

Spatially resolved gene regulatory and disease-related vulnerability map of the adult Macaque cortex

Received: 13 December 2021

Accepted: 24 October 2022

Published online: 08 November 2022

 Check for updates

A list of authors and their affiliations appears at the end of the paper

Single cell approaches have increased our knowledge about the cell type composition of the non-human primate (NHP), but a detailed characterization of area-specific regulatory features remains outstanding. We generated single-cell transcriptomic and chromatin accessibility (single-cell ATAC) data of 358,237 cells from prefrontal cortex (PFC), primary motor cortex (M1) and primary visual cortex (V1) of adult female cynomolgus monkey brain, and integrated this dataset with Stereo-seq (spatial enhanced resolution omics-sequencing) of the corresponding cortical areas to assign topographic information to molecular states. We identified area-specific chromatin accessible sites and their targeted genes, including the cell type-specific transcriptional regulatory network associated with excitatory neurons heterogeneity. We reveal calcium ion transport and axon guidance genes related to specialized functions of PFC and M1, identified the similarities and differences between adult macaque and human oligodendrocyte trajectories, and mapped the genetic variants and gene perturbations of human diseases to NHP cortical cells. This resource establishes a transcriptomic and chromatin accessibility combinatory regulatory landscape at a single-cell and spatially resolved resolution in NHP cortex.

Cortical organization across primate brains is highly similar, as exemplified by the specialized primary visual cortex and the dorsal and ventral visual streams. With a similar neocortex organization in monkey and human brain, the monkey offers a unique model for studying features of human neurodevelopment and neuropsychiatric diseases^{1–3}. The cynomolgus monkey (*Macaca fascicularis*) is one of the most studied non-human primates (NHP) in neuroscience and medicine^{4,5}. Recent advances in transgenesis and genome-editing technologies have led to successful development of new cynomolgus monkey genetic models to study human neurological disorders^{6–8}, making this species an excellent experimental NHP model for studying higher order brain function.

Single-cell genomic sequencing enables studies of the underlying diversity and regulatory mechanisms of cortical cells at an unprecedented resolution. Single cell analyses of macaque brain by extension therefore stand to provide an in-depth understanding of the human

brain and an opportunity to identify markers and molecular signatures of neuropsychiatric diseases. Earlier single-cell RNA-sequencing studies of prenatal and adult macaque brain identified functionally distinct cortical cell types as well as subtypes across multiple cortical brain areas, and gene expression variations across these cell types^{2,9}. These studies do not capture chromatin states that exert fundamental control of regulating gene expression programs and combining epigenetic analysis with gene expression profiling could therefore yield a complementary understanding of the molecular properties of brain cells. For instance, single-cell-based chromatin states assays can identify cell-type-specific transcriptional regulatory elements and predict potential master transcriptional regulators¹⁰. Similarly, a survey of chromatin states in bulk tissue from several cerebral cortical regions and hippocampus of the macaque brain revealed region-related chromatin accessibility patterns¹¹. However, a systematic characterization of single-cell-based chromatin accessibility of the

✉ e-mail: niuyy@pbr.cn; hongkuiz@alleninstitute.org; huyong@genomics.cn; liulongqi@genomics.cn; liushiping@genomics.cn; xuxun@genomics.cn

macaque cortex at a region-specific and single-cell resolution, a requirement for advancing the field, remains outstanding.

Recently, development of spatial transcriptomics technology has made it possible to assign gene expression profiles to spatial coordinates in cortical tissues^{12,13}. In the present study, we sought to understand how the transcriptional and epigenetic regulatory states differ across functionally distinct cortical areas. To achieve this goal, we performed single nuclei ATAC-seq and RNA-seq, combined with spatial enhanced resolution omics-sequencing (Stereo-seq)¹⁴ on three functional diverse cortical regions of cynomolgus monkey brain: prefrontal cortex (PFC) and primary visual cortex (V1), two distant areas at frontal lobe and occipital lobe that the differences of neurons between these two areas were recognized in human brain development, and primary motor cortex (M1), a key structure involved in locomotion. Through our massive parallel and integrative topographical analysis, we defined cell type-specific and regional-specific regulatory elements, in a spatially resolved manner, resolving single cell gene expression programs between neural cell types, in particular excitatory neurons, in different cortical areas of the NHP brain. We also applied our dataset to delineate the dynamic regulatory landscape of myelination, linking macaque cortical cell types to human neurological disease risk. We provide an interface website (<https://db.cngb.org/mba>) to present data for exploration and sharing. This publicly available database allows a user to filter cells in the atlas data by brain region, cluster, and cell type, and perform interactive searches from our macaque cortical dataset.

