EXPLORING AND DEFINING THE CELLULAR DIVERSITY OF THE BRAIN ACROSS SPECIES

By

Casey A. Thornton

A DISSERTATION

Presented to the Department of Molecular and Medical Genetics

and the Oregon Health and Science University

School of Medicine

in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

May 2022

School of Medicine

Oregon Health and Science University

CERTIFICATE OF APPROVAL

This is to certify that the Ph.D. dissertation of

Casey A. Thornton

has been approved

Dr. Andrew Adey, Mentor/Advisor

Dr. Kevin Wright, Chair

Dr. Anthony Barnes, Member

Dr. Lucia Carbone, Member

Dr. Stephen Moore, Member

Table of Contents

List of Figures

Acknowledgments

This thesis is the culmination of the love, support, and resources that have been poured into my professional development. Firstly, I thank my family who I share this accomplishment with. The path toward higher education has never been clear – presenting financial barriers that have always seemed insurmountable. My doctorate represents my family's selflessness and love, so I offer this thesis as our shared accomplishment.

Secondly, thank you to my committee. Thank you to Dr. Kevin Wright for seeing me through my qualifying exam to my defense. You have always been such a strong supporter of my professional development and work, and I aspire to approach science with as much excitement as I have seen you do.

Thank you to Dr. Anthony Barnes for always treating me like a peer, engaging honestly with my work, and asking compelling questions that made me excited to revisit my data. I have been so lucky to have you as a mentor, who has gone from teaching me at the bench to taking the time to meet with me to discuss "moonshot" projects.

Thank you to Dr. Lucia Carbone, who I have looked up to since I first interviewed as a prospective graduate student. Lucia nominated me for the OHSU Early Independence Fellowship and worked with me to develop ideas about thesis research before I ever even joined OHSU. Your support inspired me to come to OHSU, provided my first four years of funding, and capitulated my career – for which I am so grateful.

To Dr. Stephen Moore, I give special thanks because he has taught me so much about what it means to be a mentor and an advocate. First, in journal clubs, I learned that creating an environment of mutual respect creates a rich environment for discussion. When I came to him with my aspiration to become a clinical geneticist, he supported me

through one-on-one tutoring over the past year and championed me through my application process. I cannot overstate the positive impact you have made on my time as a graduate student and my career trajectory.

To my mentor, Dr. Andew Adey, I am truly honored to have been your student. For the first time in my career in science, I was treated as an expert and given the agency to pursue my scientific interests. While I often felt underqualified or even wholly unqualified to take on the challenges I was presented with, you trusted my decision-making. I have developed a newfound professional confidence and belief in my ability to succeed in my professional pursuits because of your support.

To those who have made the day-to-day task of being a graduate student possible – thank you for teaching me so much about love and support. You have all made my time in Portland an incredible experience and are my chosen family. Specifically, to my partner in life and science, Ryan, it's not easy to work alongside such an accomplished scientist. Instead of casting me in your shadow, you have elevated my work, encouraged me, and provided constant inspiration to learn and grow inside and outside of science. I cannot thank you enough for how much I have learned and grown with you.

To everyone who has made this accomplishment possible, know that my successes are our successes. I am so excited for what we will be able to achieve.

Abstract

Single-cell technologies have created a new taxonomic challenge: to create a shared language for defining individual cells within complex organisms. In this work, I complicate the taxonomic task by developing a novel assay that captures spatially resolved single-cell chromatin accessibility and also explore and define epigenomic cell states in the mouse and human brains.

First, I present a novel assay: Single-cell Combinatorial Indexing on Microbiopsies Assigned to Positions for the Assay of Transposase Accessible Chromatin, or sciMAP-ATAC, which captures spatially-resolved single-cell chromatin accessibility data from intact tissue. I applied sciMAP-ATAC to resolve the layers of the mouse and human cortices and demonstrate the ability of sciMAP-ATAC to generate a rich dataset that captures layer-specific single-cell epigenomic features. I then applied sciMAP-ATAC to define the gradient of cellular state shifts in response to cerebral ischemia within the stroke infarction area and the surrounding penumbra. I show that significant epigenetic remodeling occurs in the brain tissue surrounding infarction.

Secondly, I present the Human Glial Atlas (HGA), which includes a comprehensive and publicly available dataset of 378,270 glial cells derived from the healthy adult human brain's hippocampus and cortex regions. I provide this resource as a baseline for the unperturbed, healthy state of the human brain and provide characterizations of rare cell states. Notably, I characterize cell states that are commonly defined in disease. I demonstrate a gradient of activation of glial reactivity in the absence of disease-associated inflammation, thus revealing that persistent yet rare populations of reactive glial cell types exist in the unperturbed cortex and hippocampus.

Together, these works provide a new single-cell omic method, sciMAP-ATAC, that collects an essential data type – the spatial orientation of cells, and provide a rigorous definition of cell states in the human brain in an attempt to fill in the branches and twigs of the emerging single-cell taxonomy of the human brain.

Chapter 1: Introductions

1.1 Overview and rationale

Our understanding of the natural world begins with exploring and defining our observations. While naturalists have observed and described biological phenomena throughout human history, sequencing technologies that emerged in 1977[1] allowed us to discern the natural world through the new lens of genomics. In the next frontier, Next Generation Sequencing (NGS) technologies emerged in 2000[2] and enabled omics studies with both broader and finer resolution investigation of the genome. Broadly, NGS technologies allow us to resolve entire genomes and address inter-organismal omic variation. Finely, NGS technologies have allowed us to assess cellular intra-organismal omic variation. While *inter-organismal* studies consider genomic differences between organisms, *intra-organismal* studies focus on the genomically uniform organism and define the epigenomic variation that gives rise to the diverse cell types and cell functions that form the organ systems within a single multicellular organism. More specifically, we can characterize how single cells behave in both health and disease. Here, I explored and defined the intra-organismal epigenomic variation of cells in the human brain and model organisms – focusing on clinically translational interpretations for the improvement of human medicine. To achieve this, I used and developed novel single-cell technologies. I expanded on both the arsenal of NGS-based methods for assessing the epigenomic features of single cells and created an atlas of non-neuronal cell types to fill in the branches and twigs of non-neuronal cellular biology in the taxonomic tree of cellular heterogeneity in the human brain.

1.2 The evolution of taxonomy in the natural sciences

We understand the natural world by defining its component parts and by using these shared definitions and nomenclatures to begin to understand complex observations. This field of biological systematics focuses on the characterization of living things and the curation of nomenclature to describe them – and encompasses many similar and overlapping disciplines in the biological and philosophical sciences. Biological systematics includes the identification, taxonomy, and nomenclature of organisms and their relationship to each other. The delineation of taxonomy from biological systematics is granular and debated[3]. Taxonomy, specifically, focuses on recognizing and delimiting organisms into hierarchical categories through observation, as in the Linnaean approach, or through genetic variation, as in phylogeny. When we parse the similarities and differences between overlapping fields of study, we ask ourselves: Why is it essential to devote entire disciplines to definition? Simply, it is through language, and therefore nomenclature, that we describe the natural world. Without a system for description, scientists would not be able to use shared verbiage to communicate their observations.

Early philosophers made the first contributions to biological systematics. Aristotle bifurcated living things into '*Anima*,' animals with no red blood, and '*Enaima*,' animals with red blood, in one of the earliest (384–322 BC) attempts to create a shared language for partitioning living organisms. This classification persists today, pertaining to vertebrate and invertebrate organisms. Early modern Europe, which emerged from the medieval period, adopted Latin as the shared language for scientific naming. However, no defined system for scientific Latin existed. Carl Linnaeus addressed this need and created the modern taxonomic system in *Systema Naturae* (1735). This Linnaean taxonomy designated groups in a binomial naming system of: genus and species.

Incredibly, as initially defined by Linnaeus in 1735 (Fig. 1), the tree of life persists as the naming schema and overarching hierarchical taxonomy that taxonomists use today.

Linnaean taxonomy, however, relies on assumptions that were reasonable for the time but are no longer accurate based on our current understanding of the evolution of species. Notably, Linnaean taxonomy relies on an assumption of creationism – stipulating that the total number of species is static and finite, and also assumes that no natural order exists between species. Therefore, Linnaean taxonomy does not account for the longitudinal evolution of species over time. Instead, Linnaean taxonomy defines morphological characteristics as the critical delimiter between species. Morphology, however, is not a reliable data type for discerning species because organisms can: (i) occupy unique morphological states throughout their life cycle and (ii) also occupy cryptic states. For instance, at the beginning of the life

Figure 1. Haeckel tree of life. Constructed from the Linnaean system of taxonomy. Illustration of the tree of life, grossly divided into Plantae (left), Protista (middle), and Animalia (right)[4].

cycle, the butterfly morphologically resembles a caterpillar, while at the end of its life cycle, the butterfly is dramatically morphologically altered and winged. Additionally, species can occupy cryptic states where two or more genetically unrelated species are morphologically indistinguishable due to convergent evolution. In his time, Linnaeus

characterized and named 5,900 plants and 4,378 animals according to his system of taxonomy[5]. From this legacy, modern taxonomists have named nearly 1.5 million of the estimated 5 ± 3 million species on Earth[6] and are increasingly finding that modern genetic technologies are required to correct the antiquated assumptions of Linnaean taxonomy[7].

Sequencing technologies have revolutionized the field of biological systematics and challenged the core assumptions of Linnaean taxonomy. Darwin published his theory of evolution in 1859[8], which both challenged and supported Linnaean taxonomy. For example, Linnaeus originally introduced the order of *Anthropomorpha*, which included the closely evolutionarily related humans and apes without prior knowledge of their shared genetic ancestry. Darwin hypothesized a genetically defined tree of life (Fig. 2a) that reflected the phylogenetic system of classifying evolutionary relationships between species that reflects our modern knowledge of genomic relationships between species (Fig. 2b).
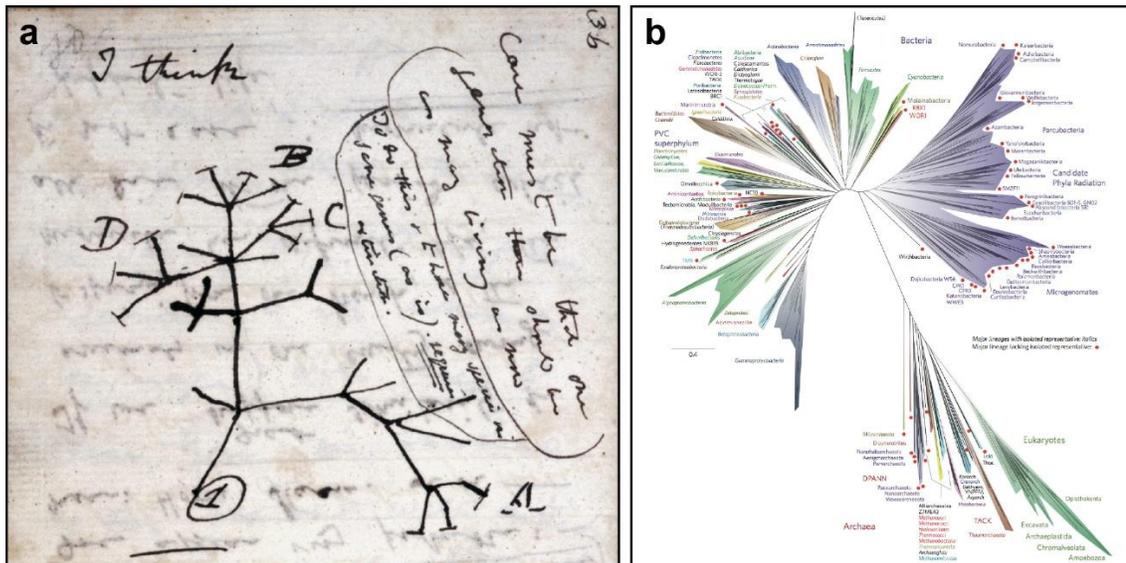
Figure 2. Evolution of the tree of life. a, Charles Darwin's hypothesized evolutionarily divided tree of life[8]. **b**, The modern phylogenetic tree of life subdivides clades based on species' evolutionary relationship[9]. Evolutionary distance also represents tree branch length.

The advent of sequencing technologies, and a new methodology for genetically defining species, called DNA barcoding, has allowed for a reinvention of the taxonomy of species, from the morphologically-based Linnaean system to the phylogenetic-based cladistic system. This inclusion of data derived from sequencing technologies has bifurcated the taxonomy field into *alpha-taxonomy*, the morphological and Linnaean delineation of species, and *beta-taxonomy*, the approach to taxonomy that incorporates data types other than morphology. DNA barcoding delineates species by sequencing a single gene that is commonly variable between species but is not variable between organisms within a species – and is the predominant beta-taxonomy approach to defining the taxon of organisms. This evolution-based genetic hierarchical taxonomy of biological types gave rise to the *taxonomic impediment*, which is broadly defined as the difficulty presented to taxonomists to both: (i) collect and interpret the enormous amount of sequencing data needed to reevaluate the taxonomic hierarchy and (ii) harmonize the genetic and morphological taxonomy[10,11].

While the field of biological systematics reels from the advent of new technologies and data types, the natural world remains unchanged. Therefore, despite the philosophical arguments between alpha and beta taxonomies – the data types, such as the DNA sequences that we are only recently able to access, have always been present in the organisms that biological systematics has sought to define. Therefore, I propose that taxonomy should not focus on the inclusion or exclusion of data types but instead on a complete mapping of heritable genetic material within the domains of biological life.

1.2.1 The epigenomic taxonomy of single cells

The discovery of a wealth of previously unexplored biological types in the Linnaean era echoes the age of single-cell omics that we are experiencing today. Linnaean taxonomy is considered one of the earliest and most successful bioinformatic crises[12]. The

increase in globalization in the mid-1700s resulted in an influx of newly identified species to Europe from the American, Asian, and African continents. And with these discoveries came new names for these species, which met no criteria other than being Latin in origin. For instance, in naming a briar rose, naturalists alternately named it "*Rosa sylvestris inodora seu canina*" and "*Rosa sylvestris alba cum rubore, folio glabro*," where the length of the name scaled with the heuristically chosen adjectives used to describe it. As naturalists faced a wealth of newly identified biological organisms, they shared no standard nomenclature, method of characterization, or hierarchy within which to place their discoveries. Historians have lauded Linnaeus for his, perhaps oversimplistic, approach to creating a taxonomic system. Linnaeus chose specific characteristics, i.e., reproductive organ structure, to create a hierarchy and naming system for all organisms – with the primary objective of creating an ontology, or shared understanding, that allowed for meaningful scientific discourse. With this first informatic challenge in mind, we now must consider the current taxonomic challenge of single cells.

A comprehensive system of single-cell taxonomical hierarchy has not been defined though there are significant efforts underway. Notably, the field of single-cell omics remains in the discovery phase, much in the same way that the identification of species blossomed before the advent of Linnaean taxonomy. Single-cell taxonomy requires an intra-organismal approach based on the assumption that organisms within a species are genetically homogenous. Defining taxons of single cells from the intra-organismal view requires reconfiguring the Linnaean schema. Linnaean taxonomy assumes that an organism is genetically homogenous and that organisms are arranged into genetically similar "types" or species. Similarly, cellular taxonomy can be viewed as a subset of Linnaean taxonomy, where the same assumptions are made about the genetic homogeneity within an organism and within a species. However, single-cell taxonomy

progresses beyond Linnaean taxonomy by characterizing the parts of organisms, or cells, based on features of the genome other than the DNA sequence or epigenomics. Therefore, in contrast to the Linnaean taxonomic hierarchy of species: domain, kingdom, phylum, class, order, family, genus, and species, the single-cell epigenomic taxonomic hierarchy can extend beyond the species classification to: organs, tissues, tissues sub-structures, and cells. Single-cell taxonomy, therefore, must rely on defining the epigenomic states of cells within organisms. There are difficulties, however, in defining a comprehensive single-cell taxonomy.

Genetic homogeneity must be established to develop a single-cell taxonomic tree. In efforts that have been undertaken to genomically characterize cells in humans, the species of homo sapiens has been assumed to be approximately homogeneous. This assumption of approximate genetic homogeneity does stand in contrast to many efforts to characterize the impact of genetic heterogeneity in the human population. In efforts to epigenomically characterize cells in mice, genetically homogenous mice strains are used to control for genomic variation. These differences in approaches between mouse and human single-cell studies highlight the field's infancy and the challenges of genetic research in humans. Human inter-individual omic variability is substantial and is a core focus of the Human Cell Atlas initiative and this work[13,14] (see Chapter 3). One consideration for formulating a comprehensive single-cell systematics paradigm is the presence of omic inter-individual variation.

Secondly, a single-cell epigenomic taxonomy must account for the mutability of cell states. For example, cells differentiate in the developing organisms, thus potentially creating a string of multiple taxonomies along the time axis. Similarly, diseases affect organs, tissues, and cells, resulting in a potential series of single-cell taxonomies along the axis of pathogenicity. However, these cell state shifts do not make single-cell

epigenomic taxonomy moot. Classical Linnaean taxonomy faces similar challenges. Linnaean taxonomy assumes a genetically static organism with a framework of categorically separated species. These assumptions were founded on the creationist view of a pre-ordained finite number of species but do not reflect our modern understanding of genetic diversity. In contrast, organisms are not genetically homogeneous and static.

Organisms are genetically mosaic[15] and accumulate genetic mutations with age[16] that can have gross effects, including cancer[17,18]. However, taxonomists today support the Linnaean system because the assumption of genetic homogeneity is an acceptable imperfection that allows for a comprehensive system for defining all organisms. Similarly, while epigenomic states of cells are variable, assumptions can be made to create a shared and interpretable taxonomic system for single cells. Here I propose that a common single-cell epigenomic taxonomy is both possible and necessary for the future of sub-organismal omic research.

1.3 The epigenome of single cells

To create a taxonomy of single cells, we must first understand the epigenome of cells and the importance of assaying single cells. Understanding epigenetics comes most easily to me when I consider that all cells in our body are approximately genetically identical, meaning that the DNA sequence in the cells of the human body is the same across all of the organ systems and tissues. Additionally, we know from the fundamental experimentation of early geneticists, such as Mendel, that the sequence of DNA that all cells share serves as the blueprint for developing an entire organism. When we consider these two foundational features of genetics, (i) that all cells within an organism share the same DNA, and (ii) this uniform DNA sequence gives rise to many disparate functional units of an organism, then it is clear that there is a biological mechanism that leads to

differential utilization of the genome between functionally distinct cell types that compromise tissues. C.H. Waddington addressed this exact biological mechanism in 1957 in his pioneering publication, "*The Strategy of the Genes*," where he proposed that cells, like marbles on a landscape, roll along the peaks and valleys toward terminally differentiated states (Fig. 3a)[19].

If we consider our first assumption that cells begin development with the same genome – then we can think of cells sitting on an entirely smooth and flat *genomic landscape* that represents their shared DNA sequence. These cells would remain stationary and homogenous. Given that cells are not homogenous, Waddington proposed that the landscape must have an irregular topology with peaks and valleys, which cause cells to probabilistically select a pathway toward differentiated states. He described the underpinning of the landscape as pegs attached to a base, representing genes, and "guy-lines," shown as lines, which represent the epigenetic "forces" exerted on genes that manipulate the genomic landscape to create the *epigenomic landscape* (Fig. 3b). These epigenomic forces on the genomes of cells give rise to the heterogeneous and complex cellular populations that comprise organisms.
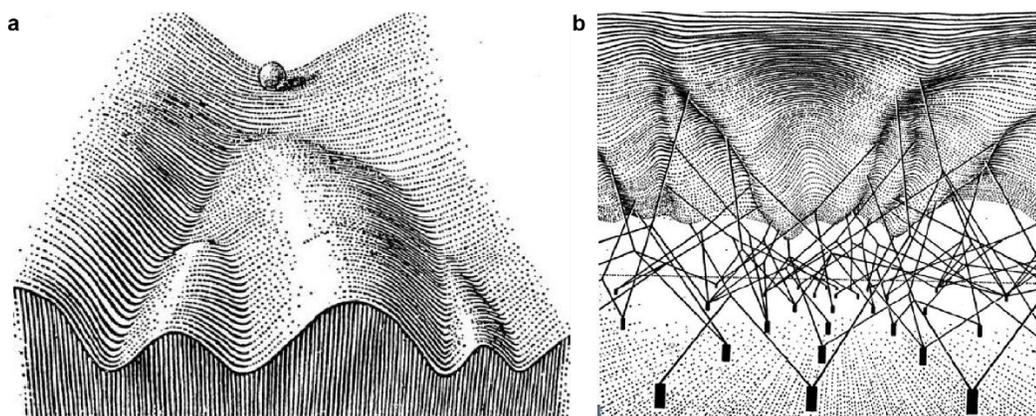
Figure 3. Waddington epigenetic landscape. a, Topology of the epigenetic landscape that cells (shown as a marble) traverse pathways from undifferentiated to differentiated cell types. **b**, Representation of the many epigenetic features (lines) which influence gene (peg) expression and define the epigenetic landscape.

Remarkably, Waddington's hypothesis remains relevant and serves as the basis for the field of epigenetics. We now know that many genome modifications regulate the fate of cells and lead to their differentiation into functional and morphological states.

1.3.1 Cell type variation

From the intra-organism perspective of taxonomy, the hierarchy of functional specification begins with organs, progresses downward to tissues, then to individual cells. Epigenetic assessment of single cells has revealed an astounding diversity between and within tissues. Many research efforts currently focus on the epigenomic characterization of cells in distinct tissues. These studies demonstrate that cells vary between tissues and confer different functions to organs in the body.

Developing a comprehensive map of the cell types that comprise the different tissues of the healthy human body is of prime interest in the research and medical communities. A well-defined inter-tissue cell atlas would allow for a comprehensive understanding of the "normal" human body and provide a baseline for knowing which cells are affected in disease states. As a result, many efforts are underway to define this "baseline."

One such effort, the Human Cell Atlas, aims to generate a comprehensive reference map of all human cells by generating organ-specific datasets[13]. Single-cell studies of the epigenome have revealed that the human body contains many distinct cell types, which can be singular or shared between tissues[20]. Single-cell epigenomics is reaching a consensus on the complete range of cell types present within each tissue (e.g., 1.5.1 Glial cell types of the central nervous system). While assessing the epigenomic variability of cells in humans is of prime interest for medicine, many single-cell studies have focused on mice as a model organism. Entire atlas level datasets of individual mice have been developed, which has not been possible in human subjects[21]. These studies provide a reference for the depth and breadth of cellular epigenomic cell types.

1.3.2 Cell state variation in health and disease

C.H. Waddington described a continuous epigenetic landscape (Fig. 3a). As previously described, cells traverse this landscape toward differentiated cell types. However, the continuity of this landscape represents all of the possible states that cells can exist in, within, and between defined cell types. Single-cell technologies have allowed for the unbiased sampling of cells from tissues and have led to the discovery and characterization of (i) defined cell types and (ii) intra-cell-type and disease-associated cell states. The distinction between cell types and states remains fluid within single-cell omics[22]. Cell types are commonly defined as the terminally differentiated states of cells, whereas cell states are defined as both the variation within cell types and epigenetically distinct cell populations that arise from a perturbation such as disease.

1.3.3 Spatial variation

Single-cell technologies have grown to include assessments of the genome, epigenome, and intra- and extra-cellular moieties, which has allowed for precise quantification of the heterogeneity of complex tissues. As technologies for assessing cellular features improve, studies capture increasingly high numbers of cells in single experiments with increasingly high informational content per cell[24–27]. These approaches, however, rely on aggregate dissociation of tissues, which results in loss of the positions of cells within the intact tissue. However, there is substantial evidence that cells can encompass a range of omic states
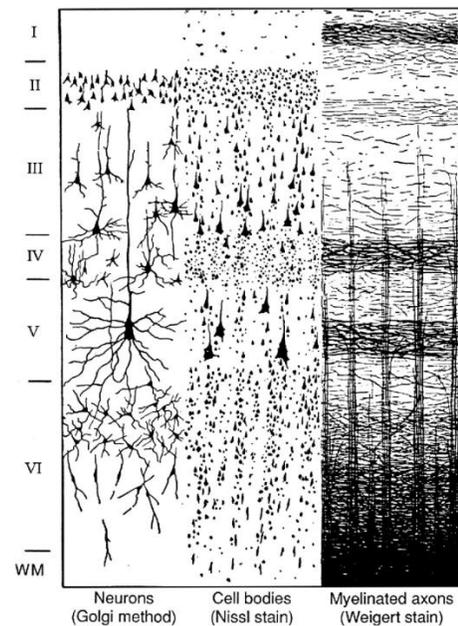
Figure 4. The structure of the cortex. Neurons (left), cell bodies (middle), myelinated axons (right)[23].

11

in response to their spatial orientation within a tissue, specifically in spatially-progressive phenotypes. As in histological assessment of oncological samples, spatial assessment of cells in the tumor microenvironment is necessary to define the spatial limits of the cancerous cells, or the margins. Additionally, the healthy tissues of the human body are comprised of spatially organized organs. For example, the brain, which represents one continuous organ, is compartmentalized into distinct functional regions. In the case of the cortex, cells are spatially stratified across six layers (Fig. 4), and this cortical layering is essential for neurotypical development[28–30]. Many spatially-resolved gene expression and transcriptomic assays have been developed to address this next challenge in resolving complex tissues – and there is an interest in developing spatially-resolved single-cell assays for many epigenomic features (see Chapter 2)[31,32]. While grouping cells by their shared epigenomic profiles reveals cellular similarities, spatially-resolved methods add a valuable data type that allows for the assessment of cell-cell interactions that allows for a comprehensive understating of the complete picture (Fig. 5).
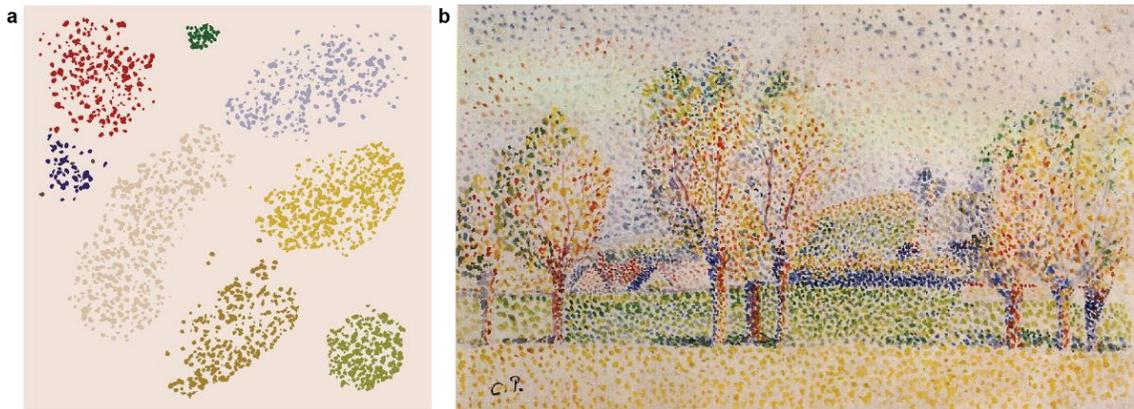
Figure 5. The addition of spatial orientation to single-cell characterization. a, Representative UMAP of single-cell, shown as points, oriented in two-dimensions based on similarity. **b**, Cells, shown as points, as in a, shown in the intact painting, revealing information displayed when cell position is known. Eragny Landscape by Camille Pissarro (1886)[33].

1.4 Chromatin accessibility as a feature of the epigenome

Epigenetics encompasses all modifications to and the use of the genome. The scope of epigenetics, which comprise the "epigenome," is best defined by the term's etymology. "Epi," in Greek, means "in addition to" or "on top of." C.H. Waddington coined the term "epigenetics" to define all features regulating cell fates other than the genome. As shown in Fig. 3b, the epigenome encompasses the guy-lines and anchors that warp the genomic landscape. In this work, I focus on the epigenetic feature of chromatin accessibility.

Each cell contains an entire genome that is organized within the nucleus. The high-level organization of the genome has been observable since 1842 when botanist Karl Nägeli identified distinct chromosomes in pollen[34]. Today, the gross observation of the organization of the human genome in chromosomes by cytogenetic assessment is a standard of human clinical care. In this approach, dark and light bands represent "closed" heterochromatin and "open" euchromatin, depending on the staining modality. While cytogenetics has been essential to human medicine, chromatin compaction at finer resolution than banding is required to identify the DNA sequences accessible to the proteins that bind DNA. The data type that provides base pair-level resolution of DNA, which is accessible and not sterically blocked by histone proteins, is chromatin accessibility. Chromatin accessibility of single cells corresponds to cell types[35–37], cell states[38,39] (see Chapter 3), and gene expression – where accessible genes are preferentially expressed, in contrast to inaccessible chromatin. Numerous chromatin accessibility assays capture accessible DNA in bulk (multi-cell) or single-cell formats. Given that individual cells demonstrate unique chromatin accessibility profiles across their entire genome, single-cell methods are required to capture the finest unit of chromatin variation. In contrast, bulk methods capture the cumulative DNA accessibility

across all cells. Therefore, the assessment of chromatin accessibility in bulk or single-cell contexts should be determined based on the specific scientific questions being asked, as no one method is appropriate for all scientific questions.

1.4.1 Single-cell ATAC-seq as a solution to interpreting cellular genomic usage

Methods for the assessment of chromatin accessibility utilize a variety of biochemical approaches to capture genomic fragments which are unoccupied by DNA binding proteins. Historically, the DNAse hypersensitivity assay used the DNAse I endonuclease enzyme to digest unbound DNA[40]. An improved methodology for capturing chromatin accessibility utilized a transposase enzyme, Tn5, which was first discovered in a study of kanamycin resistance in 1975[41]. Evolutionarily, transposases catalyze transposition, a process by which transposable elements are integrated into the genome through double-stranded cleavage of genomic DNA and ligation-mediated insertion of transposed elements[42]. Recent advancements in biochemical library preparation for NGS motivated the use of Tn5 for catalytic ligation of the addition of universal sequencing adapters to DNA sequencing libraries[43]. Additional co-option of the inability of Tn5 to catalytically cleave protein-bound DNA resulted in the use of Tn5 to capture chromatin accessibility data in an approach termed the Assay for Transposase Accessible Chromatin using sequencing (ATAC-seq)[44]. ATAC-seq utilizes the tagmentation reaction, where Tn5, nuclei, and sequencing adapters are combined to result in a purified sequencing library pool comprised of excised open chromatin fragments ligated to sequencing adapters. Many biochemical variations of the ATAC-seq methodology have been developed to facilitate single-cell assessment. Approaches vary from single-cell single-well tagmentation reactions to microfluidic isolation of single cells within reaction micelles to combinatorial indexing approaches to capturing single-cell ATAC-seq data.

In this work, I utilized the single-cell combinatorial indexed assay for chromatin accessibility using sequencing (sci-ATAC-seq)[45]. This approach relies on a probabilistic sampling approach to create a high likelihood that open, or accessible, chromatin sequencing reads can be assigned to one nucleus (Fig. 6). Isolated nuclei are randomly distributed across wells, and in the work presented here, 96-well plates were used for all sci-ATAC-seq experiments.

Uniquely indexed Tn5 tagmentation reactions are then performed in each well, such that the resulting excised accessible chromatin fragments contained in intact nuclei are uniquely indexed – as denoted by the uniquely colored well in Fig. 6 (see Transpose, Index 1 Incorporation). All tagmented nuclei are then pooled and sparsely Fluorescence-Activated Nucleus Sorted (FANS) into new

Figure 6. Single-cell combinatorial indexing assay for transposase accessible chromatin. Circles represent single nuclei. Dashed boxes represent individual reaction wells. Colors indicate unique indexes which are ligated to excised genomic open chromatin fragments. Note that Index 2 PCR is scaled by the number of Index 1 tagmentation well (15 * 96-well plates = n cells per well in Index 2 PCR).

wells (e.g., 96-well plates). The number of nuclei sorted into each well is metered such that the likelihood of any two same-indexed Index 1 cells, or same-colored cells as in Fig. 6, being sorted into the same Index 2 well is highly unlikely. We find in practice that sci-ATAC-seq barnyard experiments between mouse and human cells result in a 5% collision rate, or a 5% incidence of same-indexed cells being sorted into a single well. Nuclei sorted into wells in this second tier of indexing contain excised chromatin

fragments which are then polymerase chain reaction (PCR) amplified to duplicate the unique and genomic open chromatin regions (see Fig. 6, PCR, Index 2 Incorporation). Post-PCR, the combinatorially indexed accessible chromatin fragments are pooled, cleaned, and sequenced. This approach to single-cell ATAC-seq is highly scalable, low-cost, and presents unique opportunities to multiplex different samples based on transposase index.

1.4.2 Technical considerations of single-cell ATAC-seq experiments

Single-cell ATAC-seq, and all single-cell assay experiments, should be carefully designed and analyzed to ensure that technical variables do not influence biological conclusions derived from the data.

Firstly, bioinformatics data analysis relies on the minimization of batch effect. As single-cell assays have grown in popularity and funding, so have the number of single-cell experiments and data caches. However, data integration poses a substantial barrier to analyzing these data. The batch effect, or the non-biological factors in an experiment that can cause changes in the data produced by the experiment, can be introduced in many stages of single-cell data acquisition and analysis. Batch-effect can be introduced into the processing of multiple samples through any experimental or computational segregation of samples by *any* technical feature of the data, such as individual (e.g., patient ID, or mouse ID), biological condition/control, timepoint, sex, or age. Therefore, experiments must be carefully designed to generate a dataset that can be interpreted independently of these technical variables. Sample acquisition can, however, pose significant barriers to batch effect reduction. For example, longitudinal studies that require sequential sampling and tissue processing result in datasets where the experimental batch corresponds to the sample time point. This batch effect can be mitigated by collecting all samples across all time points and then including a mixture of

all time points in each experimental preparation. In cases where samples must be immediately processed, as in many RNA sequencing experiments, the technical batch effect is an unavoidable artifact of the dataset. In this era of single-cell analysis and meta-analysis, technical batch effects affect the conclusions derived from studies – however, the technical batch effect is not always addressed in manuscripts. As a result, it is essential, as consumers of single-cell studies, to independently assess the quality of the biological conclusions presented by single-cell studies.

Secondly, single-cell studies routinely rely on dimensionality reduction algorithms to create limited dimensional representations of the data – and the interpretability of those figures is highly debated. Again, as consumers of these studies, it falls on the reader to understand the use of these algorithms and determine whether the conclusions presented in a single-cell study are based on multi-dimensional data or two-dimensional data representation. One such dimensionality algorithm, Uniform Manifold Approximation Projection (UMAP), projects a manifold, or topology, through the multidimensional data space[46], i.e., the single-cell by feature space, in the case of single-cell studies. In single-cell publications, UMAP is commonly used to generate two- or three- dimensions for each cell such that the local space between similar cells is minimized and the global space between dissimilar cells is maximized. Plotting the cells by UMAP dimensions results in individual cells plotted in a two-dimensional or three-dimensional representation of the feature space. Importantly, all such dimensionality reduction algorithms rely on parameter variables to manipulate the minimization and maximization of cells' global and local separation within the output dimensions. Therefore, limited (e.g., two- or three-) dimensional representations of high-dimensionality data should be considered as one visualization of the relationship between cells of an infinite number of possible visualizations.

Recent investigations and opinions have raised concerns about the over-interpretation of single-cell limited-dimensionality representations[47]. However, it is essential to remember that dimensionality reductions such as UMAP are tools used to process data, and it is an academic understanding of that tool that must be maintained to interpret the output from it correctly. Therefore, it is the responsibility of those reporting single-cell results and those interpreting single-cell results to understand the strengths and limitations of the biological tools being used.

1.5 Toward a cellular ontology of the mammalian brain

The Gene Ontology (GO) initiative has developed a consistent nomenclature for describing features associated with known genes to facilitate genomic research. Since the advent of single-cell technologies, we are approaching a complete set of known cell types in mammalian brains – particularly in the human and mouse brains. Therefore, we are approaching a consistent representation of cellular kinds in the mammalian brain, or a complete cellular ontology.

The focus of this work, the central nervous system (CNS), comprises a diverse range of neuronal and non-neuronal cell types that give rise to cognition. Characterization of neuronal cell types has elucidated sub-classifications of neurons that execute disparate functions throughout the human brain[48]. In contrast, single-cell characterization of non-neuronal cells, or glia, in the mammalian brain has lagged. As a result, documentation of the regulatory elements that give rise to the multifaceted functions of glial cell types is required to create a complete map of cellular heterogeneity in the mammalian brain.

1.5.1 Glial cell types of the central nervous system

Previous work has advanced our understanding of the vital role of non-neuronal cells in the central nervous system, and single-cell studies have begun to converge on the

abundant glial cell types in the mouse and human brains. In the absence of a shared taxonomy and nomenclature for defining cell types and states, there are many names for cell types and many features used to determine the naming schema used in each study.

Historically, glial cell types have been grossly grouped into astrocytes, microglia, oligodendrocytes, oligodendrocyte progenitor cells, and vascular cells. In the healthy white matter, axonal projections of neurons receive trophic factors through astrocytes and oligodendrocytes[20]. In this reciprocal relationship, neurons release the neurotransmitter glutamate that is taken up by oligodendrocytes and astrocytes. Additionally, oligodendrocytes myelinate neuronal projections, thus allowing for rapid signal transduction[21]. Astrocytic membranes cover 90% of the brain vasculature, making astrocytes the primary conduit for glucose, iron, and other essential molecules and a vital part of the blood-brain barrier[20]. Additionally, astrocytes and microglia are critical for synaptogenesis, synapse pruning, and elimination[22]. Astrocytic release of D-serine is essential for long-term potentiation and neuronal plasticity, lending to the tripartite synapse model of symbiotic reliance between astrocytes and neurons[23,24].

While many essential functions are associated with these gross cell types, single-cell omic characterization of glia has, revealed that these cell types might be an oversimplification of a diverse range of glial states. Table 1 summarizes the cell types and marker genes commonly used to define them in single-cell chromatin accessibility (ATAC-seq) and gene expression (RNA-seq) studies. The astrocyte cell type is widely reported; however, variations – such as astroependymocytes and astroendothelial cell type variations have been reported – due to the similarities between astrocytes and the cells that they complex with to maintain the blood-brain barrier (BBB). BBB-associated cell types are commonly grouped and referred to as vascular cells. However, the cells included in this designation are highly variable. Vascular cells typically consist of the cell

types: endothelial cells, ependymal cells, and mural cells (pericytes and vascular smooth

muscle cells, VSMCs). Immune cell types are commonly reported as microglia, the

resident immune cells of the CNS, and perivascular macrophages, which invade the

CNS in response to inflammation. The oligodendroglial lineage is frequently reported as

oligodendrocytes and oligodendrocyte progenitor cells (OPCs). Some studies

characterize oligodendrocytes by maturity stage (from OPC to post-myelinating

oligodendrocytes) and report corresponding maturity marker genes (Table 1).

Table 1. Common gene markers for cell types identified in single-cell studies. A review of
gene markers, named according to mouse nomenclature, that are used to identify cell types.

Cell type	Sub-Classification	Marker
Astrocytes	Pan-astrocyte	Aqp4, F3, Gfap, Glt1, Aldh11, Aqp4, Mfge8[49,50]
Vascular cells	Endothelia	Flt1, Xdh, Ly6c1[49,50]
	Ependymal cells	Foxj1, Myb, Rfx2, Zmynd10[50]
	Pericytes	Kcnj8, Pdgfrb, Acta2, Tbx18[50–52]
	VSMCs	Bgn, My19[49]
Immune cells	Microglia	C1qa, C1qc, Itgam, Ctss, Cx3cr1, Aif1[49,50]
	Perivascular macrophages	Mrc1, Lyve1, Lyl1, Spic[50]
Oligodendroglial Lineage	OPC	Pdgfra, Cspg4, Dcc, Ephb2, Ptch1, Mki67, Myrf, Enpp6[49,53,54]
	Pan-oligodendrocyte	Mog, Hapln2, Sox10, Olig2[49,50,54]
	Newly Generated	Enpp6, Mbp, Galc, Plp, Mag, Mog, Olig1, Itpr2, Prom1, Gpr17, Tcf7l2, Idh1, Cnksr3, Rnf122[49]
	Pre-myelinating	H2afj, Tmem141, Cd81, Rhob[49]
	Myelinating	Opalin, Mrf,Klk6, S100b, Pmp22, S100a1, Sec11c,Ptgds, Grm3, Car2, Mobp, Mbp, Cldn11[49,50]
	Post-myelinating	Klk6, Sec11c[50]

1.5.2 Glial cell states of the central nervous system

In addition to terminally differentiated cell types, CNS cells present epigenomic variability

within these types in the healthy and diseased brain. This variability mirrors the genomic

variability between organisms within a species. Within an organism, cells exhibit

epigenomic variability within a cell type. Within a single glial cell type, different cell states

have been shown to have diverse functions between and within brain regions. Single-cell

RNA-seq studies have found regionally restricted astrocyte types that correlate to

glutamate/glycine neurotransmitter spatial patterning[25]. In contrast, oligodendrocytes, which have been reported to account for 50-70% of the glia in the cerebral cortex, exhibit no regional identity[25]. This work defines the known glial cell state diversity of the healthy human brain and provides a novel characterization of rare glial cell state populations in Chapter 3. Here, I present the glial the cell state diversity that arises from inflammation responses in the diseased tissue that results from cerebral ischemia.

1.5.2.2 Cell state shifts in cerebral ischemia

Reactive glial cell types, astrocytes, and microglia alter their morphology and function in response to inflammation signals. Here, I present the glial cell state diversity in the disease focus of this work, cerebral ischemia.

Cerebral ischemia is a focal and acute loss of blood flow, which induces epigenetic remodeling in a gradient radiating away from the site of infarction. This impaired blood flow results in a loss of oxygen and glucose to the ischemic microenvironment, which initiates a cascade of spatial and temporal pathogenic events. Accumulation of lactic acid, a dearth of ATP, and accumulation of calcium and sodium in neurons decrease the osmotic pressure, leading to cytotoxic edema and excitotoxic neuron death[26]. The products of this ischemic cascade flood the tissue microenvironment with neurotransmitters and apoptotic factors that potentiate spreading depolarization, excitotoxicity, hypoxic and nitric stress responses, and secondary phenomena like inflammation and cell death in the hours and weeks following ischemia[26–28]. In the form of ischemic stroke, this acute focal cerebral hypoperfusion leads to cell death at the site of infarction and a lesion that expands in volume over time[27]. This well-characterized disease state exhibits a spatially progressive etiology.

The central nervous system (CNS) network of tightly regulated cell-cell interactions is dysregulated upon hypoperfusion, and sustained epigenetically shifted cell states persist

in the peri-infarct area. Glial cell types in the surrounding tissue adopt novel functions, and epigenomic remodeling occurs in the injured tissue[29–32]. Previous work has shown that diverse epigenetic remodeling occurs post-ischemia[33,34]. Cells in the peri-infarct area undergo epigenomic shifts, including global transcription repression via a three-fold increase in DNA methylation[32], global histone methylation[30,31,35] and global de-acetylation[35,36] in the days and weeks post-ischemia. Glial cell types adopt reactive states, leading to the adoption of novel cell functions that promote gliosis, where glia in the tissue surrounding the infract enter immunoreactive states, adopt CNS damage associated functions, and form a protective glial scar[37,38]. Cells that persist in this ischemic lesion undergo cell state changes dependent on cell type.

