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

DISCOVERING CELL-TYPE DYNAMICS IN THE NERVOUS SYSTEM BY SINGLE-CELL TRANSCRIPTOMICS

Hannah Sophie Hochgerner

Stockholm 2017

Cover Art: *Standing on the Shoulders of Giants - Cajal Meets Single Cell Transcriptomics.*

Published by Karolinska Institutet.

ISBN 978-91-7676-619-4
Printed by E-Print AB 2017

DISCOVERING CELL-TYPE DYNAMICS IN THE NERVOUS SYSTEM BY SINGLE-CELL TRANSCRIPTOMICS

THESIS FOR DOCTORAL DEGREE (Ph.D.)

By

Hannah Hochgerner

Public Defence on **Friday, May 12, 2017, 10:00**
At **Gustaf Retzius Lecture Hall**, Berzelius väg 3, Karolinska Institutet

Principal Supervisor:
Prof. Sten Linnarsson
Karolinska Institutet
Department of Medical Biochemistry and Biophysics
Division of Molecular Neurobiology

Co-supervisor(s):
Prof. Patrik Ernfors
Karolinska Institutet
Department of Medical Biochemistry and Biophysics

Associate Prof. Jens Hjerling-Leffler
Karolinska Institutet
Department of Medical Biochemistry and Biophysics

Associate Prof. Clas Johansson
Karolinska Institutet
Department of Clinical Neuroscience

Opponent:
Prof. Arnold Kriegstein
University of California San Francisco
Department of Neurology

Examination Board:
Associate Prof. Eva Hedlund
Karolinska Institutet
Department of Neuroscience

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

Ann-Christine Syvänen
Uppsala Univeritetet
Department of Molecular Medicine

Hofstadter's Law:

It always takes longer than you expect, even when you take into account Hofstadter's Law.

To the Duck

ABSTRACT

The mammalian nervous system is arguably the most intricate system known to science. At its basis lie highly specialized single cells, specifically interacting to ensure everything from normal functionality to complex behavior and cognition. For over a century, neuroscientists have been fascinated by the diversity of cell types that make up the nervous system, and have sought ever-new strategies to characterize them. With the advance of single-cell transcriptomics, particularly RNA-seq, a new toolbox has become available for molecular cell type classification. In this thesis, I will discuss the development of relevant technologies leading up to cellular taxonomy studies, the concept of cell types on a more generalized level, and focus on cell type characterization in the context of continuous, dynamic processes such as development and maturation. Further, I will present the results of two published papers and two manuscripts, as well as preliminary data from our lab's biggest effort so far, to build an atlas of cell types across the entire nervous system.

In **paper I**, we describe previously uncharacterized heterogeneity in the CNS myelinating cell population, the oligodendrocytes (OL). We delineate the continuous maturation process from oligodendrocyte progenitors (OPCs), via a number of distinct stages, to mature OLs.

In **paper II**, we use single-cell RNA-seq to explore neurons in the sympathetic nervous system, describing seven distinct types. Retrograde and developmental tracing directly associated two of the cell types with distinct functions as erector muscle neurons.

Paper III describes the development and application of STRT-seq-2i, a 5' single-cell RNA-seq platform adapted to a high-throughput 9600-well plate. We discuss technical aspects, throughput and flexibility, as well as results from cortical samples of fresh mouse cells and human post mortem nuclei.

In **paper IV**, we performed high throughput unbiased sampling of early postnatal and adult mouse dentate gyrus, a region known for postnatal and maintained adult neurogenesis. We describe distinct stages in the developmental trajectory, holding true for the early and adult neurogenesis.

Overall, this thesis aims to shed light on molecular cell-type dynamics in different contexts, as well as discuss key concepts emerging and reevaluated along with the technological advances in the field.

LIST OF SCIENTIFIC PAPERS

I. Marques S*, Zeisel A*, Codeluppi S, van Bruggen D, Mendanha Falcão A, Xiao L, Li H, Häring M, **Hochgerner H**, Romanov RA, Gyllborg D, Muñoz-Manchado AB, La Manno G, Lönnerberg P, Floriddia E, Rezayee F, Ernfors P, Arenas E, Hjerling-Leffler J, Harkany T, Richardson WD, Linnarsson S, Castelo-Branco G.

   Oligodendrocyte heterogeneity in the mouse juvenile and adult central nervous system.

   Science 2016 June, 352(6291), 1326-1369.
   Reprinted with permission from AAAS.

II. Furlan A, La Manno G, Lübke M, Häring M, Abdo H, **Hochgerner H**, Kupari J, Usoskin D, Airaksinen MS, Oliver G, Linnarsson S, Ernfors P

   Visceral motor neuron diversity delineates a cellular basis for nipple- and pilo-erection muscle control.

   Nature Neuroscience 2016 October; 19(10), 1331-1340

III. **Hochgerner H**, Lönnerberg P, Hodge R, Mikes J, Heskol A, Hubschle H, Lin P, Picelli S, La Manno G, Ratz M, Dunne J, Husain S, Lein E, Srinivasan M, Zeisel A and Linnarsson S

   STRT-seq-2i: dual-index 5′ single cell and nucleus RNA-seq on an addressable microwell array

   Submitted Manuscript

IV. **Hochgerner H***, Zeisel A*, Lönnerberg P, Linnarsson S

   Postnatal neurogenesis is conserved in the adult mouse dentate gyrus

   Submitted Manuscript

PUBLICATIONS NOT INCLUDED IN THIS THESIS

Goldman T, Wieghofer P, Jordão MJ, Prutek F, Hagemeyer N, Frenzel K, Amann L, Staszewski O, Kierdorf K, Krueger M, Locatelli G, **Hochgerner H**, Zeiser R, Epelman S, Geissmann F, Priller J, Rossi FM, Bechmann I, Kerschensteiner M, Linnarsson S, Jung S, Prinz M

Origin, fate and dynamics of macrophages at central nervous system interfaces

Nature Immunology 2016 July; 17(7), 797-805

Zajac P, Islam S, **Hochgerner H**, Lönnerberg P, Linnarsson S

Base preferences in non-templated nucleotide incorporation by MMLV-derived reverse transcriptases

PLoS One 2013 December; 8(12), e85270

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

CNS, PNS, ENS	Central, Peripheral, Enteric Nervous System
COP	committed oligodendrocyte progenitor
CyTOF	Time Of Flight Mass Cytometry
E8 / P8	embryonic / postnatal day 8
ERCC	External RNA Controls Consortium
FACS	Fluorescence-Activated Cell Sorting
KNN	*k*-Nearest Neighbors
MARS-seq	Massively Parallel Single-Cell RNA-sequencing
MFOL	myelin forming oligodendrocyte
MMLV	Moloney Murine Leukemia Virus
MOL	mature oligodendrocyte
mRNA	messenger RNA
NFOL	newly formed oligodendrocyte
nIPC	neuronal intermediate progenitor cell
OB	olfactory bulb
OL	oligodendrocyte
OPC	oligodendrocyte progenitor cell
PCA	Principal Component Analysis
PCR	Polymerase Chain Reaction
pMN	motor neuron progenitor domain
qPCR	quantitative Polymerase Chain Reaction
RGL	radial glia-like cell
RPKM	reads per million per kilobase
RT	Reverse Transcription
RT-PCR	Reverse Transcription Polymerase Chain Reaction
SGZ	subgranular zone
smFISH	single molecule fluorescent *in situ* hybridization
STRT	Single-cell Tagged Reverse Transcription
SVZ	subventricular zone
t-SNE	t-distributed stochastic neighbor embedding
TSO	Template Switching Oligo
UMI	Unique Molecular Identifier

1 SINGLE-CELL TRANSCRIPTOMICS

1.1 INTRODUCTION

Multicellular organisms depend on a tightly regulated interplay of highly specialized cell types. This is especially true for the most complex system known to biology, the mammalian nervous system. It has therefore been a century long quest (Ramon y Cajal, 1890, 1895, 1899) to understand the components of this system, the cell types, in a variety of approaches. In order not to overlook the true cell type diversity, such approaches must inherently be carried out on a single cell level, and have as such been either laborious or biased. Due to technological breakthroughs in recent years, single-cell transcriptomics methods have turned into a pragmatic, highly informative method that allows identification and characterization of cell types on an unprecedented scale. They have revolutionized biology in a variety of fields, including for cell-type discovery, in neuroscience, immunology and cancer cell biology. Technologically, several groups step-by-step overcame a number of challenges when working with single cells: (1) reliable capture of single cells to reaction chambers, (2) capture and amplification of minute amounts of mRNA present in a single cell while reducing measurement noise and (3) multiplexing, to achieve high throughput and parallel measurements of a biologically reasonable amount of single cells.

1.2 HISTORY

Interestingly, the first study to publish transcript analysis from single cells was applied to neurons of the hippocampus by (Eberwine et al., 1992). The authors saw a need of describing previously uncharacterized differences between cell types, and performed cDNA synthesis by injecting reagents, and then extracting the cell content, with the help of a patch pipette. Already then, one key was to perform the early reactions, including capture of polyadenylated transcripts and reverse transcription, in as small a volume as possible, to efficiently recover the minute amounts of material gained from a single cell while suppressing otherwise dominating background and side reactions. A few years later, it was yet again neuroscientists that encountered a need to discriminate transcripts between single cells. Studying odorant receptors of olfactory neurons is a system where ‘one neuron, one receptor’ applies. The group of Linda Buck used targeted single-cell RT-PCR, in combination with Southern Blotting and sequencing, to detect the olfactory receptor genes differentially expressed between single neurons. This helped identify a multigene family of pheromone receptors and understand the overall combinatorial receptor coding scheme used by the olfactory system (Malnic et al., 1999; Matsunami and Buck, 1997).

Since then, both cell capture and downstream reactions have however evolved strongly and have made single-cell transcriptomics both more efficient and higher throughput. The Quake lab (Warren et al., 2006) for instance, introduced a microfluidic device to perform targeted

digital RT-PCR on single cells based on transcript limiting dilution, in an array performing a large number of PCR reactions in parallel, in small volumes. Also (Bengtsson et al., 2008) realized the need for accurate quantitative measures of targeted transcripts in single cells and presented a single-cell RT-qPCR protocol without purifications between the reactions.

Moving towards measuring a much greater range of transcripts in a single experiment, single-cell cDNA from olfactory neuronal progenitors and mature sensory neurons (Tietjen et al., 2003), and amplified mRNA from single hippocampal CA1 cells (Kamme et al., 2003) was hybridized on Affymetrix microarrays. Both groups provided more detailed molecular characterization than previously known, and described certain cellular heterogeneity not dissected in bulk analysis. Further efforts in single-cell microarray studies focused on optimizing protocols, for instance by reducing amplification bias (Kurimoto et al., 2006). Still, the cost per single cell was approximately two orders of magnitude higher than current methods, greatly limiting its widespread application.

Soon, the first full transcriptomes of single cells, generated by full-length mRNA-sequencing, were published (Tang et al., 2009). Here, the authors saw a need for a technology that allows for an even less biased approach, to both detect the expression of yet more genes than microarrays did, and obtain resolution of transcript isoforms via analysis of splice junctions. Around the same time, several highly informative studies on the diversity of CNS cell types were performed by RNA-seq in bulk, before single-cell RNA-seq could reach comparable scales (Cahoy et al., 2008; Sugino et al., 2006; Zhang et al., 2014). While these studies gave detailed expression profiles and important insights on cell type heterogeneity, they were inherently biased by relying on marker specificity when isolating cell populations by FACS sorting or immunopanning .

1.3 OVERVIEW OF TECHNOLOGIES

The early experiments presented above all included highly laborious protocols and usually required handling of single cells as separate samples. Protocols still used today, STRT-Seq, SMART-seq2 and CEL-seq, (Fig. 1) all follow the basic following workflow: single cell capture and lysis, reverse transcription of mRNAs initiated by poly-dT priming and cDNA amplification by PCR or *in vitro* transcription. These cDNA libraries are then prepared for sequencing. I will initially focus on aspects and differences in the molecular biology, and return to the non-trivial question of single-cell capture later.

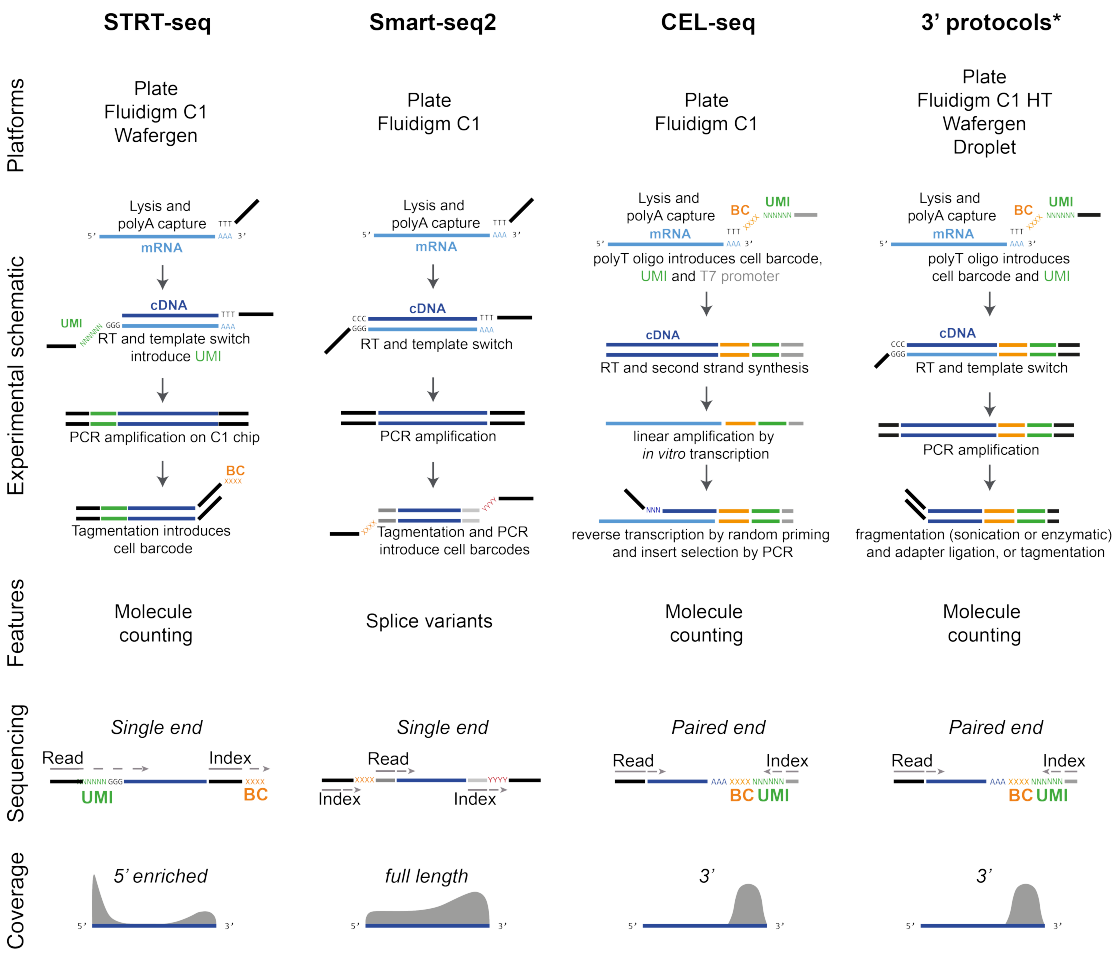

Figure 1 | Overview of the main single-cell RNA-seq technologies used today. "3' protocols" is a generalized example of technologies primarily used in commercial high-throughput platforms, such as Wafergen iCell8, Fluidigm C1 HT (High Throughput) and 10X Genomics Chromium.

