NEUROSCIENCE

A single-cell multi-omic atlas spanning the adult rhesus macaque brain

Kenneth L. Chiou[1,2]†, Xingfan Huang[3,4]†, Martin O. Bohlen[5], Sébastien Tremblay[6], Alex R. DeCasien[7], Diana R. O'Day[8], Cailyn H. Spurrell[8,9], Aishwarya A. Gogate[8,9], Trisha M. Zintel[1,2]‡, Cayo Biobank Research Unit, Madeline G. Andrews[10], Melween I. Martínez[11], Lea M. Starita[3,8], Michael J. Montague[6], Michael L. Platt[6,12,13]§*, Jay Shendure[3,8,14,15]§*, Noah Snyder-Mackler[1,2,16,17]§*

Cataloging the diverse cellular architecture of the primate brain is crucial for understanding cognition, behavior, and disease in humans. Here, we generated a brain-wide single-cell multimodal molecular atlas of the rhesus macaque brain. Together, we profiled 2.58 M transcriptomes and 1.59 M epigenomes from single nuclei sampled from 30 regions across the adult brain. Cell composition differed extensively across the brain, revealing cellular signatures of region-specific functions. We also identified 1.19 M candidate regulatory elements, many previously unidentified, allowing us to explore the landscape of cis-regulatory grammar and neurological disease risk in a cell type–specific manner. Altogether, this multi-omic atlas provides an open resource for investigating the evolution of the human brain and identifying novel targets for disease interventions.

INTRODUCTION

The cellular and molecular origins of complex human thought and behavior remain largely a mystery. Historically, proposed explanations have centered on the substantial relative size (1–3), prodigious number of cells (4), or the large cortical surface area and thickness (5) of the human brain. These explanations in isolation, however, fail to explain the many uniquely human faculties, nor do they explain the extreme variety and complexity of impairments that accompany human neurodevelopmental, neuropsychiatric, and neurodegenerative disorders (6). The human brain is composed of myriad cell types, and this cellular heterogeneity presumably contributes to our cognitive and behavioral complexity (7, 8). Supporting this hypothesis is the observation that the number of distinct cell types in the brain is positively correlated with behavioral complexity across vertebrates (9). In recent decades, it has been proposed that certain aspects of human cognition are supported by specific cell types such as von Economo neurons (10) and "mirror neurons" (11), which have been hypothesized to support social intuition and empathy, respectively. These propositions, however, remain largely untested due to gaps in our understanding of the cellular landscape of the human brain and, crucially, differences in cell type composition and regional heterogeneity among the brains of humans, nonhuman primates, and other animals.

In recent years, the application of rapidly developing single-cell technologies to the brain has begun to address these gaps. Single-cell molecular surveys of targeted regions of the mouse and human brain, for example, have revealed specialized species-specific cell types—e.g., rosehip neurons in humans (12)—and regional biases in cell type distribution and function [e.g., (13)]. Such atlases are yielding unprecedented cross-species insights into the cellular architecture supporting the structure and function of the brain (14, 15), but the general paucity of comparative nonhuman primate brain atlases has left a conspicuous gap (16). Moreover, much effort has focused on single molecular modalities (e.g., transcriptomics), typically in only one or a few regions, leaving a lacuna in our understanding of the molecular mechanisms underlying cell function across much of the primate brain.

Here, we generated a 4.2 million cell (combined) transcriptomic and epigenomic atlas across the brain of the rhesus macaque (*Macaca mulatta*), the most widely used nonhuman primate model organism for studies of human perception, cognition, aging, and neurological disease (17). These single-cell profiles derive from 30 distinct brain regions that collectively represent major cortical, subcortical, and cerebellar areas involved in sensory, cognitive, emotional, and motor functions. Many of these regions are also implicated in one or more clinically relevant neurological disorders. By integrating measures of gene expression and chromatin accessibility, we discover molecular signatures that define cell types across the macaque brain, characterize their distribution and molecular function across disparate anatomical regions, and nominate sets of cis-regulatory regions that likely contribute to mature cell fate and function across the brain.

[1]Center for Evolution and Medicine, Arizona State University, Tempe, AZ, USA. [2]School of Life Sciences, Arizona State University, Tempe, AZ, USA. [3]Department of Genome Sciences, University of Washington, Seattle, WA, USA. [4]Paul G. Allen School of Computer Science and Engineering, University of Washington, Seattle, WA, USA. [5]Department of Biomedical Engineering, Duke University, Durham, NC, USA. [6]Department of Neuroscience, University of Pennsylvania, Philadelphia, PA, USA. [7]Section on Developmental Neurogenomics, National Institute of Mental Health, Bethesda, MD, USA. [8]Brotman Baty Institute for Precision Medicine, Seattle, WA, USA. [9]Seattle Children's Research Institute, Seattle, WA, USA. [10]School of Biological and Health Systems Engineering, Arizona State University, Tempe, AZ, USA. [11]Caribbean Primate Research Center, University of Puerto Rico, San Juan, PR, USA. [12]Department of Psychology, University of Pennsylvania, Philadelphia, PA, USA. [13]Marketing Department, University of Pennsylvania, Philadelphia, PA, USA. [14]Howard Hughes Medical Institute, Seattle, WA, USA. [15]Allen Discovery Center for Cell Lineage Tracing, Seattle, WA, USA. [16]School of Human Evolution and Social Change, Arizona State University, Tempe, AZ, USA. [17]ASU-Banner Neurodegenerative Disease Research Center, Arizona State University, Tempe, AZ, USA.
*Corresponding author. Email: nsnyderm@asu.edu (N.S.-M.); shendure@uw.edu (J.S.); mplatt@pennmedicine.upenn.edu (M.P.)
†These authors contributed equally to this work.
‡Present address: Sage Bionetworks, 2901 3rd Ave. Ste. 330, Seattle, Washington, 98121, USA.
§These authors contributed equally to this work.

RESULTS

A molecular taxonomy of cell types across the primate brain

We generated single-nucleus RNA sequencing (snRNA-seq) data from 30 distinct regions across the cortex, subcortex, cerebellum, and brainstem ($N = 5$ animals; three females) using three-level single-cell combinatorial indexing RNA-seq (sci-RNA-seq3) (Fig. 1A and table S1) (18, 19). With the original sci-RNA-seq3 protocol (20), we generated 1,008,204 single-nucleus transcriptomes from 110 age-, sex-, and hemisphere-matched samples representing 28 brain regions of 10 year-old (mid-adult aged) macaques ($N = 3$ animals; two females). Over the course of the study, we implemented improvements in nuclei isolation and preservation (21), which increased nuclear transcriptome recovery by ~60% [median unique molecular indices (UMIs), before = 202, after = 320] and, consequently, the number of nuclei passing our UMI threshold. With the improved protocol, we generated an additional 1,702,081 single-nucleus transcriptomes from the right hemisphere of two animals, the vast majority ($N = 1,579,908$) of which were sampled from 27 brain regions of a single 10-year-old female macaque. Altogether, after applying quality control (QC) filters (Materials and Methods and figs. S1 and S2), we recovered transcriptome profiles for 2,583,967 nuclei (median UMI per cell = 265, median genes expressed per cell = 221; table S2).

Controlling for batch effects across sequencing runs (Materials and Methods and fig. S4), we jointly clustered single-cell profiles across all sampled brain regions to identify 17 molecularly distinct cell types, which we refer to as "cell classes" (Fig. 1, B and C). On the basis of established cell markers (fig. S5 and table S3), we annotated these 17 cell classes as either: (i) neuronal cells, including cortical glutamatergic neurons (*CAMK2A*), cortical GABAergic neurons (*GAD1* and *GAD2*), basket cells (*GRID2* and *SORCS3*), other cerebellar neurons (primarily granule cells; *GRM4*), medium spiny neurons (*DACH1*, *PPP1R1B*, and *BCL11B*), serotonergic neurons (*TPH2*), dopaminergic neurons (*TH* and *DBH*); or (ii) non-neuronal cells, including microglia (*DOCK2*), oligodendrocyte precursor cells (OPCs; *VCAN*), astrocytes (*ALDH1A1* and *GFAP*), oligodendrocytes (*MOG* and *MBP*), vascular cells (*CFH*), and ependymal cells (*FOXJ1*). Our broad survey also captured four rare cell populations that, to our knowledge, have not yet been identified in other studies: three *RBFOX3*+ (NeuN+) neuron-like populations (marker genes: *APOA2*, $N = 7055$ cells; *F5*, $N = 880$; *KIR2DL1/2*, $N = 84$) and one *RBFOX3*− microglia-like population (marker gene: *KIR3DL1/2*+, $N = 44$ cells, also *P2RY12*+/*PTPRC*+/*ENTPD1*+). Given their rarity, we removed these four cell populations from downstream analyses. Hierarchical clustering of cell classes by the top 50 principal components of gene expression largely recapitulated broad ontogenetic relationships, with most neuronal classes clustering together (dopaminergic neurons being the exception) and the two mesoderm-derived classes (microglia and vascular cells) clustering together (fig. S6A). Further, analysis of pathways associated with gene expression of each cell class revealed known molecular physiological processes characteristic of cell classes (tables S4 and S5). For example, we found that gene expression in glutamatergic and GABAergic neurons were associated with synapse assembly/function, oligodendrocyte gene expression was associated with myelination, and microglia gene expression was associated with immune processes including inflammation. Dopaminergic and serotonergic neurons had high activation levels of expected pathways such as tyrosine hydroxylase/catecholamine and serotonin/melatonin biosynthesis, respectively.

By sampling across a broad range of anatomical regions within the same individuals, we were able to characterize cellular composition across 30 distinct brain regions—to our knowledge, the most regionally expansive nonhuman primate single-cell brain atlas to date (Fig. 1, D and E). The distribution of major cell classes was balanced between sexes and hemispheres (fig. S7) but differed extensively across regions, reflecting the cellular makeup underlying region-specific functions (Fig. 1E). Unsupervised hierarchical clustering of brain regions according to cell class composition for the most part conformed to broader anatomical categorizations, with regions of the cortex, subcortex, brainstem, and cerebellum usually grouping together (fig. S6B), which was also the case when clustering regions based on the top 50 principal components of gene expression (fig. S6B). Two of these four broad regional classes were composed primarily of a single cell class: In the cortex ($N = 16$ regions, table S6), glutamatergic neurons were the most abundant cell type (mean = 63.7% of all cells per sample) and outnumbered GABAergic neurons by almost fourfold (Fig. 1E; mean = 17.4%), while the cerebellum ($N = 2$ regions) was composed almost entirely of cerebellar neurons (mean = 85.1%). In contrast, the subcortex ($N = 8$ regions) and brainstem ($N = 4$ regions) were more heterogeneous with respect to their cellular composition, with samples from these regions containing roughly equal proportions of glutamatergic neurons (mean$_{subcortex}$ = 25.1%; mean$_{brainstem}$ = 25.5%), GABAergic neurons (mean$_{subcortex}$ = 20.2%; mean$_{brainstem}$ = 23.0%), and oligodendrocytes (mean$_{subcortex}$ = 18.5%; mean$_{brainstem}$ = 25.5%). We further subdivided the cortical and subcortical samples into "region subclasses" based on neuroanatomical groups (table S1), within which variation in cellular composition was more limited (Fig. 1E). For instance, in the subcortex, medium spiny neurons (MSNs) comprised around half of the cells in the basal ganglia [nucleus accumbens (NAc) mean = 44.7%; caudate nucleus (CN) mean = 60.0% MSNs], while the thalamus was enriched for GABAergic neurons [lateral geniculate nucleus (LGN) mean = 55.7%; mediodorsal thalamic nucleus (mdTN) mean = 43.8%; ventrolateral thalamic nucleus (vlTN) mean = 28.6%].

Our broad survey also captured two rarer, but important, cell classes: dopaminergic and serotonergic neurons. These two classes of neurons collectively represented less than 0.3% of all profiled cells (dopaminergic = 0.14%; serotonergic = 0.12% of all cells) and 0.5% of all neurons (dopaminergic = 0.19%; serotonergic = 0.17% of all neurons), suggesting that targeted approaches that enrich for these cells [e.g., (22, 23)] are necessary to identify transcriptional variation among subtypes. Dopaminergic neurons, which are found primarily in the substantia nigra pars compacta at low frequency [1.1% of cells sampled in the midbrain (MB) versus mean 0.1% in other sampled regions], are involved in a range of important processes, including voluntary movement, reinforcement learning, and addiction, and their loss is a neuropathological hallmark of Parkinson's disease (24). We found that serotonergic neurons were most abundant in the brainstem (mean 0.35% in the four brainstem regions versus mean 0.09% in other sampled regions), where they play a major role in sleep, mood, and appetite, and are key targets of pharmacological therapies for major depressive disorder in humans (25).
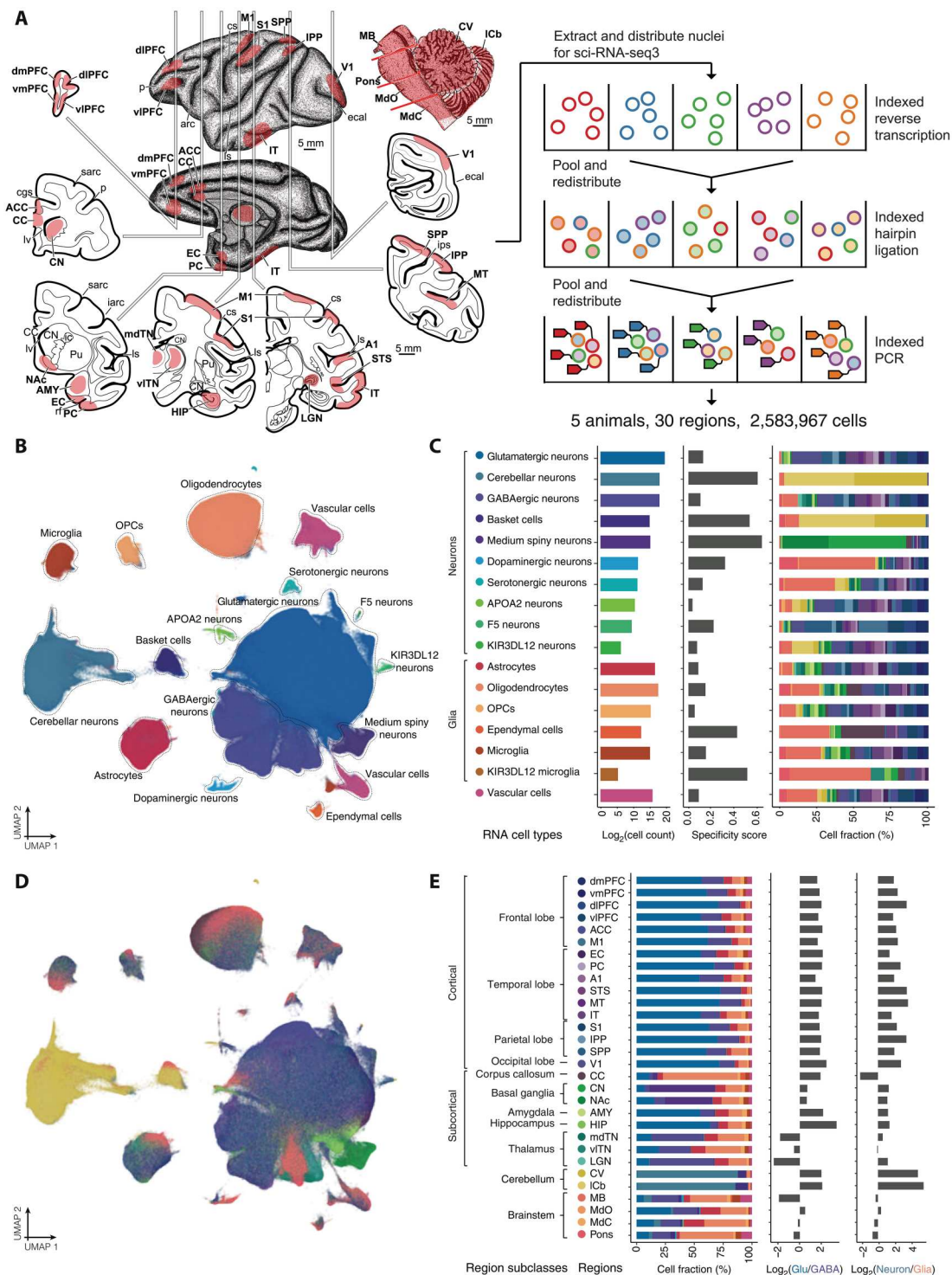
Fig. 1. Experimental setup and summary of the Macaque Brain Atlas snRNA-seq dataset. (**A**) Schematic of biopsied brain regions for sci-RNA-seq3 experiment. A full list of sampled regions is provided in table S1. arc, arcuate sulcus; cgs, cingulate sulcus; cs, central sulcus; ecal, external calcarine sulcus; iarc, inferior arcuate sulcus; ic, internal capsule; ips, intraparietal sulcus; ls, lateral sulcus; lv, lateral ventricle; p, principal sulcus; rf, rhinal fissure; sarc, superior arcuate sulcus. (**B**) UMAP visualization of all snRNA-seq profiled cells colored by cell type [with color code shown in (C)]. (**C**) Barplots showing the log$_2$-transformed cell counts (left), regional specificity score (middle), and regional composition [right, with color code shown in (E)] of each cell type. (**D**) UMAP visualization of all snRNA-seq cells colored by cell type [with color code shown in (E)]. (**E**) Barplots showing the cell type composition [left, with color code shown in (C)], log$_2$-transformed ratio of glutamatergic neurons and GABAergic neurons (middle), and log$_2$-transformed ratio of neurons and glial cells (right) of each region. Regions are organized by the regional subclass to which they belong.

Regional variation in cell subtype composition

To characterize heterogeneity within cell classes, we partitioned the dataset and repeated preprocessing and clustering separately for each of the 17 cell classes. Collectively, we identified 112 distinct clusters (fig. S8 and table S7), using a coarser clustering criterion, and 397 distinct clusters (table S8) using a more fine-grained criterion, which captured neuronal and non-neuronal diversity across the primate brain (Fig. 2A). We refer to clusters at the coarser level as "cell subtypes" and clusters at the finer-grained level as "cell subclusters." We identified extensive heterogeneity in glutamatergic (39 subtypes) and GABAergic (20 subtypes) neurons primarily found in the cortex and some regions of the subcortex [e.g., hippocampus (HIP) and thalamus], while neurons derived from other noncortical brain regions (e.g., cerebellum and striatum) were transcriptionally distinct and relatively homogeneous within those regions (Fig. 2A). This is due in part to the large number of specialized neurons present in some of these regions, including granule and Purkinje cells in the cerebellum, and medium spiny neurons in the basal ganglia (table S7).

Our systematic approach also allowed us to characterize and compare the regional cellular distribution of non-neuronal subtypes, including those of glia, which have not often been the focus on most single-cell atlases to date (Fig. 2A and table S9). Overall, we identified six astrocyte, two microglial, seven oligodendrocyte, and six vascular cell subtypes, the latter including endothelial cells, smooth muscle cells, pericytes, and both perivascular and meningeal fibroblasts (fig. S8 and table S7) (26). We compared cell subtypes to published datasets using a non-negative least squares (NNLS) approach (19) and found broad correspondence with subtypes observed in human cortical (14), human brain vascular (26), and macaque hippocampal atlases (fig. S10, A to E) (27).

To identify cell subtypes or subclusters that were specific or biased toward a single region or set of regions, we calculated a measure of "regional specificity" using the Jensen-Shannon divergence statistic (Materials and Methods) (28, 29). Overall, glial subtypes were more evenly distributed across all regions compared to neuronal subtypes (Fig. 2A). This is reflected in lower Jensen-Shannon specificity scores for glial subtypes (mean = 0.20; median = 0.15; range = [0.04,0.81]) compared to cortical neurons (mean = 0.31; median = 0.18; range = [0.08,0.89]). A number of cell subtypes, both neuronal and non-neuronal, were highly region specific. For instance, oligodendrocyte subtype 8, the rarest oligodendrocyte subtype (N = 3439 cells; 1.5% of oligodendrocytes) overwhelmingly derived from the highly myelinated corpus callosum (CC; 93.0% of these cells; Fig. 2A and table S9). Among cortical neurons, GABAergic interneuron subtypes generally exhibited a lower median regional specificities than to glutamatergic neuron subtypes, although there were a number of interneuron subtypes specific to the thalamus (cluster 6) or brainstem (clusters 3 and 16), discussed below.

Given that the regional specificity of excitatory neuronal subtypes has been explored in depth in other studies [e.g., (30, 31)], we focus here instead on populations that are vital for neuronal signal transduction but for which cellular diversity has not previously been explored across the macaque brain. Specifically, we concentrated on the regional diversity of interneurons, because they are important components of long-range circuitry and have been characterized in a few regions across mice, monkeys, and humans [e.g., (12, 14, 15, 32)], allowing us to both benchmark our atlas and extend current knowledge to understudied regions. We also examine regional distribution among astrocytes, which are crucial for maintaining neuronal homeostasis (33) and are implicated in neurological disorders (34) but have been relatively understudied at the single-cell level.

We pursued three main approaches to dissect the regional heterogeneity within interneuron and astrocyte subtypes, discussed in further detail below: (i) quantification of cell subtype composition to identify nuanced differences in detailed regions within the cortex; (ii) identification of regionally specific gene expression programs by analyzing region specific subtypes of interneurons; and (iii) in the case of minimal region specific subtypes, leveraging a recently developed statistic to identify region-specific gene expression patterns in astrocyte subtypes in a cell subtype–agnostic fashion.

Within specific regions of the cortex, cell subtype composition differences become more subtle and require focused quantification. As a first approach, for every sufficiently abundant interneuron and astrocyte subtype in the cortex (>100 cells), we calculated the $\log_2$-transformed ratio of cell subtype composition in a region, compared to the average composition of that subtype across all cortical regions (Fig. 2B). Within the five most abundant interneuron subtypes, we not only note general balance across all cortical regions but also observe a relative enrichment of cluster 2 ($PVALB^+$) in the occipital lobe [primary visual cortex (V1)] with depletion in regions within the temporal lobe and depletion of cluster 5 ($ADARB/PAX6^+$) in V1. In the superior temporal sulcus (STS) and middle temporal visual area (MT), there is a strong depletion of astrocyte subtype 3 ($LUZP2/GPC5^+$) but an enrichment of subtype 6 ($KCNIP4/RBFOX1^+$).