Additionally, post-ischemic myelin depletion is attributed to a dearth of mature differentiated myelinating oligodendrocytes and an accumulation of oligodendrocyte precursor cells (OPCs) in the infarct penumbra[39]. After an ischemic injury, OPCs proliferate normally but fail to migrate and differentiate into mature oligodendrocytes. This accumulation of OPCs and spatiotemporal depletion of mature oligodendrocytes in the penumbra leads to impaired trophic factor transmission from oligodendrocytes to neurons and neuron death[38,40,41].

While cerebral ischemia is an acute hypoperfusion event, a reduction in blood flow to watershed areas of the human brain is also associated with aging. In human patients, this chronic cerebral hypoperfusion (see 1.5.3 Discrepancies between the mouse and human brain) contributes to neurodegenerative injuries in the brain[42,43]. However, there is no comprehensive understanding of single-cell epigenomic cell state shifts that occur after cerebral ischemia. The chronic form of cerebral hypoperfusion has been shown to increase with patient age and has long been hypothesized to be the etiological basis of a brain lesion termed white matter hyperintensities (WMH)[44,45]. Despite the prevalence of

acute and chronic hypoperfusion injuries, no targeted clinical therapies currently exist. As the population age increases, the incidence of hypoperfusion injuries will increase, thus creating a need to characterize pathogenic cell-type-specific epigenetic state shifts to identify candidates for targeted therapies.

1.5.3 Discrepancies between the human and mouse brain

The mouse brain serves as a robust model for the human brain, and it allows for longitudinal and perturbation studies that would not be possible in human subjects. Taxonomically, humans and mice both fall within the mammalian class; however, evolutionarily, humans and mice are distantly related. Investigation into these genomes reveals 70% similarity between the protein-coding DNA sequences, which account for 1.5% of these genomes overall. This loose-genomic relationship is borne out when we compare the anatomy of the human and mouse brain. Humans are the only species with a larger white matter volume than grey matter volume[55]. This white matter is also the most recently evolved feature of the human brain and is thought to be uniquely involved in psychological disorders[56]. Unsurprisingly, some pathologies that are associated with the human brain are not observed in the mouse model organism.

One such pathology is leukoaraiosis, or white matter hyperintensities (WMH). In humans, WMH accumulate in advanced-age patients at a prevalence of 50.9% in healthy persons aged 44 to 48 years and 95% in healthy persons aged 60 to 90 years – and are associated with an increased risk of dementia, chronic depression, stroke, brain atrophy, and abnormal gait[57,58]. These WMHs present as high-intensity signals on a T2-weighted MRI and fluid-attenuated inversion recovery (FLAIR) sequences; however, their etiology is not known[59]. WMH resemble ischemic lesions in that they have low-myelin density and occur in low-perfusion regions of the brain[59]. Currently, hypotheses suggest that age-related and cerebral small vessel disease (CSVD)-related

hypoperfusion cause slight decreases in cerebral perfusion, which result in brief asymptomatic cerebral ischemic events, which present as WMHs[59,60]. While this ubiquitous and compelling pathology seems to be an age-related feature in humans, these same features are not observed in mice – likely due to the limited size of the murine white matter, the resulting lack of low-perfusion, or "watershed" areas on the mouse white matter, and their limited lifespan[61]. While humans experience age-related white matter pathologies that are unsurprisingly not recapitulated in mouse models, acute cerebral ischemia also varies between humans and mice.

The anatomical discrepancy in white matter volume is relevant to mouse models of cerebral ischemia. Ischemic strokes are experienced ubiquitously throughout the United States and the world[62,63], where 75–89% of strokes occur in individuals over the age of 65[60] and are highly correlated with heart disease, high blood pressure, and aging in humans. Additionally, cerebral ischemia induces epigenetic remodeling in cells in a gradient radiating away from the site of infarction[64,65]. Mouse white matter is limited to the corpus callosum and induced white-matter ischemic injury models commonly result in infarct expansion that encompasses both the white and grey matters[66,67]. While a mouse model of white-matter cerebral ischemia can reveal many critical biological components of hypoperfusion injury in humans, mice are an imperfect model, particularly for the study of white matter pathologies[68].

Studies that have explored the cellular diversity in the human and mouse brain have also revealed chromatin accessibility, transcriptomic, and functional differences between comparable mouse and human cell populations in health and disease. While investigations into murine neuronal circuits have provided a framework for the mammalian brain, human astrocytes vary significantly from murine astrocytes[69]. Human astrocytes are larger with more branched processes which make contact with 100x more

synapses, and also have been shown to have 5x faster calcium wave propagation than mouse astrocytes[70]. While the functional effect of these differences between human and mouse astrocytes is not fully understood, we do know that pathologies such as Alexander's Disease (AxD), caused by a mutant form of GFAP, dramatically inhibit neurodevelopment in humans but do not present with neurodevelopmental deficiencies in AxD mouse models[70]. This disease case of a morphologically and functionally distinct astrocyte response in humans that does not occur in mice suggests that we should practice caution when assuming that mice and humans share cell types and states. Additionally, in the research presented here and in a meta-analysis of numerous chromatin accessibility and transcriptomic studies, I find that oligodendrocytes demonstrate clear maturity stages in the mouse brain, which have not been observed in the human brain (see 3.5.7 Oligodendrocytes occupy a continuous chromatin accessibility landscape with few distinct sub-states).

Chapter 2: Spatially mapped single-cell chromatin accessibility

This chapter contains a modified version of material that appeared in the author's publication: Thornton, Casey A., et al. "Spatially mapped single-cell chromatin accessibility" Nature Communications 12, 1274 (2021). Copyright © 2021, The Author(s).

2.1 Authors and affiliations

Casey A. Thornton[1], Ryan M. Mulqueen[1], Kristof A. Torkenczy[1], Andrew Nishida[1], Eve G. Lowenstein[1], Andrew J. Fields[1], Frank J. Steemers[2], Wenri Zhang[3], Heather L. McConnell[4], Randy L. Woltjer[5], Anusha Mishra[4,6], Kevin M. Wright[7], Andrew C. Adey[1,6,8,9,*]

1) Molecular and Medical Genetics, Oregon Health & Science University, Portland, OR.

2) Illumina Inc. San Diego, CA

3) Anesthesiology and Peri-Operative Medicine, Oregon Health & Science University, Portland, OR.

4) Jungers Center for Neurosciences Research, Department of Neurology, Oregon Health & Science University, Portland, OR.

5) Department of Pathology, Oregon Health & Science University, Portland, OR.

6) Knight Cardiovascular Institute, Oregon Health & Science University, Portland, OR.

7) The Vollum Institute, Oregon Health & Science University, Portland, OR.

8) CEDAR, Oregon Health & Science University, Portland, OR.

9) Knight Cancer Institute, Oregon Health & Science University, Portland, OR.

* To whom correspondence should be addressed: adey@ohsu.edu

2.2 Author contributions

A.C.A. and C.A.T. conceived of the idea. C.A.T. performed all experiments described with assistance from R.M.M. and A.J.F.; F.J.S., K.M.W., and A.M. contributed to experimental design and data interpretation. C.A.T. performed data processing, and analysis with assistance from K.A.T., R.M.M., A.N. and E.G.L. W.Z., and H.M. performed stroke surgeries. R.W. identified, isolated, and cryopreserved human primary visual cortex with assistance from C.A.T. C.A.T. and A.C.A. wrote the manuscript with input from all authors.

2.3 Abstract

High-throughput single-cell epigenomic assays can resolve cell type heterogeneity in complex tissues; however, spatial orientation is lost. Here, we present single-cell combinatorial indexing on Microbiopsies Assigned to Positions for the Assay for Transposase Accessible Chromatin, or sciMAP-ATAC, as a method for highly scalable, spatially resolved, single-cell profiling of chromatin states. sciMAP-ATAC produces data of equivalent quality to non-spatial sci-ATAC and retains the positional information of each cell within a 214-micron cubic region, with up to hundreds of tracked positions in a single experiment. We apply sciMAP-ATAC to assess cortical lamination in the adult mouse primary somatosensory cortex and in the human primary visual cortex, where we produce spatial trajectories and integrate our data with non-spatial single-nucleus RNA and other chromatin accessibility single-cell datasets. Finally, we characterize the spatially progressive nature of cerebral ischemic infarction in the mouse brain using a model of transient middle cerebral artery occlusion.

2.4 Introduction

Heterogeneous cell types coordinate in complex networks to generate emergent properties of tissues. These cell types are not evenly dispersed across tissues, creating spatially localized functionality. In many disease states, this becomes more apparent as the affected organ experiences spatially progressive etiologies. For example, following cerebral ischemic injury, astrocytes and microglia enter reactive states that are metered by proximity to the site of infarction[71], but this spatial information has, so far, been difficult to assess. Single-cell technologies have advanced cell type and state characterization efforts by enabling the isolation of signals from individual cells within a sample, thus resolving the heterogeneity of complex tissues. Applications of single-cell technologies have identified novel cell types with characteristic -omic signatures in the highly complex tissue of the brain[21,49]. In the cerebral cortex, specifically, cells form an intricate layered hierarchical structure comprised of both neuronal and glial cell types that generate sensory, motor, and associational percepts[72]. Layer-specific gene expression profiles of cortical neurons and astrocytes have been characterized by spatial transcriptomic approaches and immunohistochemical (IHC) staining; however, spatially mapped epigenetic states of cortical cells have yet to be directly assayed without relying on the data integration[73–75].

To address this challenge, several strategies have emerged to assay transcription either directly in situ or a regional manner. The former techniques utilize fluorescence in situ hybridization (FISH)[76–78] or in situ RNA sequencing[32,79]. While powerful, FISH methods require the use of a defined probe set and are limited to the identification of DNA and RNA sequences. In contrast, technologies that utilize array-based mRNA barcoding do not require a defined set of genes and operate similarly to single-cell RNA-seq methods[80,81], thus allowing for whole transcriptome profiling. Initial iterations of these

platforms capture regional transcription over multiple cells; however, higher resolution variants may facilitate single-cell resolution. Unfortunately, these platforms rely on the relatively easy access to mRNA molecules that can be released from the cytoplasm and hybridized to barcoding probes, making the expansion into nuclear epigenetic properties challenging. With the wealth of epigenetic information that resides in the nucleus and the value it can add to characterizing complex biological systems[82–84], we sought to address this challenge by harnessing the inherent throughput characteristics of single-cell combinatorial indexing assays[85,86].

Here, we present single-cell combinatorial indexing from Microbiopsies with Assigned Positions for the Assay for Transposase Accessible Chromatin (sciMAP-ATAC). sciMAP-ATAC preserves the cellular localization within intact tissues and generates thousands of spatially resolved high-quality single-cell ATAC-seq profiles. As with other "sci-" technologies, sciMAP-ATAC does not require specialized equipment and scales nonlinearly, enabling high-throughput potential. Building upon multiregional sampling strategies[87,88], where several regions are isolated, we reasoned that the sample multiplexing capabilities of combinatorial indexing could be utilized to perform high-throughput sampling at resolutions approaching those of array-based spatial transcriptional profiling, all while retaining true single-cell profiles. Unlike multiregional sampling, we perform high-density microbiopsy sampling, ranging from 100 to 500 µm in diameter, on cryosectioned tissue sections, between 100 and 300 µm in thickness, to produce up to hundreds of spatially mapped punches of tissue, each producing a set of single-cell chromatin accessibility profiles (Fig. 7a). We demonstrate the utility of sciMAP-ATAC by profiling the murine and human cortex, where distinct cell type compositions and chromatin profiles are observed based on the spatial orientation of the punches, and further extend the platform to characterize cerebral ischemic injury in a

mouse model system, where cell type compositions and epigenetic states are metered

by proximity to the injury site (Fig. 8).
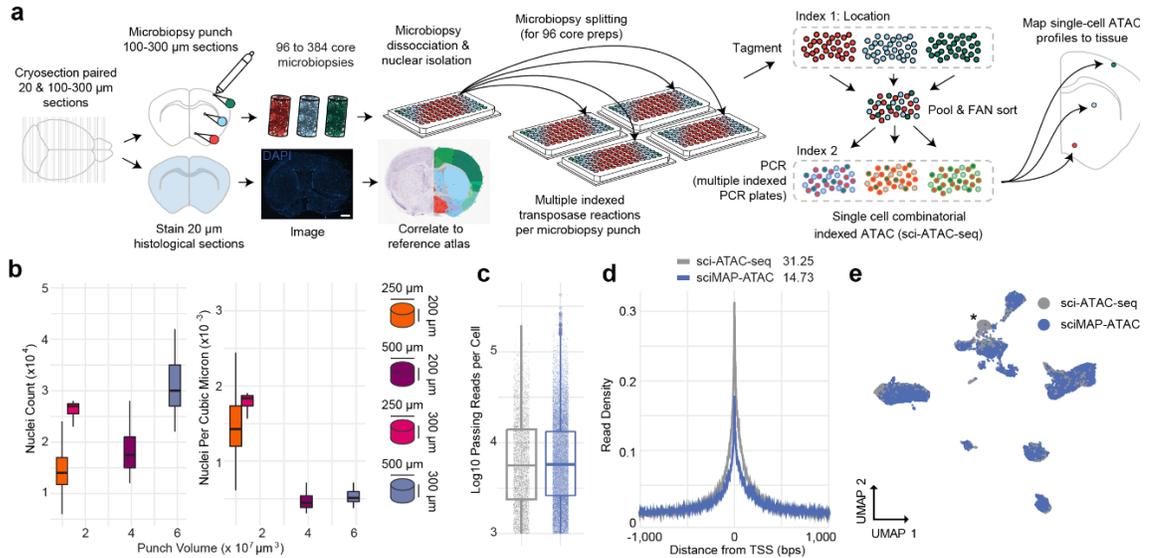
Figure 7. sciMAP-ATAC schematic and performance. a, sciMAP-ATAC workflow. Cryosectioning of alternating 20 μm (histological) and 100–300 μm (sciMAP-ATAC) slices are obtained. Thin (20 μm) slices are stained and imaged for use in spatial registration (scale bar, 1 mm) to a reference atlas (Allen Mouse Brain Atlas: http://atlas.brain-map.org/atlas?atlas=1&plate=100960312, ref. 25). Thick (100–300 μm) slices are carried through high-density microbiopsy punching (100–500 μm diameter) in the cryostat chamber. Punches are placed directly into wells of a microwell plate for nuclei isolation, and washed prior to splitting into multiple wells for indexed transposition and the sci-ATAC-seq workflow. **b,** Four punch volumes were assessed for nuclei yield using either a 250 or 500 μm diameter punch on a 200 or 300 μm thick section. Total nuclei isolated for each punch is shown on the left, and normalized for tissue voxel volume on the right, representing the efficiency of extraction from each punch, for punches with dimensions 250 × 200 μm (n = 48), 250 × 300 μm (n = 15), 500 × 200 μm (n = 46), and 500 × 300 μm (n = 7). Center line represents median, lower and upper hinges represent first and third quartiles, and whiskers extend from hinge to ±1.5 × IQR. **c,** Passing reads per cell from sci-ATAC-seq (n = 4102 cells examined from a single mouse brain experiment) and sciMAP-ATAC (n = 15,552 cells examined from two independent mouse brain experiments), which are comparable at the level of depth sequenced. Center line represents median, lower and upper hinges represent first and third quartiles, whiskers extend from hinge to ±1.5 × IQR, individual cells represented as colored dots. **d,** ATAC read signal at transcription start sites (TSSs) and surrounding base pairs (bps) for sci-ATAC-seq and sciMAP-ATAC. Enrichment for sci-ATAC-seq is greater than that of sciMAP-ATAC, likely due to increased processing time of isolated nuclei prior to transposition. **e,** UMAP of sciMAP-ATAC and sci-ATAC-seq libraries from mouse brain group closely together. Asterisk indicates a population of 734 cells, derived from spinal cord, which was not sampled during microbiopsy punching. Source data are provided as a Source data file.
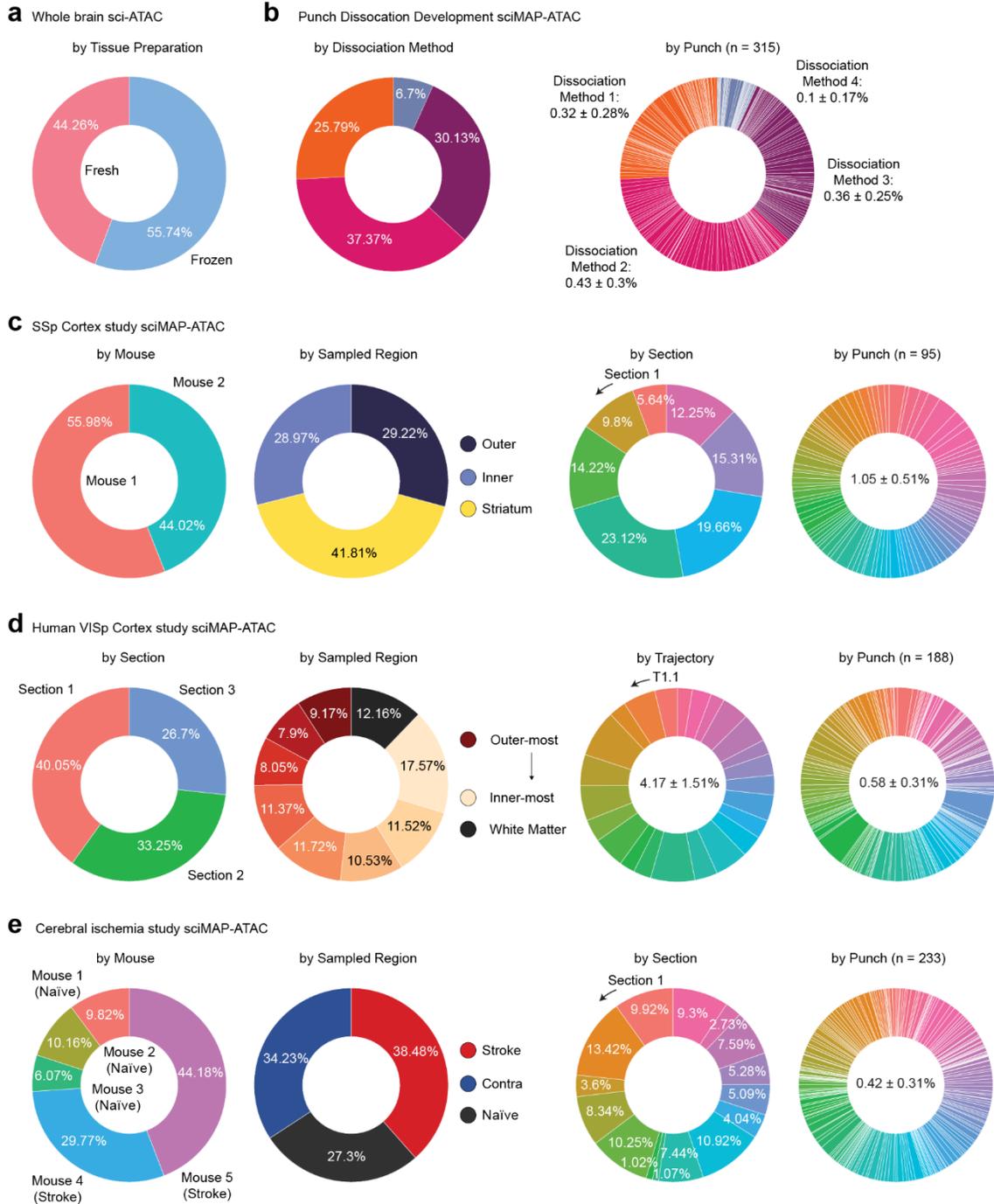
Figure 8. Overview of single-cell ATAC profiles produced across experimental conditions. Quality-passing single-cell ATAC-seq profiles for each experimental condition or spatially resolved punch (for sciMAP-ATAC) as a percentage of the experiment (or mean ± SD) for: **a,** sci-ATAC-seq on fresh vs. frozen mouse whole brain hemisphere; **b,** sciMAP-ATAC development across four dissociation methods and punches; colored by each dissociation method (n = 315 individual punches); **c,** sciMAP-ATAC on mouse SSp by biological replicate, class of sampled region, individual section, and punch (n = 95 individual punches).; **d,** Human VISp sciMAP-ATAC by section, position of sampled region, individual trajectory (T) and punch (n = 188 individual punches); and **e,** sciMAP-ATAC on a mouse model of cerebral ischemia by biological replicate, class of sampled region, section and punch (n = 233 individual punches).

2.5 Results

2.5.1 Single-cell combinatorial indexed ATAC-seq from microbiopsy punches

Single-cell ATAC-seq requires the isolation and processing of nuclei such that the nuclear scaffold remains intact to facilitate library preparation via transposition in situ; it also requires that the chromatin structure is maintained to produce a chromatin accessibility signal. We and others have explored methods for tissue preservation that are compatible with single-cell ATAC-seq[85,89]; however, we sought to confirm that these strategies are compatible with freezing techniques used for cryosectioning and IHC staining of tissue. We tested our workflow on mouse whole brain samples by processing one hemisphere using flash-freezing methods designed for tissue freezing medium (TFM) embedding and cryosectioning (see 2.7 Methods) and processing the paired hemisphere as fresh tissue. Our previously established non-spatially resolved sci-ATAC-seq workflow[89] was performed on both hemispheres, including pooling post-transposition for sorting, PCR amplification, and sequencing. Flash-frozen and fresh nuclei produced nearly identical passing reads per cell at the depth they were sequenced, along with comparable fractions of reads present in a set of aggregate mouse ATAC-seq peaks (FRiS; 0.93 and 0.91 for fresh and frozen, respectively; Fig. 9a,b).

We then explored techniques for cryosectioning flash-frozen TFM-embedded tissue at thicknesses compatible with microbiopsy punching. Typically, cryosectioning is used to produce sections for imaging applications, and thicker sectioning results in tissue fracture. Drawing on past literature[90], we carried out a series of experiments testing several sectioning thicknesses and punch diameters, followed by nuclei isolation and debris cleanup on flash-frozen, embedded mouse brain microbiopsy punches. We found

that holding cryo-chamber and chuck temperatures at −11 °C improves the flexibility of the fragile flash-frozen tissue while maintaining adherence of embedded tissue to the sample mount, thus allowing for uninterrupted sectioning of alternating 100–300 µm sections for punching and paired 20 µm sections for histology (Fig. 7a). This approach facilitates the acquisition of both sections for microbiopsy punching and paired sections compatible with IHC staining and high-resolution microscopy. Cryopreservation of 100–300 µm/20 µm slide decks at −80 °C allow for long-term sample storage and the ability to test hypotheses by staining after analysis of the spatially resolved chromatin accessibility profiles; however, we note that sections stored for ~3 months result in an overall loss of quality in transcription start site (TSS) enrichment.

Microbiopsy punching of 100–300 µm sections performed within a cooled chamber (see 2.7 Methods) allows for the isolation of microscopic pellets of nuclei that readily dissociate in nuclear isolation buffer (NIB) after mechanical dissociation by trituration. We observed minimal loss after pelleting and washing nuclei, an important step for the removal of mitochondria, which can deplete the available pool of transposase because of the high transposition efficiency into mitochondrial DNA[91]. Nuclei isolation, as measured by nuclei per cubic micron, was more efficient for volumetrically smaller punches (Fig. 7b). This implies that smaller punches dissociate more readily because of a higher surface area to volume ratio. Thus, higher resolution punches yield more nuclei, respective of volume.
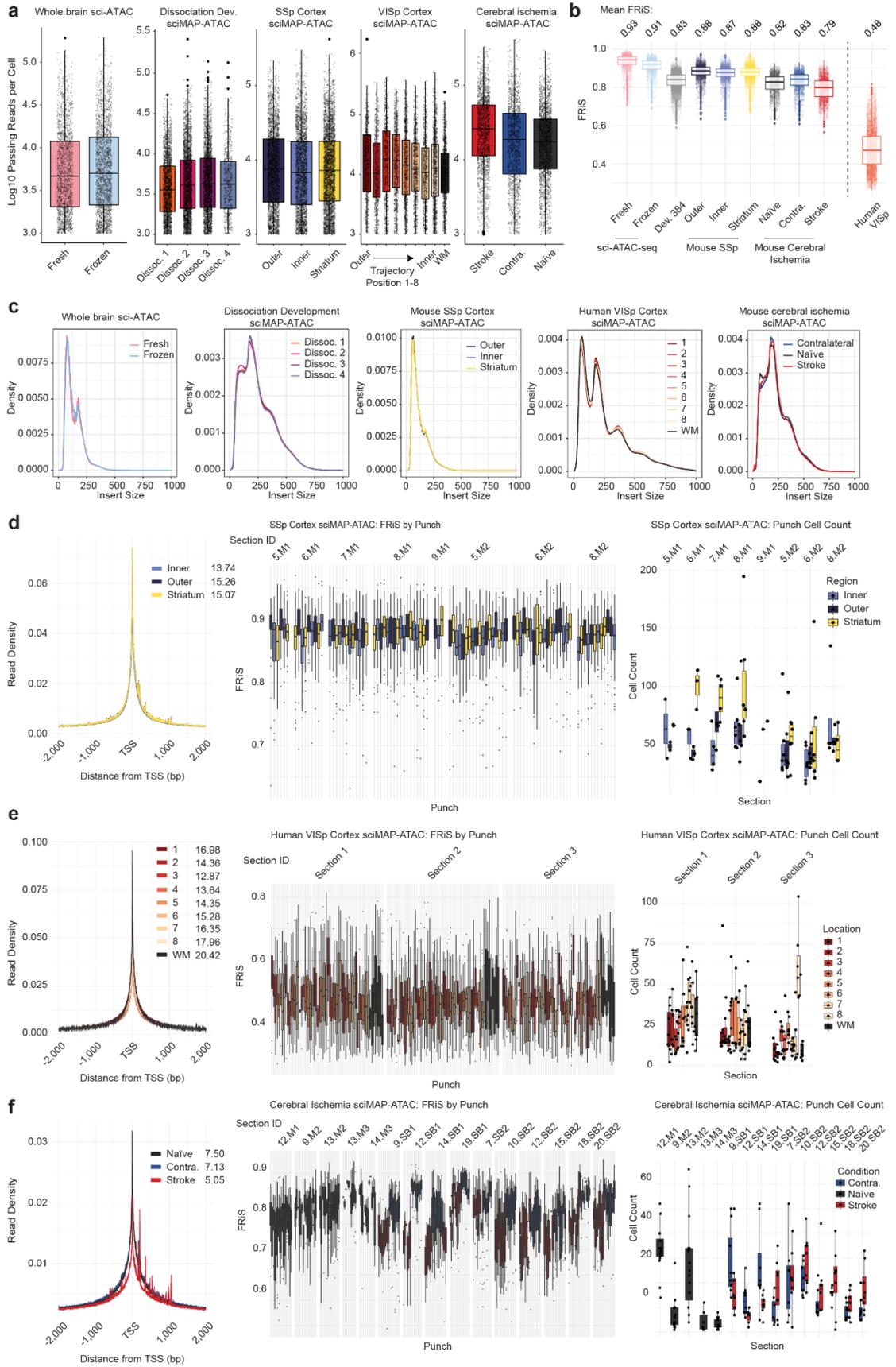
Figure 9. Quality metrics across all experiments. a, Log10 passing reads obtained per cell at the depth of sequencing for all experiments: mouse whole brain sciATAC (n = 4,569 cells), mouse dissociation development sciMAP-ATAC (n = 8,011 cells), mouse SSp cortex sciMAP-ATAC (n = 7,779 cells), mouse VISp cortex sciMAP-ATAC (n = 4,547 cells), and mouse cerebral ischemia sciMAP-ATAC (n = 5,081 cells); as described in Fig. 8. Center line represents median, lower and upper hinges represent first and third quartiles, whiskers extend from hinge to ± 1.5x IQR, individual cells represented as dots. **b,** The fraction of reads present in a reference set of peaks (FRiS) for all cells in each experiment as in a. The master list of peaks for mouse are aggregated from ATAC-seq data produced by the ENCODE project, and for human it is from a single study on DNAse hypersensitivity[28]. Center line represents median, lower and upper hinges represent first and third quartiles, whiskers extend from hinge to ± 1.5x IQR, individual cells represented as colored dots. **c,** Insert size distributions for all experiments. **d-f,** Left: aggregate read density at transcription start sites (TSSs) and surrounding base pairs (bps) present in the genome with TSS enrichment values listed by each class calculated using the ENCODE method; middle: FRiS distributions for all cells within each punch produced in the experiment split by section and mouse cerebral ischemia sciMAP-ATAC (); and right: Punch distributions of cell counts for each category within the experiment split by section, for mouse SSp (d, n = 7,779 cells examined over 95 independent punches taken from 8 sections), human VISp (e, n = 4,547 cells examined over 188 independent punches taken from 3 sections) and mouse cerebral ischemia (f, n = 5,081 cells examined over 233 independent punches taken from 15 sections) experiments. Center line represents median, lower and upper hinges represent first and third quartiles, whiskers extend from hinge to ± 1.5x IQR, individual cells represented as dots.

2.5.2 sciMAP-ATAC performance and quality assessment

We applied these techniques to perform sciMAP-ATAC, where we tested four methods of punch dissociation (see 2.7 Methods). We utilized a workflow similar to our established sci-ATAC-seq method, with each indexed transposition reaction performed on an individual punch, for a total of 384 transposition reactions performed in four 96-well plates. Reactions were pooled, and indexed nuclei were distributed via fluorescence-assisted nuclei sorting (FANS) to wells of four new 96-well plates for indexed real-time PCR, followed by pooling and sequencing. The resulting library produced 8011 cells passing filters, for an estimated doublet rate of 2.5% based on the total indexing space of 384 × 384 (see 2.7 Methods) and a mean of 12,052 passing reads per cell (unique reads, aligned to autosomes or X chromosome at q10 or higher; Fig. 9a) at the depth sequenced and potential to reach 23,830 mean passing reads per cell with additional sequencing (see 2.7 Methods). This is comparable to the mean passing reads per cell from the whole brain sci-ATAC-seq library at 11,987 (projected

mean passing reads of 24,672 and 32,029 for fresh and frozen preparations, respectively; Fig. 7c and Fig. 9a). We observed a mean of 112 passing cells per punch. This could be increased if additional PCR plates were sorted, as the pool of indexed nuclei was not depleted during FANS. A comparison between the four dissociation methods enabled us to identify an optimal means of punch processing that produced the highest cell counts per punch with high-quality cell profiles (see 2.7 Methods; Figs. 8b and 9a), which was used for all subsequent experiments. Across all sciMAP-ATAC datasets produced in this study on healthy mouse brain tissue, we achieve a TSS enrichment of 14.73, within the "acceptable" range prescribed by ENCODE (10–15, mm10 RefSeq annotation) and just shy of "ideal" (>15). This is substantially below that of our sci-ATAC-seq preparation, with a TSS enrichment of 31.25; however, we note that enrichment of more than double the "ideal" standard is exceptionally high (see 2.7 Methods, Fig. 7d). In line with the lower TSS enrichment in sciMAP-ATAC, we also observed a reduction in the fraction of reads present in a mouse reference peak set (FRiS; see 2.7 Methods), with a mean ranging from 0.83 to 0.87, compared to 0.91 and 0.93 for sci-ATAC-seq (Fig. 9b). Finally, we performed an integrated analysis across these preparations that revealed negligible batch effects (Fig. 7e and Fig. 11a, b). We observed a single exception in the form of a population of cells present only in the nonspatial dataset, which, upon inspection, were determined to be spinal cord-derived interneurons (Fig. 11c, d) and not present in coronal sections that were used in spatial experiments. Taken together, with improvements and validation on sample preparation, cryosectioning, nuclei isolation, and the general sci-ATAC-seq protocol, we generated a robust method to obtain the spatial information that we sought to test in a complex system.

2.5.3 sciMAP-ATAC in the adult mouse somatosensory cortex

To establish the ability of sciMAP-ATAC to characterize single cells within a spatially organized tissue, we applied the technique to resolve murine cortical lamination within the primary somatosensory cortex (SSp). We harvested intact whole brain tissue from three wild-type C57/Bl6J adult male mice, flash-froze the tissue and prepared whole-brain slide decks of 200 μm microbiopsy slides, each interspersed with three 20 μm histological slides. To orient sections to intact mouse brain and to establish the quality of histological section prepared according to the sciMAP-ATAC protocol, we stained nuclei using DAPI and IHC stained for SATB2 to resolve cortical layers (see 2.7 Methods, Fig. 10a). DAPI imaging was then matched to the adult mouse Allen Brain Reference Atlas[92], which enabled the determination of the SSp location within adjacent sections for punch acquisition. SATB2 imaging demonstrated the quality of histological sections across diverse fixation protocols (4% PFA post-fixation for 10 min and 70% ethanol post-fixation for 30 s) and generated a high signal-to-noise ratio canonical for SATB2 IHC staining[93] (Fig. 10b). Microbiopsy punches were then taken from three regions: (i) outer (L2–4) SSp cortical layers, (ii) inner (L5 and 6) SSp cortical layers, and (iii) throughout the striatum. The striatum is rich in glia and is absent of cortical glutamatergic neurons and cortical lamination. Therefore, the striatum punches served as a negative control for these features and also bolstered single-cell glial cell type identification. In total, 96 individual tissue punches were obtained, split evenly between the three categories over eight coronal sections spanning the SSp (Fig. 10a). After nuclei isolation, each well of the plate containing a single punch was split across four wells, resulting in four 96-well plates for subsequent indexed transposition, providing four tagmentation technical replicates for each punch. Post-transposition, nuclei were pooled and distributed to two 96-well PCR plates for the second tier of indexing and then sequenced (see 2.7 Methods).
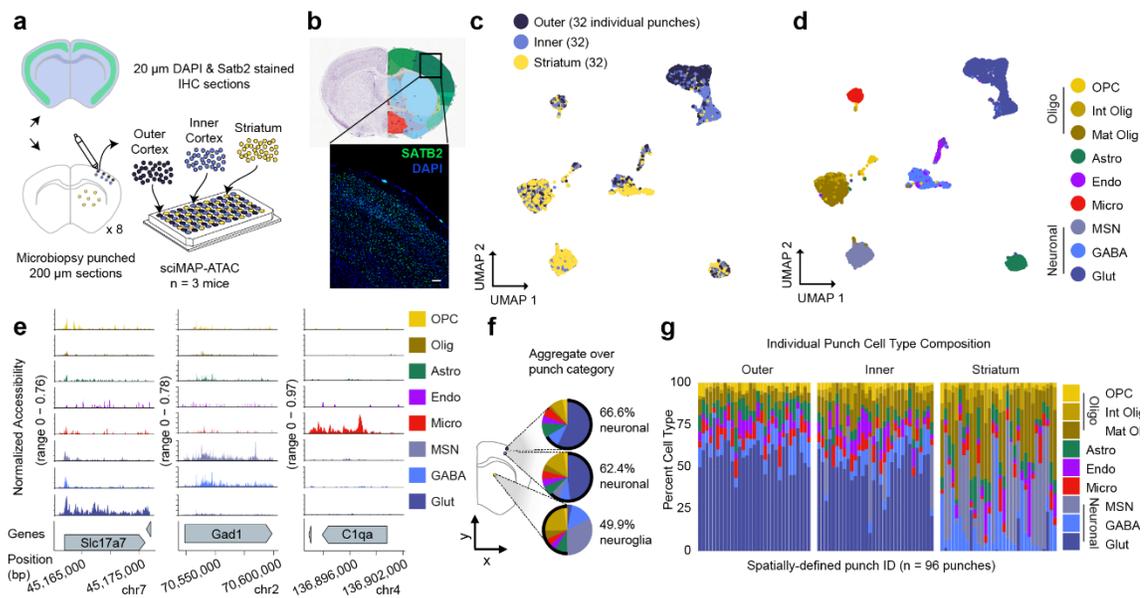
Figure 10. sciMAP-ATAC reveals spatially distinct cell type composition in the mouse somatosensory cortex. a, Experiment schematic of sciMAP-ATAC in the mouse somatosensory cortex. **b,** DAPI and SATB2 staining of SSp cortex from sciMAP-ATAC histological section (scale bar, 50 µm) in reference to matched rerference atlas image (Allen Mouse Brain Atlas: http://atlas.brain-map.org/atlas?atlas=1&plate=100960312, ref. 25). **c,** UMAP of 7779 cells colored by punch location category. Each category contains cells from 32 spatially distinct tissue punches. **d,** UMAP as in **c**, colored by cell type (OPC oligodendrocyte precursor cells, Int Olig intermediate oligodendrocytes, Mat Olig mature oligodendrocytes, Astro astrocytes, Endo endothelia, Micro microglia, MSN medium spiny neurons, GABA GABAergic (inhibitory) neurons, Glut glutamatergic (excitatory) neurons). **e,** ATAC-seq profiles for cells aggregated by cell type for marker genes; colored by cell type as in **d**. **f,** Aggregate cell type composition over punches belonging to the broad region categories; colored by cell type as in **d**. **g,** Cell type composition for each of the 96 individual punches split by broad region category; colored by cell type as in **d**. Source data are provided as a Source data file.

We processed the raw sequence data (see 2.7 Methods), which resulted in 7,779 cells passing quality filters (estimated doublet rate of 4.9%; see 2.7 Methods). Our mean passing reads per cell was 17,388, with a projected total passing mean reads per cell of 37,079 ("Methods"), a TSS enrichment ranging from 13.74 to 15.26, and nucleosomal banding present in the library insert size distribution (Fig. 9a–d). A median of 81 single-cell profiles was obtained per punch, with little bias for punch target region or section (Fig. 9d). Subsequent peak calling, topic modeling, and dimensionality reduction (see 2.7 Methods) revealed cell groupings that were either mixed between the three regional categories or highly enriched for cells derived from the cortex, which was further divided

39

by outer versus inner punch location (Fig. 10c, Fig. 11e, f and Fig. 9). The overlay of

spatial data on the UMAP projection fits with our expectation that glutamatergic

(excitatory) neurons are cortex exclusive, displaying an absence of punch-to-punch

cross-talk or contamination. In addition, these cells were integrated with prior sciMAP-

ATAC and sci-ATAC-seq experiments where excitatory neuron clusters were also

dominated by cortex-derived punches, with a shared spatial bias between upper and

lower punch positions. This demonstrates that spatial datasets can be integrated with

nonspatial datasets to provide additional spatial information to those datasets, using

label transfer or other analysis techniques (Fig. 11a, b).
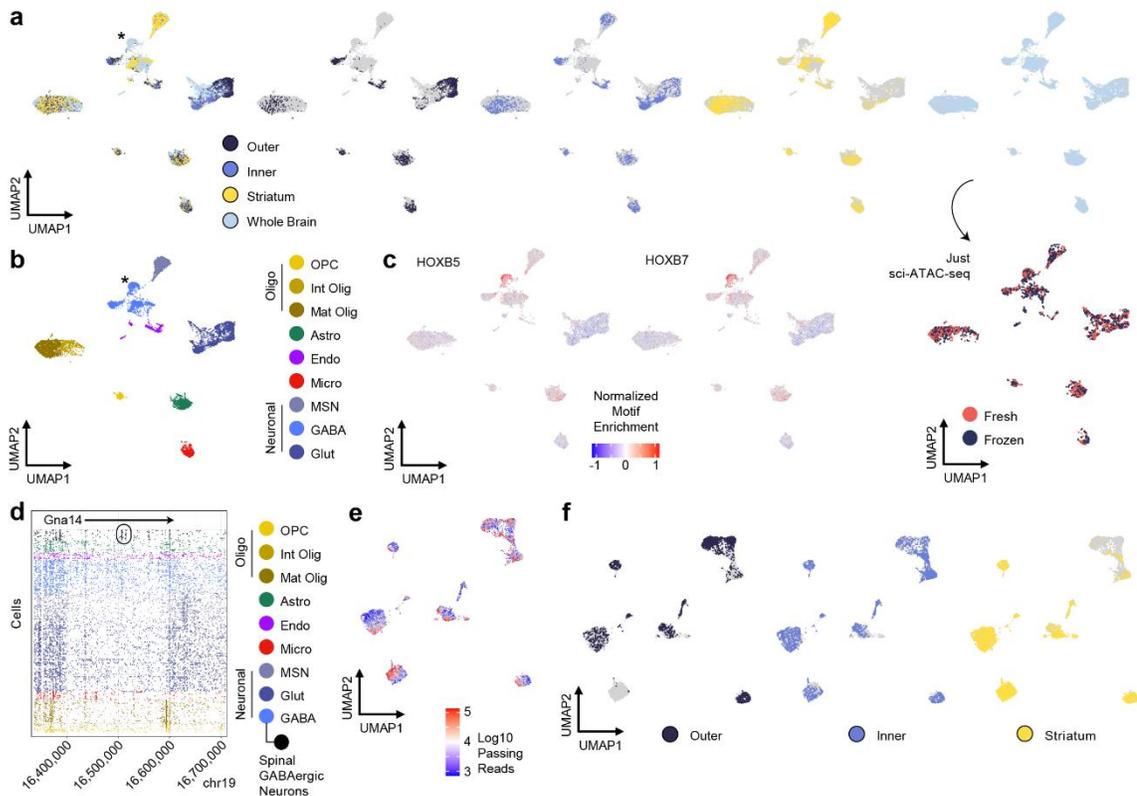
Figure 11. Extended analysis of the mouse somatosensory cortex sciMAP-ATAC dataset. a, Integration of all healthy mouse brain sci-ATAC-seq and sciMAP-ATAC datasets visualized in a UMAP. From left to right: all cells colored by the regional category of punch position (outer cortex, inner cortex, striatum) for the SSp experiment and then cells from whole brain experiments (sciMAP-ATAC and sci-ATAC-seq). Asterisk indicates the population of cells only present in the whole brain dataset. Cells are grayed out except for those from punches taken from the outer cortex, inner cortex, stratum and then whole brain. Below the whole brain

panel, cells derived from the sci-ATAC-seq experiment on fresh and frozen brain hemispheres are indicated. **b,** The same integrated UMAP with cells colored by identified cell type, as defined in Fig. 10d. Asterisk indicates the population of GABAergic neurons only present in whole brain datasets that represent spinal cord derived interneurons. **c,** HOXB5 and HOXB7 are two example motifs that exhibit increased accessibility in the spinal cord derived interneuron population. **d,** ATAC reads for cells (rows) are shown for the *Gna14* locus with cells colored by cell type, as defined in Fig. 10d, with the addition of the spinal GABAergic neuron subcluster. The cluster representing spinal cord derived interneurons is shown in black with the uniquely accessible loci circled. **e,** UMAP of the SSp dataset with cells colored by log10 passing read counts. **f,** UMAP of the SSp dataset with cells grayed out except for each of the three regional punch categories; outer cortex, inner cortex, and striatum.

