defined by three regions of development, Pax6, Gsx2, and Nkx2.1 . Pax6 is upregulated in the neuro-epithelium similarly as in spinal cord development, and maintained by Wnt3a expression from the dorsal side of the early neural tube, where Pax6 expression is limited to the more dorsal part of the neural tube [53]. Simultaneously, Nkx2.1 is expressed on the ventral side of the neural tube, establishing a dorsal Pax6 and ventral Nkx2.1 expression gradient. Cells at the border between Pax6 and

Nkx2.1 start expressing the transcription factor Gsx2, where an expression gradient arises with Nkx2.1 expressing ventrally, Gsx2 at the boundary between ventral and in the dorsal region, and Pax6 expressing dorsally [54]. The dorsal expression of Gli3 is needed to initiate expression of Emx2, with expression occurring in the dorsal posterior part of the telencephalon, while Pax6 expression is restricted more anteriorly [55]. This initial pattering reflects the transient domains that will form the medial ganglionic eminence (MGE), marked by expression of Nkx2.1, the lateral ganglionic eminence (LGE) marked by expression of Gsx2, and the caudal ganglionic eminence (CGE) marked by expression of Emx1/2.

1.6 GLIOGENESIS AND THE FORMATION OF OLIGODENDROCYTES

Oligodendrogenesis occurs at several timepoints and in different structures in the brain and spinal cord. In mouse brain development three distinct waves of oligodendrocyte formation can be distinguished, generated from three separate brain structures mentioned before, the MGE (Nkx2.1) at E12.5, the LGE (Gbx2) at E15.5, and CGE (Emx1) postnatally, whereas in the spinal cord only the first two waves are distinguished [56]. This indicates a genetic regulatory diversity, which could potentially generate genetically and perhaps functionally distinct oligodendrocyte (OL) subtypes. Spatio-temporal differences, and a varying signaling environment provide discriminatory circumstances on which oligodendrocyte sub-type evolution could capitulate, however, recent studies indicate that oligodendrocyte progenitor cells (OPCs) from different progenitor sites in the brain seem similar [56], however the exact degree of difference remains unclear.

OPCs are generated from radial glial cells in the differentiated neuro-epithelium, from the VZ, and recent studies indicate that the SVZ also contains progenitors with OPC generating capacity [57]. Studies in the spinal cord and early neural tube have provided in-depth knowledge into the generation of spinal cord OPCs from the progenitor of motor neuron (pMN) domain in the patterned neural tube, the first occurrence of OL generation. The pMN domain is flanked by the p2 domain at the dorsal side and the p3 domain at the ventral side of the neural tube, and is characterized by the expression of the bHLH transcription factor Olig2, Pax6, and Nkx6.2, giving rise to the motor neuron lineage generating RG cells [58]. However, after motor neuron production, the RG cells in the pMN domain switch to produce glial cells instead. Olig2 is essential in the pMN domain and its expression is maintained in the OL lineage where it is necessary to initiate OL fate and OL maturation [59]. Olig2 is then not only necessary for motor neuron production but also for initiation and maintenance of OL fate, reflecting the complicated structure of the initiation of the OL transcriptional program.

Olig2 expression is restricted to the pMN domain through repressive interactions with Irx3 expressed more dorsally, and Nkx2.2 expressed more ventrally. Olig2, Nkx6.2, and Nkx2.2 expression is activated and maintained by the Shh gradient from the ventral side of the neural tube, while Irx3 is repressed by Shh, generating the distinctive pattern giving rise to the pMN domain [58].

OL specification involves a combination of expression patterns allowing precise control of OL generation. Olig2 and in a lesser degree Olig1 are required for the proper specification of OLs. Olig2 is also necessary for the proper specification of motor neurons from the pMN domain, thus Olig2 is required for both motor neuron specification and OL specification in the pMN domain. Moreover, Ascl1 is necessary for OL specification, downregulating expression of Dlx1 and Dlx2. Thus, Dlx1/2 expression is incompatible with OL specification [60]. Additionally, the pMN domain, and other domains within the patterning regions of the developing brain have both neural and glial potential. In the pMN domain, and in other regions as well, neural or glial fate is governed by proneural factors such as neurogenins, amongst others. Neurogenins, like neurogenin1 or neurogenin2 suppress glial fate in the pMN domain and are required for the formation of motorneurons [61].

The neuroglial switch entails processes involving several sets of proteins of which the precise actions are still unclear. Aside from downregulating the inhibitory actions of Dlx1/2 on glial fate, additional groups of transcription factors are necessary to initiate a gliogenic state in radial glial cells. Nuclear factor I A (NFIA) is a progliogenic protein promoting both astrocyte and oligodendrocyte fates while inhibiting neurogenesis. The duration of NFIA is important in specification of oligodendrogenesis, as the expression has to be transient; astrocyte fate is associated with continuous expression of NFIA. Furthermore, Notch signaling and expression of Sox9 is also associated with gliogenic fate, together with the less specific Sox8 transcription factor presumably setting in motion epigenetic changes involving the histone deacetylases HDAC1, and HDAC2, among others [50].

Oligodendrocyte fate is not solely determined by Olig2 evidenced by the fact that this transcription factor is also expressed in the specification of neurons in the pMN domain, as well as maturation of certain types of astrocytes. Although Olig2 interacts in a cross-repressive fashion with patterning factors such as Irx3 and Nkx2.2, recent studies indicate that Olig2 directly activates pro-oligodendrocyte factors such as Sox10. Sox 8, 9, and 10 (SoxE) are transiently co-expressed but only Sox10 will remain activated [62–65]. Sox10 and

Sox9 are activators of Pdgf receptor α (Pdgfra), vital for the mitotic, migratory, and survival properties displayed by the oligodendrocyte precursor cell (OPC) state, a stable proliferative adult stem cell at the root of the OL lineage [66]. Nkx2.2 is also upregulated, although in a slightly delayed fashion compared to Sox10, and maintained in the lineage, forming a feed-forward loop with Olig2 and Sox10 [67]. Current studies conclude that the differentiation of OPCs towards OL-lineage cells is postmitotic once committed to differentiation, and thus OPCs are migratory cells that proliferate during development and adulthood to populate the brain, primed for differentiation.

1.7 PROGENITOR MAINTENANCE AND THE OLIGODENDROCYTE MATURATION

OPC maintenance involves an interplay of negative regulators of differentiation and pro-differentiation factors. Notch signaling is found to be essential, in combination with expression of Hes5, a repressive bHLH protein able to render Sox10 inert through binding and subsequent formation of inert complexes [68]. Furthermore, Hes5 is able to recruit histone deacetylases to repress certain sites in the genome. Another mechanism of repression of the myelinating program is through the interaction of Sox5 and Sox 6. Competition of Sox5/6 with Sox10 for specific pro-differentiation binding sites indicates a repressive function for these factors in maintaining the progenitor state. Indeed, studies have found that Sox5/6 initiate the positioning of repressive complexes to myelin genes, repressing expression even in the event of Sox10 binding to its targets [69,70]. Bmp signaling might also be implicated in the maintenance of the progenitor state through the highly expressed Id2 and Id4 proteins. Id proteins lack a DNA binding domain but can dimerize and therefore are able to sequester targets such as Olig1/2 and Ascl1 [71]. Furthermore, studies have shown that canonical Wnt signaling can inhibit OL differentiation, which might explain the strong upregulation of the Wnt inhibitor APC, and the upregulation of the Wnt effector Tcf7l2, however, activation or repression through Wnt is context dependent and can vary in different differentiation stages [72].

Terminal OL differentiation involves a multitude of factors working in concert to establish the myelinogenic program. Although OPCs will spontaneously differentiate in vitro, maturation of OLs can be slowed or stopped in vitro and in vivo by targeting specific factors. One of the factors involved in OL differentiation initiation is Nkx2.2, although this factor is rapidly downregulated during maturation, studies indicate that Nkx2.2 might work as a repressor of factors that negatively regulate OL differentiation, thereby enabling the OL cell

to establish the myelinogenic program [73,74]. Differentiation of OLs requires HDAC1 and HDAC2 and several other epigenetic modulators, however HDACs can also act to inhibit differentation. HDAC1 and HDAC2 are believed to act in concert with the zinc finger transcription factor YY1, aside from direct activation of myelin genes by the zinc finger transcription factor [75]. Other factors are the thyroid hormone receptor (TR), and retinoic acid receptors (RAR), both strongly expressed in OLs [72,76].

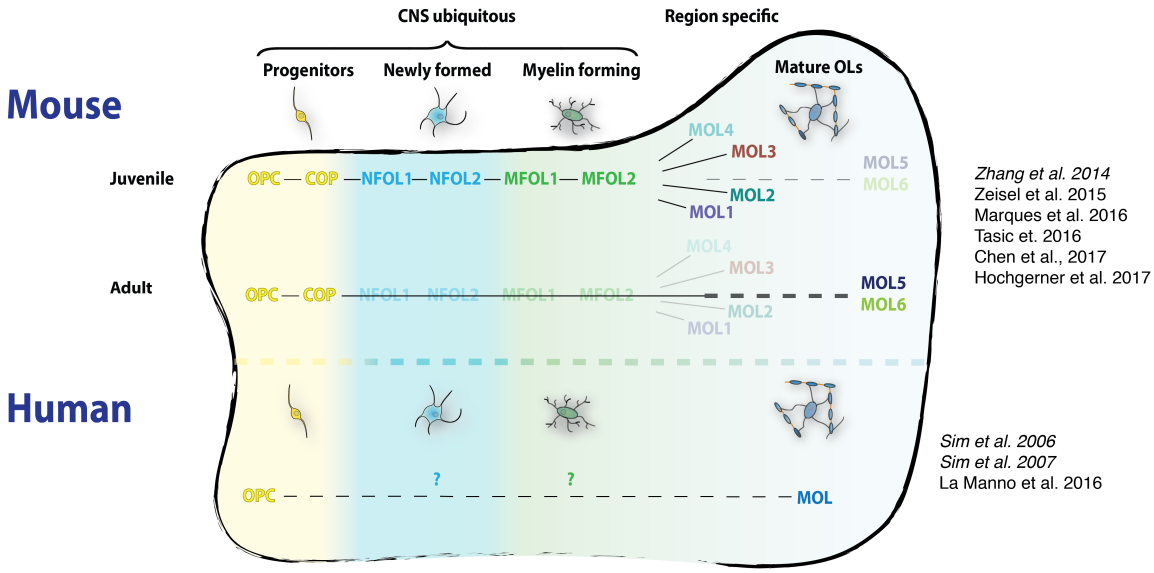

Figure 3 Overview of known oligodendrocyte subtypes in the brain.

The major role of Sox10 in OL maturation is evident in the binding pattern involving many myelin genes, as well as involving many binding sites facilitating the binding of multiple molecules of Sox10. Regulation of OL maturation by Sox10 requires high levels of Sox10 to be present in the cell, and additionally requires cooperation of factors such as Olig proteins and Nkx2.2, interestingly, the role of Olig1 in OL maturation is bigger whereas Olig2 has a specification dominated role [58,65]. Additional factors important for OL differentiation include a range of zinc finger transcription factors, such as Zfp488, and Zfp191 [77].

Recent single-cell sequencing approaches indicate expression of additional zinc finger protein families which might play a role in OL maturation. The myelin gene regulatory factor (Mrf/Myrf) is also expressed during OL differentiation, and recent studies have shown that this factor is crucial for proper OL differentiation and establishment of the myelinogenic program [78]. The exact gene regulatory program involving OL differentiation is not precisely known, novel studies will elucidate additional factors and interplay between them, including epigenetic regulators that serve as actors.

2 OLIGODENDROCYTES IN DISEASE: MULTIPLE SCLEROSIS

Multiple sclerosis (MS) is a chronic inflammatory multi-causal condition that has genetic as well as environmental aspects. It is presumed that MS is caused by an autoimmune response, and this insight has led to many disease-modifying drugs that act on parts of the immune system, as well as processes governing cell trafficking [79]. The ultimate result of the disease is demyelination of the brain and spinal cord, causing neurological damage and dysfunction. The disease has three main stages, the first is the pre-clinical stage, the stage in which the disease is triggered in an presumed multi-causal trigger involving genetics and environment. The second stage is the relapsing-remitting (RRMS) stage, in which symptoms of neurological dysfunction arise in detectible inflammatory lesions in the brain and spinal cord, the symptoms eventually resolve but patients often relapse. This stage can then worsen into a progressive clinical stage, in which the symptoms do not resolve but progressively worsen into secondary progressive MS (SPMS) or a primary progressive MS (PPMS) if the patient never enters the RRMS stage [80].