Interneurons are the primary drivers of inhibitory control through the release of GABA (γ-aminobutyric acid) and thus strongly affect neural circuitry. Inappropriate development of GABAergic interneurons and subsequent loss of inhibitory regulation contribute to disorders of neurodevelopment, including epilepsy and autism (35, 36). Despite their importance, the molecular identities and distribution of interneuron subtypes across the adult primate brain remain relatively unknown outside of a few regions (15, 32, 37). Our snRNA-seq sample captured 371,548 GABAergic interneurons corresponding to 20 subtypes. As a second approach, we focused on gene markers of region-specific interneuron subtypes. Eleven interneuronal subtypes were primarily found in the cortex and could be assigned to four primary interneuron groups that are conserved between mouse and human brains (32), marked by SST, $PVALB$, VIP, and $LAMP5$ expression (fig. S11). Compared to the cortex, the brainstem and thalamus had a unique distribution of interneuron subtypes (Fig. 2C). Thalamic interneurons, which use feed-forward inhibition to relay and tune visual inputs to thalamocortical neurons, expressed high levels of $NTNG1\ and\ RNF220$ (Fig. 2D), which is indicative of long-range interneurons in the first-order relay nuclei of the thalamus (38). Sampling across the striatum, which is a critical part of the reward pathway and the largest part of the basal ganglia, a recent single-cell study identified a molecularly unique primate interneuron (32), which was most similar to our GABAergic cluster 18 and represented 15% of interneurons in the CN (Fig. 2A).

Astrocytes, the second most abundant non-neuronal cell type in our dataset, are multifaceted support cells of the brain that perform a variety of tasks related to neuronal homeostasis. These tasks can vary across brain regions (33), and astrocyte dysfunction has been
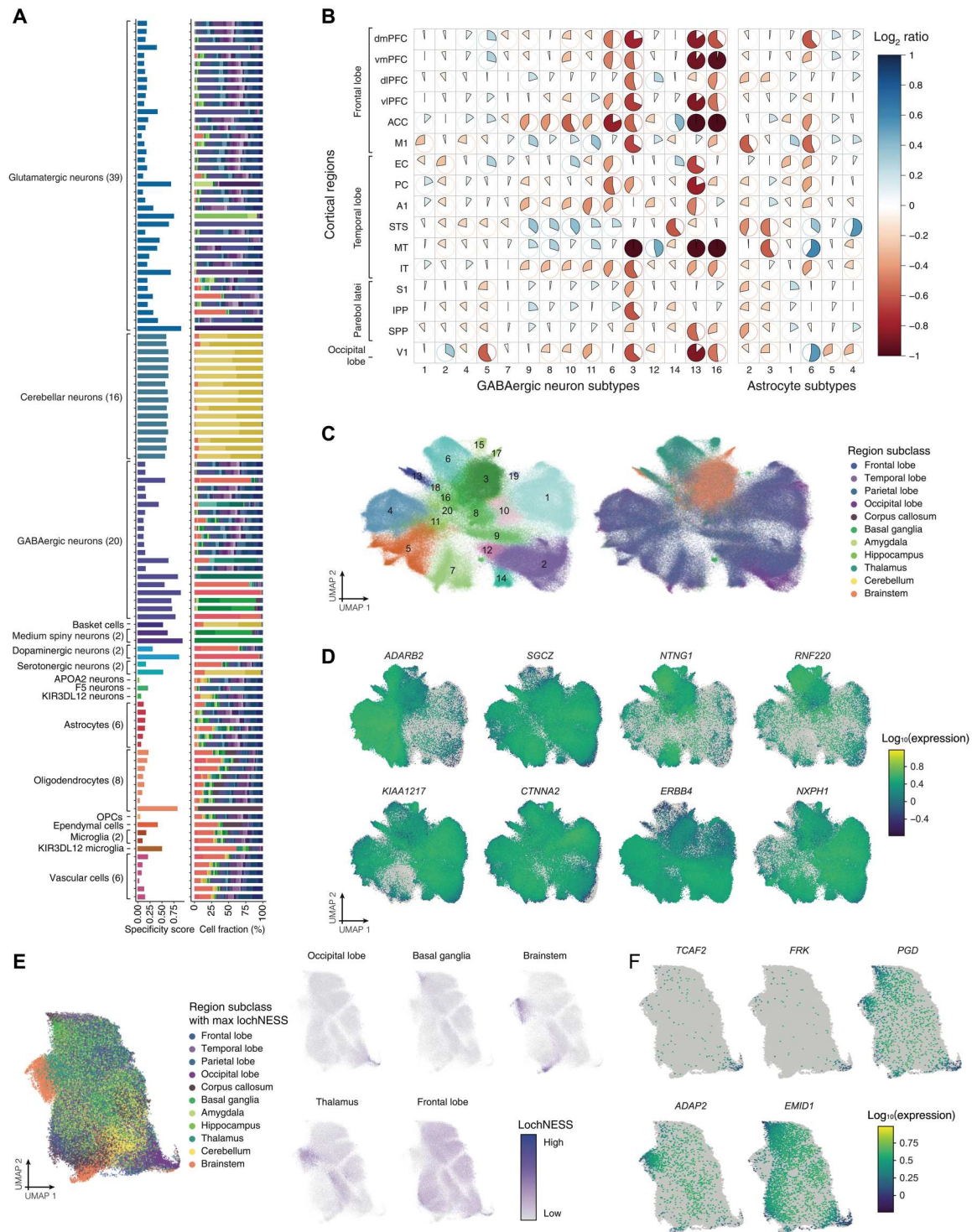
Fig. 2. Cell subtype distribution and variation across the brain. (**A**) Barplots showing the region specificity score (i.e., Jensen-Shannon divergence statistic) and composition for cell subtypes (with color code shown in Fig. 1E). (**B**) Heatmap showing scaled log$_2$ ratios of GABAergic neuron and astrocyte subtype compositions within cortical region, compared to the average across all regions. Cell subtypes with at least 100 cells profiled are shown in the order of abundance (*x* axis, left to right) in the cortical regions organized by region subclasses (*y* axis). The color and direction of each pie correspond to relative enrichment (blue, clockwise) and depletion (red, anticlockwise) of a cell subtype in a region. Log$_2$ ratios were capped at positive and negative 2 before scaling. (**C**) UMAP visualizations of GABAergic neurons colored by cell subtype (left) and regional subclass (right). (**D**) UMAP visualizations of GABAergic neurons colored by cell subtype marker gene expression. (**E**) UMAP visualization of astrocytes colored by the region with the highest lochNESS, indicating enrichment of a region subclass in the cell's transcriptional vicinity. LochNESS distribution in a few example regions (occipital lobe, basal ganglia, brainstem, thalamus, and frontal lobe) is highlighted in separate panels as examples. (**F**) UMAP visualizations of astrocytes colored by lochNESS-derived region-related marker genes.

linked to neurological diseases, including Alzheimer's disease (*39*). Given these regional differences, we examined whether astrocyte subtypes exhibited regional biases in macaque, similar to what has been observed in the mouse brain (*40*). However, while astrocyte subtypes were widely distributed across multiple regions, the cell clusters did not correspond neatly to regions of origin, making claims about inter-region differences in cell composition difficult to systematically analyze across the many regions profiled. To address this complexity, as a third approach, we adapted our recently developed statistic, lochNESS (*41*), to quantitatively measure regional enrichment in each cell's "neighborhood" of transcriptionally similar cells. Briefly, for each cell, we tally the number of cells from each brain region in its neighborhood and calculate a focal regional enrichment score (fig. S12A and Materials and Methods). We illustrate the utility of this approach by calculating the lochNESS score on astrocytes at the level of brain region subclasses. Each cell had 11 lochNESS scores calculated, one for each region subclass, with each such score quantifying the enrichment of the given region subclass in a cell's transcriptional vicinity. We then identified the most enriched region subclass in a cell's neighborhood and examined the regional heterogeneity agnostic to the cluster-assigned subtype labels (Fig. 2E). We also extended lochNESS to identify genes whose expression can be predicted by lochNESS scores for given regions. We modeled gene expression as a function of the lochNESS scores of each region in each cell using generalized linear regression (Materials and Methods and table S10). The resulting set of genes with significant positive associations with a region's lochNESS score have higher expression in, and are putatively markers for, cell subtypes in that region.

Using this approach, we identified markers for astrocytes in specific regions (e.g., *TCAF2* and *FRK* in the occipital lobe) and in combinations of regions (e.g., *PGD* in the brainstem, basal ganglia, and thalamus), which we would not have identified if we focused solely on discrete, computationally defined clusters (Fig. 2F). This strategy thus facilitates the identification of more complex region-specific gene expression patterns. For example, *EMID1*, which is a marker for a subpopulation of astrocyte-like NG2 cells (*42*), is more highly expressed in astrocytes in the cortex but not in the thalamus, brainstem, or cerebellum. In contrast, *ADAP2*, which is involved in protection from RNA virus infections (*43*), is highly specific to a subset of astrocytes found in the thalamus (Fig. 2F). LochNESS can thus provide a more nuanced approach to identifying regionally biased cell subtypes and gene expression than conventional clustering. While we focused on astrocytes in this example, lochNESS could be iteratively applied to regions within a subclass in each cell class, e.g., for all glutamatergic neurons across all cortical regions or oligodendrocytes across all subcortical regions (fig. S12, B to D).

Joint analysis of single-nucleus transcriptomic and epigenomic data
To complement our transcriptomic dataset and identify key regulatory genomic regions in brain cells, we applied three-level single-cell combinatorial indexing ATAC-seq (sci-ATAC-seq3) (*44*, *45*) to profile single-nucleus assay for transposase-accessible chromatin (ATAC) with sequencing (snATAC-seq) epigenomes from nearly all of the brain regions represented in our snRNA-seq dataset. To maximize comparability among datasets, we used 110 of the same age-, sex-, and hemisphere-matched tissue samples (representing the same three animals) profiled in our snRNA-seq dataset. To

ensure that the snRNA-seq and snATAC-seq datasets captured the same heterogeneous populations of cells, we homogenized tissue samples on dry ice before separately preparing separate nuclei isolations for each library type (Materials and Methods). Together, the snATAC-seq samples represented 28 of the 30 regions (*n* = 3 animals; MB and MT snATAC-seq data were not generated). After QC (Materials and Methods and fig. S13), the total number of nuclei profiled was 1,587,880 and ranged from 5100 [in the closed medulla (MdC)] to 114,410 [in the inferior temporal gyrus (IT)] nuclei per region (median = 63,739 nuclei per region). We called peaks on a per-sample basis and combined them across all samples based on genomic overlap, resulting in (after filtering) a combined set of 1,192,873 candidate cis-regulatory elements (cCREs) spanning 24.4% (725 Mb) of the genome.

We first applied uniform manifold approximation and projection (UMAP) dimensionality reduction and Leiden clustering to the batch-corrected epigenomic data (Fig. 3A) and identified 42 clusters which, based on promoter accessibility, could be assigned to most major cell classes found across the brain (Fig. 3B). However, given that unsupervised approaches to cell type identification are consistently more sensitive using single-cell RNA-seq or snRNA-seq data (*46*), we drew from our transcriptionally defined cell annotations to assign cell labels to our snATAC-seq nuclei. To integrate the datasets, we used the graph-linked unified embedding (GLUE) approach (*47*) and generated a unified transcriptomic and epigenomic embedding of 4,171,847 nuclei (Fig. 3, C and D). Subsequent cell type predictions based on our multimodal integration assigned the majority of snATAC-seq nuclei to a cell class (73.7% with confidence ≥ 0.95; fig. S14) and captured all of the major cell classes (Fig. 3D) with the exception of serotonergic and dopaminergic neurons, which are relatively rare and fairly specific to the MB (which as noted above was not sampled in our snATAC-seq data). The regional distribution of cell classes captured from snATAC-seq and snRNA-seq data were highly concordant, both within regions (Fig. 3E) and overall (Fig. 3F), which demonstrates that our homogenization and nuclei isolation protocols captured the same heterogeneous populations of cells in the same regions across both modalities.

The gene regulatory landscape of the rhesus macaque brain
We leveraged the snRNA-based cell class annotations (Fig. 3G) to explore heterogeneity in cell type–specific gene regulation across the brain. To do so, we partitioned all unique snATAC-seq reads by predicted cell class (Fig. 3G) and then called peaks separately for each partition using a similar peak calling approach to that used for the overall dataset, thereby generating an inventory of putative cCREs derived from each cell class in isolation (Materials and Methods). Across 11 cell classes with snATAC-seq–assigned nuclei, we identified an average of 210,572 peaks per cell class, ranging from 99,323 in microglia to 425,738 in cortical GABAergic neurons (Fig. 4A). On average, for any given cell class, these peaks covered 7.7% of the genome, and 28.8% were found >2 kb from the nearest gene or promoter (Fig. 4A).

Transcription factor regulatory networks
Multimodal integration of cell-specific snATAC-seq and snRNA-seq data allowed us to examine the cis- and trans-regulatory links between chromatin accessibility and gene expression within
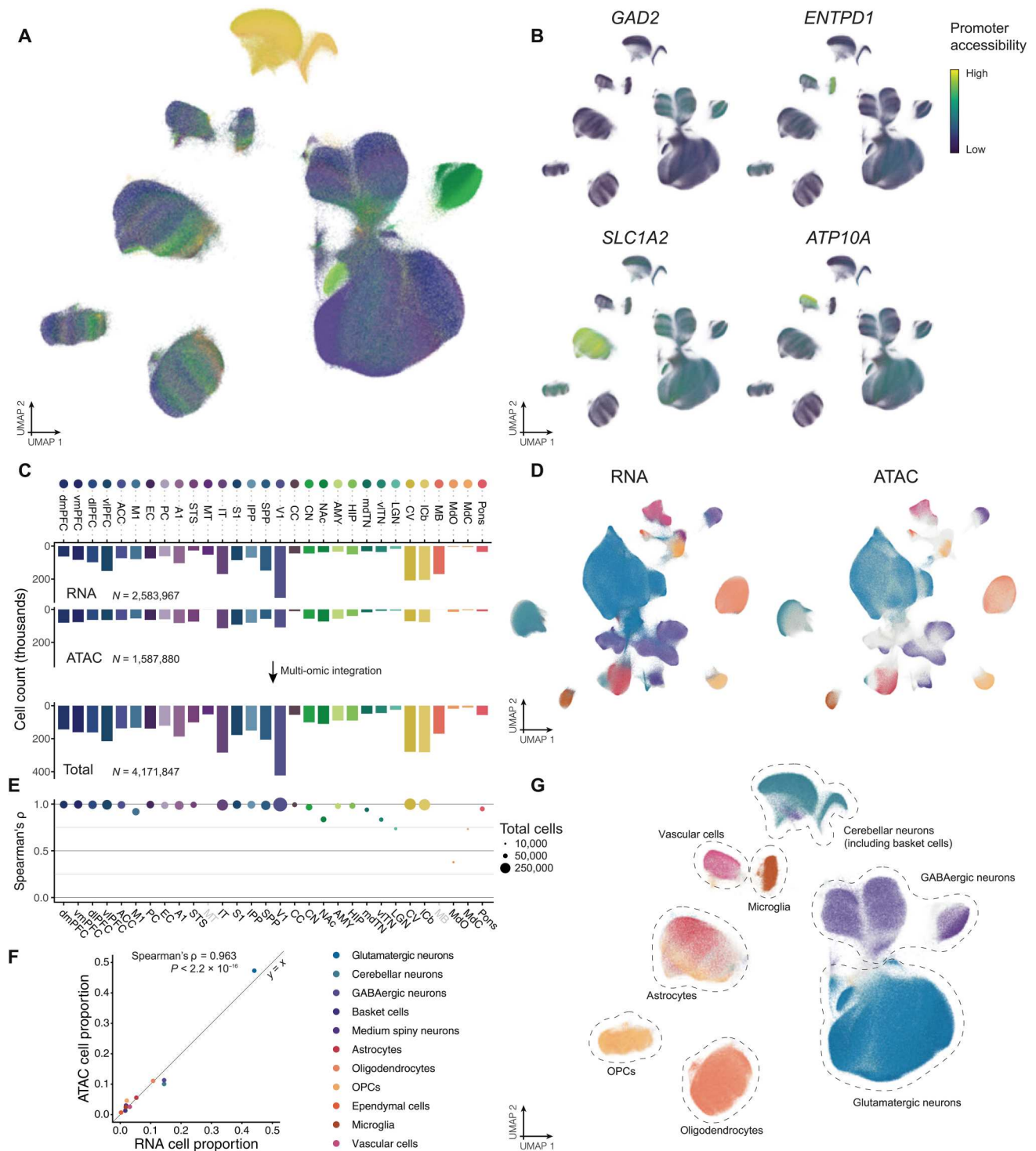
Fig. 3. Generation of the Macaque Brain Atlas sci-ATAC-seq dataset and identification of cell classes. (**A**) UMAP visualization of all snATAC-seq cells colored by brain region [with color code shown in (C)]. (**B**) UMAP visualizations of promoter accessibility scores of cell markers (*GAD2*: GABAergic neurons, *ENTPD1*: microglia, *SLC1A2*: astrocytes, *ATP10A*: vascular cells) reveal high specificity. (**C**) Barplots showing nuclei counts by brain region of the snRNA-seq, snATAC-seq, and integrated datasets. (**D**) UMAP visualizations of integrated multimodal data, with cell classes colored separately for (left) snRNA-seq and (right) snATAC-seq nuclei [with color code shown in (F)]. (**E**) Spearman's rank correlation coefficients showing the correlation between cell class proportions in the snRNA-seq and snATAC-seq datasets within each region (representing data generated from the same homogenized sample). (**F**) Scatterplot showing the correlation between cell class proportions in the overall snRNA-seq and snATAC-seq datasets (combined across brain regions). (**G**) Integration-derived cell class annotations visualized over the same snATAC-seq UMAP visualization shown in (A) [with color code shown in (F)].
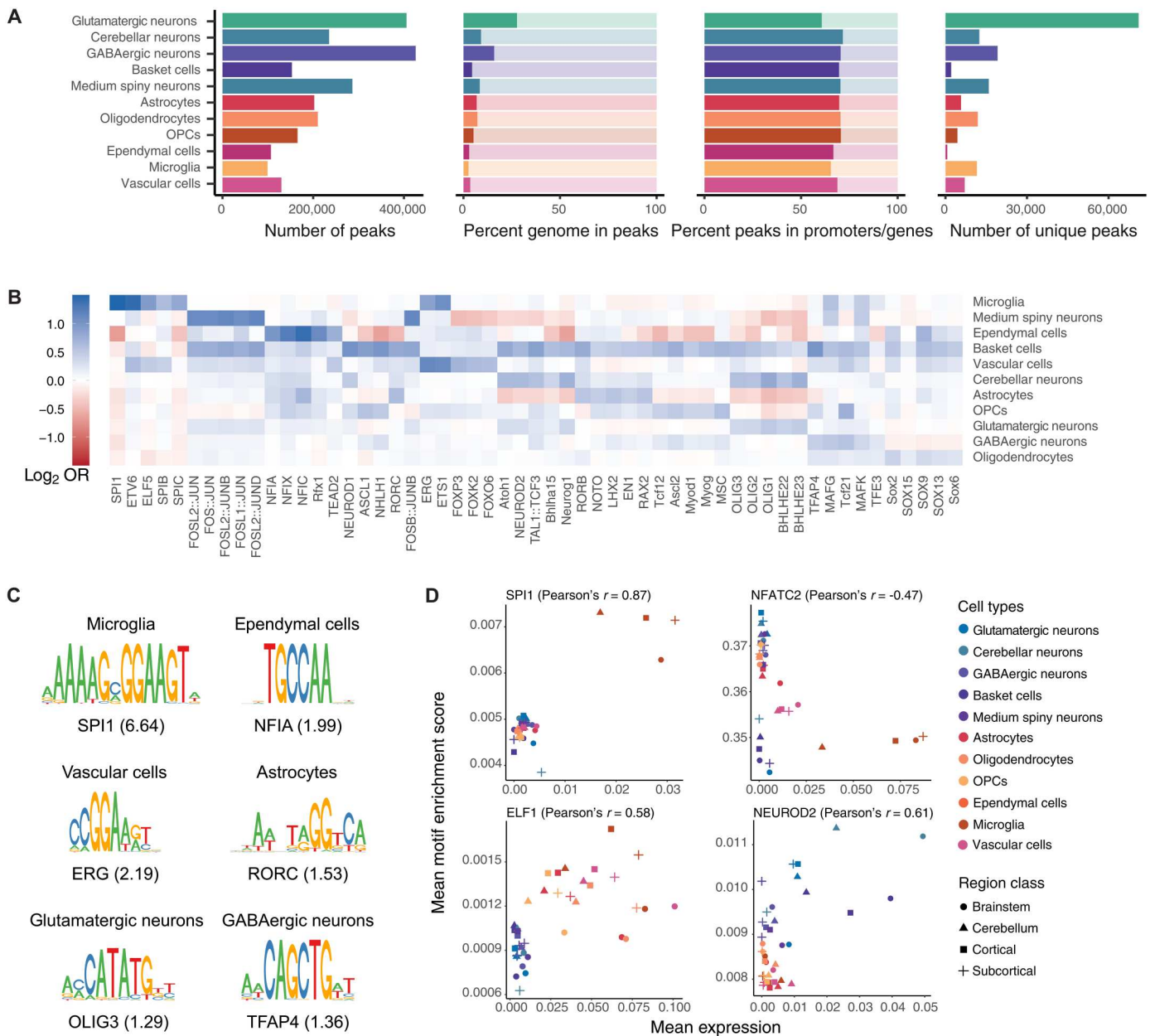
Fig. 4. Enrichment of TF binding site motifs in candidate regulatory elements. (**A**) Barplots showing summary statistics for peak sets called separately on reads derived from cells assigned to each of 11 cell classes. (**B**) Heatmap showing enrichment (log$_2$ OR) of TF binding motifs among cell classes. The top five most-enriched nonredundant TF motifs (all P_{adj} < 0.05) are shown per cell class, ordered from left to right by increasing P_{adj}. Log$_2$ OR color ranges are capped at ±1.5. (**C**) Position weight matrices of the most-enriched TF motifs for six example cell classes. ORs are shown in parentheses. (**D**) Scatterplots showing correlation between snATAC-seq accessibility of TF binding motifs and snRNA-seq gene expression of corresponding TF genes within cell classes in regional classes for four example TFs.

individual cell types. We first examined putative trans-regulatory factors within cell classes and subtypes. Transcription factors (TFs) are key trans-regulatory proteins that control cell differentiation and function during neurodevelopment (*48–51*) and have been implicated in myriad neurodegenerative diseases (*52–54*). The extremely high cell type specificity of some nuclear TFs have also made them useful targets for identifying and enriching rarer cell types before single-cell sequencing (*55–57*).