Results

Single-cell transcriptomics define brain cortical cellular taxonomy in adult cynomolgus monkey

We apply Smart-seq2 based snRNA-seq and droplet-based DNBelab C4 RNA-seq¹⁵ to PFC, M1 and V1 of the macaque neocortex (Fig. 1a). After quality filtering and cell clustering, we obtained a 11,194 single-nuclei Smart-seq2 transcriptome (6,715 from PFC, 1,943 from M1 and 2,536 from V1) and a 127,003 single-nuclei DNBelab C4 transcriptome (30,844 from PFC, 62,263 from M1 and 33,896 from V1) (Supplementary Data 1). Based on known marker genes for cortical cell types, all clusters were annotated as excitatory neuron (EX), inhibitory neuron (IN) and non-neuronal cells, including oligodendrocytes (OLI), oligodendrocyte precursor cells (OPC), astrocyte (AST), microglia (MIC) and endothelia (ENDO) (*SLC17A7* for EX, *GADI* for IN, *MOBP* for OLI, *SLCIA2* for AST, *APBB1IP* for MIC and *FIT1* for ENDO) (Supplementary Fig. 1a). Given the characteristic neocortical laminar organization and subcortical projections^{16,17}, we adopted the human brain single nuclei RNA-seq dataset from multi-cortical areas^{18,19} (Allen Cell Types Database-Human Multiple Cortical Areas, <https://portal.brain-map.org/atlas-and-data/rnaseq/human-multiple-cortical-areas-smart-seq>) to assign layer and projection pattern onto our transcriptomic excitatory neurons (Supplementary Fig. 1b). Thus, the transcriptomic EX subtypes were defined both by the cortical layers (L2/3, L4/5, L5 and L6) and neuronal projection within the cerebral hemispheres, or telencephalon, i.e., intra-telencephalic (IT), extra-telencephalic (ET), near-projecting (NP) and corticothalamic (CT) types (Fig. 1b, c, Supplementary Fig. 2, Supplementary Data 2). We further examined marker gene expression of mouse cortical EX-subtypes^{20,21} in our data, whose enriched pattern is highly congruent between macaque and mouse (e.g., *CUX2* enriched in L2-4 IT with adjusted $P < 1.00 \times 10^{-300}$, *RORB* enriched in L4/5 IT with adjusted $P < 1.00 \times 10^{-300}$, *FEZF2* enriched in L5 ET with adjusted $P < 1.00 \times 10^{-300}$ and *NXPH4* enriched in L6b with adjusted $P = 0.001$, Supplementary Fig. 1c), providing strong support for stratified EX subtypes identification. Notably, although the L4 layer is assumed to be absent in M1, an L4-like IT layer was recently identified in human M1, and in mouse, an L4/5 IT layer was also labeled by a combination of L4 and L5 marker expression, such as *CUX2*, *RORB* and *FEZF2*^{16,20}. We confirmed that our L4/5-like IT layer in macaque M1 also

expressed *CUX2* and *RORB* and is potentially located between deep L3 and superficial L5 (Supplementary Fig. 1d).

It's well known that cortical IN roughly fall into two major branches corresponding to their developmental origins in the caudal ganglionic eminence (CGE) and medial ganglionic eminence (MGE), respectively. *ADARB2*²⁸, a marker gene of CGE-derived IN, had differential expression in VIP, LAMP5 and RELN sub clusters, while *LHX6*, a marker gene of MGE-derived IN, revealed specific expression in SST and PVALB sub clusters. These relationships indicate that, during corticogenesis, VIP, LAMP5 and RELN IN are derived from CGE, while SST and PVALB are derived from MGE. The chandelier cell (ChC) subtype of PVALB2 inhibitory neuron can further be demarcated by the marker genes *UNCSB* and *RORA*¹⁶ (Fig. 1b, c). By gene ontology (GO) analysis of differentially expressed genes (DEGs), we found that the major cell types were in concordance with the expected corresponding biological processes (Supplementary Fig. 2f), solidifying our cell type assignments.

Linking chromatin accessibility to transcriptome in monkey cortical cell types

We applied modified combinatorial barcoding-assisted single-cell ATAC-seq¹⁰ and droplet-based DNBelab C4 ATAC-seq to tissue samples from three cortical areas of the cynomolgus monkey brain: PFC, M1 and V1 (Fig. 1a). To exclude low-quality cells, we filtered snATAC-seq data using a cutoff of 3,000 unique nuclear fragments per cell and a TSS enrichment score of 3. We then processed a total of 220,040 qualified single cells for further analysis: 72,714 cells from PFC, 70,050 cells from M1 and 77,276 cells from V1. The cells exhibited a 12,823 median fragment depth per nucleus and a median fraction of reads in peak regions at 55% (Supplementary Data 1). We performed iterative latent semantic indexing and batch corrections by ArchR 1.01²² with the top 15,000 accessible windows, then used shared nearest neighbor (SNN) clustering by Seurat V3 to separate cells into 23 distinct clusters. Through promoter accessibility and gene activity score, calculated by ArchR of brain cell marker genes, we manually annotated the identity of the cell clusters (Fig. 1d, Supplementary Fig. 3a-d). Using this approach, we characterized seven major cortical cell populations, including EX (accessible at *NEFH* and *SLC17A7*), IN (accessible at *NEFH* and *GADI*), OLI (accessible at *MOBP*), OPC (accessible at *PDGFRA*), MIC (accessible at *AIFI*), AST (accessible at *SLC7A10*) and ENDO (accessible at *CLDN5*). The four IN subclusters could be further assigned to VIP, LAMP5, SST and PVALB, respectively, due to the distinct open peaks at the promoter of these genes (Fig. 1e).