We identified eleven clusters over eight broad cell type groups corresponding to glutamatergic neurons, GABAergic (inhibitory) neurons, GABAergic medium spiny neurons (MSNs; also referred to as spiny projection neurons (SPNs)), oligodendrocyte precursor cells (OPCs), newly formed or intermediate oligodendrocytes, mature oligodendrocytes, astrocytes, microglia, and endothelial cells based on the chromatin accessibility signature of regulatory elements proximal to marker genes (see 2.7 Methods; Fig. 10d, e and Supplementary Data 1). GABAergic neurons subdivide into non-layer-specific cortical GABAergic neurons and striatum-derived MSNs. In contrast, glutamatergic neurons separate along the dorsal-ventral axis, as determined by punch position. This recapitulates known neuronal cell state biology, where glutamatergic pyramidal neurons express cortical layer(s)-specific markers that define the spatially defined cortical layers. Within the SSp-derived cells, we observed 66.6%, 62.4%, and 49.9% of cells corresponding to neurons in the inner cortex, outer cortex, and striatum, respectively. These equate to glia to neuron ratios (GNRs) of 0.50, 0.60, and 1.00 from the inner cortex, outer cortex, and striatum, respectively, which correspond to previously reported mouse cerebral cortex and striatum GNRs of 0.66 and 0.97, respectively[94]. In addition to coarse cell type characterization across the major punch categories, we determined cell type composition for each individual spatially resolved punch (Fig. 10g). For cortical punches, little variance was observed within the outer and inner punch categories; however, we did observe increased variability in the proportion of MSNs in

the striatum punches, ranging between 2.78% and 72.64%, suggesting a non-even distribution of these cells, which is confirmed by MSN cell type marker, *Drd1*, in situ hybridization in adult C57BL/6J striatum (Allen Mouse Brain Atlas)[92].

2.5.4 Analysis of individual punch sciMAP-ATAC profiles and spatial comparisons

We next characterized the single-cell ATAC profiles produced from a single tissue punch. We isolated cell profiles that were from punch F5 (n = 90 cells), an inner cortex punch, and performed the same analysis as above using the set of peaks called on the full dataset. This produced a set of topic weights that contained a clear structure and were associated with specific cell types (Fig. 12a). This was also clear in the UMAP projection, with three primary clusters of cells identified (Fig. 12b). Two of these groups were dominated by one cell type, including glutamatergic neurons and GABAergic neurons, with the third group comprised predominantly of glial cell types.
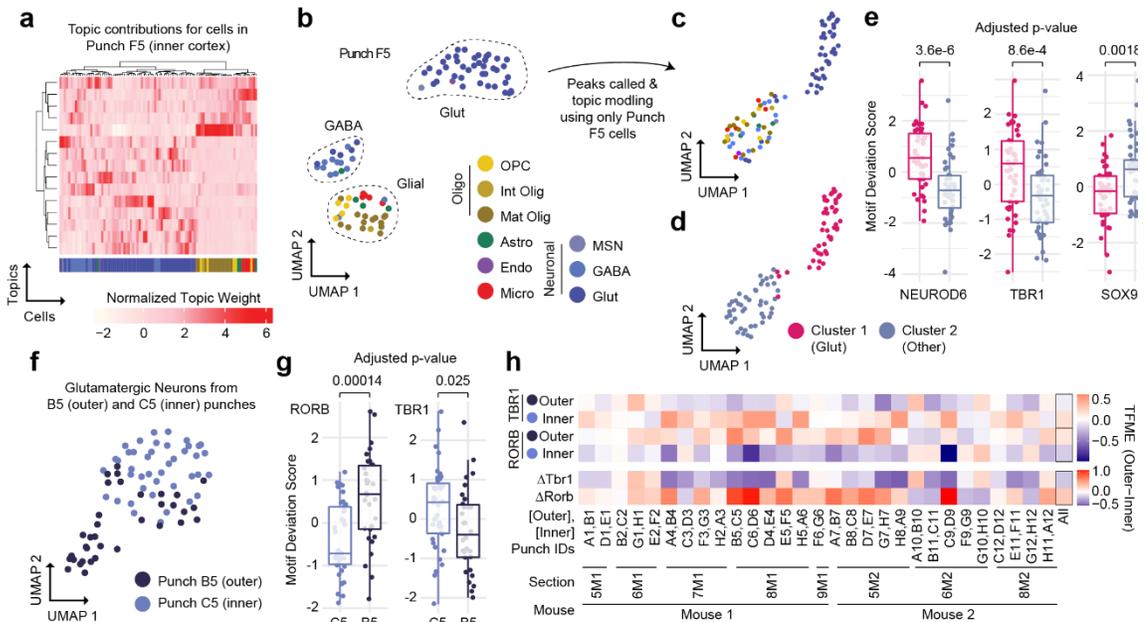
Figure 12. sciMAP-ATAC enables the analysis and comparison of cells and cell types from individual spatial positions. a, Topic weight matrix for cells present only in a single punch (F5, inner cortex punch), annotated by cell type (bottom); colored by cell type from the

42

full dataset (Fig. 10d). **b,** UMAP of cells from punch F5 showing spatially distinct groupings for cell type; colored by cell type from the full dataset (Fig. 10d). **c,** Isolated analysis of cells from Punch F5 for peak calling, topic modeling, and visualized via UMAP; colored by cell type from the full dataset (Fig. 10d). **d,** Two major clusters identified from the isolated analysis of punch F5 (Glut glutamatergic (excitatory) neurons). **e,** Transcription factor motif enrichments for the isolated analysis of punch F5, indicating that cluster 1 ($n = 44$ cells) is made up of glutamatergic neurons and cluster 2 ($n = 45$ cells) is made up of other cell types. Center line represents median, lower, and upper hinges represent first and third quartiles, whiskers extend from hinge to $\pm 1.5 \times$ IQR, individual cells represented as colored dots. **f,** UMAP of all glutamatergic neuron cells from two adjacent punches (C5, inner cortex, and B5, outer cortex) after topic modeling on the isolated cell profiles. **g,** Transcription factor motif enrichments for glutamatergic cells from adjacent punches from inner cortex ($n = 39$ cells) and outer cortex ($n = 30$ cells) shown in **f**; colored by individual punch as in **f**. Two-sided Mann–Whitney U test with Bonferroni–Holm correction. Center line represents median, lower and upper hinges represent first and third quartiles, whiskers extend from hinge to $\pm 1.5 \times$ IQR, individual cells represented as colored dots. **h,** Motif enrichments across glutamatergic neurons across all punch pairs. TFME transcription factor motif enrichment. Source data are provided as a Source data file.

We then took the examination of this individual punch further by performing all aspects

of the analysis, including peak calling, on only the cell profiles present in punch F5. From

those 90 cells, we were able to call 8460 peaks which were sufficient to perform topic

modeling and UMAP visualization and identify two distinct clusters: one comprised of

glutamatergic neurons and the second containing all other cell types, based on the cell

type identities established in the analysis of the full dataset (Fig. 12c, d). A comparison

of global motif enrichment between the two clusters revealed elevated NEUROD6 and

TBR1 and depleted SOX9 motif accessibility in the cluster comprised of glutamatergic

neurons, suggesting very coarse cell type class assignment can be performed on data

from a single punch analyzed in isolation (Fig. 12e). Further resolution of cell types on

such a small number of cells, especially without leveraging larger peak sets, is not likely

feasible simply due to the low abundance of certain cell types—for example, there was

only one endothelial cell present in punch F5. However, it is unlikely that individual

punches would be profiled alone in an experiment and the throughput provided in

sciMAP-ATAC enables the identification of low-abundance cell types in the aggregate

dataset, which can be used when performing analysis on individual punch positions.

Finally, we explored whether we could identify and characterize spatially distinct chromatin properties from a single cell type present within two adjacent punches. We isolated cells that were identified as glutamatergic neurons in two punches, C5 (inner cortex) and B5 (outer cortex), that were immediately adjacent with 83 and 65 total cells and 42 and 35 glutamatergic cells, respectively. Similar to the single punch analysis, we produced a counts matrix including only these cells and used the full set of peaks to perform topic analysis and visualization using UMAP, which showed a clear separation between the two locations (Fig. 12f). We then assessed global motif accessibility, which revealed clear enrichment for motifs associated with upper or lower cortical layers, including RORB, enriched in the outer cortex, and TBR1, enriched in the inner cortex (Fig. 12g). To systematically assess this spatial TF motif enrichment (TFME), we applied this same analysis to the glutamatergic cell populations identified in every pair of inner and outer cortical punches. This produced a consistent pattern with very few punch pairs deviating from the expected enrichment pattern (Fig. 12h).

2.5.5 Spatial trajectories of single-cell ATAC-seq in the human cortex

With the ability to probe spatial single-cell chromatin accessibility established in the mouse cortical lamination experiment, we next deployed sciMAP-ATAC on human brain tissue to profile lamination in the adult primary visual cortex (VISp) using an equivalent voxel-diameter resolution of 215 cubic microns. Samples of human VISp tissue were obtained from an adult (60-year male) with no known neurodegenerative disorders at 5.5 hours postmortem. Samples were oriented and flash-frozen in TFM prior to storage at −80 °C. The sample was cryosectioned using the same alternating thick (200 μm) and thin (20 μm) pattern as previously described. We designed and implemented a 250 μm diameter punch schematic across three adjacent 200 μm sections to produce 21 distinct trajectories comprised of eight punches spanning the cortex, with an additional 20

44

punches distributed in the subcortical white matter for a total of 188 spatially mapped

tissue punches (Fig. 13a,b). In total, 4547 cells passed quality filters with a mean of

30,212 reads per cell (estimated mean of 98,274 passing reads per cell with additional

sequencing; see 2.7 Methods, Figs. 9a and 14a), a mean TSS enrichment of 15.80—

more than twice the "ideal" ENCODE standard for bulk ATAC-seq datasets (>7, GRCh38

RefSeq annotation), a FRiS of 0.45 using a human reference dataset[95], and prominent

nucleosomal banding (see 2.7 Methods, Fig. 9b, c, e).
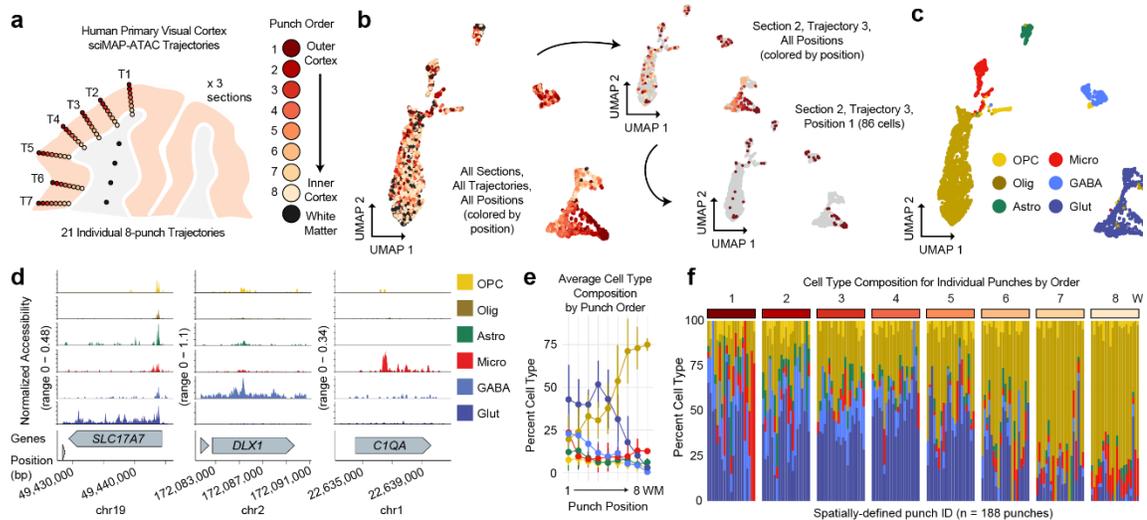
Figure 13. sciMAP-ATAC trajectories through the human primary visual cortex.
a, sciMAP-ATAC punching schematic showing one of three adjacent sections from one
individual. A total of 21 eight-punch trajectories (T) spanning the cortex were
produced. **b,** UMAP of cells colored by position within their respective trajectory as in **a**. Top
right shows the same UMAP with all cells grayed out with the exception of cells from the third
trajectory from section 2. Bottom right shows all cells grayed out with the exception of cells
from a single punch; the outermost cortical position (1) from the third trajectory of the second
section. **c,** UMAP as in **b** colored by cell type (OPC oligodendrocyte precursor cells, Olig
oligodendrocytes, Astro astrocytes, Micro microglia, GABA GABAergic (inhibitory) neurons,
Glut glutamatergic (excitatory) neurons). **d,** ATAC-seq profiles for cells aggregated by cell type
for marker genes; colored by cell type as in **c**. **e,** Aggregate cell type composition across the 21
trajectories (*n* = 4547 cells over 188 independent punches); colored by cell type as in **d**. Data
are presented as mean values ± SD. **f,** Cell type composition for each of the 188 individual
punches split by trajectory position. Punches from the WM indicated by an asterisk are
aggregated by section. Colored by cell type as in **d**. Source data are provided as a Source
data file.

Cell profiles were generated as described in prior experiments, which resulted in six

distinct clusters representing the major cell types (Fig. 13c, d). Similar to the murine

cortex, glutamatergic neurons exhibited the most distinct spatial patterning with a clear gradient spanning cortical trajectories (Fig. 13a–c), which was also determined to be the most significant (Moran's I test Bonferroni corrected p-value = 0.87 × 10−4, see: 2.7 Methods, Supplementary Table 1). Further sub-clustering of GABAergic interneurons revealed minimal spatial bias across four distinct subtypes comprised of two MGE-derived and two CGE-derived clusters (Fig. 14b–e). Each of the 21 individual trajectories through the cortex produced similar distributions of cells through UMAP projections with a lack of glutamatergic neurons present in the punches obtained from subcortical white matter (Supplementary Data 2). Our astrocyte to neuron ratio (0.15:1) was low yet comparable to the previously published snRNA-seq of the human VISp (0.12:1)[96]. Average cell type composition along these trajectories revealed the expected pattern of an increased proportion of oligodendrocytes and decreased glutamatergic neuron abundance as the trajectory approached or entered the subcortical white matter region (Fig. 13e). Individual punches largely matched the corresponding average position profile (1–8, WM), with higher variability at the first punch where some trajectories overlapped the pial surface of the cortex (Fig. 13f).
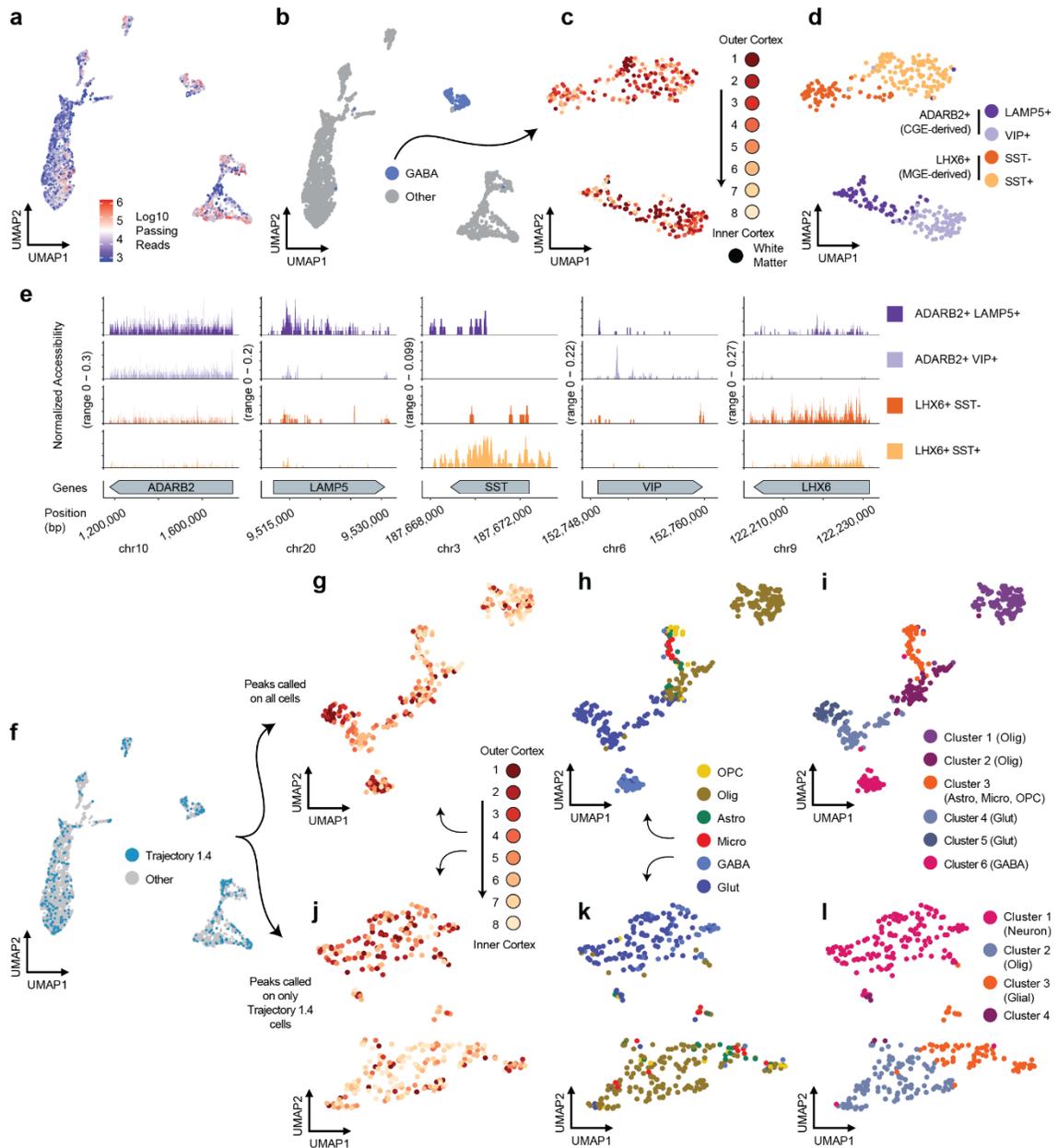
Figure 14. Extended analysis of the human primary visual cortex sciMAP-ATAC dataset.
a, UMAP of all cells from the experiment colored by log10 passing read counts. **b,** UMAP of the full dataset with all cells grayed out except for those identified as GABAergic neurons. **c,** UMAP of GABAergic neurons analyzed using topic modeling individually colored by punch position. **d,** Four interneuron clusters identified, including two MGE-derived and two CGE-derived cell types. **e,** Aggregate ATAC-seq profiles for marker genes for each of the interneuron cell types. **f,** UMAP of the full dataset with all cells grayed out except for those belonging to the fourth trajectory, of 8 consecutive punches, on the first section (Trajectory 1.4). **g,** UMAP of cells from Trajectory 1.4 that were processed using peaks from the full VISp dataset colored by the punch position; **h,** the cell type classification as determined from the full dataset; and **i,** the six clusters that were identified. **j,** UMAP of cells from Trajectory 1.4 that were processed using peaks called using only those cells, colored by the punch position; **k,** the cell type classification as determined from the full dataset; and **l,** the four clusters that were identified.

47

2.5.6 Integration of sciMAP-ATAC with scTHS-seq and snRNA-seq reveals epigenetic spatial patterning concordant with transcriptional neuronal subtypes

Previously, Lake et al. produced single-cell transposase hypersensitivity (scTHS-seq, an assay for chromatin accessibility similar to ATAC-seq) and single-nucleus RNA-seq from the human VISp[96]. We integrated our sciMAP-ATAC dataset with each of these using Seurat[97] and visualized the joint UMAP projections with cell type information, along with the positional breakdown of glutamatergic neurons (Fig. 15 a,b). The joint manifold for each integration largely agreed, with the exception of a population of cells in our sciMAP-ATAC dataset that did not co-embed with any cell types present in the snRNA-seq dataset. These cells represent all of the cell types called within the sciMAP-ATAC dataset and cluster clearly with their cell types in the sciMAP-ATAC analysis on its own, suggesting that it may be an effect of the gene activity score intermediate that is used for co-embedding with ATAC-based data (see 2.7 Methods).
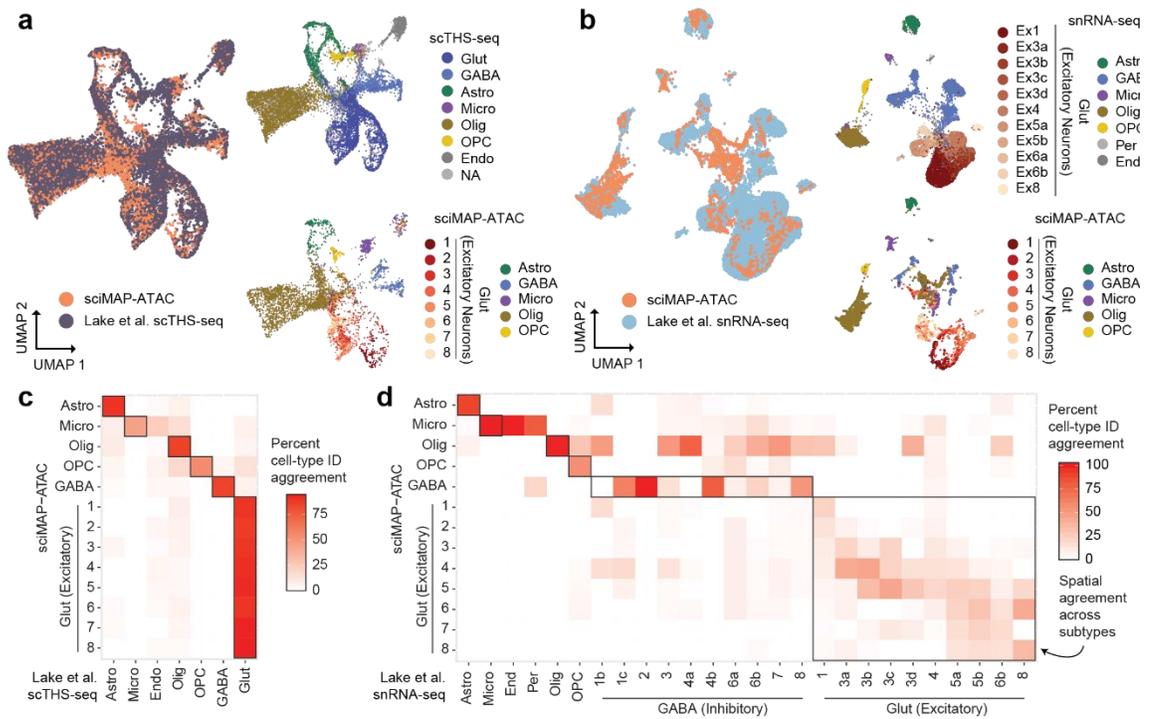
Figure 15. Integration of sciMAP-ATAC with snRNA-seq and scTHS-seq human VISp datasets. a, Co-embedding of sciMAP-ATAC and scTHS-seq cell profiles from Lake et al.[29] using Signac[75] in a joint UMAP. Top right shows only scTHS-seq cells colored by cell type identified in Lake et al.[29] and bottom shows sciMAP-ATAC cells colored by our called cell types as in Fig. 13c, except for glutamatergic neurons which are colored by spatial positions 1–8 (Glut glutamatergic (excitatory) neurons, GABA GABAergic (inhibitory) neurons, Astro astrocytes, Micro microglia, Olig oligodendrocytes, OPC oligodendrocyte precursor cells, Endo endothelial cells, NA not applicable—no cell type provided). **b,** Co-embedding of sciMAP-ATAC and snRNA-seq transcriptional profiles from Lake et al.[29] using Signac. Top right shows only snRNA-seq cells. Abbreviations as in **a**, but with the addition of Per = pericytes, and glutamatergic (excitatory) neurons (Ex) are colored by subtype identified in Lake et al.[29]. Bottom right shows only sciMAP-ATAC cells, with glutamatergic neurons colored by spatial position 1–8. **c,** Confusion matrix representing the percent agreement in predicting the cell type of a cell from one dataset using the other between sciMAP-ATAC and scTHS-seq cells. **d,** As in **c**, but between sciMAP-ATAC and snRNA-seq. Spatial agreement between excitatory neuron subtypes identified in the snRNA-seq data correspond to the spatial positioning of cells within our sciMAP-ATAC dataset. Source data are provided as a Source data file.

To directly assess the performance of the dataset integration, we used the joint manifold to perform cell-type label transfer, effectively using one assay's cell type identities to predict the others, and compared the overlap in the form of a confusion matrix. For the scTHS-seq integration, the top concordance was between the two corresponding cell types in nearly every case, including across all eight of the spatial glutamatergic neuron cell sets within the sciMAP-ATAC dataset that all corresponded to the single

49

glutamatergic cell type in the scTHS-seq dataset (Fig. 15c). One exception was the association of a subset of microglia within the sciMAP-ATAC dataset with the endothelial cell population identified in the scTHS-seq dataset, which is a population we did not define. This suggests that a portion of our cells identified as microglia are likely endothelial cells. Integration with snRNA-seq data also produced concordance for the majority of cell types (Fig. 15d), with the exception of a group of cells spanning all cell types that did not co-embed as cleanly and thus projected into the center of the UMAP. The snRNA-seq data provided in Lake et al. includes a more granular breakdown of glutamatergic neurons when compared to the single classification provided for scTHS-seq cells. Within the confusion matrix where cell types were predicted across modalities, we observed a clear spatial progression that corresponded to the subtypes of glutamatergic neurons identified by snRNA-seq, which Lake et al. previously identified as being enriched for layer-specific transcripts. The concordance between these subtypes and our spatial assignments confirms that sciMAP-ATAC spatially registers biological features of single cells from structured tissue.

2.5.7 Spatial excitatory neuron epigenetic patterning at the individual trajectory level

Using our cell-type assignments, we isolated all human VISp glutamatergic neurons and split them by position along their respective trajectories (Fig. 16a and Supplementary Data 3). We examined ATAC signal at layer-specific marker genes broken down by each spatially distinct category, which revealed increased accessibility at genes associated with outer cortical layers within the outer cortical punches and vice versa (Fig. 16b). We next selected all cells from the centermost trajectory of section 1 (T1.4, n = 358 cells) and performed an isolated analysis using peaks called on the full dataset for topic analysis, cluster identification, and visualization with UMAP (Fig. 16c and Fig. 14f, i). A

clear separation was observed between major cell types across six clusters, with two

distinct clusters of oligodendrocytes, two clusters of glutamatergic neurons, one cluster

comprised of GABAergic neurons, and finally, a cluster made up of all other cell types

(astrocytes, endothelial, and OPCs). When performing the analysis in isolation using

only T1.4 cells for peak calling, we identified 16,493 peaks that were used for

subsequent analysis to produce four clusters with notably less cell type separation than

when leveraging the set of peaks from the full dataset (Fig. 14j,l). The first cluster was

comprised of both glutamatergic and GABAergic neurons, the second was primarily

oligodendrocytes, the third included oligodendrocytes, as well as the majority of cells

from all other nonneuronal cell types, with the fourth cluster comprised of only a handful

of cells with no dominant cell type. In line with the previous assessment of a single

punch from the mouse SSp, cell type separation can be distinct for major cell types

when leveraging larger peak sets than the limited number that can be called on small

cell count datasets. This supports the assertion that computational improvements to

enable peak calling on low cell count datasets can substantially boost analytical power[98].
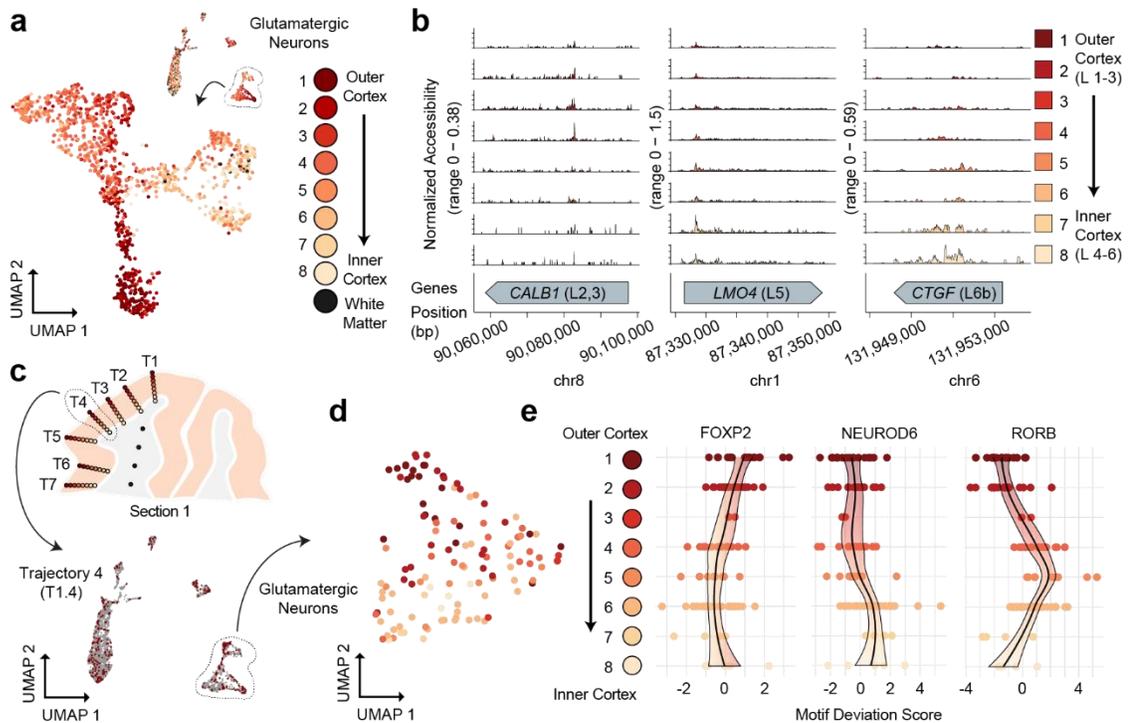
Figure 16. sciMAP-ATAC shows spatial epigenetic patterns of glutamatergic neurons.
a, Isolation and UMAP visualization of human VISp glutamatergic neurons from all cells (top right), colored by punch position. An interactive, three-dimensional UMAP embedding is available as Supplementary Data 4. **b,** ATAC-seq profiles for glutamatergic neurons along trajectory positions for layer (L)-specific marker genes CALB1 (layers 2 and 3), LMO4 (layer 5), and CTGF (layer 6b); colored by punch position as in **a**. **c,** Cells from section 1, Trajectory 4 (T1.4, top) are shown in color on the UMAP of all cells, with other cells shown in gray (bottom); colored by position as in **a**. **d,** UMAP of glutamatergic neurons from Trajectory 1.4 after topic modeling on the isolated cells; colored by position along the trajectory as in **a**. **e,** DNA-binding motif enrichment for layer-specific factors for Trajectory 1.4 shown in **d**, with cells split by their positions along the trajectory. Source data are provided as a Source data file.

Finally, we isolated only cells determined to be glutamatergic neurons based on the full dataset cell type assignment within Trajectory 1.4 (n = 121 cells). We assessed these cells again using the full peak set through the same analysis workflow (see 2.7 Methods). As in the UMAP projections on cells from the full experiment, these cells were positioned along a gradient that reflected their position along the trajectory (Fig. 16d). We then assessed the global accessibility of DNA-binding motifs that captured spatially distinct enrichments through the trajectory, reflecting the expected pattern of transcription factor (TF) activities through cortical layers (Fig. 16e). This included enrichment for FOXP2 motif accessibility in the outer cortical layers, slightly increased

52

accessibility for NEUROD6 toward the inner cortex, and increased accessibility for RORB motifs in punches 4–6 along the trajectory, corresponding to canonical cortical layer 4 RORB expression. Taken together, sciMAP-ATAC is capable of producing high-quality single-cell ATAC-seq profiles from human postmortem tissue with a spatial resolution capable of identifying the major components of cortical lamination, with the capability to characterize a single spatial trajectory through the cortex.

2.5.8 sciMAP-ATAC in a mouse model of cerebral ischemia

Cerebral ischemia produces a complex spatially progressive phenotype with extensive tissue alterations and shifts in cell type abundance and epigenetic states[99–104]. Cerebral ischemic infarction induces gliosis, a process in which glia in the surrounding tissue enter reactive states that are potentially aimed at restoring tissue homeostasis but can involve the loss of normal function (or adoption of a damaging function) and form a glial scar. Many components involved in the ischemic cascade are well studied, including factors that promote postischemic inflammation (e.g., IRF1, NF-kB, ATF2, STAT3, EGR1, and CEBPB) and prevent postischemic inflammation and neuronal damage (e.g., HIF-1, CREB, C-FOS, PPARα, PPARγ, and P53)[105]. Reactive gliosis can be characterized by increased GFAP expression in astrocytes and increased IBA1 in microglia. Myelination depletion is a hallmark of cerebral ischemic injury due to acute oligodendrocyte cell death and impaired OPC differentiation[106,107]. Far less is known, however, about glial cell state transitions in the area surrounding ischemic infarction in the brain. We reasoned that our sciMAP-ATAC technology could reveal, with cell type and spatial specificity, the epigenetic alterations that occur to accompany and/or drive the ischemic cascade and postischemic pathology.

To accomplish this, we used a transient middle cerebral artery occlusion (MCAO) mouse model of ischemic injury with reperfusion (see 2.7 Methods; Fig. 17a). Each ischemic

(n = 2 animals) and naive (n = 3) brain was flash-frozen 3 days after surgery, embedded in TFM, sectioned, alternating between 200 μm for sciMAP-ATAC and 20 μm for IHC for IBA1 (microglia), GFAP (astrocytes, Fig. 17b), and counterstained using DAPI. We used these images to define the infarct area by the absence of GFAP-positive astrocytes while being surrounded by reactive astrocytes exhibiting increased GFAP signal at the infarct border (Fig. 18a). We next defined two axes for targeting the sciMAP-ATAC punches, the first progressing from the pial surface of the cortex to the striatum, all within the infarct core (punch position axis 1–4), and the second progressing from the infarct core toward the infarct border (punch position axis 5–8). GFAP immunolabeling was absent in the infarct core (punch positions 5–7) but increased at the infarct border in punch position 8, recapitulating known features of glial scar formation surrounding the infarct area. We then performed sciMAP-ATAC on the 200 μm sections along each axis to produce 5081 cells with a mean passing reads per cell of 33,832 (estimated mean passing reads per cell of 225,670 with further sequencing) and a mean of 26.6 high-quality cell profiles per punch (Figs. 9a, f and 18b). TSS enrichment for this preparation was notably lower than previous preparations ranging from 5.05 (stroke hemisphere) to 7.50 (naive brain), which we suspect is due to several factors (Fig. 9e). The first is that the stroke hemisphere contained many dead or dying cells that exhibit reduced ATAC signal, which we describe in more detail below, and the second is that these sections were stored for >3 months prior to sciMAP-ATAC processing, suggesting that long-term storage of sections may result in a reduction in data quality. Despite the reduced TSS enrichment and comparably lower FRiS (0.79–0.82; Fig. 9b), we called 140,772 accessible genomic loci that were used in subsequent analysis.
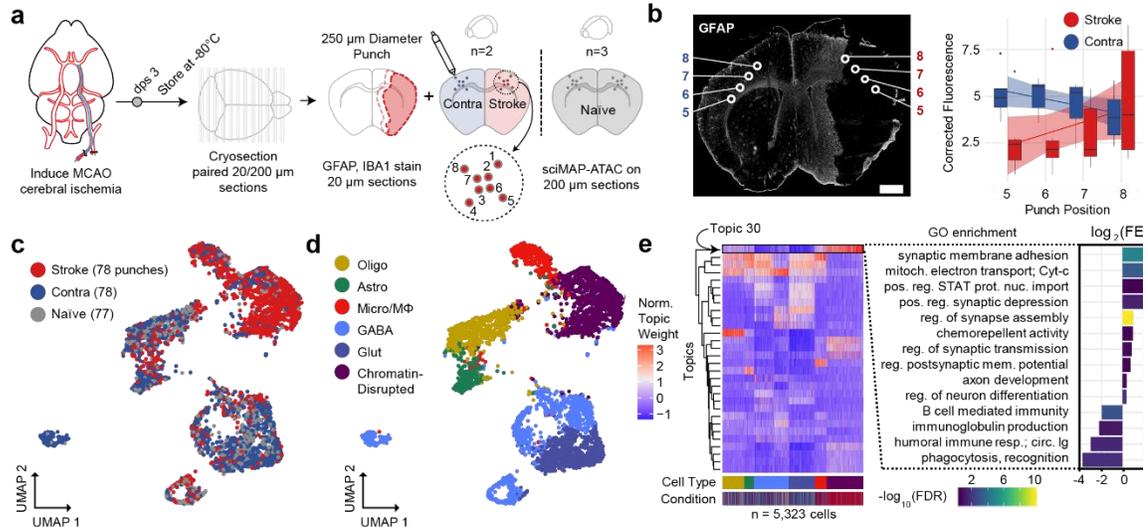
Figure 17. sciMAP-ATAC applied to a mouse model of ischemic injury. a, Experimental design using a mouse MCAO model of ischemic injury. Mice were sacrificed 3 days post-surgery (dps) and brains flash-frozen in TFM. Alternating thin (20 µm) and thick (200 µm) sections were processed using IHC to define infarction (red outline) and peri-infarct area (pink outline) and sciMAP-ATAC punching schematic, respectively. **b,** GFAP IHC of a 20 µm coronal section of an ischemic mouse brain. Punch positions along the 5–8 axis (core-to-border) are indicated. Background corrected GFAP fluorescence along the 5–8 axis is shown to the right for stroke and contralateral hemispheres (n = 10). Data are presented as linear fitted model ± SEM; boxplot center line represents median, lower and upper hinges represent first and third quartiles, and whiskers extend from hinge to ±1.5 × IQR, (scale bar, 1 mm) **c,** UMAP of cells colored by the three conditions. **d,** UMAP as in **c**, colored by clusters assigned to cell type (Olig oligodendrocytes, Astro astrocytes, Micro/MΦ microglia/macrophage, GABA GABAergic (inhibitory) neurons, Glut glutamatergic (excitatory) neurons). **e,** Cell × topic matrix colored by normalized topic weights, as in **c**, **d** and annotated by conditions and cell type as given at the bottom reveals substantially divergent topic weighting in cells from the stroke punches (left). Topic 30, enriched specifically in the stroke cells belonging to the chromatin-disrupted cluster, has peaks enriched for ontologies associated with ischemic injury with reperfusion. Colored by −$\log_{10}$ false discovery rate (FDR) Q-value, height by $\log_2$ fold enrichment (right). Source data are provided as a Source data file.

We performed topic modeling, followed by clustering, cell type identification, and

visualization of the cell × topic matrix (Fig. 17c–e), which revealed comparable cell-type

proportions across biological samples with exceptions for microglia/macrophages and a

chromatin-disrupted cluster that were highly enriched within the infarct. We profiled cell

type proportions along both of the axes (Fig. 18c); however, the pial to striatum axis

(punch positions 1–4) in stroke hemisphere samples is completely within the infarct core.

In contrast, the infarct core-to-border axis (punch positions 5–8) progresses from the

center of the infarct to the glial scar along the infarct border, capturing a transition zone

of reactive gliosis, and is the spatial trajectory that we focus on in our subsequent

analysis.
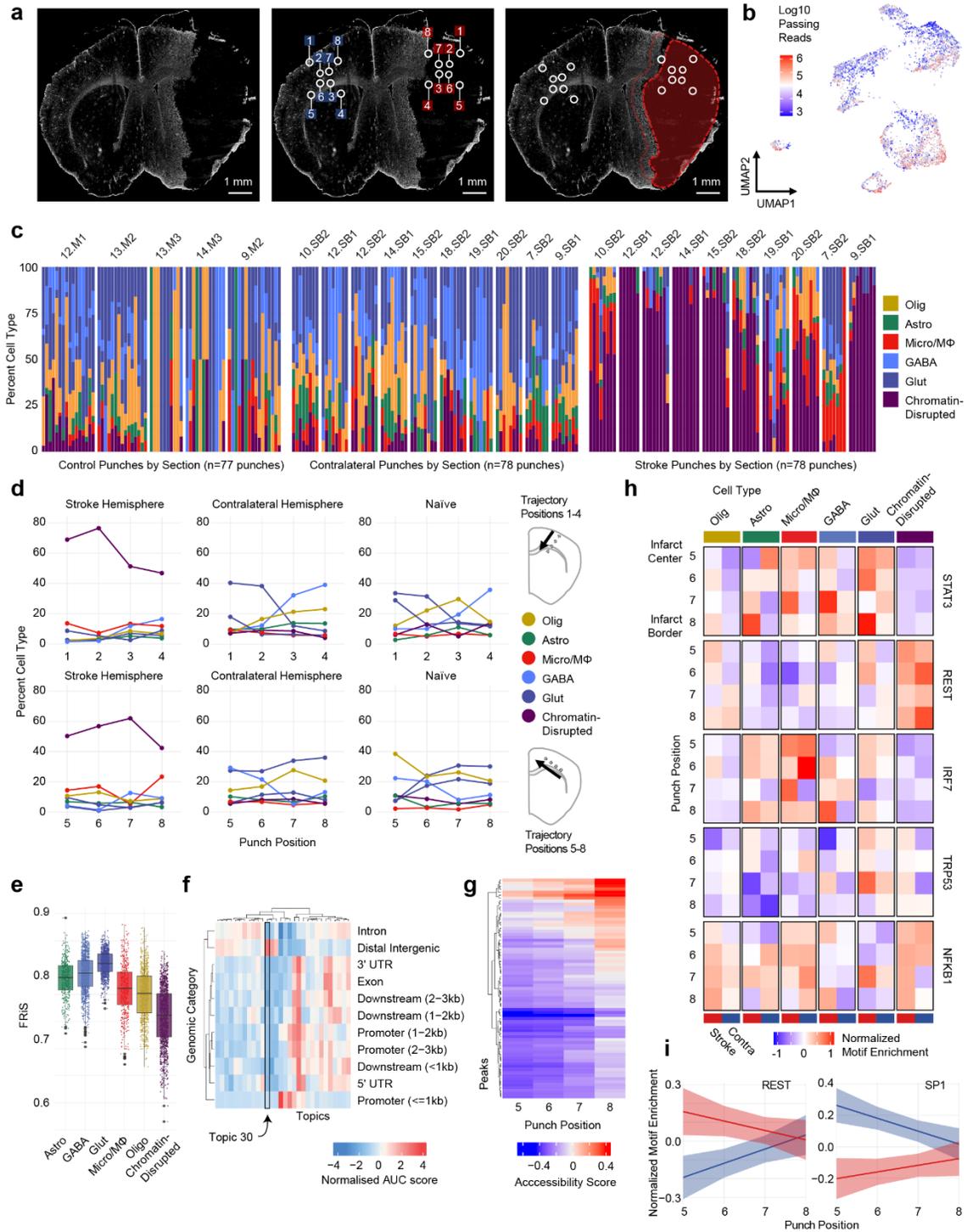
Figure 18. **Extended analysis of the cerebral ischemia sciMAP-ATAC dataset. a,** Representative GFAP immunostaining of a histological section from a stroke brain (left), with punch positions and labels shown (middle), and punch positions with the stroke region overlaid in red (right). **b,** UMAP of cells from the cerebral ischemia experiment colored by the log10 passing read counts. **c,** Cell type composition for each punch in the experiment grouped by individual section and more broadly by category; colored by cell type as defined in Fig. 17d. **d,** Aggregated cell type composition for the 1-4 axis (top) and 5-8 axis (bottom) split by category of tissue; colored by cell type as defined in Fig. 17d. **e,** FRiS values for cells split by cell type indicating a substantial decrease in FRiS for the chromatin-disrupted cluster (n = 5,081 cells); colored by cell type as defined in Fig. 17d. Center line represents median, lower and upper hinges represent first and third quartiles, whiskers extend from hinge to ± 1.5x IQR, individual cells represented as colored dots. **f,** Enrichment for topics with respect to genomic category showing that Topic 30, which is elevated in cells within the chromatin-disrupted cluster, is enriched for distal intergenic regions – further supporting a global laxing of chromatin, likely due to cell death. **g,** Regulatory elements that change significantly and uniformly along the 5-8 axis. *P* value of the two-way ANOVA from the interaction of regulatory element site enrichment per punch by condition without multiple comparison correction ($p < 0.05$). **h,** Motif enrichment along the 5-8 axis for stroke and contralateral hemispheres split by cell type as defined in Fig. 17d. **I,** REST and SP1 normalized motif enrichment along the 5-8 axis shows opposite trends between the two factors as well as for each factor between the stroke and contra hemispheres. Data are presented as linear fitted model ± SEM. Source data are provided as a Source Data file.