To identify candidate trans-acting regulatory networks in each cell class, we carried out TF binding motif enrichment analysis on each set of cell class–specific peaks, defined as the subset of a cell class's cCREs that did not overlap with any peaks called in other cell classes (Materials and Methods and Fig. 4A). Cell-class–specific cCREs were highly enriched for many TF binding motifs that are likely involved in cell-specific gene regulation (Fig. 4B and table S11), including many motifs previously implicated (Fig. 4C). For instance, microglial cCREs contained 6.6-fold more binding sites

of the nuclear TF SPI1 (also known as PU.1) than expected by chance ($P_{adj} = 1.22 \times 10^{-284}$; Fig. 4, B and C). In addition to such canonical examples, we identified numerous motifs that distinguish relatively similar cell classes. For instance, the TF binding motif for NFE2, from the nuclear respiratory factor (NRF) TF family, was most enriched [odds ratio (OR) > 2] in cCREs in both medium spiny neurons (OR = 3.07, $P_{adj} = 1.77 \times 10^{-87}$) and basket cells (OR = 2.34, $P_{adj} = 8.76 \times 10^{-7}$), while the binding motif for NEUROD1 was most enriched (OR > 2) in cCREs of basket cells (OR = 2.04, $P_{adj} = 3.53 \times 10^{-29}$), where this TF is necessary for basket cell terminal differentiation and, consequently, axon growth and inhibitory circuit formation (58).

We also characterized TF binding motif enrichment at the cell subtype level. To do this, we extended our multimodal integration and label-transferring approach to each cell class independently by tabulating the reads per-cell falling within cell class–specific cCREs described above for all cells of a given cell class (Fig. 3D). We then integrated the data with corresponding snRNA-seq data of the same cell class using GLUE (Materials and Methods). The resulting integrated embeddings for each cell class were then used as the basis for predicting cell subtypes, which we carried out on all snATAC-seq cells within each class (fig. S15).

Since cell subtypes are preselected to already share broadly similar chromatin accessibility profiles, identifying peaks that are specific to a single subtype—similar to our approach at the cell class level—was not feasible and left most cell subtypes with no or very few unique peaks to analyze. As an alternative strategy, we carried out differential accessibility analyses among cell subtypes to identify peaks that were predictive of each individual cell subtype within a given cell class (Materials and Methods). We then identified TF binding motifs enriched in highly differentially accessible regions within cell subtypes (table S12). For example, we observed numerous TF binding motifs ($N = 433$, $P_{adj} < 0.05$) that were enriched within highly accessible peaks in Purkinje cells, a GABAergic neuron type of the cerebellum that is implicated in autism spectrum disorders (ASDs). In our snRNA-seq dataset, of all tested diseases (59), genes associated with autism (DOID:12849) were overrepresented (Fisher's exact test, OR = 10.2, $P_{adj} = 8.52 \times 10^{-16}$) among the top 100 Purkinje cell marker genes, including *RORA* [fold change (FC) = 331.9], *AUTS2* (FC = 43.1), and *SHANK2* (FC = 13.8) (table S13). Correspondingly, we found that TF motifs enriched in differentially accessible peaks included RORA, four members of the early growth response (EGR) family (EGR1 to EGR4), and CTCF (EGR1, EGR3, and CTCF were among the top five TF motifs ranked by OR; RORA ranked 182nd). RORA is a regulator of circadian rhythm that exhibits decreased expression in ASD brains and may play a role in ASD pathogenesis (60, 61). EGR family TFs have been implicated in the disruption of human-specific developmental programs in autism (62). CTCF is an insulator protein that regulates chromatin structure and may play a critical role in maintaining dendrite structure in Purkinje cells (63) and *CTCF* is also a risk gene for ASD (64).

Given that families of TFs have similar binding motifs, it is often difficult to identify the specific TF in a given family that is responsible for enrichment in cell type–specific cCREs. To identify the most likely TF, we therefore used our recently developed approach (45, 65) that uses the computationally paired snRNA-seq and snATAC-seq data. Briefly, this approach relies on the assumption that TFs will be highly expressed in cell types where they play a key role, while their associated motif should be enriched (or depleted) in that cell's cCREs, indicating TF activation (or repression). Overall, we compared the accessibility of 369 TF binding motifs and their corresponding gene's expression across the cell classes in four region subclasses, with 189 TFs showing positive Pearson's correlation between gene expression and accessibility of the cognate motif and 180 showing negative correlation (fig. S16A and table S14). Among the TFs with largest positive or negative Pearson's correlation values were strong cell class–specific activators and repressors (Fig. 4D and fig. S16B). For instance, *SPI1*, which has been identified as a candidate gene for Alzheimer's disease via various functional genetics approaches (66), shows a strong activating effect with high expression of the *SPI1* gene and high accessibility for the SPI1 binding motif in microglia. In contrast, NFATC2 has a repressing effect in microglia and vascular cells, as shown by high expression of the *NFATC2* gene associated with lower NFATC2 motif binding in those cell types. We also found evidence for a clear distinction between neurons and non-neuronal cells at two TFs, with ELF1 functioning as a non-neuronal specific activator and NEUROD2 as a neuron-specific activator. In addition, we note that FLI1, an activator in vascular and microglia cell types, and ELF1 have motif sequences similar to SPI1 (fig. S16C), but their activating effects affect a broader set of cell types.

The cis-regulatory landscape of brain cell variation

We next sought to characterize cis-regulatory interactions between cCREs and proximate genes in the rhesus macaque brain. We used two complementary analyses to scan for interactions using our integrated multimodal dataset. First, we used the regulatory inference framework of GLUE (47), which leverages the unified feature embedding (i.e., joint integration of snRNA-seq genes and snATAC-seq peaks in a common data space) generated during GLUE integration to assess similarity between peaks and genes. Putative regulatory interactions are defined as a high cosine similarity between peak and gene feature embeddings in the unified data space, with statistical significance assessed by permutation (47). Second, we used a metacell-based approach to aggregate snRNA-seq transcriptomes and snATAC-seq epigenomes into multimodal metacells based on k-means clustering of the unified cell embeddings and then used logistic regression to model the relationship between gene expression and chromatin accessibility within a given metacell (67). In contrast to the GLUE regulatory score, the logistic regression analysis enabled us to differentiate between positive and negative regulatory interactions between peaks and genes. We considered peak-gene pairs to be putatively regulatory if $P_{adj} < 0.05$ for both analyses (Fig. 5A and fig. S17). For each cell class, we also scanned for differentially accessible peaks using both a regularized logistic regression and a t test, testing accessibility in a given cell class against accessibility in all other cell classes. We consider cCREs with differentially high accessibility (regularized LR coefficient > 0, $\log_2$ FC > 0, and t test $P_{adj} < 0.05$; fig. S18) to be candidate regulators of cell type–specific genes (Fig. 5A).

We focused our analysis on the 6000 most variable genes in our snRNA-seq dataset and tested all snATAC-seq peaks that fell within 150 kb of the gene promoter (defined as TSS extended 2 kb upstream). In total, we tested 223,752 peak-gene pairs (151,083 unique peaks, 5765 unique genes), of which 142,324 peak-gene pairs (63.6%) met our criteria for being considered candidate cis-regulatory interactions (table S15). A total of 128,741 peaks
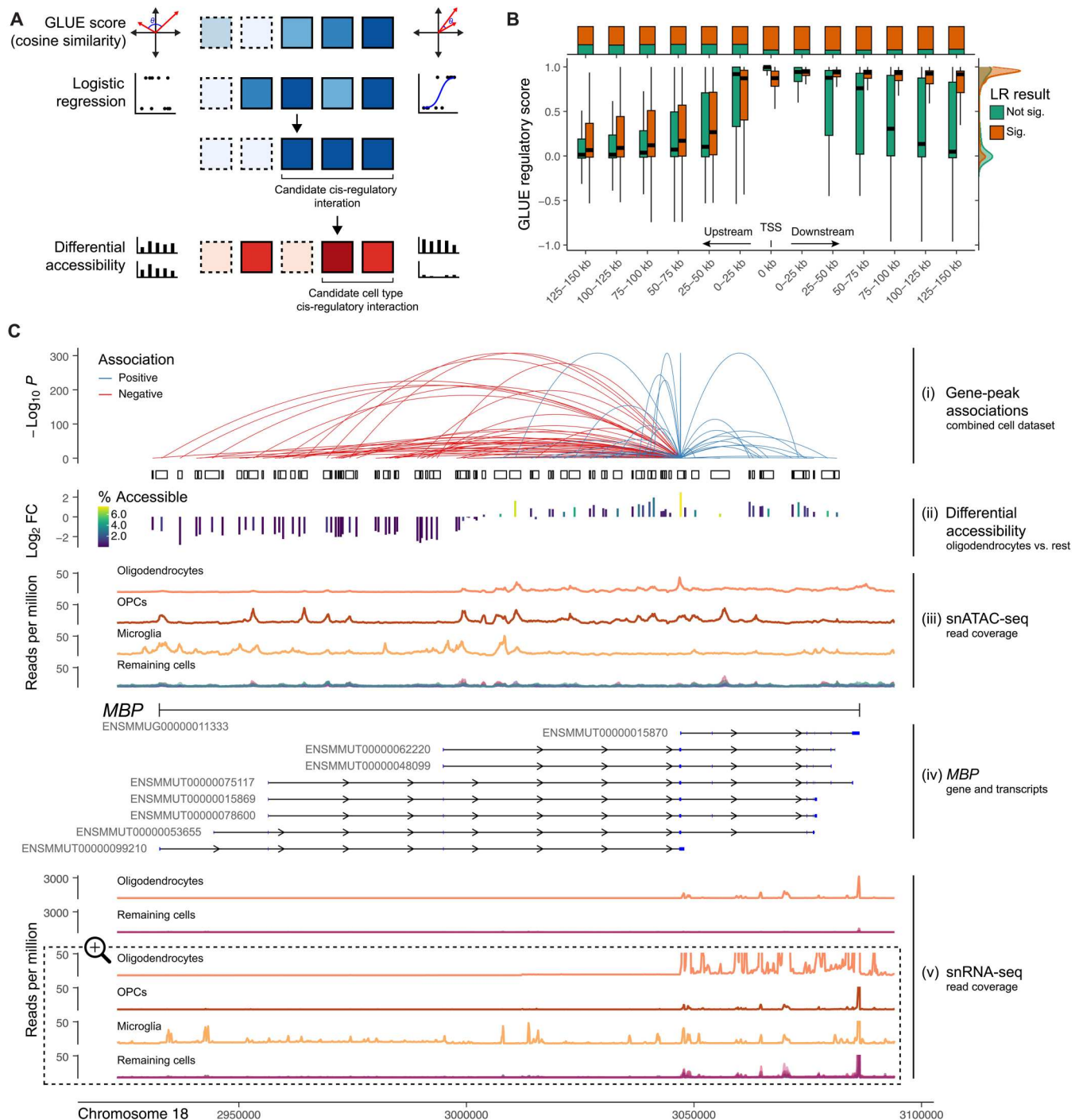
Fig. 5. The landscape of cis-regulatory interactions in the Macaque Brain Atlas. (**A**) Schematic outlining criteria for identification of cCREs. Squares represent peak: gene pairs, darker colors symbolize stronger evidence for a given measure, and solid borders represent statistically significant measures. (**B**) Distribution of gene-peak GLUE regulatory scores binned according to the minimum signed distance (left: upstream, right: downstream) between peaks and gene TSSs. Distributions are shown separately according to whether the gene:peak pair also exhibited a significant association ($P_{adj} < 0.05$) based on the metacell-based logistic regression analysis. The ratio between significant and not-significant gene-peak pairs for proximity bins according to the logistic regression model is shown in the top margin, while the distribution of GLUE regulatory scores is shown in the right margin. (**C**) Candidate regulatory elements are shown in relation to, from top to bottom, (i) the strength of and direction of inferred regulatory links connecting peaks to *MBP* expression (based on metacell logistic regression analysis). The height of links represents the strength ($-\log_{10} P$ values) of evidence for regulatory connections and the color symbolizes the direction of the relationship; (ii) the differential accessibility ($-\log_2$ FC) of peaks in oligodendrocytes relative to all other cell classes; (iii) the distribution of normalized snATAC-seq reads by cell class; (iv) gene and transcript boundaries of *MBP* and its known isoforms in the rhesus macaque genome, with exons shown in blue; (v) the distribution of normalized snRNA-seq reads by cell class. Oligodendrocyte reads are shown in relation to all other cell classes on the upper portion of the plot. On the bottom portion, the *y* axis is magnified ×60 and cropped to highlight more subtle differences among cell classes.

(85.2%) that we evaluated were cCREs for at least one gene, and 4811 genes (83.5%) that we evaluated had at least one cCRE.

Of all peak-gene pairs, 132,805 (93.3%) involved a peak that was highly differentially accessible in at least one cell class, thereby fulfilling our criteria for being considered candidate cis molecular interactions regulating cell type–specific markers. cCREs were highly differentially accessible in a maximum of seven cell classes, with 37% exclusive to a single cell class and 88% highly differentially accessible in one to three cell classes.

The vast majority (133,496 or 93.8%) of candidate regulatory interactions were positively associated (i.e., had a positive effect size in the metacell logistic regression)—this held true whether peaks were upstream (13,650 of 14,575 or 93.7%), downstream (116,939 of 124,592 or 93.9%), or overlapping (2907 of 3157 or 92.1%) the gene's transcription start site (TSS). For peak-gene pairs where the peak was upstream of the TSS, the GLUE regulatory scores were highest (indicating high similarity between peak and gene feature embeddings) when peaks were in closer proximity to the TSS (Fig. 5B). For peaks downstream of the TSS, GLUE regulatory scores remained high across all distances, with only a modest decrease farther from the TSS (Fig. 5B). This result was particularly notable for peaks that had significant, mainly positive, associations between accessibility and gene expression, likely reflecting (i) higher global accessibility across the gene body resulting from higher expression of the gene (as opposed to distal regulation) and/or (ii) methodological limitations of using a single gene-wide TSS (i.e., the most upstream TSS of all isoforms), thereby ignoring variation in TSS positioning among isoforms, which likely vary in their usage across tissues and contexts (68).

Using the cell class–specific gene expression and cCRE peak sets, we repeated our integration, regulatory inference, and differential accessibility workflows on each cell class individually. We tested a mean of 72,914 peak-gene pairs (range: 45,539 to 114,200) per cell class and identified a mean of 11,442 peak-gene pairs (range: 881 to 41,966) showing evidence of regulatory interactions (fig. S19 and table S16).

To illustrate how these maps of putative interactions might be useful to investigate the regulatory landscape at the level of an individual locus, we focused on the myelin basic protein (MBP) gene (Fig. 5C), which encodes one of the most abundant proteins in central nervous system myelin (69, 70), has a range of splice isoforms (71), and is a canonical marker of oligodendrocytes. MBP is located on chromosome 18 (positions 2,932,531 to 3,086,873) on the rhesus macaque (Mmul_10) genome and has eight annotated mRNA isoforms (Ensembl). In humans, classic MBP isoform 3 (18.5 kDa) predominates in adult myelin (71).

In our global peak set (all cells), 94 peaks fell within 150 kb of the MBP promoter and were included in our analysis. Of these peaks, 83 (88.3%) were identified as candidate regulators of MBP (crMBP), with 38 crMBPs (45.8%) positively associated with MBP expression. Of all crMBPs, only one was not located within the MBP gene boundaries—it was, however, located less than 2 kb upstream within the likely promoter region.

In accordance with the well-known status of MBP as an oligodendrocyte marker, we found that MBP was differentially expressed in oligodendrocytes, with detected expression in 80.9% of cells and 1434-fold higher expression than all other cells averaged together. Fine-grained inspection of normalized read distributions from oligodendrocyte nuclei revealed the highest densities of snRNA-seq reads corresponding to the polyadenylation site (position 3,086,373) and snATAC-seq reads corresponding to the TSS (position 3,046,976) of a single transcript, ENSMMUT00000015870, indicating that it is likely the dominant MBP isoform expressed in adult macaque oligodendrocytes.

By examining the genomic-distance relationships between crMBPs and the dominant MBP transcript in adult oligodendrocytes, we found that all 16 crMBPs that either overlapped or were downstream of the isoform's TSS were positively associated with MBP expression. Among the 67 crMBPs that were located upstream of the TSS, 22 (32.8%) were positively associated with MBP expression, while 45 (67.2%) were negatively associated. Several of these negatively associated crMBPs corresponded with sci-ATAC-seq3 peaks in other cell types, particularly OPCs and microglia (Fig. 5C). However, the accessibility landscape of OPCs is overall more similar to that of oligodendrocytes across the region upstream of the TSS of the dominant isoform, with greater accessibility at most peaks except for that of the promoter of the dominant isoform (Fig. 5C). As OPCs play a critical role in myelinogenesis by giving rise to oligodendrocytes (72), these crMBPs likely serve as critical markers of the OPC-oligodendrocyte transition, during which the expression of this gene, and this isoform in particular, is massively up-regulated.

Evolutionary conservatism and divergence of candidate regulatory elements

Human brain specialization could be driven in part by changes in cell type composition and function following the evolutionary divergence of cell type–specific regulatory elements. To evaluate evolutionary divergence/conservatism, we tested whether cCREs that were differentially accessible across cell classes (table S15) or unique to cell classes (table S17) were associated with regions that underwent rapid evolution in the human lineage [i.e., human accelerated regions (HARs) (73); human ancestor quickly evolved regions (HAQERs) (74)] or are differentially accessible between humans and chimpanzees in cerebral organoids (DA$_{HC}$ regions) (75). We also tested whether cCREs were enriched for these evolutionarily salient regions (relative to all peaks, called on the global set) and did not detect any enrichments (table S17). We detected an enrichment of DA$_{HC}$ regions among differentially accessible cCREs (OR = Inf; P_{adj} = 0.046) and a depletion of DA$_{HC}$ regions among glutaminergic neuron-specific cCREs (OR = 0.263; P_{adj} = 0.046) (table S17). These findings indicate that regulatory areas exhibiting differential accessibility across cell types in the adult macaque brain may also exhibit differential accessibility in the developing human versus chimpanzee brain, while cell class–specific regions may be more conserved during primate brain development. Neither HARs nor HAQERs were enriched among differentially accessible or cell class–specific cCREs (P_{adj} > 0.05), which was expected given that intercell class variation in orthologous gene expression is generally well-conserved in primates (76). We did identify six differentially accessible or cell class–specific cCREs that overlapped with HARs or HAQERs (table S17). There was one cCRE that was differentially accessible in macaque medium spiny neurons, cerebellar neurons, and vascular cells that overlapped with two HAQERs. These cCREs are also: (i) located in the 1q21.1-2 region containing the NBPF gene cluster that contains several human-specific segmental duplications; (ii) areas of open chromatin in the developing human brain; and/or (iii) regions showing differential chromatin

accessibility between human and chimpanzee cerebral organoids (*74*). While the presence of HAQERs in these regions suggests that they may function differently in humans versus macaques, the human sequences for these HAQERs do not appear to exhibit significantly greater enhancer activity compared to nonhuman primate sequences (*74*).

Enrichment of disease heritability among candidate regulatory elements

Last, we used our cCREs to identify cell type–associated regulatory networks that may drive polygenic disease risk. We tested for enrichment of disease trait heritability using the linkage disequilibrium score regression (LDSC) tool (*77*, *78*), after lifting over macaque cCREs to human genome coordinates (*28*). We tested a total of 53 phenotypes relevant to neurological diseases, disorders, syndromes, behaviors, or other traits (table S18) and examined enrichment among cell class cCREs called separately in each of 11 cell classes.

Our results broadly recapitulated several known roles of cell classes in neurological disease (Fig. 6 and table S19). For example, sites associated with cardioembolic stroke (OR = 32.2) or ischemic stroke (OR = 9.2) were enriched ($P_{adj} < 0.05$) only in vascular cells, which play a crucial role in forming and maintaining the blood-brain barrier (*79*). We also found that Alzheimer's disease–associated sites were enriched only in microglia—a result replicated using loci from three independent genome-wide association studies (GWAS) (OR range: 13.9–15.0)—consistent with the prominent role of microglia proliferation and activation in Alzheimer's disease (*80*).

Across all cell classes, basket cells were enriched for the greatest number ($N = 37$) of GWAS phenotypes, including disorders such as schizophrenia (OR range: 5.9 to 6.2), bipolar disorder (OR range: 5.6 to 6.2), and major depressive disorder (OR range: 5.1 to 5.3) and, most strongly, epilepsy (OR = 9.0)—a disease that basket cells have been connected to in animal models and some genetically linked human forms of the disease (*81*).

Other notable results included the enrichment of multiple sclerosis-associated sites among open regions in microglia (OR = 46.6), highlighting the outsized role of these immune cells in the etiology of the disease and as a putative therapeutic target (*82*, *83*). In multiple sclerosis, disease-associated microglia alter their transcriptional profiles and may contribute to neuroinflammatory processes underpinning this autoimmune disorder (*83*). We also found enrichment of Parkinson's disease–associated sites among open regions in the glial OPC, oligodendrocyte, and astrocyte cell classes (OR range: 7.0 to 8.4). In Parkinson's disease, glial cells may play a major role in the progressive degeneration of dopaminergic neurons (*84*), a classic hallmark of Parkinson's disease, or in alterations to glutamatergic neurotransmission (*85*).

Last, we found that heritable sites associated with attention deficit/hyperactivity disorder (ADHD) in our analysis were enriched only among open regions of medium spiny neurons. While the magnitude of the enrichment was relatively mild (OR = 2.6, $P_{adj} = 0.031$), genetic variants associated with ADHD have been historically difficult to identify, with the first risk loci only recently reported (*86*). Medium spiny neurons have been linked to behavioral hyperactivity and disrupted attention via activation of astrocyte-mediated synaptogenesis (*87*). Our results therefore suggest that medium spiny neurons may be a promising target for prospective ADHD therapeutics warranting further study.