It is assumed that MS is a multicausal disease, even though this is not known. Additionally, it has been discovered that MS has genetic as well as environmental risk factors associated with it, and women are more affected than men [81]. A major driver of the disease seems to be the immune system, with the major cause being an autoimmune reaction mounted against cells of the oligodendrocyte lineage. Genome wide association studies have found several risk loci associated with MS, of which a certain haplotype of the HLA DRB1 gene (haplotype 1501) seems to be the most significant, however, over 200 risk factors have been found, highlighting the complexity of the disease. Aside from genetic risk, there is an increased risk associated with latitude and temperate climates.

Disease progression within MS is characterized by the occurrence of lesions in the CNS manifested by an interplay between oligodendroglial cells, and immune cells such as T-cells and B-cells, ultimately leading to axonal loss and an increase in MS symptoms in the patient. Animal models such as the encephalo-myelitis (EAE) mice model, in which an immune-response is mounted against a MOG-peptide fragment leading to autoimmune responses against mature oligodendrocytes, mimicking the “outside in” hypothesis model of multiple sclerosis, have been used to attempt to reproduce MS disease progression in mice, and have revealed an considerable role for the adaptive immune system. However, treatments that work on the EAE model, do not always work in MS, indicating an incomplete model [82].

Helper and cytotoxic T-cells are present in MS lesions, and limiting T-cells from accessing the CNS lessens the formation of new lesions considerably. Additionally, B-cells are also implicated in MS, and although known for the production of antibodies, targeted depletion of B-cells in MS seems to produce beneficial results before immunoglobulin depletion occurs, hence perhaps that other B-cell functions are of importance. Infiltration of macrophages and microglia in lesions is abundant and frequent, where phagocytic debris removal and cytokine depletion is common, however activated macrophages and microglia are also present and have implications in both lesion development and myelin regeneration [83–89].

OPCs are present during both development and adulthood, and are capable of differentiating into mature oligodendrocytes, restoring function to a great degree. However, at some point during MS progression OPCs are blocked from maturing into oligodendrocytes. Failing to remyelinate, MS lesions are not repaired and the disease quickly progresses to advanced MS stages such as progressive MS [90]. Efforts are underway to find ways of promoting OPC differentiation, without increasing inflammation, to restore remyelination capacity, and partially regain function.

2.1 BRAIN LESIONS IN MULTIPLE SCLEROSIS

Multiple sclerosis manifests itself in the brain resulting in lesions. MS lesions are insults to the brain characterized by a dysregulation of the blood brain barrier through effects such as local pro-inflammatory cytokines and chemokines such as interferon-γ (IFNγ), tumor necrosis factor-α (TNFα), and Interleukin-Iβ (IL1β). This exposure of the endothelium to pro-inflammatory cytokines disrupts the integrity of the blood brain barrier, and in turn enhances the adhesion of leukocytes, macropages, T-cells and B-cells to the endothelium and in turn enhances the occurrence of trans-endothelial migration, leading to an up regulation of major histocompatibility complex II (MHC II, oligodendrocyte loss, and neuronal degradation [91].

Lesions are characterized into a subset of regionally, histologically, and morphologically determined types, showing various differences in cell type composition, inflammatory activity, and demyelination. Here, the grey and white matter are distinct in how lesions present, but for both characterization generally entails describing the inner and outer borders of the lesion, where most lesion types have varying degrees of infiltration of immune cells, presence of reactive astrocytes, demyelination, myelin debris, and neuronal degeneration. The

earliest types of lesions to appear are white matter lesions characterized by active demyelination. At the lesion edge, the white matter lesion is crowded with debris clearing microglia and macrophages, and the lesion is infiltrated with lymphocytes, such as $CD8^+$ T cells and $CD20^+$ B cells, and to a lesser extent $CD4^+$ T cells [92,93]. In primary and secondary progressive MS inflammation is less frequent, leading to a different type of lesion composition termed inactive lesion. This type of lesion is characterized by clear borders with a low density of cells and axons, well defined demyelination, reactive astrocyte gliosis, and variable microglial activation but mainly at the outside of the lesion [92–94]. Other lesions include chronic active plaques and slow expanding lesion showing varying degrees of activity, but mainly at the lesion border. Chronic active lesions are more common in patients with longer disease duration or in secondary progressive MS, and are characterized by macrophages in both the lesion edge and the center, however most macrophages are found at the lesion edge. Slow expanding regions contain inactive centers, activated microglia at the lesion edge, and few macrophages where some myelin debris is detected, indicating a slow rate of demyelination [88,95]. Additionally, normal appearing white matter lesions are lof a diffuse nature, where axonal damage and demyelination occurs. These lesions are infiltrated with macrophages, acitivated microglia, as well as lymphocytes [94].

Grey matter lesions present differently to white matter lesions, and are usually demyelinated to a greater extent compared to white matter lesions. Grey matter lesions in the spinal cord and cortex might be exposed to different immunological stimuli compared to white matter. In fact, grey matter sites are predominantly found in sulci and are close to the meninges, exposing the tissue to insults originating from inflammatory infiltrates from the meninges. Conversely, although the extent of demyelination occurring in the grey matter extensive compared to the white matter, remyelination occurs at a faster rate [96]. Lesions in the grey matter are categorized in different parts, based on location. First off all, type III lesions are the most common grey matter lesions, located in subpial areas and frequent due to commonly occuring meningeal inflammatory infiltrates. Other grey matter lesion types are type I, at the cortico-subcortical border; type II lesions, which are perivenous lesions; and type IV lesions comprise the width of the cortex without reaching into the white matter [94].

2.2 GENETIC FACTORS AND THE INVOLVEMENT OF THE IMMUNE SYSTEM

Genetic factors play a role in MS, and genome wide association studies (GWAS) have identified more than 200 risk factors individually contributing to disease pathogenesis [97,98].

These factors are thought to have only minor individual contributions to the disease, however, familial studies reveal a heritable component. The age-adjusted risk is higher for siblings (3%) and parents (2%), and children (2%) than for second- and third-degree relatives. The risk for half-siblings is lower than for full siblings, and the risk for monozygotic twins is 35% [99]. The majority of risk factors found through genome wide association studies are related to the immune system. Such as several polymorphisms on the HLA genes, as well as TNFRSF1A a tumor necrosis factor receptor, several interferon regulatory factor (IRF) family genes. Several genes involved in the activation of proliferation of T cells are major polymorphisms involved in the disease. How these polymorphisms interact and contribute to the disease is lesser known. The work in **paper II** in this thesis details the expression of some of these genes near or on the loci in relation to the cell types found in both disease and healthy animals [98,100,101].

2.2.1 T cells

During maturation in the thymus, T cells are specified through recombination of gene segments in a somatic form of gene recombination involving a set of gene segments called variable (V) diversity (D) and (J) joining genes, generating a huge variety of T cell receptor specificities. While maturing, T cells start to express both CD4 and CD8 MHC co-receptors while being subjected to either negative or positive selection pressures, resembling a micro-evolutionary principle. Positive selection follows after MHC recognition of the TCR resulting in T cells that are positive for either CD4 or CD8 recognizing either MHCII or MHCI respectively. Strong self-MHC recognizing T cells are removed through negative pressures, and the autoreactive T-cells are released to the periphery, completing the process called thymic education. To prevent T cells that have developed an affinity for recognizing "self", meaning autoreactive T cells capable of generating autoimmune attacks, regulatory T cells (Tregs) positive for forkhead box protein 3 (FoxP3) are selected in the thymus. $CD4^+$ $FoxP3^+$ Tregs are selected for their strong recognition of self, however, these T cells are capable of suppressing other immune cells and thereby inducing immune homeostasis [102,103]. Before recognizing their cognate antigen, a T cell is referred to as naïve. However, the T cell quickly differentiates when it does recognize its antigen, through a cascade of signaling events leading to the differentiation and subsequent generation of both effector T cells, mediating and interacting with the environment as well as and antigen presenting cells, but at the same time generating memory T cells that remain and are reactivated when its antigen is encountered. T cells can polarize into different types of effector T cells depending on the co-stimulatory molecules in the environment as well as different kinds of cytokines. As stated

before CD4 and CD8 T cells recognize MHCII and I respectively, leading to a more mediating role for CD4 T cells such as releasing cytokines, and a direct killing role for CD8 T cells, although there exist exceptions to this grouping [104].

T cell involvement in MS is commonly thought to stem from the interplay of Tregs with $CD4^+$ and $CD8^+$ T cells, either wrongly activated or insufficiently regulated by Tregs. The common belief is that auto reactive T cells are activated and find their way into the parenchyma, crossing the blood brain barrier where they are reactivated by resident antigen presenting cells, leading to the establishment of an inflammatory milieu, recruiting additional T cells and macrophages resulting in a lesion. T cells have been found in early lesion, and several therapies rely on blocking this immune activation successfully thereby preventing the formation of new lesions, however, neurodegeneration is not prevented over the longer term [105,106].

The periphery and CNS of MS patients has been found to have increased numbers of $CD4^+$ IL-17 expressing T cells (T helper 17 cells or T_H 17) as well as $CD8^+$ T cells. Other types of T cell effector cells found in the CNS implicated in MS are INFγ secreting $CD4^+$ T cells (T_H 1 cells) and granulocyte macrophage colony-stimulating factor (GM-CSF). It is possible that these T effector cells are aberrantly activated in MS, however spontaneous differentiation of T_H 1 and T_H 17 does not readily occur. However, due to the presence of antigen presenting cells such as B cells and myeloid cells initiating pro-inflammatory interactions and releasing cytokines such as IL-12, IL-6, IL-23, IL-1β, and TNF, a pro-inflammatory milieu is created that might be permissive for aberrant activation of T_H 1 and T_H 17 cells [107–109].

As antibody levels in healthy individuals is generally low, MS patients often exhibit contrasting levels of antibody levels within the CNS. B-cells from MS patients have an elevated level of pro-inflammatory cytokine productions, with a deficiency in regulatory cytokines such as IL-10. The difference in B-cell activity between MS patients and healthy individuals might indicate that B-cells are capable of aberrant T cell activation. B cells expressing CD20 have become a target for anti-CD20 therapies, although antibody levels do not lower it is though that treatment contributes to a lowering of the possibility of aberrant activation [94,105].

SINGLE-CELL TRANSCRIPTOMICS

The advent of massively parallel next generation sequencing (NGS) technologies, pushed the envelope on the resolution of which biological phenomena could be studied. Recently, a new revolution in RNA-sequencing is pushing the resolution even further towards the unit of biology, the single cell. Single-cell transcriptomics techniques allow measurement of the transcriptomes of thousands of single cells using massively parallel sequencing techniques in combination with sophisticated barcoding techniques. Over time, many different approaches emerged that implement some form of barcoding techniques measuring 5'-end, 3'-end, or full-length RNA counts [110–116].

The analysis of tens of thousands of genes in thousands of cells requires compatible computation methods. A common approach would be to perform a linear dimensional reduction method such as PCA, followed by non-linear visualization methods such as t-SNE or UMAP [117,118]. After these dimensional reduction methods, data is clustered to generate groups of transcriptomically similar cells in the reduced space, or to generate a developmental or lineage ordering of cells along an imputed trajectory called "pseudo time" [119].

2.3 A BRIEF HISTORY OF A YOUNG FIELD: EMERGING TECHNIQUES

The field of single cell genomics has been a continuation of the trend to evolve sequencing workflows that retain more of the RNA sequences in massively parallel RNA sequencing approaches. However, as making incremental advances in sensitivity approached the unit of the cell, the nature of the resulting data fundamentally changed due to the leap forward in resolving power. Suddenly, tissues could be dissected into individual components, and new patterns emerged to be discovered through a plethora of analysis techniques. Since its advent in 2009 single-cell sequencing technologies have increased the throughput several magnitudes from tens of cells to a million cells using later drop-seq and micro-well technologies [110–112,114,115,120,121]. The first experiments by Tang et. al. relied on cell isolation using mouth-micropipette, enabling researches to isolate single cells through careful mechanical labor. The individual cells were lysed in separate tubes containing a detergent, where further step were made, generating single cell resolved transcriptomes compatible with massively parallel RNA sequencing. Subsequent advances have built on this technique through refining the protocol developed in the Surani lab [121].

Notable improvements to the single-cell workflow were added by a multitude of labs, CEL-seq and STRT-seq improved the technique by optimizing reagents and the implementation of cell barcodes [110,122], but later also introduced the concept of unique molecular identifiers (UMIs) [114]. UMIs are random sequences added to the pre-amplification cDNA construct, allowing molecular labeling of transcripts. UMIs are effective to remove noise generated by the amplification steps in the protocol. A drawback of the approach is the loss of full transcript length information, as only the 5' end of the mRNA is sequenced. However, these improvements allowed multiplexing to be performed early on in the workflow, pooling cells and thereby reducing the labor and reagents required to generate a library, since no additional reagents are needed in the individual lysis buffer. Aside from STRT- and CEL-Seq, SMART-seq and SMART-seq II do not use UMIs, instead optimizations were added to the single cell workflow while retaining full-length transcript reads. In spite of the added noise from the lack of UMIs in the workflow, SMART-seq II became the benchmark for sensitive single-cell sequencing [111,112]. At the same time microfluidics chip approaches such as the commercial fluidigm C1 chip, allowed cell capture and lysis on chip where small precise volumes could be delivered to reaction chambers on chip, dramatically reducing complexity of library set up, as most of the workflow was automated in a proprietary device [123,124].