In 2011, our group published the first multiplex-compatible single-cell RNA-seq protocol STRT-seq (Single-cell Tagged Reverse Transcription Sequencing), where molecules derived from different cells received cell barcodes during reverse transcription (Islam et al., 2011). This allowed for pooling of samples early on and thereby not only an unprecedented throughput, but also a strategy to increase the amount of starting material before amplification, reducing both amplification cycles and bias. Figure 1 shows an updated and optimized version of STRT-seq, also known as C1-STRT, compatible with use on the Fluidigm C1 microfluidic platform and single molecule counting. Further adaptations in STRT-seq are discussed in more detail below, and shown in Figure 2.

Soon after STRT-seq, Smart-seq (Ramsköld et al., 2012) was published, then updated to the now commonly used Smart-seq2 (Picelli et al., 2013) in 2013. In both STRT-seq and Smart-seq2, poly-dT oligos with adapter sequences capture polyadenylated transcripts, which are reverse transcribed by an MMLV-derived reverse transcriptase. This class of transcriptases generates a poly-cytosine overhang at the 3' terminal, to which a poly-guanidine template switching oligo (TSO) can anneal. The reverse transcriptase then switches template to the TSO and elongates its complementary strand, rendering a full-length RNA-DNA hybrid with adapter sequences on either end of the transcript, ready for PCR-amplification. Smart-seq2

now introduces PCR adapters along the entire transcript length using Tn5 transposase, by a process called tagmentation, discussed in more detail below. Fragments generated in this way are amplified, introducing a cell barcode sequence. Smart-seq2 reads are thus spread along the full length of the transcript. STRT-seq introduces a single index sequence and sequencing adapter by tagmentation and 5' fragments are enriched, leading to a 5'-biased read coverage.

CEL-seq (Cell Expression by Linear amplification and Sequencing) (Hashimshony et al., 2012) follows a slightly different logic since cell barcode, adapter and a T7 promoter are introduced by the poly-dT oligo on the 3' end of the polyadenylated transcript. Amplification is linear on pooled samples by in vitro transcription (IVT), and amplified RNAs are fragmented and ligated with a second adapter. 3' fragments holding both adaptors are selected for by PCR. CEL-seq was also adapted to the high-throughput droplet microfluidic platform and inDrop (Klein et al., 2015).

Since the development of new high-throughput platforms, most notably Drop-seq (Macosko et al., 2015), new 3' protocols have been presented, combining the advantage of tagging molecules early using the poly-dT oligo, reverse transcription with template switching and amplification by PCR. A similar approach is used in the commercialized droplet microfluidic platform by 10X Genomics Chromium, as well as the Wafergen iCell8 microchip (schematically outlined, with slight variations between the practical implementations in Figure 1 "3' protocols").

Since STRT-seq enriches for 5' fragments rather than full-length transcripts, the main readout is quantitative expression of any given gene, and to a certain extent alternative promoter usage (manuscript from our group (Karlsson *et al.*) in revision). Similarly, CEL-seq enriching for 3' sites, gives mainly quantitative gene expression output. Smart-seq2, providing full-length coverage, also allows for conclusions on transcript variants generated by alternative splicing or a more in-depth analysis of SNPs, for instance for the study of RNA editing (Blow et al., 2004) or allelic expression differences (Deng et al., 2014); all of which may have functional influences on the cell. This type of analysis of course requires deeper sequencing compared to purely quantitative expression studies. Importantly, CEL-seq and droplet 3' protocols require paired-end sequencing, roughly doubling sequencing costs compared to STRT-seq or Smart-seq2.

1.4 THE EVOLUTION OF STRT-SEQ

As mentioned, the first iteration of STRT-seq was published in 2011 (Islam et al., 2011) and for the first time allowed multiplexing and a more streamlined protocol than previously, at the same time improving capture and reducing bias. Briefly, cells are lysed and polyadenylated molecules captured using a poly-dT primer. Reverse transcription is carried out with an MMLV-derived reverse transcriptase, capable of template switching at the 5' end. The template switching oligo (TSO), annealing to the cytosine cap, introduces a known *n*-base pair-long cellular barcode to the molecule. It thus became possible to pool any amount of

samples with distinguishable cellular barcodes (*i.e.* in the case of an $n = 6$ base pairs-long barcode, a theoretic 4^6) for downstream cDNA amplification and sequencing library preparation. In practice, this number was limited by the amount of cells reasonably sampled (often manually) to a multiwell plate in any given experiment, as well as the number of (rather costly) different TSOs available to the researcher. Importantly however, this early pooling allowed for fewer cycles of amplification, reducing associated bias and improving the quantitative nature of the data.

In 2013, this method was adapted to the Fluidigm C1 microfluidic platform (C1-STRT, see below), with aside from different cell handling discussed below, small alterations in the molecular biology. Due to the nature of the C1 platform, it became necessary to introduce the known cellular barcode after cDNA amplification, during the tagmentation reaction for sequencing library preparation. Tagmentation is carried out with the help of the hyperactive Tn5 transposase that is loaded with oligonucleotide adapters, containing the *n*-base pair-long cellular barcode. Added to the sample, Tn5 simultaneously fragments cDNA molecules and inserts the adapters (tags), used for cell identification as well as sequencing, in the site of fragmentation. The length of these fragments is determined by the relative abundance of loaded Tn5 transposase, generating tagged libraries of sequencing-compatible lengths.

More importantly, C1-STRT introduced the use of a molecular barcode called UMI (Unique Molecular Identifier), consisting of a degenerate sequence of *k*-nucleotides, and introduced via the TSO during reverse transcription. This UMI allows the distinction and counting of a theoretic 4^k molecules[1] originally captured in the cell of the same transcript or template-switching site within the transcript. This further improved the quantitative nature of single-cell RNA-seq experiments and to a great extent eliminated amplification-borne biases. It now became possible to rely on an absolute count of unique molecules rather than reads (standardized to *e.g.* reads per million per kilobase [RPKM]) or, in the best case, counts approximated with the help of spiking in known amounts of exogenous standardized molecules (External RNA Control Consortium, ERCC) (Jiang et al., 2011). The application of UMIs was quickly adapted also by other protocols, such as CEL-seq2 (Hashimshony et al., 2016) and the CEL-seq-like MARS-seq (Jaitin et al., 2014), and has become standard for new protocols such as Drop-seq, the Chromium and iCell8 platforms. Conceptually and practically however, it has not been feasible to adapt the use of UMIs to full-length protocols as Smart-seq2 yet.

[1] More accurately, the maximum distinguishable number of molecules is 4^k minus collision events, where closer to saturation the probability of randomly using the same UMI rises.

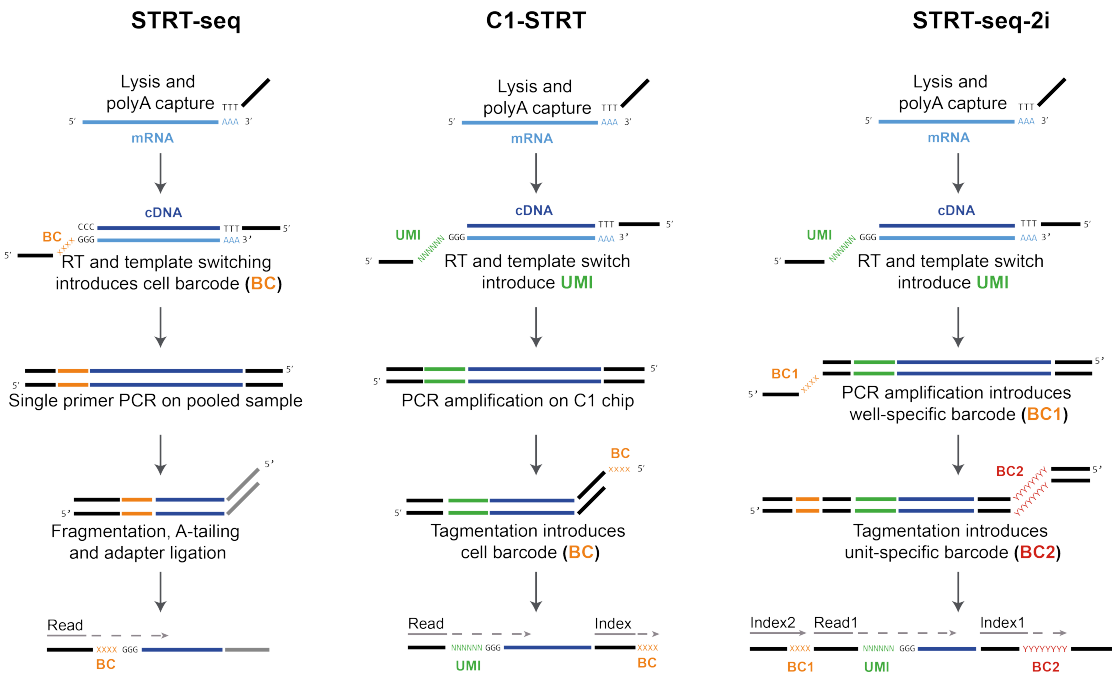

Figure 2 | Overview of the experimental workflow across three STRT-seq protocols.

Here, I will present an update of STRT-seq called STRT-seq-2i (Fig. 2, Paper III). This 5' dual-indexing protocol is adapted to increase the throughput of STRT-seq to process thousands of cells in one experiment, while retaining some of the advantages of STRT-seq, including cell imaging, UMI molecule counting and flexible sample handling. In this protocol, the UMI is introduced, as before, during reverse transcription by the TSO. The first round of cell barcoding is achieved during cDNA amplification with the help of indexed PCR primers 5' of the transcript. Due to the layout of the platform, samples differentially indexed in this first round are then pooled, and each pool receives another index during tagmentation reaction, as above. A single library of dually indexed and countable molecules is thus generated, that unlike other current high-throughput methods, is compatible with single-end Illumina HiSeq sequencing.

1.5 SINGLE-CELL CAPTURE

An important aspect in moving towards higher throughput that has not been addressed so far, is the method development in reliably capturing single cells to reaction chambers suitable for downstream processing. It is of course key to any single-cell experiment to capture (1) viable cells, since cell integrity at the time of collection determines mRNA recovery and quality, (2) obtain a biologically reasonable sample size from a single experiment and (3) true singlets, since doublets or multiplets skew downstream data. None of these have reached perfection; yet technological advances have made great improvements, discussed in more detail below.

Doublets can occur in any approach, both due to technological failure or statistical probabilities utilized by the method. Nevertheless, prevalence and ease of detection varies between methods, since not all allow for manual or automated microscopic inspection.

Missed doublets can skew analysis in several ways, for instance by appearing as high quality cells (greater mRNA yield) or as intermediates between cell states. This may be more detrimental in studies that measure cell continua (e.g. developmental trajectories) as opposed to very distinct cell states, where doublets are additive hybrids of such distinct types and can be excluded from analysis.

1.5.1 Cell integrity and viability

Besides when working with cultured cells or cells naturally in suspension (such as blood- or bone marrow-derived cells), single cell capture is usually preceded by proteolytic enzyme or physical dissociation. The ease of dissociation is strongly tissue dependent, with epithelial tissues for instance posing greater challenges in terms of sturdiness, requiring more harsh protocols. On the other hand, softer tissues, such as the brain, may contain cells highly sensitive to perturbations, and the major challenge is to preserve integrity and viability by using as soft and short protocols as possible. In any case, for maximum viability and data quality it is essential to keep the time of removal from the native environment until cell lysis minimal, and preserve the cells' native state e.g. applying physiological solutions and cold, to slow metabolism.

In the particular case of sampling sensitive neuronal cell types from adult animals, which much of the work in this thesis is based on, great efforts were initially invested to maximize viability. Briefly, we used soft enzymatic digestion by papain on tissue pieces microdissected from vibratome sections, followed by careful physical trituration with fire polished glass pipettes. All steps except digestion were carried out on ice, in well-oxygenated artificial cerebrospinal fluid (aCSF). When appropriate, aCSF was supplemented with lower Ca^{2+} and excess Mg^{2+}, to reduce neuronal excitability especially in later postnatal animals. These measures, in combination with quick and experienced handling, were a prerequisite for good quality data, independent of the downstream platform used. Nevertheless, no dissociation technique can be fully damage or bias free; especially to cells with long processes or complex arbors, and more sensitive cell types may never be viably captured in any dissociation-based sampling.

1.5.2 Cell collection

Figure 3 provides an overview of cell capture methods (adapted from (Kolodziejczyk et al., 2015)). In the early days of single-cell biology, manual or automated picking of cells from a suspension or dissociated from a tissue was carried out under a (fluorescent) microscope (Hashimshony et al., 2012; Islam et al., 2011; Usoskin et al., 2014). This represents a reliable method to isolate true single cells, but is highly labor intense and limits the sample throughput for each collection by the time frame that does not compromise cell viability. Further, different groups used laser capture microdissection (Tietjen et al., 2003), by applying a laser pulse to single cells in thinly cut, frozen tissue sections and collecting them in the reaction well. Again, this is a rather labor intense method and can quickly compromise sample quality, but has the advantage of retaining accurate spatial information that may be of

great interest during analysis (Zechel et al., 2006). Another low-throughput collection method is through aspiration of the cell cytoplasm using a patch-pipette, from more-or-less intact tissue. This method has been used for downstream quantitative RT-PCR and microarray analysis (Rossier et al., 2015; Subkhankulova et al., 2010), and recently been published for single-cell RNA-seq (Patch-seq (Cadwell et al., 2016; Fuzik et al., 2016)). Importantly, this allows for coupling of electrophysiological recording, and morphology, with gene expression analysis, adding a layer of information highly relevant especially to neuroscientists.

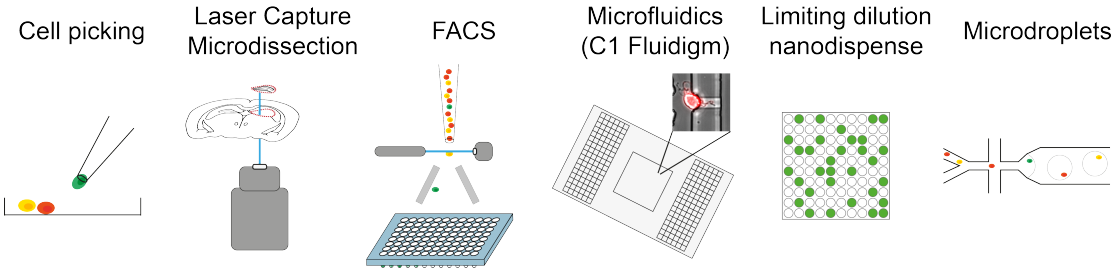

Figure 3 | A variety of approaches to perform cell capture for single-cell RNA-seq.

Much throughput was achieved when single cells could reliably be FACS-sorted to multiwell plates (96- or 384-well format). Given a viable, clean single-cell suspension and an accurate FACS operator and instrument, this is a very fast process. Since such plates are usually not compatible with use in microscopy, visual inspection of true single-cell capture is not possible. Routinely, cells are directly sorted to a buffer containing reagents for immediate cell lysis (Picelli et al., 2013). This limits the time cells are exposed to adverse environments before capture.

Overcoming some of these challenges, Fluidigm released the C1 platform in 2013, which based on microfluidic valves loads single cells to capture sites. These can be microscopically inspected for viability, fluorescence or presence of more than one cell. Downstream lysis, reverse transcription and PCR are also carried out on the same chip, in nanoliter reaction chambers. Adapted versions of Smart-seq, STRT-seq and CEL-seq all became available on the C1, with further improvements in sensitivity and efficiency compared to standard multiwell plates. With 96 capture sites and a cell capture efficiency of 50-80%, strongly dependent on the type of cell suspension loaded, large-scale studies still required a great number of separate experiments (Zeisel et al., 2015). Further, a certain capture bias based on cell size could be observed in highly heterogeneous tissue-derived suspensions. For instance, the small oligodendrocyte progenitor cells (OPCs) were absent from a comprehensive cortex data set where cells were collected on a medium-sized chip (for 10-17µm diameter cells) (Zeisel et al., 2015), while they could reliably be recovered in another study that used small-sized chips (5-10µm) (Marques et al., 2016).