DISCUSSION

Understanding the cellular architecture of the primate brain is crucial both for understanding the evolution of human cognition and behavior and for identifying mechanisms underlying neurological disorders. In service of these goals, we used snRNA-seq and snATAC-seq to derive a molecular atlas spanning the adult rhesus

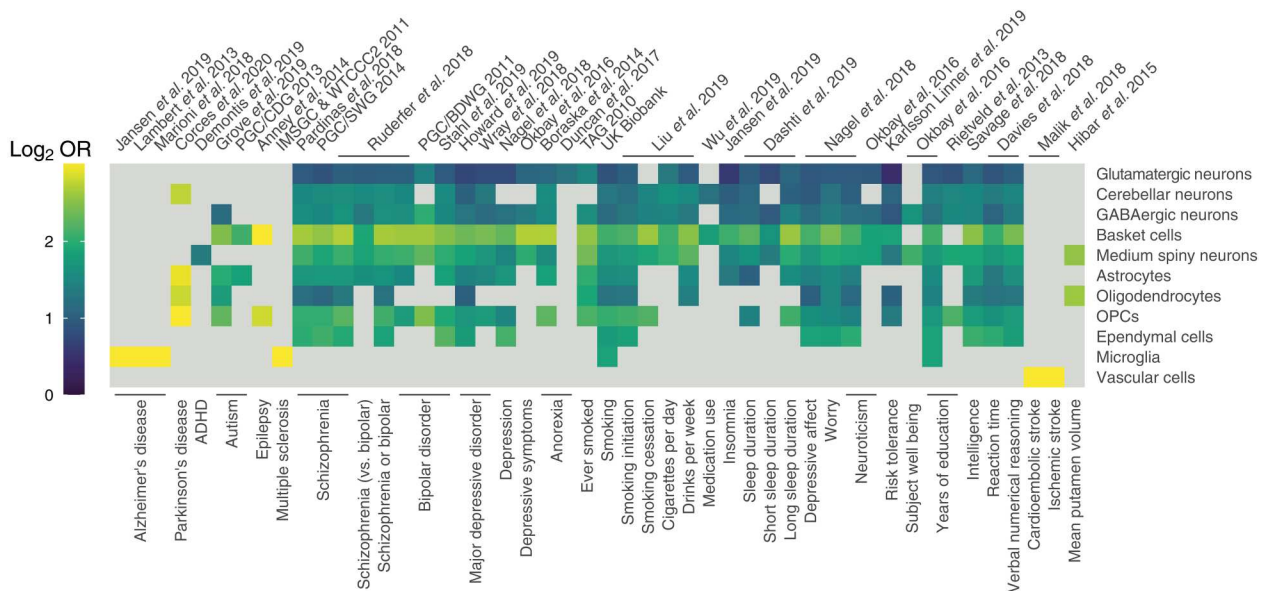**Fig. 6. Enrichment of heritable disease-relevant sites among candidate regulatory elements.** The heatmap displays heritability enrichment (log$_2$ OR) of diseases among cell class snATAC-seq peaks for tested diseases, syndromes, and phenotypes. Only results passing a threshold of $P_{adj} < 0.05$ are shown. The log$_2$ OR color range is capped at 3.0.

macaque brain, comprising data from over 4 million cells profiled from 30 brain regions. On the basis of our multimodal molecular data, we identified 112 distinct molecular cell types or subtypes and characterized their distribution across the macaque brain, adding to the growing number of primate single-cell molecular brain atlases (*15*, *32*). The data are freely available (NeMO archive, nemo:dat-rtmm5q2) and browsable (CELLxGENE viewer, https://cellxgene.cziscience.com/collections/8c4bcf0d-b4df-45c7-888c-74fb0013e9e7) and will serve as a rich resource for the neuroscience and neurogenomics communities.

In generating a multiregion transcriptomic and epigenomic atlas of the most widely used nonhuman primate in neuroscience, we (i) identified all of the major brain cell classes and many cell types that have been previously reported (Figs. 1 and 2); (ii) quantified regional distribution of cell types and subtypes within individuals, which allowed us to identify compositional differences in samples collected at the same time and from the same animals (Fig. 2); (iii) identified rare and regionally specific cell types (e.g., Purkinje cells), which may facilitate the development of molecular tools such as cell type–specific viral vectors that, in combination with new technologies such as CellREADR (Cell access through RNA sensing by Endogenous ADAR) (*88*), may enable precise targeting of cell types based on their unique patterns of chromatin accessibility and gene expression; (iv) characterized multiple trans- and cis-regulatory mechanisms that differentiate cell classes and subtypes (Figs. 4, A to D, and 5); and (v) identified numerous associations between genetic risk for neurological disorders and the epigenomic states of specific cell types (Fig. 6).

Notably, this single-cell atlas of the adult primate is generated from samples collected from healthy adults. The paired nature of the dataset, with regions sampled from the same individual brains, avoids many of the inter-individual variables (e.g., genotype and environment) that can affect neurological development and function. The atlas may thus be a valuable resource for characterizing molecular features that play a role in myriad neurological disorders. The relatively few unique individuals sampled also represent a limitation of the current study—we currently know very little about how brains of healthy individuals differ in cell composition and function and what that confers for disease susceptibility and/or progression. Given continuing improvements in cost and throughput of single-cell sequencing, characterizing multiregion cellular variation across many healthy individuals is becoming not only a possibility but also an emerging priority for the field.

To our knowledge, these data represent the largest and most comprehensive multimodal molecular atlas in any nonhuman primate to date and provide a resource for exploring how the heterogeneous molecular and cellular composition of the brain gives rise to the behavioral complexity of primates including humans. We anticipate that these data will also provide a critical and much-needed molecular and neurobiological map of complex human-relevant social behavior and disease, as well as an extensive substrate for comparative analyses across animal brains.

MATERIALS AND METHODS
Study population and sample collection
All animals sampled in this study are rhesus macaques (*M. mulatta*) from the semi-free-ranging colony on the island of Cayo Santiago, Puerto Rico. Maintained by the Caribbean Primate Research Center

(CPRC) within the University of Puerto Rico, the Cayo Santiago macaque colony has been largely continuously studied since its founding in 1938 (*89*). All present-day macaques are descended from an initial founder population of 409 animals and have since maintained an outbred population structure despite generations of isolation (*90*). Apart from being provisioned with commercial feed and occasionally subject to capture-and-release sampling, the macaques otherwise live in naturalistic conditions, subject to minimal intervention and manipulation, as approved by Institutional Animal Care and Use Committee. The study used animals that needed to be removed from Cayo Santiago (*91*) and were immediately euthanized. Standardized tissue collection and sample archiving were coordinated by the Cayo Biobank Research Unit (CBRU), which provided the brain samples used in this study (*92*, *93*).

Procedures for necropsy, brain removal, and dissection followed those previously described for this population (*93*) and are briefly outlined here. Following veterinary euthanasia, brains were perfused with sterile saline, removed from the cranium, and hemisected into left and right hemispheres using a long single-edge razor blade. After sectioning off the cerebellum/brainstem from each hemisphere, the cerebral hemispheres were placed on custom molds (designed either for left or right hemispheres) and coronally sectioned into 11 roughly 5-mm-thick blocks, numbered in order rostral to caudal. All 12 blocks (with the cerebellum/brainstem considered block 12) were then sealed in Whirl-pak bags, flash-frozen in liquid nitrogen vapor, and archived in ultralow −80°C freezers. The interval between euthanasia and permanent storage of frozen tissue averaged 51 min, with an SD of 5.8.

All procedures were performed in accordance with the National Institutes of Health (NIH) Guide for the Care and Use of Laboratory Animals and were approved by the Institutional Animal Care and Use Committee at the University of Puerto Rico (protocol #338300). Five macaques were included in this study (table S2). The vast majority of the data were derived from four 10-year-old macaques, which are considered middle-aged adults in this population (*93*, *94*).

Region selection and biopsy
Frozen brain blocks were placed on a dissection tray over dry ice to keep tissue frozen during biopsy collection. Individual blocks were then moved from the dry ice to a tray sitting on wet ice, allowing for tissues to be acutely warmed to the point that biopsies could be taken from targeted structures. Biopsies were made using a cutting spoon (Fine Science Tools Inc., catalog no. 10360-13). Dissected brain regions are listed in table S1, and approximate locations for biopsy are illustrated in Fig. 1A. For a given structure, attempts were made to minimize inclusion of off-target surrounding tissues (e.g., white matter underlying a targeted gray matter structure). Below, we document the most common block numbers where structures were located. Because of interindividual differences and/or variation in sectioning, regions of interest were sometimes identified and dissected from adjacent blocks based on neuroanatomical landmarks. Alternate block numbers are therefore also documented below.

The most anterior block sampled for this study (block 2) contained gray matter for the dorsomedial (dmPFC), ventromedial (vmPFC), dorsolateral (dlPFC), and ventrolateral prefrontal cortices (vlPFC). dmPFC and vmPFC were defined as being on the

medial side of block 2. The dmPFC biopsy was pulled from the gray matter in the ~top half of the medial edge of the block. A space along the medial edge was left to separate dmPFC from vmPFC. The vmPFC biopsy was pulled from the medial ventral half of the tissue block. Biopsies of dlPFC came from the cortical tissues surrounding the dorsal lateral portion of the block that included the superior and inferior portions of the principal sulcus. Samples from vlPFC came from the ventral and lateral portion of the block. As was the case on the medial side, a portion of the cortex was left between each lateral biopsy to avoid overlap (Fig. 1A).

Block 3 (sometimes 4) contained biopsies for the anterior cingulate cortex (ACC), CC, and head of the CN. The biopsy for ACC was the gray matter sitting between the CC, which is ventral to ACC and the cingulate sulcus (cs), which sits dorsal to the cingulate gyrus. CC was defined as the white matter track sitting ventral to the ACC and medial to the lateral ventricle. The CN was the gray matter sitting ventrolateral to the lateral ventricle and surrounded on all other sides by white matter. The CN was the only biopsy in the second block that was scooped out of the block face to minimize inclusion of any white matter sitting anteriorly past the CN within the block (Fig. 1A).

Block 5 (sometimes 4 or 6) contained the amygdala (AMY), entorhinal cortex (EC), perirhinal cortex (PC), and NAc. The NAc is located ventral to the caudate, internal capsule, and putamen (Pu). Furthermore, in fresh-frozen tissue, there was a slightly darker color to the NAc. The tissue making up the NAc was scooped out of the block face. Similarly, the AMY was identified as ventral to the Pu, medial to the ventral portion of the claustrum, and dorsal to the EC. The AMY was also scooped out to minimize the inadvertent collection of neurons within the hippocampus (HIP). Last, the EC and PC were collected, the delineation between the two was the rhinal fissure (Fig. 1A).

Blocks 5 and 6 (sometimes 4 or 7) contained tissue that were biopsied to represent cortical regions primary motor cortex (M1), primary somatosensory cortex (S1), primary auditory cortex (A1), superior temporal cortex (STS), and IT. Subcortical structures that were biopsied included mdTN, vlTN, LGN, and HIP. The delineation between M1 and S1 was the central sulcus and was taken from the approximate central third of the lateral portion of each respective gyrus. Within a case, attempts were made to biopsy from approximately the same putative mototopic and somatotopic regions. A1 biopsies were taken from the dorsal portion of the superior temporal gyrus, which is within the lateral sulcus (ls) (i.e., inferior operculum). The gray matter forming the STS sits ventral to the superior temporal gyrus and dorsal to IT. IT was defined as the gray matter forming the lateral portion of the inferior temporal cortex. mdTN sits bilaterally on midline, within the thalamus. It is bound by ventricles dorsally, laterally by the centrolateral thalamic nucleus and ventrally by the centromedial thalamic nucleus. vlTN is bound by the centrolateral thalamic nucleus medially, body of the CN dorsally, and the reticular thalamic nucleus laterally. Biopsies for mdTN were taken from the central and central medial portions of the nucleus, while vlTN biopsies were taken from the central portion of the nucleus. In both cases, this was in an effort to avoid inclusion of other thalamic nuclei. The LGN is a six-layered structure that is easily observed on the coronal face of fresh-frozen slabs. When observed, the biopsy was scooped out. Like the LGN, the HIP was defined by its classic cytoarchitectonic features within

the medial temporal lobe. For biopsies, efforts were made to not include EC, which sits ventral and ventromedial to HIP (Fig. 1A).

Block 7 (sometimes 6 or 8) contained tissues representing the superior posterior parietal (SPP), inferior posterior parietal (IPP), and the middle temporal visual area (MT). SPP biopsies were from the gray matter of the superior lobule. The intraparietal sulcus sits between SPP and IPP. Therefore, IPP biopsies were taken from the gray matter of the second, more lateral lobule. Last, area MT was defined by the gray matter of the insular cortex, bound on its medial edge by white matter of the extreme external capsule and laterally by the superior and inferior operculum divided by the STS (Fig. 1A) (95, 96).

The final cerebral block, block 11, contained the visual cortex. Biopsies from V1 were taken from the dorsolateral surface gray matter above the external calcarine sulcus (Fig. 1A).

The hemisected cerebellum/brainstem block was dissected as follows. First, the cerebellum was dissected off and the cerebellar vermis (CV) was separated from the lateral cerebellar cortex (lCb). Next, the remaining brainstem was dissected such that the MB block was separated by making a cut from just behind the inferior colliculus to the top of the basilar pons. Next, the pons was separated from the medulla by making a cut from the stria medullaris (approximate center of the fourth ventricle) to the base of the pons. A final cut at the base of the fourth ventricle to separate the open medulla (MdO) from the MdC (Fig. 1A).

To allow for the profiling of multiple genomic modalities from the same representative cell populations, we pulverized all biopsies on dry ice to homogenize and divide tissue for downstream experiments. We followed the tissue pulverization procedures described by Domcke et al. (97) to achieve a powder consistency on a sterile aluminum foil work surface. Once sufficiently pulverized, we stirred the sample thoroughly and then divided the sample using the folded edge of foil as a funnel into new 1.5-ml prechilled and prelabeled microcentrifuge tubes. Foil and tubes were set on aluminum trays or tube racks set on dry ice to keep powdered tissue frozen throughout this process. We divided samples into roughly a 2:1 ratio given the expected efficiencies/yields for snRNA-seq and snATAC-seq protocols, respectively. Pulverized tissue was stored at −80°C up until processing for downstream library preparation procedures.

snRNA-seq data generation

To profile single-nucleus gene expression, we performed snRNA-seq using the sci-RNA-seq3 approach (19), which is the improved version of the original sci-RNA-seq protocol (18). For two of the three experimental batches in our dataset, we used a protocol closely adhering to the sci-RNA-seq3 protocol described by Cao et al. (19). For the third batch, we used the improved protocol ("tiny sci") described by Martin et al. (21). Sample order was randomized between the first two batches, and within the third batch, to minimize batch effects and other technical artifacts.

For the first two batches, we slightly modified the protocol described by Cao et al. (19, 20) for a different tissue type and smaller input amounts. Briefly, we added 50 μl of cell lysis buffer to pulverized tissue in a 1.5-ml microcentrifuge tube and then homogenized the tissue using 5 to 10 strokes with a disposable ribonuclease (RNase)–free plastic pestle (Fisherbrand, catalog no. 12-141-364). We then added another 950 μl of cell lysis buffer, mixed by pipette, and then transferred the suspension through a 70-μm cell strainer (pluriSelect, catalog no. 43-10070-70) into a 15-ml conical

tube containing 5 ml of ice-cold 4% paraformaldehyde. Nuclei were fixed in 4% paraformaldehyde for 15 min with occasional mixing, washed once in 1-ml ice-cold nuclei wash buffer, and then suspended in 200-μl nuclei wash buffer. Nuclei were counted by mixing with 1 μM YOYO-1 iodide (Thermo Fisher Scientific, catalog no. Y3601) using a Countess II FL automated cell counter (Life Technologies), divided into tubes in 100-μl aliquots, and then flash-frozen in liquid nitrogen.

For nuclei fixed with paraformaldehyde, library construction was similar to the sci-RNA-seq3 method from Cao *et al.* (*19*) with minor modifications including the substitution of Quick Ligase (New England Biolabs) for 10 min at 25°C for the second index step, instead of T4 DNA ligase (NEB) for 180 min at 16°C. For tagmentation, we used N7 adaptor-loaded Tn5 from QB3 MacroLab at the University of California Berkeley in tagmentation buffer (2× TD) as previously described in Corces *et al.* (*98*): 20 mM tris-HCl (pH 7.5), 10 mM MgCl$_2$, and 20% (vol/vol) dimethylformamide. Libraries were sequenced on a NextSeq or NovaSeq platform (Illumina) (read 1: 34 cycles, read 2: 100 cycles, index 1: 10 cycles, index 2: 10 cycles).

For the [dithiobis (succinimidyl propionate)] (DSP)/methanol nuclei isolations and library construction based on Martin *et al.* (*21*), we used hypotonic lysis buffer solution B (with bovine serum albumin) for small volume tiny sci-RNA-seq3 nuclei isolation methods. For sci-RNA-seq3 library construction, we loaded ~20,000 nuclei per index 1 reverse transcriptase (RT) well in a 384 RT-well experiment with mouse and human brain added as separate QC nuclei and nuclei from cell lines human embryonic kidney (HEK) 293T (RRID:CVCL_0063) and NIH/3 T3 (RRID:CVCL_0594) combined as barnyard controls per RT plate. Nuclei from all RT plates were pooled and redistributed to ligation plates for the second index as previously published; after the addition of the second index, nuclei were again repooled for their final distribution of 4000 nuclei per well before second strand synthesis, protease digestion, tagmentation, and polymerase chain reaction all on this final third index plate.

snRNA-seq preprocessing

snRNA-seq sequencing reads were processed into a gene-by-nucleus expression matrix of UMI counts following the methods described by Cao *et al.* (*19*). We used largely an identical pipeline which, briefly, (i) converts base calls to fastq files with bcl2fastq/v.2.20 (RRID:SCR_015058) (Illumina), (ii) removes adapter sequences using Trim Galore/v.0.6.7 (RRID:SCR_011847) (*99*), (iii) aligns trimmed reads to a reference genome with STAR/v.2.7.6 (RRID:SCR_004463) (*100*), (iv) extracts mapped reads, (v) removes duplicates, and (vi) generates UMI counts for exonic and intronic regions of each gene, tabulated according to the unique three-level barcode design in sci-RNA-seq3. We used the rhesus macaque reference genome (Mmul_10) (*101*) and annotation, obtained from Ensembl (version 101) (RRID:SCR_002344). We extended the 3′ untranslated region annotations of genes and transcripts by 500 bp to avoid misclassifying genic reads as intergenic. The remainder of our pipeline followed the procedures described by Cao *et al.* (*19*). After generating the count matrix, we removed all nuclei with UMI counts < 100.

For each sample, we imported gene-by-nucleus count matrices into the AnnData/v.0.8.0 (RRID:SCR_018209) (*102*) framework and then ran Scrublet/v.0.2.3 (RRID:SCR_018098) (*103*)

(expected_doublet_rate = 0.05) to calculate doublet scores. We marked nuclei as doublets if they had Scrublet doublet scores > 0.20. For each sample, we additionally marked nuclei as doublets using per-sample thresholds determined by Scrublet and adjusted by eye as necessary to separate bimodal peaks visualized on the Scrublet doublet score histogram (fig. S1).

To further identify potential doublet nuclei, we used an iterative clustering strategy (*104*) implemented with Scanpy/v.1.9.1 (RRID:SCR_018139) (*105*). First, we combined all nuclei into a single AnnData object and filtered nuclei to those with UMI ≥ 100, number of expressed genes < 2500, and a percentage of reads mapping to the mitochondrial genome < 5%. We then removed all non-autosomal genes, genes located on unplaced scaffolds, and unexpressed genes. Next, we normalized the data to the total UMI per nucleus, logarithmized the data, and subsetted the data to the 10,000 most variable genes. For each cell, we regressed out total UMI counts per nucleus and then mean-centered and scaled the data. The dimensionality of the data was then reduced by principal components analysis (PCA) (50 components). To further reduce the dimensionality, we ran a UMAP (using umap-learn/v.0.5.2) (RRID:SCR_018217) analysis (*106*) with BBKNN/v.1.5.1 (RRID:SCR_022807) (*107*) to simultaneously correct for batch differences. For the BBKNN integration, we set neighbors_within_batch = 10 (given three batches, tantamount to UMAP n_neighbors = 30), used the cosine distance metric, and used the PyNNDescent/v.0.5.6 algorithm (RRID:SCR_022806) (*108*). We then ran UMAP using the settings min_dist = 0, spread = 1.0, and n_components = 10 to facilitate clustering (https://umap-learn.readthedocs.io/en/latest/clustering.html). For data visualization only (not clustering), we ran a similar BBKNN/UMAP pipeline with neighbors_within_batch = 5 (for three batches, tantamount to UMAP n_neighbors = 15), min_dist = 0.25, spread = 1.0, and n_components = 2. To cluster the data, we exported and imported the 10-dimensional UMAP matrix into Monocle3/v.1.2.9 (RRID:SCR_018685) (*19*) in R/v.4.0.2 (RRID:SCR_001905) (*109*) and then implemented the Leiden-clustering workflow in Monocle3 with a relatively high-resolution setting (resolution = 1 × 10^{-4}). For each cluster, we then calculated the mean Scrublet doublet score and marked all clusters with a mean Scrublet doublet score > 0.15 as doublet clusters (fig. S1).

After identifying doublets as described above, we removed all marked doublets and repeated the normalization, dimensionality reduction, and clustering procedures almost exactly as described above, with the only difference being a coarser cluster resolution setting in Monocle3 (resolution = 1 × 10^{-5}). We confirmed adequate removal of doublet cells by observing the clean separation of distinct cell types and the absence of clusters expressing obviously ambiguous marker gene profiles (fig. S1).

Removal of sci-RNA-seq cell contamination

During the course of cell type identification (see the following section), we observed the presence of two distinct clusters of cells (fig. S2A) with expression profiles resembling embryonic progenitors (markers genes, unknown cluster 1: *ASPM*, *CENPE*, *CENPF*, *MKI67*; unknown cluster 2: *COL1A1*, *COL1A2*, *FN1*, *VIM*), an unusual finding in adult primate brain samples. Because these were present in relatively large proportions in some samples (~25%)—but at low levels overall (2.2%)—and because our sci-RNA-seq experiments included control samples of exogenous

(i.e., non-macaque brain) origin (specifically, a fetal mouse brain positive control and a "barnyard" sample consisting of mixed human HEK293T and mouse NIH/3 T3 cells), we tested for the presence of contaminating nuclei of exogenous origin. We identified and removed contaminating cells as follows.

Because the only non-macaque samples included in all experiments were the control samples of either human or mouse origin, we used BBSplit/v.38.38 (RRID:SCR_016965) (110) to assign reads to the macaque, human, or mouse genomes. BBSplit is a competitive aligner that maps to several references simultaneously, assigning reads to the genome with the best unambiguous match. We used the following reference assemblies from Ensembl version 101: Mmul_10 (rhesus macaque), GRCh38.p13 (human), and GRCm38.p6 (mouse). After indexing the three references simultaneously using BBSplit, we aligned 10 million randomly sampled unique (de-duplicated) reads for each sample using default settings in BBSplit, which partitioned reads assigned to each genome into separate fastq files. Unmapped and ambiguous reads were directed to additional fastq files that were not used. Using a similar demultiplexing workflow to the sci-RNA-seq3 preprocessing pipeline, we tabulated reads-per-cell for each of the three genomes and calculated summary statistics.

After filtering to only cells with ≥10 unambiguously assigned reads by BBSplit, we observed that discernible fractions of exogenous reads (reads unambiguously assigned to human or mouse) were specific to certain barcodes from the first round of sci-RNA-seq barcoding [reverse transcription (RT)], indicating that a low level of cross-well contamination of cells or barcoded primers likely occurred at this stage (fig. S2C). We also observed that, after filtering to cells passing all previous QC filters, our clustering and annotation workflow had partitioned exogenous cells into two clusters corresponding to human and mouse cells respectively, with no discernible exogenous contamination in other annotated cell types (fig. S2B). After removing the entirety of the two exogenous clusters from the dataset (N = 58,443 cells), we reexamined the distribution of exogenous read fractions across RT barcodes and confirmed that human and mouse cells were effectively removed (fig. S2C).