To link the transcriptome states to open chromatin, we co-embedded the snRNA-seq and snATAC-seq data of cortical cells with Seurat V3²³. We converted the accessible peak of snATAC-seq data to gene activity score using ArchR and then anchored this analysis to snRNA-seq gene expression data using Seurat V3. With this approach, cells from snATAC-seq were positioned close to cells with matching snRNA-seq data assignments (Fig. 1f). Our snATAC-seq data contained similar sub clusters corresponding to the same major brain cell types as the snRNA-seq. Notably, pericytes (PERI) can be distinguished from ENDO cell types both in transcriptomics and in epigenetic nuclei; and the epigenetic LAMP5 IN subtype fell into two sub clusters, one that co-clustered with the transcriptomic LAMP5 subtype, while the other co-clustered with the reelin (RELN) expressing transcriptomic subtype (Fig. 1g). The epigenetic IN PVALB-ChC subtype can be resolved by co-embedding with transcriptomic PVALB-ChC. To further support the concordance of our cell type identification between the snRNA-seq and snATAC-seq datasets, we looked at marker gene expression in the identified transcriptomic cell types and gene activity scores in the corresponding epigenetic cell types (Supplementary Fig. 4). Further, we transferred the subtype annotation of snRNA-seq to epigenetic excitatory neurons in the integrated analysis (Fig. 1g, Supplementary

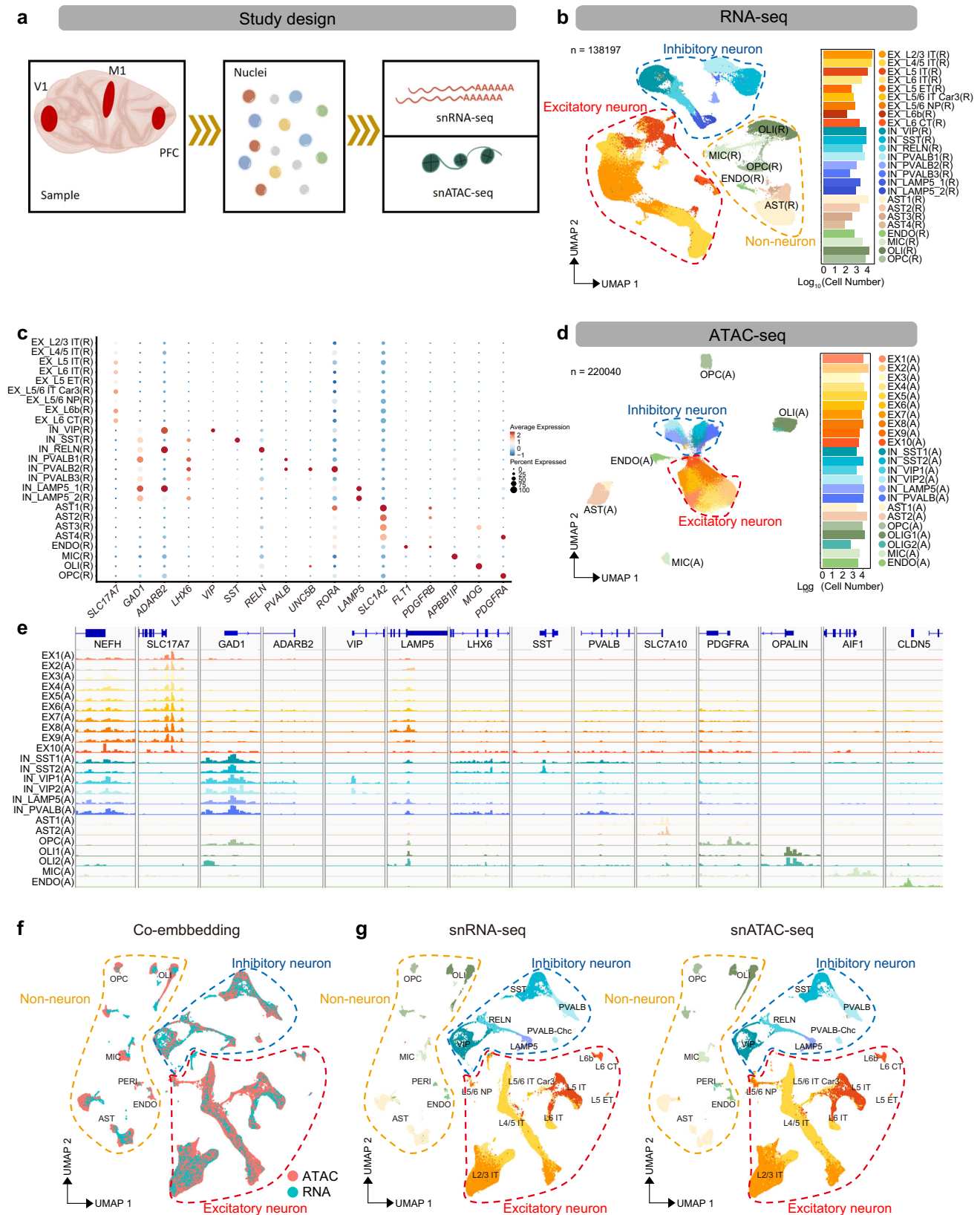


Fig. 3e, f). To confirm congruence between the transcriptomic and epigenetic EX subtypes, we performed differential gene expression and differential accessible peak tests for layer/projection-defined EX subtypes. The DEGs of EX subtypes for snRNA-seq cells were also found with chromatin accessibility in the corresponding subtypes of snATAC-seq cells. (Supplementary Fig. 5a, b, Supplementary Data 3).