One year later two different studies published a technique using microfluidics based on nano-sized aqueous droplets suspended in oil [115,125]. This technique could be set up in a lab using pumps and syringes and was relatively affordable compared to previous methods. Additionally, using droplet-based techniques, it was now possible to up to the range of thousands of cells per experiment [120]. The company 10x genomics have commercialized the droplet technique into their chromium platform, which has quickly become the standard method for generating single-cell sequencing libraries [126]. More recently, split-pool barcoding enables researchers to generate single-cell sequencing libraries of cells in the range of millions of individual transcriptomes. Split-pool ligation-based transcriptome sequencing (SPLiT-seq) relies on combinatorial barcoding, achieved through the pooled ligation-based RNA labeling, followed by a remixing of cells to one pool from which the second pooling will be performed, leading to rounds of labeling followed by recombination of pools of cells, ad infinitum. This technique can be performed using basic laboratory equipment (pipette, PCR machine) and can theoretically scale to the whole organ, or even organism level. The drawback of the technique (at the present) is that the technique is generally less sensitive compared to other high-throughput methods such as droplet-based techniques [127].

2.4 STRT-SEQ, SMART-SEQ II, AND CHROMIUM BASED CHEMISTRIES

The main difference between single-cell chemistries is the choice of full length or UMI based data generation. These protocols vary in the way adapter and primers are designed, dictating the information measured. Here the focus lies on the chemistries used in the papers in this thesis, in Paper I STRT-seq has been used; Paper II features Smart-seq2 data; Paper III features STRT-seq data; and finally in Paper IV, 10x genomics chromium V2 and V3 chemistries have been used. The STRT-seq method performs first strand synthesis of the RNA-template using a barcoded oligodT primer, but including a template-switching event. The template switching is achieved through the use of a specific reverse transcriptase, that, when finalizing the first strand synthesis will add a sequence of three cytosine nucleotides. The primer containing the UMI sequence is introduced, tagging the individual transcript, and as a result individual molecular counts can be detected. The UMI primer anneals to the previously generated 3' end (the 5' end of the RNA molecule) and facilitates elongation of the second strand. Amplification occurs through the use of a single primer located at both end of the cDNA amplicon. A cell barcode is then introduced though tagmentation with Tn5 transposase, and only the fragment is only sequenced from one end [114,128].

The Smart-seq2 protocol is comparable to the STRT-seq protocol, but generates full-length libraries. The main difference is found in the template switching where in contrast to STRT-seq, no UMI barcode is added in Smart-seq2. The library amplification steps are similar and both STRT-seq and Smart-seq2 generate libraries of full-length cDNA, however Smart-seq2 features both the P1 and P2 primers in the amplicon, additionally the tagmentation process is symmetrical in Smart-seq2, adding a barcode at both ends of the tagmented fragments making it possible to read the fragmented full-length transcript using next generation sequencing approaches. While Smart-seq2 theoretically retains the intron-exon structure, although with imperfect information, STRT-seq omitted this information, and only retains information about the 5' end of the transcript [111,112].

The first steps of the chromium library generation occur in a nano-liter volume, where the cell is suspended in oil captured in an aqueous bubble kept intact by hydrostatic forces. A hydrogel bead containing immobilized polydT primers carrying the UMI and cell-barcodes in one single sequence on its surface accompanies the cell, and lysis is initiated inside the droplet freeing RNA transcripts, causing them to bind to the hydrogel surface through annealing with the polydT primers on its surface. First strand synthesis occurs including template switching similar to STRT-seq and Smart-seq2 protocols. At this point the droplets surface tension it broken, mechanically, and the synthesized barcoded cDNA molecules are

pooled for subsequent full-length amplification. Here, the amplicon is cut enzymatically and the second read sequence is added through end-repair, A-tailing, and ligation steps, followed by PCR based addition of illumina compatible bridge amplification primers with a sample index added (see fig x) [126].

2.5 ANALYSIS OF SINGLE-CELL RNA SEQUENCING DATA

As a field, single cell resolved data analysis has only recently emerged, therefore standardization is hard and the growing number of analytic tools tailored to single cell data, in combination to the changing number of techniques and features to analyse (RNA-seq, ATAC-seq, Cut and Tag, bi-sulphite sequencing, gene dynamics inference such as velocity, gene regulatory network inference, lineage inference, and more) all require different approaches and tools, complicating standardization. Several initiatives started in recent years to bring standardized approaches or platforms in the hands of researchers. Platforms such as Seurat, and later Scanpy contributed to forming a platform of joint and standardized approaches in the programming languages R and Python respectively, Seurat added adoption of a multitude of tools over the years, including batch correction, negative binomial regression based normalization, integration of distinct datatypes, as well as many optimized versions of existing algorithms to facilitate data analysis with ever increasing numbers of cells [129–132].

2.5.1 Raw data and quality control

Raw data is normally processed through automated pipelines that align the raw RNA reads on a reference genome, and assign cell barcodes to cells in a process termed demultiplexing. Several pipelines exist and tools for this crucial step are expanding, some more popular pipelines are Cellranger, inDrops, emptyDrops, and Kallisto BUS tools [133–136]. Ultimately, data will be converted into a count or read matrix, depending on the use of UMIs or full-length sequencing technique. The matrix has the format of features (number of detected counts/reads per gene) and cells (barcodes). During the read alignment and cell detection, read counts have to be accurately added to cells. Complications in this process can arise from ambient RNA present in the whole sample or adhering to cells, but also from the accidental capturing of two or more cells in the same droplet or well. A complication with barcode reads can arise when barcodes are mutated during amplification steps or when read errors occur. To combat this, most barcodes are designed to be at least 2 or more Hamming distances away from other barcode sequences, so that when a registered barcode is not on the curated list, an algorithm can calculate the probability of the barcode origin through incorporating read

statistics such as base quality scores, after which a barcode can be reassigned to a valid barcode given it exceeds a certain probability threshold [126].

Filtering of the count matrix occurs through removing cells (barcodes) that show hallmarks of low quality transcriptomes. This is achieved by assessing the amount of counts a cell is assigned (count depth) as well as the amount of genes in which a count is detected (expressed genes), and the fraction of counts assigned to mitochondrial genes. Cells that underwent apoptosis during the experiment and preparation stages generally tend to exhibit low count numbers, as well as a low amount of genes expressed, the caveat being that small cells or quiescent cells might exhibit similar characteristics. In general, a dataset that captured many different cell types will also feature a large variation in cell sizes, whereby the total RNA content of the cells will fluctuate depending on the total cell volume, as so, care should be taken when setting thresholds for these values. The range at which cells are expected to lie in terms of number of genes expressed, is estimated to be around 500-5000 depending on the cell type, and thus size, but also the technology used. In the papers in this thesis, the number of genes was found to approximately be within this range. The third metric, fraction of mitochondrial counts can help differentiate better between high or low quality cells and is believed to measure cell stress, where cells that have a high proportion of mitochondrial counts tend to be in the process of apoptosis, or have a damaged membrane causing cytoplasm and cytoplasmic RNA to leak out while retaining the mitochondria, thus increasing the fraction of mitochondrial counts. However, this needs to be taken with care as a high fraction of mitochondrial counts might indicate respiratory processes, or in the case of oligodendrocytes, high numbers of mitochondria are expected to facilitate support for neurons.

2.5.2 Normalization and dimensional reduction

Normalization in single cell data requires different approaches compared to bulk-RNA seq data due to the intrinsically zero-inflated nature of the data. Next to normalization by total cell counts where one simply takes the size factor of the cell to be the sum of all the counts in the cell, after which the genes in the cell will be scaled proportional to the total library size of the cell. Other more specialized approaches have been developed, such as a method that uses a deconvolution approach, where cells are pooled by summing the expression of a group of cells together leading to fewer zeros, after which a size factor estimate is obtained through regression and recombination of cells in different pools [137]. Non-linear normalization techniques such as SCtransform developed by the Satija lab fit a non-negative binomial

model to the data where the residuals of the model are considered as the normalized gene expression data [130]. Common steps involving non-linear or linear models to normalize data include the correction of technical as well as biological effects, for example the linear model ComBat, originally developed for batch correction in microarray data, but adapted to single cell sequencing data, can perform both normalization and batch correction removing effects such as cell cycle or other biological or technical factors [138]. One caveat is unwanted variation in the data can be confounded with favored patterns, thus leading to the removal of potentially important variation.

After data normalization, variable features are selected and the dimensionality of the data is reduced. Standard feature selection methods include binning the features by mean expression magnitude and extracting the highly variable genes through extracting a threshold number of variable features per bin. The biological structure captured through single-cell sequencing is thought to be well approximated in a low dimensional space [139] which is commonly reduced through principal component analysis (PCA) [140]. PCA attempts to capture the variance in components, where each component describes the most variance not captured by previous components, leading to a ranking of components where the first component describes the most variance in the dataset diminishing with every subsequent dimension.

Non-linear dimensional reduction methods are mainly used for visualization, unlike PCA, the lower dimensional space is transformed in a way that does not conserve a uniform meaningful distance between data points, one of the great benefits of PCA reduced dimensional spaces. However, non-linear dimensional reduction techniques are excellent visualization tools, allowing the user to project the data into a 2 or 3 dimensional space, while retaining much of the complexity of the data, unlike PCA, which quickly loses complexity in the first two principal components when reducing the dimensions of a complex dataset [117,118].

2.5.3 Clustering

Grouping cells into biologically meaningful states is a problem that ultimately has many valid solutions, all producing differing results. The first clustering algorithms to be used for single cell data were adapted from classical machine learning approaches, algorithms such as k-means clustering attempt to partition the cells in a reduced space such that a user given *k* number of centroids are created, cells are the iteratively assigned to the nearest cluster centroid, after which centroids are updated, and cells are reassigned to the updated centroid positions [141]. Another classical method is hierarchical clustering, where cells are split according to certain criteria, such as single-, complete-, or average-linkage, and Wards

metric. In effect, any set of divisive rules can be used in the algorithm, additionally the splitting can be initiated from the top down (divisive) as well as bottom up (agglomerative), producing a dendrogram depicting the hierarchy of the data as a result [142].

K-Nearest neighbor graphs (KNN-graphs) are widespread in use in the single cell genomics field, due to their scalability and belief that KNN-graphs approximate the topology of the underlying data well. To create a KNN-graph, one has to assign a *k* number nearest neighbors to all data points most similar in the dataset, generating a network of data points. KNN-graphs lie at the basis of community detection algorithms, the most popular method for clustering single-cell data at present. The Louvain clustering algorithm is geared towards optimizing the amount of expected connections between cells based on the total connections of the cells, called modularity optimization, and hence does not need a set *k* [143]. The Louvain algorithm is efficient and scales well to large numbers of cells. Additionally, the optimized modularity function contains a resolution parameter, allowing more fine tuned control over the clustering. Moreover, a similar and faster community detection modularity optimizing algorithm called Leiden is increasing in popularity [144].

2.5.4 Batch correction

Single cell resolution data is now more accessible, and large atlas projects as well as vast collections of smaller experiment based datasets are publically available. This is not only a great contribution to data democratization, as well as a boon to hypothesis generation by combining originally disparate data into one, possibly revealing new biology. However, data generation through differing means results in unwanted variation in the data. In fact, even within a set of experiments, day-to-day variations in conditions and batches of cell preparation bring unwanted variation. In the medical fields, patient-to-patient variation adds in an unwanted manner, where cell types might exhibit different responses in disease conditions, age, background, lifestyle, or even simply differing circadian rhythms all contribute to making data comparison challenging. Batch correction is essential to facilitate researchers in making use of interesting patterns in data.

A variety of tools have been developed to correct batch effects in data. Traditional methods, adapted from micro-array analysis have been used, such as ComBat [138] are useful tools in removing batch effects, however, do not work in situations where complex batch correction is necessary, as these tools are inherently unable to address the zero inflated nature of single cell data. Currently the most popular and accessible tool by far is the correction method implemented in Seurat 3 [132]. This method is the merger of the pioneering mutual nearest

neighbor correction technique (MNN) [145] where correction vectors between matching cells in the different batches are found and corrected, and the previous implementation of Seurat 2 batch correction [131], where the canonical correlation analysis (CCA) implementation of finding a shared feature space between datasets is implemented to greatly facilitate the matching cells to subsequently calculate a set of correction vectors, effectively aligning the datasets to perform downstream analysis on. CCA batch effect correction can effectively find shared feature spaces between greatly varying datasets including the alignment of data from different organisms to study the evolution and conservation of cell types.