Our discovery of contamination at the level of RT wells called into question whether similar contamination occurred elsewhere in our dataset. Any such contamination is more difficult to detect as contaminating cells would be of macaque origin. Because contamination is specific to RT barcodes, however, and because samples were loaded with multiple RT barcodes (typically at least 6), we scanned our dataset for barcodes for which proportions of identified cell types (see the following sections) deviated noticeably from those of other barcodes for a given sample.

To perform these scans, we computed the Jensen-Shannon divergence statistic using the "JSD" function from the philentropy package (RRID:SCR_022805) (111) in R. We calculated the Jensen-Shannon divergence by comparing cell class or cell subtype proportions for a given RT barcode to mean cell class or cell subtype proportions across all barcodes of a given sample. Because cell class and cell subtype counts/proportions tend to be sparser for barcodes/wells with fewer cells, leading to heightened risks of false specificity, we focused on barcodes with both higher cell counts (>10,000) and high specificities (cell class specificity > 0.15 and cell subtype specificity > 0.075) based on visual inspection of both distributions (fig. S3). By these criteria, we identified two RT barcodes that represented likely mixtures of multiple samples. For

these barcodes, we marked cells as potential contamination and, for downstream analyses, excluded them from analyses of regional proportions, specificity, and enrichment.

snRNA-seq cell type and cell subtype identification

To identify cell types, we visualized the expression of canonical marker genes (table S3) on normalized, log-transformed gene expression data using Scanpy. Most clusters were readily assigned to well-characterized cell types in this manner. To aid in the classification of more nuanced cell types, we determined top marker genes using logistic regression and t test marker gene methods implemented via the "rank_genes_groups" function in Scanpy. For each discrete cell type, we ran marker gene tests by testing gene expression in a given cell type against gene expression in all other cells in our dataset.

On the basis of canonical markers and data-derived marker genes, we identified 17 parent cell types (not including the two cells of exogenous origin, see the section above), which we refer to as cell classes. In all but two cases, our parent cell types corresponded with partitions identified through our clustering using Monocle3 (q value threshold = 0.05). In two cases, we considered clusters assigned to the same partition to be discrete parent cell types because they exhibited clear separation in our global analysis while clearly expressing canonical markers of known cell types (dopaminergic and serotonergic neurons; table S3) yet did not effectively segregate when their assigned partition (the partition also including GABAergic neurons) was analyzed separately.

To identify pathways associated with up-regulated and/or down-regulated genes in each cell class, we conducted pathway enrichment analysis using two complementary approaches. First, we computed pathways activity levels (PALs) for each cell class across 52,041 pathways from the Oncobox Pathway Databank (OncoboxPD; RRID:SCR_023723). We used the curated set of pathway annotations for human genes (based on HGNC nomenclature) downloaded from OncoboxPD (112), which we converted to macaque ENSEMBL (version 101) gene identifiers through biomaRt (RRID:SCR_019214). We calculated PALs using oncoboxlib/v.1.2.3 (RRID:SCR_023722) (113), with aggregated (pseudobulked) RNA profiles for each cell class within each brain region as input. When calculating the PALs for each cell class, we use all the remaining cell classes as "control" samples. For each cell class, we then calculated the average PAL across all brain regions. When summarizing these results, we focused on a set of nonredundant databases [Qiagen 1.5, Biocarta 1.2, Kyoto Encyclopedia of Genes and Genomes (KEGG) Adjusted 1.4, PathBank 1.0, Reactome 1.3, and NCI 1.2] and pathways with the 50 highest and 50 lowest average PALs (table S4). Second, we tested for enriched Gene Ontology (GO; RRID:SCR_002811) biological processes (BP), molecular functions (MF), and cellular components (CC) using GO annotations for the rhesus macaque reference genome (Mmul_10) downloaded from ENSEMBL (verison 101) via biomaRt (RRID:SCR_019214). For each cell class, we used the t test statistic from our marker gene analysis as input, after filtering out mitochondrial genes and genes from unplaced scaffolds. We performed GO enrichment analysis using threshold-independent Kolmogorov-Smirnov (KS) tests implemented in topGO (RRID:SCR_014798), which corrects for the correlated graph structure of the underlying GO database (114). KS enrichment tests were run using the "weight01" algorithm in topGO. We tested for

enriched GO classes separately for up-regulated and down-regulated genes and separately for each cell class and GO namespace (BP, MF, and CC). We considered GO classes to be significantly enriched where $P_{adj} < 0.01$ (table S5).

To identify cell subtypes, we partitioned the data by cell class and reanalyzed each data partition individually. For each cell class–specific analysis, we repeated a preprocessing, dimensionality reduction, and clustering analysis that largely followed the pipeline described above for our global analysis, with the following exceptions. After normalizing and log-transforming the data, we identified the 2000 most variable genes for each given cell type and subset the data to those highly variable genes. Because we observed that differences in total UMI among batches resulted in artifactual clusters being identified downstream (even after batch correction with BBKNN, a problem we did not observe in our global analysis), we regressed out total UMI counts per nucleus separately for each batch. We then combined residual values from all batches before mean centering and scaling for PCA and UMAP analysis. For Leiden clustering, we used the same resolution parameter (resolution = 1×10^{-5}) for most cell types but, in four cases, defaulted to partitions identified using Monocle3 (q value threshold = 0.05) after observing small clusters with unusually high UMI. We considered clusters/partitions identified in this manner to be cell subtypes. To explore cell-cell variation with even higher granularity, we performed another round of clustering using a relatively fine resolution parameter (resolution = 1×10^{-4}). In all cases, this resolution parameter also maximized the modularity of the Leiden clustering algorithm and was thus automatically selected by Monocle3 as the ideal clustering resolution. We refer to clusters identified in this manner as cell subclusters.

As with our global (all cell classes combined) analysis, for each cell subtype, we identified top marker genes using logistic regression and t test marker gene methods implemented in Scanpy. In addition we used an NNLS approach (19) to identify correlations between cell subtypes and annotated labels in reference datasets (table S7 and fig. S10).

In addition, we scanned for gene-disease associations that were enriched among the top 100 marker genes for each cell subtype. We used gene-disease associations from the DISEASES database (RRID:SCR_015664) (59) and used Fisher's exact test to identify overrepresented disease associations among the top 100 marker genes for a given cell subtype, using all macaque genes in our analysis as the background (table S13).

Because of the large number of cell subclusters identified and smaller number of cells per subcluster, we used a neighborhood-based procedure to identify marker genes for each subcluster. For cells of a given subcluster, we identified neighboring cells using the batch-corrected weighted adjacency matrix determined with BBKNN, keeping all cells with weights > 0 after excluding cells from the same given subcluster. We then calculated marker genes using this set of neighboring cells as the reference, using the t test method implemented in Scanpy (table S8).

For all cell subtypes and cell subclusters, we assessed the reliability of each annotation using a machine learning classification approach with fivefold cross-validation. For each cell class, we used the LinearSVC function from scikit-learn/v.1.0.2 (RRID:SCR_002577) (115) to train a classifier using normalized, log-transformed, and scaled expression data. After aggregating predictions across the five partitions, we assessed cell subtype and subcluster "quality scores" (classification accuracy) as the frequency of annotated cells assigned the correct label by the classifier (tables S7 and S8 and fig. S9).

Cell composition and regional heterogeneity analysis

To assess the specificity of cell classes and/or subtypes, we calculated the Jensen-Shannon divergence statistic using the "JSD" function from the philentropy package (RRID:SCR_022805) (111) in R. We calculated the Jensen-Shannon divergence by comparing, for a given cell class or cell subtype, the cell type's count distribution across brain regions to the count distribution (combining all cell types per region) of the entire dataset combined (29).

To measure regional heterogeneity within cell types, we extended our recently developed statistic, lochNESS (41), to quantitatively measure enrichment of each region subclass or region within each cell's neighborhood. For each cell type, we define lochNESS of $cell_n$ for $region_m$ as

$$\text{lochNESS}_{cell_n,region_m} = \frac{\text{\# of cells from } region_m \text{ in kNNs of } cell_n}{k} \Big/ \frac{\text{\# of cells from } region_m \text{ in cell type}}{N}$$

where N is the total number of cells in the cell type, and k is the number of nearest neighbors for $cell_n$.

For each cell type, the calculation results in a cell × region matrix, where each row can be separately visualized. For a summarizing visualization, each cell can be colored by the region with the largest lochNESS. In addition, when we focus on a subset of regions (e.g., just the cortical regions), we calculate a normalized lochNESS that is comparable across the regions of interest

$$\text{lochNESS}^*_{cell_n,region_m} = \text{lochNESS}_{cell_n,region_m} \Big/ \sum_{m=1}^{M} \text{lochNESS}_{cell_n,region_m}$$

where M is the number of regions or region subclasses of interest.

To identify genes that are expressed with regional bias, we fit a regression model for each gene to identify regions with significant nonzero correlation with gene expression as implemented in Monocle3 (19). The model for each cell type is

$$\begin{aligned} log(expression) = \beta_0 &+ \beta_1 \times \text{lochNESS}_{region_1} \\ &+ \beta_2 \times \text{lochNESS}_{region_2} + \ldots \\ &+ \beta_m \times \text{lochNESS}_{region_m} \end{aligned}$$

where β_0 is the intercept and $\text{lochNESS}_{region_i}$ is a vector of lochNESS across all cells in the cell type.

Hierarchical clustering of cells and regions

We used Scanpy to cluster cell classes and brain regions based on the top 50 principal components of gene expression. Because of our use of BBKNN for batch correction in our main workflow, our PCA was not actually corrected for batch. To rectify this, we first used the harmonpy/v.0.0.5 (RRID:SCR_022798) implementation of Harmony (RRID:SCR_022206) (116) to generate a batch-corrected PCA matrix (convergence after two generations). We then used the Scanpy "dendrogram" function to perform hierarchical clustering using the batch-corrected PCA embedding. To visualize uncertainty, we performed 1000 bootstrap iterations in which we resampled cells randomly with replacement and computed new dendrograms. We then used the "DensiTree" function (117) implemented in the

phangorn/v.2.6.3 (RRID:SCR_017302) (118) R package to visualize trees. We performed this procedure using both cell class and brain region as labels (fig. S6, A and B).

For brain regions, we also performed hierarchical clustering using the cell proportion (cell class × brain region) matrix. We used the "hclust" function in R to cluster using the "complete" method based on Euclidean distances. To again visualize uncertainty, we resampled all cells in our dataset 1000× with replacement and then repeated calculation of cell class proportions and hierarchical clustering. We visualized the final tree with "DensiTree" (fig. S6B).

snATAC-seq data generation

To profile single-nucleus chromatin accessibility, we performed snATAC-seq using the sci-ATAC-seq3 approach (45), which is the improved version of the original sci-ATAC-seq protocol (44). We followed the protocol of Domcke et al. (97), with slight modifications. Briefly, we added 50 µl of Omni-ATAC lysis buffer to pulverized tissue and homogenized the tissue with 5 to 10 strokes with a disposable RNase-free plastic pestle (Fisherbrand, catalog no. 12-141-364). We then added another 950 µl of Omni-ATAC lysis buffer, mixed by pipette, incubated on ice for 3 min, and then transferred the suspension to a new 15-ml conical tube containing 5 ml of ATAC resuspension buffer (ATAC-RSB) with 0.1% Tween 20. We then pelleted the nuclei, removed the supernatant, and resuspended the pellet in 1 ml of 1× Dulbecco's phosphate buffered saline (DPBS). We then transferred the suspension through a 70-µm cell strainer (pluriSelect catalog no. 43-10070-70) into a 15-ml conical tube containing 4 ml of 1× DPBS and 140 µl of 37% formaldehyde (final concentration, 1% formaldehyde). We then incubated the nuclei for 10 min with occasional mixing. The fixation was then quenched with 250 µl of 2.5 M glycine, incubated for 5 min at room temperature, and then incubated for another 15 min on ice. We then pelleted the nuclei, removed the supernatant, and resuspended the pellet in 2-ml freezing buffer. Nuclei were counted by mixing with 1 µM YOYO-1 iodide (Thermo Fisher Scientific, catalog no. Y3601) using a Countess II FL automated cell counter (Life Technologies), divided into tubes in 50-µl aliquots, and then flash-frozen in liquid nitrogen.

Frozen fixed nuclei were prepared for the sci-ATAC-seq3 library similar to Domcke et al. (45). Omni-ATAC lysis buffer [10 mM NaCl, 3 mM MgCl$_2$, 10 mM tris-HCl (pH 7.4), 0.1% IGEPAL CA-630, 0.1% Tween 20, and 0.01% digitonin] was used to permeabilize fixed nuclei before diluting samples with ATAC-RSB [10 mM NaCl, 3 mM MgCl$_2$, and 10 mM tris-HCl (pH 7.4)] supplemented with 0.1% Tween 20. Approximately 200,000 nuclei per sample were spread across four wells for tagmentation as previously described. Barnyard control for each set of experiments included mouse cell line (CH12-LX; RRID:CVCL_0211) and human pancreas as a QC tissue.

Our combined snATAC-seq dataset encompasses data prepared using five sci-ATAC-seq3 experimental runs (i.e., library preparation/sequencing batches). Sample order was randomized between batches to ensure balance of brain regions, sex, and hemispheres between runs and to minimize batch effects.

snATAC-seq preprocessing

snATAC-seq sequencing reads were processed into a peak-by-nucleus count matrix following the methods described by Domcke et al. (45). We followed largely an identical pipeline

which, briefly, (i) converts base calls to fastq files with bcl2fastq/v.2.20 (Illumina), (ii) removes adapter sequences using Trimmomatic/v.0.39 (RRID:SCR_011848) (119), (iii) aligns trimmed reads to a reference genome with bowtie2/v.2.4.1 (RRID:SCR_016368) (120), (iv) calculates nonduplicate fragment endpoints for each cell, (v) calls peaks using MACS2/v.2.2.7.1 (RRID:SCR_013291) (121, 122) and merges peaks across samples to create a merged BED file, (vi) tabulates reads from merged peaks and annotated TSSs (±1 Kb around each TSS) for QC, (vii) separates cell barcodes from background barcodes by fitting a mixture of two negative binomials (noise versus signal), and (viii) assembles a sparse matrix tabulating reads per cell barcode falling within the master set of peaks and within gene bodies extended by 2 kb upstream. We used the rhesus macaque reference genome (Mmul_10) (101) and annotation, obtained from Ensembl (version 101), and merged peaks across all samples (encompassing five library preparation and sequencing batches) to create a global set of peaks. After binarizing UMI counts, we filtered the peak-by-nucleus matrix to include only nuclei with ≥ 1000 binarized UMI, less than 100,000 binarized UMI, and ≥ 30% fraction of reads in peaks (fig. S13).

We identified and removed doublets using a similar iterative clustering approach to that described for our single-nucleus RNA dataset (fig. S13). Briefly, we ran Scrublet/v.0.2.3 (103) on each sample individually and marked doublets using both a universal threshold (Scrublet doublet score > 0.20) and a per-sample threshold determined by Scrublet and checked and adjusted (if necessary) by eye. We then performed a similar preprocessing, dimensionality reduction, and clustering pipeline to identify clusters with relatively high Scrublet doublet scores (mean Scrublet doublet score > 0.15). We lastly removed all nuclei marked as doublets based on the described criteria before concatenating all singlet nuclei across all samples together.

Our snATAC-seq preprocessing, dimensionality reduction, and clustering pipeline likewise tracked closely to our snRNA-seq analysis, with minor modifications to accommodate best practices for ATAC-seq data. Briefly, we filtered the data to remove peaks that were not accessible in a minimum of five cells as well as peaks that were located on non-autosomal or unplaced scaffolds in the macaque genome. We then filtered the data to the top 100,000 variable features. We performed latent semantic analysis (LSI) on the resulting peak-by-cell matrix to reduce the dimensionality of the data. We performed term frequency/inverse document frequency (TF-IDF) normalization followed by singular value decomposition (SVD) following previously described procedures (45) to reduce the data to 50 PCA dimensions. L2 normalization was then performed on the last 49 principal components, thereby excluding the first principal component, which tends to capture read depth (45). TF-IDF, SVD, and L2 normalization procedures were implemented using scikit-learn/v.1.0.2 (RRID:SCR_002577) (115). The L2-normalized PCA matrix was then reduced further and clustered using an identical BBKNN/UMAP/Monocle3 approach to that used for our snRNA-seq data. Doublet-derived clusters were also marked for removal using an identical threshold (mean Scrublet doublet score > 0.15).

After marking and removing doublets from our data, we repeated our preprocessing, dimensionality reduction, and clustering pipeline. After observing clear separation of distinct cell classes, we used MUON/v.0.1.2 (RRID:SCR_022804) (123) to calculate promoter accessibility scores by tabulating binarized UMI counts

within the region 2000 bp upstream of a TSS. Because at the time of this analysis MUON did not factor in DNA strand information, we ran the function "count_fragments_features" separately for + and − strand genes, using the "upstream_bp" or "downstream_bp" arguments as necessary to tabulate counts in the correct upstream region [extending from the TSS to (TSS − 2000 bp) or (TSS + 2000 bp), respectively] (https://github.com/scverse/muon/issues/59). We used Scanpy to normalize and visualize resulting promoter accessibility scores (Fig. 3B). We provisionally classified nuclei based on promoter accessibility scores of known marker genes.

Integration of snRNA-seq and snATAC-seq data

We used GLUE implemented in scglue/v.0.2.3 (RRID:SCR_022803) (*47*) to integrate our snRNA-seq and snATAC-seq datasets. To run scglue, we followed preprocessing procedures in Scanpy recommended by the scglue authors for both our snRNA-seq and snATAC-seq data, after filtering out doublets as described above. For snRNA-seq data, we identified the top 2000 most variable genes, then normalized, log-transformed, and scaled the data using default parameters in Scanpy. We then reduced the dimensionality of the data to the top 100 principal components using PCA, based on the top 2000 variable genes and the automatic SVD solver selected by Scanpy. For snATAC-seq data, we used the LSI implementation in scglue to reduce the data to the top 100 principal components, with the number of power iterations set to 15.

We then used scglue to compute a prior guidance graph and propagated highly variable snRNA-seq features (genes) to identify highly variable snATAC-seq features (peaks) based on the guidance graph. We then built and trained the GLUE integration model using the PCA and LSI embeddings, respectively, as the first encoding transformation, modeling raw counts of both snRNA-seq and snATAC-seq data using the negative binomial model, using the batch correction option to correct for sequencing batches, and using the previously computed prior guidance graph as input. As all nuclei from this study were included (totaling over 4 million nuclei), this analysis was particularly computationally demanding. We performed this analysis on a machine with 1.5 TB RAM, accelerated by 4 Tesla V100 (NVIDIA) GPUs.

After training a GLUE model, we validated effective integration by calculating integration consistency scores using scglue (fig. S14A). We then calculated integrated cell and feature embeddings for both snRNA-seq and snATAC-seq data using scglue. After projecting all cells to a unified embedding, we performed UMAP dimensionality reduction using the same procedures as described previously, with one exception. Because the unified GLUE embedding was already batch-corrected, we computed the neighborhood graph using the Scanpy "neighbors" function rather than BBKNN, with n_neighbors = 15.

To transfer cell class labels from our snRNA-seq data to our snATAC-seq data, we used the "transfer_labels" function in scglue, which computes shared nearest neighbors between reference (snRNA-seq) and query (snATAC-seq) nuclei, weighted by the Jaccard index. Jaccard indices are then normalized per query nucleus to form a mapping matrix, which is then multiplied by one-hot-encoded reference labels. The reference label with the highest score is then assigned as the predicted cell type, with the highest score retained as the confidence score. For label transfer, because a subset of our snRNA-seq data was derived from samples that were unprofiled in our snATAC-seq data, we limited our reference RNA-seq dataset to only those nuclei deriving from samples profiled in both snRNA-seq and snATAC-seq experiments. We then retained 100,000 nuclei from withheld (unmatched) snRNA-seq samples as a query dataset to evaluate label transfer accuracy. For snATAC-seq label transfer, we used all snATAC-seq nuclei as a query dataset. We used previously assigned parent cell types for our snRNA-seq dataset as reference labels. For our snATAC-seq query nuclei, we retained all predicted cell class labels with a label transfer confidence score ≥ 0.95. At this threshold, the error rate in our evaluation dataset was 0.43% (fig. S14B).

Identification of cCREs

To scan for cCREs underlying differential expression among brain cells, we used two complementary approaches. First, we used the integrative GLUE regulatory inference approach implemented in scglue/v.2.0.3 (*47*), which calculates regulatory scores based on the cosine similarities between multi-omics data features in an integrated space. Second, we used a metacell approach to construct multi-omic samples (determined via k-means clustering in integrated space) with aggregated (pseudobulk) gene expression and chromatin accessibility counts, which we then modeled using logistic regression. Last, we calculated differentially accessible peaks using a similar workflow to our snRNA-seq marker gene analysis.

To calculate GLUE regulatory scores, we performed a second integration of our snRNA-seq and snATAC-seq datasets, following an identical pipeline except including the top 6000 most variable genes (rather than 2000). This allowed us to identify putative gene:peak regulatory connections and to generate an integrated feature embedding for a greater number of genes and genomic regions. We constructed a window graph between inferred promoters—which we calculated as the region from the strand-specific TSS extended upstream 2000 bp—and peaks using the "window_graph" function, with the window size set to 150 kb and a distance-decaying weight, as recommended by the scglue authors. We then used the previously computed window graph and feature embeddings to perform the regulatory inference analysis using the "regulatory_inference" function, with the alternative hypothesis set to "greater" to perform a one-sided test.

To determine the directionality of putative regulatory relationships, we used a second approach based on metacell identification and logistic regression (Fig. 5A). We use the "get_metacells" function to generate multi-omic (snRNA-seq/snATAC-seq) metacells based on k-means clustering of their integrated cell embeddings. As our snRNA-seq dataset included 2,583,967 single-cell transcriptomes, we set k (n_meta) to 10,335 to target a mean of roughly 250 RNA transcriptomes per metacell. After identifying metacells in this manner, we summed (pseudobulked) gene expression per metacell. For each gene:peak pair tested in our GLUE regulatory inference, we then performed a logistic regression modeling accessibility of each individual snATAC-seq cell in a given metacell (1: open, 0: closed) as a function of $\log_2$CPM-normalized gene expression for that snATAC-seq cell's respective metacell. Logistic regressions were performed in R/v.4.0.2.

We considered candidate cis-regulatory relationships to be gene:peak pairs for which false discovery rate–adjusted $P < 0.05$ for both the GLUE regulatory inference and metacell-based logistic regression tests. We classified candidate cis-regulatory relationships as

positive or negative relationships based on the sign of their logistic regression coefficients (β values) (fig. S17).