1.5.3 High throughput platforms

Another capture strategy is based on limiting dilution, where a highly diluted cell suspension is loaded onto a well- or droplet-based platform to reduce the likelihood of multiple cells getting captured to the same site or droplet. At the same time, this increases the incidence of

empty capture sites, making this approach applicable only to platforms with reasonably high throughput. The likelihood of the capture of a single cell ($x = 1$) is thus based on a Poisson distribution ($P(x) = \frac{\lambda^x e^{-\lambda}}{x!}$, where λ is the average number of events). For example, Wafergen introduced the iCell8 chip with 5,184 single reaction wells, where cell suspension and downstream reagents are dispensed in nanoliter volumes, targeted directly into single wells. Loading an average of $\lambda = 1$ cells per dispense volume in the cell suspension would thus give a probability of $x = 1$ cell/well of $P(1) = \frac{1^1 e^{-1}}{1!} = 0.37$, or 37% (1,907 single cell wells). At the same time, the probability of empty wells $P(0)$ is also 37%, while the probability of multiplets divides, *i.e.* $P(2)$ is 18%, and so forth. Combined with an imaging system to detect empty wells or doublets after dispense, the system can then target all successful events and, in an ideal case, still achieve throughput of over 1,800 cells per experiment; far exceeding methods discussed above. In practice, this number is still dependent on the nature of the cell suspension, such that for instance highly heterogeneous suspensions derived from brain tend to dispense less regularly than cultured cells and achieve single-cell loading rates closer to 25-30%. In this thesis, I will present an upgrade of this platform, based on a chip of 9,600 wells (Paper III). Here, we demonstrate further increased throughput not only by limiting dilution dispense, but also by targeted FACS sorting to the single wells of the chip.

As mentioned above, recently developed droplet microfluidic platforms make use of limiting dilution in a similar fashion (Klein et al., 2015; Macosko et al., 2015; Zheng et al., 2016). In these systems, two aqueous streams containing cell suspension on the one hand, and microbeads (loaded with barcoded poly-dT oligos) on the other hand, are flowed in microfluidic channels and merged at a junction, where they encounter an oily carrier solution and form droplets of highly uniform size. By tuning the flow rate and stability, many thousands of such droplets can be formed per minute. In order to avoid merging of several cells or barcoded beads in one droplet, both components are diluted to such an extent that the likelihood of multiplet events significantly decreases, *i.e.* the aim of average events $\lambda \ll 1$. This is important since there is no manual inspection or selection of single-cell and single-bead containing wells possible at a later point. Since both events of a single-cell and single-bead need to overlap, the resulting likelihood of early droplet microfluidic systems approaches a Double Poisson distribution. Hence the fraction of droplet containing exactly one cell and one bead will be the product of two Poissons, *i.e.* ideally 0.37^2 = 14%. Through technical advances however, the reliable loading of a single bead per droplet has since made it possible to surpass the Poisson probability. It is thus the combination of applying strongly limited dilution on a massively efficient droplet generator that makes droplet microfluidics the most successful platform in terms of throughput, today.

1.6 CHALLENGES AND FUTURE DEVELOPMENTS

Certainly, the discussed advances in single-cell technologies will further strengthen their role in studies in a number of fields. For instance, Seq-Well (Gierahn et al., 2017), a portable high-throughput platform has high potential to be adapted to clinical applications, where flexibility and low costs are major factors. Intriguingly, other recent approaches potentially allow for yet higher throughput by applying several rounds of split-and-pool, in combination with *in situ* reverse transcription, on cells initially sampled in bulk (Cao et al., 2017; Rosenberg et al., 2017). The emergence of such clever statistics-aided multiplexing approaches make it difficult to envision an end to further developments, with the ultimate goal of maximizing flexibility and throughput for the user, while becoming less labor intense and costly.

At the same time, the rate-limiting factor may become downstream sequencing and analysis. Not every study may need the simultaneous measurement of all genes in ten thousands of single cells, nor will all budgets allow sequencing these at adequate depth. For each case, the study design should be such that it not only takes into account advantages of available technologies, but also performs power analysis, *i.e.* weighing benefits of sampling more cells or samples, versus deeper sequencing.

It can further be argued that with increased understanding of cell types and their transcriptional landscapes, it will no longer be necessary to perform all studies at (a) single-cell resolution or (b) transcriptome-wide coverage. Rather, it could become feasible to build statistical classifiers based on existing datasets and, for instance, approximate the qualitative and quantitative cell-type composition of a new bulk dataset using Bayesian binomial distributions. By the same token, once we have acquired systematic data-driven understanding of modular and hierarchical gene expression, we could measure – depending on context – a set of only several hundreds of genes by a reliable targeted approach, rather than more costly unbiased sampling of full transcriptomes. This may be further facilitated by *in situ* technologies that can already today measure dozens or hundreds of genes in the same sample, such as multiplexed single-molecule FISH (smFISH) (Chen et al., 2015; La Manno et al., 2016; Lubeck et al., 2014), *in situ* sequencing (Ke et al., 2013; Lee et al., 2014) or spatial transcriptomics (Ståhl et al., 2016), and have the added advantage of retaining (almost) full spatial information.

Another goal of the field has long been to perform several distinct methods in the same cell. Such simultaneous measurements could include genome sequencing (Dey et al., 2015), or the probing of chromatin accessibility and the epigenome by ChIP-seq (Rotem et al., 2015), ATAC-seq (Buenrostro et al., 2015; Cusanovich et al., 2015), Hi-C (Nagano et al., 2013) or DNA-methylation (Guo et al., 2013, 2014). Today, single-cell protocols published on these methods generate fairly sparse data, and will need further optimizations before rising up to the standards achieved in single-cell RNA-seq, yet alone be combined in a single assay. It remains to be seen which direction the field will move into; certainly however, researchers will continue to benefit from creative technological advances yet to come. Where appropriate,

single-cell technologies are likely to become a gold standard for measuring differential gene expression and regulation in heterogeneous tissues, for developmental and functional studies, as well as in clinical applications.

1.7 ANALYSIS OF SINGLE-CELL RNA-SEQ DATA FOR CELL TYPE DISCOVERY

As emphasized previously, the main advantage of using single-cell RNA-seq for cell type discovery, besides its great throughput, is the unbiased approach it allows the researcher to take. This enables studying known tissues more deeply than previously, or entirely unknown samples without prior knowledge of their composition. As such, to discover the underlying structure of any such data, analysis needs to be unsupervised, albeit not uninformed. After initial analysis steps such as extraction of quality reads, demultiplexing and mapping (discussed in (Grün and van Oudenaarden, 2015)), single-cell RNA-seq data specifically consists of a large transcriptome-wide matrix of normalized expression values or actual molecule counts. Given this highly dimensional expression space, the researcher needs to take decisions on (1) quality control, (2) the choice of genes or dimensionality reduction and (3) clustering.

1.7.1 Quality control

In single cell datasets sampled from complex or sensitive tissues, the variation in terms of quality between cells can be large. For instance, cells undergoing apoptosis are more prone to RNA degradation, which is practically reflected in their number of molecules and detected genes. Further, such cells have been observed to upregulate their mitochondrial genes, while ribosomal genes are more abundant in healthy, actively transcribing cells. Such measures are thus frequently used to perform initial quality control on the cell level, with cells falling under a certain threshold to be removed from all further analysis. While this may vary between experiments and require certain assumptions and experience, downstream analysis and clustering tends to become more straightforward when not contaminated by poor quality cells.

1.7.2 Gene selection

Since by far not all genes are informative in any given assay, it is usually recommendable to perform gene (or feature) selection. First, to reduce noise, genes that are expressed in very low numbers across all cells may be removed, as are genes that do not show correlation with any other genes above a certain threshold in a correlation matrix. Importantly, selecting informative, differentially expressed genes only, has been a common approach. The variability in expression of any given gene between single cells depends on its expression level. Therefore, genes that are more variable than expected given their expression level can be considered informative. Practically, plotting the coefficient of variation against mean

expression, such informative genes would stand out as 'noisy' and are used for downstream analysis (Zeisel et al., 2015).

In certain cases, it can be appropriate to eliminate genes that may introduce batch effects or be considered confounders, such as apoptosis or cell cycle genes (Stegle et al., 2015), or the genes involved in X chromosome inactivation in females. Some cell functions, such as cilia in ependymal cells (Zeisel et al., 2015), are encoded by large modules of genes and are unambiguously detected. On the other hand, the effect of single genes that contribute disproportionally to the function of a cell may need to be enhanced, such as olfactory receptor genes in olfactory neurons. Eliminating or enhancing certain genes is usually based on some knowledge of the data structure or specific questions that need to be addressed, and may thus be introduced only in a later iteration of the analysis.

Even after poor quality cells have been excluded and genes downselected to the most relevant ones underlying the structure of the data, the researcher is left with a highly multidimensional dataset. An additional commonly applied approach to perform dimensionality reduction is principal component analysis (PCA). PCA orthogonally transforms sets of possibly correlated variables into sets of uncorrelated variables. These so-called principal components are then ordered by the fraction of variance they explain in the dataset (Grün and van Oudenaarden, 2015). Using PCA for dimensionality reduction, downstream analysis would then be performed on the top *n* principal components that explain most of the variation of the dataset. Analyzing the first two to three principal components can be enough to identify the general data structure and major cell type clusters (Pollen et al., 2014; Treutlein et al., 2014).

1.7.3 Clustering

Distinguishing cell types in such a dataset is achieved by grouping single-cell transcriptomes into clusters, which are more similar within the group than to other datapoints, in an unsupervised fashion. In general terms, any clustering method thus needs to meet the challenge of identifying meaningful groups of similar cells, correlated by their overall gene expression profiles, represented by distance measures (e.g. Euclidian, Correlation, Jaccard or Cosine) in highly multidimensional space. Clustering to identify cell types in single cell data comes in a variety of flavors. A classical clustering approach, *k*-means clustering aims to partition datapoints into *k* clusters (where *k* is user-defined), by iteratively adjusting the position of a centroid and assigning datapoints to their closest mean. While this is a highly efficient, fast method, it easily fails if the shape of data is not best described by a centroid. Hierarchical (agglomerative) clustering, on the other hand, builds a hierarchal 'tree' of similar cells ('leaves'), by connecting similar points, two at a time, until the 'root' of the dataset (the two most variable points) are connected.

BackSPIN (Marques et al., 2016; Zeisel et al., 2015) is a two-way clustering algorithm, which simultaneously clusters cells and genes. In several rounds, cells and genes in an expression matrix are sorted according to similarity and split at an optimal breaking point, until a stopping conditions allows no further splits. In each round, genes are assigned to the

clusters they are most highly expressed in. In this way, genes assigned to oligodendrocytes in an initial split will not affect the split of neurons in a later round, for instance. BackSPIN has been successfully implemented in a number of datasets with cell numbers in the order of 10^3-10^4. With the emergence of technologies able to process 10^3 cells in a single experiment, graph-based clustering algorithms have proven successful. Several studies (Levine et al., 2015; Macosko et al., 2015; Shekhar et al., 2016) have implemented versions of KNN-based clustering on large datasets. In principle, a KNN-graph is a binary matrix representation of pair-wise similarities, based on a distance measure. In this thesis, we used Louvain Jaccard mutual-KNN clustering on the dentate gyrus datasets. Starting from Euclidian distance a KNN matrix is generated, connecting only pairs of cells that are mutual *k* nearest neighbors (where *k* is user defined). Based on the network each cell finds itself in (*i.e.* how many neighbors each pair of cells share with each other), Jaccard weighs the strength of these connections. The Louvain algorithm is then implemented to detect communities that are strongly interconnected, the clusters. Visualization can be implemented in different ways (we use t-SNE, discussed below). Importantly, this clustering allows visualizing connections between cells using edges, revealing additional information about inter-cluster relationships.

t-distributed stochastic neighbor embedding (t-SNE) has been a highly popular approach to visualize single-cell datasets, making use of its efficient nonlinear dimensionality reduction. Briefly, it iteratively reduces the divergence between probability distributions of the original high dimensional space and the low dimensional representation. In the case of single cell datasets, this leads to a visual representation where similar cells are modeled by nearby, and dissimilar cells by distant points. t-SNE can thus be used as a visual implementation, where the results of independent clusterings, such as BackSPIN clusters or Louvain Jaccard KNN clusters and edges can be represented.

1.7.4 Further notes

Other challenges in the analysis of single-cell RNA-seq datasets lie in its inherently substantial noise levels. These can be biologically intrinsic, due to the stochastic, bursting nature of transcription and differences in lysis efficiency between different cell types. Further, technical variations are introduced by variations in RNA capture and cDNA conversion, library quality and sequencing depth. Genes that do not get detected due to low efficiencies lead to false-negatives and patterns of bimodal expression. Normalization using spike-in molecules or UMIs (discussed in Chapter 1.4) can help account for certain technical variations. The contribution of zero values, however, is more difficult to handle. Due to the noisy nature of single-cell datasets it can be difficult to distinguish spurious from real clusters, especially considering that cluster sizes can vary by orders of magnitude in any given datasets. Researchers may choose to downsample datasets for each cluster to contain similar numbers of cells, and thus eliminate spurious clusters, or arbitrary noise-induced subdivisions of larger clusters (Wagner et al., 2016). Usually, both iterative clustering and several different analyses need to be performed based on informed decisions when dissecting the structure of such datasets.

In summary, despite increasing understanding and all efforts in developing new solutions to ever more complex datasets, there is no community consensus on standardized quality control, gene selection or clustering. While rough guidelines on this subject may be useful, especially to researchers entering the field or just peripherally dealing with such a dataset, it is doubtful that a one-size-fits-all fully automatized pipeline can be reasonably employed. In addition, it is likely that some researcher-induced bias will remain, especially in the interpretation of the data. For instance, manual interpretation and reiteration of certain analysis steps is commonly practiced. Here, the wealth of literature generated over decades using more low-throughput or biased methods is absolutely instrumental for data interpretation. This is also necessary to maintain biological relevance. For instance, while deeper sampling may reveal further details of variation between cells, most researchers agree that there must be a certain ground truth in cell types, that is, subdivisions must be meaningful, and a hierarchy cannot end up in single-cell clusters. It has become common practice to validate single-cell RNA-seq data using complementary or independent methods, such as different single-cell platforms, *in situ* stainings of RNA or protein, animal lines or assays of physiology and function. After all, the power of any dataset comes down to its validity across methods and biological functionality.

2 CELL TYPE DISCOVERY IN THE NERVOUS SYSTEM

2.1 WHAT IS A CELL TYPE?

The definition of what makes a cell type can be approached from a variety of perspectives. Personally, it brought me back to my undergraduate course in The Philosophy of Language, where starting from Aristotle, I began to understand the difficulties of defining the membership of a certain object to a distinct concept. Aristotle's classical theory of concepts states that they have a definitional structure, that is, a list of features that must be necessary and sufficient for the membership in the class, covered by the concept. Wittgenstein and his contemporaries argued that such a definition ignores the problem of fuzzy memberships. For instance, a dog with three legs should still be considered a dog, even if it does not follow the definitional structure. Instead, they describe membership as closeness to a prototype, the most central member of the concept. If an individual is close enough to such a typical member, it needs to be considered a member of the relevant class. Importantly, such concepts cannot be learned in isolation, but always only in combination with surrounding concepts. In order to distinguish a whale from a fish, the concept of either of them have to be understood first. In yet more abstract terms, Frege reserved the word concept for a philosophical or mental representation that would allow us to draw inferences about the types of entities we encounter. In this way, the concept of a dog is philosophically distinct from actual things in the world grouped by the concept 'dog'. Here, perhaps, the most central member of a concept is not necessary.