A caveat with neighbor finding tools is that similar yet distinct cell populations are at risk of being merged into one hybrid cell type. Recently, harmony [146] was developed, an algorithm that iteratively attempts to correct data in a lower dimensional space by using an adapted *k*-means algorithm with a soft-thresholded clustering that favors clusters with cells from multiple datasets and penalized several factors at once, user specified. Correction vectors are then iteratively applied until convergence, while exhausting relatively few computational resources. Making it a favorable algorithm to integrate over large datasets of millions of cells.

Several other integration methods have been developed, increasingly using deep neural networks. In sum, all techniques have unique weaknesses, and it is up to the user to choose the right approach fitting their data [147]. In **paper IV** I have adapted the Seurat 3 approach and extended it to integrate cells using hybrid additive distance matrices, aligning several samples and batches in one alignment event using PCA projection or combined with CCA, as well as to integrate different expression representations of the data.

2.5.5 Trajectory inference strategies and considerations

Grouping cell states together is informative to find specific gene expression profiles or markers associated to cell populations, even in the context of development, grouping cells into clusters can help reveal populations with certain characteristics such as stem cells, intermediate states and mature cell states. However, especially in developing tissues continuous processes are dominant, complicating the meaning of discrete cluster representational view of cells. As such, several techniques have been developed to identify continuous processes such as cell lineages from single cell data, attempting to capture cell states, including the possible origins and destinations along the cell landscape. Single cell RNA sequencing data is generally a snapshot capturing of dynamical processes in the context of development. In order to reconstruct dynamics of a developmental landscape from snapshot data several assumptions have to be made.

The first attempt to reconstruct developmental progressions from snapshot data is Monocle [148] introducing the concept of a reconstructed pseudotemporal ordering of cells termed pseudotime. To generate the pseudotemporal ordering in monocle, cells were analysed in a reduced dimensional space through independent component analysis, where a weighted graph is computed with cells as vertexes and edges represent distance in ICA space. To assign cells to the weighted graph a minimal spanning tree is computed and cells are assigned to the tree by proximity to their neighbors, minimizing transcriptional noise during the embedding process. Later iterations used a method termed reversed graph embedding as implemented in monocle 2 [149]. Another early method for trajectory inference is wanderlust, embedding cells in a KNN network and traversing the network through graph walks guided by waypoint cells designed to reflect the topology of the underlying developmental process. The walks along the graph are translated into pseudo-orderings and pseudotime [150]. Another notable early lineage inference method, diffusion pseudotime, adapts the use of diffusion maps, KNN graphs and random walk derived topology, to assess potential paths through the transcription based manifold, to derive a diffusion map. Distances in the diffusion map, together with diffusion components potentially inform about possible trajectories in the data, where the diffusion components themselves should reflect major axis of linear or non-linear transitions along a lineage if presented with developmental data [151].

Lineage branching points or other unconventional structures in the manifold such as cyclic trajectories due to the cell cycle or convergent paths taken in development are hard to extract. Strategies involve the modeling of complex iteratively updating tree based graphs such as the stream package [152], or monocle [149], or KNN based algorithms using pseudotime to direct edges in a asymmetric way such as palantir [153].

In order to accurately reconstruct dynamic processes it is vital to sample all states of the system. Dynamic processes might be transient and thus likely to be missed when low cell numbers are present, failing to capture these relatively rare cell states. Violations in assumptions can lead to differing results in the inference of trajectories, where Weinreb et. al. demonstrates through simulations of putative biological processes what might be limits in the KNN sampling derived method for inferring trajectories and cell dynamics from snapshot data [154,155].

In an effort to extract more information from snapshot data, La Manno et. al. exploited the process of splicing of RNAs in order to get an approximation of recent versus older RNA from a single cell. The assumption being that RNA reads carrying in them intronal sequences are less likely to have been spliced due to time constraints imposed by the transcription

machinery, therefore these unspliced transcripts likely reflect a later state of the cell, not yet manifested fully in the spliced transcriptome. Modeling the gene dynamics of single cells through the estimation of the degradation rate and the amount of spliced and unspliced reads, leads to the approximation of a steady state model of individual gene dynamics where RNA levels in a cell can be regarded as maintained when the cell produces exactly the amount of transcripts needed to overcome the degradation rate of older transcripts, being induced by the overcoming of the degradation rate, or down regulated by under-producing RNA, thereby lowering the pool of RNAs in the cell [156]. Recent frameworks such as scVelo developed by Bergen et. al. in the Theis lab have built on this concept to extend the model to a generalized kinetic model of transcriptional dynamics [157]. The RNA velocity principle allows for the inclusion of cell displacement estimation, which when applied to the KNN approach can generate asymmetric transition probability graphs likely to follow markovchain dynamics. In **paper IV** I have implemented this approach to estimate the direction of dynamic processes and used this in both lineage estimation efforts and gene regulatory network inference.

3 RESULTS

3.1 PAPER I

Using single cell RNA sequencing we revealed that the oligodendrocyte population, previously considered to be a functionally homogeneous population in the central nervous system (CNS), show distinct cell states in the adult and juvenile adult mouse brain.

Data obtained from 5072 cells of the oligodendrocyte lineage in the CNS sampled from 10 regions of the mouse juvenile, and adult were subjected to clustering using a custom clustering approach BackSPIN, reveal thirteen distinct populations of which 12 represent a continuous trajectory from OPC towards mature oligodendrocytes (mOLs). We could identify a narrow differentiation path from OPCs towards mOLs through pseudo-temporal analysis and *t*-Distributed stochastic neighbor embedding (t-SNE).

OPCs revealed themselves as being distinct from other OL cells, while still retaining expression of genes such as Fabp7 and Hes1 traditionally associated with radial glial cells, possibly indicating a short developmental window from the radial glia into OPCs. We found that differentiation committed oligodendrocyte precursor cells (COP) that lost the expression of OPC markers such as Pdgfra and Cspg4, as well as expressed Bmp4 and Gpr17, among others. We found specific markers for all intermediate states towards the mOLs, termed newly formed 1 and 2, myelin-forming oligodendrocytes 1 and 2. mOLs exhibited 6 states, of which 3 presented with unique markers. Additionally, we identified a population of pericyte like cells, which we termed vascular and lepto-meningeal cells were transcriptomically distinct from OPCs except for the expression of Pdgfra and Cspg4. Lineage tracing of the immature OL marker Itpr2 confirmed that they are generated by OPCs. Furthermore, we confirmed that complex wheel running of mice leads to a 50% increase of Itpr2 and Sox10 expressing oligodendrocytes compared to non-runners.

We found that among the juvenile and adult taken samples, OPCs and COPs were present in all regions indicating a common developmental trajectory. Furthermore, the proportion of cycling cells differed between the juvenile (30%) and adult (3%), additionally, we could find that regions differed in the proportion of oligodendrocyte subtypes, as well as regions in the brain of juvenile mice still populated by immature oligodendrocytes. Indicating different rates of maturation in the various brain regions.

Overall our results reveal an unknown heterogeneity underlies the oligodendrocyte lineage, with a strong implication for diseases related to oligodendrocyte such as multiple sclerosis

and other demyelinating lineages. Diving deeper into the data, I could reveal that oligodendrocytes express various non-coding RNAs along the pseudotime trajectory, indicating a somewhat strong regulatory control, especially since most of the non-coding RNAs seems to be conserved between mouse and human (Figure 4).

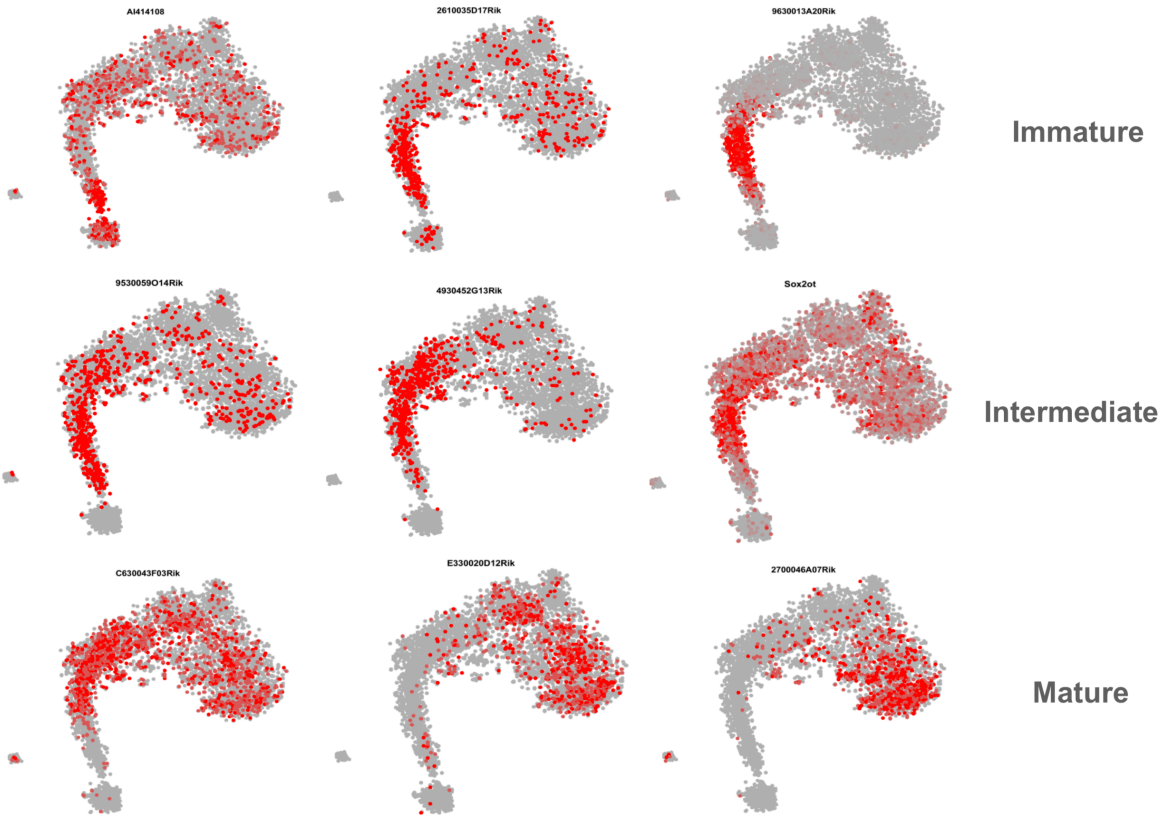

Figure 4. tSNE plots of the top differentially expressed ncRNAs along the pseudotime axis. Unpublished data

This study shows the transcriptional landscape of the oligodendrocyte lineage for the first time, and as such provides an coherent detailing of the differentiation path of the oligodendrocyte

3.2 PAPER II

The use of single-cell resolved transcriptomics techniques in combination with the experimental autoimmune encephalo-myelitis (EAE) mice model reveals that oligodendrocytes (OLs) are not passive in a disease context, but instead respond actively by antigen presentation, among other responses.

The EAE mouse model aims to mimic several aspects of multiple sclerosis (MS), in which an immune-response is mounted against a MOG-peptide fragment leading to autoimmune responses against mature oligodendrocytes, mimicking the "outside in" hypothesis model of multiple sclerosis, with the aim to elucidate any differences between control mice and EAE mice, so that we could investigate possible disease mechanisms in more detail.

To effectively resolve disease states and health states we developed an approach that leveraged a spatial auto-correlation metric, the Moran's I, using distances obtained from the diffusion map. This allowed spatial correlation of genes based on the manifold as approximated by the diffusion algorithm, to select the highest correlating genes on the manifold to generate a new diffusion map. We applied this filtering algorithm to significantly reduce the number of genes and improve the features of the manifold. We obtained an KNN network and Louvain clustering revealed oligodendrocyte precursor cells (OPCs), mature oligodendrocytes (MOLs), microglial cells, and VLMCs. Clustering within each level1 cluster revealed a total of fourteen distinct OL lineage populations, including 4 different OPC clusters, one committed OPC (COP) cluster, one newly formed OL (NFOL) cluster, and 8 MOL clusters. We found that 2 OPC populations were nearly exclusively enriched in the EAE model, as well as 5 of the MOL clusters, revealing disease specific cell states in the OL lineage.

We observed a marked difference between the EAE condition versus the control condition. Not only could we discern differences in mature oligodendrocytes, including a completely unique Plin4 expressing cluster with no obvious counterpart in the control condition, but clear differences also appeared in the OPC clusters. We found expression of immunoprotective and adaptive immunity genes in the disease specific oligodendrocytes.