For all peaks, we also tested for marker peaks (peaks with differentially accessibility) using logistic regression and t test marker gene methods implemented via the "rank_genes_groups" function in Scanpy. Similar to our snRNA-seq marker genes analysis, we ran marker peak tests by testing chromatin accessibility in a given cell type against accessibility in all other cells in our dataset. In addition, to validate marker peaks, we used a second logistic regression approach implemented via the "FindMarkers" function in Seurat/v.4.1.1 (RRID:SCR_016341) (*124*). In contrast to the logistic regression in Scanpy, the Seurat implementation is not a regularized procedure and is thus able to control for latent variables and to calculate P values. To reduce computational burden, we ran "FindMarkers" on a dataset with 1000 cells per cell class. As we found that output statistics (regularized logistic regression coefficient in Scanpy and log FC in Seurat) were highly concordant (fig. S18), we report Scanpy results here as they included all possible cells. We considered peaks to be differentially accessible if the regularized logistic regression coefficient > 0, the log FC > 0, and the t test P_{adj} < 0.05.

snATAC-seq cell subtype analysis

To mitigate peak calling biases while allowing us to probe more nuanced regulatory variation within cell populations, we called a new set of cell class–specific peaks for each cell type with assigned cells, skipping rarer cell types for which no snATAC-seq nuclei passed our prediction threshold above.

Following scglue integration and assignment of snATAC-seq cells to cell classes, we created cell class–specific pseudobulk epigenomes by aggregating all nonduplicate fragment endpoints for each cell class. These cell class–level ATAC-seq data were then used for peak calling using MACS3/v.3.0.0a6 (*121*, *122*), with the same peak calling parameters that we used for each sample and batch described in the "snATAC-seq preprocessing" section above ("-g 2.7e9 --call-summits --nomodel"). For each cell class, we repeated steps from our snATAC-seq data generation pipeline to tabulate reads from newly called peaks and to assemble sparse count matrices matrix tabulating reads per cell barcode falling within the master set of peaks and within gene bodies extended by 2 kb upstream. We then imported peak-by-nucleus count matrices into the AnnData/v.0.8.0 (*102*) framework.

To assign cell subtypes for our snATAC-seq data (fig. S15), we repeated preprocessing, data integration, label transfer, and regulatory inference procedures described above on each cell class individually. In contrast to our global joint analysis, we only included snRNA-seq nuclei deriving from samples that were profiled in both snRNA-seq and snATAC-seq experiments and used the top 6000 most variable genes in our snRNA-seq analysis and used the snATAC-seq peak sets specific to each cell type. The remainder of our preprocessing and data integration procedures followed the same pipeline described previously for our global integration analysis. For label transfer, we also followed largely the same procedures as for our global label transfer pipeline. We did not, however, use a label transfer confidence score threshold under the assumption that snATAC-seq nuclei would, on average, be assigned to the correct cell subtype and, if incorrect, would be assigned to a closely related cell subtype (i.e., a neighboring subtype in the integrated multidimensional cell space). For metacell-based regulatory

inference, we varied the settings for k based on dataset size to target a mean of 50 transcriptomes per metacell.

Evolutionary analysis of candidate regulatory regions

We tested whether cCREs that were differentially accessible across cell classes (called on the global dataset; N = 2838 of 3404 unique cCREs; table S15) or cell class–specific (called within each of N = 7 cell classes; N = 469 of 5875 unique cCREs across all cell classes; table S16) were associated with regions that underwent particularly rapid evolution in the human lineage. We first obtained genomic locations for two sets of evolutionarily salient regions, including HARs (*73*) and HAQERs (*74*). The latter differ from the former in that they are not limited to highly conserved regulatory elements (*74*). We converted all HARs and HAQERs from the human genome (hg19 or hg38, respectively) to the rhesus macaque reference genome (Mmul_10) (*101*) using UCSC's liftOver/v.302 (RRID:SCR_018160) tool (*125*), allowing for multiple output regions. We similarly obtained macaque genomic locations for regions reported to exhibit differential accessibility between human and chimpanzee cerebral organoids (DA$_{HC}$) (*75*) using this method (from hg19 to Mmul_10). Most regions were successfully converted for all datasets, although success rates were lower for HAQERs, as expected [HARs: N = 2737 submitted, N = 6 failed, N = 2731 (99.8%) successfully converted into N = 2733 regions; HAQERs: N = 1581 submitted, N = 383 failed, N = 1198 (75.8%) successfully converted into N = 1417 regions; DA$_{HC}$: N = 17,935 submitted; N = 261 failed; N = 17,674 (98.5%) successfully converted into N = 17,744 regions]. We used Fisher's exact tests to test for overlap between (i) differentially accessible cCREs and HARs, HAQERs, and DA$_{HC}$ peaks (N = 3 tests) and (ii) cell class–specific cCREs and HARs, HAQERs, and DA$_{HC}$ peaks (N = 21 tests). Using the methods above, we also tested whether regulatory peaks (cCREs) were enriched for evolutionarily salient gene sets, relative to all detected peaks called on the global dataset [3404 of 1,189,415 (0.3%) peaks are cCREs]. P values were adjusted across all tests using the Benjamini-Hochberg method (table S17).

TF binding site enrichment

For enrichment analyses at the cell class level, we focused on peaks that were deemed accessible in one and only one cell class, which we called "cell class–unique peaks." We identified these peaks using BedTools/v.2.30.0 ("intersect −v") (RRID:SCR_006646) (*126*) to find all peaks in a cell class that did not overlap with any peak called in another cell class. The number of peaks identified in this manner ranged from 655 (in ependymal cells) to 71,049 (in glutamatergic neurons). We tested for enrichment of TF binding motifs in cell class–unique peaks compared to the background of the rhesus macaque genome while controlling for GC content, implemented in the monaLisa/v.1.3.1 (RRID:SCR_022802) (*127*) in R/v.4.1.0 (table S11). We used the JASPAR 2018 (RRID:SCR_003030) nonredundant vertebrate core position weight matrices (*128*).

At the cell subtype level, we tested for enrichment using the top differentially accessible peaks among subtypes of the same cell class, excluding peaks with regularized logistic regression coefficients < 0 (table S12). We retained the top first percentile of marker peaks, ranked according to their regularized logistic regression coefficients.

Disease heritability enrichment

We calculated enrichment of disease-associated variants in cell class–specific accessible chromatin regions using LDSC (RRID:SCR_022801) (Fig. 6 and table S19) (*77*, *78*). Because the trait-associated loci are annotated in the human genome, we converted all peaks (at the combined level as well as each individual cell class level) from the rhesus macaque genome coordinates to GRCh37 using UCSC's liftOver/v.302 (RRID:SCR_018160) tool (*125*). We followed the standard pipeline using the 1000 Genomes baseline model and precomputed .sumstats files. A list of phenotypes tested can be found in table S18.

Methods summary

Briefly, we collected fresh-frozen brains from five adult rhesus macaques that were part of the free-ranging CPRC research colony on Cayo Santiago. We focused our atlas on 30 anatomically defined regions that are associated with key cognitive, behavioral, and disease traits. To allow for the profiling of multiple genomic modalities from the same representative cell populations, we pulverized all samples on dry ice to homogenize and divide tissue for single-nucleus sequencing. We generated snRNA-seq data from 2,583,967 nuclei spanning a total of 30 unique regions from both hemispheres of the brain and paired those data with snATAC-seq data from 1,587,880 28 regions across 28 unique regions. These data were generated using sci-RNAseq3 (*19*) and sci-ATACseq3 (*45*) combinatorial indexing. Single-nucleus libraries were deeply sequenced and processed using a uniform protocol that included extensive QC filters (figs. S1 and S2).

Using Leiden-clustering on snRNA-seq nuclei (*19*), we identified 17 primary cell classes and then iteratively clustered each cell class for deeper annotation of cell subtypes. Whenever external data were available, we validated our cell classifications using an NNLS approach (*19*) to identify correlations between cell subtypes and annotated labels in reference datasets. We then identified marker genes for each cell class and subtype, characterized the regional distribution and expression of each cell class and subtype across the brain, and identified cell-specific enrichment of disease-associated genes.

To connect snATAC-seq profiles to snRNA-seq nuclei, we used the GLUE integration approach (*47*), which allowed us to annotate all snATAC nuclei based on the cell classes and subtypes identified in our snRNA-seq data. These connections allowed us to carry out a range of analyses, including TF binding site enrichment, linking TF enrichment to and TF expression within cell types, and identifying cell-specific regulatory links between cCREs and nearby genes. Last, following coordinate liftover between the primate and human genomes, we used LDSC (*77*, *78*) to quantify enrichment of neurological disease-associated variants in cell class biased cCREs.

Raw sequencing data and the annotated count matrices are available through NeMO (RRID:SCR_016152), and protocols for data generation are on protocols.io (DOI:10.17504/protocols.io.9yih7ue and DOI:10.17504/protocols.io.be8mjhu6); and scripts to process samples and recreate all analyses are available on GitHub (Data and Materials Availability).

Supplementary Materials

This PDF file includes:
Figs. S1 to S19

Legends for tables S1 to S19

Other Supplementary Material for this manuscript includes the following:
Tables S1 to S19
List of consortium authors

REFERENCES AND NOTES

1. R. I. M. Dunbar, S. Shultz, Evolution in the social brain. *Science* **317**, 1344–1347 (2007).
2. A. Navarrete, C. P. van Schaik, K. Isler, Energetics and the evolution of human brain size. *Nature* **480**, 91–93 (2011).
3. C. Darwin, *The Descent of Man, and Selection in Relation to Sex* (John Murray, 1871).
4. S. Herculano-Houzel, C. E. Collins, P. Wong, J. H. Kaas, Cellular scaling rules for primate brains. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 3562–3567 (2007).
5. K. L. Grasby, N. Jahanshad, J. N. Painter, L. Colodro-Conde, J. Bralten, D. P. Hibar, P. A. Lind, F. Pizzagalli, C. R. K. Ching, M. A. B. McMahon, N. Shatokhina, L. C. P. Zsembik, S. I. Thomopoulos, A. H. Zhu, L. T. Strike, I. Agartz, S. Alhusaini, M. A. A. Almeida, D. Alnæs, I. K. Amlien, M. Andersson, T. Ard, N. J. Armstrong, A. Ashley-Koch, J. R. Atkins, M. Bernard, R. M. Brouwer, E. E. L. Buimer, R. Bülow, C. Bürger, D. M. Cannon, M. Chakravarty, Q. Chen, J. W. Cheung, B. Couvy-Duchesne, A. M. Dale, S. Dalvie, T. K. de Araujo, G. I. de Zubicaray, S. M. C. de Zwarte, A. den Braber, N. T. Doan, K. Dohm, S. Ehrlich, H.-R. Engelbrecht, S. Erk, C. C. Fan, I. O. Fedko, S. F. Foley, J. M. Ford, M. Fukunaga, M. E. Garrett, T. Ge, S. Giddaluru, A. L. Goldman, M. J. Green, N. A. Groenewold, D. Grotegerd, T. P. Gurholt, B. A. Gutman, N. K. Hansell, M. A. Harris, M. B. Harrison, C. C. Haswell, M. Hauser, S. Herms, D. J. Heslenfeld, N. F. Ho, D. Hoehn, P. Hoffmann, L. Holleran, M. Hoogman, J.-J. Hottenga, M. Ikeda, D. Janowitz, I. E. Jansen, T. Jia, J. Jockwitz, R. Kanai, S. Karama, D. Kasperaviciute, T. Kaufmann, S. Kelly, M. Kikuchi, M. Klein, M. Knapp, A. R. Knodt, B. Krämer, M. Lam, T. M. Lancaster, P. H. Lee, T. A. Lett, L. B. Lewis, I. Lopes-Cendes, M. Luciano, F. Macciardi, A. F. Marquand, S. R. Mathias, T. R. Melzer, Y. Milaneschi, N. Mirza-Schreiber, J. C. V. Moreira, T. W. Mühleisen, B. Müller-Myhsok, P. Najt, S. Nakahara, K. Nho, L. M. O. Loohuis, D. P. Orfanos, J. F. Pearson, T. L. Pitcher, B. Pütz, Y. Quidé, A. Ragothaman, F. M. Rashid, W. R. Reay, R. Redlich, C. S. Reinbold, J. Repple, G. Richard, B. C. Riedel, S. L. Risacher, C. S. Rocha, N. R. Mota, L. Salminen, A. Saremi, A. J. Saykin, F. Schlag, L. Schmaal, P. R. Schofield, R. Secolin, C. Y. Shapland, L. Shen, J. Shin, E. Shumskaya, I. E. Sønderby, E. Sprooten, K. E. Tansey, A. Teumer, A. Thalamuthu, D. Tordesillas-Gutiérrez, J. A. Turner, A. Uhlmann, C. L. Vallerga, D. van der Meer, M. M. J. van Donkelaar, L. van Eijk, T. G. M. van Erp, N. E. M. van Haren, D. van Rooij, M.-J. van Tol, J. H. Veldink, E. Verhoef, E. Walton, M. Wang, Y. Wang, J. M. Wardlaw, W. Wen, L. T. Westlye, C. D. Whelan, S. H. Witt, K. Wittfeld, C. Wolf, T. Wolfers, J. Q. Wu, C. L. Yasuda, D. Zaremba, Z. Zhang, M. P. Zwiers, E. Artiges, A. A. Assareh, R. Ayesa-Arriola, A. Belger, C. L. Brandt, G. G. Brown, S. Cichon, J. E. Curran, G. E. Davies, F. Degenhardt, M. F. Dennis, B. Dietsche, S. Djurovic, C. P. Doherty, R. Espiritu, D. Garijo, Y. Gil, P. A. Gowland, R. C. Green, A. N. Häusler, W. Heindel, B.-C. Ho, W. U. Hoffmann, F. Holsboer, G. Homuth, N. Hosten, C. R. Jack Jr, M. Jang, A. Jansen, N. A. Kimbrel, K. Kolskår, S. Koops, A. Krug, K. O. Lim, J. J. Luykx, D. H. Mathalon, K. A. Mather, V. S. Mattay, S. Matthews, J. M. Van Son, S. C. McEwen, I. Melle, D. W. Morris, B. A. Mueller, M. Nauck, J. E. Nordvik, M. M. Nöthen, D. S. O'Leary, N. Opel, M.-L. P. Martinot, G. B. Pike, A. Preda, E. B. Quinlan, P. E. Rasser, V. Ratnakar, S. Reppermund, V. M. Steen, P. A. Tooney, F. R. Torres, D. J. Veltman, J. T. Voyvodic, R. Whelan, T. White, H. Yamamori, H. H. H. Adams, C. Bis, S. Debette, C. Decarli, M. Fornage, V. Gudnason, E. Hofer, M. A. Ikram, L. Launer, W. T. Longstreth, O. L. Lopez, B. Mazoyer, T. H. Mosley, G. V. Roshchupkin, C. L. Satizabal, R. Schmidt, S. Seshadri, Q. Yang; Alzheimer's Disease Neuroimaging Initiative; CHARGE Consortium; EPIGEN Consortium; IMAGEN Consortium; SYS Consortium; Parkinson's Progression Markers Initiative, M. K. M. Alvim, D. Ames, T. J. Anderson, O. A. Andreassen, A. Arias-Vasquez, M. E. Bastin, B. T. Baune, J. C. Beckham, J. Blangero, D. I. Boomsma, H. Brodaty, H. G. Brunner, R. L. Buckner, J. K. Buitelaar, J. R. Bustillo, W. Cahn, M. J. Cairns, V. Calhoun, V. J. Carr, X. Caseras, S. Caspers, G. L. Cavalleri, F. Cendes, A. Corvin, B. Crespo-Facorro, J. C. Dalrymple-Alford, U. Dannlowski, E. J. C. de Geus, I. J. Deary, N. Delanty, C. Depondt, S. Desrivières, G. Donohoe, T. Espeseth, G. Fernández, S. E. Fisher, H. Flor, A. J. Forstner, C. Francks, B. Franke, D. C. Glahn, R. L. Gollub, H. J. Grabe, O. Gruber, A. K. Håberg, A. R. Hariri, C. A. Hartman, R. Hashimoto, A. Heinz, F. A. Henskens, M. H. J. Hillegers, P. J. Hoekstra, A. J. Holmes, L. E. Hong, W. D. Hopkins, H. E. Hulshoff Pol, T. L. Jernigan, E. G. Jönsson, R. S. Kahn, M. A. Kennedy, T. T. J. Kircher, P. Kochunov, J. B. J. Kwok, S. Le Hellard, C. M. Loughland, N. G. Martin, J.-L. Martinot, C. McDonald, K. L. McMahon, A. Meyer-Lindenberg, P. T. Michie, R. A. Morey, B. Mowry, L. Nyberg, J. Oosterlaan, R. A. Ophoff, C. Pantelis, T. Paus, Z. Pausova, B. W. J. H. Penninx, T. J. C. Polderman, D. Posthuma, M. Rietschel, J. L. Roffman, L. M. Rowland, P. S. Sachdev, P. G. Sämann, U. Schall, G. Schumann, R. J. Scott, K. Sim, S. M. Sisodiya, J. W. Smoller, I. E. Sommer, B. St. Pourcain, D. J. Stein, A. W. Toga, J. N. Trollor, N. J. A. Van der Wee, D. van 't Ent, H. Völzke, H. Walter,

B. Weber, D. R. Weinberger, M. J. Wright, J. Zhou, J. L. Stein, P. M. Thompson, S. E. Medland; Enhancing Neuroimaging Genetics through Meta-Analysis Consortium (ENIGMA)—Genetics working group, The genetic architecture of the human cerebral cortex. *Science* **367**, eaay6690 (2020).

6. J. Zeng, G. Konopka, B. G. Hunt, T. M. Preuss, D. Geschwind, S. V. Yi, Divergent whole-genome methylation maps of human and chimpanzee brains reveal epigenetic basis of human regulatory evolution. *Am. J. Hum. Genet.* **91**, 455–465 (2012).

7. T. M. Preuss, Human brain evolution: From gene discovery to phenotype discovery. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 10709–10716 (2012).

8. S. R. Y. Cajal, The Croonian lecture: La fine structure des centres nerveux. *Proc. R. Soc. Lond.* **55**, 444–468 (1894).

9. G. F. Striedter, *Principles of Brain Evolution* (Sinauer Associates Incorporated, 2005).

10. J. M. Allman, K. K. Watson, N. A. Tetreault, A. Y. Hakeem, Intuition and autism: A possible role for Von Economo neurons. *Trends Cogn. Sci.* **9**, 367–373 (2005).

11. G. Rizzolatti, M. Fabbri-Destro, L. Cattaneo, Mirror neurons and their clinical relevance. *Nat. Clin. Pract. Neurol.* **5**, 24–34 (2009).

12. E. Boldog, T. E. Bakken, R. D. Hodge, M. Novotny, B. D. Aevermann, J. Baka, S. Bordé, J. L. Close, F. Diez-Fuertes, S.-L. Ding, N. Faragó, Á. K. Kocsis, B. Kovács, Z. Maltzer, J. M. McCorrison, J. A. Miller, G. Molnár, G. Oláh, A. Ozsvár, M. Rózsa, S. I. Shehata, K. A. Smith, S. M. Sunkin, D. N. Tran, P. Venepally, A. Wall, L. G. Puskás, P. Barzó, F. J. Steemers, N. J. Schork, R. H. Scheuermann, R. S. Lasken, E. S. Lein, G. Tamás, Transcriptomic and morphophysiological evidence for a specialized human cortical GABAergic cell type. *Nat. Neurosci.* **21**, 1185–1195 (2018).

13. T. J. Nowakowski, A. Bhaduri, A. A. Pollen, B. Alvarado, M. A. Mostajo-Radji, E. Di Lullo, M. Haeussler, C. Sandoval-Espinosa, S. J. Liu, D. Velmeshev, J. R. Ounadjela, J. Shuga, X. Wang, D. A. Lim, J. A. West, A. A. Leyrat, W. J. Kent, A. R. Kriegstein, Spatiotemporal gene expression trajectories reveal developmental hierarchies of the human cortex. *Science* **358**, 1318–1323 (2017).

14. R. D. Hodge, T. E. Bakken, J. A. Miller, K. A. Smith, E. R. Barkan, L. T. Graybuck, J. L. Close, B. Long, N. Johansen, O. Penn, Z. Yao, J. Eggermont, T. Höllt, B. P. Levi, S. I. Shehata, B. Aevermann, A. Beller, D. Bertagnolli, K. Brouner, T. Casper, C. Cobbs, R. Dalley, N. Dee, S.-L. Ding, R. G. Ellenbogen, O. Fong, E. Garren, J. Goldy, R. P. Gwinn, D. Hirschstein, C. D. Keene, M. Keshk, A. L. Ko, K. Lathia, A. Mahfouz, Z. Maltzer, M. McGraw, T. N. Nguyen, J. Nyhus, J. G. Ojemann, A. Oldre, S. Parry, S. Reynolds, C. Rimorin, V. Shapovalova, S. Somasundaram, A. Szafer, E. R. Thomsen, M. Tieu, G. Quon, R. H. Scheuermann, R. Yuste, S. M. Sunkin, B. Lelieveldt, D. Feng, L. Ng, A. Bernard, M. Hawrylycz, J. W. Phillips, B. Tasic, H. Zeng, A. R. Jones, C. Koch, E. S. Lein, Conserved cell types with divergent features in human versus mouse cortex. *Nature* **573**, 61–68 (2019).

15. T. E. Bakken, N. L. Jorstad, Q. Hu, B. B. Lake, W. Tian, B. E. Kalmbach, M. Crow, R. D. Hodge, F. M. Krienen, S. A. Sorensen, J. Eggermont, Z. Yao, B. D. Aevermann, A. I. Aldridge, A. Bartlett, D. Bertagnolli, T. Casper, R. G. Castanon, K. Crichton, T. L. Daigle, R. Dalley, N. Dee, N. Dembrow, D. Diep, S.-L. Ding, W. Dong, R. Fang, S. Fischer, M. Goldman, J. Goldy, L. T. Graybuck, B. R. Herb, X. Hou, J. Kancherla, M. Kroll, K. Lathia, B. van Lew, Y. E. Li, C. S. Liu, H. Liu, J. D. Lucero, A. Mahurkar, D. McMillen, J. A. Miller, M. Moussa, J. R. Nery, P. R. Nicovich, S.-Y. Niu, J. Orvis, J. K. Osteen, S. Owen, C. R. Palmer, T. Pham, N. Plongthongkum, O. Poirion, N. M. Reed, C. Rimorin, A. Rivkin, W. J. Romanow, A. E. Sedeño-Cortés, K. Siletti, S. Somasundaram, J. Sulc, M. Tieu, A. Torkelson, H. Tung, X. Wang, F. Xie, A. M. Yanny, R. Zhang, S. A. Ament, M. M. Behrens, H. C. Bravo, J. Chun, A. Dobin, J. Gillis, R. Hertzano, P. R. Hof, T. Höllt, G. D. Horwitz, C. D. Keene, P. V. Kharchenko, A. L. Ko, B. P. Lelieveldt, C. Luo, E. A. Mukamel, A. Pinto-Duarte, S. Preissl, A. Regev, B. Ren, R. H. Scheuermann, K. Smith, W. J. Spain, O. R. White, C. Koch, M. Hawrylycz, B. Tasic, E. Z. Macosko, S. A. McCarroll, J. T. Ting, H. Zeng, K. Zhang, G. Feng, J. R. Ecker, S. Linnarsson, E. S. Lein, Comparative cellular analysis of motor cortex in human, marmoset and mouse. *Nature* **598**, 111–119 (2021).