Given that cell types in multicellular organisms fulfill specialized tasks, functionality must lay the basis of a cell type definition. Practically, it is hardly feasible to characterize the full spectrum of a cell's functions. Several separate isolated features could be recorded, such as morphology, intrinsic physiological character, or (a combination of) immunomarkers. Arguably, our closest description of cell function would be based on a full description of the functional units it contains, best approximated by proteins. Today, even though it's now possible to measure surface antigens in the dozens on single cells by CyTOF (Bendall et al., 2011, 2014a), proteins are still too complex to measure in a wide, unbiased spectrum at meaningful cellular resolution. Therefore, it is useful to exploit the Central Dogma of molecular biology and measure a unit that is simpler to recover, preserve, read and interpret; mRNA. A cell type would thus be defined as the function of the activity of all genes, measured as mRNA, at any given moment. This approach would not give us a perfect, but precise enough idea of the multitude of underlying functions of the cell.

Coherent with progressing functional specialization in multicellular organisms, the number of cell types has changed during evolution. (Arendt et al., 2016) therefore argue for an evolutionary definition of cell types, allowing cell types to be compared and delineated within and between species. Here, cell types are evolutionary units, even though distinct evolutionary processes may bring about similarity between these units. At the core of this lay

the modular functional organization and cell type specific traits (‘apomeres’), brought about by a unique core regulatory complex, made up of transcription factors, that continuously evolve and distinguish a new cell type from their evolutionary sister cell type. Further, cell types must be rebuilt each generation, are thus encoded in the genome and share common developmental history. Therefore, both knowledge of modifications to the genome during evolution, and how this relates to the developmental lineage, is necessary to understand cell types.

Fitting the discussion of philosophical concepts and (modular) gene expression as underlying functional features, the idea of cell types as attractors has been discussed for several decades (Kauffman, 1969). Cell types are thus concepts, attracting cells moving in cell state space to stable configurations. In the context of gene expression, this space is highly multidimensional. Even reducing the problem to simple binary expression, the number of possible states would be close to $2^{25,000}$. Yet, we observe that the large majority of such states is uninhabited, explained by the act of attractors, binding cells to local minima in cell state space. The underlying mechanism of such attractors would be explained by gene regulatory networks, which restrict attractor space by imposing a set of rules (Enver et al., 2009). These rules are implemented by entities such as transcription factors, simultaneously regulating entire modules of genes in hierarchical and combinatorial fashion. Waddington’s epigenetic landscape (Waddington, 1939, 1940, 1957) is frequently used to visualize this concept (Fig. 4A); especially in the context of cells moving down a trajectory of differentiation, from a state of high potential, possibly under the influence of external perturbations or signals, to a state of low potential (functionally specialized attractor). Underneath this landscape, the gene regulatory network builds a scaffold, directing possible routes and intermediate attractors (Fig. 4B).

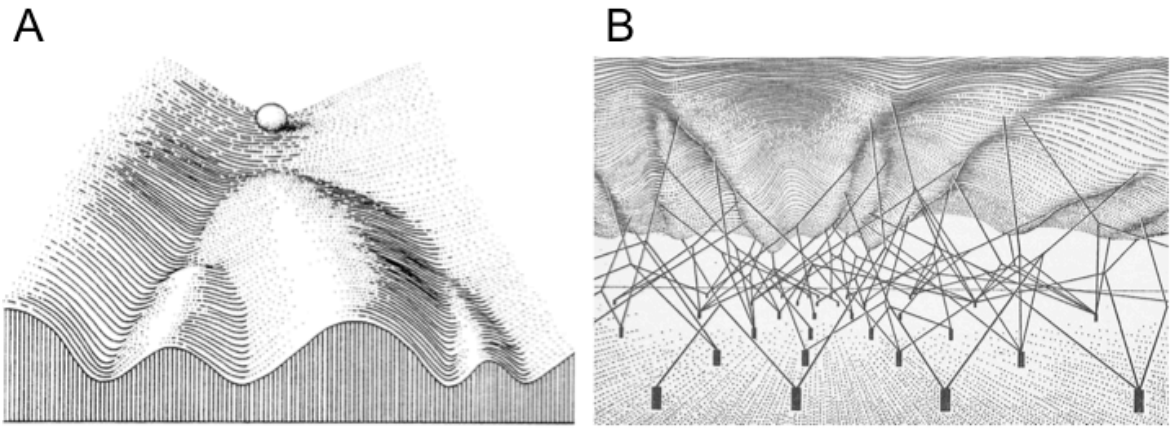

Figure 4 | Cell types as attractors, illustrated for the case of differentiation by Waddington’s classic epigenetic landscape (Waddington, 1939)**. (A) A cell moves ‘down’ a valley differentiation, from a state of high potential, to one of several possible attractor cell types. (B) The underlying gene regulatory network is dynamic, and provides a scaffold of possible routes for the differentiating cell.**

Returning to pragmatic cell type definitions, (Fishell and Heintz, 2013) opted for what they called, for the lack of an agreed universal definition; an ‘operational’ cell type definition, as a

shared, stable, ground state broadly dictating its functional capacities. They also argue that the most objective methodology would be to measure this ground state by profiling gene expression. Their molecular ground state would then identify cells distinctly from other cell types and determine their functional capabilities. The ground state signature is established after overcoming a number of hurdles, such as exclusion of activity-dependent genes and stochastic gene expression. Then, no further subcategorisation should be possible that would divide the type into additional stable, defined subtypes of cells. Yet, cells of the same type could still dynamically alter their molecular profiles, stochastically, or depending on the current context and activity. As such, individual cells need not have identical expression to fall into the same cell type. For instance, olfactory neurons, each specifically expressing a different single receptor gene (Buck and Axel, 1991) would still be part of the same cell type. The same way, Purkinje cells developmentally patterning parasagittally into stripes of *Aldoc*$^{+}$ and *Aldoc*$^{-}$ (zebrin) cells (Gravel and Hawkes, 1990) would represent fine-tuned functional diversification, rather than separate cell types.

2.2 WHY DO WE NEED TO DISCUSS CELL TYPES IN NEUROSCIENCE?

It has thus been a century-long fascination for neuroscientists to describe the great cell type diversity of the mammalian nervous system, in order to understand the building blocks that make it up. There are several underlying functional reasons why such diversity is needed.

Firstly, the high degree of specialization of nervous system cell types becomes particularly apparent thinking of the complex tasks that need to be executed. To a certain degree, it can be assumed that the number of specialized cell types reflects the computational complexity any given substructure needs to perform. This has been thoroughly studied in the peripheral nervous system of simple organisms (Bargmann and Marder, 2013), and likely holds true moving towards more complex CNS circuits made up of a larger number of distinct cell types.

Second, however, the ‘circuit is certainly more than the sum of its parts’ (Fishell and Heintz, 2013). Neurons must be specified as distinct cell types, at least transiently during development, in order to form proper short- and long-distance connection. Thus at least some of the molecular cellular diversity observed in the brain may not exist to confer distinct cell-intrinsic functions, but to ensure proper connectivity.

A further intriguing function of nervous system cell-type diversity is in its combinatorial, modular organization. This way, neurons that are part of different functional systems and localized in different subregions are able to respond to a single neuromodulator by expressing the same, or similar receptors. Downstream signaling pathways and behavioral outcomes, however, may vary greatly, depending on the specific function of the cell type. Further, there are great combinatorial possibilities by coexpressing other receptors within the same cell type. Neuromodulators, including neuropeptides and small mediators such as serotonin,

prostaglandins and even peripheral peptides, can in this way achieve global modulations and a wide scope of action (Fishell and Heintz, 2013).

Neurons do not exist in isolation or as static entities, so the overall goal should be to achieve a more contextual classification. Single-cell transcriptomics can help to infer biological properties and functions on neural cell types, as in the example of neuromodulator-related gene expression above. Detecting expression of such genes gives clues to the circuitries the neuromodulator and the particular cell type are involved in. In a similar way, detecting cell-surface receptors that influence neuronal activity, or the transcription factors that regulate their expression, may also strongly contribute to understanding the cell-types functionality in context, in a data-driven way, rather than relying on prior hypotheses.

The function of a great proportion of genes expressed and cell types discovered, however, has usually not been characterized. Therefore, once molecular cell identities are established, (combinations of) specific new and known markers reflecting key biological components can be used to profile a certain cell type under different conditions. For instance, engineered mouse models, such as transgenics, Cre- or other recombinase lines allow examining cell types more comprehensively by their physiology and activity. This can be combined with virus transductions and transsynaptic labeling to further study connectivity. Behavioral outcomes can be studied using optogenetic or chemogenetic approaches (Poulin et al., 2016). In summary, discoveries made by single-cell transcriptomics, and cell-type taxonomy of the nervous system generated this way, opens up an unprecedented array of opportunities to answer neuroscientists' questions.

3 DEVELOPMENT AND DYNAMICS OF THE NERVOUS SYSTEM

The functionality of the nervous system depends on the precise interconnections of many million neurons, made up of a yet unknown number of distinct cell types. Underlying this sheer complexity are tightly regulated developmental processes, made up of a combination of spatial and temporal patterns of triggering external sources, and intrinsic cellular factors responsive to the inducers (summarized from (Kandel et al., 2000)).

3.1 NEURULATION

Developmentally, the nervous system arises from the ectoderm in the gastrula stage, when a sheet of cells along the dorsal midline, called the neural plate, begin to fold into a tubular structure during neurulation. The neural tube formed during this process will give rise to the spinal cord caudally, and the brain rostrally. The extent of proliferation of these neural precursor cells is rapid, and non-uniform along the rostrocaudal axis; giving rise to specialized regions of the mature nervous system. Rostrally, the forebrain, midbrain and hindbrain initially form three vesicles. The forebrain vesicle later further divides to give rise to telencephalon and diencephalon, and the hindbrain vesicle divides to metencephalon and myelencephalon. These four subdivisions as well as midbrain and spinal cord make up the six major regions of the mature central nervous system. The hippocampal formation, discussed in more detail in postnatal development in this thesis (Paper IV), arises from the telencephalon, as do other structures such as cortex, amygdala and basal ganglia; studied here in the context of oligodendrocyte heterogeneity (Paper I). Also the diencephalon, represented by thalamus (zona incerta) and hypothalamus, the midbrain and spinal cord were included in the oligodendrocyte study.

3.2 NEURONAL PATTERNING

An individual cell's fate in this differentiation process is, as mentioned, determined in part by the long range and local signals it is exposed to (largely a consequence of its position), and in part by its developmental gene expression profile of surface receptors and transcription factors. On the signaling side, the fate of neural cells is controlled by two independent systems, patterning the neural plate along its medial-to-lateral (later dorsoventral) axis, and along the anteroposterior axis. Patterning along the dorsoventral axis mainly results in the two anatomically and functionally segregated neuronal circuits of the spinal cord; the sensory (dorsal) and motor (ventral) circuits. To give an example in the ventral half, sonic hedgehog (SHH) secreted from notochord and floorplate is both necessary and sufficient for the induction of most cell types. It acts as a morphogen; determining the different cell fates of

motor- or interneuron in a concentration-dependent manner. Dorsal tube patterning is achieved by secretion of BMPs, giving rise to neural crest cells, roof plate cells and dorsal interneurons. Of note, neural crest later gives rise to a diverse cell lineage, including the peripheral nervous system discussed in the context of cell type discovery in the sympathetic ganglion in this thesis. Briefly, during neurulation, neural crest cells from the neural tube begin to migrate along different paths through the periphery. Medially migrating cells form the sympathetic ganglia, while others form as diverse lineages as the adrenal medulla, sensory neurons of the dorsal root ganglia, parasympathetic and enteric neurons as well as melanocytes.

Turning to the second, anteroposterior direction of patterning of the neural tube, this is achieved in several stages. While induction of posterior neural tissue is achieved early on by retinoic acid and genes belonging to the fibroblast growth factor (FGF) family, patterning of the hindbrain serves as a useful model to understand how rostrocaudal organization into segments is specifically achieved. Here, members of the *Hox* gene family are differentially and combinatorially expressed between rostrocaudal swellings, or rhombomeres. Specific expression patterns are restricted by discrete boundaries between these rhombomerers and lead to preferential development of specialized neurons in each segment, such as trigeminal or facial motor neurons.

On top of ensuring differentiation of appropriate numbers and types of neural types discussed above, the specificity of functional synaptic connections needs to be established to form the mature nervous system. Briefly, this is initially achieved by the complex process of axon guidance, mediated by environmental cues and the corresponding specific axonal receptors. Then, selective connections need to be established between the axon and its target, the axon differentiates its growth cone into an axon terminal and the target cell establishes the postsynaptic apparatus. As previously discussed processes, also this is critically dependent on fine-tuned intercellular interactions, leading to changes in gene expression and functional specialization. In our study of the sympathetic nervous system (Paper II), for instance, we showed that although born embryonically, certain neuronal types first differentiate upon target innervation, several days into postnatal development. This suggests that differentiation to specialized mature cell types can be instructed by factors produced by the target.

3.3 GLIOGENESIS

Having focused on the specification of neurons by patterning, parallels have been observed also for non-neuronal macroglial cells: astrocytes and oligodendrocytes (reviewed in (Rowitch and Kriegstein, 2010)). Given that the proportions of these cells increase alongside increased neurological complexity in evolution underlines their crucial, diverse functionality. In contrast to microglia or brain vascular cells derived from mesoderm, astrocytes and oligodendrocytes, like neurons, are derived from the neuroepithelium. In embryonic development, radial glia cells are considered the primary progenitor cell, giving rise to both

neurons and macroglia. In mice, neurogenesis begins around E9, at which point mechanisms are in place that actively repress gliogenic programs. Later, however, the same progenitor domains switch from generation of neurons to gliogenesis. This switch is achieved both by the downregulation of proneural factors such as *Neurog2* (NGN2) or the NOTCH program and activation of pro-gliogenic programs such as SOX9.

Data especially from ventral spinal cord development indicates that similar segmented organizing principles diversifying neurons in development hold true for glia, too. For instance, much like in the development of motor neurons, SHH is both necessary and sufficient for production of oligodendrocytes in the spinal cord, in a time- and concentration-dependent manner, sharpening and maintaining ventral domain boundaries. Also, OLIG proteins control both motor neuron and oligodendrocyte development in the ventral pMN domain, in a mutually exclusive fashion with the development of astrocytes and certain interneurons, restricted to other developmental domains. For dorsal regions of the spinal cord as well as the forebrain, less is known as to whether this segmental model holds true for gliogenesis. Further, partly owing to a lack of marker genes unambiguously distinguishing astrocyte progenitors from radial glia, less is known about the timing and patterning of astrogenesis. Thus, the relationship between mature astrocytes and their origin in the neural tube, as well as their functional heterogeneity remains less studied.

3.4 POSTNATAL AND ADULT NEUROGENESIS

Oligodendrocytes are continuously replenished in the postnatal and adult mammalian brain through the proliferation of oligodendrocyte progenitor cells (OPCs). Astrocytes, on the other hand, are largely considered quiescent, but can rapidly proliferate upon injury. Certainly, however, neurogenesis is largely considered to be restricted to embryonic development, with some exceptions of continued early postnatal developmental neurogenesis. As early as the 1960s, Joseph Altman and colleagues performed an array of seminal studies ((Altman and Das, 1965), reviewed in (Altman, 2011)) based on autoradiography with ^{3}H-thymidine in rats, guinea pigs and cats. Here, dividing cells incorporate the radioactive epitope in cell division and can be detected in tissue sections as label retaining cells. Further, early progeny of thus labeled cells retains low levels of ^{3}H-thymidine, allowing tracing of newly generated cells to a certain extent. Soon, the group found that generation of the most abundant neurons of the brain, cerebellar granule neurons, first peaks in the first postnatal week in rats and mice, and their progenitors remain proliferative until postnatal day 21. Similarly, up to 90% of olfactory bulb granule cells, generated in the subventricular zone (SVZ), were shown to be born postnatally. Intriguingly, Altman and colleagues also found that the SVZ stays neurogenic, feeding newborn neurons via the rostral migratory stream to the olfactory bulb throughout adulthood. Similar dynamics applied to the dentate gyrus granule cells of the hippocampal formation. Only a small proportion, approximately 15% of granule cells, was found to be generated embryonically. Instead, their production peaks in the first postnatal week in rat. As

in the SVZ, neurogenesis then slowly decreases, but is maintained at a low rate in the dentate gyrus subgranular zone (SGC) also in adult animals.