To find transcriptional regulators that specify cortical cell type identity, we performed transcription factor (TF) motif enrichment analysis on epigenetic cell types (Supplementary Fig. 5c, Supplementary Data 4). This analysis showed that major cell type-enriched TF binding motifs in mouse and human brains^{24,25} are present in the corresponding macaque cortical cell types. NEUROD family and NEUROG

Fig. 1 | Single nucleus transcriptome and chromatin accessibility of macaque cortex (Source data are provided as a Source Data file). **a** Schematic workflow of single-nucleus isolation from the prefrontal cortex (PFC), primary motor cortex (M1) and primary visual cortex (V1) for snATAC-seq and snRNA-seq, sample processing, library generation and downstream analysis. **b** Uniform manifold approximation and projection (UMAP) of snRNA-seq data showing all clusters colored by cell types (N = 138,197 cells) (left), bar plot showing the number of cells per cell type (right). **c** Dot plots of expression values for marker genes of neuronal and non-neuronal cell types in snRNA-seq data. **d** UMAP visualization of snATAC data

showing all clusters colored by cell types (N = 220,040 cells) (left), bar plot showing the number of cells per cell type (right). **e** Integrative genomics viewer (IGV) plots showing color-coded aggregate read density for cells within each cell cluster at neuronal and non-neuronal cell-type-specific marker genes. **f** Co-embedding of snRNA-seq and snATAC-seq clusters across PFC, M1 and V1 brain regions using Seurat V3 (N = 347,682 cells). **g** UMAP projection of snRNA-seq (N = 138,197 cells) and snATAC-seq cells (N = 209,485 cells) from (f). Cell clusters in each UMAP are color-coded and annotated according to cell identity.