16. J. R. Ecker, D. H. Geschwind, A. R. Kriegstein, J. Ngai, P. Osten, D. Polioudakis, A. Regev, N. Sestan, I. R. Wickersham, H. Zeng, The BRAIN Initiative Cell Census Consortium: Lessons learned toward generating a comprehensive brain cell atlas. *Neuron* **96**, 542–557 (2017).

17. Rhesus Macaque Genome Sequencing and Analysis Consortium, R. A. Gibbs, J. Rogers, M. G. Katze, R. Bumgarner, G. M. Weinstock, E. R. Mardis, K. A. Remington, R. L. Strausberg, J. C. Venter, R. K. Wilson, M. A. Batzer, C. D. Bustamante, E. E. Eichler, M. W. Hahn, R. C. Hardison, K. D. Makova, W. Miller, A. Milosavljevic, R. E. Palermo, A. Siepel, J. M. Sikela, T. Attaway, S. Bell, K. E. Bernard, C. J. Buhay, M. N. Chandrabose, M. Dao, C. Davis, K. D. Delehaunty, Y. Ding, H. H. Dinh, S. Dugan-Rocha, L. A. Fulton, R. A. Gabisi, T. T. Garner, J. Godfrey, A. C. Hawes, J. Hernandez, S. Hines, M. Holder, J. Hume, S. N. Jhangiani, V. Joshi, Z. M. Khan, E. F. Kirkness, A. Cree, R. G. Fowler, S. Lee, L. R. Lewis, Z. Li, Y.-S. Liu, S. M. Moore, D. Muzny, L. V. Nazareth, D. N. Ngo, G. O. Okwuonu, G. Pai, D. Parker, H. A. Paul, C. Pfannkoch, C. S. Pohl, Y.-H. Rogers, S. J. Ruiz, A. Sabo, J. Santibanez, B. W. Schneider, S. M. Smith, E. Sodergren, A. F. Svatek, T. R. Utterback, S. Vattathil, W. Warren, C. S. White, A. T. Chinwalla, Y. Feng, A. L. Halpern, L. W. Hillier, X. Huang, P. Minx, J. O. Nelson, K. H. Pepin, X. Qin, G. G. Sutton, E. Venter, B. P. Walenz, J. W. Wallis, K. C. Worley, S.-P. Yang,

S. M. Jones, M. A. Marra, M. Rocchi, J. E. Schein, R. Baertsch, L. Clarke, M. Csürös, J. Glasscock, R. A. Harris, P. Havlak, A. R. Jackson, H. Jiang, Y. Liu, D. N. Messina, Y. Shen, H. X.-Z. Song, T. Wylie, L. Zhang, E. Birney, K. Han, M. K. Konkel, J. Lee, A. F. A. Smit, B. Ullmer, H. Wang, J. Xing, R. Burhans, Z. Cheng, J. E. Karro, J. Ma, B. Raney, X. She, M. J. Cox, J. P. Demuth, L. J. Dumas, S.-G. Han, J. Hopkins, A. Karimpour-Fard, Y. H. Kim, J. R. Pollack, T. Vinar, C. Addo-Quaye, J. Degenhardt, A. Denby, M. J. Hubisz, A. Indap, C. Kosiol, B. T. Lahn, H. A. Lawson, A. Marklein, R. Nielsen, E. J. Vallender, A. G. Clark, B. Ferguson, R. D. Hernandez, K. Hirani, H. Kehrer-Sawatzki, J. Kolb, S. Patil, L.-L. Pu, Y. Ren, D. G. Smith, D. A. Wheeler, I. Schenck, E. V. Ball, R. Chen, D. N. Cooper, B. Giardine, F. Hsu, W. J. Kent, A. Lesk, D. L. Nelson, W. E. O'brien, K. Prüfer, P. D. Stenson, J. C. Wallace, H. Ke, X.-M. Liu, P. Wang, A. P. Xiang, F. Yang, G. P. Barber, D. Haussler, D. Karolchik, A. D. Kern, R. M. Kuhn, K. E. Smith, A. S. Zwieg, Evolutionary and biomedical insights from the rhesus macaque genome. *Science* **316**, 222–234 (2007).

18. J. Cao, J. S. Packer, V. Ramani, D. A. Cusanovich, C. Huynh, R. Daza, X. Qiu, C. Lee, S. N. Furlan, F. J. Steemers, A. Adey, R. H. Waterston, C. Trapnell, J. Shendure, Comprehensive single-cell transcriptional profiling of a multicellular organism. *Science* **357**, 661–667 (2017).

19. J. Cao, M. Spielmann, X. Qiu, X. Huang, D. M. Ibrahim, A. J. Hill, F. Zhang, S. Mundlos, L. Christiansen, F. J. Steemers, C. Trapnell, J. Shendure, The single-cell transcriptional landscape of mammalian organogenesis. *Nature* **566**, 496–502 (2019).

20. J. Cao, J. Shendure, sci-RNA-seq3. *protocols.io* (2020), doi:10.17504/protocols.io.9yih7ue.

21. B. K. Martin, C. Qiu, E. Nichols, M. Phung, R. Green-Gladden, S. Srivatsan, R. Blecher-Gonen, B. J. Beliveau, C. Trapnell, J. Cao, J. Shendure, Optimized single-nucleus transcriptional profiling by combinatorial indexing. *Nat. Protoc.* **18**, 188–207 (2023).

22. T. Kamath, A. Abdulraouf, S. J. Burris, J. Langlieb, V. Gazestani, N. M. Nadaf, K. Balderrama, C. Vanderburg, E. Z. Macosko, Single-cell genomic profiling of human dopamine neurons identifies a population that selectively degenerates in Parkinson's disease. *Nat. Neurosci.* **25**, 588–595 (2022).

23. J. Ren, A. Isakova, D. Friedmann, J. Zeng, S. M. Grutzner, A. Pun, G. Q. Zhao, S. S. Kolluru, R. Wang, R. Lin, P. Li, A. Li, J. L. Raymond, Q. Luo, M. Luo, S. R. Quake, L. Luo, Single-cell transcriptomes and whole-brain projections of serotonin neurons in the mouse dorsal and median raphe nuclei. *eLife* **8**, e49424 (2019).

24. W. Poewe, K. Seppi, C. M. Tanner, G. M. Halliday, P. Brundin, J. Volkmann, A.-E. Schrag, A. E. Lang, Parkinson disease. *Nat. Rev. Dis. Primers.* **3**, 17013 (2017).

25. M. Fakhoury, Revisiting the serotonin hypothesis: Implications for major depressive disorders. *Mol. Neurobiol.* **53**, 2778–2786 (2016).

26. A. C. Yang, R. T. Vest, F. Kern, D. P. Lee, M. Agam, C. A. Maat, P. M. Losada, M. B. Chen, N. Schaum, N. Khoury, A. Toland, K. Calcuttawala, H. Shin, R. Pálovics, A. Shin, E. Y. Wang, J. Luo, D. Gate, W. J. Schulz-Schaeffer, P. Chu, J. A. Siegenthaler, M. W. McNerney, A. Keller, T. Wyss-Coray, A human brain vascular atlas reveals diverse mediators of Alzheimer's risk. *Nature* **603**, 885–892 (2022).

27. Z.-Z. Hao, J.-R. Wei, D. Xiao, R. Liu, N. Xu, L. Tang, M. Huang, Y. Shen, C. Xing, W. Huang, X. Liu, M. Xiang, Y. Liu, Z. Miao, S. Liu, Single-cell transcriptomics of adult macaque hippocampus reveals neural precursor cell populations. *Nat. Neurosci.* **25**, 805–817 (2022).

28. D. A. Cusanovich, A. J. Hill, D. Aghamirzaie, R. M. Daza, H. A. Pliner, J. B. Berletch, G. N. Filippova, X. Huang, L. Christiansen, W. S. DeWitt, C. Lee, S. G. Regalado, D. F. Read, F. J. Steemers, C. M. Disteche, C. Trapnell, J. Shendure, A single-cell atlas of *in vivo* mammalian chromatin accessibility. *Cell* **174**, 1309–1324.e18 (2018).

29. Y. E. Li, S. Preissl, X. Hou, Z. Zhang, K. Zhang, Y. Qiu, O. B. Poirion, B. Li, J. Chiou, H. Liu, A. Pinto-Duarte, N. Kubo, X. Yang, R. Fang, X. Wang, J.-Y. Han, J. Lucero, Y. Yan, M. Miller, S. Kuan, D. Gorkin, K. J. Gaulton, Y. Shen, M. Nunn, E. A. Mukamel, M. M. Behrens, J. R. Ecker, B. Ren, An atlas of gene regulatory elements in adult mouse cerebrum. *Nature* **598**, 129–136 (2021).

30. J. Berg, S. A. Sorensen, J. T. Ting, J. A. Miller, T. Chartrand, A. Buchin, T. E. Bakken, A. Budzillo, N. Dee, S.-L. Ding, N. W. Gouwens, R. D. Hodge, B. Kalmbach, C. Lee, B. R. Lee, L. Alfiler, K. Baker, E. Barkan, A. Beller, K. Berry, D. Bertagnolli, K. Bickley, J. Bomben, T. Braun, K. Brouner, T. Casper, P. Chong, K. Crichton, R. Dalley, R. de Frates, T. Desta, S. D. Lee, F. D'Orazi, N. Dotson, T. Egdorf, R. Enstrom, C. Farrell, D. Feng, O. Fong, S. Furdan, A. A. Galakhova, C. Gamlin, A. Gary, A. Glandon, J. Goldy, M. Gorham, N. A. Gorioounova, S. Gratiy, L. Graybuck, H. Gu, K. Hadley, N. Hansen, T. S. Heistek, A. M. Henry, D. B. Heyer, D. Hill, C. Hill, M. Hupp, T. Jarsky, S. Kebede, L. Keene, L. Kim, M.-H. Kim, M. Kroll, C. Latimer, B. P. Levi, K. E. Link, M. Mallory, R. Mann, D. Marshall, M. Maxwell, M. McGraw, D. McMillen, E. Melief, E. J. Mertens, L. Mezei, N. Mihut, S. Mok, G. Molnar, A. Mukora, L. Ng, K. Ngo, P. R. Nicovich, J. Nyhus, G. Olah, A. Oldre, V. Omstead, A. Ozsvar, D. Park, H. Peng, T. Pham, C. A. Pom, L. Potekhina, R. Rajanbabu, S. Ransford, D. Reid, C. Rimorin, A. Ruiz, D. Sandman, J. Sulc, S. M. Sunkin, A. Szafer, V. Szemenyei, E. R. Thomsen, M. Tieu, A. Torkelson, J. Trinh, H. Tung, W. Wakeman, F. Waleboer, K. Ward, R. Wilbers, G. Williams, Z. Yao, J.-G. Yoon, C. Anastassiou, A. Arkhipov, P. Barzo, A. Bernard, C. Cobbs, P. C. de Witt Hamer, R. G. Ellenbogen, L. Esposito, M. Ferreira, R. P. Gwinn, M. J. Hawrylycz, P. R. Hof, S. Idema, A. R. Jones, C. D. Keene, A. L. Ko, G. J. Murphy, L. Ng, J. G. Ojemann, A. P. Patel, J. W. Phillips,

D. L. Silbergeld, K. Smith, B. Tasic, R. Yuste, I. Segev, C. P. J. de Kock, H. D. Mansvelder, G. Tamas, H. Zeng, C. Koch, E. S. Lein, Human neocortical expansion involves glutamatergic neuron diversification. *Nature* **598**, 151–158 (2021).

31. A. E. Trevino, F. Müller, J. Andersen, L. Sundaram, A. Kathiria, A. Shcherbina, K. Farh, H. Y. Chang, A. M. Pașca, A. Kundaje, S. P. Pașca, W. J. Greenleaf, Chromatin and gene-regulatory dynamics of the developing human cerebral cortex at single-cell resolution. *Cell* **184**, 5053–5069.e23 (2021).

32. F. M. Krienen, M. Goldman, Q. Zhang, R. C. H. Del Rosario, M. Florio, R. Machold, A. Saunders, K. Levandowski, H. Zaniewski, B. Schuman, C. Wu, A. Lutservitz, C. D. Mullally, N. Reed, E. Bien, L. Bortolin, M. Fernandez-Otero, J. D. Lin, A. Wysoker, J. Nemesh, D. Kulp, M. Burns, V. Tkachev, R. Smith, C. A. Walsh, J. Dimidschstein, B. Rudy, L. S. Kean, S. Berretta, G. Fishell, G. Feng, S. A. McCarroll, Innovations present in the primate interneuron repertoire. *Nature* **586**, 262–269 (2020).

33. B. A. Barres, The mystery and magic of glia: A perspective on their roles in health and disease. *Neuron* **60**, 430–440 (2008).

34. C. F. Valori, G. Guidotti, L. Brambilla, D. Rossi, Astrocytes: Emerging therapeutic targets in neurological disorders. *Trends Mol. Med.* **25**, 750–759 (2019).

35. S. Wiebe, A. Nagpal, V. T. Truong, J. Park, A. Skalecka, A. J. He, K. Gamache, A. Khoutorsky, I. Gantois, N. Sonenberg, Inhibitory interneurons mediate autism-associated behaviors via 4E-BP2. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 18060–18067 (2019).

36. G. Miyoshi, Y. Ueta, A. Natsubori, K. Hiraga, H. Osaki, Y. Yagasaki, Y. Kishi, Y. Yanagawa, G. Fishell, R. P. Machold, M. Miyata, *FoxG1* regulates the formation of cortical GABAergic circuit during an early postnatal critical period resulting in autism spectrum disorder-like phenotypes. *Nat. Commun.* **12**, 3773 (2021).

37. T. E. Bakken, C. T. van Velthoven, V. Menon, R. D. Hodge, Z. Yao, T. N. Nguyen, L. T. Graybuck, G. D. Horwitz, D. Bertagnolli, J. Goldy, A. M. Yanny, E. Garren, S. Parry, T. Casper, S. I. Shehata, E. R. Barkan, A. Szafer, B. P. Levi, N. Dee, K. A. Smith, S. M. Sunkin, A. Bernard, J. Phillips, M. J. Hawrylycz, C. Koch, G. J. Murphy, E. Lein, H. Zeng, B. Tasic, Single-cell and single-nucleus RNA-seq uncovers shared and distinct axes of variation in dorsal LGN neurons in mice, non-human primates, and humans. *eLife* **10**, e64875 (2021).

38. Q. Zhang, C. Sano, A. Masuda, R. Ando, M. Tanaka, S. Itohara, Netrin-G1 regulates fear-like and anxiety-like behaviors in dissociable neural circuits. *Sci. Rep.* **6**, 28750 (2016).

39. H. Mathys, J. Davila-Velderrain, Z. Peng, F. Gao, S. Mohammadi, J. Z. Young, M. Menon, L. He, F. Abdurrob, X. Jiang, A. J. Martorell, R. M. Ransohoff, B. P. Hafler, D. A. Bennett, M. Kellis, L.-H. Tsai, Single-cell transcriptomic analysis of Alzheimer's disease. *Nature* **570**, 332–337 (2019).

40. M. Y. Batiuk, A. Martirosyan, J. Wahis, F. de Vin, C. Marneffe, C. Kusserow, J. Koeppen, J. F. Viana, J. F. Oliveira, T. Voet, C. P. Ponting, T. G. Belgard, M. G. Holt, Identification of region-specific astrocyte subtypes at single cell resolution. *Nat. Commun.* **11**, 1220 (2020).

41. X. Huang, J. Henck, C. Qiu, V. K. A. Sreenivasan, S. Balachandran, R. Behncke, W.-L. Chan, A. Despang, D. E. Dickel, N. Haag, R. Hägerling, N. Hansmeier, F. Hennig, C. Marshall, S. Rajderkar, A. Ringel, M. Robson, L. Saunders, S. R. Srivatsan, S. Ulferts, L. Wittler, Y. Zhu, V. M. Kalscheuer, D. Ibrahim, I. Kurth, U. Kornak, D. R. Beier, A. Visel, L. A. Pennacchio, C. Trapnell, J. Cao, J. Shendure, M. Spielmann, Single cell, whole embryo phenotyping of pleiotropic disorders of mammalian development. bioRxiv, 2022.08.03.500325 (2022). https://doi.org/10.1101/2022.08.03.500325.

42. D. Kirdajova, L. Valihrach, M. Valny, J. Kriska, D. Krocianova, S. Benesova, P. Abaffy, D. Zucha, R. Klassen, D. Kolenicova, P. Honsa, M. Kubista, M. Anderova, Transient astrocyte-like NG2 glia subpopulation emerges solely following permanent brain ischemia. *Glia* **69**, 2658–2681 (2021).

43. Q. Shu, N. J. Lennemann, S. N. Sarkar, Y. Sadovsky, C. B. Coyne, ADAP2 is an interferon stimulated gene that restricts RNA virus entry. *PLOS Pathog.* **11**, e1005150 (2015).

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

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

46. T. Stuart, A. Butler, P. Hoffman, C. Hafemeister, E. Papalexi, W. M. Mauck 3rd, Y. Hao, M. Stoeckius, P. Smibert, R. Satija, Comprehensive integration of single-cell data. *Cell* **177**, 1888–1902.e21 (2019).

47. Z.-J. Cao, G. Gao, Multi-omics single-cell data integration and regulatory inference with graph-linked embedding. *Nat. Biotechnol.* **40**, 1458–1466 (2022).

48. M. Ishibashi, S. L. Ang, K. Shiota, S. Nakanishi, R. Kageyama, F. Guillemot, Targeted disruption of mammalian hairy and Enhancer of split homolog-1 (HES-1) leads to up-regulation of neural helix-loop-helix factors, premature neurogenesis, and severe neural tube defects. *Genes Dev.* **9**, 3136–3148 (1995).

49. J. Collignon, S. Sockanathan, A. Hacker, M. Cohen-Tannoudji, D. Norris, S. Rastan, M. Stevanovic, P. N. Goodfellow, R. Lovell-Badge, A comparison of the properties of Sox-3 with Sry and two related genes,Sox-1 and Sox-2. *Development* **122**, 509–520 (1996).

50. N. Gaiano, J. S. Nye, G. Fishell, Radial glial identity is promoted by Notch1 signaling in the murine forebrain. *Neuron* **26**, 395–404 (2000).

51. V. Palma, A. R. i. Altaba, Hedgehog-GLI signaling regulates the behavior of cells with stem cell properties in the developing neocortex. *Development* **131**, 337–345 (2004).

52. Y. Kitamura, S. Shimohama, T. Ota, Y. Matsuoka, Y. Nomura, T. Taniguchi, Alteration of transcription factors NF-κB and STAT1 in Alzheimer's disease brains. *Neurosci. Lett.* **237**, 17–20 (1997).

53. B. A. Citron, J. S. Dennis, R. S. Zeitlin, V. Echeverria, Transcription factor Sp1 dysregulation in Alzheimer's disease. *J. Neurosci. Res.* **86**, 2499–2504 (2008).

54. P. C. Tiwari, R. Pal, The potential role of neuroinflammation and transcription factors in Parkinson disease. *Dialogues Clin. Neurosci.* **19**, 71–80 (2017).

55. PsychENCODE Consortium, S. Akbarian, C. Liu, J. A. Knowles, F. M. Vaccarino, P. J. Farnham, G. E. Crawford, A. E. Jaffe, D. Pinto, S. Dracheva, D. H. Geschwind, J. Mill, A. C. Nairn, A. Abyzov, S. Pochareddy, S. Prabhakar, S. Weissman, P. F. Sullivan, M. W. State, Z. Weng, M. A. Peters, K. P. White, M. B. Gerstein, A. Amiri, C. Armoskus, A. E. Ashley-Koch, T. Bae, A. Beckel-Mitchener, B. P. Berman, G. A. Coetzee, G. Coppola, N. Francoeur, M. Fromer, R. Gao, K. Grennan, J. Herstein, D. H. Kavanagh, N. A. Ivanov, Y. Jiang, R. R. Kitchen, A. Kozlenkov, M. Kundakovic, M. Li, Z. Li, S. Liu, L. M. Mangravite, E. Mattei, E. Markenscoff-Papadimitriou, F. C. P. Navarro, N. North, L. Omberg, D. Panchision, N. Parikshak, J. Poschmann, A. J. Price, M. Purcaro, T. E. Reddy, P. Roussos, S. Schreiner, S. Scuderi, R. Sebra, M. Shibata, A. W. Shieh, M. Skarica, W. Sun, V. Swarup, A. Thomas, J. Tsuji, H. van Bakel, D. Wang, Y. Wang, K. Wang, D. M. Werling, A. J. Willsey, H. Witt, H. Won, C. C. Y. Wong, G. A. Wray, E. Y. Wu, X. Xu, L. Yao, G. Senthil, T. Lehner, P. Sklar, N. Sestan, The PsychENCODE project. *Nat. Neurosci.* **18**, 1707–1712 (2015).

56. K. E. Prater, K. J. Green, W. Sun, C. L. Smith, K. L. Chiou, L. Heath, S. Rose, C. D. Keene, R. Y. Kwon, N. Snyder-Mackler, E. E. Blue, J. E. Young, A. Shojaie, B. Logsdon, G. A. Garden, S. Jayadev, Human microglia show unique transcriptional changes in Alzheimer's disease. *Nat. Aging.* **3**, 894–907 (2023).

57. E. Gerrits, L. A. A. Giannini, N. Brouwer, S. Melhem, D. Seilhean, I. Le Ber; Brainbank Neuro-CEB Neuropathology Network, A. Kamermans, G. Kooij, H. E. de Vries, E. W. G. M. Boddeke, H. Seelaar, J. C. van Swieten, B. J. L. Eggen, Neurovascular dysfunction in GRN-associated frontotemporal dementia identified by single-nucleus RNA sequencing of human cerebral cortex. *Nat. Neurosci.* **25**, 1034–1048 (2022).

58. A. Pieper, S. Rudolph, G. L. Wieser, T. Götze, H. Mießner, T. Yonemasu, K. Yan, I. Tzvetanova, B. D. Castillo, U. Bode, I. Bormuth, J. I. Wadiche, M. H. Schwab, S. Goebbels, NeuroD2 controls inhibitory circuit formation in the molecular layer of the cerebellum. *Sci. Rep.* **9**, 1448 (2019).