To uncover disease factors, we performed non-negative matrix factorization. The number of factors (rank) we reduced by calculating the mutual information remaining between factors, rank estimation was performed using the "elbow" to estimate the best rank with the least average mutual information. The factor analysis revealed a number of factors relating to genes involved in the OL differentiation, progenitor state, and includes factors specific for

several clusters found during differentiation, such as COP and NFOLs, and several MOL clusters. Additionally, we discerned three factors associated with the disease model, these factors involve, an factor explaining MHC-I involved expression, as well as a factor specific for the Plin4 population uniquely identified in the disease condition, and lastly we found a factor describing expression of interferon response genes, Serpin family gene expression involved in immunoprotection, and highly surprisingly, MHC-II expression.

We investigated the possible expression of MHC-II genes in both OPCs and mature oligodendrocytes using RNA-scope and immune-histocytochemistry, in which we were able to validate the Plin4 population, as well as increased Serpina3n, and MHC-I expression in the EAE model, and in addition we could validate MHC-II expression in OL lineage cells, including human oligodendrocyte lineage cells expressing OLIG2 and the MHC-II complex.

Probing further using a published protocol to induce immune responses in OPCs, we could determine that OPCs can be induced into the MHC-II expressing state and that they subsequently exhibit phagocytic capabilities. Additionally, we performed co-culture experiments with T cells of 2D2 mice, which possesses a population of T cells specific for MOG 35-55 peptide, and could demonstrate a marked increase in proliferation of 2D2 memory T cells co-cultured in the presence of immune activated OPCs treated with interferon-γ and the MOG 35-55 peptide.

The phagocytic nature of the immunocompetent OPC and the strong evidence of MHC-II mediated antigen presentation suggests that the OPC exhibits a feedback mechanism with the T cell, eliciting a response, possibly indicating that the “inside out” hypothesis regarding the emergence of multiple sclerosis might be true, perhaps shifting the oligodendrocyte population back into view as a possible immunological player in this disease.

3.3 PAPER III

Bulk-RNA sequencing of cells from the brain and spinal cord, of time points E13.5, and postnatal day (P) 7 of the Pdgfra-H2B-GFP mouse model revealed great transcriptional differences between the E13.5 and P7 timepoints of the Pdgfra^{+} samples. The bulk data analysis revealed an increased expression of myelination associated genes in the P7 obtained sample, and conversely, the E13.5 time point was enriched for pattering genes, early neuroblast genes, and neuronal progenitor genes. Single-cell sequencing of Pdgfra-GFP+ E13.5, E17.5, and postnatal day (P) 7 cells using Smart-seq2 (1514 post-QC) from Pdgfra-H2B-GFP and Pdgfra-CreERT-RCE (LoxP-GFP) mice uncovered a multitude of populations explaining the great transcriptional differences observed between the timepoints in the bulk RNA-seq experiment.

The E13.5 time point revealed several groups of cells expressing Pdgfra in development. Detailed analysis of the genes expressed in these groups revealed populations of radial-glial like cells, pericyte-like cells, vascular and leptomenigeal-like cells (VLMCs), several populations of neuroblasts (NPs), and early choroid plexus-like cells. The majority of pattering genes found in the E13.5 bulk RNA sample could be deconvolved in the single-cell sequencing sample, where most of the genes could be attributed to the found populations, such as *Otx1/2* for cluster NP1a/b, *Lhx6* for NP3, *Lhx1*, *Ebf1-3*, *Neurod2/6*, for NP2, eVLMC/PLC/Radial-Glia (*Meox2*, *Wwtr1*, *Hes1*). Additionally, great differences between spinal-cord and brain derived cells became apparent, as cells from the spinal-cord expressed many head/tail axis-associated genes from the Hox gene-cluster.

The P7 time point consisted almost entirely of cell clusters of OPCs, OL-lineage cells, and VLMC cells. The OPC cluster could be split in two clusters, and, although the clusters are very similar, we could find some differentially expressed genes such as Resp18, Fos and Egr1. However, it is not clear if these differences are due to stress related responses or true biological variation. We also found committed oligodendrocyte precursor cells (COPs) and newly formed oligodendrocytes (NFOLs), forming a gradually transitioning trajectory towards maturing OLs.

To elucidate probable relationships between the cells, we created an approach to generate a k-nearest-neighbor (KNN)-network of cells, single-cell nearest neighbor network embedding (SCN3E), by leveraging the diffusion mapping algorithm, a non-linear dimensional reduction technique, robust to the high dimensionality problem and dropouts. A diffusion map is in fact

a transition matrix approximating the relative similarities between cells, while at the same time it can be considered a markov-chain. Connectivity between two points, or, the transition probability, is defined as the probability of landing on that neighboring point in a one-step random walk.

To approximate the manifold better we clustered the data into 100 clusters, for each of which a pseudotime was calculated using the R-package DPT-destiny, turning each cluster into a beacon in the manifold from which a pseudotime "signal" would emerge. For each cluster we chose a cell that was spatially the most distant from the previous "beacon", ultimately resulting in 100 pseudotimes with locations and relative orderings for each cell. I then performed principle component analysis (PCA) on the pseudotime matrix, effectively extracting the main features of the manifold through exploiting the fact that dissimilar manifold in regards to pseudotime flow would be emphasized in components of variation.

To make the KNN-algorithm robust we attempted to scale the number of nearest neighbours depending on the topology of the manifold by limiting the number of nearest neighbours to only allow the neighbours corresponding to the 1% closest distance values. Pearson correlation between the neighbours, further dictates the weights of the edges in the final network. We then performed Louvain clustering on the KNN graph.

Using these approaches, we could detect a sub-cluster of a neural progenitor population expressing *Olig1/2*, *Ptprz1*, *Bcan*, *Rfx4*, and *Nes*, directly adjacent to the identified OPC population in the KNN network, as well as the radial glial like population, possibly constituting pre-OPCs (Figure 5). Further inspection in previously published data of the embryonic midbrain including several timepoints during development (La Manno, et. al. 2016) using the algorithms described above revealed a similar population within the radial glial populations, expressing all markers above, thus indicating a cell state of committed glia precursor cells.

We then delineated different waves of OPCs a using Pdgfra-CreERT-RCE (LoxP-GFP) mouse. Injection of tamoxifen at E12.5 resulted in GFP expressing cells depending on the expression of Pdfra, allowing us to isolate the first wave OPCs from later wave OPCs. Cells were collected at P7-8 and analysed. We found that not only did the E13.5 timepoint give rise to OPCs we also found VLMC/pericytes like cells generated from the same cell subpopulations expressing Pdgfra at these early timepoints. However, no significant differences between the different developmental waves of OPC generation could be found from a transciptomic perspective using MAST, a hurdle-model based differential expression

package. OPCs from spinal cord and brain did not exhibit many significant differences, except Hoxc8 and Plp1 genes, we did see other genes differentially expressed although not significantly. Thus, we concluded that regardless of the origin of OPCs, be it spinal cord or brain, as well as different waves of OPC generation during development, convergence towards a transcriptionally homogeneous state occurs upon attainment of the OPC identity.

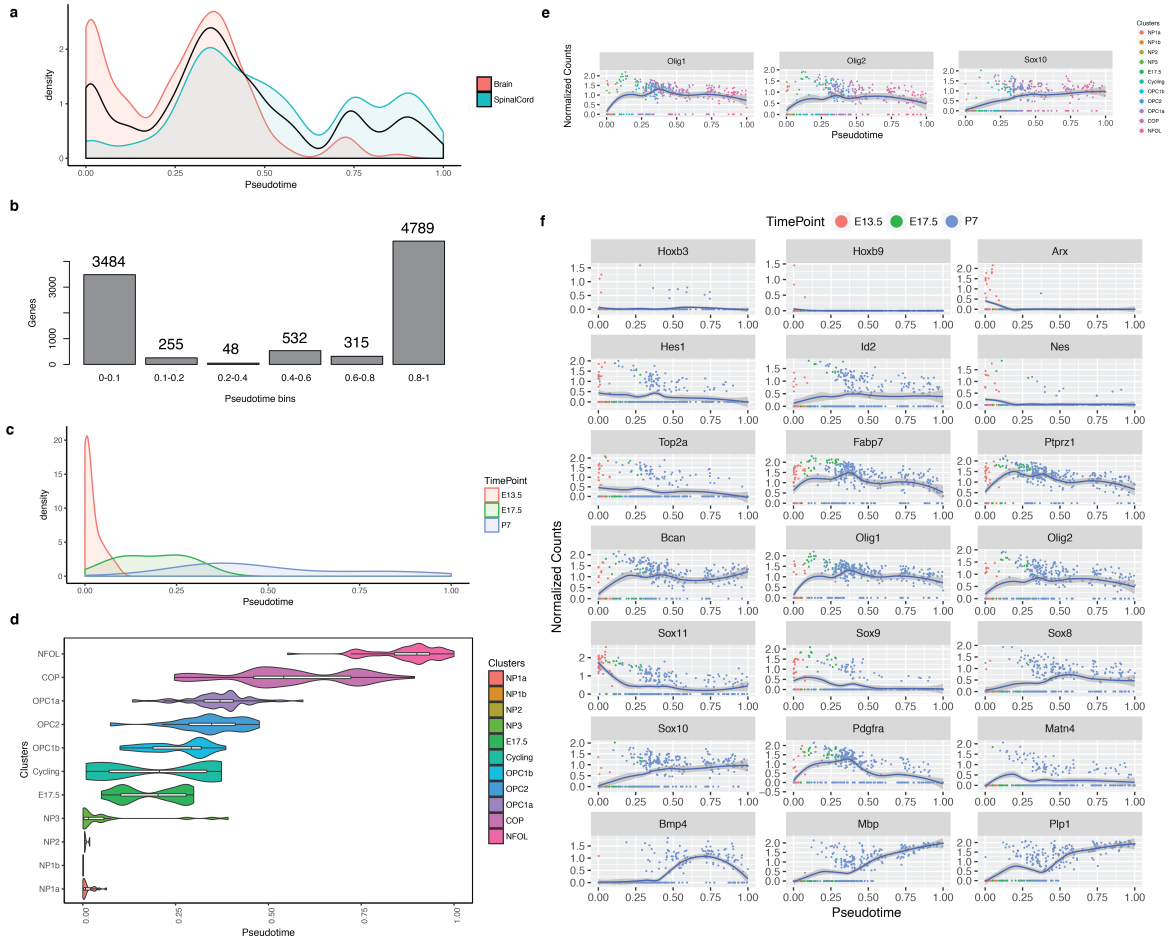

Figure 5 | Pseudotime trajectory reveals progenitors of the OL lineage in the E13.5 timepoint. a, Density plot showing differences in Spinal Cord and Brain derived cells along pseudotime trajectory. The black line depicts the density of all cells. b, Number of genes at least 2 fold differentially expressed between Spinal Cord and Brain. Cells are pseudotime binned according to density peaks along pseudotime. c, Density plot showing the distribution of the three different timepoints. d, Violin plot illustrating the distribution of the cell clusters along pseudotime. e-f, Pseudotime plot depicting expression along pseudotime (e) Scatterplot showing the expression patterns along pseudotime for essential oligodendrocyte lineage genes. (f). Scatter plot along pseudotime. Unpublished

3.4 PAPER IV

We dissociated human fetal brain tissue from week 7-10 post conception after which we performed single-cell RNA sequencing on the 10x genomics platform. We obtained approximately 25 000 cells from an unbiased dissociation.

Analysis reveals several clusters of radial glial cells, as well as, endothelial cells, several kinds of neuroblasts such as excitatory neurons, inhibitory neurons, motor neurons and, additionally, several clusters of glial cells, such as glial progenitor cells, SPARCL1 expressing early astrocytes or transit amplifying cells, and surprisingly, OPCs expressing *PDGFRA*, *SOX10*, *OLIG1*, and *OLIG2*. Interestingly, the pre-OPC population identified in the mouse data of **Paper II** was present in the human data and seems to be the precursor population for glial cells including OPCs, according to the manifold approximation and KNN-network embedding. We calculated a RNA velocity inspired diffusion map on which we attempt to trace the preceding populations possibly generating OPCs. We stratified the obtained collection of progenitors along a pseudotime into 4 bins, after which we defined the conditional mutual information (CDI) of transcription factors between spliced and unspliced counts, we then generated a network of CDI measured weights across all progenitors. Furthermore, we scored the top regulated genes and their putative regulator in individual pseudotime bins by defining the RNA velocity associated gene shifts for the cluster, retaining only the genes that show velocity shifts. This resulted in a list of transcription factors and possible targets associated to each population. Revealing the transcriptional actors in the different populations allowed insight into the possible key transitions in regulatory repertoire needed to undergo cell state transition into ultimately OPCs.