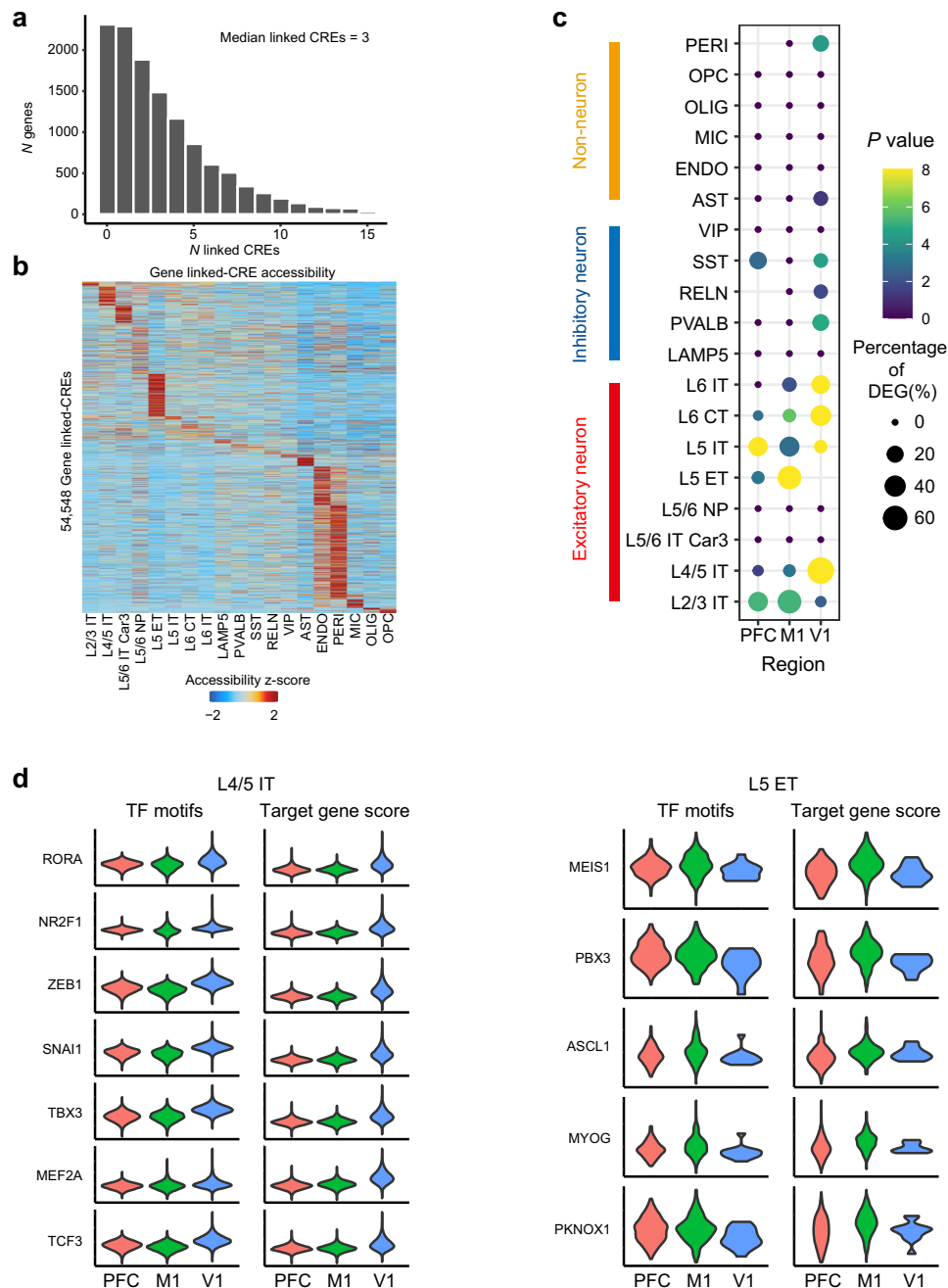


Fig. 2 | Transcriptional regulation of the cell type-specific and area-specific gene expression (Source data are provided as a Source Data file). **a** Histogram showing the number of genes that linked to *cis*-regulatory elements (CREs). **b** Heatmap showing row-normalized gene-linked CREs in each cell type of snATAC-seq cells. **c** Dot plots of overlapped genes between differentially accessible peaks (DA peaks)-targeted gene and the differentially expressed genes (DEGs) of each

cortical area in neuronal and non-neuronal cell types. Dot size represents the percentage of overlap genes of area DEGs. A one-sided hypergeometric test was performed for gene set overlap significance. **d** Violin plots of motif enrichment and targeted gene score of selected TFs that had significant upregulated in V1 of snATAC-seq L4/5 IT type excitatory neuron (left) and in M1 of snATAC-seq L5 ET type excitatory neuron (right).

family motifs were enriched in EX clusters, consistent with their roles in EX specification and synapses forming²⁶. Likewise, we found that SOX family (SOX TFs contribute to OLI migration, specification, maintenance of OPC state)^{27–29} motifs were enriched in OLI (e.g. SOX4, SOX9 and SOX17 with adjusted $P < 1.00 \times 10^{-300}$), while PU.1 (a myeloid master regulator that controls microglial development and function)³⁰ motif were enriched in MIC (adjusted $P < 1.00 \times 10^{-300}$). Finally, LHX2 (known astroglialogenesis regulators)³¹ motif were enriched in AST (adjusted $P < 1.00 \times 10^{-300}$), and forkhead (FOX) transcription factor family (endothelial gene expression regulators)³² motifs were enriched in endothelial cells (e.g. FOXA2 and FOXF1 with adjusted $P < 1.00 \times 10^{-300}$). These findings demonstrate that regulatory signatures for major cortical cell types are conserved across species. Moreover, we identified differentially enriched TF motifs within neuron subtypes, such as MEF2B /MEF2C /MEF2D (adjusted $P < 1.00 \times 10^{-300}$) in parvalbumin (PVALB) INs, and ZNF238 (adjusted $P < 1.00 \times 10^{-300}$) and ZBTB42 (adjusted $P < 1.00 \times 10^{-300}$) in somatostatin (SST) INs. (Supplementary Fig. 5c) In addition, we found 12 transcription factors that were enriched in EX subtypes as predicted by their binding motifs (Supplementary Fig. 5d), including cortical neurogenesis regulation co-factors MEIS2³³ in L2/3 IT (adjusted $P < 1.00 \times 10^{-300}$), regional identity and laminar patterning transcription factor PBX1³⁴ in L5/6 NP (adjusted $P = 6.33 \times 10^{-5}$), neurodevelopment regulator NR4A2³⁵ (adjusted $P = 3.55 \times 10^{-30}$) and NR2F1³⁶ (adjusted $P = 0.0025$) in L5/6 IT Car3, and cortical development transcription factor NFIA³⁷ in L6 CT (adjusted $P < 1.00 \times 10^{-300}$) and L6b (adjusted $P = 7.38 \times 10^{-48}$).

In summary, the snRNA-seq shows a better resolution for cell subtype identification in the present study. Our results demonstrate robust congruence between epigenetic and transcriptomic data for classifying cortical cell types, enabling further integrative analysis.

Areal heterogeneity of excitatory neuron revealed by integrative single-cell analysis

Recent single-cell studies have resolved area-specific transcriptomic features of cortical EX in mice¹⁷, and the molecular and epigenetic signatures in the developing human brain^{38,39}. However, the epigenetic features that correspond to area-specific feature in adult cortex (monkey or human) are yet to be clarified. We were therefore intrigued to investigate the area-specific transcriptomic and epigenetic features in our dataset.