Along with this progression, we found that the stroke hemisphere had diminished neural cell types (depletion of glutamatergic and GABAergic neurons, oligodendrocytes, and astrocytes), as well as a progressive increase in cells within a cluster exhibiting globally disrupted chromatin structure up to punch position 7 and a drop at punch position 8 upon entering the infarct border (Fig. 18d). This state is predominantly characterized by globally increased chromatin accessibility, with a decrease in TSS enrichment, a decrease in FRiS, and an increase in reads falling within distal intergenic regions, which is likely caused by cell death (Fig. 18e, f). In addition to the global effects on chromatin structure, the chromatin-disrupted cell population also showed strong enrichment in one of the topics (Topic 30; Fig. 17e, left). A gene ontology (GO) enrichment analysis of the peaks that define topic 30 revealed that cells within the ischemic hemisphere undergo a chromatin state shift as a result of the ischemic cascade, which leads to enrichment for processes canonically associated with ischemia (Fig. 17e, right). Most notably, positive regulation of synaptic membrane adhesion, synaptic depression, assembly, transmission, and membrane potential were all enriched in ischemia-derived cells,

indicating that CNS synaptogenesis is upregulated in a subset of cells 3 days post ischemia[108,109]. In addition, while the percentage of microglia increased in the stroke condition (13.2%) as compared to contralateral (6.7%) and naive (4.3%), depletion of immune response processes (B-cell-mediated immunity, humoral immune response mediated by circulating immunoglobulins) was seen in ischemia-derived cells. This recapitulates previous findings that acute ischemic immune response is followed by post-stroke immunodepression and dysregulation[110,111].

2.5.9 Spatially progressive chromatin features in cerebral ischemia

To directly characterize the relationship between space and epigenetic state in cerebral ischemia, we assessed TF DNA-binding motif enrichments for each cell and performed a regression for all cells across the infarct core-to-border axis (punch positions 5–8) in the stroke and contralateral hemispheres. We used the difference between linear model coefficients for paired affected (stroke) and unaffected (contralateral) hemispheres along with the significance of the hemisphere motif enrichment differences to identify TFs that undergo spatially progressive regulatory changes (see 2.7 Methods). In total, we identified 95 TF motifs that were significantly altered with a spatial component, many of which have been previously reported as key factors identified in cerebral ischemia (Fig. 19a, b). KLF9, a member of the Kruppel-like factor family, demonstrated the most significant increase in accessibility with proximity to the peri-infarct area. The 17 KLF family TFs are key factors in neuronal development, plasticity, and axon regeneration and are ubiquitously expressed in the CNS. Several KLF family members, namely KLF2, 4, 5, 6, and 11, have been specifically linked to cerebral ischemia pathogenesis[112,113]. Notably, KLF2 and KLF11 have been shown to contribute to the protection of the blood-brain barrier in cerebral ischemia[114–116]. However, as DNA-binding motifs within the KLF family are similar, members of the KLF family other than KLF9 may be driving this motif

accessibility change. Finally, we assessed the accessibility of individual elements and identified 73 accessible chromatin sites that varied significantly through the 5–8 axis of spatial progression (see 2.7 Methods; Fig. 18g).
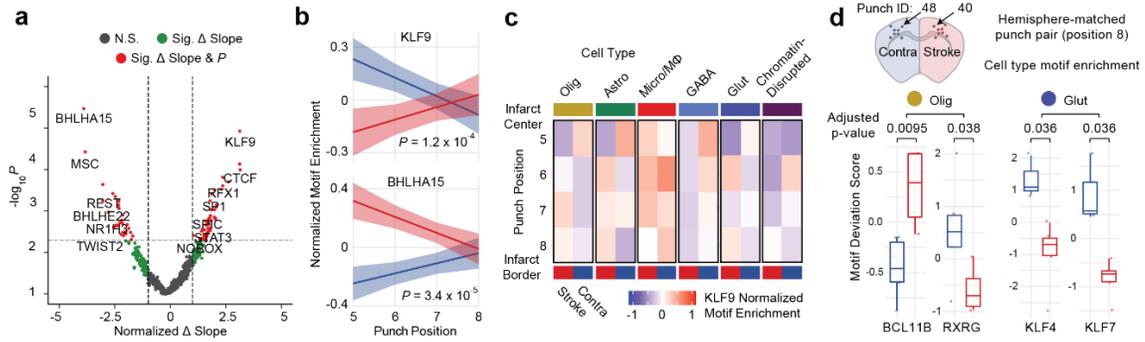
Figure 19. Spatially progressive epigenetic remodeling in ischemic injury. a, Volcano plot of Z-scored transcription factor (TF) motif enrichment slope change across punch positions 5–8 (Δslope = slope$_{stroke}$ − slope$_{contralateral}$) by −log10 p value of the two-way ANOVA from the interaction of TF motif enrichment per punch by condition (stroke, contralateral) without multiple comparison correction. Colored by significance (N.S. not significant, Sig. Δ slope = significant change in slope, Sig. Δ Slope and p = significant change in slope and significant p value). **b,** Top hits for significantly different changes in TF motif enrichment over space as compared between stroke (red) and contralateral (blue); KLF9 (top) and BHLHA15 (bottom). −Log10 p value of the two-way ANOVA from the interaction of TF motif enrichment per punch by condition without multiple comparison correction. Data are presented as linear fitted model ± SEM. **c,** KLF9 TF motif enrichment over space reveals cell type contribution to KLF9 enrichment from infarct core to peri-infarct area. Cell types as defined in Fig. 17d. **d,** Comparison of TF motif enrichment at the infarct border (punch position 8) between stroke (punch 40) and contralateral (punch 48) single-cell profiles. Oligodendrocyte (Olig) TF motif enrichment shown for BCL11B and RXRG for punch 40 (n = 4 cells) and punch 48 (n = 6 cells). Glutamatergic neuron (Glut) TF motif enrichment shown for KLF4 and KLF7 for punch 40 (n = 5 cells) and punch 48 (n = 3 cells). Two-sided Mann–Whitney U test with Bonferroni–Holm correction. Center line represents median, lower and upper hinges represent first and third quartiles, whiskers extend from hinge to ±1.5 × IQR, individual cells represented as colored dots. Source data are provided as a Source data file.

We next explored the cell-type specificity of the KLF9 motif accessibility changes (Fig. 19c). In the stroke hemisphere chromatin-disrupted cell subset, we observed a reduction in KLF9 motif accessibility in all punch positions except punch position 8, at the infarct border, with all cell types other than microglia showing a reduction in accessibility at the center of the infarct core (punch position 5). Uniquely, microglia are largely unaffected and have comparable KLF9 TF-binding motif enrichment at the infarct core in comparison to the contralateral hemisphere. In addition to KLF9, we also identified

STAT3 as varying significantly over space (Fig. 18h), which was also an enriched GO

term in stroke cells (Fig. 17e). STAT3 has been extensively studied in the JAK/STAT3

pathway, which is a key regulator of apoptosis in cerebral ischemia injuries with

reperfusion[117], as well as an initiator of reactive astrogliosis under diverse conditions[118].

Accordingly, we found that STAT3 was largely absent from astrocytes in punches

positions 5–7, but was enriched in the reactive astrocytes at the infarct border zone at

punch position 8. In contrast, we find that the RE1-silencing factor (REST) is significantly

elevated at the ischemic core and decreases with proximity to the infarct border.

Accordingly, REST has been shown to form a histone deacetylase complex that is a

director repressor of SP1 in cerebral ischemia, a TF we identify as varying significantly

over space, in the opposite direction of REST[104] (Fig. 18i).

Finally, we sought to characterize chromatin accessibility profiles of cells isolated from a

single punch at the glial scar (Fig. 19d). To do this, we isolated two punches (punch 40

and punch 48), both originating from the same section (15.SB2), from punch position 8

of the stroke (punch 40) and contralateral hemisphere (punch 48). We processed the

cells in isolation as described in prior individual punch analyses, using the peak set from

the full experiment. We performed DNA-binding motif enrichment analysis across all

cells[119] and then performed cell-type-specific comparisons for a glial (oligodendrocyte)

and neuronal (glutamatergic neuron) cell type. In oligodendrocytes, 56 TF motifs were

significantly different between the stroke and contralateral hemisphere, many of which

(44; 78.6%) corresponded to higher enrichment in stroke as compared to contralateral.

Specifically, we found BCL11B (CTIP2), a negative regulator of glial progenitor cell

differentiation, to be significantly increased at the glial scar (Fig. 19d, left)[120]. Conversely,

we found RXRG, a positive regulator of OPC differentiation and remyelination, to be

significantly depleted (Fig. 19d, left)[121]. Together these findings indicate an impaired

ability of OPCs to differentiate into mature oligodendrocytes at the glial scar. In glutamatergic neurons, we found neuron-associated TFs such as NEUROD2 to be significantly depleted in the stroke hemisphere, which corresponds with decreased neuronal cell types at punch position 8 in the stroke hemisphere. In accordance with our infarct core-to-border axis (punch positions 5–8) analysis, we found that seven of the KLF family of TFs (KLF2–4, 6–8, and 12) were significantly depleted in glutamatergic neurons at the glial scar in the stroke hemisphere (Fig. 19d, right; KLF4 and KLF7 shown). Interestingly, previous studies have found that in response to cerebral ischemia, KLF4, 5, and 6 are induced in astrocytes, while KLF2 is depleted in endothelia and induced in microglia[122]. With these data, we identify that motif enrichment for many members of the KLF family not only significantly varies over space across all cell types we also indicate novel depletion of multiple KLFs, specifically in glutamatergic neurons at the ischemic glial scar.

2.6 Discussion

sciMAP-ATAC provides a low-cost, highly scalable, hypothesis-independent approach to acquiring spatially resolved epigenomic single-cell data with the use of immediately available commercial tools. In addition, sciMAP-ATAC is translatable to any tissue, culture, or model system compatible with cryosectioning. While many methods rely on signal-to-noise optical detection of densely packed molecules and computationally intensive spatial reconstruction, sciMAP-ATAC encodes nuclear localization directly into each library molecule, allowing for rapid subsetting of cells by localization and mapping of cells across vector space in 3D between adjacent sections. We demonstrate the use of sciMAP-ATAC to profile the murine somatosensory cortex, as well as multi-punch trajectories through the human primary visual cortex, recapitulating known marker gene progression through cortical layering and cell type composition based on the category

and positioning of spatially registered microbiopsy punches. We further show the utility

of sciMAP-ATAC to resolve the progressive epigenomic changes in a cerebral ischemia

model system, revealing distinct trends in chromatin accessibility, cell-type composition,

and cell states along the axes of tissue damage and altered morphology. Application of

sciMAP-ATAC to other highly structured systems or tissues with a gradient of disease

phenotype will be particularly valuable areas for this technology. The primary limitation of

sciMAP-ATAC is that punches are currently performed manually and registered with

adjacent imaged sections post-punching. This limits the precision of desired punch

positions, as well as throughput; however, automated processing of tissue sections

using robotics[123], where punch patterns are designed on adjacent imaged sections and

registered to the target section, will enable high precision, as well as increased

throughput into the range of thousands. Furthermore, as spatial transcriptomic

technologies evolve, they may enable the acquisition of chromatin accessibility

information; however, substantial technical hurdles must first be overcome, and profiles

produced would be in aggregate over the feature size and not necessarily single cell.

Finally, here we applied the sciMAP strategy to assess chromatin accessibility; however,

it can, in theory, be applied to any single-cell combinatorial indexing technique to enable

spatially registered single-cell genome[86], transcriptome[124], chromatin folding[125],

methylation[126], or multi-omic[24,26,127] assays.

2.7 Methods

2.7.1 Mouse brain and human VISp cortex sample preparation

A step-by-step protocol describing the sciMAP-ATAC methods can be found at

Protocols.io. All animal studies were approved by the Oregon Health and Science

University Institutional Animal Care and Use Committee. Male C57Bl/6 J mice aged 8

weeks were purchased from Jackson Laboratories for the mouse whole brain sciATAC,

punch dissociation development sciMAP-ATAC, and mouse SSp cortex sciMAP-ATAC experiments. All mouse cages were kept on a 12 h light/dark cycle at a temperature of 70 °F and within a humidity range of 30–70%. Animals were sacrificed by carbon dioxide primary euthanasia and cervical dislocation secondary euthanasia. Animals were immediately decapitated, intact brain tissue was harvested, washed in ice-cold phosphate-buffered saline (PBS; pH 7.4), submerged in TFM (Cat. TFM-C) within a disposable embedding mold (Cat. EMS 70183). Human VISp cortex samples were provided by the Oregon Brain Bank 5.5 h postmortem and were submerged in TFM. The use of human samples in this study falls under the NIH defined Exempt Human Subjects Research, under Exemption 4 (https://humansubjects.nih.gov/). Embedded mouse whole brain and human VISp cortex samples were flash-frozen in liquid nitrogen-cooled isopentane by lowering the sample into the isopentane bath without submerging within 5 min of embedding. Samples were immediately transferred to dry ice, paraffin wrapped to delay sample dehydration, and stored in an airtight container at −80 °C.

2.7.2 Mouse cerebral ischemia model

Two C57BL/6 9-week-old (P63) female mice were placed under isoflurane anesthesia (5% induction, 1.5% maintenance) in 30% oxygen-enriched air. Body temperature was maintained at 37 ± 0.5 °C throughout the procedure. Middle cerebral artery (MCA) occlusion was performed using a previously described method by Longa et al. with slight modifications[128]. Briefly, a laser Doppler flowmeter (Moore Instruments) probe was affixed over the right parietal bone overlying the MCA territory to monitor changes in cerebral blood flow. A midline incision was made, the right common carotid artery (RCCA) bifurcation was exposed by gentle dissection, and the external carotid artery (ECA) was permanently ligated distal to the occipital artery using electrocautery, such that a short ECA stump remained attached to the bifurcation. The RCCA and internal

carotid arteries (ICA) were temporarily closed with reversible slip knots before an
arteriotomy was made in the ECA stump. A silicone-coated 6.0 nylon monofilament was
inserted into the ICA via the arteriotomy and gently advanced to the ICA/MCA bifurcation
to occlude CBF to the MCA territory, and confirmed by a laser Doppler signal drop of
<30% of baseline. After 60 min occlusion, the filament was gently retracted, the ECA
permanently ligated, the slip knot of the common carotid artery removed, and the
incision sites sutured closed. The mice exposed to MCAO were euthanized 3 days after
the MCAO procedure, intact brain tissue harvested, washed in ice-cold PBS (pH 7.4),
submerged in TFM, and flash-frozen in liquid nitrogen-cooled isopentane. Samples were
paraffin wrapped and stored at −80 °C, and intact embedded whole mouse brains were
sectioned at the time of the experiment.

2.7.3 Sample sectioning

All embedded samples were sectioned in a cryostat (Leica CM3050) at −11 °C chuck
and chamber temperature and collected on Superfrost Plus microscope slides
(Fisherbrand, Cat. 22-037-246). Sectioning was performed in sets of one section at 100–
300 µm paired with three sections at 20 µm to generate sets of four slides consisting of
microbiopsy (1) and histology (3) sections at one section per slide. Slide boxes were
sealed with paraffin to prevent sample dehydration and stored long term at −80 °C.

2.7.4 Mouse whole brain coronal section immunohistochemistry and
mapping

To determine the mouse brain atlas coordinate of each coronal microbiopsy section, the
histological section immediately adjacent to each microbiopsy section was fixed in 4%
PFA for 10 min and counterstained using 300 µM DAPI (Thermo Fisher, Cat. D1306) in
1× (pH 7.4) PBS (Thermo Fisher, Cat. 10010023) for 5 min. Slides were rinsed with 1×
PBS and mounted in Fluoromount-G (Thermo Fisher, Cat. 00-4958-02). Slides stained

for Satb2 were equilibrated to room temperature and circumscribed with a hydrophobic barrier pen (Invignome, Cat. GPF-VPSA-V). Sections were washed twice with PBS for 10 min then blocked for 1 h at room temperature in permeabilization/blocking buffer comprised of PBS with 10% normal goat serum (NGS, Jackson ImmunoResearch, Cat. 005-000-121), 1% bovine serum albumin (BSA, Millipore, Cat. 126626), 0.3% Triton X-100 (TX-100, Sigma, Cat. 11332481001), 0.05% Tween-20 (Sigma, Cat. P1379), 0.3 M glycine (Sigma, Cat. G7126), and 0.01% sodium azide (Sigma, Cat. S2002). During the blocking step, the primary antibody rabbit anti-Satb2 (Abcam Cat. ab92446) was diluted 1:1000 in a buffer containing PBS, 2% NGS, 1% BSA, 0.01% TX-100, 0.05% Tween-20, and 0.01% sodium azide. The diluted primary antibody was applied to sections and then incubated overnight at 4 °C. The primary antibody was washed from the sections five times with PBS for 5 min at room temperature. Secondary antibody AF488 goat anti-rabbit (Thermo Fisher Cat. A32731) was prepared by diluting 1:1000 in the same buffer used to dilute primary antibodies. Sections were incubated with the diluted secondary antibody for 1 h in the dark at room temperature. Secondary antibodies were washed from the sections three times with PBS for 5 min; then nuclei were counterstained with DAPI for 10 min at room temperature. After DAPI staining, sections were washed an additional two times then glass coverslips were mounted with ProLong Diamond Anti-Fade Mounting Medium (Thermo Fisher, Cat. P36961). Slides were imaged on a Zeiss ApoTome AxioImager M2 fluorescent upright microscope and processed using Fiji software (v1.52p)[129]. Coronal section images were mapped to the Adult Mouse Allen Brain Atlas[92] according to anatomical regions.

2.7.5 Mouse cerebral ischemia immunohistochemistry and mapping

One of the histological sections corresponding to each microbiopsy section was stained for GFAP to identify the infarct. Slides were equilibrated to room temperature and

circumscribed with a hydrophobic barrier pen. Sections were washed twice with PBS for 10 min and then blocked for 1 hour at room temperature in permeabilization/blocking buffer comprised of PBS with 10% normal donkey serum, 1% BSA, and 0.05% TX-100. The sections were next incubated in a primary antibody solution comprised of 1:1000 goat anti-GFAP (Abcam, ab53554) and 1:5000 rabbit anti-Iba1 (Fujifilm Wako, NCNP24) diluted in PBS with 1% NGS, 0.1% BSA, and 0.005% TX-100 overnight at 4 °C. The sections were then washed three times with PBS for 5 min each at room temperature and next incubated for 2 h at room temperature in a secondary antibody solution containing 1:500 donkey anti-goat conjugated to Alexa Fluor 488 (Invitrogen) and 1:500 donkey anti-rabbit conjugated to Alexa Fluor 555 (Invitrogen) prepared in the same buffer as the primary antibodies. Following the secondary incubation, sections were washed three times with PBS for 5 min each, counterstained with DAPI for 10 min, washed an additional two times for 5 min each, then coverslipped with Fluoromount-G. Slides were imaged on a Zeiss AxioScan.Z1 Slide Scanner and processed using Fiji software (v1.52p). Coronal cerebral ischemia section images were mapped to the Adult Mouse Allen Brain Atlas[92] according to anatomical regions using the DAPI channel, as described above.

Immunohistochemistry fluorescence was quantified using Fiji software (v1.52p). Punch positions were mapped to regions of interest (ROIs), along with three negative naive ROIs for each image. Corrected total fluorescence was calculated as the difference between the integrated density (ROI area × mean fluorescence) of an ROI for a given punch and the average integrated density of negative naive ROIs. GFAP-corrected total fluorescence was plotted using geom_boxplot and geom_smooth, method lm using ggplot (v3.2.1) in R (v3.5.1).

2.7.6 Mouse whole brain dissociation and nuclei isolation

To evaluate the effect of flash-freezing on chromatin accessibility in mouse brain tissue, we evaluated single-cell chromatin accessibility profiles from an intact mouse brain, in which one hemisphere was flash-frozen, and one hemisphere remained unfrozen. Both hemispheres were processed in parallel and underwent dissociation and nuclear isolation. Tissue was diced in NIB (10 mM Tris HCl, pH 7.5 [Fisher, Cat. T1503 and Fisher, Cat. A144], 10 mM NaCl [Fisher, Cat. M-11624], 3 mM MgCl2 [Sigma, Cat. M8226], 0.1% IGEPAL [v/v; Sigma, I8896], 0.1% Tween-20 [v/v, Sigma, Cat. P7949], and 1× protease inhibitor [Roche, Cat. 11873580001]) in a petri dish on ice using a chilled razor blade. Diced tissue was transferred to 2 mL chilled NIB in a 7 mL Dounce-homogenizer on ice. The tissue was incubated on ice for 5 min and then homogenized via 10 gentle strokes of the loose pestle (A) on ice, a 5-min incubation on ice, then ten gentle strokes of the tight pestle (B) on ice. The homogenate was then strained through a 35 µm strainer and centrifuged at 500 × g for 10 min. Samples were aspirated, resuspended in 5 mL of ice-cold NIB, and nuclei were counted on a hemocytometer. Samples were diluted to 500 nuclei per 1 µL to facilitate tagmentation reaction assembly at ~5000 nuclei per 10 µL of NIB.

2.7.7 Tissue microbiopsy acquisition and nuclear isolation

Tissue microbiopsies were acquired from 100–300 µm sections. Punches were isolated in four experiments: (1) mouse dissociation development sciMAP-ATAC (384 punches), (2) mouse SSp cortex sciMAP-ATAC (96 punches), (3) mouse cerebral ischemia sciMAP-ATAC (240 punches), and (4) human VISp cortex sciMAP-ATAC (192 punches; for details refer to Fig. 8). Microbiopsy coronal sections were acclimated to −20 °C in a cryostat (Leica CM3050), and microbiopsy punch tools (EMS, Cat. 57401) were cooled on dry ice prior to punching to prevent warming of the tissue. Microbiopsy punches were

acquired according to the location identified from section atlas mapping, and frozen microbiopsies were deposited directly into 100 µL of ice-cold NIB in a 96-well plate. Punch deposition into each well of the 96-well plate was visually confirmed under a dissecting microscope. To facilitate tissue dissociation and nuclear isolation, 96-well plates of microbiopsy punches were then gently shaken (80 r.p.m.) while covered for 1 h on ice. We then tested mechanical dissociation by varying the number of triturations performed via multichannel pipette per well (punch dissociation development sciMAP-ATAC). We found the following averaged metrics across the four dissociation methods: 15 triturations (26 cells per punch, 5679 unique passing reads per cell, 0.844 FRis), 30 triturations (35 cells per punch, 7189 unique passing reads per cell, 0.835 FRis), 60 triturations (28 cells per punch, 7611 unique passing reads per cell, 0.827 FRis), and 100 triturations (8 cells per punch, 7611 unique passing reads per cell, 0.821 FRis). Given that 60-trituration mechanical dissociation yielded the highest number of cells per punch, with otherwise comparable metrics, we proceeded with 60 triturations for all future experiments. Post-mechanical dissociation, sample plates were then centrifuged at $500 \times g$ for 10 min. While nuclear pellets were not visible, we found that aspiration of 90 µL of supernatant and resuspension in an added 30 µL of NIB results in a final isolated nuclear volume of 40 µL with ~15,000 nuclei per well (for microbiopsy punching conditions: 200 µm section, 250 µm diameter microbiopsy punch used in the human VISp and mouse cerebral ischemia preparations). Nuclei were split across four 96-well plates such that nuclei were aliquoted to 10 µL, or ~3750 nuclei per well. This enabled four independent indexed transposase complexes to be utilized for each individual punch, or 384 uniquely indexed transposition reactions in one experiment. To calculate the approximate resolution for each preparation, we took the cubed root of the cylindrical volume.

2.7.8 Location indexing via tagmentation

Transposase catalyzed excision of the chromatin accessible regions via tagmentation results in the addition of unique molecular identifiers (indexes) for each tagmentation reaction. Uniquely indexed transposase adapter sequences are reported in [Supplementary Table 2](). To encode microbiopsy punch location into library molecules, we recorded the corresponding tagmentation well within each 96-well plate to the user-identified microbiopsy punch location. The incorporation of location information is therefore inherently encoded by the first tier of indexing in our established sci-ATAC-seq method. Tagmentation reactions were assembled at 10 μL of isolated nuclei at 500 nuclei per 1 μL, 10 μL 2× tagmentation buffer (Illumina, Cat. FC-121-1031), and 1 μL of 8 μM loaded indexed synthesized Tn5 transposase was added per well (see Picelli et al. for transposase synthesis protocol)[130]. As an alternative to Tn5 synthesis, EZTn5 transposase (https://www.lucigen.com/EZ-Tn5-Transposase/) can be purchased commercially and diluted, salt adjusted, and loaded with sci indexes according to the sciMAP-ATAC protocol[131]. Each assembled 96-well plate of tagmentation reactions was incubated at 55 °C for 15 min. For the mouse whole brain sci-ATAC-seq preparation on fresh and frozen tissue, as well as the sciMAP-ATAC preparations, four 96-well plates of tagmentation were used (384 uniquely indexed tagmentation reactions). For whole-brain sci-ATAC-seq preparation on fresh and frozen tissue experiment, tagmentation wells were pooled separately for fresh and frozen hemisphere samples. For the microbiopsy punch-derived experiments, all reactions were pooled post-tagmentation.

2.7.9 Combinatorial indexing

To lyse nuclei and release bound transposase, PCR plates are prepared with protease buffer (PB), primers, and sparsely sorted nuclei and then incubated. Uniquely indexed PCR primer sequences are reported in [Supplementary Table 2](). Post-denaturation, the

remaining PCR reagents are added, and the incorporation of the PCR primers results in the incorporation of the secondary index for single-combinatorial indexing. For the denaturation step, 96-well PCR plates of 8.5 µL PB (30 mM Tris HCl, pH 7.5, 2 mM EDTA [Ambion, Cat. AM9261], 20 mM KCl [Fisher, Cat. P217 and Fisher, Cat. A144], 0.2% TX-100 [v/v], 500 µg/mL serine protease [Fisher, Cat. NC9221823]), 1 µL 10 mM indexed i5, and 1 µL indexed i7 per well were prepared. Pooled tagmented nuclei were stained by adding 3 µL of DAPI (5 mg/mL) per 1 mL of sample. Each sample was then FAN sorted using BD FACSDiva software (v8.0.1) on a Sony SH800 FACS machine at 22 events per well per 96-well Tn5 plate (e.g., 88 for 384 indexes) into prepared 96-well plate(s). Event numbers were selected based on the expected success rate of events as actual cells for a given target cell doublet rate (see "Doublet rate estimations" section below). Across the sciMAP-ATAC experiments, four PCR plates (384 uniquely indexed wells) were utilized for the initial punch-derived sci-ATAC-seq preparation from whole brain-derived punches, two PCR plates (192 uniquely indexed wells) were used for the mouse SSp cortex experiment, one full and one partial plate (128 uniquely indexed wells) for the human VISp experiment, two plates (192 uniquely indexed wells) for the mouse cerebral ischemia experiment, and finally, two PCR plates (192 uniquely indexed wells) were utilized for the nonspatial whole-brain sci-ATAC-seq preparation on fresh and frozen tissue. Transposase denaturation was performed by sealing each sorted plate and incubating at 55 °C for 15 min. Plates were immediately transferred to ice post-incubation, and 12 µL of PCR mix (7.5 µL NPM [Illumina Inc. Cat FC-131-1096], 4 µL nuclease-free water, and 0.5 µL 100× SYBR Green) was added to each well. For each experiment, plates were then sealed and PCR amplified on a BioRad CFX real-time cycler running CFX Manager (v3.1) software, using the following protocol: 72 °C for 5:00, 98 °C for 0:30, cycles of (98 °C for 0:10, 63 °C for 0:30, 72 °C for 1:00, plate read, 72 °C for 0:10) for 18–22 cycles. PCR plates were transferred to 4 °C once all wells reached

mid-exponential amplification on average. Each PCR plate is then pooled at 10 µL per well, and DNA libraries are isolated using a QIAquick PCR Purification column. Each pooled PCR plate library is then quantified using a Qubit 2.0 fluorimeter, diluted to 4 ng/µL with nuclease-free water, and quantification of library size performed on Agilent Bioanalyzer using a dsDNA high sensitivity chip. Libraries were then sequenced on a NextSeqTM 500 sequencer (Illumina Inc.) running NextSeq500 NCS (v4.0) software loaded within a range of 1.2–1.6 pM with a custom sequencing chemistry protocol (read 1:50 imaged cycles; index read 1:8 imaged cycles, 27 dark cycles, 10 imaged cycles; index read 2:8 imaged cycles, 21 dark cycles, 10 imaged cycles; read 2:50 imaged cycles) using custom sequencing primers supplied in Supplementary Table 2.

2.7.10 Doublet rate estimations

An important factor in single-cell studies is the expected doublet or collision rate. This manifests in droplet-based platforms as two cells being encapsulated within the same droplet, thus having the same cell barcode for their genomic information. This is tunable by the number of cells or nuclei loaded onto the instrument, with typical doublet rates targeted to be at or <5%. This is also true for combinatorial indexing workflows, where doublets are present in the form of two cells or nuclei with the same level 1 index—which is the transposase index for ATAC—that end up in the same level 2 indexing well (i.e., the PCR well). This results in an identical pair of indexes for the two cells. This rate, like with droplet methods, is also tunable by altering the number of indexed cells or nuclei that are deposited into each well, with a typical experiment targeting at or below a 5% doublet rate. This rate is approximated by leveraging the "birthday problem" formulation in statistics, where the transposase index space (days in the year) and number of indexed nuclei per well (number of people at each table) are taken into account. These predictions assume that there is complete mixing of nuclei prior to distribution and that

the distribution is unbiased, which are reasonable given the single-nuclei suspension and use of flow sorting for the distribution process, and hold up when compared to empirical data produced by multispecies cell mixing experiments[85,86,126] (i.e., barnyard experiments, typically mixing human and mouse cells). However, in the case of sciMAP-ATAC, nuclei are directly isolated and then indexed within the same well, making a true barnyard experiment not feasible. Any experiment that would use tissue punches from two different species into different wells would not capture doublets because of the de facto unique indexes for each species imparted by the different wells for the first level of indexing. We, therefore, assumed that the assumptions that have been made and tested for standard sci-ATAC-seq and related combinatorial technologies also apply to sciMAP-ATAC, as the novel components of the workflow are in the processing prior to the combinatorial indexing stages.

With our set of 384 unique transposase indexes and the sorting of 88 nuclei per well across experiments, this would result in a doublet rate (i.e., two nuclei of the same transposase index ending up in the same PCR well) of 10.5% if the yield of sorted nuclei was perfect. However, we favor speed over precise quantification during the sorting step, as the actual number of sorted cells does not matter as long as it ends up being below the target number. We have found that using our fast-sorting workflow, of the target number of events that are sorted, only between 25 and 50% are true nuclei. The rest of the events are empty droplets. We also note that these droplets do not contain ambient chromatin based on human-mouse mixing experiments[86]. Using the high end of the ~50% true nuclei yield, the expected doublet rate is 5.4%, in line with other commercially available single-cell platforms. When factoring in the actual yield with respect to single-cell profiles produced, the doublet rate is even lower. For example, the punch dissociation development sciMAP-ATAC preparation produced 8012 single-cell

profiles over 384 unique indexed transposition wells, for an average of just under 21 cells produced per well out of the 88 events that were sorted—a 23.7% yield. The final expected doublet rate is, therefore, most accurately calculated according to 21 indexed nuclei produced per well with a transposase index space of 384 for a doublet rate of 2.5%, which is well within the accepted range.

2.7.11 Sequence data processing

Fastq files were generated from BCL files using bcl2fastq (Illumina Inc., v2.19.0). Fastq files were aligned, filtered, and analyzed primarily using the "scitools" software (github.com/adeylab/scitools)[89], which includes wrappers for numerous external tools. Raw sequence reads had their index combinations matched to a whitelist of expected indexed using "scitools fastq-dump", which allows for a hamming distance of two and produces error-corrected fastq files. These were then aligned to a mouse or human reference genome (mm10 or hg38) via bwa mem (v0.7.15-r1140)[132] and sorted using "scitools align". PCR duplicate removal and filtering for quality ten aligned autosomal and chromosome X reads (i.e., excluding mitochondrial, chromosome Y, and unanchored contigs) was performed using "scitools rmdup" with default parameters and plotted using "scitools plot-complexity". Projections of passing reads given increased sequencing depth were performed using "scitools bam-project" on the pre-duplicate removed bam file, which generates a model for every single cell based on sampling reads and calculating the passing read percentage that empirically falls within 2% accuracy[86]. Bam files were then filtered to only contain cell barcodes that contained a minimum of 1000 passing reads and a percent unique reads <80 (any overly complex cell libraries may be doublets and were therefore excluded). For the human VISp dataset, cells were also filtered to have a TSS enrichment (per cell calculation) of 2 (see section "Quality metric calculations").

2.7.12 Chromatin accessibility analysis

The filtered bam file was used for chromatin accessibility peak calling for each of the five experiments individually, as well as on a combined bam file from the mouse whole brain sciATAC-seq, mouse punch dissociation development sciMAP-ATAC, and mouse SSp cortex sciMAP-ATAC experiments for the combined dataset analysis. Peak calling was run using the wrapper function "scitools callpeak", which utilized macs2 (v2.1.1.20160309) for peak calling, and then filtering and peak extension to 500 bp (ref.[133]). Called peaks from mouse whole brain sciATAC-seq, mouse punch dissociation development sciMAP-ATAC, and mouse SSp cortex sciMAP-ATAC datasets were merged to generate a union peak set that was used to compare sciATAC-seq and sciMAP-ATAC clustering. Peak bed files and filtered bam files were then used to construct counts matrix of cells × peaks. Latent Dirichlet Allocation using the package cisTopic (v0.2.0)[134] was performed using the scitools wrapper function "scitools cistopic". Topic enrichments for region type annotations (Fig. 14g) were annotated using cisTopic function annotateRegions, using the Bioconductor package TxDb.Hsapiens.UCSC.hg38.knownGene (v3.4.7) and annotation database org.Mm.eg.db (v3.8.2). The topic by annotation heatmap was plotted using cisTopic function signaturesHeatmap. The cells × topics matrix was biclustered and plotted using "scitools matrix-bicluster", which utilizes the Heatmap function in the ComplexHeatmap package (v1.20.1) in R (v3.5.1)[135]. Two-dimensional visualization was performed using UMAP via "scitools umap" and plotted using "scitools plot-dims". Visualization of topic weights on the UMAP coordinates was performed using "scitools plot-dims" with -M as the cells × topics matrix. Clustering was performed on the cells × topics matrix using the package Rphenograph (v0.99.1) in R (v3.5.1), which employs Louvain clustering and was executed using the wrapper function "scitools matrix-pg"[136]. In addition to topic analyses, we utilized ChromVAR (v1.4.1)[119] to assess the global motif accessibility

profiles of cells using the wrapper function "scitools chromvar" on the bam file with added read group tags using "scitools addrg". Boxplots illustrating TFME per cell were generated using values from the ChromVAR deviations_scores matrix and plotted using geom_boxplot from the package ggplot (v3.2.1) in R (3.5.1), where lower and upper hinges indicate the first and third quartiles, the center line indicates median, upper, and lower whiskers indicate 1.5 times the inner quartile range (IQR). Data points beyond the end of the whiskers are plotted individually. All boxplot comparison significance calculations were performed using the compare_means function in the ggpubr package (v0.2.5), indicating paired = FALSE and p.adjust.method set to Bonferroni–Holm correction in R (v3.5.1).

2.7.13 Quality metric calculations

To generate tagmentation site density plots centered around TSSs, we first subset filtered experiment bam files into respective annotations. We used the alignment position (chromosome and start site) for each read to generate a bed file that was then fed into the BEDOPS closest-feature command mapped the distance between all read start sites and TSSs (v2.4.36, ref.[137]). From this, we collapsed distances into a counts table respective to experiment and annotation and generated a percentage of read start site distances within each counts table. We plotted these data using R (v3.6.1) and ggplot2 (v3.3.2) geom_line function (default parameters) subset to 2000 base pairs around the start site to visualize enrichment. TSS enrichment values were calculated for each experimental condition using the method established by the ENCODE project (https://www.encodeproject.org/data-standards/terms/enrichment), whereby the aggregate distribution of reads ±1000 bp centered on the set of TSSs is then used to generate 100 bp windows at the flanks of the distribution as the background and then through the distribution, where the maximum window centered on the TSS is used to

calculate the fold enrichment over the outer flanking windows. The fraction of reads in a defined read set (FRiS) was used as an alternative to the fraction of reads in peaks for two major reasons. The first is that FRiP is highly dependent on the number of peaks that are called, which is, in turn, highly dependent on (a) the number of cells profiled, and (b) the depth of sequencing. One can increase FRiP values by sequencing a library more deeply or profiling larger numbers of cells at the same depth without reflecting any difference in underlying data quality. Second, peak calling on a population of cells favors peaks in high abundance cell types, as they make up more of the data going into the peak calling. Therefore, cells of a cell type that is lower abundance will have fewer peaks called that are specifically associated with that cell type, owing to the dominance of signal by the more abundant cell type and consequently reducing the FRiP of those cells. Using FRiS instead largely avoids the challenges associated with peak calling by leveraging a comprehensive reference dataset. For the mouse FRiS calculations, we aggregated peaks that are available from mouse bulk ATAC-seq and DNAse hypersensitivity experiments provided by the ENCODE project, followed by peak collapsing, resulting in 2,377,227 total peaks averaging 744.9 bp. For the human dataset, we used a human reference dataset for DNAse hypersensitivity[95] that contains 3,591,898 loci defined as TF footprints with an average size of 203.9 bp leading to the lower FRiS values when compared to the aggregate mouse ATAC-seq peak dataset.

2.7.14 Cell type identification

The identified clusters were assigned to their respective cell type by examining the chromatin accessibility profile of marker genes that correspond to known cell types. Gene regions were plotted using "scitools plot-reads" using the filtered bam file, and genome track plots were generated using *CoveragePlots* from the analysis suite of tools, *Signac* (v0.2.5, https://github.com/timoast/signac). Additional support for identified cell

types was performed by assessing the chromVAR results for global motif accessibility. Marker genes used for cell-type identification included: *Gfap*, *Glul*, and *Agt* for astrocytes, *Col19a1* for all neuronal cell types, *Gad1*, *Gad2*, *Pvalb*, *Dlx1*, and *Dlx2* for GABAergic neurons, *Slc17a7*, *Drd1*, *Drd2*, *Bcl11b* (*Ctip2*), and *Ppp1r1b* for GABAergic MSNs, also referred to as SPNs, *C1qa*, *C1qc*, and *Cx3cr1* for microglia, *Mrc1* for macrophages within the microglia cluster, *Kdr* and *Flt1* for endothelia, *Olig1* for all oligodendrocyte cell types, *Top2a* and *Cspg4* (*NG2*) for OPCs, *Fyn*, and *Prox1* for newly formed oligodendrocytes, and *Mobp*, *Mog*, *Cldn11*, and *Prox1* for mature myelinating oligodendrocytes.

2.7.15 Gene ontology enrichment analysis

GO enrichment analysis was performed for the genomic regions defined within topic 30, the topic enriched in ischemia-specific cells. Single nearest genes to topic 30 regions were identified using GREAT (v4.0.4) for reference genome mm10 (ref.[138]). GO term statistical overrepresentation for GO biological processes was calculated using Panther (v14) binomial test with false discovery rate (FDR) correction for overrepresentation of topic 30 genomic regions in comparison to all mouse (mm10) genes. Data were plotted using *ggplot* (v3.2.1) plotting function *geom_barplot* in R (v3.5.1) with height corresponding to log2 GO term fold enrichment and colored by GO term −log10 FDR *Q*-value.

2.7.16 Transcription factor and site enrichment throughout trajectories

TF motif enrichment analysis was performed using chromVAR (v1.4.1) in R (v3.5.1) on all cells derived from ischemia mouse models, including the ischemic (stroke) hemisphere and contralateral (contra) hemisphere. For the cells × TFME matrix, cells were annotated by the punch they were derived from, and a linear regression of TFME as a function of punch location for each cell using the base function *lm* in R (v3.6.1).

Slopes of the linear model for the ischemic and contralateral hemispheres were defined as the coefficient of the fit. The statistical significance of the interaction between TFME over space and disease condition (stroke versus contralateral hemisphere) was calculated by performing an analysis of variance (ANOVA, *anova* base R v3.6.1) on the interaction of hemisphere on the linear regression defined by TFME as a function of punch position (TFME ~ punch × hemisphere$_{(stroke/contra)}$), and slopes were compared using the lsmeans package function *lstrends* (v2.30-0). Slopes were compared between the stroke and contralateral hemispheres by taking the difference between the slopes (Δ slope = slope$_{stroke}$ − slope$_{contra}$). The change in slope was z-scored to center and scale TFME difference, where z-score Δ slope is equal to two standard deviations from the mean. Volcano plot of $-\log_{10}p$ value by Δ slope was generated using the package EnhancedVolcano (v1.4.0) in R (v3.5.1). Line plots vignettes were generated by plotting volcano plot data using *ggplot* (v3.2.1) plotting function *geom_smooth*, method *lm*. Heatmaps illustrating cell-type-specific TFME over space were generated by subsetting ischemia mouse model cells by cell type and plotting TFME by punch, compared between stroke and contralateral hemispheres using package ComplexHeatmap (v2.0.0) in R.

Analysis of putative regulatory elements was performed by assessing the ATAC peak probabilistic weight per cell (*cisTopic* predictive distribution) across cells derived from punches taken from the infarct core to infarct border axis (punch positions 5–8) in the stroke and contralateral hemispheres, aggregated across all MCAO mice. This was performed similarly to TFME described above, where ATAC peak probability per cell was averaged by punch position (punch positions 5–8). ATAC peak probability along the 5–8 axis was fit to a linear model, and the slope in the stroke hemisphere as compared to the slope in the contralateral hemisphere in order to generate significance and delta-slope

values. We found that 3852 peaks out of 104,773 total peaks (4.8%) vary significantly across the 5–8 axis in MCAO stroke hemispheres in contrast to the contralateral hemispheres. In order to identify putative regulatory elements which are associated with the progressive gradient of glial reactivity from the infarct core to the infarct border in stroke hemispheres, we subset our spatially significant peak set to those which uniformly increase or decrease along the 5–8 axis in stroke hemispheres. We found 73 sites that uniformly increase with increasing proximity to the infarct border and no sites which uniformly decrease. We report these 73 spatially significant peaks as a reference for future MCAO regulatory element studies.

2.7.17 Moran's I spatial autocorrelation analysis

We performed a Moran's I test to assess spatial autocorrelation between punch locations, wherein a higher Moran's I value signifies a higher chance of cells from the same punch location being nearby in Euclidean space. Cells sourced from white matter punches were excluded. The test was performed using the same 27 topic weight matrix used for UMAP projections. Cells were split by assigned types and processed in parallel in R (v4.0) using a modified version of the "Moran_*I*" method in the function *graph_test* in monocle3 (v0.2.3.0)[139]. Briefly, we used a bootstrapping method wherein each punch location (1–8) was randomly assigned a new location for all punches and all trajectories, such that all cells from the same punch still shared the same location. The Moran's *I* value was calculated for 1000 iterations using this random location reassignment strategy. The resulting null distribution was then compared to our true punch location Moran's I using the *pnorm* function to perform an unpaired one-sided (lower.tail = FALSE) z-test. To account for multiple testing, we applied a Bonferroni correction to the p values.