Given the revolutionary nature of these studies, they were not initially widely accepted by the community. It was mainly through the introduction of new, simpler technologies, the use of BrdU to generate label retaining cells, on the one hand, and a quest and discovery of new molecular markers to label proliferating, differentiating and maturing neurons, on the other hand, that propelled the field forward. Only recently, significant turnover of dentate gyrus granule cells could also be shown for adult humans, using ^{14}C cell birth dating (Spalding et al., 2013). Since, the function of adult neurogenesis has been further explored. In the hippocampus for instance, adult generated neurons have extraordinary plasticity during maturation and contribute, among others, to pattern separation in spatial memory formation. Lack of neurogenesis has been linked to depressive behavioral patterns, while physical activity and enriched environments lead to increased neurogenesis in lab animals (reviewed in (Gonçalves et al., 2016).

3.5 HOW CAN SINGLE-CELL TRANSCRIPTOMICS HELP UNTANGLE DEVELOPMENT?

Cell identities can be transient in time and as they progress along (developmental) trajectories, or as they oscillate in cell cycle. Bulk studies need to either rely on the synchronization of a population, or isolation of substates based on predefined marker genes. Apart from overlooking subtle differences across the spectrum, this neglects that in terms of gene expression, dynamic cellular phenotypes are frequently characterized by continuous, rather than sharply changing marker expression levels. Here, the great advantage of single-cell approaches becomes immediately apparent, as they can without bias measure continuous spectra at cellular resolution. This is because single-cell approaches of a reasonable size are more likely to stochastically sample cells positioned along the entire trajectory.

Single-cell transcriptomics has therefore become instrumental and generated great insights, especially in the context of development. A landmark study examined alveolar differentiation in lung development, discovered unknown specific markers and reconstructed molecular steps of progenitor cell maturation (Treutlein et al., 2014). The first example in the developing brain dissected cortical development, identified rare progenitors, early neurons and transient intermediates in an unbiased way and envisioned the possibility to reconstruct entire developmental lineages by single-cell RNA-seq (Pollen et al., 2014). Single-cell RNA-seq in combination with time-resolved whole embryo dissections also greatly aided the study of the classic model organism in development, *C. elegans*, to allow new conclusions about the emergence of endoderm and ectoderm across phyla (Hashimshony et al., 2014). Recently, the development of the ventral midbrain was elucidated in detail in human and mouse; particularly relevant for the field of Parkinson's disease research, where knowledge of the molecular dynamics is of great interest to develop cell replacement therapies (La Manno et al., 2016).

In development, the cell ground state is determined shortly after the cell exits cell cycle, followed by a critical period of maturation until the stabilization of gene expression in the mature cell type (Fishell and Heintz, 2013). During differentiation, progenitors undergo temporarily regulated changes reflected by their expression of transcription factors in combination with external morphogen gradients. The variety and combinatorial possibilities in both of these aspects can generate a large diversity of nervous system ground-state identities, as reflected in the molecular profiles of differentiating and mature cells. Nevertheless, it is possible to overestimate the true number of final cell types by projecting from the developmental perspective. For instance, a great machinery of developmental control may be necessary to place the right number of motor neurons in the right position of the spinal cord ventral horn and establish correct projections and connectivity, while molecularly the resulting motor neurons may be largely of one single type.

Dynamic processes such as differentiation are reflected in the molecular profile of single cells. In a typical developmental dataset, there are thus several analysis methods applied that help introduce an understanding of its underlying dynamics. Since no actual temporal information is usually collected in a single-cell snapshot-type experiment, researchers resort to 'pseudotemporal' ordering. These types of approaches give a scalar measure of a cell's progress along a temporal trajectory, by fitting an optimal path based on the similarity of the cells in the dataset. Examples include Monocle, Wanderlust and Wishbone (Bendall et al., 2014b; Setty et al., 2016; Trapnell et al., 2014), some allowing splits in the dynamic trajectory to represent differentiation to separate fates, or user defined starting points, to infer directionality. Necessarily, these methods assume smooth temporal transitions between states, and that all intermediate states are represented in the dataset (Wagner et al., 2016). Visualization of continuous datasets by t-SNE provides an initial intuition on the underlying dynamics (Marques et al., 2016).

In our studies of developmental dynamics in the dentate gyrus (Paper IV), we additionally applied a mutual KNN-based clustering, which by connecting mutual *k* nearest neighbors by edges allows inference on the succession of states across a trajectory. Groups of most similar cells remain intact clusters, and closely related clusters will appear to have multiple edges between them, often linked by small numbers of cells in intermediate states, suggesting a progression. This may help remove uncertainties on what kind of transitions are possible in a dataset. Importantly, however, the directionality of successions cannot be established by any such algorithms, but must be imposed with the help of additional information on the biology of the dynamic system.

4 RESULTS

4.1 PAPER I: OLIGODENDROCYTE HETEROGENEITY IN THE MOUSE JUVENILE AND ADULT CENTRAL NERVOUS SYSTEM.

Oligodendrocytes allow for rapid signal propagation along projections in the central nervous system by ensheathing axons with myelin and providing insulation to neuronal conduction. Heterogeneity in this population, besides the oligodendrocyte progenitor (OPC) and mature oligodendrocyte, was largely uncharacterized. In this study, we sampled more than 5000 oligodendrocytes across 10 brain regions by single-cell RNA-seq. We describe 12 distinct cell states as a continuum of precursor cells to mature myelinating cells, with several intermediate stages along the maturation trajectory, characterised by distinct molecular profiles.

Strikingly, OPCs as well as all stages in maturation were molecularly homogenous across brain regions, whereas the final mature states were more diverse. Nevertheless, the degree of myelination in each region determined the proportions of the separate populations present. For instance, while OPCs and mature oligodendrocytes were similarly represented across the regions, intermediate stages were largely absent from less myelinated areas; likely reflecting a lower rate of proliferation and maturation compared to highly myelinated areas.

The study reveals new insights into the diversity and maturation dynamics of this highly prevalent neural cell population and may contribute to a better understanding of demyelinating diseases, leading to new therapeutic approaches.

4.2 PAPER II: VISCERAL MOTOR NEURON DIVERSITY DELINEATES A CELLULAR BASIS FOR NIPPLE- AND PILO-ERECTION MUSCLE CONTROL

The autonomic nervous system, consisting of the sympathetic and parasympathetic divisions, controls involuntary functions. Interestingly, its neurons regulating diverse actions through smooth and cardiac muscle fibers and glands have long been considered non-specific. In recent years, however, evidence has emerged about selective action of separate components, such as thermoregulation via sweat glands or increase in cutaneous vasoconstriction, pilo-erection and nipple erection during hypothermia.

Using single-cell RNA-seq on the stellate and thoracic sympathetic nervous system, we shed light on the transcriptome-wide molecular landscape of sympathetic neurons, previously confined to a small number of known markers. We identified seven distinct sympathetic neuronal populations, broadly characterised as noradrenergic or cholinergic. Based on this, the study focused on further dissecting the separate circuitry, and specialized cell types that underlay differential activation of nipple- and pilo-erection. Retrograde tracing from the respective target organs together with combinatorial staining with markers identified through

the single-cell RNA-seq data revealed two distinct, target-specific populations as erector muscle neurons. In addition, differentiation of these cell types was found to happen only postnatally, upon target innervation, even though the neurons were born embryonically.

Among others, this study represents an intriguing view of the potential of single-cell transcriptomics beyond cellular taxonomy; combining new insights into functional delineation and developmental dynamics with underlying cell type diversity.

4.3 PAPER III: STRT-SEQ-2I: DUAL-INDEX 5' SINGLE CELL AND NUCLEUS RNA-SEQ ON AN ADDRESSABLE MICROWELL ARRAY

In an attempt to adapt 5′ single-cell RNA-seq to meet new demands in the field, we developed a 9600-well multiarray in collaboration with Wafergen. The study is largely focused on the molecular design and adaptation of our previously published STRT-seq method to this new format, towards a 5′ dual-index protocol called STRT-seq-2i. This platform allows for greater flexibility and high throughput while keeping costs low and staying compatible with post-capture cell imaging. We tested the technical performance of the method and collected data both after single-cell dispense by limiting dilution, as well as targeted FACS sorting, yielding up to ~2800 or ~7500 single cells per plate, respectively.

We further reproduce previous data of cell type heterogeneity in the adolescent mouse somatosensory cortex. Importantly, we also present data from human neuronal nuclei, extracted post-mortem from the middle temporal gyrus, sorted by NeuN+ staining, frozen, shipped and thawed before dispense and processing. This opens up new opportunities to study such rare and precious samples at high throughput.

In the light of the rapid advancement of droplet microfluidic platforms that easily achieve much higher throughput, the competitive edge of this type of platform thus lies in its flexibility, the compatibility with imaging, FACS and index sorting. Further, up to eight samples can be processed in parallel on the same chip, keeping per-sample costs low for biologists interested in medium-sized studies. In addition, we argue that due to its flexible nature, the platform will prove useful also in the context of scaling up other single-cell applications, such as full-coverage transcriptomics by Smart-seq, whole-genome applications or chromatin and transcription factor assays.

4.4 PAPER IV: POSTNATAL NEUROGENESIS IS MOLECULARLY CONSERVED IN THE ADULT MOUSE DENTATE GYRUS

The dentate gyrus is one of two regions in the brain described to have maintained neurogenesis in the adult and has therefore been of great interest also in the context of regenerative medicine. Initial, developmental neurogenesis to a large extent takes place early postnatally. Here, we therefore examined neuronal lineage dynamics in the mouse dentate

gyrus in independent samplings of thousands of single-cell transcriptomes across postnatal development into adulthood; from P8 to P68.

In contrast to previous single-cell RNA-seq studies in this tissue, which describe neurogenesis as a molecular continuum, we found that it proceeds through a series of discrete, well-defined cell types (or states). We found populations of quiescent (radial glia-like, RGL), distinct from astrocytes, and dividing (neuronal intermediate progenitor cells, nIPC) cells. Further, we describe a maturation trajectory through early neuroblasts stages, where frequencies of early and more mature neurons change drastically across age. Interestingly, all stages of neurogenesis were molecularly indistinguishable during early postnatal neurogenesis compared to adult. Importantly, we describe a number of marker genes more specific to separate stages of maturation than some of the ones described in the literature. This opens possibilities to study progenitors and maturation with higher accuracy than previously.

4.5 PRELIMINARY RESULTS: DISSECTING THE NERVOUS SYSTEM BY SINGLE-CELL RNA-SEQ

With single-cell RNA-seq technologies rapidly advancing, allowing for unprecedented throughput and good data quality at lower per-cell cost, it has now become feasible to perform high-throughput studies to dissect molecular cell types across the entire brain. This would allow gaining fascinating insight into molecular cell-type diversity across the nervous system and providing a resource to researchers interested in particular regions, types or genes. This scale of study also allows to answer more systematic questions, partly yet to be formulated.

We thus set out to systematically and without bias dissect cell-type heterogeneity in the whole mouse adolescent nervous system, and performed dissections at coarse anatomical resolution covering 19 distinct regions in the central, peripheral and enteric nervous system (Fig. 5). In total, we sampled around 400,000 cells on the 10X Genomics Chromium platform, sequenced at a depth of about 50,000 reads per cell, putting us in front of unprecedented challenges in analysis. We used the graph-based clustering Louvain Jaccard KNN algorithm (described in Chapter 1.7.3) and have identified 659 distinct cell types, manually curated into broader categories, and reanalyzed in an unbiased fashion using a trained classifier. Deeper analysis and validation of cell types is yet incomplete, and the number of cell types is likely to increase as we refine the clustering, but also decrease as we eliminate artefacts and merge cell types that are the same across brain regions. Figure 5 gives an overview of the contribution of the broader cell-type categories to each region. Immediately evident, the proportion of glia to neurons was mainly dependent on the relative contribution of oligodendrocytes, directly reflecting the degree of myelination in any given region. As expected, oligodendrocytes were more prevalent in ventral and caudal regions, reflecting both their relative degree of maturation (oligodendrocytes develop earlier caudally) and the eventual degree of myelination of those tissues.

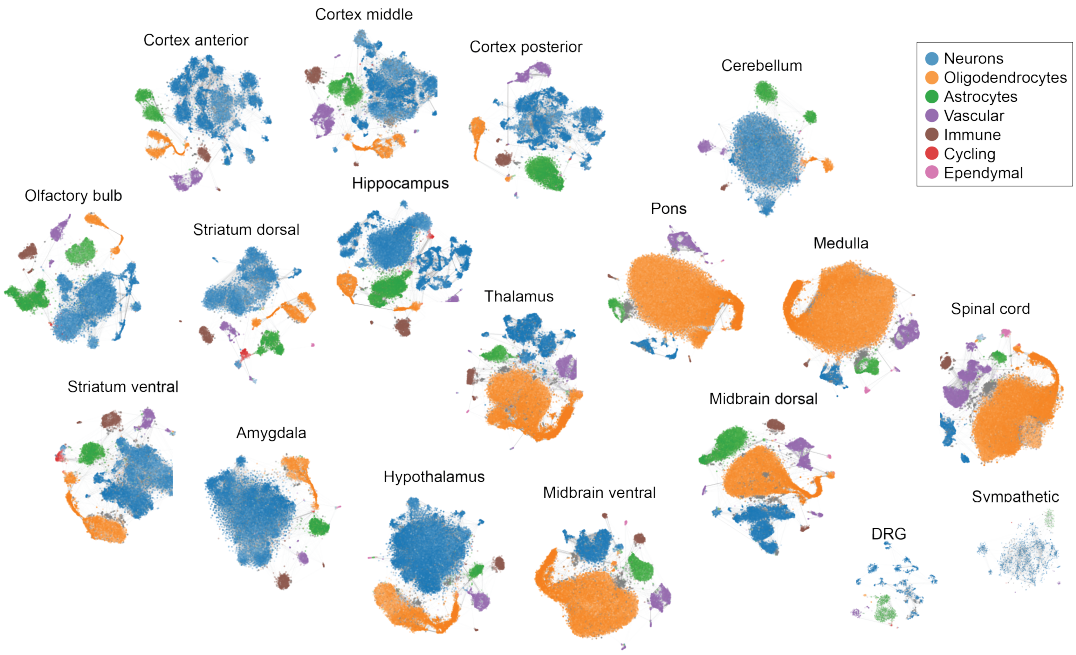

Figure 5 | Overview of sampled regions in and outside the central nervous system by t-SNE, each dot represents a cell, broad cell-type category stains the graphs.

Summarizing some highlights found in preliminary analysis of this dataset, we discovered great regional heterogeneity of neurons, while the diversity of glial, microglial and vascular cell types was lower and most types were broadly distributed across most regions, especially within the central nervous system; discussed below.

4.5.1 Neuronal diversity across the CNS, hippocampus as example

Across the full dataset, the greatest diversity of cell types is undoubtedly among the neuronal clusters. Although many histological features are shared across the nervous system, clustering of neurons revealed a strong regional bias, indicating the specialization of molecular cell types required to carry out distinct functions. It is too soon in our analysis to give a confident, detailed picture of neuronal cell type diversity, but I have chosen to show hippocampus neurons as an example of both dynamic, continuous and distinct cell types.

We sampled the full hippocampal formation (excluding entorhinal cortex), included the temporally resolved dentate gyrus dataset (P12-P35, Paper IV), in addition to deeper sampling of CA1 interneurons by a targeted FACS approach (based on vGat2-tdTomato mouse (Ogiwara et al., 2013)). Figure 6A visualizes 5000 neurons (downsampled from the full dataset), overlayed by colors representing Louvain Jaccard KNN cluster identity. Non-neurons such as astrocytes, radial glia-like cells and nIPCs were excluded from this analysis. Besides the dentate gyrus granule cell lineage, pyramidal cells from CA1, 2 and 3 make up the main excitatory cell type. Coherent with previous single-cell studies (Tasic et al., 2016; Zeisel et al., 2015), inhibitory interneurons showed the greatest degree of diversity, with distinct types broadly categorized by their expression of known markers such as *Sst, Pvalb, Htra3, Npy, Vip,* or *Cnr1* (Fig. 6B).