We performed an inter-regional comparison of the transcriptomic features across PFC, M1 and V1 using droplet-based DNBelab C4 RNA-seq cells (127,003 cells), and identified 699, 125 and 388 genes that were area-specifically enriched in EX, IN and non-neuronal cell types, respectively (Supplementary Data 5, Supplementary Fig. 6a–c). We found most of the area-enriched genes of excitatory neurons (66.4% in PFC, 73.8% in M1 and 63.5% in V1) are restricted in single EX subtype, such as *KCNH8* in L4/5 IT of V1 (adjusted $P < 1.00 \times 10^{-300}$); while less than 10% of area-enriched genes (5.4% in PFC, 4.1% in M1 and 8.2% in V1) are broadly up-regulated in more than three subtypes, such as *SNTG2* in L2/3 IT (adjusted $P < 1.00 \times 10^{-300}$), L4/5 IT (adjusted $P < 1.00 \times 10^{-300}$), L5 IT (adjusted $P < 1.00 \times 10^{-300}$) and L6 IT (adjusted $P = 1.23 \times 10^{-225}$) of M1, and none of the area-enriched genes were found to be present in all EX subtypes (Supplementary Fig. 6d, e). We found genes differentially expressed in PFC and V1 areas of adult macaque bulk tissue⁴⁰ are expressed in EX-subtypes of corresponding layers, including *DPYD*, which is upregulated in PFC EX subtypes, and *ZFPM2*, *MAML3* and *VAI3*, which are upregulated in V1 EX subtypes (Supplementary Fig. 6e). Moreover, genes reported as differentially expressed in human excitatory neurons of PFC versus V1^{41,42} were similarly enriched in macaque, including *L3MBTL4* (adjusted $P < 1.00 \times 10^{-300}$ in L4/5 IT, adjusted $P = 3.58 \times 10^{-6}$ in L5/6 IT Car3) and *TLL1* (adjusted $P = 5.12 \times 10^{-199}$ in L5 IT), which were upregulated in PFC; and *PDE3A* (adjusted $P < 1.00 \times 10^{-300}$ in L2/3 IT and L4/5 IT), *TRPC3* (adjusted $P < 1.00 \times 10^{-300}$ in L4/5 IT, adjusted $P = 3.10 \times 10^{-236}$ in L5 IT) and *PCDH7* (adjusted $P < 1.00 \times 10^{-300}$

in L4/5 IT), which were upregulated in V1 (Supplementary Fig. 6e), indicating a conserved area specialization between human and macaque cortex. Of note, we not only identified layer and neuron projection specificity in previous known area-enriched genes for PFC or V1, but also identified new area-specific candidate EX genes, such as polypeptide N-acetylgalactosaminyltransferase like 6 (*GALNTL6*, adjusted $P < 1.00 \times 10^{-300}$) and beta-1,4-N-acetylgalactosaminyltransferase 3 (*B4GALNT3*, adjusted $P < 1.00 \times 10^{-300}$) in PFC, syntrophin gamma 2 (*SNTG2*, adjusted $P < 1.00 \times 10^{-300}$) and prune homolog 2 with BCH domain (*PRUNE2*, adjusted $P < 1.00 \times 10^{-300}$) in M1, and potassium voltage-gated channel subfamily H member 8 (*KCNH8*, adjusted $P < 1.00 \times 10^{-300}$), heparanase 2 (*HPSE2*, adjusted $P < 1.00 \times 10^{-300}$), and ADAM metalloproteinase with thrombospondin type 1 motif 17 (*ADAMTS17*, adjusted $P < 1.00 \times 10^{-300}$) in V1 (Supplementary Fig. 6e).

To identify area-specific chromatin opening states, we performed an inter-regional comparison of chromatin accessibility using the snATAC-seq cells with consensus cell type identification between snATAC-seq independent annotation and co-embedding assignment (209,485 cells) and found 83,110 differential accessible sites (DA peaks) between PFC, M1 and V1, across 9 subtypes of EX (Supplementary Data 6). Among of them, 6,347 (7.64%), 15,885 (19.11%) and 42,528 (51.17%) of differential accessible sites were located within 0 to 1 kb, 1 kb to 10 kb and 10 kb to 100 kb genomic distance to the nearest genes, respectively. Using the matched cell type of single cell transcriptomic and chromatin accessibility data, we built the links of the *cis*-regulatory elements (CREs) to targeted gene expression by ArchR in each cell type (see methods). In total, we identified 54,548 significant associated CRE-gene pairs, with a median of 3 CREs linked to each of 12,484 genes, similar as reported in human brain cell types⁴³ (Fig. 2a), and most of the gene-linked CREs showed a cell type-specific distribution (Fig. 2b). For all neuron and non-neuronal cell types, we found a significant overlap between CRE-targeted genes and cell types DEGs (Supplementary Fig. 7), demonstrating a key role of chromatin accessible sites in defining cell type specific transcriptomics. Next, we sought to elucidate the peak-gene regulation on areal heterogeneity. For most excitatory neuron subtypes, there is a significant overlap between DA peak-targeted genes and the differentially expressed genes among PFC, M1 and V1 (Fig. 2c), indicating an important role of differentially accessible sites in area-specialized transcriptomics.

To further explore the transcriptional regulation of the area-specific expression, we first measure transcription factor binding site enrichment in each EX-subtypes using chromVAR and analyzed the differentially enriched TF binding motif across PFC, M1 and V1. We showcase transcription factors in L4/5 IT and L5 ET, we found the binding motif enrichment and significantly increased targeted gene score of ZEB1, MEF2A, RORA, NR2F1, SNAI1, TBX3 and TCF3 in V1 of L4/5 IT (adjusted $P < 1.00 \times 10^{-300}$ for motif and targeted gene score), some of these have previously been reported for motif enrichment in human developing V1³⁸, and MEIS1 (adjusted $P = 1.81 \times 10^{-31}$ and 8.62×10^{-17} for motif and targeted gene score), PBX3 (adjusted $P = 3.86 \times 10^{-25}$ and 3.18×10^{-9} for motif and targeted gene score), ASCL1 (adjusted $P = 3.17 \times 10^{-22}$ and 7.87×10^{-10} for motif and targeted gene score), MYOG (adjusted $P = 1.99 \times 10^{-26}$ and 1.12×10^{-12} for motif and targeted gene score) and PKNOX1 (adjusted $P = 1.29 \times 10^{-32}$ and 1.16×10^{-10} for motif and targeted gene score) in M1 of L5 ET (Fig. 2d). Then, we constructed the TFs regulated genes network using peak-gene links and TF binding motif. The area-specific DEGs were identified among the TFs targeted genes (Supplementary Fig. 8a) and the area-enriched TF binding motif can be linked to the distal-regulated genes (Supplementary Fig. 8b).