2.7.18 Integration with snDrop-seq and scTHS-seq data

We applied cross-data-modality integration based in canonical correlation analysis (CCA) to coanchor our sciMAP-ATAC-seq dataset with publicly available snDrop-seq and scTHS-seq visual cortex datasets[96,140] (Fig. 15). For single-cell chromatin accessibility data, we used Signac (v1.1.0)[140] to perform latent semantic indexing (LSI) on the filtered chromatin accessibility matrices (for both scTHS-seq and sciMAP-ATAC-seq) and calculated the normalized LSI loadings scores (using dims: 2:30) for anchor weighting. We then created gene activity matrices using the R package Cicero (v1.3.4.10)[141]. Similarly, we preprocessed the snDrop-seq expression matrix using the standard Seurat 3 (v3.2.1) workflow, where we filtered for variable features (5000 features), scaled and normalized data, and reduced dimensions via PCA and UMAP. For RNA-ATAC integration, we first learned the transfer anchors based on the gene activity and expression data by applying FindTransferAnchor (with the parameters dims = 1:30 and reduction = "cca"). We then used TransferData (weight.reduction = atac[["lsi"]]), to project scRNA-seq data labels onto sciMAP-ATAC-seq cells. We finally created a confusion count matrix based on the label matches between the snDrop-seq predicted and sciMAP-ATAC-seq labels. Using a similar method for feature imputation at variable genes, we transferred the scRNA-seq data onto the sciMAP-ATAC-seq cells and performed PCA on the combined datasets, followed by visualization via UMAP[46]. We applied a matching CCA-based strategy to coanchor scTHS-seq and sciMAP-ATAC-seq cells, using 70,832 overlapping accessibility sites between the datasets. For label transfer, we used the normalized LSI loadings scores for anchor weighting of the scTHS-seq data and then compared labels via a confusion matrix (Fig. 15).

2.7.19 Statistics

Data are presented as mean ± SEM unless otherwise specified. Statistical analysis was generally performed by a two-sided, unpaired Wilcoxon nonparametric test or two-way ANOVA and the Bonferroni method of correction for pairwise multiple comparisons, or as specified in the figure legends. Significance was assigned to $p < 0.05$. All analyses were performed using R version 3.5.1 or scitools scripts (github.com/adeylab/scitools) unless otherwise specified. Plots were generated primarily using R ggplot2 version 3.2.1.

2.8 Data availability

Raw and processed single-cell library sequencing data, as well as single-cell metadata, have been submitted to the National Center for Biotechnology Information Gene Expression Omnibus (GEO) under the accession code GSE164849. All other data supporting the findings of this study are available with the article and its Supplementary Information files and from the corresponding author upon reasonable request. Source data are provided with this paper.

2.9 Code availability

Data analysis and plotting was performed using functions contained within the publicly available scitools software suite of single-cell analysis tools (github.com/adeylab/scitools).

2.10 Acknowledgements

We thank members of the Adey, O'Roak, and Wright labs for their support; Kylee Rosette for assistance with animal husbandry; Eleonora Juarez for discussion on protocol development; Anthony P. Barnes and Brian J. O'Roak for helpful discussions on experimental design; and Dominica Cao and Brooke DeRosa for IHC staining. This work was supported by the NIH Brain Initiative, National Institute for Drug Abuse

Chapter 3: An atlas of glial chromatin accessibility in the human cortex and hippocampus

This chapter contains a modified version of material that appeared in the author's submitted manuscript: Thornton, Casey A., et al. "An atlas of glial chromatin accessibility in the human cortex and hippocampus."

3.1 Authors and affiliations

Casey A. Thornton[1], Sonia N. Acharya[1], Ruth V. Nichols[1], Andrew Nishida[1], Ryan M. Mulqueen[2], Brendan L. O'Connell[1], Andrew J. Fields[1], Randy Woltjer[3], Anusha Mishra[4,5], Andrew C. Adey[1,2,5,6,*]

1) Molecular and Medical Genetics, Oregon Health & Science University, Portland, OR.

2) CEDAR, Oregon Health & Science University, Portland, OR.

3) Department of Pathology, Oregon Health & Science University, Portland, OR.

4) Jungers Center for Neurosciences Research, Department of Neurology, Oregon Health & Science University, Portland, OR.

5) Knight Cardiovascular Institute, Oregon Health & Science University, Portland, OR.

6) Knight Cancer Institute, Oregon Health & Science University, Portland, OR.

* To whom correspondence should be addressed: adey@ohsu.edu

3.2 Author contributions

A.C.A. and C.A.T. conceived of the idea. S.N.A., C.A.T. and R.V.N. performed all experiments with assistance from A.J.F.; R.W. and A.M. contributed to experimental design and data interpretation. C.A.T. performed data processing, analysis, and

interpretation with assistance from R.M.M, A.N, and A.C.A. C.A.T. and A.C.A. wrote the

manuscript with input from all authors.

3.3 Abstract

The human central nervous system is a complex tissue comprising neuronal and non-neuronal cell types that orchestrate cognition. Single-cell characterization of neurons has revealed cellular sub-types with diverse functions; however, non-neuronal cells have not been characterized to the same resolution. This discrepancy in our understanding of neuronal and glial molecular diversity is an obstacle to creating a complete map of cell state heterogeneity and, therefore, glial function in the human brain. In this study, we focused on the cortex and hippocampus of the human brain in nine cognitively healthy adults. We assessed the cellular diversity of glia using the single-cell chromatin accessibility assay sci-ATAC-seq to epigenetically characterize 378,270 cells with 779,149 open chromatin regions. By selecting against NeuN immunolabelling, we achieved an 88.47% glial identity in cells processed across cortex and hippocampus samples. After sub-clustering of glial cell types, we uncovered intra-cell-type heterogeneity and multiple activated glial cell states in the healthy adult brain that are typically associated with the disease. Taken together, we produced a high cell coverage reference dataset of healthy human glia with profound heterogeneity of cell states.

3.4 Introduction

Single-cell approaches have enabled characterization of this cellular heterogeneity via assessment of genomic and epigenomic features[127,142]. Notably, single-cell chromatin accessibility characterizes cells by the organization of chromatin, providing insight into gene regulatory control and allowing for the assessment of epigenetically distinct cellular states, including those associated with health and disease. Single-cell characterization of neurons has been of prime interest in neuroscience, as neurons vary by regional specificity throughout the brain, exhibit variable morphologies, and assemble circuits that perform diverse information processing functions[48]. While single-cell chromatin accessibility has expanded the taxonomic classification of neurons in the human brain, the characterization of glia in the human brain is incomplete.

Glia in mammalian brains, like neurons, demonstrate disparate functions and morphologies both inter- and intra- regionally and also adopt heterogeneous states in tissue environments experiencing inflammatory pathologies. For instance, astrocytes and microglia are essential for synaptogenesis, synapse pruning, and elimination in the healthy CNS[143], yet also adopt reactive states in response to inflammation. Previous work has categorized reactive states of glia based on cellular function and injury type in various mouse models, characterizing reactive astrocytes in mice into distinct A1 immune responsive and A2 phagocytic states and reactive microglia into M1 pro-inflammatory and M2 anti-inflammatory states[143]. Though these characterizations of glial cell types have provided a basic framework for describing the heterogeneity of reactive glial cells, such a binary categorization is considered overly simplistic. Instead, as glial reactivity is induced in response to a wide array of disparate pathological contexts, the reactive glia themselves are expected to showcase just as broad a heterogeneity depending on the specific context[144,145]. To understand the heterogeneity of reactive glial

cells, however, it is necessary to first possess a rigorous and complete catalog of glial cell types and cell states in the cognitively healthy brain.

In this study, we used the single-cell combinatorial indexing assay for transposase accessible chromatin using sequencing (sci-ATAC-seq)[85,89] to generate a glia-enriched atlas of 378,270 cells from the human brain. We focused on two regions, the frontal cortex (middle frontal gyrus) and hippocampus, from nine cognitively healthy individuals in order to assess baseline glial heterogeneity. We present our findings on major glial cell types, including oligodendrocyte progenitor cells (OPCs, also known as polydendrocytes), astrocytes, microglia, and oligodendrocytes. In these gross cell types, we identify sub-types that have been previously described using other methodologies as well as novel sub-classifications with distinct epigenetic states. Importantly, we found that key features of astrocytes and microglia that have been described in disease states of the brain are present in low abundance in healthy aging individuals, supporting the premise that high cell numbers are required to identify rare cell states. We believe that these findings provide a necessary baseline for future investigations into diseases of the brain.

3.5 Results

3.5.1 Sci-ATAC-seq enriched for glial cells in human post-mortem brain tissue

We performed sci-ATAC-seq on two regions of the human brain: the frontal cortex (middle frontal gyrus) and hippocampus, in order to assess heterogeneity in the glial chromatin landscapes (Fig. 23a). Nuclei were isolated from cryopreserved samples collected from nine cognitively normal post-mortem individuals (4 male, 5 female) ranging from 37-73 years of age to identify shared glial heterogeneity between individuals (Fig. 23b, Supplementary Table 3), and then stained with an anti-NeuN

antibody with a secondary GFP antibody to enable the enrichment of NeuN-negative nuclei. We applied sci-ATAC-seq and sequencing across fifteen experimental preparations, generating chromatin accessibility profiles of 378,270 cells (Fig. 23c, Fig. 20). We identified 779,149 open chromatin regions from the complete dataset with a mean number of 20,520 unique reads per cell (Fig. 23d, Fig. 21a,c,e) and an ideal global transcription start site (TSS) enrichment of 17.85, according to ENCODE standards[146] (ideal TSS ≥ 7, Fig. 23e, Fig. 21b,d,f).
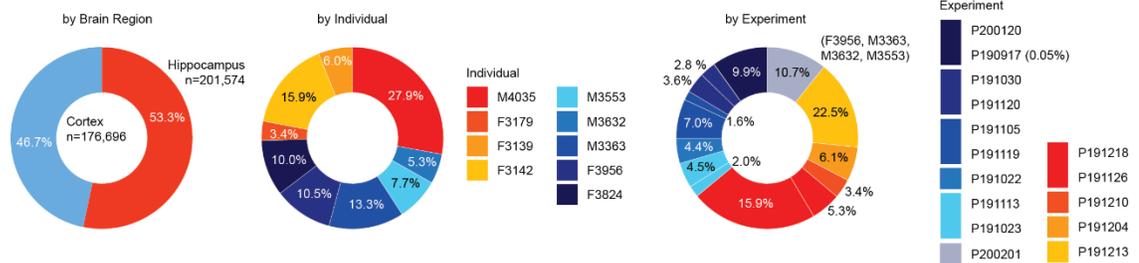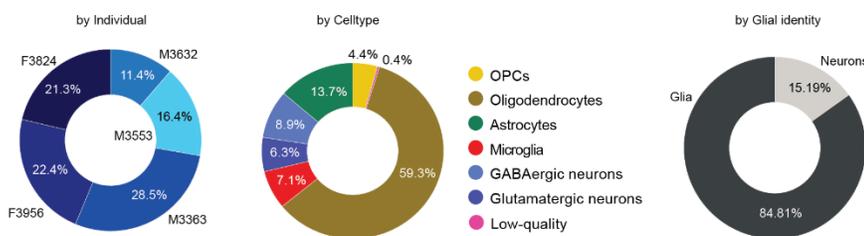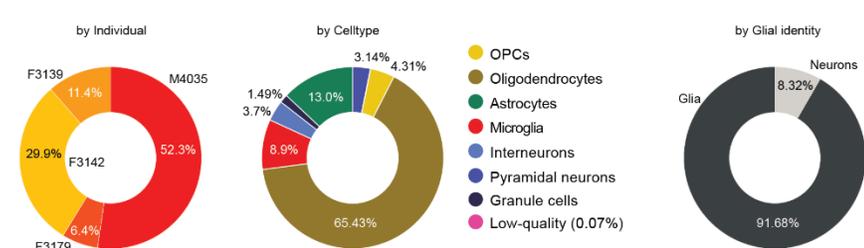
Figure 20. Overview of single-cell ATAC profiles produced across experimental conditions. Quality-passing single-cell ATAC-seq profiles for all, cortex and hippocampus cells (or mean ± SD) for **a,** All human glial atlas cells (n=378,270 cells); **b,** Cortex-derived cells (n=176,696 cells); **c,** Hippocampus-derived cells (n=201,574 cells).

We initially performed our analysis respective of brain regions to identify canonical cortex and hippocampus cell types within the glial-enriched dataset. For each region, we

performed topic-based dimensionality reduction[134], followed by cluster identification on the cell by topic matrix, and then visualization of chromatin accessibility in canonical cell type marker genes on a UMAP projection. In line with expectations, a small number of neuronal nuclei made it through the NeuN-negative selection based on the gating used during sorting. In the cortex, we identified six gross cell types that encompass several merged clusters; the first two corresponding to glutamatergic and GABAergic neurons, and the remaining four representing the four primary glial cell types: oligodendrocytes, OPCs, astrocytes, and microglia (Fig. 23f). In the hippocampus, we identified the same four broad glial cell types: oligodendrocytes, OPCs, astrocytes, and microglia, as well as three neuron cell types: dentate granule cells, pyramidal neurons, and inhibitory neurons (Fig. 23g).
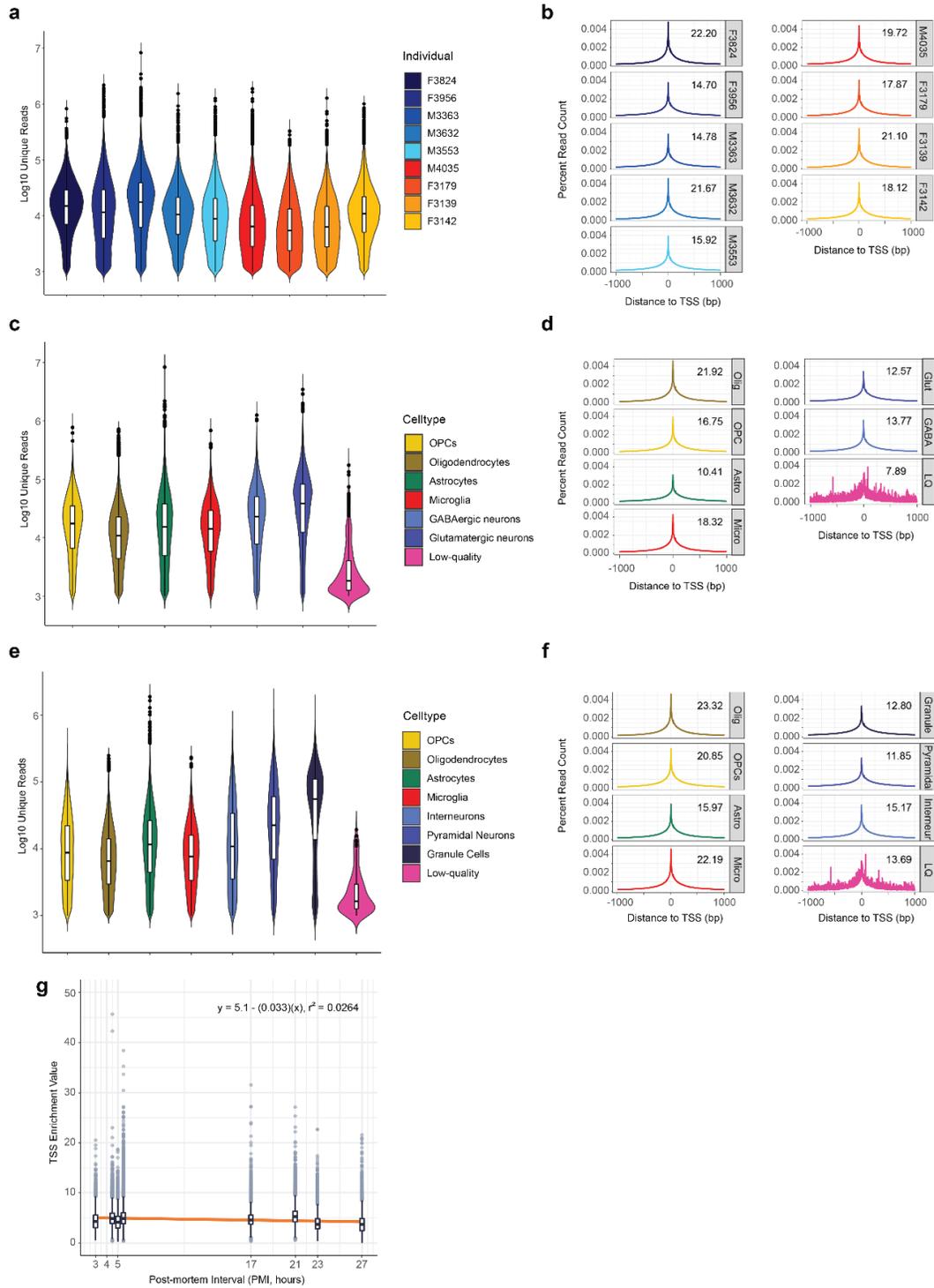
Figure 21. Quality assessment of single-cell ATAC-seq profiles. a, Passing reads per cell split by individual from cortex (blues) and hippocampus (oranges). **b,** Transcription start site (TSS) enrichment scores split by individual. **c,** Cortex passing reads per cell split by cell type. **d,** TSS enrichment scores of cortex cells split by cell type. **e,** Hippocampus passing reads per cell split by cell type. **f,** TSS enrichment scores of hippocampus cells split by cell type. **g,** TSS enrichment values of individual cells by post-mortem interval.
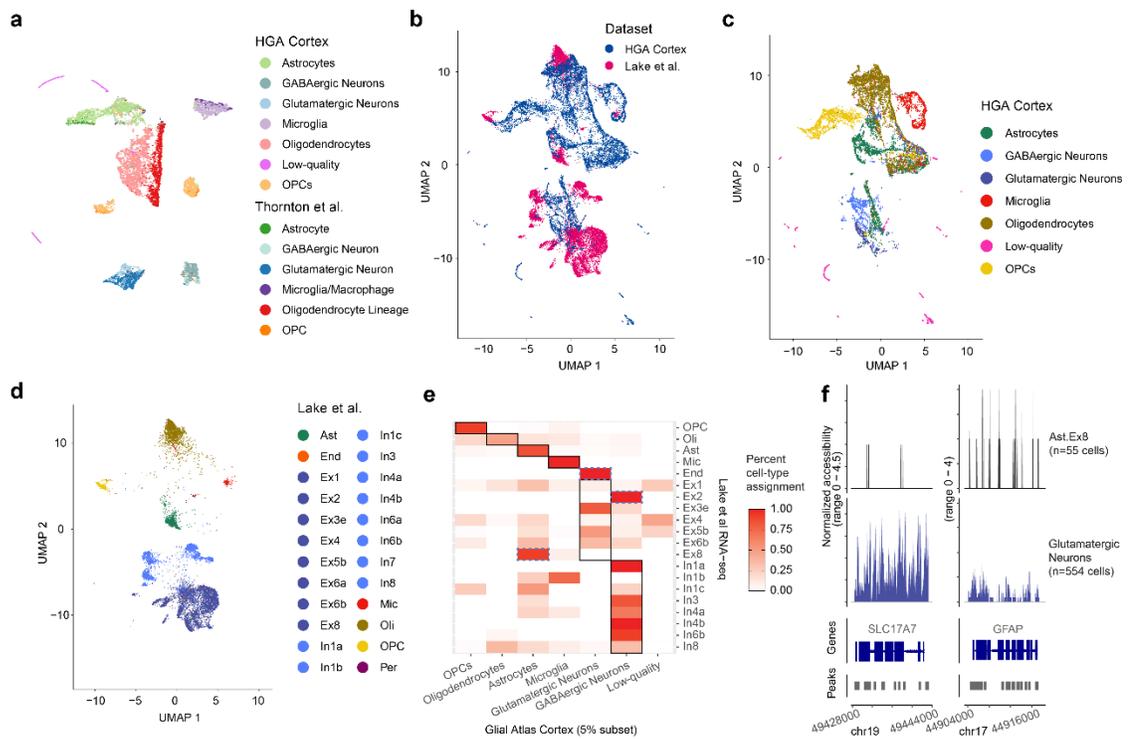
Figure 22. Cell type annotation validation by single-cell integrations. a, Human Glia Atlas (HGA) 5% cortex subset integration with Thornton *et al.* sci-ATAC-seq mouse primary somatosensory cortex cells (n=10,326 HGA cortex cells, n=4,547 Thornton *et al.* cells). **b,** UMAP of Seurat integration of HGA sci-ATAC-seq frontal cortex 5% cellular subset (blue, n=10,326 cells) and Lake *et al.* mouse frontal cortex (pink, n=10,319 cells), colored by dataset. **c,** UMAP, as shown in b, of only HGA cortex sci-ATAC-seq cells, colored by HGA annotated cell types (n=10,326 cells). **d,** UMAP, as shown in b, of only Lake *et al.* scRNA-seq cells, colored by Lake *et al.* annotated cell types (n=10,319 cells). **e,** Confusion matrix heatmap of percent assignment of predicted cell type label transfer from Lake *et al.* scRNA-seq to HGA cortex sci-ATAC-seq cells. Left-to-right diagonal marked by black outline to predicted cell type assignment agreement. Dotted blue outline signifies cell type assignment disagreement (off-diagonal). **f,** ATAC-seq profiles of HGA cortex astrocytes at excitatory neuron marker gene *SLC17A7*, split by predicted cell type (e.g., HGA astrocytes predicted to be astrocytes, HGA astrocytes predicted to be Excitatory Neuron 1, etc.), (n=2,415 HGA astrocyte cells).

To confirm the validity of our cell type assignment, we integrated a subset of our cortex dataset (5% downsampling of each cell type, n=10,326 cells) with the previously generated Thornton *et al.* adult human primary somatosensory cortex sci-ATAC-seq dataset (n=4,547 cells)[147] and, separately, the Lake *et al.* adult human frontal cortex snRNA-seq dataset (n=10,319, Fig. 22)[148]. We found perfect agreement across all cell types between the sci-ATAC-seq libraries from Thornton *et al.* and our cortex dataset (Fig. 22a) and high concordance with Lake *et al.* snRNA-seq after label transfer of cell-

type identity (Fig. 22e). In both integrations, cells that were annotated as low-quality due to low TSS and low unique reads demonstrated no clustering or label transfer cell type prediction with any previously identified cell type (Fig. 21c-f, Fig. 22a,c). As a result, these low-quality cells were excluded from future analysis (n=746). We then assessed the validity of cell type assignment discrepancies. This included cells that we had annotated as astrocytes but that Lake *et al.* label transfer predicted to be excitatory neurons, which we found exhibited no accessibility for the canonical excitatory neuron marker *SLC17A7* and clear accessibility in the astrocyte marker *GFAP* (Fig. 22e,f), suggesting that the issue may lie in the ATAC-to-RNA label transfer process and not with our assignments. As a result of our confirmed cell type annotation, we achieved 88.47% glial cell identity in a high-powered single-cell dataset derived from the human cortex and hippocampus.
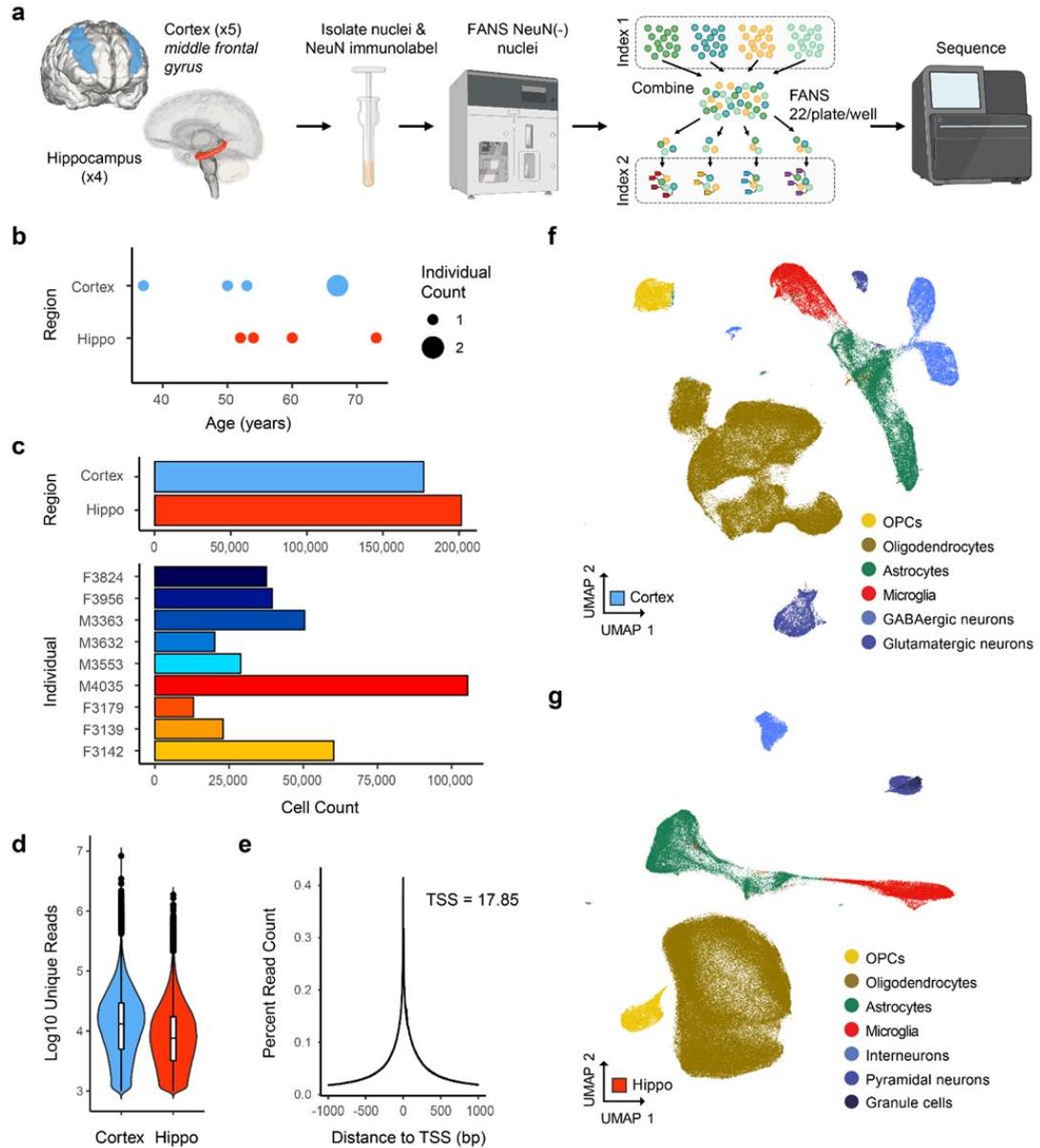
Figure 23. Chromatin accessibility profiling of 378,270 glia in the adult human cortex and hippocampus. a, Human glial atlas data generation workflow. Nuclei were isolated from nine post-mortem samples (5 cortex, 4 hippocampus, "hippo") and immunolabeled with NeuN. NeuN-negative nuclei were then flow sorted into wells for sci-ATAC-seq combinatorial indexed transposition, single-cell library generation and sequencing. **b,** Dot plot of post-mortem sample count by donor age and region. **c,** Barplot of total number of cells assayed split by region sampled (top) and individual sampled (bottom). **d,** Violin plot of number of passing reads per cell, split by region sampled. Center line represents median, lower and upper hinges represent first and third quartiles, whiskers extend from hinge to ±1.5 × IQR, individual cells represented as colored dots (n=378,270 cells). **e,** ATAC read signal at transcription start sites (TSSs) and surrounding base pairs (bps) for all cells. TSS of 17.85 for all cells. **f,** UMAP of cortex cells colored by gross cell type (n=176,696 cells). **g,** UMAP of hippocampus cell colored by gross cell type (n=201,574 cells).

93

Data from each regional glial cell type were then merged to produce cell-type-specific

combined datasets, using a union peak set between the two regions (n=779,149 peaks).

This resulted in the following datasets: oligodendrocytes (236,680 combined: 104,799

cortical, 13,1881 hippocampal cells), OPCs (16,534 combined: 7,842 cortical, 8,692

hippocampal cells), astrocytes (49,992 total: 24,151 cortical, 21,246 hippocampal cells),

and microglia (30,478 combined: 12,541 cortical, 17,937 hippocampal cells). Topic-

based dimensionality reduction and UMAP visualization was then performed on each of

the cell-type count matrices, revealing varied levels of individual-specific clustering

across the cell types (Fig. 24b,d,f). We found that oligodendrocytes, astrocytes, and

microglia clustering were driven by individual subjects, while OPCs revealed no

individual effect (Fig. 24a, 5c). To allow for inter-individual assessment of glial cell type

heterogeneity, we corrected for this individual-based batch effect by projecting cells from

each individual into shared embeddings using Harmony[149]. The resulting

oligodendrocyte, astrocyte, and microglia Harmony-weighted cell-by-topic matrix that we

utilized in the following cell type analysis resulted in clustering with improved

homogeneity of individuals across clusters (Fig. 24c,e,g). In all subsequent cell type-

specific analyses, we generated gene activity scores for each cell from chromatin cis-co-

accessibility networks[141] and assigned the most significantly enriched gene for each

cluster, or marker gene, as the cluster identifier.
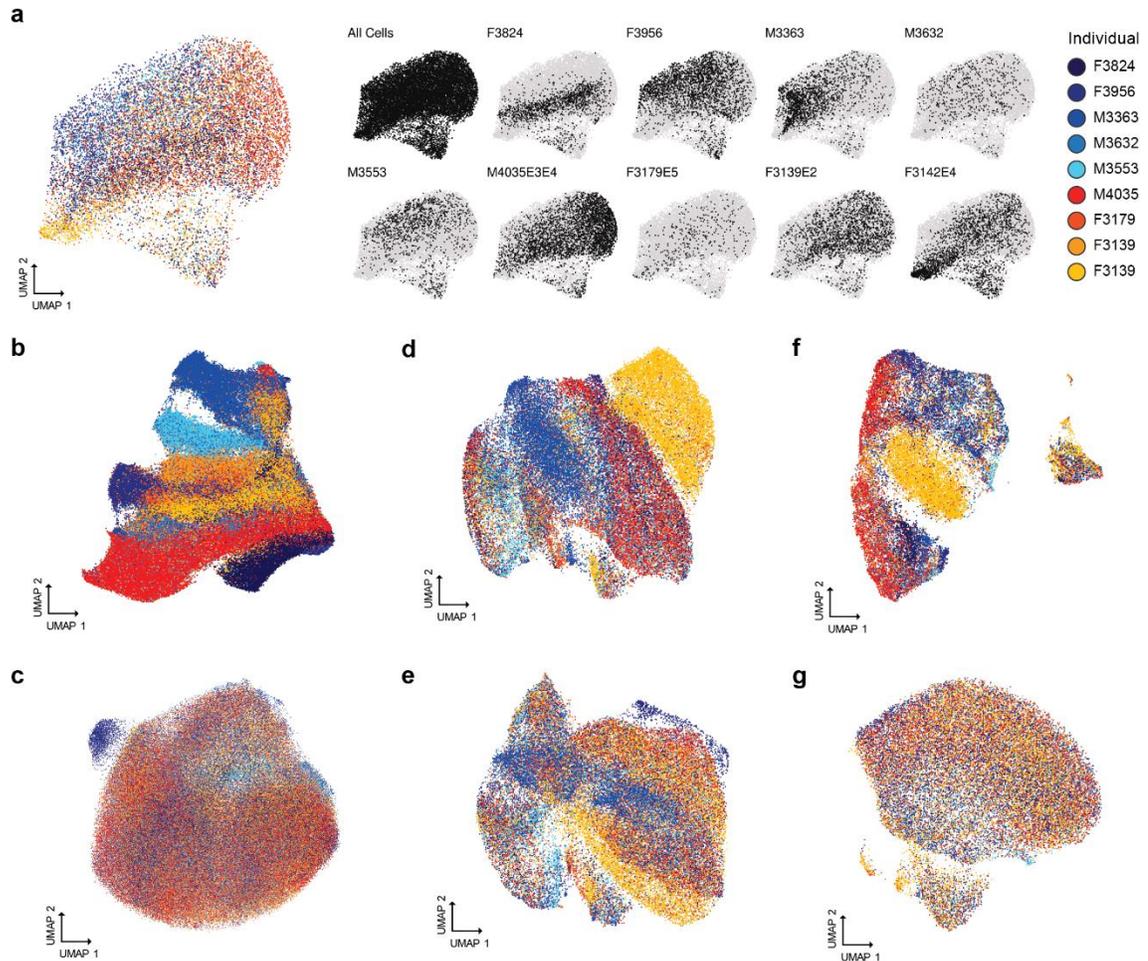
Figure 24. Individual-based batch-effect. Pre- and post-Harmony glial cell type UMAPs colored by individual. a, (Left) Combined OPC UMAP colored by individual (n=16,534 cells). (Right) Individuals mapped separately (black) onto all other OPC cells (grey). **b,** Pre-Harmony combined oligodendrocyte UMAP (n=236,680 cells). **c,** Post-Harmony combined oligodendrocyte UMAP (n=236,680 cells). **d,** Pre-Harmony combined astrocyte UMAP (n=49,992 cells). **e,** Post-Harmony combined astrocyte UMAP (n=49,992 cells). **f,** Pre-Harmony combined microglia UMAP (n=30,478 cells). **g,** Post-Harmony combined microglia UMAP (n=30,478 cells).

3.5.2 Astrocytes exhibit a wide range of sub-states stratified by *GFAP* gene activity

Astrocytes in the CNS have diverse functions, ranging from complexing with vascular

cells for blood-brain barrier maintenance to interacting with neurons for synapse

formation, pruning, and elimination. We identified astrocytes in the cortical and

hippocampal datasets from defined clusters (Fig. 25a) with elevated accessibility of the

markers: glial fibrillary acidic protein (*GFAP*) and angiotensinogen (*AGT*) as compared to non-astrocytic cell types (Fig. 25b)[150]. Initial processing of astrocytes revealed contaminating hippocampal dentate granule neurons, which were removed, resulting in a combined astrocyte dataset of 49,992 cells (48.3% cortex, 51.7% hippocampus, Fig. 26). We then identified eleven clusters using our Harmony-weighted cell by topic matrix, each labeled by its top marker gene: AST-*SNHG14*, AST-*C1orf61*, AST-*GFAP*, AST-*KCNJ3*, AST-*FOXG1*, AST-*5S-rRNA*, AST-*SNHG14*-2 (also marked by *SNHG14*, delineated with '-2'), AST-*KCNT1*, AST-*TPRN*, AST-*HIC1*, AST-*GPM6A* (Fig. 25c,g). These marker genes highlight some that have previously been identified in resting healthy and reactive disease-associated states. For instance, lncRNA-small nucleolar RNA host gene 14, *SNHG14*, is upregulated in astrocytes following stroke and promotes neuron survival while reducing inflammation[151]; however, the role of *SNHG14* in unperturbed astrocytes remains poorly understood. A complete list of top cluster marker genes for all cell types along with significance values is available in Supplemental Table 4.
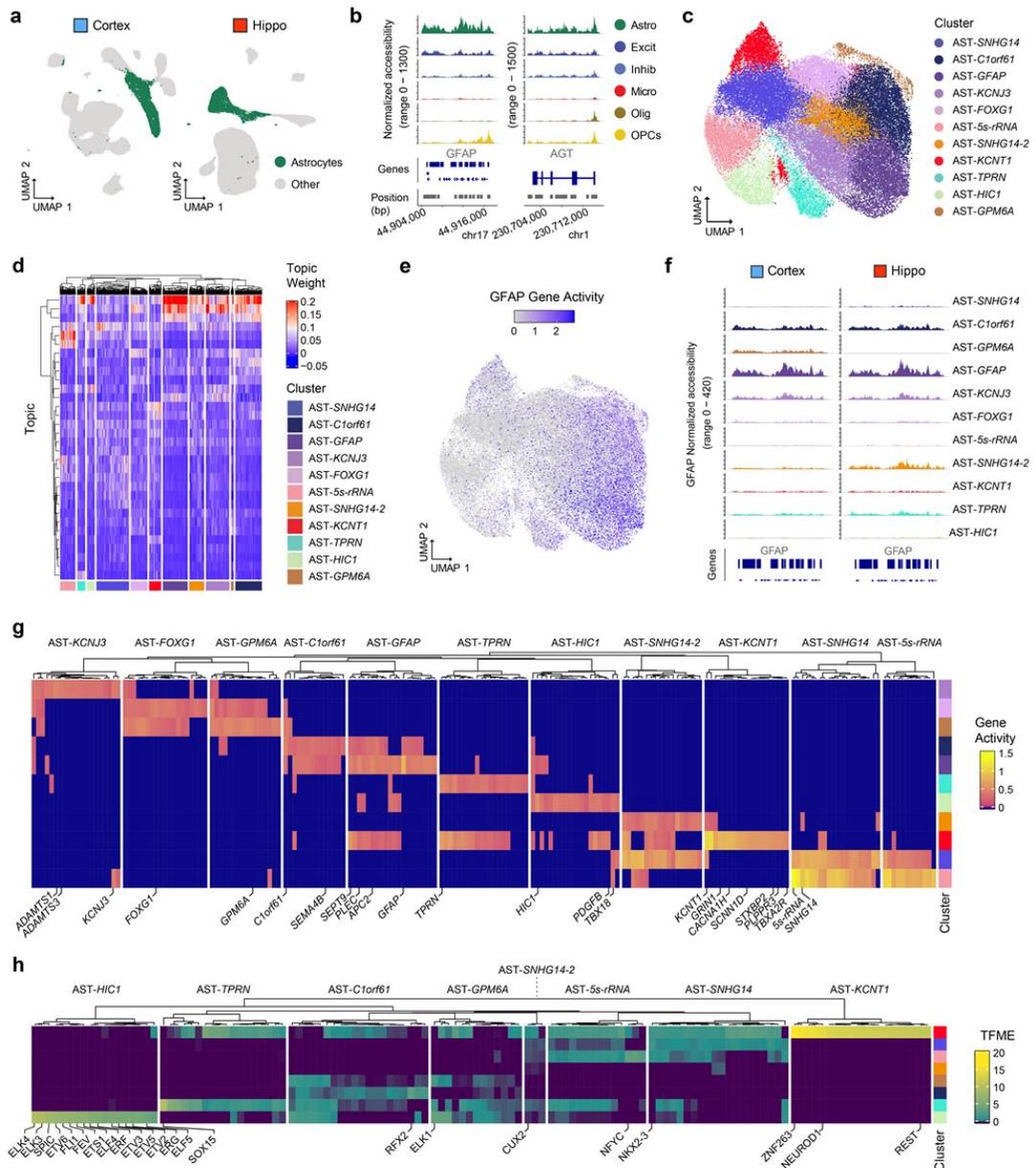
Figure 25. *GFAP* **delineates eleven astrocyte clusters into resting and reactive cell states. a,** Cortex (left) and hippocampus (right) UMAPs with astrocyte annotated cells highlighted (green) in contrast to all cells (grey). **b,** ATAC-seq profiles of astrocyte marker genes *GFAP* and *AGT* of all cells from cortex and hippocampus (n=378,270 cells). **c,** UMAP of combined cortex and hippocampus astrocytes colored by cluster (n=49,992). **d,** Astrocyte harmony-weighted topic matrix split and colored by cluster, as in c. **e,** UMAP of astrocytes colored by *GFAP* gene activity (n=49,992 cells). **f,** ATAC-seq profiles of astrocyte marker gene *GFAP* of astrocytes from cortex (left) and hippocampus (right) split by cluster, as in c (n=49,992 cells). **g,** Heatmap of gene activities for the top gene markers identified for each astrocyte cluster. All values are available in Supplementary Table 4. **h,** Heatmap of transcription factor motif enrichment scores for the top transcription factor markers for each astrocyte cluster. Clusters AST-*GFAP*, AST-*KCNJ3*, and AST-*FOXG1* had no significantly enriched gene markers. All values are available in Supplementary Table 4.

We found that for individual topics, *i.e.,* sets of associated open chromatin regions, accessibility was enriched in separate clusters, thus establishing the differential chromatin accessibility landscapes between putative astrocyte subtypes (Fig. 25d). We visualized these clusters using UMAP in three dimensions, where we observed a 3D topology that is not visible in the 2D UMAP, where for instance, cluster AST-*KCNT1* is separated from the bulk of astrocytes on the UMAP 3 axis (Fig. 28a). Assessment of individual patient distribution between clusters revealed that AST-*SNHG14-2* and AST-*GPM6A* were represented predominantly by one individual. Individual M3363 comprised 49.5% of AST-*SNHG14*-2 cells and F3956 comprised 98% of AST-*GPM6A* (Fig. 27a). As *SNHG14* is upregulated in ischemic reperfusion injuries, cluster AST-*SNHG14*-2 may represent an individual-specific inflammatory astrocyte subtype.
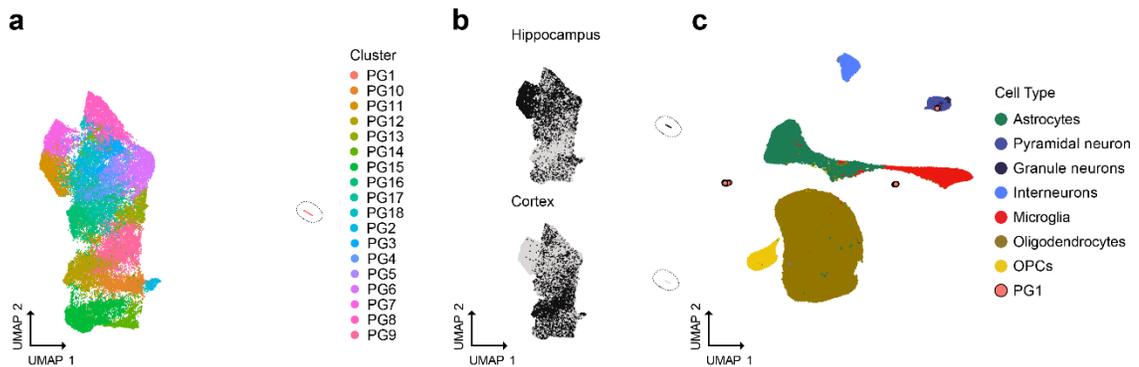
Figure 26. Neuronal contamination in astrocytes based on Harmony-weighted cisTopic matrix clustering. a, Louvain clustering of combined astrocytes. Outlier cluster identified as cluster PG1 (n= 50,312 cells). Dashed circle indicates cluster PG1. **b,** UMAP of combined astrocytes colored by all astrocytes (left), hippocampus-derived astrocytes (upper right), and UMAP of cortex-derived astrocytes (lower right), (n= 50,312 cells). Dashed circle indicates cluster PG1. **c,** Outlier PG1 cells (pink) on hippocampus UMAP colored by cell type (n=378,270 cells).