4 CONCLUSIONS AND POINTS OF PERSPECTIVE

The work presented in this thesis represents the accumulated work of several years in the field of neuroscience, where we have attempted to dive deeper into the complexities of brain development, disease, and brain composition. **Paper I** contributed greatly to the field of oligodendrocyte biology and neurodevelopment, where we revealed the oligodendrocytes in unprecedented detail for the first time. The work in **Paper II** revealed that the oligodendrocyte lineage seems to communicate and interact with the immune system rather than be subject to it. In **Paper III** we show that OPCs are a seemingly homogeneous population of cells, regardless of their origins. Additionally we show that OPCs as well as VLMCs are generated from the E13.5 lineage expressing Pdgfra, and continue to show what precursor cells are most likely to be precursors to the OPC cell state. In **Paper IV** we attempt to disentangle the complicated structures of the developing human forebrain glial lineage. We sequence post conception week 8 – 10 in the human forebrain, which is to our knowledge the first time that this time window has been captured in the resolution we provide. Furthermore, for the first time, we detect OPCs at week 8 in development where we continue to dissect the possible origin of these early human OPCs.

The understanding of developmental processes in the brain and throughout the organism necessitates the reconstruction of the evolutionary process that formed it. Cell and developmental biology is increasingly becoming an exact science, reaching the point where the modeling of a single cell might become possible in the near future.

The recent emergence of multimodal single cell data is necessary to capture all facets of cell dynamics, the regulation of which depends on a multitude of factors, such as the 3d conformation of the genome in the cell. The spatial context in which the chromosomes operate dictates how genes, promoters, enhancers, and non-regulatory elements interact together. Recent advances have transformed previously bulk techniques such as ATAC-seq, chromatin conformation capture techniques such as Hi-C, as well as Cut and Tag which allows researchers to probe the chromatin of a single cell. The field of microscopy is making similar strides and is currently the most promising way to study dynamical biological phenomena at the single-cell level.

The digitization of biology moves the field slowly but surely into a position to merge with more exact sciences, leading to more and more research being conducted as computation. The complexity of multimodal data will need more advanced algorithms to find patterns, and make accurate predictions. Recent advances in machine learning have not yet made serious

impact in the way biological research is conducted today. Initiatives such as the human cell atlas help spearhead accessible and open data, broadening the so far narrow cracked opening of the door to all researchers in the biological fields. Medicine can benefit greatly from evolving models of the cell. On can envision applications brought about by the dynamic information from tools such as RNA velocity and metabolic labeling. When completed by lineage tracing such as the sophisticated techniques that CRISPR can bring, a new paradigm can enter the stage. As far as my work will go into the future, it seems straightforward that the learning algorithms developed in the past five years can have both input and output in terms of RNA velocity data, when these algorithms attempt to converge on the unspliced transcriptomic state or any other simultaneously measured modalities of the same cell, we could asses if these predictions are generalizable and perhaps predictive of perturbations. Such a system could in theory allow *in silico* perturbations to increasingly be a viable alternative to experiment, and thus capable of aiding in the generation of scientific hypotheses.

5 ACKNOWLEDGEMENTS

First of all thank you **Gonçalo!** For all the opportunities and advice you have given me. I have learned a lot over the years. Thank you for allowing me to be independent and going off in a tangent, but also for the endless patience, and for all the trust right from the beginning. It was a wonderful time.

Carsten, thank you for being my co-supervisor and letting me be a teacher for the first time, and in Japan of all places! It was a great experience to stand in front of a class. **Sten,** thank you for being my co-supervisor, even though I did not come by with a lot of questions, I am confident you would have answered them all. Also your labs online atlases were a treasure trove!

Thank you **Rahul Satija** for the honor of being my opponent. Thank you **Thora Káradóttir** and **Thomas Perlmann,** for giving me your expertise as board members, and thank you **Eva Hedlund** for all the above and for coordinating. Thank you **Jens Hjerling-Leffler**, for the willingness to step in as chairperson in a socially distant time.

Thank you **Ernest Arenas** for the chats and always being ready to give a speech at the dissertation dinners. Thanks to all the rest of the PI's at MolNeuro, **Patrick Enfors, Ulrika Marklund,** and **Per Uhlén,** for making the department such a great place to be, full of collaboration and discussion. Thank you **Alessandra Nanni** for all the assistance over the years. Thank you **Ahmad** and **Johnny** for being the steady hand in the lab.

And then to all my colleagues in the GCB lab! Former members, thank you **Sam**, for always being straight forward, and the occasional glimpse of your dark humor, **Sueli**, for tirelessly getting me fit for the lab. **Ana,** for giving me the opportunity to contribute to that great paper. **Elisa**, for your showing of dedication, character, and good chats.

Thank you **Tony** for your unwavering belief in my understanding of Spanish, and for letting your feelings for your emails to me hitting a wall slide. **Marek**, your climbing is legendary. Your introduction to the sport taught me that feet are strictly optional to use! **Eneritz,** for the chats in our far away office that have always been good, and thank you for being the data protector! **Fabio,** for keeping those board games coming. **Mukund,** for telling me all about fasting. **Petra,** see you next time at the gym! **Mandy**, for always being a step ahead in the administrative things so I can ask you. **Leslie**, for becoming a dedicated educator during labmeetings. And thank you to the other great members and ex-members of the group, **Tanja, Chao, Bastien, Shreya, Karl, Prachi, Cassandra, Florian, Leipa, Negi, and Carmen.**

Alessando. How you always find those new places to eat is beyond me, but they were always good! Thank you for hanging out with me in my short visit in time in Japan, having coffee ready, for never being able to sneak up on me because you are just always audible from a distance, always being willing to dive into scientific thought experiments, all that home cooking, and the open ear. It was a great time! Also, if I could have a PhD party now, I'll be

on that boat already! **Ibrahim**, you are the only one that made me worry about the size of the portions when going out to eat. Thank you for giving me meaning back to the word potato, digging my coffee, and general willingness to have a chat and chill to discuss the real things in life, next to finding this guy. **Khom,** thanks for always being around but never in sight, it's good to know you are out there. But when I see you its always good, and I get free book recommendations. **Lars**, I'm amazed at your ability to both creatively find your way through steps in making a microscope, and through steps on the dance floor alike. **Alejandro,** thank you sir! Thank you **Viktoria**, for indulging me in chatting, even if it stretched into your bedtime. And for only taking one corner of the couch so I can have all the rest ☺.

Thank you **Emelie,** for sharing that great trip to Japan, getting excited about Pikachu chopsticks, and for sharing the pain of that oyster incident. **Kimberly,** for not only tolerating the continued noise my computer made when crunching numbers, but also that the office got turned into a chat factory sometimes. **Peter**, for al the help you provide, and going to extra lengths. **Gioele**, for all the late night coffee talks we had discussing physics, cell types and more. I miss it. **Miri**, for the occasional nice chat. **Emilia,** my most loyal customer, and for convincing me that I can turn semi-professional in the coffee business after my PhD. **Elin,** for always bringing the good vibes, and for returning the threat of beating each other in a game of badminton, but never actually going. **Camiel**, for making the trip to Japan awesome, the occasional nugget of history wisdom. **Jokubas**, for that fresh humor so dry it stings. **Natalie**, for all the fun we had together. **Simone**, for the occasional Italian swearing while coding. **Shanzheng** for the great questions at any seminar, and your wonderful attitude. **Lauri**, for always striving to push the conversation. **Hermany**, for your spirit, attitude to life and giving me a place to escape the horrors that is KI-housing locations. **Carmen**, for the warm reception and making me feel at home the moment I joined MolNeuro. **Dagmara**, for still showing up at my half-time, and upholding the level of sarcasm in the unit. **Alejandro,** for reading this far. **Kasra**, for your chillness and literally being a mini golf court owner. **Fatima**, for all the amazing cake creations I've seen over the years. **Ana J** for helping me straighten out the champagne glass incident. **Enrique**, for the patience when I came to you when learning coding, **Daniel** for being the invisible force keeping the old MolNeuro clean. **Karol** for the endless enthusiasm translating into decibels. **Alejandro**, letting me help you scare people in your basement. **Markus**, for all the silly jokes too many too count. **Job**, for all the great talks about everything and nothing. **Boris**, I'm still wanting to see you perform on stage. **Shigeaki**, for the openness and just doing awesome things in science, **Ivar**, I still remember you and Lauri eating those Swedish specialties. **Carolina**, for always being genuine. **Mingdong**, for the occasional "passerby" chat, they were good, **Connla**, for always walking that fine line between random talk and utter nonsense. And thank you **Razieh**, **Lisa**, **Milda**, **Anastassia**, **Leonardo**, **Lisbeth**, **Amparo**, **Roland, Josè, Jinan, Dmitry, Lijuan, Amit, , Alca, Anna Muñoz, Michael, Laura, Anneke, Dongoh, Jussi, Daohua, Göran, Hannah, Nathan, Yishu** for making or having made MolNeuro such a great place to do research.

Thanks my friends from home, **Lars**, **Johan**, **Simone**, **Imke**, and **Tsz Shan**.

So much thanks to **Ka Wai** supporting me tirelessly through writing this book.

Thanks to my **brother,** my family, my **Mom** for putting up with me doing a PhD so long and far from home. My **Dad**, for always supporting me in what I do and giving me so many different news bits involving my work.

My **grandma**, who is always my biggest believer.

6 REFERENCES

1. Achim, K. & Arendt, D. Structural evolution of cell types by step-wise assembly of cellular modules. *Curr. Opin. Genet. Dev.* **27**, 102–108 (2014).

2. Kristan, W. B. Early evolution of neurons. *Curr. Biol.* **26**, R949–R954 (2016).

3. Putnam, N. H. *et al.* Sea anemone genome reveals the gene repertoire and genomic organization of the eumetazoan ancestor. *Siencemag* 86–94 (2007). doi:10.1126/science.1139158

4. Srivastava, M. *et al.* The Trichoplax genome and the nature of placozoans. *Nature* **454**, 955–960 (2008).

5. Srivastava, M. *et al.* The Amphimedon queenslandica genome and the evolution of animal complexity. *Nature* **466**, 720–726 (2010).

6. Ryan, J. F. *et al.* The genome of the ctenophore Mnemiopsis leidyi and its implications for cell type evolution. *Science (80-.).* **342**, (2013).

7. Arendt, D. *et al.* The origin and evolution of cell types. *Nat. Rev. Genet.* **17**, 744–757 (2016).

8. Regev, A. *et al.* The human cell atlas. *Elife* **6**, 1–30 (2017).

9. Alié, A. & Manuël, M. The backbone of the post-synaptic density originated in a unicellular ancestor of choanoflagellates and metazoans. *BMC Evol. Biol.* **10**, (2010).

10. Mitchell, M. J. & King, M. R. NIH Public Access. **11**, 1–23 (2014).

11. Ryan, T. J. & Grant, S. G. N. The origin and evolution of synapses. *Nat. Rev. Neurosci.* **10**, 701–712 (2009).

12. Chiu, J., Desalle, R., Lam, H., Meisel, L. & Coruzzi, G. Molecular Evolution of Glutamate Receptors : A Primitive Signaling Mechanism that Existed Before Plants and Animals Diverged. 826–838 (1998).

13. Chen, G. Q., Cul, C., Mayer, M. L. & Gouaux, E. Functional characterization of a potassium-selective prokaryotic glutamate receptor. *Nature* **402**, 817–821 (1999).

14. Vickaryous, M. K. & Hall, B. K. Human cell type diversity, evolution, development, and classification with special reference to cells derived from the neural crest. *Biol. Rev.* **81**, 425 (2006).

15. Arendt, D., Tosches, M. A. & Marlow, H. From nerve net to nerve ring, nerve cord and brain-evolution of the nervous system. *Nat. Rev. Neurosci.* **17**, 61–72 (2016).

16. Marlow, H., Matus, D. Q. & Martindale, M. Q. Ectopic activation of the canonical wnt signaling pathway affects ectodermal patterning along the primary axis during larval development in the anthozoan Nematostella vectensis. *Dev. Biol.* **380**, 324–334 (2013).

17. Sebé-Pedrós, A. *et al.* Cnidarian Cell Type Diversity and Regulation Revealed by Whole-Organism Single-Cell RNA-Seq. *Cell* **173**, 1520-1534.e20 (2018).

18. Marlow, H., Roettinger, E., Boekhout, M. & Martindale, M. Q. Functional roles of Notch signaling in the cnidarian Nematostella vectensis. *Dev. Biol.* **362**, 295–308

(2012).

19. Watanabe, H. *et al.* Sequential actions of β-catenin and Bmp pattern the oral nerve net in Nematostella vectensis. *Nat. Commun.* **5**, 5536 (2014).