Collectively, our findings reveal overlap in inter-regional transcriptomic profiles and chromatin accessibility and identified transcriptional regulation in areal heterogeneity of excitatory neuron.

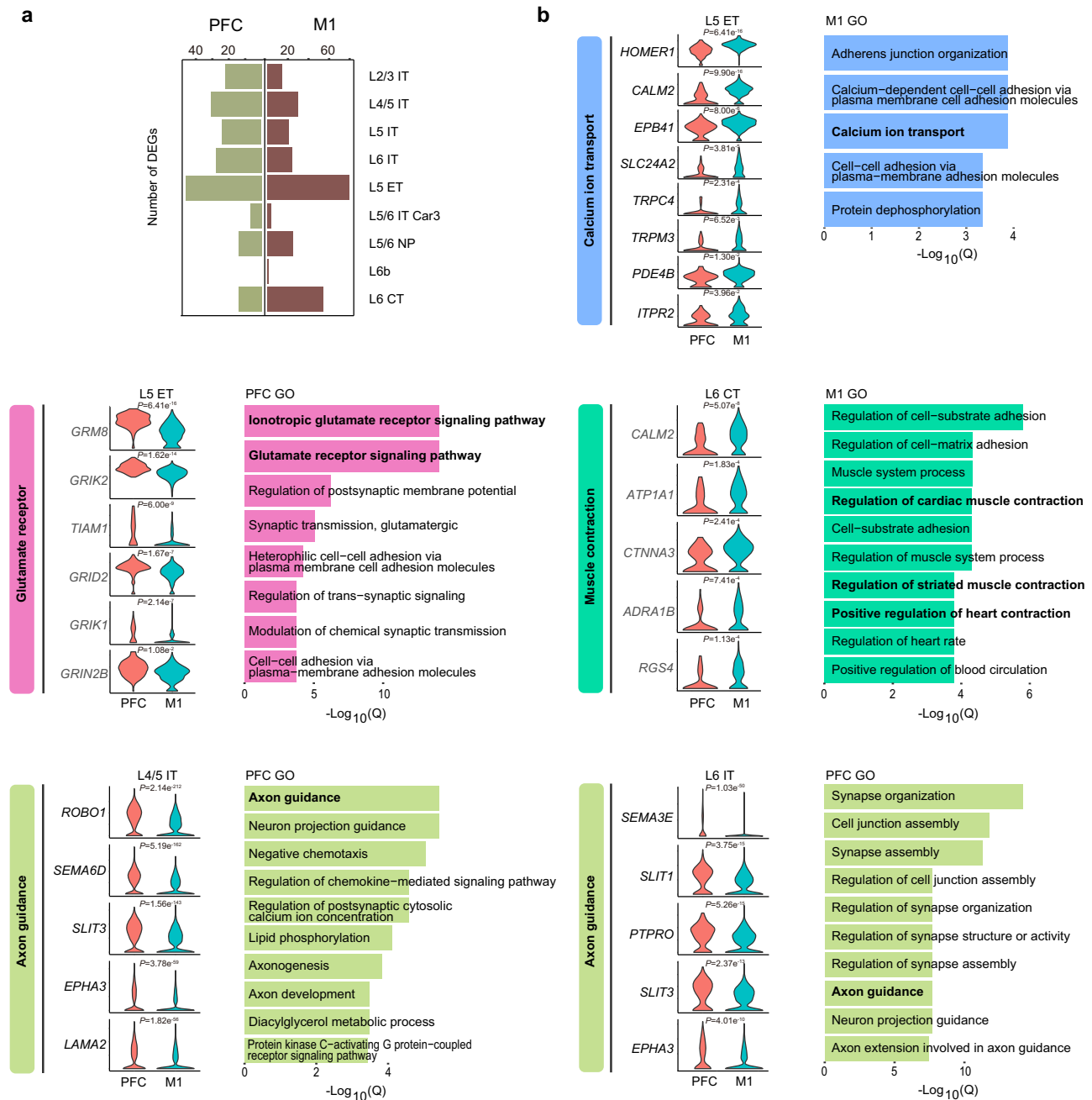


Fig. 3 | Specialized transcriptomics of excitatory neuron between macaque prefrontal cortex and motor cortex (Source data are provided as a Source Data file). **a** Number of differentially expressed genes between PFC and M1 in each subtype of excitatory neurons. **b** Gene ontology terms enriched among differentially expressed genes of PFC and M1, and violin plots showing expression of

selected DEGs between PFC and M1 of L5 ET, L6 CT, L4/5 IT and L6 IT type excitatory neuron, respectively. Unpaired two-samples one-sided Wilcoxon test was used to perform inter-regional gene expression comparison and with Bonferroni correction.