For our initial analysis, we assessed the distribution of *GFAP* gene activity across astrocytes. The GFAP protein is an intermediate filament that is expressed widely, at varying levels, in healthy astrocytes and is used as a canonical marker of the cell type. Expression of the *GFAP* gene is often elevated in reactive astrocytes[144,152]. While reactive states are primarily associated with inflammation and disease pathologies, rodent studies suggest that reactive astrocytes also exist in the normal brain, particularly

with aging[153]. Thus, we explored whether astrocytic reactive states exist in the healthy

human brain as well. In our data, we observed one cluster that was marked by *GFAP*

(AST-*GFAP*), and we found elevated *GFAP* gene activity in multiple clusters and

corresponding elevated chromatin accessibility in both cortex and hippocampus (Fig.

25e,f). Additionally, we assessed the gene activity of vimentin (*VIM*), an alternative

marker of astrocyte reactivity[144,154], and observed concordance between *GFAP*-high and

VIM-high clusters (Fig. 28a,b). Therefore, we subdivided our analysis into characterizing

GFAP-high: AST-*C1orf61,* AST-*GFAP,* AST-*KCNJ3,* AST-*SNHG14-2,* and AST-*GPM6A*;

and *GFAP*-low: AST-*SNHG14,* AST-*FOXG1,* AST-5*S-rRNA,* AST-*KCNT1,* AST-*TPRN,*

AST-*HIC1* astrocytic subtypes.
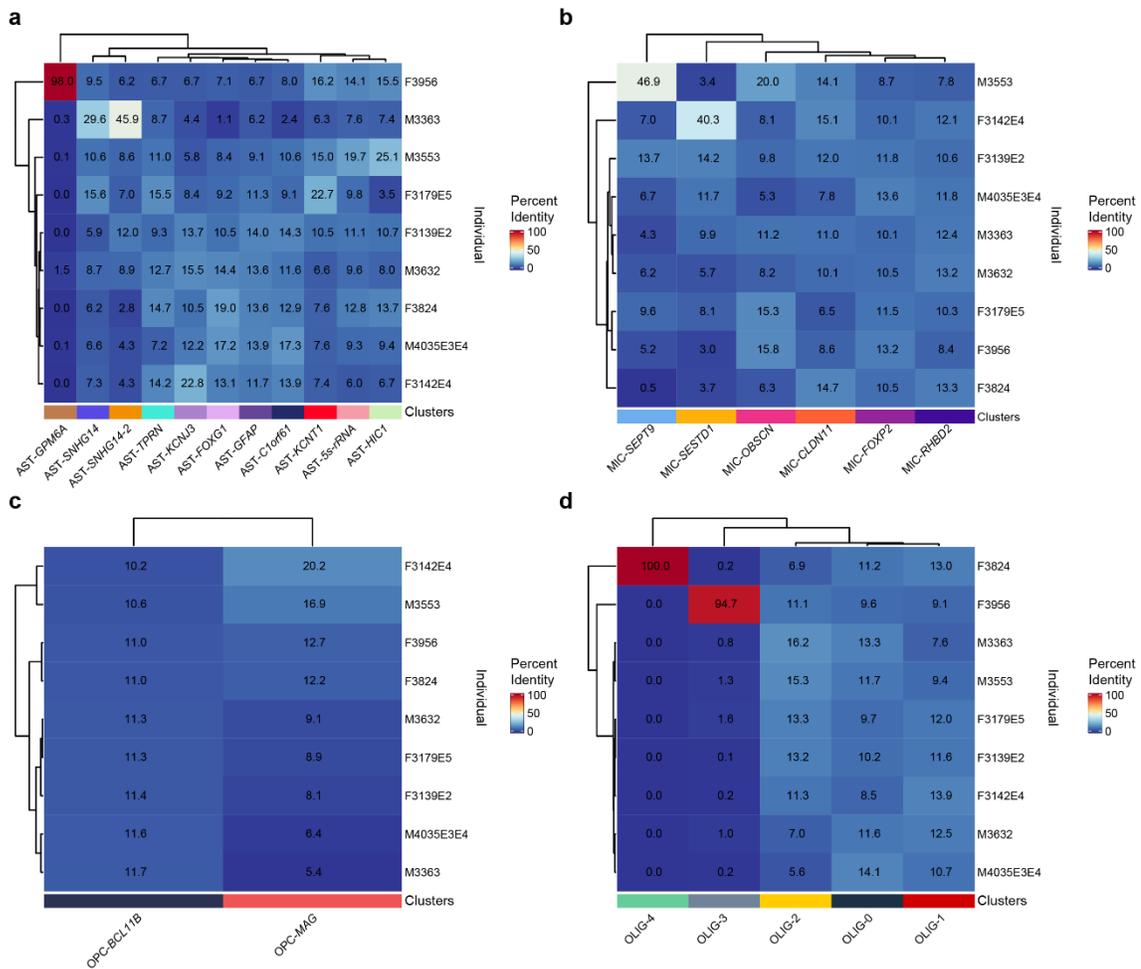
Figure 27. Percent identity of each individual per cluster. Cell counts normalized per individual to account for total number of cells for each individual. **a**, Astrocytes (n=49,992 cells), **b**, Microglia (n=30,478 cells), **c**, OPCs (n=16,534 cells), **d**, Oligodendrocytes (n=236,680 cells).

3.5.3 Resting state astrocytes vary from ion channel-enriched to blood brain barrier-associated sub-states

Assessment of *GFAP*-low astrocytes revealed two clusters with marker genes and transcription factors that were distinct from the other astrocyte clusters: AST-*KCNT1* and AST-*HIC1* (Fig. 25g,h). AST-*KCNT1* gene activity revealed strong enrichment of genes associated with the synaptic function (Fig. 25g). Top gene markers encoding ionotropic channels included *KCNT1*, a sodium-activated potassium channel; *CACNA1H*, a voltage-gated calcium channel; *SCNN1D*, a subunit of the sodium channel ENaC; as well as receptors *GRIN1*, a glutamate NMDA receptor; and *TBXA2R*, thromboxane A2

receptor[155]. Additionally, AST-*KCNT1* was marked by genes involved in cell signaling, such as *CABP2*, a calcium-binding protein; *PLPP3*, a modulator of bioactive phospholipids involved in astrocytic paracrine signaling[156]. These markers have been previously identified in astrocytes and are associated with synaptic regulation[157–159]. Specifically, recent work has shown that *SCNN1D* is expressed in a subset of fibrous astrocytes[160], and *PLPP3* is highly expressed in low-*GFAP* expressing astrocytes[156]. Together, these markers delineate AST-*KCNT1* as a *GFAP*-low astrocytic subtype that has elevated expression of synaptic signaling channels, receptors, and mediators.

In AST-*KCNT1*, we also found enrichment of transcription factor motifs REST and NEUROD1 (Fig. 25h). The transcription factor REST binds to NEUROD1/4 loci and has been shown to be highly expressed in a subset of astrocytes actively signaling with neurons[161,162]. Astrocytes are already known to be essential for the maintenance of synapses, as calcium-dependent release of D-serine from astrocytes is essential for long-term potentiation of synapses[163]. Additionally, recent work assessing transcriptomic heterogeneity of astrocytes identified a subset of ionotropically active astrocytes that aligns with our characterization of AST-*KCNT1*[164], thus compounding evidence that astrocytes may have an ionotropically active sub-state in at least the cortex and hippocampus.

After non-NeuN enrichment of cortex and hippocampus tissues, we expected to identify cell types that support the blood-brain barrier, such as astrocytes, as well as sparse non-astrocytic cell types such as pericytes, vascular smooth muscle cells (VSMCs), and endothelial cells, which form the brain vasculature. Analysis of cluster AST-*HIC1* revealed diagnostic marker genes of mural cells, *TBX18* and *PDGFB*, which include pericytes and VSMCs (Fig. 25g, Fig. 26c,d)[165–167]. Inspection of endothelial identity revealed no AST-*HIC1* specificity of the endothelial cell marker *FLT1*, though a small

subset of AST-*HIC1* cells did display elevated chromatin accessibility at the promoter of

endothelial cell marker *ICAM2* (Fig. 28e,f). Transcription factor motif enrichment (TFME)

analysis of all astrocytes revealed AST-*HIC1* to be uniquely enriched for E26

transformation-specific (ETS) family transcription factors, which are involved in vascular

development (Fig. 25h)[168]. Transcription factor ETS-1 has been shown to be associated

with endothelial cells in angiogenesis[169]. Our data reveal the similarities between

vascular cells and astrocytes in the healthy human cortex and hippocampus, as well as

the distinct differences highlighted by the lack of shared transcription factors and gene

markers.
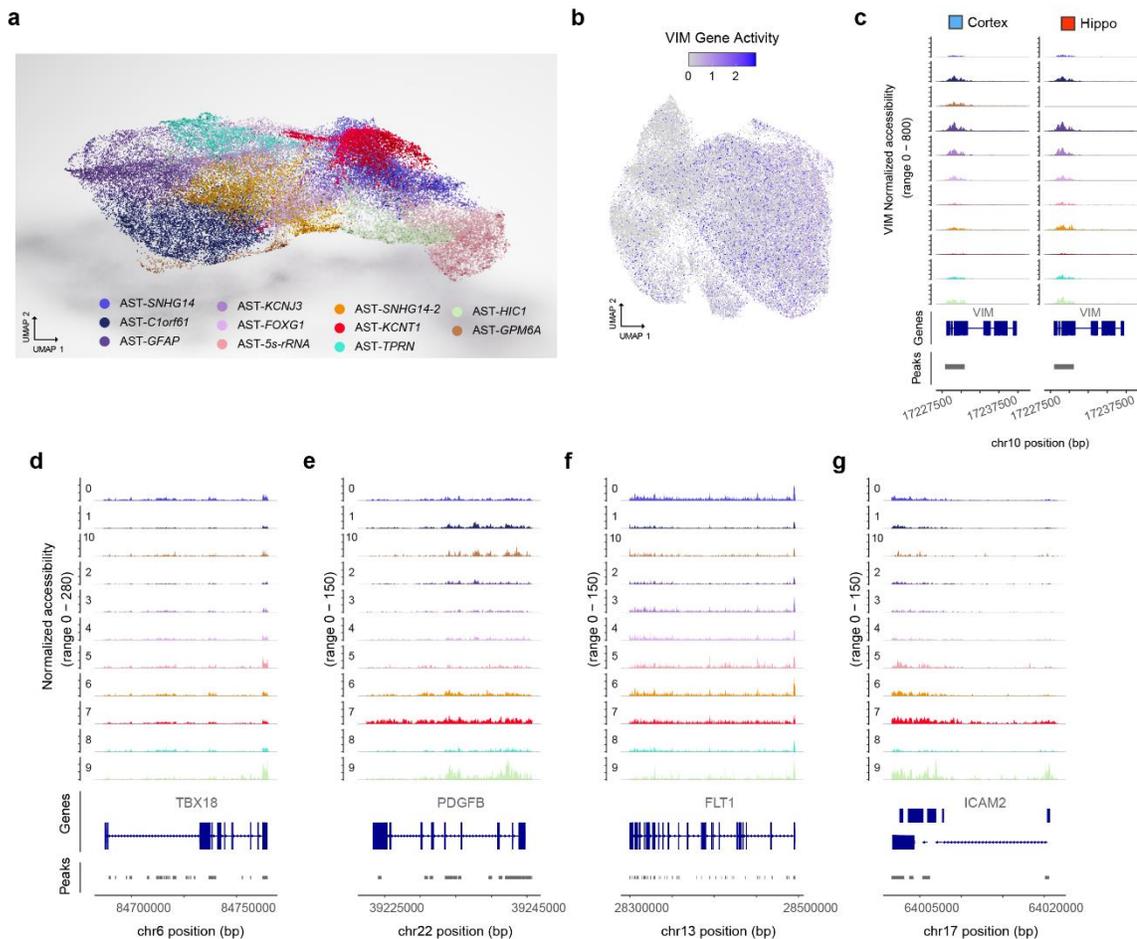
Figure 28. Feature of astrocytic heterogeneity. a, 3D UMAP of combined astrocytes colored by clusters as in Fig. 25c (n=49,992 cells). **b,** 2D combined astrocyte UMAP colored by vimentin, *VIM*, gene activity. **c,** ATAC-seq profiles of astrocyte marker gene *VIM* of astrocytes from cortex (left) and hippocampus (right) split by cluster, as in a (n=49,992 cells). **d-g,** ATAC-

seq profiles of combined astrocytes for mural cell genes *TBX18* and *PDGFB;* and endothelial genes *FLT1* and *ICAM2,* split by cluster, as in a.

3.5.4 Multiple reactive astrocytic states associated with disease pathologies are present in the healthy human CNS

In *GFAP*-high clusters, we identified key features of reactive astrocytic states. Although reactive astrocytic states are likely to be as varied as the contexts that trigger them, previous work has described two such states, A1 and A2, which are specifically induced by lipopolysaccharide (LPS) injection and middle cerebral artery occlusion (MCAO), respectively[170]. While both states show elevated *GFAP*, the A1 state is characterized by increased immune activity induced by reactive microglia, while the less characterized A2 state is marked by increased metabolism and cell-cycling[143,170,171]. In our *GFAP*-high clusters, we identified markers of the A1 state in cluster AST-*GFAP* and the A2 state in clusters AST-*C1orf61* and AST-*KCNJ3*. We found that the top marker genes of AST-*GFAP* included not only *GFAP,* but also *PLEC*, the gene that encodes plectin, a cytoskeleton protein that aids in remodeling of intermediate filaments, including *GFAP*, to morphologically alter astrocytes for glial scar or, more accurately, astrocyte border formation (Fig. 25g)[172–175]. Additionally, elevated plectin mRNA is associated with increased mobility in A2 astrocytes. Similarly, other top marker genes of AST-*GFAP* included *APC2*, which encodes a cytoskeletal protein required for intermediate filament reorganization in astrocyte migration[176], and *SEPT9*, which is a key marker of the reactive amoeboid microglial cells (AMC) and also facilitates microtubule-dependent transport[177,178].

The remaining *GFAP*-high clusters, AST-*C1orf61* and AST-*KCNJ3*, exhibited gene markers that have been implicated in A2 state astrogliosis. A top marker gene of AST-*C1orf61* is *SEMA4B,* which regulates astrogliosis post-brain injury[179] and is associated with synaptogenesis after MCAO[164]. AST-*KCNJ3* is marked by *ADAMTS1* and

ADAMTS3, which are both upregulated following MCAO[180,181]. Together, we found that the top marker genes for *GFAP*-high clusters are associated with known reactive astrocyte states associated with neuroinflammation and cerebral ischemic events. Strikingly, we found these reactive astrocytes in healthy adult human brain tissue – leading us to consider that reactive astrocytic states may be features of normal aging and physiology[153].

3.5.5 Microglia occupy multiple reactive amoeboid and resting ramified microglial chromatin sub-states

The other primary reactive glial cell type in the brain is microglia, which arise from primitive myeloid precursors in the yolk sac and infiltrate the brain early in embryonic development[182]. Microglia are the resident immune cells of the CNS, resembling peripheral macrophages in function and adopting reactive states in response to inflammation[145,183]. The cell type annotation of the cortex and hippocampus datasets revealed 30,478 microglia, identified by elevated chromatin accessibility of canonical markers *C1QA*, *C1QC*, and *CX3CR1* (41.1% cortex, 58.9% hippocampus, Fig. 29a,b). Clustering of the Harmony-weighted microglia revealed stratification across six clusters labeled by their top marker gene: MIC-*RHBD2*, MIC-*FOXP2*, MIC-*OBSCN*, MIC-*CLDN11*, MIC-*SEPT9*, MIC-*SESTD1* (Fig. 29c, left), distributed evenly across individuals, except in the cases of cluster MIC-*CLDN11*, which was overrepresented for individual M3353 (46.9%) and cluster MIC-*SESTD1*, which was overrepresented for individual F3142 (40.3%, Fig. 27b). Plotting the six microglial clusters on the regional cortex and hippocampus UMAPs revealed MIC-*SESTD1* to be embedded within the hippocampal astrocytes (Fig. 29c, right). Investigation into the microglial identity of MIC-*SESTD1* and the astrocyte clusters confirmed that MIC-*SESTD1* is highly accessible for the diagnostic microglial marker *C1QC*, while all annotated astrocytes lacked *C1QC*

accessibility (Fig. 29d), thus confirming the microglial identity of cluster MIC-*SESTD1*.

Similarly, the top marker gene of cluster MIC-*CLDN11*, *CLDN11,* is considered an

oligodendrocyte marker gene. However, we found high chromatin accessibility for

microglia marker *C1QC* in MIC-*CLDN11* and a lack of *C1QC* accessibility across all

oligodendrocyte clusters, thus also confirming the microglial identity of MIC-*CLDN11*.

Claudin-like (CLDN) family transmembrane tight junction genes have been found to be

upregulated in microglia in systemic inflammation[184] and, therefore, may represent a

sparse *CLDN11*-expressing microglial population.
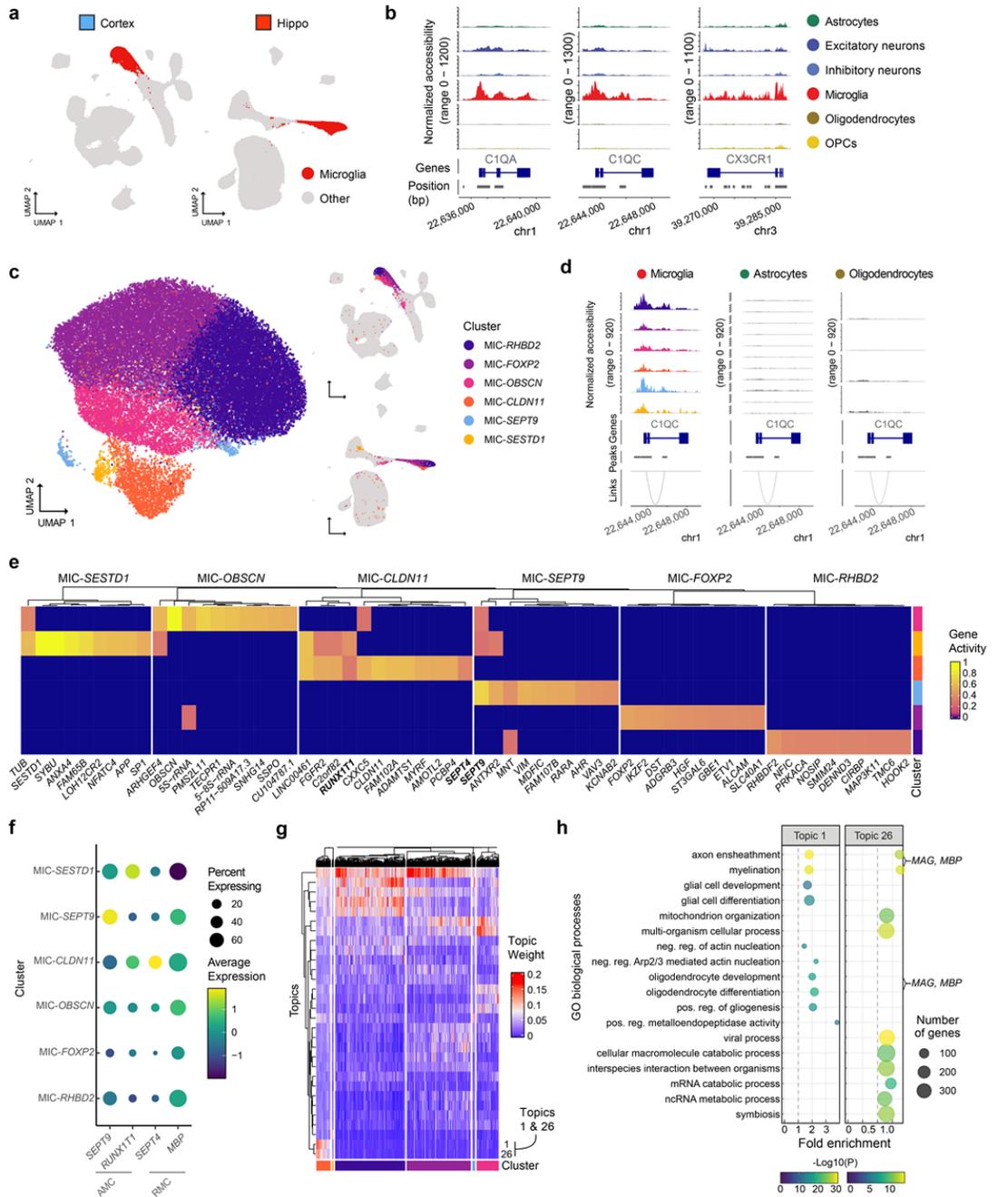
Figure 29. Microglia clusters exhibit features of resting RMC and reactive AMC states. a, Cortex (left) and hippocampus (right) UMAPs with microglia annotated cells highlighted (red) in contrast to all cells (grey). **b,** ATAC-seq profiles of microglia marker genes *C1QA, C1QC,* and *CX3CR1* of all cells from cortex hippocampus (n=378,270 cells). **c,** UMAP of the combined cortex and hippocampus microglia colored by cluster (left, n= 30,478), all cortex cells colored by microglia clusters (top right, n=176,696 cells), all hippocampus cells colored by microglia clusters (bottom right, n=201,574 cells). **d,** ATAC-seq profiles of microglia marker genes *C1QA* (left) and *C1QC* (right) with tracks split by microglia and astrocyte clusters, colored by microglia clusters, as in c (n=30,478 microglia cells, n=49,992 astrocyte cells). **e,** Heatmap of gene activities for the top gene markers identified for each microglia cluster. All values are available in Supplementary Table 4. **f,** Dot plot of gene activity of two AMC-specific genes (*RUNX1T1*

and *SEPT9*) and two RMC-specific genes (*SEPT4* and *MBP*) split by microglia cluster, as in c. Dot size scaled by percent of cells in cluster with gene activity expression. **g,** Harmony-weighted cisTopic matrix split by microglia cluster (n=30,78 cells). **h,** Dot plot of GO term enrichment for microglia cluster MIC-*CLDN11* enriched topics 1 and 26. Dot size scaled to number of genes in the test set with the annotation and colored by GO term -log10 GO term Bonferroni corrected p-values. Select genes highlighted in text, *MAG* and *MBP*, underlying GO term. All GO enrichment values available in Supplementary Table 5.

Previous work has found that microglia stratify into two morphological states: resting ramified microglial cells (RMC) and reactive amoeboid microglial cells (AMC), which are thought to correspond to quiescent sensing activity and macrophage-like phagocytic activity, respectively[177,185–187]. Although this binomial stratification is likely to be a crude underestimation of microglial morphology and diversity[145], nonetheless, it serves as a useful compass to probe our dataset for key genes associated with each category. In the present study, we found that identified microglia clusters were differentially enriched for AMC- and RMC-associated genes identified in a previous AMC/RMC transcriptome analysis[177]. Notably, we found that *SEPT9*, a marker of AMC microglia, was found to be the top marker gene of MIC-*SEPT9* and was enriched in clusters MIC-*FOXP2* and MIC-*SESTD1*, while *SEPT4*, a marker of RMC microglia, was enriched in cluster MIC-*OBCSN* (Fig. 29e). To assess the AMC versus RMC identity of the microglia clusters, we assessed the percentage of cells expressing key markers of the two states, AMC: *SEPT9* and *RUNX1T1*, and RMC: *SEPT4* and *MBP.* MIC-*SEPT9* was significantly enriched for the AMC marker *SEPT9* in a majority of cells, while MIC-*SESTD1* was uniquely enriched for AMC marker *RUNX1T1* (Fig. 29f). This analysis provides single-cell context to previous bulk, morphologically driven microglial studies by revealing that AMC and RMC microglia do not exist in two homogenous states. Our findings demonstrate two AMC states, clusters MIC-*SEPT9* and MIC-*SESTD1*, and three RMC states, MIC-*RHBD2*, MIC-*FOXP2*, and MIC-*OBSCN*, which are shared across all individuals.

Notably, MIC-*CLDN11* exhibited elevated gene activity for both AMC and RMC markers. Investigation into unique features of MIC-*CLDN11* revealed that two Harmony-weighted topics were specifically enriched in MIC-*CLDN11*, topics 1 and 26 (Fig. 29g). We hypothesized that MIC-*CLDN11* might occupy an intermediate AMC/RMC state that shares chromatin accessibility from both morphologies. To test this hypothesis, we performed a gene ontology (GO) enrichment analysis (Fig. 29h). We found that topics 1 and 26 were associated with biological processes, including those related to oligodendrocytes (axon ensheathment, myelination), viral processes, and gliogenesis. Previous transcriptomic analysis of RMC microglia has found that canonically oligodendrocyte-associated genes, *MOBP*, *MOG*, *MBP*, *MAG*, and *OPALIN* have elevated expression in RMC as opposed to AMC microglia (7.99-623.18 fold change)[177]. In our topic analysis, we found that *MAG* and *MBP* were two of the genes leading to the oligodendrocyte-associated GO terms, thus establishing the RMC-like chromatin state of MIC-*CLDN11.*

3.5.6 Oligodendrocyte progenitor cells occupy a gradient of cellular states toward differentiation

OPCs are a multipotent cell type in the human brain that serve multiple functions as precursors of oligodendrocytes. However, OPCs are also capable of differentiating into other cell types and are therefore also referred to as polydendrocytes[188]. In the human brain, OPCs persist throughout adulthood and play a multifaceted role beyond oligodendrocyte generation: they are electrically active cells that form synapses with neurons, support the blood-brain barrier, and may even play an immune role[189].

In our study, we identified 16,534 OPCs in the cortical and hippocampal datasets from defined clusters with elevated accessibility of the markers *OLIG1* and from integration with public scRNA-seq and scATAC-seq datasets (47.4% cortex, 52.6 % hippocampus,

Fig. 31a, Fig. 22). We found that the number of OPCs was comparable between

individuals with minimal variation by age (1837.1 ± 367.9 cells, mean ± s.e., Fig. 30a).

As in previous cell type analyses, we selected cell profiles for OPCs from both the cortex

and hippocampus and then performed topic-based dimensionality reduction and UMAP

visualization (Fig. 31b). While all other glial cell types were separated in the UMAP

space by the individual, the combined cortex and hippocampus OPC UMAP did not

exhibit a detectable, individual-based batch-effect between clusters (Fig. 24a, Fig. 27c).

Given that individuals were processed in multiple experiments, this lack of batch-effect

indicated that OPCs did not exhibit the same degree of person-to-person-based variation

that other glial cell types (astrocytes, microglia, and oligodendrocytes) exhibited (Fig.

24b,d,f). This may be due to the undifferentiated progenitor state of the OPCs, where the

OPC progenitor pool is more homogenous than the other mature differentiated glial cell

types. Therefore, we did not use Harmony batch-correction on the combined OPC

dataset and performed analysis on the unweighted cell by topic matrix.
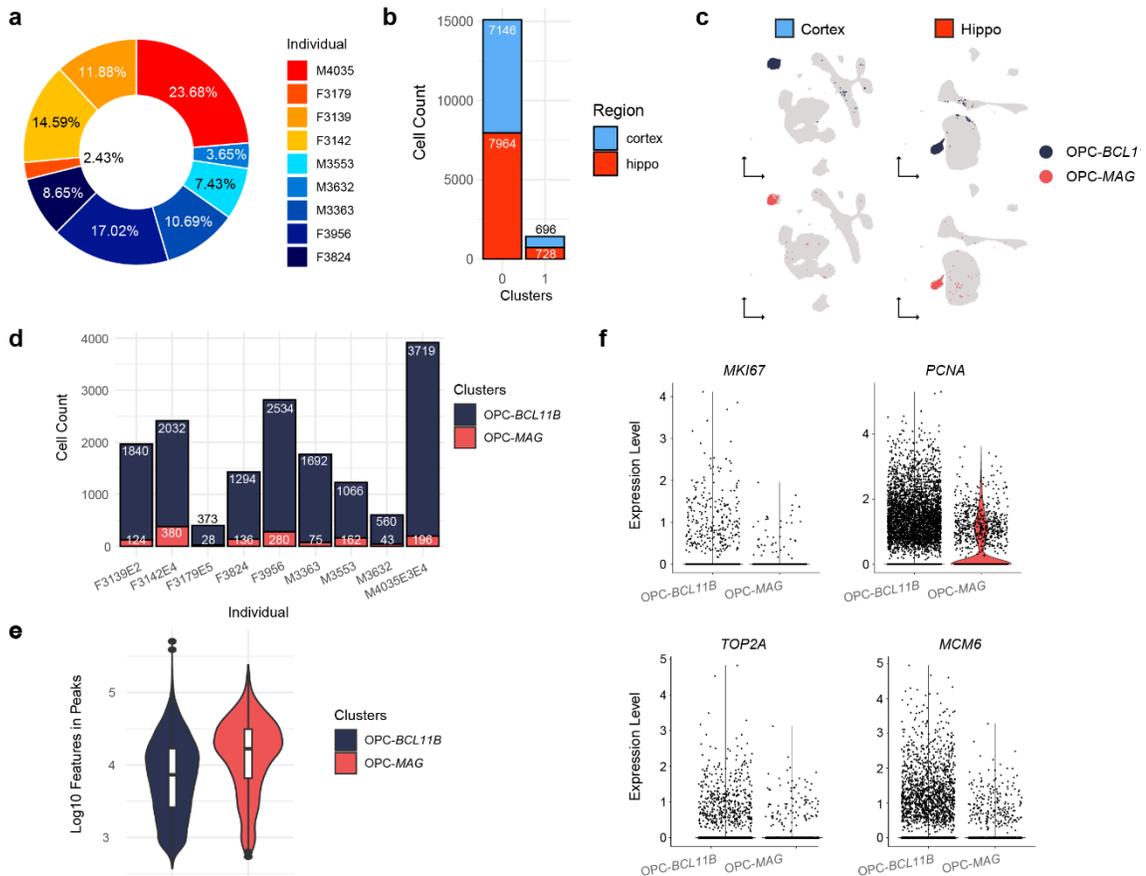
Figure 30. Features of OPC heterogeneity. a, Percent OPC cell count per individual patient. **b,** Bar chart of OPC cell count per brain region. **c,** OPC clusters plotted on regional cortex (left) and hippocampus (right) UMAPs. **d,** OPC cluster cell count across all nine individuals. **e,** Violin plot of Log10 Features in peaks between clusters OPC-*BCL11B* (right, blue) and OPC-*MAG* (right, coral) **f,** Gene activity violin plot of cell cycle marker genes: *MKI67*, *PCNA*, *TOP2A*, *MCM6*.

From the 16,534 OPCs derived from both regions, we identified two clusters labeled by their top marker gene: OPC-*BCL11B* and OPC-*MAG*, which did not coincide with a region (cortex or hippocampus) or localization within the cortex and hippocampus UMAPs (Fig. 30b,c). Visualizing these clusters in 3D revealed a previously hidden spatial stratification between the two clusters (Fig. 31c). We found that 91.4% of OPC cells fell into cluster OPC-*BCL11B* (7,146 cortex; 7,964 hippocampus; 15,110 total cells) and 8.6% into OPC-*MAG* (696 cortex; 728 hippocampus; 1,424 total cells). The percent of cells belonging to each cluster was also consistent across all nine individuals (OPC-*BCL11B*: 91.3 ± 1.3%; OPC-*MAG*: 8.7 ± 1.3%, mean ± s.e., Fig. 30d) and OPC-*BCL11B*

and OPC-*MAG* did not stratify by the quality metric of the number of features of peaks per cell (Fig. 30e).
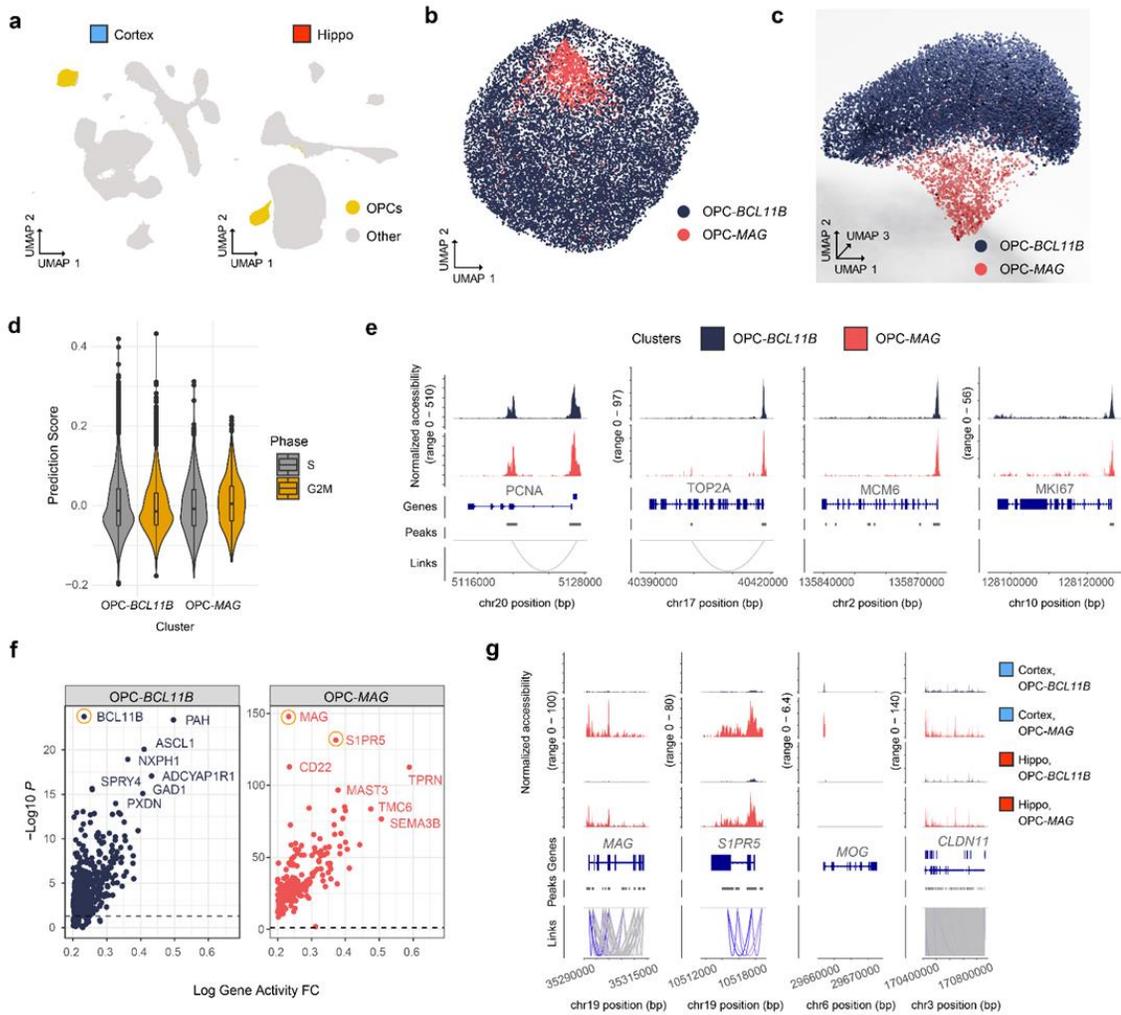
Figure 31. OPC clusters are differentially enriched for oligodendrocyte maturation genes. a, Cortex (left) and hippocampus (right) UMAPs with OPCs annotated cells highlighted (yellow) in contrast to all cells (grey). **b,** Combined OPC UMAP of 16,534 cells colored by cluster identity. **c,** 3D combined OPC UMAP plotted along UMAP1 and UMAP3 axes, colored by OPC cluster, as in b (n=16,534 cells). **d,** Violin plot of S (grey) and G2M (yellow) cell cycle phase gene set enrichment score for each OPC cluster, as in b. **e,** ATAC-seq profiles of cell cycle marker genes *PCNA*, *TOP2A*, *MCM6*, and *MKI67* split by OPC clusters. **f,** Scatter plot of gene activity fold change by –log10 P value between OPC-*BCL11B* (left, blue) and OPC-*MAG* (right, coral). Dotted line represents P vale threshold of 0.05. Each point represents a unique gene. Marker genes highlighted in the text, *BCL11B, MAG, S1PR5*, circled in yellow. **g,** ATAC-seq profiles of oligodendrolineage differentiation marker genes *S1PR5* and *MAG*; and mature oligodendrocyte markers *MOG* and *CLDN11*, between OPC-*BCL11B* and OPC-*MAG* clusters and brain region: cortex and hippocampus.

111

A previous single-cell transcriptomic study that assessed glial heterogeneity in the adult mouse brain identified two clusters of OPCs that were hypothesized to correspond to cycling and non-cycling cells[37]. To address this hypothesis in our data, we assessed our combined OPCs for the enrichment of genes associated with cell cycling. We found that our clusters were not defined by any cell cycle phase (Fig. 31d). Assessment of key cell cycle markers such as *MKI67*, *PCNA*, *TOP2A*, and *MCM6* also revealed no significant cluster bias in read coverage (Fig. 31e) or in gene activity (Fig. 30f)[190]. Additionally, our sci-ATAC-seq FANS gating strategy stratifies nuclei by DAPI-stained DNA content, thus allowing for the selection of quiescent diploid nuclei in G_0 phase. As a result, we did not find that clusters OPC-*BCL11B* and OPC-*MAG* were defined by cell cycle stages.

To identify features that stratify clusters OPC-*BCL11B* and OPC-*MAG*, we assessed differential gene activities between them. Markers known to be associated with OPC-to-oligodendrocyte maturation were found to be significantly enriched in OPC-*MAG*. The two most significantly enriched markers, *S1PR5* and *MAG,* are both associated with pro-oligodendrocyte OPC differentiation (Fig. 31f). The expression levels of sphingosine-1-phosphate receptor (S1PRs) 1, 2, 3, and 5 are modulated in OPCs by PDGF, a canonical OPC marker, which titers S1PR expression in different stages of oligodendrolineage cell maturation[191]. *S1PR5* becomes the most highly expressed *S1PR* as OPCs differentiate into oligodendrocytes[191–193]. Additionally, *MAG*, or myelin-associated glycoprotein, is expressed in pro-oligodendrocyte OPCs (Fig. 31f)[194]. We found that both *S1PR5* and *MAG* had elevated chromatin accessibility in OPC-*MAG* in both the cortex and hippocampus. However, mature oligodendrocyte markers *MOG* and *CLDN11* were not comparably accessible (Fig. 31g). This marks OPC-*MAG* as a post-OPC and pre-oligodendrocyte cellular state. In contrast, cluster OPC-*BCL11B*, which contained the majority of OPCs, lacked oligodendrocyte-associated gene activity. The

top marker gene, *BCL11B* (*CTIP2*), is associated with many cell types throughout the brain, primarily GABAergic neurons and astrocytes; however, *BCL11B* expression has also been shown to promote oligodendrolineage progression[120]. We found *BCL11B* to be suppressed upon OPC-to-oligodendrocyte differentiation between OPC-*BCL11B* and OPC-*MAG* (Fig. 31g). Therefore, our analysis identified cluster OPC-*BCL11B* as an undifferentiated progenitor state and OPC-*MAG* as a differentiation-committed oligodendrocyte precursor (COP) population[39]. Taken together, we demonstrate that, with sufficient cellular sampling, it is possible to resolve distinct OPC cellular states in the adult human brain that correspond to the genomic reorganization accompanying oligodendrolineage differentiation.

3.5.7 Oligodendrocytes occupy a continuous chromatin accessibility landscape with few distinct sub-states

Oligodendrocytes are the glial cell type in the CNS that produce myelin, which surrounds neuronal axon tracts and allows for saltatory conduction of action potentials. This myelin production remains plastic throughout the lifetime of the cell and is essential for learning and memory[195]. Additionally, oligodendrocytes provide metabolic support to the neuronal processes they ensheath and also regulate neuronal axon sprouting and synaptogenesis[196]. Investigations into spatially and functionally mediated oligodendrocyte heterogeneity have revealed distinctions between oligodendrocytes that are hypothesized to be driven by extrinsic signals such as synaptic activity and perturbations in disease states, such as multiple sclerosis (MS), as opposed to multiple differentiation trajectories[38,54]. Here, we sought to robustly assess the heterogeneity of oligodendrocyte chromatin accessibility in the unperturbed, healthy, adult human brain with high cell numbers in order to define the baseline of oligodendrocyte heterogeneity in the cortex and hippocampus.
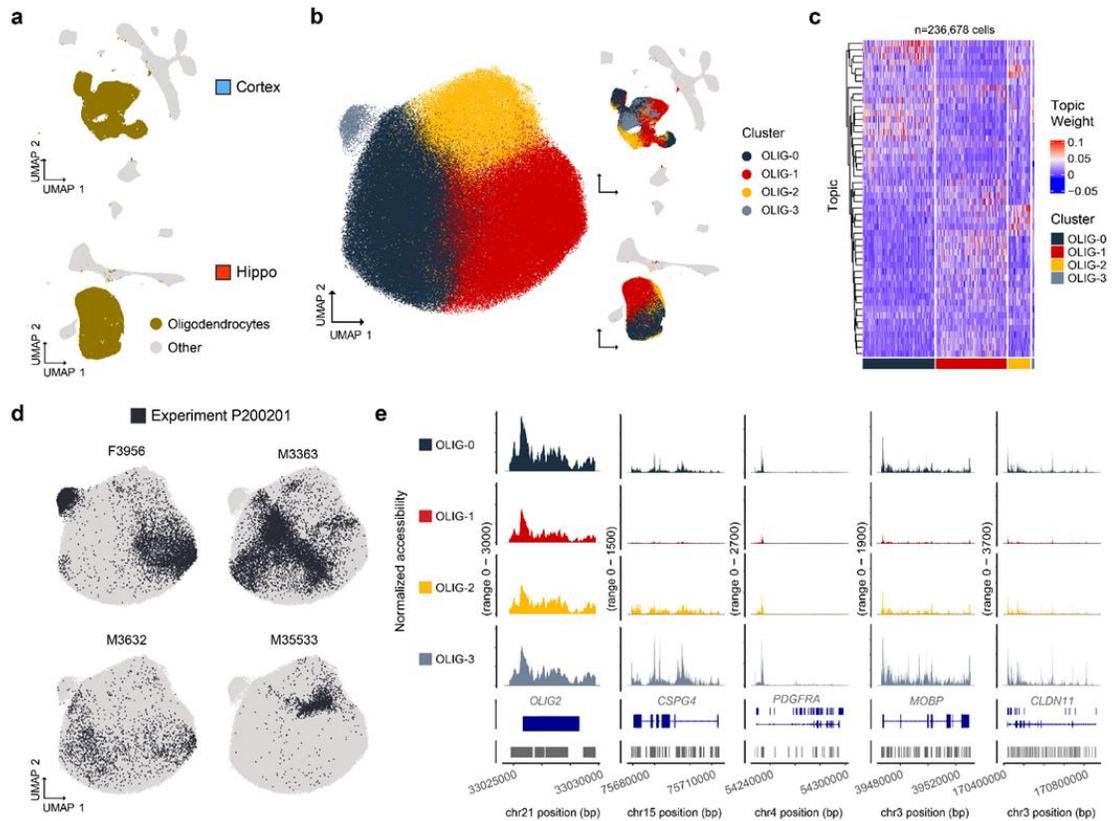
Figure 32. Oligodendrocytes occupy largely homogenous chromatin state. a, Cortex (top) and hippocampus (bottom) UMAPs with oligodendrocyte annotated cells highlighted (tan) in contrast to all cells (grey). **b,** UMAP of combined cortex and hippocampus oligodendrocytes colored by cluster (n=236,6878 cells). Cluster OLIG-4 excluded (n=2 cells). **c,** Harmony-weight cisTopic matrix split by oligodendrocyte clusters, as in b (n=236,678 cells). **d,** Combined oligodendrocyte UMAP with cells derived from experiment P200201 colored in black, split by individual patient cells from experiment P200201. **e,** ATAC-seq profiles of oligodendrolineage markers genes: *OLIG2, CSPG4, PDGFRA, MOBP* and *CLDN11* split by oligodendrocyte clusters (n=236,678 cells)*.*

We identified 236,680 oligodendrocytes in the cortical and hippocampal datasets from

defined clusters with elevated accessibility of the markers *OLIG1*, *OLIG2* and from

integration with public scRNA-seq and scATAC-seq datasets (44.3% cortex, 55.7%

hippocampus, Fig. 32a, Fig. 22). Clustering of the Harmony-weighted oligodendrocytes

stratified the cells into five clusters (OLIG-0-4, Fig. 32b); however, we found no

significantly enriched marker genes for any cluster. OLIG-4 contained only two cells,

8.4e-4% of oligodendrocytes, from one individual (F3824) and, as a result, was excluded

from future analyses. Notably, the bulk of oligodendrocytes occupied a single large

grouping, clusters OLIG-0-2, and a second small grouping, cluster OLIG-3 (3170 cells, 1.34% of oligodendrocytes). Assessment of the Harmony-weighted cell by topic matrix across the clusters revealed no clearly defined topic enrichments between clusters, implying largely uniform chromatin accessibility across all oligodendrocytes (Fig. 32c).