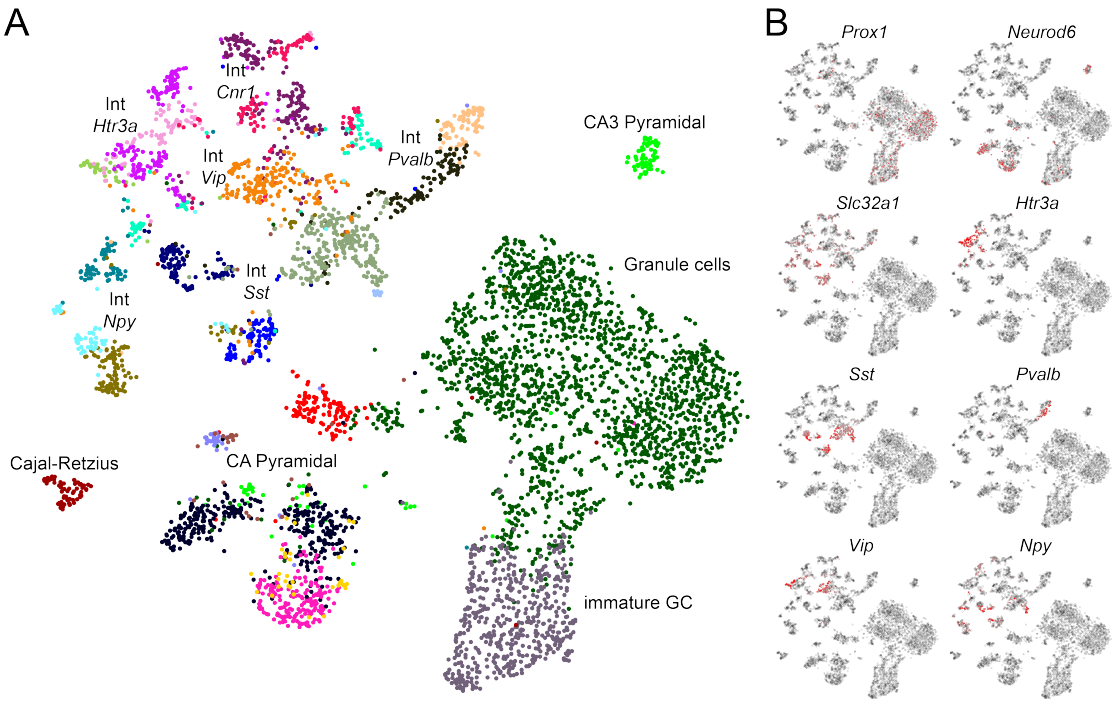

Figure 6 | Accumulated hippocampus data, including dentate gyrus sampling from P12-P35 (Paper IV) and CA1 FACS-sorted interneurons P27 and P60. Dentate gyrus granule cells form a maturation trajectory, inhibitory interneurons show greatest diversity, with all major cortical interneuron types described before (Zeisel et al., 2015) **represented.**

4.5.2 Oligodendrocyte maturation conserved across the CNS

Cells of the oligodendrocyte (OL) lineage made up the largest number of sampled cells in the dataset, especially owing to the deep sampling of strongly myelinated hindbrain regions such as the pons and medulla. In total, almost 200,000 cells belonged to the oligodendrocyte lineage. Figure 7A is a t-SNE plot of over 12,000 OLs with equal sampling across Louvain Jaccard KNN clusters. Very similarly to our previously published study of ~5000 oligodendrocytes across 10 regions, OLs followed a maturation trajectory from oligodendrocyte progenitor cells (OPCs) via committed progenitors (COPs), newly formed (NFOL) and myelin forming (MFOL) oligodendrocytes to several mature types, not yet characterized in more detail. Markers for each of these stages were again found highly conserved and restricted (Fig. 7B). In addition, we found the myelinating Schwann cells of the PNS (derived from dorsal root ganglia and sympathetic system) to be a separate type from all OLs in the CNS, characterized by a separate cassette of gene expression (e.g. *Mpz*).

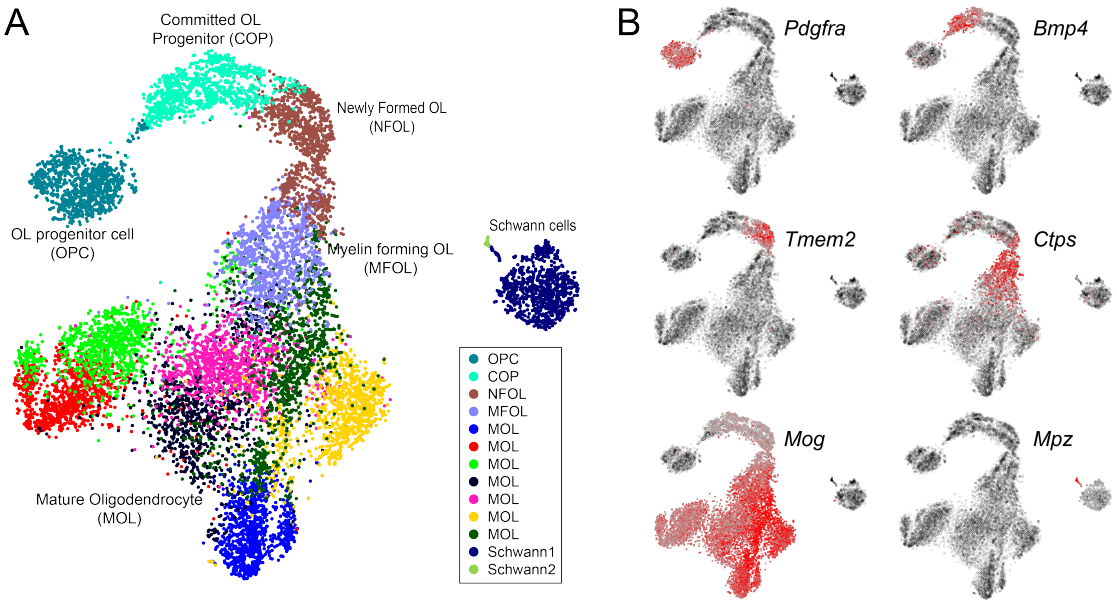

Figure 7 | Oligodendrocyte (OL) clusters follow the same main dynamics described previously (Paper I), across deeper regional sampling. (A) Louvain Jaccard KNN clusters stain t-SNE plot of a downsampled OL population, derived from all sampled regions. (B) t-SNE stained by previously identified markers along the OL maturation trajectory.

Plotting the region of origin on the same t-SNE (Fig. 8A), especially early maturation stages were highly conserved across the sampled regions. This is particularly striking for OPCs, where no regional differences were identified across the cluster. Further, strongly myelinated areas were more likely to have ongoing maturation, with cells spread across the trajectory (eg. pons, medulla, thalamus and hypothalamus), rather that restricted to OPC and mature populations. We observed polarity to some extent in the maturation trajectory between the hindbrain regions of pons and medulla versus the rest of sampled CNS regions. While expression of some genes seems to explain part of this polarity (Fig. 8B, *Cd81, Il1rap*), we cannot rule out a batch effect, since pons and medulla were both sampled with a newer version of 10X Chromium kits than the rest. Further, mature OLs of the heavily myelinated spinal cord formed a separate trajectory, strongly supported by the restricted expression of *Klk6* (Fig. 8B).

It should be noted that a pitfall of t-SNE visualization lies in exaggerating or underestimating distances, dependent on the number of datapoints plotted per type. Here, for instance, the overall difference in gene expression space between OPCs and the remaining OLs appears smaller than the difference among the mature clusters. This is not reflected in actual gene expression, where mature types are overall highly similar, merely distinguished by expression level differences, rather than entirely separate markers, as is the case for OPCs.

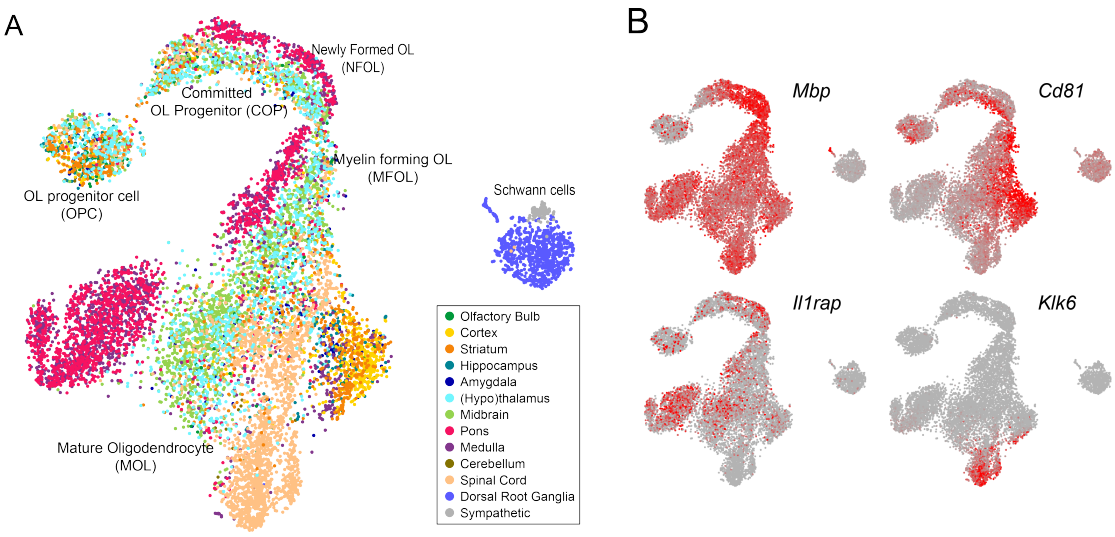

Figure 8 | Regional differences between oligodendrocytes. (A) t-SNE nodes stained by region of origin (summarized to broader categories for clarity and visualization). OL progenitors (OPCs) are similarly represented across regions and do not form subclusters. Along the maturation trajectory, some regional differences emerge, possibly influenced by batch effects. (B) General and more restricted markers explaining some of the regional variation in mature OLs.

4.5.3 Astrocytes heterogeneity and developmental patterning

Among glial cells, astrocytes showed the greatest heterogeneity (Fig. 9A). All cells (except three *Foxj1*+ ependymal cell clusters included in this analysis) expressed pan-astrocyte marker *Aldoc* (Fig. 9C). Two main 'clouds' of astrocytes were distinguished by their enriched expression of *Agt* or *Mfge8*, and each of these types (although not fully overlapping) had a subpopulation of fibrous *Gfap*+ astrocytes (Fig. 9C, D). Bergman glia of the cerebellum and olfactory bulb (OB) ensheathing glia formed distinct clusters and were characterized by the specific expression of several genes (Fig. 9B). In olfactory ensheathing glia we found modular expression of astrocyte genes as well as OL genes (e.g. *Plp1*) possibly sharing a role in ensheathing, but not other genes characterizing the OL lineage such as *Mbp* and *Mog*, prominently involved in myelination. Uncharacteristically of other glia, they also express cell adhesion molecules enriched in endothelial types (eg. *Vtn*) and the neuropeptide *Npy,* as reported previously (Ubink et al., 2003). A small group of radial glia-like cells (RGLs) from the hippocampus and dentate gyrus clustered closely with other astrocytes, emphasizing their glial identity as neuronal progenitors (Kriegstein and Alvarez-Buylla, 2009). In this dataset commonly used RGL marker *Fabp7* (also know as *Blbp*) did not distinguish RGLs from other astrocytes, mainly due to the prominent expression of *Fabp7* in cerebellum Bergman glia. Expression of *Tfap2c*, a dentate gyrus RGL marker identified in Paper IV, remained unique to this population (Fig. 9C).

Returning to the two main populations (Astro-*Mfge8* and Astro-*Agt*), we could see a strong regional bias of *Mfge8*+ astrocytes in the forebrain. More specifically, this type was almost entirely restricted to regions developmentally derived from the telencephalon. The *Agt*+ population comprised a mixture of diencephalon, midbrain, hindbrain and spinal cord astrocytes.

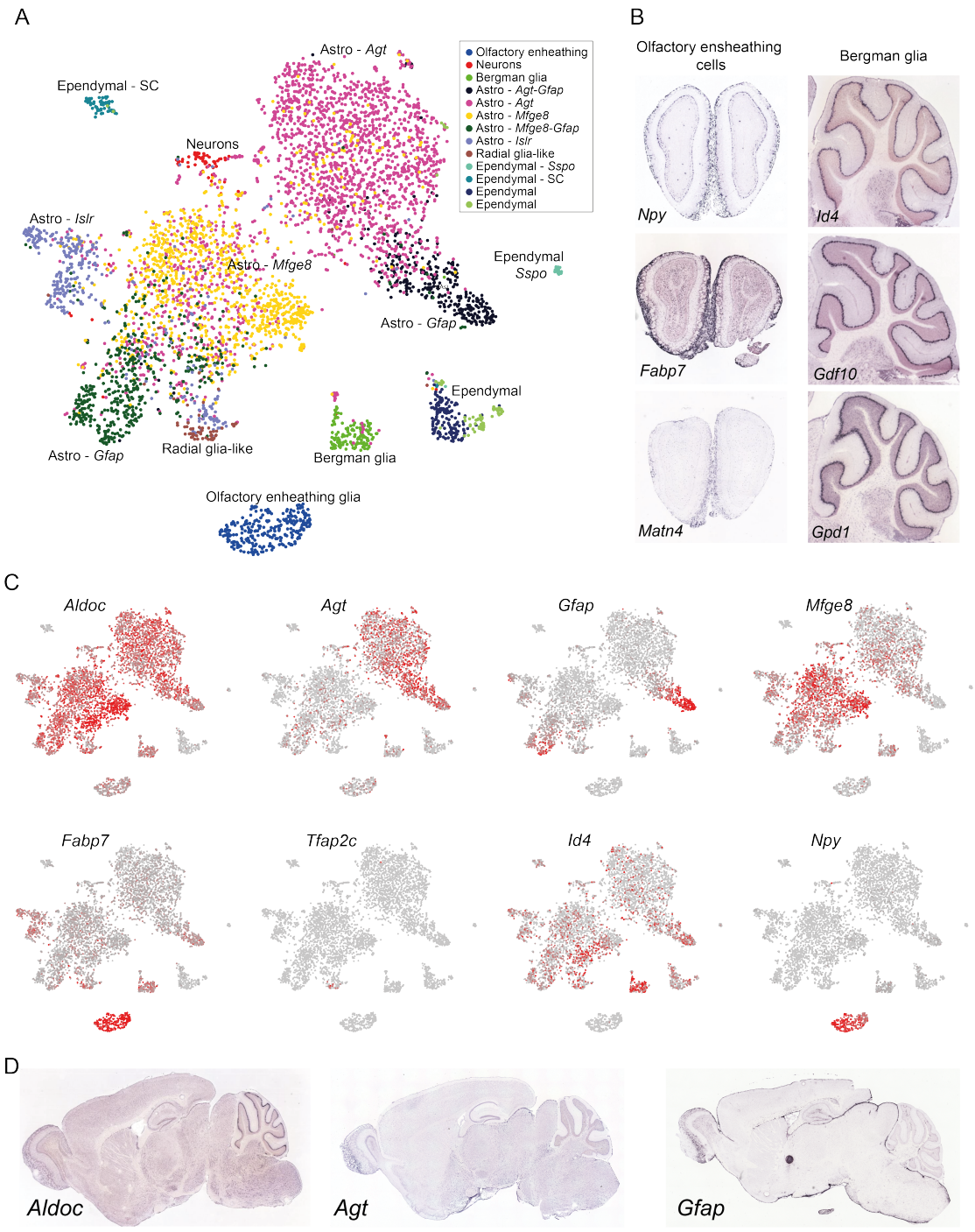

Figure 9 | Astrocytes across the central nervous system. (A) t-SNE plot of 5000 cells from downsampled astrocyte and ependymal populations, colored by Louvain Jaccard KNN cluster, (preliminary names). Some apparent regional heterogeneity. (B) *In situ* hybridisations (Allen Mouse Brain Atlas) of distinct astrocyte types in olfactory bulb (olfactory ensheathing glia) and cerebellum (Bergman glia), with identified markers. (C) Expression of marker genes across, or specific to, astrocyte subpopulations, visualized by t-SNE. (D) Allen Mouse Brain Atlas *in situ* hybridizations of general astrocyte markers. *Aldoc* is expressed across the brain, *Agt* is enriched in posterior structures. *Gfap* is expressed by fibrous astrocytes and enriched in white matter and pial surfaces. Image credit (B) and (D): Allen Institute.