59. S. Pletscher-Frankild, A. Pallejà, K. Tsafou, J. X. Binder, L. J. Jensen, DISEASES: Text mining and data integration of disease-gene associations. *Methods* **74**, 83–89 (2015).

60. A. Nguyen, T. A. Rauch, G. P. Pfeifer, V. W. Hu, Global methylation profiling of lymphoblastoid cell lines reveals epigenetic contributions to autism spectrum disorders and a novel autism candidate gene, RORA, whose protein product is reduced in autistic brain. *FASEB J.* **24**, 3036–3051 (2010).

61. A. Sayad, R. Noroozi, M. D. Omrani, M. Taheri, S. Ghafouri-Fard, Retinoic acid-related orphan receptor alpha (RORA) variants are associated with autism spectrum disorder. *Metab. Brain Dis.* **32**, 1595–1601 (2017).

62. X. Liu, D. Han, M. Somel, X. Jiang, H. Hu, P. Guijarro, N. Zhang, A. Mitchell, T. Halene, J. J. Ely, C. C. Sherwood, P. R. Hof, Z. Qiu, S. Pääbo, S. Akbarian, P. Khaitovich, Disruption of an evolutionarily novel synaptic expression pattern in autism. *PLOS Biol.* **14**, e1002558 (2016).

63. T. Hirayama, E. Yuuki, E. Tarusawa, S. Saito, N. Nakayama, N. Hoshino, S. Nakama, T. Fukuishi, Y. Kawanishi, H. Umeshima, K. Tomita, Y. Yoshimura, N. Galjart, K. Hashimoto, N. Ohno, T. Yagi, CTCF loss induces giant lamellar bodies in Purkinje cell dendrites. *Acta Neuropathol. Commun.* **10**, 172 (2022).

64. J. Chang, S. R. Gilman, A. H. Chiang, S. J. Sanders, D. Vitkup, Genotype to phenotype relationships in autism spectrum disorders. *Nat. Neurosci.* **18**, 191–198 (2015).

65. D. Calderon, R. Blecher-Gonen, X. Huang, S. Secchia, J. Kentro, R. M. Daza, B. Martin, A. Dulja, C. Schaub, C. Trapnell, E. Larschan, K. M. O'Connor-Giles, E. E. M. Furlong, J. Shendure, The continuum of *Drosophila* embryonic development at single-cell resolution. *Science* **377**, eabn5800 (2022).

66. A. G. Efthymiou, A. M. Goate, Late onset Alzheimer's disease genetics implicates microglial pathways in disease risk. *Mol. Neurodegener.* **12**, 43 (2017).

67. R. Fang, S. Preissl, Y. Li, X. Hou, J. Lucero, X. Wang, A. Motamedi, A. K. Shiau, X. Zhou, F. Xie, E. A. Mukamel, K. Zhang, Y. Zhang, M. M. Behrens, J. R. Ecker, B. Ren, Comprehensive analysis of single cell ATAC-seq data with SnapATAC. *Nat. Commun.* **12**, 1337 (2021).

68. S. K. Leung, A. R. Jeffries, I. Castanho, B. T. Jordan, K. Moore, J. P. Davies, E. L. Dempster, N. J. Bray, P. O'Neill, E. Tseng, Z. Ahmed, D. A. Collier, E. D. Jeffery, S. Prabhakar,

L. Schalkwyk, C. Jops, M. J. Gandal, G. M. Sheynkman, E. Hannon, J. Mill, Full-length transcript sequencing of human and mouse cerebral cortex identifies widespread isoform diversity and alternative splicing. *Cell Rep.* **37**, 110022 (2021).

69. J. M. Boggs, Myelin basic protein: A multifunctional protein. *Cell. Mol. Life Sci.* **63**, 1945–1961 (2006).

70. G. Harauz, V. Ladizhansky, J. M. Boggs, Structural polymorphism and multifunctionality of myelin basic protein. *Biochemistry* **48**, 8094–8104 (2009).

71. G. Harauz, J. M. Boggs, Myelin management by the 18.5-kDa and 21.5-kDa classic myelin basic protein isoforms. *J. Neurochem.* **125**, 334–361 (2013).

72. K. C. Dietz, J. J. Polanco, S. U. Pol, F. J. Sim, Targeting human oligodendrocyte progenitors for myelin repair. *Exp. Neurol.* **283**, 489–500 (2016).

73. R. N. Doan, B.-I. Bae, B. Cubelos, C. Chang, A. A. Hossain, S. Al-Saad, N. M. Mukaddes, O. Oner, M. Al-Saffar, S. Balkhy, G. G. Gascon; Homozygosity Mapping Consortium for Autism, M. Nieto, C. A. Walsh, Mutations in human accelerated regions disrupt cognition and social behavior. *Cell* **167**, 341–354.e12 (2016).

74. R. J. Mangan, F. C. Alsina, F. Mosti, J. E. Sotelo-Fonseca, D. A. Snellings, E. H. Au, J. Carvalho, L. Sathyan, G. D. Johnson, T. E. Reddy, D. L. Silver, C. B. Lowe, Adaptive sequence diver-gence forged new neurodevelopmental enhancers in humans. *Cell* **185**, 4587–4603.e23 (2022).

75. S. Kanton, M. J. Boyle, Z. He, M. Santel, A. Weigert, F. Sanchís-Calleja, P. Guijarro, L. Sidow, J. S. Fleck, D. Han, Z. Qian, M. Heide, W. B. Huttner, P. Khaitovich, S. Pääbo, B. Treutlein, J. G. Camp, Organoid single-cell genomic atlas uncovers human-specific features of brain development. *Nature* **574**, 418–422 (2019).

76. H. Suresh, M. Crow, N. Jorstad, R. Hodge, E. Lein, A. Dobin, T. Bakken, J. Gillis, Conserved coexpression at single cell resolution across primate brains. bioRxiv, 2022.09.20.508736 (2022). https://doi.org/10.1101/2022.09.20.508736.

77. B. K. Bulik-Sullivan, P.-R. Loh, H. K. Finucane, S. Ripke, J. Yang; Schizophrenia Working Group of the Psychiatric Genomics Consortium, N. Patterson, M. J. Daly, A. L. Price, B. M. Neale, LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).

78. H. K. Finucane, B. Bulik-Sullivan, A. Gusev, G. Trynka, Y. Reshef, P.-R. Loh, V. Anttila, H. Xu, C. Zang, K. Farh, S. Ripke, F. R. Day; ReproGen Consortium; Schizophrenia Working Group of the Psychiatric Genomics Consortium; RACI Consortium, S. Purcell, E. Stahl, S. Lindstrom, J. R. B. Perry, Y. Okada, S. Raychaudhuri, M. J. Daly, N. Patterson, B. M. Neale, A. L. Price, Partitioning heritability by functional annotation using genome-wide asso-ciation summary statistics. *Nat. Genet.* **47**, 1228–1235 (2015).

79. C. Yang, K. E. Hawkins, S. Doré, E. Candelario-Jalil, Neuroinflammatory mechanisms of blood-brain barrier damage in ischemic stroke. *Am. J. Physiol. Cell Physiol.* **316**, C135–C153 (2019).

80. C. J. Bohlen, B. A. Friedman, B. Dejanovic, M. Sheng, Microglia in brain development, homeostasis, and neurodegeneration. *Annu. Rev. Genet.* **53**, 263–288 (2019).

81. X. Jiang, M. Lachance, E. Rossignol, Involvement of cortical fast-spiking parvalbumin-positive basket cells in epilepsy. *Prog. Brain Res.* **226**, 81–126 (2016).

82. E. Kamma, W. Lasisi, C. Libner, H. S. Ng, J. R. Plemel, Central nervous system macrophages in progressive multiple sclerosis: Relationship to neurodegeneration and therapeutics. *J. Neuroinflammation* **19**, 45 (2022).

83. S. Voet, M. Prinz, G. van Loo, Microglia in central nervous system inflammation and multiple sclerosis pathology. *Trends Mol. Med.* **25**, 112–123 (2019).

84. A. V. Domingues, I. M. Pereira, H. Vilaça-Faria, A. J. Salgado, A. J. Rodrigues, F. G. Teixeira, Glial cells in Parkinson´s disease: Protective or deleterious? *Cell. Mol. Life Sci.* **77**, 5171–5188 (2020).

85. L. Iovino, M. E. Tremblay, L. Civiero, Glutamate-induced excitotoxicity in Parkinson's disease: The role of glial cells. *J. Pharmacol. Sci.* **144**, 151–164 (2020).

86. D. Demontis, R. K. Walters, J. Martin, M. Mattheisen, T. D. Als, E. Agerbo, G. Baldursson, R. Belliveau, J. Bybjerg-Grauholm, M. Bækvad-Hansen, F. Cerrato, K. Chambert, C. Churchhouse, A. Dumont, N. Eriksson, M. Gandal, J. I. Goldstein, K. L. Grasby, J. Grove, O. O. Gudmundsson, C. S. Hansen, M. E. Hauberg, M. V. Hollegaard, D. P. Howrigan, H. Huang, J. B. Maller, A. R. Martin, N. G. Martin, J. Moran, J. Pallesen, D. S. Palmer, C. B. Pedersen, M. G. Pedersen, T. Poterba, J. B. Poulsen, S. Ripke, E. B. Robinson, F. K. Satterstrom, H. Stefansson, C. Stevens, P. Turley, G. B. Walters, H. Won, M. J. Wright; ADHD Working Group of the Psychiatric Genomics Consortium (PGC); Early Lifecourse & Genetic Epidemiology (EAGLE) Consortium; 23andMe Research Team, O. A. Andreassen, P. Asherson, C. L. Burton, D. I. Boomsma, B. Cormand, S. Dalsgaard, B. Franke, J. Gelernter, D. Geschwind, H. Hakonarson, J. Haavik, H. R. Kranzler, J. Kuntsi, K. Langley, K.-P. Lesch, C. Middeldorp, A. Reif, L. A. Rohde, P. Roussos, R. Schachar, P. Sklar, E. J. S. Sonuga-Barke, P. F. Sullivan, A. Thapar, J. Y. Tung, I. D. Waldman, S. E. Medland, K. Stefansson, M. Nordentoft, D. M. Hougaard, T. Werge, O. Mors, P. B. Mortensen, M. J. Daly, S. V. Faraone, A. D. Børglum, B. M. Neale, Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. *Nat. Genet.* **51**, 63–75 (2019).

87. J. Nagai, A. K. Rajbhandari, M. R. Gangwani, A. Hachisuka, G. Coppola, S. C. Masmanidis, M. S. Fanselow, B. S. Khakh, Hyperactivity with disrupted attention by activation of an astrocyte synaptogenic cue. *Cell* **177**, 1280–1292.e20 (2019).

88. Y. Qian, J. Li, S. Zhao, E. Matthews, M. Adoff, W. Zhong, X. An, M. Yeo, C. Park, X. Yang, B.-S. Wang, D. Southwell, Z. J. Huang, Programmable RNA sensing for cell monitoring and manipulation. *Nature* **610**, 713–721 (2022).

89. M. J. Kessler, R. G. Rawlins, A 75-year pictorial history of the Cayo Santiago rhesus monkey colony. *Am. J. Primatol.* **78**, 6–43 (2016).

90. A. Widdig, L. Muniz, M. Minkner, Y. Barth, S. Bley, A. Ruiz-Lambides, O. Junge, R. Mundry, L. Kulik, Low incidence of inbreeding in a long-lived primate population isolated for 75 years. *Behav. Ecol. Sociobiol.* **71**, 18 (2017).

91. R. Hernandez-Pacheco, D. L. Delgado, R. G. Rawlins, M. J. Kessler, A. V. Ruiz-Lambides, E. Maldonado, A. M. Sabat, Managing the Cayo Santiago rhesus macaque population: The role of density. *Am. J. Primatol.* **78**, 167–181 (2016).

92. C. Testard, L. J. N. Brent, J. Andersson, K. L. Chiou, J. E. N.-D. Valle, A. R. DeCasien, A. Acevedo-Ithier, M. K. Stock, S. C. Antón, O. Gonzalez, C. S. Walker, S. Foxley, N. R. Compo, S. Bauman, A. V. Ruiz-Lambides, M. I. Martinez, J. H. P. Skene, S. E. Horvath, C. B. R. Unit, J. P. Higham, K. L. Miller, N. Snyder-Mackler, M. J. Montague, M. L. Platt, J. Sallet, Social connections predict brain structure in a multidimensional free-ranging primate society. *Sci. Adv.* **8**, eabl5794 (2022).

93. K. L. Chiou, A. R. DeCasien, K. P. Rees, C. Testard, C. H. Spurrell, A. A. Gogate, H. A. Pliner, S. Tremblay, A. Mercer, C. J. Whalen, J. E. N.-D. Valle, M. C. Janiak, S. E. B. Surratt, O. González, N. R. Compo, M. K. Stock, A. V. Ruiz-Lambides, M. I. Martínez; Cayo Biobank Research Unit, M. A. Wilson, A. D. Melin, S. C. Antón, C. S. Walker, J. Sallet, J. M. Newbern, L. M. Starita, J. Shendure, J. P. Higham, L. J. N. Brent, M. J. Montague, M. L. Platt, N. Snyder-Mackler, Multiregion transcriptomic profiling of the primate brain reveals signatures of aging and the social environment. *Nat. Neurosci.* **25**, 1714–1723 (2022).

94. K. L. Chiou, M. J. Montague, E. A. Goldman, M. M. Watowich, S. N. Sams, J. Song, J. E. Horvath, K. N. Sterner, A. V. Ruiz-Lambides, M. I. Martínez, J. P. Higham, L. J. N. Brent, M. L. Platt, N. Snyder-Mackler, Rhesus macaques as a tractable physiological model of human ageing. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **375**, 20190612 (2020).

95. R. Desimone, L. G. Ungerleider, Multiple visual areas in the caudal superior temporal sulcus of the macaque. *J. Comp. Neurol.* **248**, 164–189 (1986).

96. R. T. Born, D. C. Bradley, Structure and function of visual area MT. *Annu. Rev. Neurosci.* **28**, 157–189 (2005).

97. S. Domcke, A. J. Hill, R. M. Daza, C. Trapnell, D. A. Cusanovich, J. Shendure, sci-ATAC-seq3. *protocols.io* (2020), doi:10.17504/protocols.io.be8mjhu6.

98. M. R. Corces, A. E. Trevino, E. G. Hamilton, P. G. Greenside, N. A. Sinnott-Armstrong, S. Vesuna, A. T. Satpathy, A. J. Rubin, K. S. Montine, B. Wu, A. Kathiria, S. W. Cho, M. R. Mumbach, A. C. Carter, M. Kasowski, L. A. Orloff, V. I. Risca, A. Kundaje, P. A. Khavari, T. J. Montine, W. J. Greenleaf, H. Y. Chang, An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nat. Methods* **14**, 959–962 (2017).

99. F. Krueger, F. James, P. Ewels, E. Afyounian, B. Schuster-Boeckler, *TrimGalore* (2021; https://zenodo.org/record/5127899).

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

101. W. C. Warren, R. A. Harris, R. A. Haukness, I. T. Fiddes, S. C. Murali, J. Fernandes, P. C. Dishuck, J. M. Storer, M. Raveendran, L. W. Hillier, D. Porubsky, Y. Mao, D. Gordon, M. R. Vollger, A. P. Lewis, K. M. Munson, E. DeVogelaere, J. Armstrong, M. Diekhans, J. A. Walker, C. Tomlinson, T. A. Graves-Lindsay, M. Kremitzki, S. R. Salama, P. A. Audano, M. Escalona, N. W. Maurer, F. Antonacci, L. Mercuri, F. A. M. Maggiolini, C. R. Catacchio, J. G. Underwood, D. H. O'Connor, A. D. Sanders, J. O. Korbel, B. Ferguson, H. M. Kubisch, L. Picker, N. H. Kalin, D. Rosene, J. Levine, D. H. Abbott, S. B. Gray, M. M. Sanchez, Z. A. Kovacs-Balint, J. W. Kemnitz, S. M. Thomasy, J. A. Roberts, E. L. Kinnally, J. P. Capitanio, J. H. P. Skene, M. Platt, S. A. Cole, R. E. Green, M. Ventura, R. W. Wiseman, B. Paten, M. A. Batzer, J. Rogers, E. E. Eichler, Sequence diversity analyses of an improved rhesus macaque genome enhance its biomedical utility. *Science* **370**, eabc6617 (2020).

102. I. Virshup, S. Rybakov, F. J. Theis, P. Angerer, F. Alexander Wolf, anndata: annotated data. bioRxiv, 2021.12.16.473007 (2021). https://doi.org/10.1101/2021.12.16.473007.

103. S. L. Wolock, R. Lopez, A. M. Klein, Scrublet: Computational identification of cell doublets in single-cell transcriptomic data. *Cell Syst.* **8**, 281–291.e9 (2019).

104. C. Qiu, J. Cao, B. K. Martin, T. Li, I. C. Welsh, S. Srivatsan, X. Huang, D. Calderon, W. S. Noble, C. M. Disteche, S. A. Murray, M. Spielmann, C. B. Moens, C. Trapnell, J. Shendure, Sys-tematic reconstruction of cellular trajectories across mouse embryogenesis. *Nat. Genet.* **54**, 328–341 (2022).

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

106. L. McInnes, J. Healy, UMAP: Uniform Manifold Approximation and Projection for Di-mension Reduction. arXiv:1802.03426 [stat.ML] (2020).

107. K. Polański, M. D. Young, Z. Miao, K. B. Meyer, S. A. Teichmann, J.-E. Park, BBKNN: Fast batch alignment of single cell transcriptomes. *Bioinformatics* **36**, 964–965 (2020).

108. W. Dong, C. Moses, K. Li, "Efficient k-nearest neighbor graph construction for generic similarity measures" in *Proceedings of the 20th International Conference on World Wide Web* (Association for Computing Machinery, York, NY, USA, 2011; http://dx.doi.org/10.1145/1963405.1963487), *WW '11*, pp. 577–586.

109. R Core Team, R: A language and environment for statistical computing (2013).

110. B. Bushnell, "BBMap: a fast, accurate, splice-aware aligner" (Lawrence Berkeley National Lab, 2014), (available at https://osti.gov/servlets/purl/1241166).

111. H.-G. Drost, Philentropy: Information theory and distance quantification with R. *J. Open Source Softw.* **3**, 765 (2018).

112. M. A. Zolotovskaia, V. S. Tkachev, A. A. Guryanova, A. M. Simonov, M. M. Raevskiy, V. V. Efimov, Y. Wang, M. I. Sekacheva, A. V. Garazha, N. M. Borisov, D. V. Kuzmin, M. I. Sorokin, A. A. Buzdin, OncoboxPD: Human 51 672 molecular pathways database with tools for activity calculating and visualization. *Comput. Struct. Biotechnol. J.* **20**, 2280–2291 (2022).

113. M. Sorokin, N. Borisov, D. Kuzmin, A. Gudkov, M. Zolotovskaia, A. Garazha, A. Buzdin, Algorithmic annotation of functional roles for components of 3,044 human molecular pathways. *Front. Genet.* **12**, 617059 (2021).

114. A. Alexa, J. Rahnenführer, T. Lengauer, Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. *Bioinformatics* **22**, 1600–1607 (2006).

115. F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, É. Duchesnay, Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).

116. I. Korsunsky, N. Millard, J. Fan, K. Slowikowski, F. Zhang, K. Wei, Y. Baglaenko, M. Brenner, P.-R. Loh, S. Raychaudhuri, Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* **16**, 1289–1296 (2019).

117. R. R. Bouckaert, DensiTree: Making sense of sets of phylogenetic trees. *Bioinformatics* **26**, 1372–1373 (2010).

118. K. P. Schliep, phangorn: Phylogenetic analysis in R. *Bioinformatics* **27**, 592–593 (2011).

119. A. M. Bolger, M. Lohse, B. Usadel, Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).

120. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).

121. Y. Zhang, T. Liu, C. A. Meyer, J. Eeckhoute, D. S. Johnson, B. E. Bernstein, C. Nusbaum, R. M. Myers, M. Brown, W. Li, X. S. Liu, Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).

122. J. M. Gaspar, Improved peak-calling with MACS2. bioRxiv, 496521 (2018). https://doi.org/10.1101/496521.

123. D. Bredikhin, I. Kats, O. Stegle, MUON: Multimodal omics analysis framework. *Genome Biol.* **23**, 42 (2022).

124. R. Satija, J. A. Farrell, D. Gennert, A. F. Schier, A. Regev, Spatial reconstruction of single-cell gene expression data. *Nat. Biotechnol.* **33**, 495–502 (2015).

125. A. S. Hinrichs, D. Karolchik, R. Baertsch, G. P. Barber, G. Bejerano, H. Clawson, M. Diekhans, T. S. Furey, R. A. Harte, F. Hsu, J. Hillman-Jackson, R. M. Kuhn, J. S. Pedersen, A. Pohl, B. J. Raney, K. R. Rosenbloom, A. Siepel, K. E. Smith, C. W. Sugnet, A. Sultan-Qurraie, D. J. Thomas, H. Trumbower, R. J. Weber, M. Weirauch, A. S. Zweig, D. Haussler, W. J. Kent, The UCSC Genome Browser Database: Update 2006. *Nucleic Acids Res.* **34**, D590–D598 (2006).

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

127. D. Machlab, L. Burger, C. Soneson, F. M. Rijli, D. Schübeler, M. B. Stadler, monaLisa: An R/Bioconductor package for identifying regulatory motifs. *Bioinformatics* **38**, 2624–2625 (2022).

128. A. Khan, O. Fornes, A. Stigliani, M. Gheorghe, J. A. Castro-Mondragon, R. van der Lee, A. Bessy, J. Chèneby, S. R. Kulkarni, G. Tan, D. Baranasic, D. J. Arenillas, A. Sandelin, K. Vandepoele, B. Lenhard, B. Ballester, W. W. Wasserman, F. Parcy, A. Mathelier, JASPAR 2018: Update of the open-access database of transcription factor binding profiles and its web framework. *Nucleic Acids Res.* **46**, D260–D266 (2018).

Science Advances

A single-cell multi-omic atlas spanning the adult rhesus macaque brain

Kenneth L. Chiou, Xingfan Huang, Martin O. Bohlen, Sébastien Tremblay, Alex R. DeCasien, Diana R. O'Day, Cailyn H. Spurrell, Aishwarya A. Gogate, Trisha M. Zintel, Cayo Biobank Research Unit, Madeline G. Andrews, Melween I. Martínez, Lea M. Starita, Michael J. Montague, Michael L. Platt, Jay Shendure, and Noah Snyder-Mackler

View the article online
https://www.science.org/doi/10.1126/sciadv.adh1914
Permissions
https://www.science.org/help/reprints-and-permissions