Mapping of the satellite cluster, OLIG-3, onto the cortex and hippocampus regional UMAPs revealed clustering of OLIG-3 cells within the cortex and hippocampus oligodendrocyte clusters and not in any other cell type (Fig. 32b, right). Upon investigation, we found OLIG-3 to be comprised of cells derived from one experiment (94% experiment P200201) and from one of the three individuals in experiment P200201, F3956, comprising 95% of the cells in OLIG-3 (Fig. 32d). OLIG-3 did not, however, represent all cells from individual F3956. In contrast, cells from individual F3956 were stratified across all clusters (35.1% OLIG-0, 34.9% OLIG-1, 15.2% OLIG-2, and 14.8% OLIG-3, Fig. 32d). Therefore, OLIG-3 represented a single-individual oligodendrocyte population that did not correspond with any experimental batch. Assessment of canonical oligodendrocyte lineage markers in our dataset revealed accessibility of pan-oligodendrocyte marker *OLIG2* across all clusters (Fig. 32e). Interestingly, OLIG-3 had elevated chromatin accessibility of oligodendrocyte progenitor markers *CSPG4* (*NG2*) and *PDGFRA* in addition to mature oligodendrocyte markers *MOG* and *CLDN11*, in comparison to other oligodendrocyte clusters (Fig. 32f). Additionally, we found elevated *TOP2A* chromatin accessibility in OLIG-3, which is associated with cell proliferation (Fig. 33e). Due to OLIG-3 representing a single sampling of one individual, these cells may represent a patient-specific functional, spatial, or disease-associated state.
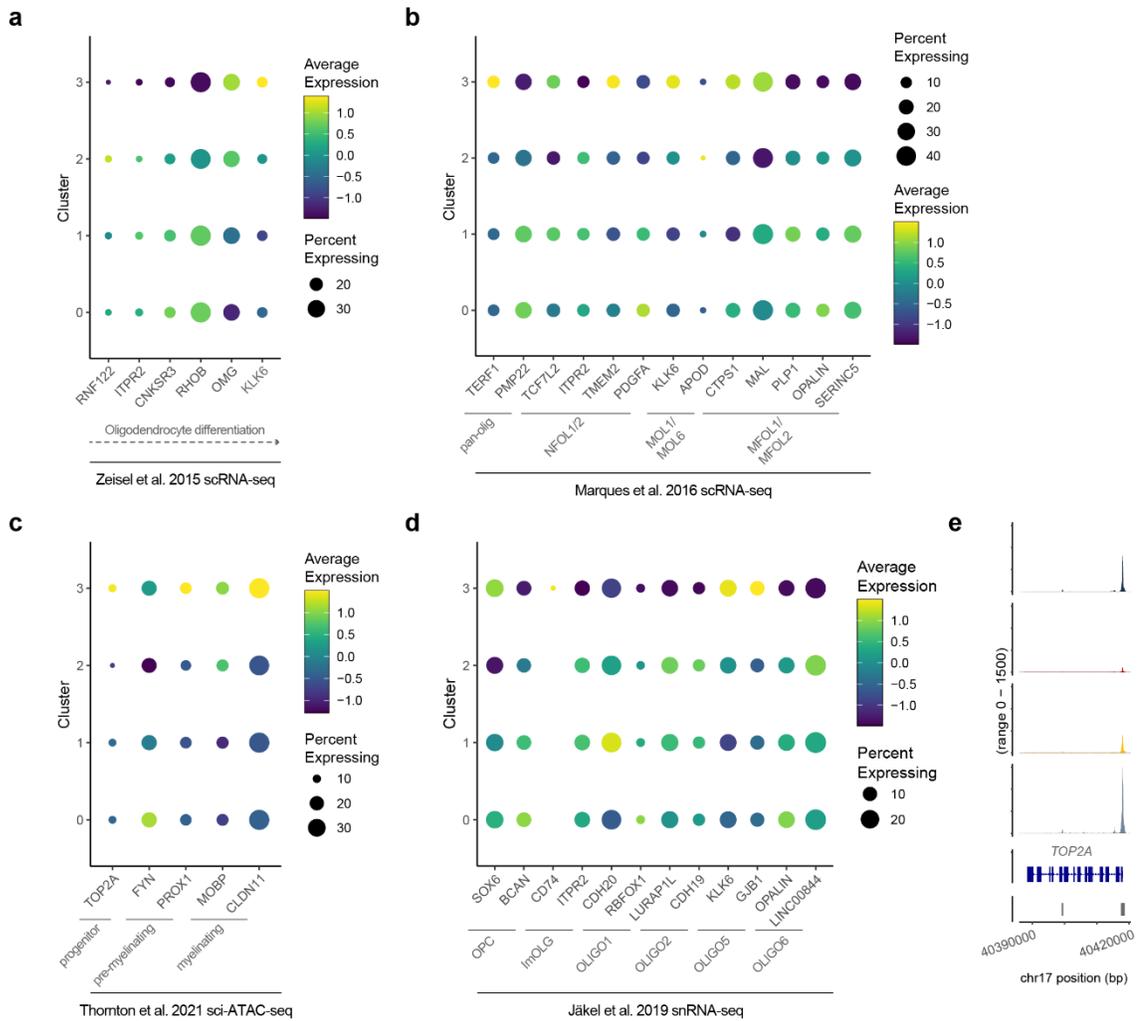
Figure 33. Comparison of oligodendrocyte ATAC profiles with previous oligodendrocyte characterizations. a, Dot plot of oligodendrolineage maturation markers derived from Zeisel *et al.* scRNA-seq in adult mouse, split by oligodendrocyte clusters, as in Fig. 32b. Markers progress from OPC (left) to myelinating oligodendrocyte (right). **b**, Dot plot of oligodendrolineage maturation markers derived from Marques *et al.* scRNA-seq in adult mouse, split by oligodendrocyte clusters, as in Fig. 32b. Markers progress from pan-oligodendrocytes markers (left) to clusters defined in Marquez *et al.* NFOL1/2 = Newly Formed Oligodendrocyte clusters 1 and 2, MOL1/6 = Mature Oligodendrocyte clusters 1 and 6, MFOL1/2 = Myelin Forming Oligodendrocytes clusters 1 and 2. **c**, Dot plot of oligodendrolineage maturation markers used in Thornton *et al.* sci-ATAC-seq in adult mouse, split by oligodendrocyte clusters, as in Fig. 32b. **d**, Dot plot of oligodendrolineage maturation markers derived from Jäkel *et al.* scRNA-seq in adult human oligodendrocytes derived from control and MS white matter tissue, split by oligodendrocyte clusters, as in Fig. 32b. Markers progress from OPC (left) to clusters defined in Jäkel *et al.* ImOLG = Immune oligodendroglia, OLIG1,2,5,6 = Numbered oligodendrocyte clusters. **e**, ATAC-seq profiles of combined oligodendrocytes for cell cycling marker *TOP2A,* split by cluster, as in Fig. 32b.

Given that no genes were significantly differentially active between any of the

oligodendrocyte clusters, further establishing the homogeneity of oligodendrocytes in

116

both the cortex and hippocampus, we compared our gene activities to previously identified oligodendrocyte sub-population markers. Previously, single-cell transcriptomic studies have revealed that adult mouse oligodendrocytes robustly stratify by oligodendrocyte maturation stage[37,39,50,197]. However, this robust stratification is not similarly observed in adult humans[38,147]. We confirm this in our human database as we found no clear stratification of oligodendrocytes gene activity by maturation marker (Fig. 33a-c). In human white matter oligodendrocytes, a previous study by Jäkel *et al.* found healthy- and MS-associated sub-populations[38]. We compared gene activities of our oligodendrocyte clusters with Jäkel *et al.* markers, but we found low concordance (Fig. 33b). Taken together, we demonstrate that the homogenous state of oligodendrocytes is in contrast to other glial cell types assayed in this study.

3.6 Discussion

In this study, we demonstrate the utility of single-cell chromatin accessibility profiling of cells in characterizing intra-cell type diversity, and we provide a curated dataset that provides additional context to ongoing investigations of glial heterogeneity from an epigenetic perspective. These datasets are available in an interactive form at the UCSC genome browser that includes downloadable analysis objects, as well as in raw form via the Neuroscience Multi-omic Archive (NeMO).

In astrocytes, we found eleven sub-types stratified by *GFAP* accessibility, a marker associated with astrocytic reactivity. Bifurcating clusters into *GFAP*-high and *GFAP*-low clusters revealed five *GFAP*-high clusters, including cells from multiple individuals, with features previously found to be associated with neuroinflammation pathologies such as cerebral ischemia and multiple sclerosis. However, we found disease-associated characteristics in the healthy adult human CNS, thus establishing the natural diversity of astrocytes along a gradient of reactivity. Additionally, characterization of the six *GFAP*-

low clusters revealed functionally distinct resting astrocyte clusters, ranging from astrocytes enriched for ion transport channels to those enriched for BBB support. Furthermore, future analyses of astrocytes in diseased human brains are likely to expand upon the classifications described in this study, revealing increasingly diverse reactive astrocyte states, as dictated by the specific context.

In microglia, we address long-standing hypotheses associated with resting ramified and reactive amoeboid subtypes. We identified enriched features of ramified and amoeboid microglia in our data and found that in contrast to a bimodal AMC/RMC classification of microglia, we found six clusters, each demonstrating variable features of the subtypes. In this analysis, we demonstrate the utility of single-cell studies to provide additional information to previously proposed cell states.

Within the oligodendroglia lineage, we found a clear delineation between oligodendrocytes and OPCs in the cortex and hippocampus but observed minimal subtype heterogeneity. OPCs occupied a de-differentiated progenitor pool with a small percentage of cells entering differentiation. Oligodendrocytes were found to have minimal chromatin accessibility variation between clusters, in direct contrast to the heterogeneity observed in astrocytes, which were subdivided into functional domains. This homogeneity also highlights the potential differences between genome organization, assayed through chromatin accessibility, and actual transcription, assayed through RNA sequencing, as previous transcriptomic oligodendrocyte sub-classifications were not reproduced in this study.

Our study is, however, limited. In order to account for the individual-based batch effect, integration between tissue from individuals was performed. In addition to eliminating signals that corresponded to batch-effect, this integration eliminates the ability to cluster cells by the individual, as well as by cortical and hippocampal tissues. As a result, we

were unable to assess the heterogeneity of glial cell types in the context of these conditions. Future studies assessing the region-specific features of a cell type such as astrocytes with sufficient numbers of individuals could provide insight into the hypothesis that hippocampal and cortical astrocytes exhibit disparate cellular states.

Additionally, the glial sub-type classifications in this study are descriptive of the observed glial heterogeneity across nine patients, ranging from 37-73 years of age. Glial reactivity increases with age[153,198,199]; however, in the present study, an insufficient number of individuals at distinct ages were assessed to address hypotheses pertinent to this topic. This study provides a foundation for future investigation of the epigenetic landscape of glia in the context of normal aging.

3.7 Methods

3.7.1 Human sample collection

All samples were provided by the Oregon Brain Bank. Five middle frontal gyrus cortex samples were collected from two female, and three male deceased individuals, and four hippocampus samples were collected from three male and one female deceased individual. Individual ages range from 37-73 years old. Samples were collected by an OHSU neuropathologist, placed into a labeled cassette, and cryopreserved in an airtight container in a -80°C freezer. The duration of time between the patient time of death and brain biopsy sample freezing, or post-mortem interim (PMI), ranged from 3-27 hours. We find no significant correlation between PMI and ATAC data quality as assessed by transcription start site enrichment (Figure 21g). Detailed information on each sample is provided in Supplementary Table 3.

3.7.2 Nuclei isolation

We performed nuclei isolation on all fresh-frozen human samples. Approximately 0.1 mg of tissue (2-3 mm^3) was resected from the sample cassette using a razor blade on dry ice. Resected tissue was transferred to 2 ml chilled NIB (10 mM HEPES-KOH, pH 7.2 [Fisher, Cat. BP310-1], 10 mM NaCl [Fisher, Cat. M-11624], 3 mM MgCl$_2$ [Sigma, Cat. M8226], 0.1% IGEPAL [v/v; Sigma, I8896], 0.1% Tween-20 [v/v, Sigma, Cat. P7949], and 1x protease inhibitor [Roche, Cat. 11873580001]) in a 7 ml Dounce-homogenizer on ice. Minor NIB formulation variations were used in two experiments, which utilized 20 mM Tris in place of 10 mM HEPES (see Supplementary Table 3). The tissue was incubated on ice for 5 min. then homogenized via 5 gentle strokes of the loose pestle (A) on ice, a 5 min. incubation on ice, then 5 gentle strokes of the tight pestle (B) on ice. The homogenate was then strained through a 35 µm strainer and centrifuged at 500 × g for 10 min. at 4°C to pellet. Samples were aspirated, resuspended in 1 ml of ice-cold blocking buffer (0.5%. Bovine serum albumin (BSA), 5% normal goat serum [Fisher, Cat. 50062Z], in 1x PBS [VWR, Cat. 75800-992]), and incubated for 15 min. on ice.

3.7.3 Immunolabeling of neuronal nuclei

Neuronal nuclei were labeled using mouse monoclonal anti-NeuN primary antibody (Sigma, Cat. MAB377) conjugated to a fluorescent goat anti-mouse IgG Alexa Fluor 647 secondary antibody. The conjugated staining solution (2X 2 µg/ml MAB377:AF647) was shaken on an orbital shaker at 80 rpm for 10 min. in a light-protected container. Control solutions of primary-only (2X 2 µg/ml MAB377) and secondary-only (2X 2 µg/ml AF647) staining solutions were prepared as negative controls. Nuclei were stained by combining equal volumes of 2X conjugated staining solution and nuclei incubated in blocking buffer. Negative controls were prepared by combining equal volumes of nuclei incubated in blocking buffer with i) antibody-free blocking buffer, ii) primary-only staining solution, and

iii) secondary-only staining solution. Nuclei staining reactions, including negative

controls, were incubated in light-blocking microcentrifuge tubes for 1 hour at 4°C on an

orbital shaker at 80 rpm. Post-incubation, 1 ml of 1x PBS was added to all reactions,

which were then centrifuged at 500 × g for 10 min. at 4°C to pellet. Samples were

aspirated, resuspended in 400 µl of ice-cold NIB, and counterstained using DAPI

(Thermo Fisher, Cat. D1306) for all reactions, excluding the antibody-free staining

reaction negative control, for a final concentration of 5 µg/ml DAPI.

3.7.4 Glial enrichment by FANS

Each sample was fluorescence-activated nuclei sorted (FANS) using BD FACSDiva

software (v8.0.1) on a Sony SH800 FACS machine into two tubes of 500 µl NIB, for

neuronal NeuN$^+$ and glial NeuN$^-$ nuclei populations. DAPI-stained nuclei we used to

define the singlet nuclei population and all negative controls were used to set a

background fluorescence threshold. Nuclei stained with the conjugated primary and

secondary antibody staining solution were sorted into NeuN$^+$ and NeuN$^-$ collection tubes

based on the threshold that divided the bimodal AF647 fluorescence. The NeuN$^-$ nuclei

suspended in NIB were diluted to 500 nuclei/µl to facilitate tagmentation reaction

assembly at 5,000 nuclei per 10 µl of NIB.

3.7.5 Nuclei tagmentation

Transposase catalyzed excision of chromatin accessible regions via tagmentation

results in the addition of unique molecular identifiers (indexes) for each tagmentation

reaction. Uniquely indexed transposase adapter sequences are reported in

Supplementary Table 6. Tagmentation reactions were assembled from 7.5 µl of isolated

nuclei at 500 nuclei/µl and 2.5 µl fresh 4X TAPS-TD buffer (132 mM TAPS pH=8.5, 264

mM potassium acetate, 40 mM magnesium acetate, and 64% dimethylformamide). We

use TAPS (N-[Tris(hydroxymethyl)methyl]-3-aminopropanesulfonic acid) for added

buffering at elevated temperatures[130]. 1 µl of 8 µM loaded indexed synthesized Tn5 transposase was added per well (see Picelli *et al.* for transposase synthesis protocol)[130]. As an alternative to Tn5 synthesis, EZTn5 transposase (https://www.lucigen.com/EZ-Tn5-Transposase) can be purchased commercially, diluted, salt adjusted, and loaded with sci indexes[200]. Each assembled 96-well plate of tagmentation reactions was incubated at 55°C for 15 min. We targeted 50,000-100,000 cells per individual. In total, we used 6-18 96-well tagmentation plates, where each well contains a unique combination of transposase indexes per individual (Supplementary Table 3). A total of 107 tagmentation plates were used across all nine individuals, totaling 10,272 (107 x 96) uniquely indexed wells. We found that NIB-HEPES extended the integrity of the nuclei from frozen brain tissue during tagmentation and FANS as compared to NIB-Tris. To preserve tagmented nuclei for later PCR amplification, we cryopreserved tagmented nuclei at -20°C in a modified Nuclear Freezing Buffer (NFB: 10 mM HEPES-KOH, pH 7.2, [Fisher, Cat. BP310-1] 25% glycerol, 5 mM Mg(OAc)$_2$, 0.1 mM EDTA, 5 mM DTT, 1× protease inhibitor [Roche, Cat. 11873580001] , 1:2500 superasin [Ambion])[201].

3.7.6 Combinatorial indexing via PCR

To lyse nuclei and release bound transposase, PCR plates were prepared beforehand with SDS buffer (0.25 µl BSA [NEB, Cat. B9000S], 0.5 µl 1% SDS, 7.75 µl dH$_2$0 per well) and primers (2.5 µl i5 and 2.5 µl i7 PCR indexing primers at 10 µM). Uniquely indexed PCR primer sequences are reported in Supplementary Table 6. Pooled tagmented nuclei were stained by adding 3 µl of DAPI (5 mg/ml) per 1 ml of sample. Each sample was then FAN sorted using BD FACSDiva software (v8.0.1) on a Sony SH800 FACS machine at 15 events per well per 96-well Tn5 plate (e.g., 88 for 384 indexes) into the prepared 96-well plate(s). Event numbers were selected based on the expected success rate of events as actual cells for a given target cell doublet rate[86]. In total, we used 6-20

96-well PCR plates, where each well contained a unique combination of forward and reverse primers per individual (Supplementary Table 3). In total, 109 PCR plates were used across all nine individuals, totaling 10,464 (109 x 96) uniquely indexed wells. Transposase denaturation was performed by sealing each sorted plate and incubating at 55°C for 15 min. Plates were immediately transferred to ice post-incubation, and 11.5 μl of PCR mix (1 μl Kapa Hi Fi [Roche Diagnostics, cat 07958846001], 5X GC buffer 5 μl, 10mM dNTP 0.75 μl, 0.25 μl Sybr 100x, 5 μl dH_2O) was added to each well. For each experiment, plates were then sealed, and one plate was PCR amplified on a BioRad CFX real-time cycler running CFX Manager (v3.1) software, using the following protocol: 72°C for 5:00, 98°C for 0:30, cycles of (98°C for 0:10, 63°C for 0:30, 72°C for 1:00, plate read, 72°C for 0:10) for 15–22 cycles. PCR plates were transferred to 4°C once all wells reached mid-exponential amplification on average on the CFX. The remaining plates from the same experiment were amplified for the number of cycles determined on the CFX and were transferred to 4°C. Each PCR plate was pooled at 10 μl per well, and DNA libraries were isolated using a QIAquick PCR Purification column. Each pooled PCR plate library was then quantified using a Qubit 2.0 fluorometer, diluted to 4 ng/μl with nuclease-free water, and quantification of library fragment size was performed on an Agilent Tapestation using a D1000 or D5000 tape. Libraries were then sequenced on a NextSeq™ 500 sequencer (Illumina Inc.) running NextSeq500 NCS (v4.0) software loaded within a range of 1.2–1.6 pM with a custom sequencing chemistry protocol (read 1:50 imaged cycles; index read 1:8 imaged cycles, 27 dark cycles, 10 imaged cycles; index read 2:8 imaged cycles, 21 dark cycles, 10 imaged cycles; read 2:50 imaged cycles). Custom sequencing primers are provided in Supplementary Table 6.

3.7.7 Sequencing data processing

FASTQ files were generated from BCL files using bcl2fastq (Illumina Inc., v2.19.0). FASTQ files were aligned, filtered, and analyzed primarily using the 'scitools' software suite (github.com/adeylab/scitools)[202], which includes wrappers for numerous external tools. Raw sequence read index combinations were matched to a whitelist of expected indexes using '*scitools fastq-dump*', which allows for a hamming distance of two from sci indexes and produces error-corrected FASTQ files. These were then aligned to the human reference genome (hg38) via BWA-MEM (v0.7.15-r1140)[132] and sorted using '*scitools align*'. PCR duplicate removal and filtering for quality ten aligned autosomal and chromosome X reads (i.e., excluding mitochondrial, chromosome Y, and unanchored contigs) was performed using '*scitools rmdup*' with default parameters and plotted using '*scitools plot-complexity*'. BAM files were filtered to only contain cell barcodes that contained a minimum of 1000 passing reads and a percent unique read of ≤80%, as any overly complex cell libraries may be doublets and were therefore excluded.

3.7.8 Regional chromatin accessibility processing

Chromatin accessibility peak calling was performed on the filtered BAM file comprised of the aligned reads from the sci-ATAC-seq data collected from all individuals. Peak calling was run on pseudo-bulk data (irrespective of single cells) to call read pile-ups, or peaks, using the wrapper function '*scitools callpeak*', which utilized MACS2 (v.2.2.7.1) for peak calling, filtering, and peak extension to 500 bp[133]. The 779,149 called peak set and filtered BAM files of the cortex sample data and hippocampus sample data, respectively, were then used to construct region-specific cortex and hippocampus, peak × cellID count matrix of the count occurrence of reads within peak regions per cell. Latent Dirichlet Allocation (LDA) dimensionality reduction of the peak × cellID count matrix was performed using the package cisTopic (v0.2.0)[134] using the scitools wrapper function

'*scitools cistopic*' for models 20, 30, and 40. Topic count was selected using

'*selectModel*' based on the second derivative of model perplexity. Two-dimensional

visualization of regional datasets was performed using '*scitools umap*' and plotted using

'*scitools plot-dims*'. Louvain clustering was performed on the topic × cellID matrix using

the wrapper function '*scitools matrix-pg*' that utilizes the R (v4.0.0) package

Rphenograph (v0.99.1)[136]. A tabix fragment file was generated using samtools and tabix

(v1.7). The counts matrix and tabix files were then input into a Signac (v1.0.0)

SeuratObject for chromatin accessibility analysis[140].

3.7.9 Cell type annotation

The clusters defined in the cortex and hippocampus UMAPs were assigned to their

respective cell type by examining the chromatin accessibility profiles of marker genes

that correspond to known cell types and also through integration with public single-cell

human brain datasets. Tn5 insertion frequency at marker genes from regional filtered

BAM files was plotted using '*scitools plot-reads*' and from using the R (v4.0.0) package

Signac (v1.0.0) function '*CoveragePlot*.' Marker genes used for cell-type identification

included: *GFAP, GLUL,* and *AGT* for astrocytes, *SLC17A7* and *DKK3* for glutamatergic

neurons, also referred to as pyramidal neurons, *PROX1* and *DSP* for dentate granule

cells, *GAD1, GAD2, DLX1,* and *DLX2* for GABAergic neurons, also referred to as

interneurons, *C1QA* and *C1QC* for microglia, *OLIG1* and *OLIG2* for all oligodendrocyte

cell types, *OLIG1* for OPCs, and *OLIG1* and *OLIG2* for oligodendrocytes. A cluster of

cells identified as having low unique reads per cell and low TSS (Fig. 21) was labeled as

"Low-quality" and was excluded from future analysis.

3.7.10 ATAC-ATAC single-cell integration

Combined processing of the Thornton *et al.*[147] human primary somatosensory (SSp)

cortex sciMAP-ATAC dataset and a subset of the cortex dataset from this study was

performed to validate cell type annotations. The cortex SeuratObject was subset to 5%

by randomly subsampling barcodes from the count matrix (n=10,326 cells). We found

the sciMAP-ATAC peak set (n=132,695 peaks) to be primarily a subset of the peaks

called from all cells in this study (n=779,149 peaks), with 96% of sciMAP-ATAC peaks

being encompassed in our peak set. As a result, we used the peaks from this study to

generate a peak × cellID count matrix for all cells from the 5% cortex subset (n=10,326

cells) and sciMAP-ATAC human SSp dataset (n=4,547 cells). This count matrix (14873

cellIDs x 779,149 peaks) was imported into a Signac (v1.0.0) SeuratObject with cell type

annotations from both studies imported as metadata. Dimensionality reduction was

performed using the package cisTopic (v0.2.0)[134] function '*runWarpLDAModels*', for

models 20, 25, 28, 30, 33, 35, and 38, selecting the topic count using '*selectModel*'

based on the second derivative of model perplexity. The topic × cellID matrix was then

projected into two-dimensional space using R (v4.0.0) package uwot (v0.1.8) function

'*umap*' and plotted using '*scitools plot-dims*', colored by cell types identified in Thornton

et al. and from cell type annotations in this study (Fig. 22a).

3.7.11 RNA-ATAC cell-type label transfer

Canonical correlation analysis (CCA)-based single-cell dataset integration between Lake

et al.[148] human frontal cortex DROP-seq RNA dataset (n=10,319) and the 5% cortex

subset dataset from this study was also performed to validate cell type annotations. The

Lake *et al.* transcriptomic dataset was imported into a Seurat (v3.2.1) SeuratObject and

processed using the standard Seurat 3 (v3.2.1) workflow. The SeuratObject was created

using '*CreateSeuratObject*' using the expression matrix and cell metadata, scaled and

normalized using '*ScaleData*' and '*NormalizeData*' based on the top 5,000 variable

features that were identified using the function '*FindVariable*' specifying

'*nfeatures=5000*'. Dimensionality reduction and clustering were performed using

'*RunPCA*', '*FindNeighbors*', '*FindClusters*', and '*RunUMAP*'. The 5% cortex subset from this study was imported into a Signac (v1.0.0) SeuratObject and processed using the standard Signac (v1.0.0) workflow. Dimensionality reduction was performed using '*RunTFIDF*', '*FindTopFeatures*', '*RunSVD*', and '*RunUMAP*'. Clustering was performed using '*FindNeighbors*', and '*FindClusters*' with '*resolution=1.2*'. In order to integrate our ATAC data with Lake *et al.* RNA data, we generated a gene activity matrix from the 5% cortex subset ATAC SeuratObject from cis-coaccessibility networks generated through the Signac wrapper of cicero (v1.3.4.10)[141] and the Signac function '*build_gene_activity_matrix*'. Gene activities were normalized using '*normalize_gene_activities*'.

To integrate the processed RNA and ATAC SeuratObjects, we learned the transfer anchors based on the gene activity and expression data using '*FindTransferAnchors'* and then used '*TransferData'* to project the RNA cell type labels onto the ATAC cells as described in the 'Integrating scRNA-seq and scATAC-seq data' Seurat vignette (https://github.com/satijalab/seurat/blob/master/vignettes/atacseq_integration_vignette.R md). The UMAP of the integrated datasets were plotted using '*scitools plot-dims*' colored by dataset of origin (Fig. 22b), cell type annotations from this study (Fig. 22c), and cell type annotations defined in Lake *et al.* (Fig. 22d). The confusion matrix of the RNA cell type labels projected onto the ATAC cells was plotted using the R (v4.0.0) package ggplot2 (v3.3.2) function '*geom_tile*' on the row normalized confusion matrix (Fig. 22e). Discrepancies in cell type assignment (5% cortex subset astrocytes predictively labeled as 'Ex8' Lake *et al.* excitatory neurons) were assessed for astrocyte and glutamatergic neuron chromatin accessibility at marker genes using Signac (v1.0.0) function '*CoveragePlot*'.

3.7.12 Quality assessment

We calculated the transcription start site (TSS) density for all data (Fig. 23e) and for each cell type (Fig. 21b,d,f) identified from each region. To calculate cell type TSS, we subset the filtered BAM of all cells by annotated cell type. TSS enrichment values were calculated using the method established by the ENCODE project[146] (https://www.encodeproject.org/data-standards/terms/enrichment), whereby the aggregate distribution of reads ±1000 bp centered on the set of TSSs is then used to generate 100 bp windows at the flanks of the distribution as the background and then through the distribution, where the maximum window centered on the TSS is used to calculate the fold enrichment over the outer flanking windows. We plotted these data using R (v4.0.0) package ggplot2 (v3.3.2) function '*geom_line'* using default parameters. The total number of unique reads per cell was plotted using '*geom_violin*' and '*geom_boxplot*'. Violin width represents the mirrored continuous distribution of cell values. Boxplot lower and upper hinges indicate first and third quartiles, the center line indicates the median, upper, and lower whiskers indicate 1.5 times the inner quartile range (IQR). Data points beyond the end of the whiskers are plotted individually.

3.7.13 Cell type chromatin accessibility processing

Cell type-specific SeuratObjects were generated from cortex and hippocampus-derived cells based on cell type annotation. To achieve this, the cell-type-specific duplicate removed and filtered BAM for all cells from the cortex and hippocampus was split by cell type annotation into cell type-specific BAMS using '*scitools bam-split*'. Cell x peak count matrices were generated from the cell-type-specific BAM and all peaks were called on the entire dataset (n=779,149 peaks). These count matrices and the tabix fragment file were then input into a Signac (v1.0.0) SeuratObject for chromatin accessibility analysis, resulting in four combined cortex and hippocampus glial cell-type-specific SeuratObjects:

astrocytes, microglia, oligodendrocytes, and OPCs. LDA dimensionality reduction of the

peak × cellID count matrix was performed using the package cisTopic (v0.2.0) was

performed using the scitools wrapper function '*scitools cistopic*' for models 25, 28, 30,

33, 35, 37, and 40, selecting the topic count using '*selectModel*' based on the second

derivative of model perplexity. Two-dimensional visualization was performed using

UMAP via '*scitools umap*' colored by individual and plotted using '*scitools plot-dims*' (Fig.

24a,b,d,f). All genome track coverage plots were generated using R (v4.0.0) package

Signac (v1.0.0) plotting function '*CoveragePlot*'. Coverage plots that compare Tn5

insertion site frequency between cell types were plotted with '*scale.factor=1e8*' to scale

different plots equivalently and allow for inter-SeuratObject comparisons.

3.7.14 Batch-effect correction using Harmony

We found that experimental bias was minimally present in the 2D UMAPs generated by

region, for the cortex and hippocampus and did not present a barrier to cell type

identification. However, we did find that upon cell-type-specific sub-clustering, individual

and experimental batches were non-randomly distributed across the UMAP (Fig. 24b,d,f)

in three cell types: astrocytes, microglia, oligodendrocytes. In order to reduce the

partitioning of cells by batch, we used the Seurat (v3.3.2.1) wrapper for Harmony[149]

(v1.0) on each cell-type-specific dataset to integrate experimental batches. The function

'*RunHarmony*' was run on astrocyte, microglia, and oligodendrocyte cell-type-specific

SeuratObjects, using options '*group.by.vars=Experiment*', to remove the effect of

experimental bias, and '*reduction='cistopic'*' in order to perform integration on the entire

cell x topic cisTopic matrix. Dimensionality reduction using UMAP was reperformed on

the Harmony matrix using the function '*RunUMAP*'. We found that Harmony mitigated

the batch effect in the 2D UMAP projections (Fig. 24c,e,g). Visualization of Harmony-

weighted topic x cell matrices was performed using the Heatmap function in the ComplexHeatmap package (v2.5.5) in R (v4.0.0)[135].

3.7.15 Batch-corrected cell type chromatin accessibility processing

For each cell type, clustering was performed with Seurat (v3.2.1) '*FindNeighbors*' and '*FindClusters*' functions on the topic × cellID data frame. For the '*FindClusters*' function call, resolutions ranging from 0.01 to 0.6 were run for all cell types. We observed robust clustering across multiple resolutions in all cell types. As high resolutions generate artificially high cluster numbers, the clustering resolution was set at the lowest resolution that resulted in this robust clustering and is provided as 'seurat_clusters' in the metadata of all cell-type-specific SeuratObjects. To assess the distribution of individuals across all clusters in each cell type, we calculated the percent that each individual contributes to the cell count of a given cluster based on the normalized individual cell count (normalized by the total number of cells for each individual within a given cell type). These values, or the contributions of each individual to each cluster, were plotted using the '*Heatmap*' function in the ComplexHeatmap package (v2.5.5) in R (v4.0.0, Fig. 27).

3.7.16 Cell type-specific cluster marker gene analysis

The most significantly elevated markers of each cell-type-specific cluster were identified in order to elucidate the biological underpinning on cell type-specific cellular sub-states. We identified cis-coaccessibile networks (CCANs) in single cells using the Cicero (v1.3.4.10) function '*generate_ccans*' and, from those chromatin accessible interactions, generated a normalized gene activity matrix for each cell type using the function '*build_gene_activity_matrix*' and '*normalize_gene_activities*'. Gene activity was plotted on the cell type UMAP using Seurat (v.3.2.1) function '*FeaturePlot*' (Fig. 25e, Fig. 28b), and plotted as violin density plots using '*VlnPlot*' (Fig. 30f). Marker genes for each cluster were identified by cluster in one by rest comparisons that were generated using

Seurat (v.3.2.1) function '*FindMarkers*' using options: '*test.use* = '*LR*'', '*only.pos* = *T*', and '*logfc.threshold=0.2*' with '*latent.vars* = '*nCount_peaks*'' to account for read depth. All marker gene activities are provided with fold-change and Bonferroni corrected p-value in Supplementary Table 4. Heatmaps of gene activity values of the top marker genes were plotted using the '*Heatmap*' function in the ComplexHeatmap package (v2.5.5) in R (v4.0.0), split by cluster (Fig. 21g, 29e). Scatter plots of gene activity values of top marker genes were plotted using ggplot2 (v3.3.2) '*geom_scatter*' (Extended Data Fig. 24f). Gene activity dot plots were plotted using '*geom_point*' with size corresponding to the percent of cells expressing a given gene and color corresponding to average gene activity.

3.7.17 Cell type-specific cluster maker transcription factor analysis

We used the Seurat (v.3.2.1) chromVAR package wrapper function '*RunChromVAR*' to quantify transcription factor (TF) motif occurrence in each peak and to calculate transcription factor motif enrichment (TFME) in each cell, using the JASPAR set of motifs (v2020_0.99.10). Marker transcription factors for each cluster were identified by cluster in one by rest comparisons that were generated using Seurat (v.3.2.1) function '*FindMarkers*' using options: '*test.use* = '*LR*'', '*only.pos* = *T*', and '*logfc.threshold=0.2*' with '*latent.vars* = '*nCount_peaks*'' to account for read depth. Marker TF activities of the top markers of cell-type-specific clusters were plotted using the '*Heatmap*' function in the ComplexHeatmap package (v2.5.5) in R (v4.0.0), split by cluster. All marker TFs are provided with fold-change and Bonferroni corrected p-value in Supplementary Table 4.

3.7.18 Gene ontology enrichment analysis

GO enrichment analysis was performed on the genomic regions defined within Harmony-weighted topic 1 (n=8,579 peaks) and 26 (n=8,157 peaks) of the microglial SeuratObject, which were enriched in microglial cluster MIC-*CLDN11*. The single

nearest gene within 1,000 kb to topic1 (n= 4,926 genes) and 26 (n= 5,408 genes)

regions were identified using GREAT (v4.0.4) for reference genome hg38[203]. Statistical

overrepresentation for GO biological processes was calculated using GREAT (v4.0.4)

binomial test with Bonferroni correction of p-values (≤ 0.05) for overrepresentation of

topic 1 and 26 genomic regions in comparison to all human (hg38) genes. We used

Revigo on the GREAT output to summarize them by removing redundant GO terms,

specifying the parameters of a medium output list, and providing p-values associated

with GO terms[204]. Data were plotted using ggplot2 (v3.3.2) plotting function '*geom_point*'

in R (v4.0.0) with size corresponding to the number of genes in the test set with the

annotation and colored by GO term -log10 GO term Bonferroni corrected p-values.

3.7.19 Cell cycle analysis

We performed cell cycle phase scoring on OPC cells in order to determine the impact of

cell cycle heterogeneity on OPC clustering by testing for the enrichment of canonical cell

cycle genes in OPC clusters. Cell cycle scoring was based on gene activity of G2/M and

S phase marker enrichment scoring from a human mid-gestation neocortical

development cell cycle marker list from Polioudakis *et al.*[205]. We used the Seurat (v3.2.1)

function '*CellCycleScoring*' to calculate the S phase score, G2M phase score, and

Phase prediction for each cell. All values are available in the OPC SeuratObject

metadata. Phase scores for each cell were plotted using '*geom_violin*' and

'*geom_boxplot*' from the package ggplot2 (v3.3.2) in R (v4.0.0). Violin width represents

the mirrored continuous distribution of cell values. Boxplot lower and upper hinges

indicate first and third quartiles, centerline indicates the median, upper, and lower

whiskers indicate 1.5 times the inner quartile range (IQR). Data points beyond the end of

the whiskers are plotted individually.

3.8 Data and materials availability

Analysis was performed using functions contained within the scitools software (github.com/adeylab/scitools). Additional code used to analyze the data will be made available as a Github repository (github.com/thorntca). Raw and processed single-cell library sequencing data will also be made available through the NeMo archive and UCSC Cell Browser (includes Seurat Object Download links).

3.9 Acknowledgements

Chapter 4: Conclusions

4.1 Growing the taxonomic tree of single cells

Here, I present my work exploring and defining the intra-organismal epigenomic variation of cells in human and mouse brains. First, I developed a novel method for studying the intra-organismal epigenomic heterogeneity of cells while maintaining cellular spatial orientation. I applied this method, sciMAP-ATAC, in multiple contexts. Using sciMAP-ATAC, I resolved the cortical layering structure in the primary visual cortex in the human brain and the primary somatosensory cortex in the mouse brain. This application of sciMAP-ATAC in cortical regions served as a validation of sciMAP-ATAC by confirming that we could identify the layer-specific features of the cortex. The development of sciMAP-ATAC represents methods-development-driven science and applying sciMAP-ATAC to the known spatially resolved structure of the cortex served as a validation experiment. In this work, I generated data that adds to the cellular taxonomy of the healthy human and mouse brain tissue sub-structure level of the taxonomic hierarchy of single-cells.

Additionally, I applied sciMAP-ATAC to a mouse model of cerebral ischemia and was able to identify the response of specific cell types in the space surrounding focal hypoperfusion. This study revealed regulatory elements and transcription factors that vary, i.e., increase or decrease, with spatial relation to the infarction. This application of sciMAP-ATAC in cerebral ischemia highlights how previously observed functional and morphological changes in tissues in disease are tied to epigenomic changes and thus require spatially-resolved single-cell methods to assess them.

Secondly, I explored and defined cells in the human cortex and hippocampus and provided a publicly explorable atlas of cells that can be used for future basic and

translational research. We focused on non-neuronal cells in order to balance the abundance of single-cell neuron epigenomic data and cellular characterizations. In the context of single-cell epigenomic taxonomy, this work offers characterization in taxon: Domain: Eukarya, Kingdom: Animalia, Phylum: Chordata, Class: Mammalia, Order: Primates, Family: Hominids, Genus: Homo, Species: Homo Sapiens, Organ: Brain, Tissues: Cortex and Hippocampus, Tissue Sub-Structure: All, cortical layers and hippocampal regions, Cell: Glia, Cell state: All. Tissue sub-structure selection was not performed and therefore was not assigned *a priori*. Cell selection for non-neuronal cells was performed, however, and therefore our classification and naming of cell types and cell states pertained to all non-neuronal cells. Without a taxonomical nomenclature for this work, I was able to develop my own based on Linnaeas' example; designating a binomial English naming system of (i) cell type, as commonly used in glial research (see Table 1) and (ii) cell state as defined by the name of the top gene marker. While this naming schema has been sporadically employed in single-cell epigenomic characterization studies, it is not a standardized approach. Therefore, the human glial atlas is one of many works which identifies cell states but does not represent a shared taxonomy.

This human glial atlas corroborates the breadth of previously identified cell types and provides new and essential representations and interpretations of glial cell states based on cellular chromatin accessibility. I substantiate findings in the field that show a gross homogeneity of oligodendrocytes between tissue sub-substructures in the brain and spatially within tissue sub-structures in the brain (e.g., along the rostral-caudal axis within the cortex)[37]. I established low regional specificity in oligodendrocytes coupled with high homogeneity and low inter-individual heterogeneity in oligodendrocyte

progenitors. I also identified a rare (0.37% of all cells) population of early-differentiating oligodendrocyte precursors present across all individuals.

Conversely, we demonstrate robust heterogeneity within the other canonical glial cell types in the healthy human brain. Thus, adding context to glial heterogeneity studies, which primarily focus on disease states. I identified several subpopulations of astrocytes across all healthy individuals (including A1 and A2 reactive cell states) that exhibit signatures of neuroinflammation, suggesting that this may represent a normal component of aging and not a cell state specifically associated with disease. Additionally, we directly address the long-standing hypothesis regarding the bimodality of ramified and reactive amoeboid states in microglia, identifying multiple subtypes that bridge the two states.

4.1.1 The strength and limitations of single-cell assays in glial characterization

The methods used to capture features of all cells, and specifically glia, significantly impact the scope and conclusions that can be derived from a study of cellular heterogeneity. As previously noted, the field of glial research has provided our current understanding of glial states and activity in healthy and diseased tissues. As referenced in our analysis of microglia (see Section 3.5.5), previous research has characterized microglia by morphology, identifying ramified microglial cell (RMC) and amoeboid microglial cell (AMC), by staining and imaging methodologies[185]. Additionally, these morphological states are correlated with functional states, e.g., cellular macrophage-like phagocytic activity in the AMC, and quiescent sensing activity in the RMC, morphologies. Morphological and functional characterizations of glia are, however, limited in that the partitioning of different cell states is limited to the physical presentations of cells.

In this work, I demonstrate that cells can vary epigenomically within a morphological subtype. For example, I found two clusters of microglia that demonstrated two distinct chromatin landscapes (MIC-*SEPT9*, MIC-*RUNX1T1*) that shared markers for the AMC morphological subtype and found statistically significant marker genes which differentiated MIC-*SEPT9* from MIC-*RUNX1T1*. Therefore, single-cell epigenomic methodologies are necessary to distinguish morphologically similar but epigenomically dissimilar cell states.