Taken together, these results suggest a certain extent of developmental patterning of astrocytes that is maintained in the mature brain, yet to be investigated and validated further. Perhaps this patterning is more strongly reflected by distinct cell types than that of

oligodendrocytes, studied in greater detail through development (reviewed in (Rowitch and Kriegstein, 2010)).

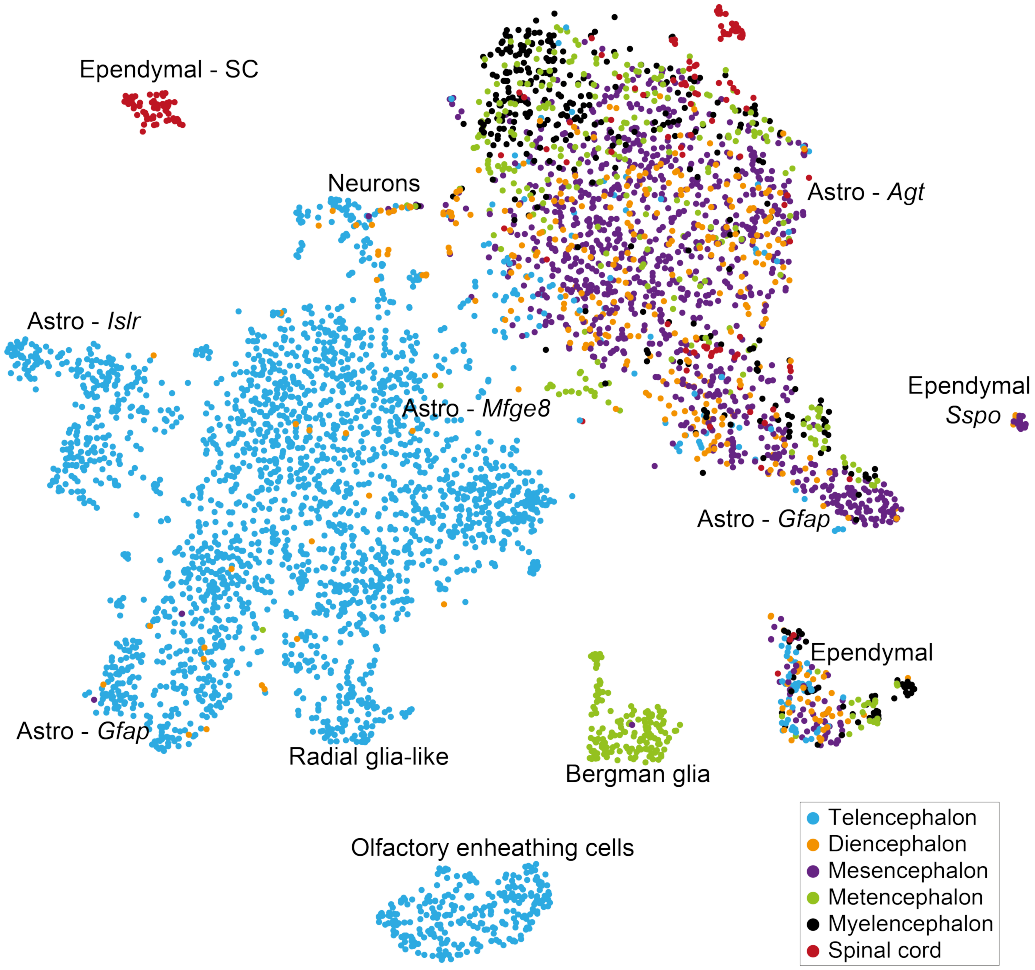

Figure 10 | t-SNE plot of astrocyte types, stained by developmental origin of the sampled region each cell was derived from. Between the two 'main clouds' of astrocytes (the *Agt* and *Mfge8* clusters in Fig. 9A), the major distinction is based on telencephalon versus rest of the CNS. Two clusters of ependymal cells included here split between spinal cord and rest of the CNS.

4.5.4 Conclusions

Although still in the early phases of exploration and validation, this dataset already promises a wealth of questions to be addressed on an unprecedented systematic scale. Importantly, our lab simultaneously generated a detailed temporally resolved dataset of embryonic development, covering the entire brain and sampling every gestational day from E7.0 to E18.5. Combined, these datasets have the potential to further resolve dynamics in development and maturation, shedding light on the multitude of cell types that make up the complex mammalian nervous system. It will require independent validation by us and others to untangle many of these novel findings. Importantly, as for understanding any single-cell transcriptomics data, we are strongly building on the plethora of pioneering literature generated through decades of studying the nervous system. After all, if we are to see further, it is by standing on the shoulders of giants.

ACKNOWLEDGEMENTS

Sten, thank you for creating – and giving me this opportunity to work in – a group and an environment unique in science. Your vision, approach and personality are a rare, lucky combination.

To my co-supervisors, thank you **Patrik** for your invaluable inputs and wide perspective on several projects and collaborations, as well as ever entertaining lunch discussions. Thank you **Jens** for your discussions and expertise on keeping these neurons alive. Thank you **Clas** for your outstanding dedication and everything you taught me.

Thank you **Arnold Kriegstein** for taking your time out of an incredibly busy schedule to be my opponent, it's a true honor! To **Eva Hedlund**, **Thomas Perlmann** and **Ann-Christine Syvänen**, thank you for your expertise on my examination board, and **Mitya** for narrating this big day for me as defence chairman.

To the very best group out there, thank you Linnarsson lab members. To **Amit**, thank you for your cookies, your tireless efforts, creativity and being the best teacher and collaborator one could ask. **Anna Jo**, what would this place be without you? Thank you for your help, comments, organization, distraction, passion. To **Peter**, for your never-ending patience, and excellent harpsichord and organ concerts. Thank you **Gioele** for your inspiration, enthusiasm and lots of fun memories! Thanks **Simone C** for your great spirit and knowing everything there is to know in a lab, like, ever. To **Lars**, it's a great pleasure you joined the lab, bringing new ideas inside and outside of science (including some of the best music Molneuro has heard) and slowing down lunches considerably. Thank you **Job** for new perspectives, challenging us biologists. Thank you **Anna Ju** for your help, perfectionism and staying on top of all the oligos. **Marcela**, thank you for your smiles, warmth and hard work, **Karolina** for your support and fun spirit and **Simone P** for sharing all your knowledge on molecular biology, running and food. New members **Emelie**, **Anastasius**, **Henrik** and **Samaneh** for keeping the single-cell spirit going. Thank you past members, **Kasper** for sharing office and laughs, **Pawel**, **Una** and **Saiful** who I've had the pleasure to work with in the early days and **Abeer** for all your help.

A heartfelt thank you to Molneuro for being a fantastic workplace, in light and tough times and to all Molneuro PIs, **Ernest, Patrik, Jens**, **Goncalo**, **Ulrika** and **Per**, thank you for your commitment to building and maintaining this spirit. Without you, **Alessandra**, **Johnny** and **Ahmad** the place would fall apart, thank you for your dedication!

Dear **Dagmara**, thank you for not running with me, not swimming with me, and not watching GG with me. Thank you for your friendship, I will never forget (and always judge you by) our very first conversation. To **Ana F**, thank you for your kindness, honesty, deep and light conversations and smiles. Thank you **Martin** for excellent suspensions, laughs and your easygoing spirit. **Ivar**, thank you for all your useful (and useless) advice, and **Lili** for

being the cheerful, sweet office mate you are. Thank you **Carolina** for your never-ending energy and active-wear-Sundays, **Ana M** for all your kindness and help, **Sueli** and **Fatima** for your warmth, smiles and food. Thank you **Shigeaki** and **Boris** for the late-lunch company (and Boris for introducing me to chocolate cake at Meridionale). Thank you **Karol** for smoothies, lots of laughter and the unforgettable frequency of your voice, **Mandy** for your sweet spirit and **Daniel** for your trashy, danceable tunes. Thank you **Alca** and **Dahoua** for fun conversations, **Lauri** for sharing hummus, **Viktoria** for keeping the Austrian in me alive, **Elisa, Carmen** and **Jana** for your warmth, help and enthusiasm. Thank you **Mitya** for the best impersonation of a Soviet officer KI has seen and **Enrique** for keeping the good old days alive. **Ruani**, thank you for making Molneuro my first scientific home at KI. Thank you **Alessandro** for your skills and sarcasm, **Moritz** for continuously crossing paths through these years and **Roman** for giving Vienna a chance. **Pia**, **Hind**, **Manuel**, **Carlos**, **Blanchi**, **Anneke**, **Nathan**, **Göran**, **Connla**, **Songbai**, **Erik**, **Sam**, **David**, **Eneritz**, **Dawei**, **Tanja**, **Puneet**, **Nobuyuki, Yiwen**, **Kasra**, **Khom**, **José**, **Laura**, **Dongoh**, **Marek**, **Kaneyasu** all you lovely people at Molneuro, past and present, thank you for making this a great place to spend so much of our time at.

Thank you **Nadia**, **Therese** and **Martina** for always staying on top of everything, even my last minute orders. To **Jaromir**, thank you for your mean FACS skills (and crazy stories), half of what we did just wouldn't have worked without you.

To **Milena**, thank you for lunches, coffees, talks, distraction and laughs. CMM veterans **Oskar** and **Daniel**, thank you for being social at L8:01 and the sporadic but fun beers and lunches since. **Yu Li**, thank you for your sweet spirit, it was lovely working with you. Thank you **Gilad** and **Anna** for your support in critical situations, and lovely dinners. Thank you **Natasha** for so briefly sharing office and lab with me and yet so quickly becoming friends.

A special thank you goes to my early scientific influencers and mentors, **Oliver Kim** for your enthusiasm and insights into the fascinating world of molecular biology. **Alex McQuillian** for making chemistry fun, for your support and friendship. My very first PI, idol and mentor **Ortrun Mittelsten-Scheid**, you opened this world to me. Thank you for everything.

The gang, **Matteo**, **Lorenz**, **Ann-Kathrin**, **Bas**, **Sarah**, **Sóley** and **Moritz**, thank you for sharing the first Stockholm years with me, and continuing to be this amazing group of friends that meet once a year and just click. Thank you **Anna** for these many years of friendship that have connected us. Dear **Peter**, without you none of this could have happened. Thank you for the time we had, your enthusiasm and support through the years.

To my family. My idols, support crew, playmates, role models, distraction. **Mama**, **Papa**, **Jakob**, **Anna**, **Teresa**, **Matthäus**, **Markus**, **Sonja**, **Leonie**, **Samuel**, **Benjamin** and **Werner**, thank you for opening the world to me, and welcoming me back home. Thank you for sharing what lays ahead. **Amit**, I could not be more happy, thankful and excited.

REFERENCES

Altman, J. (2011). The Discovery of Adult Mammalian Neurogenesis. In Neurogenesis in the Adult Brain I, (Tokyo: Springer Japan), pp. 3–46.

Altman, J., and Das, G.D. (1965). Autoradiographic and histological evidence of postnatal hippocampal neurogenesis in rats. J. Comp. Neurol. *124*, 319–335.

Arendt, D., Musser, J.M., H Baker, C. V, Bergman, A., Cepko, C., Erwin, D.H., Pavlicev, M., Schlosser, G., Widder, S., Laubichler, M.D., et al. (2016). The origin and evolution of cell types. Nat. Publ. Gr. *17*.

Bargmann, C.I., and Marder, E. (2013). From the connectome to brain function. Nat. Methods *10*, 483–490.

Bendall, S.C., Simonds, E.F., Qiu, P., Amir, E. -a. D., Krutzik, P.O., Finck, R., Bruggner, R. V., Melamed, R., Trejo, A., Ornatsky, O.I., et al. (2011). Single-Cell Mass Cytometry of Differential Immune and Drug Responses Across a Human Hematopoietic Continuum. Science (80-.). *332*, 687–696.

Bendall, S.C., Davis, K.L., Amir, E.-A.D., Tadmor, M.D., Simonds, E.F., Chen, T.J., Shenfeld, D.K., Nolan, G.P., and Pe'er, D. (2014a). Single-cell trajectory detection uncovers progression and regulatory coordination in human B cell development. Cell *157*, 714–725.

Bendall, S.C., Davis, K.L., Amir, E.-A.D., Tadmor, M.D., Simonds, E.F., Chen, T.J., Shenfeld, D.K., Nolan, G.P., and Pe'er, D. (2014b). Single-cell trajectory detection uncovers progression and regulatory coordination in human B cell development. Cell *157*, 714–725.

Bengtsson, M., Hemberg, M., Rorsman, P., and Ståhlberg, A. (2008). Quantification of mRNA in single cells and modelling of RT-qPCR induced noise. BMC Mol. Biol. *9*, 63.

Blow, M., Futreal, P.A., Wooster, R., and Stratton, M.R. (2004). A survey of RNA editing in human brain. Genome Res. *14*, 2379–2387.

Buck, L., and Axel, R. (1991). A novel multigene family may encode odorant receptors: A molecular basis for odor recognition. Cell *65*, 175–187.

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

Cadwell, C.R., Palasantza, A., Jiang, X., Berens, P., Deng, Q., Yilmaz, M., Reimer, J., Shen, S., Bethge, M., Tolias, K.F., et al. (2016). Electrophysiological, transcriptomic and morphologic profiling of single neurons using Patch-seq. Nat. Biotechnol. *34*, 199–203.

Cahoy, J.D., Emery, B., Kaushal, A., Foo, L.C., Zamanian, J.L., Christopherson, K.S., Xing, Y., Lubischer, J.L., Krieg, P.A., Krupenko, S.A., et al. (2008). A Transcriptome Database for

Astrocytes, Neurons, and Oligodendrocytes: A New Resource for Understanding Brain Development and Function. J. Neurosci. *28*.

Cao, J., Packer, J.S., Ramani, V., Cusanovich, D.A., Huynh, C., Daza, R., Qiu, X., Lee, C., Furlan, S.N., Steemers, F.J., et al. (2017). Comprehensive single cell transcriptional profiling of a multicellular organism by combinatorial indexing. bioRxiv 104844.

Chen, K.H., Boettiger, A.N., Moffitt, J.R., Wang, S., and Zhuang, X. (2015). Spatially resolved, highly multiplexed RNA profiling in single cells. Science (80-.). *348*, aaa6090-aaa6090.

Cusanovich, D.A., Daza, R., Adey, A., Pliner, H.A., Christiansen, L., Gunderson, K.L., Steemers, F.J., Trapnell, C., and Shendure, J. (2015). Multiplex single cell profiling of chromatin accessibility by combinatorial cellular indexing. Science *348*, 910–914.

Deng, Q., Ramsköld, D., Reinius, B., and Sandberg, R. (2014). Single-Cell RNA-Seq Reveals Dynamic, Random Monoallelic Gene Expression in Mammalian Cells. Science (80-.). *343*.

Dey, S.S., Kester, L., Spanjaard, B., Bienko, M., and van Oudenaarden, A. (2015). Integrated genome and transcriptome sequencing of the same cell. Nat. Biotechnol. *33*, 285–289.

Eberwine, J., Yeh, H., Miyashiro, K., Cao, Y., Nair, S., Finnell, R., Zettel, M., and Coleman, P. (1992). Analysis of gene expression in single live neurons. Neurobiology *89*, 3010–3014.