20. Flici, H. *et al.* An Evolutionarily Conserved SoxB-Hdac2 Crosstalk Regulates Neurogenesis in a Cnidarian. *Cell Rep.* **18**, 1395–1409 (2017).

21. Streit, A., Berliner, A. J., Papanayotou, C., Sirulnik, A. & Stern, C. D. Initiation of neural induction by FGF signalling before gastrulation. *Science (80-.).* **406**, 6–10 (2000).

22. Groves, A. K. & Labonne, C. Setting appropriate boundaries : Fate , patterning and competence at the neural plate border. *Dev. Biol.* **389**, 2–12 (2014).

23. Meulemans, D. & Bronner-fraser, M. Amphioxus and lamprey AP-2 genes : implications for neural crest evolution and migration patterns. **4962**, 4953–4962 (2002).

24. Nichane, M., Ren, X., Souopgui, J. & Bellefroid, E. J. Hairy2 functions through both DNA-binding and non DNA-binding mechanisms at the neural plate border in Xenopus. *Dev. Biol.* **322**, 368–380 (2008).

25. Khudyakov, J. & Bronner-fraser, M. Comprehensive Spatiotemporal Analysis of Early Chick Neural Crest Network Genes. 716–723 (2009). doi:10.1002/dvdy.21881

26. Grocott, T., Tambalo, M. & Streit, A. The peripheral sensory nervous system in the vertebrate head : A gene regulatory perspective. *Dev. Biol.* **370**, 3–23 (2012).

27. Sato, S. *et al.* Conserved expression of mouse Six1 in the pre-placodal region (PPR) and identi fi cation of an enhancer for the rostral PPR. *Dev. Biol.* **344**, 158–171 (2010).

28. Giacomello, E., Faraldo, M. M., Thiery, J. P. & Broders, F. Cloning and Characterization of Three Xenopus Slug Promoters Reveal Direct Regulation by Lef / ᴺʟ -Catenin Signaling *. **276**, 30350–30358 (2001).

29. Yook, J. I. *et al.* LETTERS A Wnt – Axin2 – GSK3 β cascade regulates Snail1 activity in breast cancer cells. **8**, (2006).

30. Betancur, P., Bronner-fraser, M. & Sauka-spengler, T. Assembling Neural Crest Regulatory Circuits into a Gene Regulatory Network. (2010). doi:10.1146/annurev.cellbio.042308.113245

31. Takahashi, Y., Sipp, D. & Enomoto, H. Tissue Interactions in Neural Crest Cell Development and Disease. **341**, 860–864 (2013).

32. Wilson, L. & Maden, M. The mechanisms of dorsoventral patterning in the vertebrate neural tube. *Dev. Biol.* **282**, 1–13 (2005).

33. Jessell, T. M. NEURONAL SPECIFICATION IN THE SPINAL CORD : INDUCTIVE SIGNALS AND TRANSCRIPTIONAL CODES. **1**, 20–29 (2000).

34. van Straaten, H. W., Hekking, J. W., Thors, F., Wiertz-Hoessels, E. L. & Drukker, J. Induction of an additional floor plate in the neural tube. *Acta Morphol. Neerl. Scand.* **23**, 91–97 (1985).

35. Lobjois, V., Benazeraf, B., Bertrand, N., Medevielle, F. & Pituello, F. Specific

regulation of cyclins D1 and D2 by FGF and Shh signaling coordinates cell cycle progression, patterning, and differentiation during early steps of spinal cord development. *Dev. Biol.* **273**, 195–209 (2004).

36. Choi, B. H. & Lapham, L. W. Radial glia in the human fetal cerebrum: a combined Golgi, immunofluorescent and electron microscopic study. *Brain Res.* **148**, 295–311 (1978).

37. Choi, B. H. Prenatal gliogenesis in the developing cerebrum of the mouse. *Glia* **1**, 308–316 (1988).

38. Campbell, K. & Gotz, M. Radial glia: multi-purpose cells for vertebrate brain development. *Trends Neurosci.* **25**, 235–238 (2002).

39. Gotz, M. & Huttner, W. B. The cell biology of neurogenesis. *Nat. Rev. Mol. Cell Biol.* **6**, 777–788 (2005).

40. Imai, F. *et al.* Inactivation of aPKClambda results in the loss of adherens junctions in neuroepithelial cells without affecting neurogenesis in mouse neocortex. *Development* **133**, 1735–1744 (2006).

41. Konno, D. *et al.* Neuroepithelial progenitors undergo LGN-dependent planar divisions to maintain self-renewability during mammalian neurogenesis. *Nat. Cell Biol.* **10**, 93–101 (2008).

42. Kuo, C. T. *et al.* Postnatal deletion of Numb/Numblike reveals repair and remodeling capacity in the subventricular neurogenic niche. *Cell* **127**, 1253–1264 (2006).

43. Noctor, S. C., Martinez-Cerdeno, V., Ivic, L. & Kriegstein, A. R. Cortical neurons arise in symmetric and asymmetric division zones and migrate through specific phases. *Nat. Neurosci.* **7**, 136–144 (2004).

44. Mayer, C. *et al.* Developmental diversification of cortical inhibitory interneurons. *Nature* **555**, 457–462 (2018).

45. Shimamura, K., Hartigan, D. J., Martinez, S., Puelles, L. & Rubenstein, J. L. Longitudinal organization of the anterior neural plate and neural tube. *Development* (1995). doi:8575293

46. Shimamura, K. & Rubenstein, J. L. Inductive interactions direct early regionalization of the mouse forebrain. *Development* (1997). doi:10.1016/0092-8674(93)90249-p

47. Houart, C., Westerfield, M. & Wilson, S. W. A small population of anterior cells patterns the forebrain during zebrafish gastrulation. *Nature* (1998). doi:10.1038/35853

48. Houart, C. *et al.* Establishment of the telencephalon during gastrulation by local antagonism of Wnt signaling. *Neuron* (2002). doi:10.1016/S0896-6273(02)00751-1

49. Tole, S. & Hébert, J. Telencephalon Patterning. in *Comprehensive Developmental Neuroscience: Patterning and Cell Type Specification in the Developing CNS and PNS* (2013). doi:10.1016/B978-0-12-397265-1.00018-6

50. Rowitch, D. H. & Kriegstein, A. R. Developmental genetics of vertebrate glial-cell specification. *Nature* **468**, 214–222 (2010).

51. Fuccillo, M. Temporal requirement for hedgehog signaling in ventral telencephalic patterning. *Development* (2004). doi:10.1242/dev.01349

52. Huang, X., Litingtung, Y. & Chiang, C. Region-specific requirement for cholesterol modification of sonic hedgehog in patterning the telencephalon and spinal cord. *Development* (2007). doi:10.1242/dev.000729

53. Gunhaga, L. *et al.* Specification of dorsal telencephalic character by sequential Wnt and FGF signaling. *Nat. Neurosci.* (2003). doi:10.1038/nn1068

54. Corbin, J. G. Combinatorial function of the homeodomain proteins Nkx2.1 and Gsh2 in ventral telencephalic patterning. *Development* (2003). doi:10.1242/dev.00717

55. Theil, T., Aydin, S., Koch, S., Grotewold, L. & Rüther, U. Wnt and Bmp signalling cooperatively regulate graded Emx2 expression in the dorsal telencephalon. *Development* (2002). doi:10.1093/emboj/16.13.3797

56. Kessaris, N. *et al.* Competing waves of oligodendrocytes in the forebrain and postnatal elimination of an embryonic lineage. *Nat. Neurosci.* **9**, 173–179 (2006).

57. Ostrem, B., Di Lullo, E. & Kriegstein, A. oRGs and mitotic somal translocation — a role in development and disease. *Curr. Opin. Neurobiol.* **42**, 61–67 (2017).

58. Lu, Q. R. *et al.* Common developmental requirement for Olig function indicates a motor neuron/oligodendrocyte connection. *Cell* (2002). doi:10.1016/S0092-8674(02)00678-5

59. Gokhan, S. Combinatorial Profiles of Oligodendrocyte-Selective Classes of Transcriptional Regulators Differentially Modulate Myelin Basic Protein Gene Expression. *J. Neurosci.* (2005). doi:10.1523/JNEUROSCI.1850-05.2005

60. Petryniak, M. A., Potter, G. B., Rowitch, D. H. & Rubenstein, J. L. R. Dlx1 and Dlx2 Control Neuronal versus Oligodendroglial Cell Fate Acquisition in the Developing Forebrain. *Neuron* (2007). doi:10.1016/j.neuron.2007.06.036

61. Lee, S. K., Lee, B., Ruiz, E. C. & Pfaff, S. L. Olig2 and Ngn2 function in opposition to modulate gene expression in motor neuron progenitor cells. *Genes Dev.* (2005). doi:10.1101/gad.1257105

62. Stolt, C. C. & Wegner, M. SoxE function in vertebrate nervous system development. *International Journal of Biochemistry and Cell Biology* (2010). doi:10.1016/j.biocel.2009.07.014

63. Stolt, C. C. & Wegner, M. Schwann cells and their transcriptional network: Evolution of key regulators of peripheral myelination. *Brain Research* (2016). doi:10.1016/j.brainres.2015.09.025

64. Stolt, C. C. *et al.* The Sox9 transcription factor determines glial fate choice in the developing spinal cord. *Genes Dev.* (2003). doi:10.1101/gad.259003

65. Claus Stolt, C. *et al.* Terminal differentiation of myelin-forming oligodendrocytes depends on the transcription factor Sox10. *Genes Dev.* (2002). doi:10.1101/gad.215802

66. Barres, B. A. *et al.* Cell death and control of cell survival in the oligodendrocyte lineage. *Cell* (1992). doi:10.1016/0092-8674(92)90531-G

67. Sun, T. *et al.* Olig bHLH proteins interact with homeodomain proteins to regulate cell fate acquisition in progenitors of the ventral neural tube. *Curr. Biol.* (2001).

doi:10.1016/S0960-9822(01)00441-9

68. Genoud, S. *et al.* Notch1 control of oligodendrocyte differentiation in the spinal cord. *J. Cell Biol.* (2002). doi:10.1083/jcb.200202002

69. Liu, A. *et al.* A molecular insight of Hes5-dependent inhibition of myelin gene expression: Old partners and new players. *EMBO J.* (2006). doi:10.1038/sj.emboj.7601352

70. Stolt, C. C. *et al.* SoxD Proteins Influence Multiple Stages of Oligodendrocyte Development and Modulate SoxE Protein Function. *Dev. Cell* (2006). doi:10.1016/j.devcel.2006.08.011

71. Kondo, T. The Id4 HLH protein and the timing of oligodendrocyte differentiation. *EMBO J.* (2000). doi:10.1093/emboj/19.9.1998

72. Fancy, S. P. J. *et al.* Dysregulation of the Wnt pathway inhibits timely myelination and remyelination in the mammalian CNS. *Genes Dev.* (2009). doi:10.1101/gad.1806309

73. Liu, Z. *et al.* Induction of oligodendrocyte differentiation by Olig2 and Sox10: Evidence for reciprocal interactions and dosage-dependent mechanisms. *Dev. Biol.* (2007). doi:10.1016/j.ydbio.2006.10.007

74. Fu, H. *et al.* Dual origin of spinal oligodendrocyte progenitors and evidence for the cooperative role of Olig2 and Nkx2.2 in the control of oligodendrocyte differentiation. *Development* (2002). doi:11830569

75. Marin-Husstege, M., Muggironi, M., Liu, A. & Casaccia-Bonnefil, P. Histone deacetylase activity is necessary for oligodendrocyte lineage progression. *J. Neurosci.* (2002). doi:22/23/10333 [pii]

76. Fu, H. *et al.* A Genome-Wide Screen for Spatially Restricted Expression Patterns Identifies Transcription Factors That Regulate Glial Development. *J. Neurosci.* (2009). doi:10.1523/JNEUROSCI.0160-09.2009

77. Wang, S.-Z. An oligodendrocyte-specific zinc-finger transcription regulator cooperates with Olig2 to promote oligodendrocyte differentiation. *Development* (2006). doi:10.1242/dev.02522

78. Emery, B. *et al.* Myelin Gene Regulatory Factor Is a Critical Transcriptional Regulator Required for CNS Myelination. *Cell* (2009). doi:10.1016/j.cell.2009.04.031

79. Tramacere, I. *et al.* Immunomodulators and immunosuppressants for relapsing-remitting multiple sclerosis: A network meta-analysis. *Cochrane Database of Systematic Reviews* (2014). doi:10.1002/14651858.CD011381

80. Baecher-Allan, C., Kaskow, B. J. & Weiner, H. L. Multiple Sclerosis: Mechanisms and Immunotherapy. *Neuron* **97**, 742–768 (2018).