Single-cell assays are also valuable for glial characterization because they facilitate high cell count and hypothesis-independent cell state discovery. In the case of AMC and RMC microglia, previous work assayed the bulk transcriptomes of batched AMC and RMC, respectively[177]. As a result, the aggregate transcriptome of AMC and RMC were revealed. This morphologically driven and hypothesis-dependent approach relied on the known physical characteristics of cells to discover features of the epigenome. In contrast, single-cell assays, such as sci-ATAC-seq, provide an unbiased and morphology-independent methodology for cell state discovery.

The computational characterization of single-cell omic data is one of many necessary methodological approaches required to develop a comprehensive understanding of glia. For instance, in single-cell epigenomic assay analysis, it is difficult to discern whether statistically significant omic variation, especially in non-coding regions of the genome, contributes to the morphology or function of a cell. Therefore, omic data alone cannot determine the effect of omic variation between glia, and so multiple methods, including those that epigenomically, morphologically, and functionally characterize glia, must be used in concert to describe cellular variation, and most importantly – define the features that affect how glia function and react to stimuli in-situ.

The Human Glial Atlas, presented in this work, succeeds in using sci-ATAC-seq and computational omic characterization to define robust cell state differences within glial cell types, that are shared between the human cortex and hippocampus. Additionally, the Human Glial Atlas also serves as a launching pad for future research that morphologically and functionally characterizes the glial states described in this work. In total, the development of sciMAP-ATAC, its application in healthy and diseased tissues, and the development and characterization of the Human Glial Atlas represent substantial contributions to the taxonomic characterization of single-cells within the mouse and human brains.

4.1.2 The complicated task of incorporating space into taxonomy

The classical system of Linnaean taxonomy does not account for mutable features of biological moieties. As the epigenomics of single cells is incorporated into a new epigenomic tier of the tree of life – we must consider the multitude of data types to be included in defining the many types and states of cells. Incorporating omic data types into taxonomy presents the next great taxonomic bioinformatic challenge. As Linnaeus developed the taxonomic binomial naming system to make sense of the many species being identified in his lifetime, so to, we must address the bioinformatic challenge of incorporating omic data types into an interpretable map of cellular heterogeneity.

In this work, I demonstrate that we can capture the location of cells and their chromatin accessibility in complex tissues. I show that we can resolve the different cell types and the different cell states in healthy tissue through the cortical lamina and in the disease state through the infarction and penumbra area. I posit that this spatial data type is one of many features that will be used to define epigenomic cell states. For example, in the taxonomic tree of the human brain, within the: Species: Homo Sapiens, Organ: Brain, Tissue: Cortex, Cell type: Neuron, we will use the spatial orientation of cells as one data

type that will be descriptive of distinct neuronal states in the cortical layers. For instance, in the case of oligodendrocytes, our work and others have found that the spatial orientation of oligodendrocytes within and between tissues of the human brain do not correspond to transcriptomic or chromatin accessibility differences between cells. Therefore, the spatial data type reveals that oligodendrocyte subtypes, if there are any, are not spatially distinct. In contrast, neurons, as described, do exhibit spatial variability, and as a result, the data type of spatial orientation will be used to describe distinct neuronal cell states. While spatial orientation is a new data type in single-cell omics analysis, we are poised to address this new omic challenge with the same incremental reasoning that Linnaeus was able to use long ago.

4.2 An opportunity for advancement: expanding epigenetic assessment in clinical genetics

Epigenetic assessment of cell types within diseased tissues is essential to understanding how conditions affect networks of interconnected cells and what, if any, therapeutic approaches are possible. Here I provide a perspective on the current and potential applications of single-cell omic technologies in improving patient medical care.

4.2.1 The clinical relevance of single-cell technologies

The relatively recent emergence of the field of single-cell technologies presents barriers to clinical actionability at present. The two primary barriers to single-cell assessment in clinical genetics are (i) the lack of a comprehensive understanding of epigenetic states in health and diseases and (ii) the high cost of exploratory single-cell diagnostics.

Clinical genetic testing relies on a shared knowledge of disease states. For example, cytogenetic assessment depends on understanding the number, banding, and organization of chromosomes in the normal human cell. In the case of single-cell

epigenetic evaluation in the clinical setting, too little is known, at this time, about the normal range of epigenomic features in any given cell type to allow for diagnosis or an actionable therapeutic strategy based on epigenomic feature information. However, there are cases where single-cell assessment is being trialed in the clinical setting.

Single-cell omics technologies have had the most pronounced impact on oncological care. Cancers are a group of diseases resulting from genomic instability, leading to unchecked cell growth and proliferation and dysfunction of tissues and organ systems. A marked feature of cancers is the heterogeneity of the tumor microenvironment, which has been shown to contain non-cancerous cells, cancerous cells, and infiltrating immune cell populations. This spatial heterogeneity lends itself to a spatially-resolved single-cell omic approach to understanding the impact of each cell type on tumor progression. An immense amount of focus and research has been dedicated to assessing cell genomic and epigenomic states within the tumor microenvironment. For instance, the National Cancer Institute's Human Tumor Atlas Network (HTAN) aims to generate a spatially-resolved single-cell atlas of cellular cancer progression from the pre-cancerous lesion to late-stage cancers[206]. Some single-cell approaches to oncological care are, however, in clinical trials. In the case of melanoma, single-cell transcriptomic analyses of the melanoma tumor microenvironment have revealed an immune T-cell population that is resistant to a commonly used immune checkpoint inhibitor (ICI) drug[207]. From this finding, a clinical trial is being conducted to first test melanoma patients for the abundance of ICI-resistant T cells, then treat patients with ICI-resistant T cells with a combination therapy that uses FDA-approved cancer drugs to deplete the ICI-resistant T cell population.

Single-cell omics technologies have also led to advances in identifying druggable targets in specific diseases. In the case of lung fibrosis, a condition that leads to scar tissue

accumulation in the lungs, a single-cell transcriptomics study was able to identify an immune cell population, monocyte-derived alveolar macrophages, that expanded with disease progression[208]. This mechanistic understanding of lung fibrosis has led to efforts to screen for the abundance of the disease-associated cell type and identify cell-type-specific druggable targets. Additionally, single-cell research into the expression of the defective *CFTR* gene that leads to reduced respiratory function in cystic fibrosis (CF) revealed a rare, previously unknown cell type, termed ionocytes, that produce nearly 50% of the defective CFTR protein in CF patients[209]. Similarly, this finding is now being used to develop therapeutics that target the sparse ionocyte cell population.

A second impediment to applying single-cell technologies in clinical genetic diagnostics is the high cost. The costliness of single-cell studies is due to the expense of sequencing and is also due to the specialized materials, such as microfluidics equipment, indexed oligonucleotides, and expertise that is required to perform and interpret single-cell studies. Commercial applications are, however, increasing the accessibility of single-cell experiments[210]. However, the cost decrease of single-cell experiments for research has not had a substantial impact on the clinical application of single-cell technologies. The delay between technology development in the research space and the application of these technologies in the clinical space is because the field of single-cell omics is still in the discovery phase. Clinical genetics laboratories will be able to utilize single-cell technologies using commercial platforms when known single-cell results have actionable conclusions based on human diseases. The field of single-cell omics is still defining the normal and diseased states of human tissues, and therefore utilization of single-cell technologies in clinical genetics laboratories remains cost and time prohibitive.

References

1. Sanger, F., Nicklen, S. & Coulson, A. R. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. U. S. A.* **74**, 5463–5467 (1977).

2. Brenner, S. *et al.* Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. *Nat. Biotechnol.* **18**, 630–634 (2000).

3. Michener, C. D. *et al. Systematics In Support of Biological Research.* (Division of Biology and Agriculture, National Research Council, 1970).

4. Haeckel, E. *Generelle morphologie der organismen. Allgemeine grundzüge der organischen formen-wissenschaft, mechanisch begründet durch die von Charles Darwin reformirte descendenztheorie* . **bd. 1**, (G. Reimer, 1866).

5. Müller-Wille, S. Linnaeus' herbarium cabinet: a piece of furniture and its function. *Endeavour* **30**, 60–64 (2006).

6. J., C. M., M., M. R. & E., S. N. Can We Name Earth's Species Before They Go Extinct? *Science (80-.).* **339**, 413–416 (2013).

7. Paterlini, M. There shall be order. The legacy of Linnaeus in the age of molecular biology. *EMBO Rep.* **8**, 814–816 (2007).

8. Darwin, C. & Kebler, L. On the origin of species by means of natural selection, or, The preservation of favoured races in the struggle for life. *J. Murray* (1859).

9. Hug, L. A. *et al.* A new view of the tree of life. *Nat. Microbiol.* **1**, 16048 (2016).

10. Dubois, A. Zoological nomenclature in the century of extinctions: priority vs. 'usage'. *Org. Divers. Evol.* **10**, 259–274 (2010).

11. Raposo, M. A. *et al.* On the notions of taxonomic 'impediment', 'gap', 'inflation' and 'anarchy', and their effects on the field of conservation. *Syst. Biodivers.* **19**, 296–311 (2021).

12. Godfray, H. C. J. & Knapp, S. Introduction. Taxonomy for the twenty-first century. *Philos. Trans. R. Soc. London. Ser. B, Biol. Sci.* **359**, 559–569 (2004).

13. Regev, A. *et al.* The Human Cell Atlas. *Elife* **6**, e27041 (2017).

14. Rood, J. E. *et al.* Toward a Common Coordinate Framework for the Human Body. *Cell* **179**, 1455–1467 (2019).

15. Biesecker, L. G. & Spinner, N. B. A genomic view of mosaicism and human disease. *Nat. Rev. Genet.* **14**, 307–320 (2013).

16. Sasani, T. A. *et al.* Large, three-generation human families reveal post-zygotic mosaicism and variability in germline mutation accumulation. *Elife* **8**, e46922 (2019).

17. Hao, D., Wang, L. & Di, L. Distinct mutation accumulation rates among tissues determine the variation in cancer risk. *Sci. Rep.* **6**, 19458 (2016).

18. Lee, C. A., Abd-Rabbo, D. & Reimand, J. Functional and genetic determinants of mutation rate variability in regulatory elements of cancer genomes. *Genome Biol.*

22, 133 (2021).

19. Waddington, C. H. Developmental Pathways. in *The Strategy of the Genes* 26–28 (George Allen & Unwin LTD, 1957).

20. Max, K. *et al.* A single–cell type transcriptomics map of human tissues. *Sci. Adv.* **7**, eabh2169 (2022).

21. Cusanovich, D. A. *et al.* A Single-Cell Atlas of In Vivo Mammalian Chromatin Accessibility. *Cell* **174**, 1309-1324.e18 (2018).

22. What Is Your Conceptual Definition of 'Cell Type' in the Context of a Mature Organism? *Cell systems* **4**, 255–259 (2017).

23. Finlay, B. Brain and behavioral development (II) : cortical. in

24. Ma, S. *et al.* Chromatin Potential Identified by Shared Single-Cell Profiling of RNA and Chromatin. *Cell* **183**, 1103–1116 (2020).

25. Chen, S., Lake, B. B. & Zhang, K. High-throughput sequencing of the transcriptome and chromatin accessibility in the same cell. *Nat. Biotechnol.* **37**, 1452–1457 (2019).

26. Zhu, C. *et al.* An ultra high-throughput method for single-cell joint analysis of open chromatin and transcriptome. *Nat. Struct. Mol. Biol.* (2019). doi:10.1038/s41594-019-0323-x

27. Hagemann-Jensen, M. *et al.* Single-cell RNA counting at allele and isoform resolution using Smart-seq3. *Nat. Biotechnol.* **38**, 708–714 (2020).

28. Co, M. *et al.* Shared and distinct functional effects of patient-specific Tbr1 mutations on cortical development. *bioRxiv* 2022.01.27.478064 (2022). doi:10.1101/2022.01.27.478064

29. Willsey, A. J. *et al.* Coexpression Networks Implicate Human Midfetal Deep Cortical Projection Neurons in the Pathogenesis of Autism. *Cell* **155**, 997–1007 (2013).

30. Yook, C. *et al.* A TBR1-K228E Mutation Induces Tbr1 Upregulation, Altered Cortical Distribution of Interneurons, Increased Inhibitory Synaptic Transmission, and Autistic-Like Behavioral Deficits in Mice. *Front. Mol. Neurosci.* **12**, 241 (2019).

31. Genomics, 10×. *Inside Visium spatial capture technology.* (2019).

32. Rodriques, S. G. *et al.* Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution. *Science (80-.).* **363**, 1463 LP – 1467 (2019).

33. Pissarro, C. Eragny Landscape. *Morgan Libr. Museum* (1886).

34. Ducker, S. C. & Knox, R. B. Pollen and Pollination: A Historical Review. *Taxon* **34**, 401–419 (1985).

35. Buenrostro, J. D. *et al.* Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* **523**, 486 (2015).

36. Zhang, K. *et al.* A single-cell atlas of chromatin accessibility in the human

genome. *Cell* **184**, 5985-6001.e19 (2021).

37. Zeisel, A. *et al.* Molecular Architecture of the Mouse Nervous System. *Cell* **174**, 999-1014.e22 (2018).

38. Jäkel, S. *et al.* Altered human oligodendrocyte heterogeneity in multiple sclerosis. *Nature* **566**, 543–547 (2019).

39. Sueli, M. *et al.* Oligodendrocyte heterogeneity in the mouse juvenile and adult central nervous system. *Science (80-.).* **352**, 1326–1329 (2016).

40. Boyle, A. P. *et al.* High-Resolution Mapping and Characterization of Open Chromatin across the Genome. *Cell* **132**, 311–322 (2008).

41. Berg, D. E., Davies, J., Allet, B. & Rochaix, J. D. Transposition of R factor genes to bacteriophage lambda. *Proc. Natl. Acad. Sci.* **72**, 3628–3632 (1975).

42. Hickman, A. B. & Dyda, F. DNA Transposition at Work. *Chem. Rev.* **116**, 12758–12784 (2016).

43. Adey, A. *et al.* Rapid, low-input, low-bias construction of shotgun fragment libraries by high-density in vitro transposition. *Genome Biol.* **11**, R119 (2010).

44. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213 (2013).

45. Cusanovich, D. A. *et al.* Multiplex single-cell profiling of chromatin accessibility by combinatorial cellular indexing. *Science (80-.).* **348**, 910 LP – 914 (2015).

46. Becht, E. *et al.* Dimensionality reduction for visualizing single-cell data using UMAP. *Nat. Biotechnol.* **37**, 38 (2018).

47. Chari, T., Banerjee, J. & Pachter, L. The Specious Art of Single-Cell Genomics. *bioRxiv* 2021.08.25.457696 (2021). doi:10.1101/2021.08.25.457696

48. Miterko, L. N., Lackey, E. P., Heck, D. H. & Sillitoe, R. V. Shaping Diversity Into the Brain's Form and Function   . *Frontiers in Neural Circuits*   **12**, (2018).

49. Tasic, B. *et al.* Adult mouse cortical cell taxonomy revealed by single cell transcriptomics. *Nat. Neurosci.* **19**, 335–346 (2016).

50. Zeisel, A. *et al.* Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. *Science (80-.).* **347**, 1138 LP – 1142 (2015).

51. Macosko, E. Z. *et al.* Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell* **161**, 1202–1214 (2015).

52. De La Fuente, A. G. *et al.* Pericytes Stimulate Oligodendrocyte Progenitor Cell Differentiation during CNS Remyelination. *Cell Rep.* **20**, 1755–1764 (2017).

53. Spitzer, S. O. *et al.* Oligodendrocyte Progenitor Cells Become Regionally Diverse and Heterogeneous with Age. *Neuron* **101**, 459-471.e5 (2019).

54. Floriddia, E. M. *et al.* Distinct oligodendrocyte populations have spatial preference and different responses to spinal cord injury. *Nat. Commun.* **11**, 5860 (2020).

55. Hofman, M. A. Evolution of the human brain: when bigger is better. *Front. Neuroanat.* **8**, 15 (2014).

56. Stern, P. The human prefrontal cortex is special. *Science (80-.).* **360**, 1311 LP – 1312 (2018).

57. Lin, Q. *et al.* Incidence and risk factors of leukoaraiosis from 4683 hospitalized patients: A cross-sectional study. *Medicine (Baltimore).* **96**, e7682–e7682 (2017).

58. Mortamais, M. *et al.* Spatial Distribution of Cerebral White Matter Lesions Predicts Progression to Mild Cognitive Impairment and Dementia. *PLoS One* **8**, e56972 (2013).

59. Lin, J., Wang, D., Lan, L. & Fan, Y. Multiple Factors Involved in the Pathogenesis of White Matter Lesions. *Biomed Res. Int.* **2017**, **Arti**, https://doi.org/10.1155/2017/9372050 (2017).

60. Pantoni, L. & H., G. J. Pathogenesis of Leukoaraiosis. *Stroke* **28**, 652–659 (1997).

61. Ahn, K., Lee, S.-J. & Mook-Jung, I. White matter-associated microglia: New players in brain aging and neurodegenerative diseases. *Ageing Res. Rev.* **75**, 101574 (2022).

62. Xiong, Y. Y. & Mok, V. Age-related white matter changes. *J. Aging Res.* **2011**, 617927 (2011).

63. Choi, J. C. Genetics of cerebral small vessel disease. *J. stroke* **17**, 7–16 (2015).

64. Kirmes, I. *et al.* A transient ischemic environment induces reversible compaction of chromatin. *Genome Biol.* **16**, 246 (2015).

65. Markus, H. S. Unravelling the Genetics of Ischaemic Stroke. *PLOS Med.* **7**, e1000225 (2010).

66. Sozmen, E. G., Hinman, J. D. & Carmichael, S. T. Models that matter: white matter stroke models. *Neurotherapeutics* **9**, 349–358 (2012).

67. Sozmen, E. G., Kolekar, A., Havton, L. A. & Carmichael, S. T. A white matter stroke model in the mouse: axonal damage, progenitor responses and MRI correlates. *J. Neurosci. Methods* **180**, 261–272 (2009).

68. Ventura-Antunes, L., Mota, B. & Herculano-Houzel, S. Different scaling of white matter volume, cortical connectivity, and gyrification across rodent and primate brains. *Front. Neuroanat.* **7**, 3 (2013).

69. Oberheim, N. A. *et al.* Uniquely hominid features of adult human astrocytes. *J. Neurosci.* **29**, 3276–3287 (2009).

70. Jones, J. R. *et al.* Mutations in GFAP Disrupt the Distribution and Function of Organelles in Human Astrocytes. *Cell Rep.* **25**, 947-958.e4 (2018).

71. Buscemi, L., Price, M., Bezzi, P. & Hirt, L. Spatio-temporal overview of neuroinflammation in an experimental mouse stroke model. *Sci. Rep.* **9**, 507 (2019).

72. Adesnik, H. & Naka, A. Cracking the Function of Layers in the Sensory Cortex. *Neuron* **100**, 1028–1043 (2018).

73. Fan, X. *et al.* Spatial transcriptomic survey of human embryonic cerebral cortex by single-cell RNA-seq analysis. *Cell Res.* **28**, 730–745 (2018).

74. Hevner, R. F. Layer-Specific Markers as Probes for Neuron Type Identity in Human Neocortex and Malformations of Cortical Development. *J. Neuropathol. Exp. Neurol.* **66**, 101–109 (2007).

75. Gray, L. T. *et al.* Layer-specific chromatin accessibility landscapes reveal regulatory networks in adult mouse visual cortex. *Elife* **6**, e21883 (2017).

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

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

78. Eng, C.-H. L. *et al.* Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH+. *Nature* **568**, 235–239 (2019).

79. Vickovic, S. *et al.* High-definition spatial transcriptomics for in situ tissue profiling. *Nat. Methods* **16**, 987–990 (2019).

80. Islam, S. *et al.* Quantitative single-cell RNA-seq with unique molecular identifiers. *Nat. Methods* **11**, 163 (2013).

81. Hwang, B., Lee, J. H. & Bang, D. Single-cell RNA sequencing technologies and bioinformatics pipelines. *Exp. Mol. Med.* **50**, 96 (2018).

82. Topper, M. J., Vaz, M., Marrone, K. A., Brahmer, J. R. & Baylin, S. B. The emerging role of epigenetic therapeutics in immuno-oncology. *Nat. Rev. Clin. Oncol.* (2019). doi:10.1038/s41571-019-0266-5

83. Kato, M. & Natarajan, R. Epigenetics and epigenomics in diabetic kidney disease and metabolic memory. *Nat. Rev. Nephrol.* **15**, 327–345 (2019).

84. Hwang, J.-Y., Aromolaran, K. A. & Zukin, R. S. The emerging field of epigenetics in neurodegeneration and neuroprotection. *Nat. Rev. Neurosci.* **18**, 347 (2017).

85. Cusanovich, D. A. *et al.* Multiplex single-cell profiling of chromatin accessibility by combinatorial cellular indexing. *Science (80-.).* **348**, 910–914 (2015).

86. Vitak, S. A. *et al.* Sequencing thousands of single-cell genomes with combinatorial indexing. *Nat. Methods* **14**, 302 (2017).

87. Gerlinger, M. *et al.* Genomic architecture and evolution of clear cell renal cell carcinomas defined by multiregion sequencing. *Nat. Genet.* **46**, 225 (2014).

88. Erramuzpe, A., Cortés, J. M. & López, J. I. Multisite tumor sampling enhances the detection of intratumor heterogeneity at all different temporal stages of tumor evolution. *Virchows Arch.* **472**, 187–194 (2018).

89. Sinnamon, J. R. *et al.* The accessible chromatin landscape of the murine hippocampus at single-cell resolution. *Genome Res.* **29**, 857–869 (2019).

90. Palkovits, M. Isolated removal of hypothalamic or other brain nuclei of the rat. *Brain Res.* **59**, 449–450 (1973).

91. Corces, M. R. *et al.* An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nat. Methods* **14**, 959–962 (2017).

92. Lein, E. S. *et al.* Genome-wide atlas of gene expression in the adult mouse brain. *Nature* **445**, 168–176 (2007).

93. Alcamo, E. A. *et al.* Satb2 Regulates Callosal Projection Neuron Identity in the Developing Cerebral Cortex. *Neuron* **57**, 364–377 (2008).

94. Erö, C., Gewaltig, M.-O., Keller, D. & Markram, H. A Cell Atlas for the Mouse Brain. *Front. Neuroinform.* **12**, 84 (2018).

95. Vierstra, J. *et al.* Global reference mapping of human transcription factor footprints. *Nature* **583**, 729–736 (2020).

96. Lake, B. B. *et al.* Integrative single-cell analysis of transcriptional and epigenetic states in the human adult brain. *Nat. Biotechnol.* **36**, 70 (2017).

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

98. Lal, A. *et al.* AtacWorks: A deep convolutional neural network toolkit for epigenomics. *bioRxiv* 829481 (2020). doi:10.1101/829481

99. Endres, M. *et al.* DNA Methyltransferase Contributes to Delayed Ischemic Brain Injury. *J. Neurosci.* **20**, 3175 LP – 3181 (2000).

100. Chakravarty, S. *et al.* Insights into the epigenetic mechanisms involving histone lysine methylation and demethylation in ischemia induced damage and repair has therapeutic implication. *Biochim. Biophys. Acta - Mol. Basis Dis.* **1863**, 152–164 (2017).

101. Chisholm, N. C. *et al.* Histone methylation patterns in astrocytes are influenced by age following ischemia. *Epigenetics* **10**, 142–152 (2015).

102. Faraco, G. *et al.* Pharmacological Inhibition of Histone Deacetylases by Suberoylanilide Hydroxamic Acid Specifically Alters Gene Expression and Reduces Ischemic Injury in the Mouse Brain. *Mol. Pharmacol.* **70**, 1876 LP – 1884 (2006).

103. Formisano, L. *et al.* Ischemic insults promote epigenetic reprogramming of mu opioid receptor expression in hippocampal neurons. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 4170–4175 (2007).

104. Formisano, L. *et al.* Sp3/REST/HDAC1/HDAC2 Complex Represses and Sp1/HIF-1/p300 Complex Activates ncx1 Gene Transcription, in Brain Ischemia and in Ischemic Brain Preconditioning, by Epigenetic Mechanism. *J. Neurosci.* **35**, 7332 LP – 7348 (2015).

105. Yi, J.-H., Park, S.-W., Kapadia, R. & Vemuganti, R. Role of transcription factors in mediating post-ischemic cerebral inflammation and brain damage. *Neurochem. Int.* **50**, 1014–1027 (2007).

106. Payne, S. C. *et al.* Early Proliferation Does Not Prevent the Loss of Oligodendrocyte Progenitor Cells during the Chronic Phase of Secondary Degeneration in a CNS White Matter Tract. *PLoS One* **8**, e65710 (2013).

107. Dingman, A. L. *et al.* Oligodendrocyte Progenitor Cell Proliferation and Fate after White Matter Stroke in Juvenile and Adult Mice. *Dev. Neurosci.* (2019). doi:10.1159/000496200

108. Paul, S. R., A., K. T. & E., H. C. Neocortical Neural Sprouting, Synaptogenesis, and Behavioral Recovery After Neocortical Infarction in Rats. *Stroke* **26**, 2135–2144 (1995).

109. Lindvall, O. & Kokaia, Z. Neurogenesis following Stroke Affecting the Adult Brain. *Cold Spring Harb. Perspect. Biol.* **7**, (2015).

110. Famakin, B. M. The Immune Response to Acute Focal Cerebral Ischemia and Associated Post-stroke Immunodepression: A Focused Review. *Aging Dis.* **5**, 307–326 (2014).

111. Becker, K. J. Modulation of the postischemic immune response to improve stroke outcome. *Stroke* **41**, S75–S78 (2010).

112. Apara, A. *et al.* KLF9 and JNK3 Interact to Suppress Axon Regeneration in the Adult CNS. *J. Neurosci.* **37**, 9632 LP – 9644 (2017).

113. Moore, D. L., Apara, A. & Goldberg, J. L. Krüppel-like transcription factors in the nervous system: novel players in neurite outgrowth and axon regeneration. *Mol. Cell. Neurosci.* **47**, 233–243 (2011).

114. Tang, X., Liu, K., Hamblin, M. H., Xu, Y. & Yin, K.-J. Genetic Deletion of Krüppel-Like Factor 11 Aggravates Ischemic Brain Injury. *Mol. Neurobiol.* **55**, 2911–2921 (2018).

115. Shi, H. *et al.* Kruppel-like factor 2 protects against ischemic stroke by regulating endothelial blood brain barrier function. *Am. J. Physiol. Heart Circ. Physiol.* **304**, H796–H805 (2013).

116. Yin, K.-J. *et al.* KLF11 mediates PPARγ cerebrovascular protection in ischaemic stroke. *Brain* **136**, 1274–1287 (2013).

117. Hu, G.-Q. *et al.* Inhibition of cerebral ischemia/reperfusion injury-induced apoptosis: nicotiflorin and JAK2/STAT3 pathway. *Neural Regen. Res.* **12**, 96–102 (2017).

118. O'Callaghan, J. P., Kelly, K. A., VanGilder, R. L., Sofroniew, M. V & Miller, D. B. Early activation of STAT3 regulates reactive astrogliosis induced by diverse forms of neurotoxicity. *PLoS One* **9**, e102003–e102003 (2014).

119. Schep, A. N., Wu, B., Buenrostro, J. D. & Greenleaf, W. J. chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. *Nat. Methods* **14**, 975 (2017).

120. Wang, C.-Y. *et al.* Function of B-Cell CLL/Lymphoma 11B in Glial Progenitor Proliferation and Oligodendrocyte Maturation   . *Frontiers in Molecular Neuroscience*   **11**, 4 (2018).

121. Huang, J. K. *et al.* Retinoid X receptor gamma signaling accelerates CNS remyelination. *Nat. Neurosci.* **14**, 45–53 (2011).

122. Yin, K.-J., Hamblin, M., Fan, Y., Zhang, J. & Chen, Y. E. Krüpple-like factors in the

central nervous system: novel mediators in stroke. *Metab. Brain Dis.* **30**, 401–410 (2015).

123. Yoda, T. *et al.* Site-specific gene expression analysis using an automated tissue micro-dissection punching system. *Sci. Rep.* **7**, 4325 (2017).

124. Cao, J. *et al.* Comprehensive single-cell transcriptional profiling of a multicellular organism. *Science (80-.).* **357**, 661–667 (2017).

125. Ramani, V. *et al.* Massively multiplex single-cell Hi-C. *Nat. Methods* **14**, 263–266 (2017).

126. Mulqueen, R. M. *et al.* Highly scalable generation of DNA methylation profiles in single cells. *Nat. Biotechnol.* **36**, 428–431 (2018).

127. Cao, J. *et al.* Joint profiling of chromatin accessibility and gene expression in thousands of single cells. *Science (80-.).* **361**, 1380 LP – 1385 (2018).

128. Longa, E. Z., Weinstein, P. R., Carlson, S. & Cummins, R. Reversible middle cerebral artery occlusion without craniectomy in rats. *Stroke* **20**, 84–91 (1989).

129. Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).

130. Picelli, S. *et al.* Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. *Genome Res.* **24**, 2033–2040 (2014).

131. Thornton, C. A. & Adey, A. C. sciMAP-ATAC. *Protocols.io* (2021). doi:dx.doi.org/10.17504/protocols.io.brhxm37n

132. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589–595 (2010).

133. Zhang, Y. *et al.* Model-based Analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).

134. Bravo González-Blas, C. *et al.* cisTopic: cis-regulatory topic modeling on single-cell ATAC-seq data. *Nat. Methods* **16**, 397–400 (2019).

135. Gu, Z., Eils, R. & Schlesner, M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* (2016).

136. Levine, J. H. *et al.* Data-Driven Phenotypic Dissection of AML Reveals Progenitor-like Cells that Correlate with Prognosis. *Cell* **162**, 184–197 (2015).

137. Neph, S. *et al.* BEDOPS: high-performance genomic feature operations. *Bioinformatics* **28**, 1919–1920 (2012).

138. McLean, C. Y. *et al.* GREAT improves functional interpretation of cis-regulatory regions. *Nat. Biotechnol.* **28**, 495–501 (2010).

139. Cao, J. *et al.* The single-cell transcriptional landscape of mammalian organogenesis. *Nature* **566**, 496–502 (2019).

140. Stuart, T., Srivastava, A., Lareau, C. & Satija, R. Multimodal single-cell chromatin analysis with Signac. *bioRxiv* 2020.11.09.373613 (2020). doi:10.1101/2020.11.09.373613

141. Pliner, H. A. *et al.* Cicero Predicts cis-Regulatory DNA Interactions from Single-Cell Chromatin Accessibility Data. *Mol. Cell* **71**, 858-871.e8 (2018).

142. Cao, J. *et al.* A human cell atlas of fetal gene expression. *Science (80-.).* **370**, eaba7721 (2020).

143. Lee, E. & Chung, W.-S. Glial Control of Synapse Number in Healthy and Diseased Brain. *Front. Cell. Neurosci.* **13**, 42 (2019).

144. Escartin, C. *et al.* Reactive astrocyte nomenclature, definitions, and future directions. *Nat. Neurosci.* **24**, 312–325 (2021).

145. Ransohoff, R. M. A polarizing question: do M1 and M2 microglia exist? *Nat. Neurosci.* **19**, 987–991 (2016).

146. Dunham, I., Kundaje, A. & Aldred, S. F. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).

147. Thornton, C. A. *et al.* Spatially mapped single-cell chromatin accessibility. *Nat. Commun.* **12**, 1274 (2021).

148. Lake, B. B. *et al.* Integrative Single-Cell Analysis By Transcriptional And Epigenetic States In Human Adult Brain. *Nat. Biotechnol.* **36**, 70–80 (2017).

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

150. Sherrod, M., Liu, X., Zhang, X. & Sigmund, C. D. Nuclear localization of angiotensinogen in astrocytes. *Am. J. Physiol. Integr. Comp. Physiol.* **288**, R539–R546 (2005).

151. Zhong, Y., Yu, C. & Qin, W. LncRNA SNHG14 promotes inflammatory response induced by cerebral ischemia/reperfusion injury through regulating miR-136-5p /ROCK1. *Cancer Gene Ther.* **26**, 234–247 (2019).

152. Hol, E. M. & Pekny, M. Glial fibrillary acidic protein (GFAP) and the astrocyte intermediate filament system in diseases of the central nervous system. *Curr. Opin. Cell Biol.* **32**, 121–130 (2015).

153. Clarke, L. E. *et al.* Normal aging induces A1-like astrocyte reactivity. *Proc. Natl. Acad. Sci.* **115**, E1896 LP-E1905 (2018).

154. O'Leary, L. A. *et al.* Characterization of Vimentin-Immunoreactive Astrocytes in the Human Brain. *Front. Neuroanat.* **14**, 31 (2020).

155. Bortell, N. *et al.* Astrocyte-specific overexpressed gene signatures in response to methamphetamine exposure in vitro. *J. Neuroinflammation* **14**, 49 (2017).

156. Gómez-López, S. *et al.* Neural ablation of the PARK10 candidate Plpp3 leads to dopaminergic transmission deficits without neurodegeneration. *Sci. Rep.* **6**, 24028 (2016).

157. Balderas, P. M. de O. NMDA Receptors in Astroglia: Chronology, Controversies, and Contradictions from a Complex Molecule. in (ed. D'Amato, J. R. G. H. E.-M. T. G. E.-L. C.) Ch. 4 (IntechOpen, 2018). doi:10.5772/intechopen.72975

158. Cheli, V. T. *et al.* L-type voltage-operated calcium channels contribute to astrocyte

activation In vitro. *Glia* **64**, 1396–1415 (2016).

159. Boscia, F., Elkjaer, M. L., Illes, Z. & Kukley, M. Altered Expression of Ion Channels in White Matter Lesions of Progressive Multiple Sclerosis: What Do We Know About Their Function? . *Frontiers in Cellular Neuroscience* **15**, (2021).

160. Miller, R. L. & Loewy, A. D. ENaC γ-expressing astrocytes in the circumventricular organs, white matter, and ventral medullary surface: sites for Na+ regulation by glial cells. *J. Chem. Neuroanat.* **53**, 72–80 (2013).

161. Prada, I. *et al.* REST/NRSF governs the expression of dense-core vesicle gliosecretion in astrocytes. *J. Cell Biol.* **193**, 537–549 (2011).

162. Araque, A. *et al.* Gliotransmitters Travel in Time and Space. *Neuron* **81**, 728–739 (2014).

163. Henneberger, C., Papouin, T., Oliet, S. H. R. & Rusakov, D. A. Long-term potentiation depends on release of d-serine from astrocytes. *Nature* **463**, 232–236 (2010).

164. Batiuk, M. Y. *et al.* Identification of region-specific astrocyte subtypes at single cell resolution. *Nat. Commun.* **11**, 1220 (2020).

165. Guimarães-Camboa, N. *et al.* Pericytes of Multiple Organs Do Not Behave as Mesenchymal Stem Cells In Vivo. *Cell Stem Cell* **20**, 345-359.e5 (2017).

166. Yamazaki, T. & Mukouyama, Y. Tissue Specific Origin, Development, and Pathological Perspectives of Pericytes . *Frontiers in Cardiovascular Medicine* **5**, (2018).

167. Winkler, E. A., Bell, R. D. & Zlokovic, B. V. Central nervous system pericytes in health and disease. *Nat. Neurosci.* **14**, 1398–1405 (2011).

168. Craig, M. P. & Sumanas, S. ETS transcription factors in embryonic vascular development. *Angiogenesis* **19**, 275–285 (2016).

169. Lelièvre, E., Lionneton, F., Mattot, V., Spruyt, N. & Soncin, F. Ets-1 Regulates fli-1 Expression in Endothelial Cells: IDENTIFICATION OF ETS BINDING SITES IN THE fli-1 GENE PROMOTER *. *J. Biol. Chem.* **277**, 25143–25151 (2002).

170. Liddelow, S. A. & Barres, B. A. Reactive Astrocytes: Production, Function, and Therapeutic Potential. *Immunity* **46**, 957–967 (2017).

171. Zamanian, J. L. *et al.* Genomic analysis of reactive astrogliosis. *J. Neurosci.* **32**, 6391–6410 (2012).

172. Potokar, M., Morita, M., Wiche, G. & Jorgačevski, J. The Diversity of Intermediate Filaments in Astrocytes. *Cells* **9**, 1604 (2020).

173. Dozio, V. & Sanchez, J.-C. Profiling the proteomic inflammatory state of human astrocytes using DIA mass spectrometry. *J. Neuroinflammation* **15**, 331 (2018).

174. Tian, R., Gregor, M., Wiche, G. & Goldman, J. E. Plectin regulates the organization of glial fibrillary acidic protein in Alexander disease. *Am. J. Pathol.* **168**, 888–897 (2006).

175. Sofroniew, M. V. Astrocyte Reactivity: Subtypes, States, and Functions in CNS Innate Immunity. *Trends Immunol.* **41**, 758–770 (2020).

176. Sakamoto, Y., Boëda, B. & Etienne-Manneville, S. APC binds intermediate filaments and is required for their reorganization during cell migration. *J. Cell Biol.* **200**, 249–258 (2013).

177. Parakalan, R. *et al.* Transcriptome analysis of amoeboid and ramified microglia isolated from the corpus callosum of rat brain. *BMC Neurosci.* **13**, 64 (2012).

178. Bai, X., Karasmanis, E. P. & Spiliotis, E. T. Septin 9 interacts with kinesin KIF17 and interferes with the mechanism of NMDA receptor cargo binding and transport. *Mol. Biol. Cell* **27**, 897–906 (2016).

179. Ben-Gigi, L. *et al.* Astrogliosis Induced by Brain Injury Is Regulated by Sema4B Phosphorylation. *eNeuro* **2**, ENEURO.0078-14.2015 (2015).

180. Cross, A. K. *et al.* ADAMTS-1 and -4 are up-regulated following transient middle cerebral artery occlusion in the rat and their expression is modulated by TNF in cultured astrocytes. *Brain Res.* **1088**, 19–30 (2006).

181. Lemarchant, S. *et al.* ADAMTS proteoglycanases in the physiological and pathological central nervous system. *J. Neuroinflammation* **10**, 899 (2013).

182. Ginhoux, F. *et al.* Fate mapping analysis reveals that adult microglia derive from primitive macrophages. *Science* **330**, 841–845 (2010).

183. Saijo, K. & Glass, C. K. Microglial cell origin and phenotypes in health and disease. *Nat. Rev. Immunol.* **11**, 775–787 (2011).

184. Haruwaka, K. *et al.* Dual microglia effects on blood brain barrier permeability induced by systemic inflammation. *Nat. Commun.* **10**, 5816 (2019).

185. Hirbec, H., Rassendren, F. & Audinat, E. Microglia Reactivity: Heterogeneous Pathological Phenotypes. *Methods Mol. Biol.* **2034**, 41–55 (2019).

186. Schafer, D. P. & Stevens, B. Microglia Function in Central Nervous System Development and Plasticity. *Cold Spring Harb. Perspect. Biol.* **7**, a020545–a020545 (2015).

187. Izquierdo, P., Attwell, D. & Madry, C. Ion Channels and Receptors as Determinants of Microglial Function. *Trends Neurosci.* **42**, 278–292 (2019).

188. Nishiyama, A., Komitova, M., Suzuki, R. & Zhu, X. Polydendrocytes (NG2 cells): multifunctional cells with lineage plasticity. *Nat. Rev. Neurosci.* **10**, 9–22 (2009).

189. Akay, L. A., Effenberger, A. H. & Tsai, L.-H. Cell of all trades: oligodendrocyte precursor cells in synaptic, vascular, and immune function. *Genes Dev.* **35**, 180–198 (2021).

190. Chamling, X. *et al.* Single-cell transcriptomic reveals molecular diversity and developmental heterogeneity of human stem cell-derived oligodendrocyte lineage cells. *Nat. Commun.* **12**, 652 (2021).

191. Roggeri, A. *et al.* Sphingosine-1-Phosphate Receptor Modulators and Oligodendroglial Cells: Beyond Immunomodulation. *Int. J. Mol. Sci.* **21**, 7537 (2020).

192. Jaillard, C. *et al.* Edg8/S1P5: an oligodendroglial receptor with dual function on process retraction and cell survival. *J. Neurosci.* **25**, 1459–1469 (2005).

193. Yu, N., Lariosa-Willingham, K. D., Lin, F.-F., Webb, M. & Rao, T. S. Characterization of lysophosphatidic acid and sphingosine-1-phosphate-mediated signal transduction in rat cortical oligodendrocytes. *Glia* **45**, 17–27 (2004).

194. Tiane, A. *et al.* From OPC to Oligodendrocyte: An Epigenetic Journey. *Cells* **8**, 1236 (2019).

195. Tomassy, G. S., Dershowitz, L. B. & Arlotta, P. Diversity Matters: A Revised Guide to Myelination. *Trends Cell Biol.* **26**, 135–147 (2016).

196. Fields, R. D. White matter in learning, cognition and psychiatric disorders. *Trends Neurosci.* **31**, 361–370 (2008).

197. Thornton, C. A. *et al.* Spatially mapped single-cell chromatin accessibility. *bioRxiv* 815720 (2020). doi:10.1101/815720

198. Boisvert, M. M., Erikson, G. A., Shokhirev, M. N. & Allen, N. J. The Aging Astrocyte Transcriptome from Multiple Regions of the Mouse Brain. *Cell Rep.* **22**, 269–285 (2018).

199. Niraula, A., Sheridan, J. F. & Godbout, J. P. Microglia Priming with Aging and Stress. *Neuropsychopharmacology* **42**, 318–333 (2017).

200. Mulqueen, R. M., Fields, A. J. & Adey, A. C. Loading Tn5 Enzyme with sci-protocol Oligonucleotides. *Protocols.io* (2021). doi:dx.doi.org/10.17504/protocols.io.6zahf2e

201. Cusanovich, D. A. *et al.* The cis-regulatory dynamics of embryonic development at single-cell resolution. *Nature* **555**, 538–542 (2018).

202. Adey, A. scitools. **https://gi**, (2020).

203. McLean, C. Y. *et al.* GREAT improves functional interpretation of cis-regulatory regions. *Nat. Biotechnol.* **28**, 495–501 (2010).

204. Supek, F., Bošnjak, M., Škunca, N. & Šmuc, T. REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One* **6**, e21800–e21800 (2011).

205. Polioudakis, D. *et al.* A Single-Cell Transcriptomic Atlas of Human Neocortical Development during Mid-gestation. *Neuron* **103**, 785-801.e8 (2019).

206. Rozenblatt-Rosen, O. *et al.* The Human Tumor Atlas Network: Charting Tumor Transitions across Space and Time at Single-Cell Resolution. *Cell* **181**, 236–249 (2020).

207. Jerby-Arnon, L. *et al.* A Cancer Cell Program Promotes T Cell Exclusion and Resistance to Checkpoint Blockade. *Cell* **175**, 984-997.e24 (2018).

208. Reyfman, P. A. *et al.* Single-Cell Transcriptomic Analysis of Human Lung Provides Insights into the Pathobiology of Pulmonary Fibrosis. *Am. J. Respir. Crit. Care Med.* **199**, 1517–1536 (2018).

209. Montoro, D. T. *et al.* A revised airway epithelial hierarchy includes CFTR-expressing ionocytes. *Nature* **560**, 319–324 (2018).

210. Zheng, G. X. Y. *et al.* Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* **8**, 14049 (2017).