Enver, T., Pera, M., Peterson, C., and Andrews, P.W. (2009). Stem Cell States, Fates, and the Rules of Attraction. Cell Stem Cell *4*, 387–397.

Fishell, G., and Heintz, N. (2013). The Neuron Identity Problem: Form Meets Function. Neuron *80*, 602–612.

Fuzik, J., Zeisel, A., Máté, Z., Calvigioni, D., Yanagawa, Y., Szabó, G., Linnarsson, S., and Harkany, T. (2016). Integration of electrophysiological recordings with single-cell RNA-seq data identifies neuronal subtypes. Nat. Biotechnol. *34*, 175–183.

Gierahn, T.M., Wadsworth, M.H., Hughes, T.K., Bryson, B.D., Butler, A., Satija, R., Fortune, S., Love, J.C., and Shalek, A.K. (2017). Seq-Well: portable, low-cost RNA sequencing of single cells at high throughput. Nat. Methods.

Gonçalves, J.T., Schafer, S.T., Gage, F.H., Ahn, S., Joyner, A.L., Aimone, J.B., Wiles, J., Gage, F.H., Aimone, J.B., Wiles, J., et al. (2016). Adult Neurogenesis in the Hippocampus: From Stem Cells to Behavior. Cell *167*, 897–914.

Gravel, C., and Hawkes, R. (1990). Parasagittal organization of the rat cerebellar cortex: Direct comparison of purkinje cell compartments and the organization of the spinocerebellar projection. J. Comp. Neurol. *291*, 79–102.

Grün, D., and van Oudenaarden, A. (2015). Design and Analysis of Single-Cell Sequencing Experiments. Cell *163*, 799–810.

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

Guo, H., Zhu, P., Yan, L., Li, R., Hu, B., Lian, Y., Yan, J., Ren, X., Lin, S., Li, J., et al. (2014). The DNA methylation landscape of human early embryos. Nature *511*, 606–610.

Hashimshony, T., Wagner, F., Sher, N., and Yanai, I. (2012). CEL-Seq: Single-Cell RNA-Seq by Multiplexed Linear Amplification. Cell Rep. *2*, 666–673.

Hashimshony, T., Feder, M., Levin, M., Hall, B.K., and Yanai, I. (2014). Spatiotemporal transcriptomics reveals the evolutionary history of the endoderm germ layer. Nature *519*, 219–222.

Hashimshony, T., Senderovich, N., Avital, G., Klochendler, A., de Leeuw, Y., Anavy, L., Gennert, D., Li, S., Livak, K.J., Rozenblatt-Rosen, O., et al. (2016). CEL-Seq2: sensitive highly-multiplexed single-cell RNA-Seq. Genome Biol. *17*, 77.

Islam, S., Kjällquist, U., Moliner, A., Zajac, P., Fan, J.-B., Lönnerberg, P., and Linnarsson, S. (2011). Characterization of the single-cell transcriptional landscape by highly multiplex RNA-seq. Genome Res. *21*, 1160–1167.

Jaitin, D.A., Kenigsberg, E., Keren-Shaul, H., Elefant, N., Paul, F., Zaretsky, I., Mildner, A., Cohen, N., Jung, S., Tanay, A., et al. (2014). Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types. Science *343*, 776–779.

Jiang, L., Schlesinger, F., Davis, C.A., Zhang, Y., Li, R., Salit, M., Gingeras, T.R., and Oliver, B. (2011). Synthetic spike-in standards for RNA-seq experiments. Genome Res. *21*, 1543–1551.

Kamme, F., Salunga, R., Yu, J., Tran, D.-T., Zhu, J., Luo, L., Bittner, A., Guo, H.-Q., Miller, N., Wan, J., et al. (2003). Single-Cell Microarray Analysis in Hippocampus CA1: Demonstration and Validation of Cellular Heterogeneity. J. Neurosci. *23*.

Kandel, E.R., Schwartz, J.H.H., and Jessell, T.M. (2000). The Development of the Nervous System. In Principles of Neural Science, (McGraw-Hill), pp. 1014–1148.

Kauffman, S.A. (1969). Metabolic stability and epigenesis in randomly constructed genetic nets. J. Theor. Biol. *22*, 437–467.

Ke, R., Mignardi, M., Pacureanu, A., Svedlund, J., Botling, J., Wählby, C., and Nilsson, M. (2013). In situ sequencing for RNA analysis in preserved tissue and cells. Nat. Methods *10*, 857–860.

Klein, A.M., Mazutis, L., Akartuna, I., Tallapragada, N., Veres, A., Li, V., Peshkin, L., Weitz, D.A., and Kirschner, M.W. (2015). Droplet Barcoding for Single-Cell Transcriptomics Applied to Embryonic Stem Cells. Cell *161*, 1187–1201.

Kolodziejczyk, A.A., Kim, J.K., Svensson, V., Marioni, J.C., and Teichmann, S.A. (2015).

The Technology and Biology of Single-Cell RNA Sequencing. Mol. Cell *58*, 610–620.

Kriegstein, A., and Alvarez-Buylla, A. (2009). The Glial Nature of Embryonic and Adult Neural Stem Cells. Annu. Rev. Neurosci. *32*, 149–184.

Kurimoto, K., Yabuta, Y., Ohinata, Y., Ono, Y., Uno, K.D., Yamada, R.G., Ueda, H.R., and Saitou, M. (2006). An improved single-cell cDNA amplification method for efficient high-density oligonucleotide microarray analysis. Nucleic Acids Res. *34*, e42–e42.

La Manno, G., Gyllborg, D., Codeluppi, S., Nishimura, K., Salto, C., Zeisel, A., Borm, L.E., Stott, S.R.W., Toledo, E.M., Villaescusa, J.C., et al. (2016). Molecular Diversity of Midbrain Development in Mouse, Human, and Stem Cells. Cell *167*, 566–580.e19.

Lee, J.H., Daugharthy, E.R., Scheiman, J., Kalhor, R., Yang, J.L., Ferrante, T.C., Terry, R., Jeanty, S.S.F., Li, C., Amamoto, R., et al. (2014). Highly Multiplexed Subcellular RNA Sequencing in Situ. Science (80-.). *343*, 1360–1363.

Levine, J.H., Simonds, E.F., Bendall, S.C., Davis, K.L., Amir, E.D., Tadmor, M.D., Litvin, O., Fienberg, H.G., Jager, A., Zunder, E.R., et al. (2015). Data-Driven Phenotypic Dissection of AML Reveals Progenitor-like Cells that Correlate with Prognosis. Cell *162*, 184–197.

Lubeck, E., Coskun, A.F., Zhiyentayev, T., Ahmad, M., and Cai, L. (2014). Single-cell in situ RNA profiling by sequential hybridization. Nat. Methods *11*, 360–361.

Macosko, E.Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, A.R., Kamitaki, N., Martersteck, E.M., et al. (2015). Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. Cell *161*, 1202–1214.

Malnic, B., Hirono, J., Sato, T., and Buck, L.B. (1999). Combinatorial Receptor Codes for Odors. Cell *96*, 713–723.

Marques, S., Zeisel, A., Codeluppi, S., van Bruggen, D., Mendanha Falcão, A., Xiao, L., Li, H., Häring, M., Hochgerner, H., Romanov, R.A., et al. (2016). Oligodendrocyte heterogeneity in the mouse juvenile and adult central nervous system. Science (80-.). *352*.

Matsunami, H., and Buck, L.B. (1997). A Multigene Family Encoding a Diverse Array of Putative Pheromone Receptors in Mammals. Cell *90*, 775–784.

Nagano, T., Lubling, Y., Stevens, T.J., Schoenfelder, S., Yaffe, E., Dean, W., Laue, E.D., Tanay, A., and Fraser, P. (2013). Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. Nature *502*, 59–64.

Ogiwara, I., Iwasato, T., Miyamoto, H., Iwata, R., Yamagata, T., Mazaki, E., Yanagawa, Y., Tamamaki, N., Hensch, T.K., Itohara, S., et al. (2013). Nav1.1 haploinsufficiency in excitatory neurons ameliorates seizure-associated sudden death in a mouse model of Dravet syndrome. Hum. Mol. Genet. *22*, 4784–4804.

Picelli, S., Björklund, Å.K., Faridani, O.R., Sagasser, S., Winberg, G., and Sandberg, R. (2013). Smart-seq2 for sensitive full-length transcriptome profiling in single cells. Nat.

Methods *10*, 1096–1098.

Pollen, A.A., Nowakowski, T.J., Shuga, J., Wang, X., Leyrat, A.A., Lui, J.H., Li, N., Szpankowski, L., Fowler, B., Chen, P., et al. (2014). Low-coverage single-cell mRNA sequencing reveals cellular heterogeneity and activated signaling pathways in developing cerebral cortex. Nat. Biotechnol. *32*, 1053–1058.

Poulin, J.-F., Tasic, B., Hjerling-Leffler, J., Trimarchi, J.M., and Awatramani, R. (2016). Disentangling neural cell diversity using single-cell transcriptomics. Nat. Neurosci. *19*, 1131–1141.

Ramon y Cajal, S. (1890). Sobre la existencia de células nerviosas especiales en la primera capa de las circunvoluciones cerebrales. Gac. Médica Catalana *13*, 737–739.

Ramon y Cajal, S. (1895). Histology of the Nervous System, Swanson, N., and Swanson, L.W. (trans.).

Ramon y Cajal, S. (1899). Estudios sobre la corteza cerebral humana I: corteza visual. Rev. Trim. Micrográf. Madrid *4*, 117–200.

Ramsköld, D., Luo, S., Wang, Y.-C., Li, R., Deng, Q., Faridani, O.R., Daniels, G.A., Khrebtukova, I., Loring, J.F., Laurent, L.C., et al. (2012). Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. Nat. Biotechnol. *30*, 777–782.

Rosenberg, A.B., Roco, C., Muscat, R.A., Kuchina, A., Mukherjee, S., Chen, W., Peeler, D.J., Yao, Z., Tasic, B., Sellers, D.L., et al. (2017). Scaling single cell transcriptomics through split pool barcoding. bioRxiv 105163.

Rossier, J., Bernard, A., Cabungcal, J.-H., Perrenoud, Q., Savoye, A., Gallopin, T., Hawrylycz, M., Cuénod, M., Do, K., Urban, A., et al. (2015). Cortical fast-spiking parvalbumin interneurons enwrapped in the perineuronal net express the metallopeptidases Adamts8, Adamts15 and Neprilysin. Mol. Psychiatry *20*, 154–161.

Rotem, A., Ram, O., Shoresh, N., Sperling, R.A., Goren, A., Weitz, D.A., and Bernstein, B.E. (2015). Single-cell ChIP-seq reveals cell subpopulations defined by chromatin state. Nat. Biotechnol. *33*, 1165–1172.

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

Setty, M., Tadmor, M.D., Reich-Zeliger, S., Angel, O., Salame, T.M., Kathail, P., Choi, K., Bendall, S., Friedman, N., and Pe'er, D. (2016). Wishbone identifies bifurcating developmental trajectories from single-cell data. Nat. Biotechnol. *34*, 637–645.

Shekhar, K., Lapan, S.W., Whitney, I.E., Tran, N.M., Macosko, E.Z., Kowalczyk, M., Adiconis, X., Levin, J.Z., Nemesh, J., Goldman, M., et al. (2016). Comprehensive Classification of Retinal Bipolar Neurons by Single-Cell Transcriptomics. Cell *166*, 1308–

1323.e30.

Spalding, K.L., Bergmann, O., Alkass, K., Bernard, S., Salehpour, M., Huttner, H.B., Boström, E., Westerlund, I., Vial, C., Buchholz, B.A., et al. (2013). Dynamics of Hippocampal Neurogenesis in Adult Humans. Cell *153*, 1219–1227.

Ståhl, P.L., Salmén, F., Vickovic, S., Lundmark, A., Navarro, J.F., Magnusson, J., Giacomello, S., Asp, M., Westholm, J.O., Huss, M., et al. (2016). Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. Science *353*, 78–82.

Stegle, O., Teichmann, S.A., and Marioni, J.C. (2015). Computational and analytical challenges in single-cell transcriptomics. Nat. Rev. Genet. *16*, 133–145.

Subkhankulova, T., Yano, K., Robinson, H.P.C., and Livesey, F.J. (2010). Grouping and classifying electrophysiologically-defined classes of neocortical neurons by single cell, whole-genome expression profiling. Front. Mol. Neurosci. *3*, 10.

Sugino, K., Hempel, C.M., Miller, M.N., Hattox, A.M., Shapiro, P., Wu, C., Huang, Z.J., and Nelson, S.B. (2006). Molecular taxonomy of major neuronal classes in the adult mouse forebrain. Nat. Neurosci. *9*, 99–107.

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

Tasic, B., Menon, V., Nguyen, T.N., Kim, T.K., Jarsky, T., Yao, Z., Levi, B., Gray, L.T., Sorensen, S.A., Dolbeare, T., et al. (2016). Adult mouse cortical cell taxonomy revealed by single cell transcriptomics. Nat. Neurosci. *19*, 335–346.

Tietjen, I., Rihel, J.M., Cao, Y., Koentges, G., Zakhary, L., and Dulac, C. (2003). Single-Cell Transcriptional Analysis of Neuronal Progenitors. Neuron *38*, 161–175.

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

Treutlein, B., Brownfield, D.G., Wu, A.R., Neff, N.F., Mantalas, G.L., Espinoza, F.H., Desai, T.J., Krasnow, M.A., and Quake, S.R. (2014). Reconstructing lineage hierarchies of the distal lung epithelium using single-cell RNA-seq. Nature *509*, 371–375.

Ubink, R., Calza, L., and Hökfelt, T. (2003). “Neuro”-peptides in glia: focus on NPY and galanin. Trends Neurosci. *26*, 604–609.

Usoskin, D., Furlan, A., Islam, S., Abdo, H., Lönnerberg, P., Lou, D., Hjerling-Leffler, J., Haeggström, J., Kharchenko, O., Kharchenko, P. V, et al. (2014). Unbiased classification of sensory neuron types by large-scale single-cell RNA sequencing. Nat. Neurosci. *18*, 145–153.

Waddington, C. (1939). An introduction to modern genetics (London: George Allen &

Unwin).

Waddington, C. (1940). Organisers and Genes (Cambridge, UK: Cambridge University Press).

Waddington, C. (1957). The Strategy of the Genes (London: George Allen & Unwin).

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

Warren, L., Bryder, D., Weissman, I.L., and Quake, S.R. (2006). Transcription factor profiling in individual hematopoietic progenitors by digital RT-PCR. Proc. Natl. Acad. Sci. U. S. A. *103*, 17807–17812.

Zechel, S., Huber-Wittmer, K., von Bohlen und Halbach, O., Ibáñez, C.F., Linnarsson, S., Heuckeroth, R., Johnson, E., Milbrandt, J., Li, C., Amamoto, R., et al. (2006). Distribution of the iron-regulating protein hepcidin in the murine central nervous system. J. Neurosci. Res. *84*, 790–800.

Zeisel, A., Muñoz-Manchado, A.B., Codeluppi, S., Lönnerberg, P., La Manno, G., Juréus, A., Marques, S., Munguba, H., He, L., Betsholtz, C., et al. (2015). Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. Science (80-.). *347*.

Zhang, Y., Chen, K., Sloan, S.A., Bennett, M.L., Scholze, A.R., O'Keeffe, S., Phatnani, H.P., Guarnieri, P., Caneda, C., Ruderisch, N., et al. (2014). An RNA-sequencing transcriptome and splicing database of glia, neurons, and vascular cells of the cerebral cortex. J. Neurosci. *34*, 11929–11947.

Zheng, G.X.Y., Terry, J.M., Belgrader, P., Ryvkin, P., Bent, Z.W., Wilson, R., Ziraldo, S.B., Wheeler, T.D., McDermott, G.P., Zhu, J., et al. (2016). Massively parallel digital transcriptional profiling of single cells. bioRxiv.