81. Ascherio, A. & Munger, K. L. Epidemiology of Multiple Sclerosis: From Risk Factors to Prevention--An Update. *Seminars in Neurology* (2016). doi:10.1055/s-0036-1579693

82. Kipp, M. *et al.* Experimental in vivo and in vitro models of multiple sclerosis: EAE and beyond. *Multiple Sclerosis and Related Disorders* (2012). doi:10.1016/j.msard.2011.09.002

83. Lassmann, H. Mechanisms of white matter damage in multiple sclerosis. *GLIA* (2014). doi:10.1002/glia.22597

84. Miron, V. E. *et al.* M2 microglia and macrophages drive oligodendrocyte differentiation during CNS remyelination. *Nat. Neurosci.* (2013). doi:10.1038/nn.3469

85. Ruckh, J. M. *et al.* Rejuvenation of regeneration in the aging central nervous system. *Cell Stem Cell* (2012). doi:10.1016/j.stem.2011.11.019

86. Van Der Valk, P. & Amor, S. Preactive lesions in multiple sclerosis. *Current Opinion in Neurology* (2009). doi:10.1097/WCO.0b013e32832b4c76

87. Prinz, M., Priller, J., Sisodia, S. S. & Ransohoff, R. M. Heterogeneity of CNS myeloid cells and their roles in neurodegeneration. *Nature Neuroscience* (2011). doi:10.1038/nn.2923

88. Lucchinetti, C. *et al.* Heterogeneity of multiple sclerosis lesions: Implications for the pathogenesis of demyelination. *Ann. Neurol.* (2000). doi:10.1002/1531-8249(200006)47:6<707::AID-ANA3>3.0.CO;2-Q

89. Michel, L. *et al.* B cells in the multiple sclerosis central nervous system: Trafficking and contribution to CNS-compartmentalized inflammation. *Frontiers in Immunology* (2015). doi:10.3389/fimmu.2015.00636

90. Boyd, A., Zhang, H. & Williams, A. Insufficient OPC migration into demyelinated lesions is a cause of poor remyelination in MS and mouse models. *Acta Neuropathol.* (2013). doi:10.1007/s00401-013-1112-y

91. Minagar, A. & Alexander, J. S. Bloo d-brain barrier disruption in multiple sclero sis. 540–549 (2003).

92. Frischer, J. M. *et al.* Clinical and pathological insights into the dynamic nature of the white matter multiple sclerosis plaque. *Ann. Neurol.* **78**, 710–721 (2015).

93. Machado-Santos, J. *et al.* The compartmentalized inflammatory response in the multiple sclerosis brain is composed of tissue-resident CD8+ T lymphocytes and B cells. *Brain* **141**, 2066–2082 (2018).

94. Filippi, M. *et al.* Multiple sclerosis. *Nat. Rev. Dis. Prim.* **4**, 1–27 (2018).

95. Thompson, A. J., Baranzini, S. E., Geurts, J., Hemmer, B. & Ciccarelli, O. Multiple sclerosis. *Lancet* **391**, 1622–1636 (2018).

96. Strijbis, E. M. M., Kooi, E. J., van der Valk, P. & Geurts, J. J. G. Cortical remyelination is heterogeneous in multiple sclerosis. *J. Neuropathol. Exp. Neurol.* **76**, 390–401 (2017).

97. Baranzini, S. E. & Oksenberg, J. R. The genetics of MS: From 0 to 200 in 50 years. *Trends Genet.* **33**, 960–970 (2017).

98. Cotsapas, C. & Mitrovic, M. Genome-wide association studies of multiple sclerosis. *Clin. Transl. Immunol.* **7**, 1–9 (2018).

99. Compston, A. & Coles, A. Multiple sclerosis. **359**, 1221–1231 (2002).

100. De Jager, P. L. *et al.* Meta-analysis of genome scans and replication identify CD6, IRF8 and TNFRSF1A as new multiple sclerosis susceptibility loci. *Nat. Genet.* **41**,

776–782 (2009).

101. Beecham, A. H. *et al.* Analysis of immune-related loci identifies 48 new susceptibility variants for multiple sclerosis. *Nat. Genet.* **45**, 1353–1362 (2013).

102. Caramalho, Í., Nunes-Cabaço, H., Foxall, R. B. & Sousa, A. E. Regulatory T-cell development in the human thymus. *Front. Immunol.* **6**, 1–7 (2015).

103. Jordan, M. S. *et al.* Thymic selection of CD4+CD25+ regulatory T cells induced by an agonist self-peptide. *Nat. Immunol.* **2**, 301–306 (2001).

104. Kaiko, G. E., Horvat, J. C., Beagley, K. W. & Hansbro, P. M. Immunological decision-making: How does the immune system decide to mount a helper T-cell response? *Immunology* **123**, 326–338 (2008).

105. Wingerchuk, D. M. & Carter, J. L. Multiple sclerosis: Current and emerging disease-modifying therapies and treatment strategies. *Mayo Clin. Proc.* **89**, 225–240 (2014).

106. Frischer, J. M. *et al.* The relation between inflammation and neurodegeneration in multiple sclerosis brains. *Brain* **132**, 1175–1189 (2009).

107. Annibali, V. *et al.* CD161highCD8+T cells bear pathogenetic potential in multiple sclerosis. *Brain* **134**, 542–554 (2011).

108. Abrahamsson, S. V. *et al.* Non-myeloablative autologous haematopoietic stem cell transplantation expands regulatory cells and depletes IL-17 producing mucosal-associated invariant T cells in multiple sclerosis. *Brain* **136**, 2888–2903 (2013).

109. Rasouli, J. *et al.* Expression of GM-CSF in T Cells Is Increased in Multiple Sclerosis and Suppressed by IFN-β Therapy. *J. Immunol.* **194**, 5085–5093 (2015).

110. Hashimshony, T., Wagner, F., Sher, N. & Yanai, I. CEL-Seq: Single-Cell RNA-Seq by Multiplexed Linear Amplification. *Cell Rep.* (2012). doi:10.1016/j.celrep.2012.08.003

111. Ramsköld, D. *et al.* Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. *Nat. Biotechnol.* (2012). doi:10.1038/nbt.2282

112. Picelli, S. *et al.* Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nat. Methods* (2013). doi:10.1038/nmeth.2639

113. Wu, A. R., Wang, J., Streets, A. M. & Huang, Y. Single-Cell Transcriptional Analysis. *Annu. Rev. Anal. Chem.* (2017). doi:10.1146/annurev-anchem-061516-045228

114. Islam, S. *et al.* Quantitative single-cell RNA-seq with unique molecular identifiers. *Nat. Methods* (2014). doi:10.1038/nmeth.2772

115. Macosko, E. Z. *et al.* Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* (2015). doi:10.1016/j.cell.2015.05.002

116. Hochgerner, H. *et al.* STRT-seq-2i: Dual-index 5′ single cell and nucleus RNA-seq on an addressable microwell array. *Sci. Rep.* (2017). doi:10.1038/s41598-017-16546-4

117. Van Der Maaten, L. & Hinton, G. Visualizing Data using t-SNE. *J. Mach. Learn. Res.* (2008). doi:10.1007/s10479-011-0841-3

118. McInnes, L., Healy, J., Saul, N. & Großberger, L. UMAP: Uniform Manifold

Approximation and Projection. *J. Open Source Softw.* (2018). doi:10.21105/joss.00861

119. Trapnell, C. *et al.* The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* **32**, 381–386 (2014).

120. Svensson, V., Vento-Tormo, R. & Teichmann, S. A. Exponential scaling of single-cell RNA-seq in the past decade. *Nat. Protoc.* **13**, 599–604 (2018).

121. Tang, F. *et al.* mRNA-Seq whole-transcriptome analysis of a single cell. *Nat. Methods* **6**, 377–382 (2009).

122. Islam, S. *et al.* Characterization of the single-cell transcriptional landscape by highly multiplex RNA-seq. *Genome Res.* **21**, 1160–1167 (2011).

123. https://www.fluidigm.com/. Available at: https://www.fluidigm.com/.

124. Brennecke, P. *et al.* Accounting for technical noise in single-cell RNA-seq experiments. *Nat. Methods* **10**, 1093–1098 (2013).

125. Klein, A. M. *et al.* Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell* **161**, 1187–1201 (2015).

126. https://www.10xgenomics.com/.

127. Ding, J. *et al.* Systematic comparison of single-cell and single-nucleus RNA-sequencing methods. *Nat. Biotechnol.* **38**, 737–746 (2020).

128. Islam, S. *et al.* Quantitative single-cell RNA-seq with unique molecular identifiers. *Nat. Methods* **11**, 163–166 (2014).

129. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15 (2018).

130. Hafemeister, C. & Satija, R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *bioRxiv* 1–15 (2019). doi:10.1101/576827

131. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420 (2018).

132. Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888-1902.e21 (2019).

133. Melsted, P., Ntranos, V. & Pachter, L. The barcode, UMI, set format and BUStools. *Bioinformatics* **35**, 4472–4473 (2019).

134. Lun, A. T. L. *et al.* Distinguishing cells from empty droplets in droplet-based single-cell RNA sequencing data. *bioRxiv* 1–9 (2018). doi:10.1101/234872

135. Zilionis, R. *et al.* Single-cell barcoding and sequencing using droplet microfluidics. *Nat. Protoc.* **12**, 44–73 (2017).

136. Zheng, G. X. Y. *et al.* Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* **8**, (2017).

137. Lun, A. T. L., Bach, K. & Marioni, J. C. Pooling across cells to normalize single-cell RNA sequencing data with many zero counts. *Genome Biol.* **17**, 1–14 (2016).

138. Johnson, W. E., Li, C. & Rabinovic, A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* **8**, 118–127 (2007).

139. Heimberg, G., Bhatnagar, R., El-Samad, H. & Thomson, M. Low Dimensionality in Gene Expression Data Enables the Accurate Extraction of Transcriptional Programs from Shallow Sequencing. *Cell Syst.* **2**, 239–250 (2016).

140. Pearson, K. LIII. On lines and planes of closest fit to systems of points in space. *London, Edinburgh, Dublin Philos. Mag. J. Sci.* **2**, 559–572 (1901).

141. MacQueen, J. and others. Some methods for classification and analysis of multivariate observations. *Proc. fifth Berkeley Symp. Math. Stat. Probab.* **1**, 281–297 (1967).

142. Partitioning Around Medoids (Program PAM). in 68–125 (2008). doi:10.1002/9780470316801.ch2

143. Vincent, D. B., Jean-Loup, G., Renaud, L. & Etienne, L. Fast unfolding of communities in large networks. *J. Stat. Mech. Theory Exp.* **2008**, P10008 (2008).

144. Traag, V. A., Waltman, L. & van Eck, N. J. From Louvain to Leiden: guaranteeing well-connected communities. *Sci. Rep.* **9**, 1–12 (2019).

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

146. Korsunsky, I. *et al.* Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* **16**, 1289–1296 (2019).

147. Thi, H. *et al.* A benchmark of batch-effect correction methods for single-cell RNA sequencing data. *Genome Biol.* 1–32 (2020).

148. Trapnell, C. *et al.* The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* (2014). doi:10.1038/nbt.2859

149. Qiu, X. *et al.* Reversed graph embedding resolves complex single-cell trajectories. *Nat. Methods* **14**, 979–982 (2017).

150. Bendall, S. C. *et al.* Single-cell trajectory detection uncovers progression and regulatory coordination in human b cell development. *Cell* **157**, 714–725 (2014).

151. Haghverdi, L., Büttner, M., Wolf, F. A., Buettner, F. & Theis, F. J. Diffusion pseudotime robustly reconstructs lineage branching. *Nat. Methods* **13**, 845–848 (2016).

152. Bosco, G. Lo *et al.* and mapping of omics data with STREAM. *Nat. Commun.* (2019). doi:10.1038/s41467-019-09670-4

153. Setty, M. *et al.* Characterization of cell fate probabilities in single-cell data with Palantir. *Nat. Biotechnol.* **37**, 451–460 (2019).

154. Weinreb, C., Wolock, S., Tusi, B. K., Socolovsky, M. & Klein, A. M. Fundamental limits on dynamic inference from single-cell snapshots. *Proc. Natl. Acad. Sci. U. S. A.* **115**, E2467–E2476 (2018).

155. Tritschler, S. *et al.* Concepts and limitations for learning developmental trajectories from single cell genomics. *Dev.* **146**, (2019).

156. La Manno, G. *et al.* RNA velocity of single cells. *Nature* **560**, 494–498 (2018).

157. Bergen, V., Lange, M., Peidli, S., Wolf, F. A. & Theis, F. J. Generalizing RNA velocity to transient cell states through dynamical modeling. *Nat. Biotechnol.* **38**, (2020).