Specialized transcriptomics in macaque prefrontal cortex and motor cortex

Areal and layer proximity is the strongest determinant for gene expression relationships in the primate cortex⁴, which is believed to explain why V1 has a more distinct transcription pattern when compared to PFC and M1. Previous studies have revealed fundamental differences between the anterior and posterior cerebral cortex by comparing differentially expressed genes between PFC and V1, whereas heterogeneity within the frontal cortex of primates is less well understood. Here, we analyzed differentially expressed genes between PFC and M1 in excitatory neuronal subtypes and found that 122 genes and 159 genes were upregulated in PFC and M1 across EX

subtypes, respectively. (Fig. 3a, Supplementary Data 7). Specialized functions and expressions have been reported in L5 extratelencephalic (L5 ET) neurons of primate and mouse motor cortex¹⁶. In M1 L5 ET, we identified many genes involved in calcium ion transport, such as *SLC24A2* (adjusted $P=3.81 \times 10^{-5}$), *ITPR2* (adjusted $P=0.04$), *CALM2* (adjusted $P=9.90 \times 10^{-16}$) and *PDE4B* (adjusted $P=0.01$) (Fig. 3b and Supplementary Fig. 9). In our data sets, we didn't find primate M1 L5 ET-enriched potassium ion signaling-related genes¹⁶ (including *KCNH1*, *KCNC1* and *KCNC2*) were differentially expressed between PFC and M1. To ascertain PFC- and M1-associated specialized biological functions in EX subtypes, we performed gene ontology analysis. We found that many M1-enriched

genes are involved in ‘muscle contraction’ and comprise genes encoding for the adrenergic receptor and calcium ion transport proteins, exemplified by L6 CT type (Fig. 3b and Supplementary Fig. 9), suggestive of an involvement in the long-range cortico-motor neuronal projections⁴⁴. For PFC-enriched genes in EX subtypes, we found glutamate receptor signaling pathway genes to be significantly increased in L5 ET, including *GRIK1* (adjusted $P=2.14 \times 10^{-07}$), *GRIK2* (adjusted $P=1.62 \times 10^{-14}$), *GRID2* (adjusted $P=1.67 \times 10^{-07}$), and *GRIN2B* (adjusted $P=0.01$) (Fig. 3b and Supplementary Fig. 9). We also found enrichment of many axon guidance genes implicated in cortical development and psychiatric disorders^{45–47}, such as *SLIT/ROBO* signaling members *SLITI* (adjusted $P=3.75 \times 10^{-15}$ in L6 IT), *SLIT3* (adjusted $P=1.56 \times 10^{-143}$ in L4/5 IT, adjusted $P=5.41 \times 10^{-132}$ in L5 IT and adjusted $P=2.37 \times 10^{-13}$ in L6 IT) and *ROBO1* (adjusted $P=2.14 \times 10^{-212}$ in L4/5 IT) and semaphorins *SEMA3E* (adjusted $P=1.03 \times 10^{-50}$ in L6 IT) and *SEMA6D* (adjusted $P=5.19 \times 10^{-162}$ in L4/5 IT) (Fig. 3b and Supplementary Fig. 9). These axon guidance genes might influence distinct PFC connectivity patterns.

Gradient gene expression pattern of excitatory neuron across laminar and cortical regions

Tissue level transcriptomics of macaque neocortex^{2,40} and single cell transcriptomics in multi-regions of mouse isocortex²⁰ revealed continuous variation across cortical layers and regions, especially for glutamatergic neuron types. Here, we sought to further reveal two aspects of the inter-areal diversity of excitatory neurons: (1) The gradient gene expression patterns across cortical layers; and (2) The gradient gene expression patterns across PFC, M1 and V1 areas. We first analyzed the gene expression pattern in the largest branch of excitatory neurons; the inter-telencephalon (IT) projection type (corresponded well with cortical depth), and clustered these genes into 5 gradient expression patterns across L2/3 IT, L4/5 IT, L5 IT and L6 IT using the soft clustering package Mfuzz⁴⁸ in PFC, M1 and V1, respectively (Supplementary Fig. 10a, b) to reveal gene distribution and enrichment in specific layer of cortex. Then, we compared the gene sets in consensus patterns between PFC, M1 and V1. We found that 119 genes were uniformly expressed across layers, including transcriptional factors *MEIS2*, *RFX3*, *HIVEP1*, *ARID2* and *PKNOX2*, and glutamate receptor *GRIA4* and *GRIN2B*; while 1694 genes showed distinct expression pattern between PFC, M1 and V1, and were enriched in biological processes such as cell morphogenesis, synaptic signaling and regulation of cell projection organization (Supplementary Fig. 10c, Supplementary Data 8).

When we analyzed the gradient gene expression patterns across PFC, M1 and V1 areas, we obtained 6 clustered patterns in each excitatory neuron subtype (Supplementary Fig. 10d, e, Supplementary Data 9). These patterns reflect the cortical area-preferred gene expression, for example, pattern 1 and pattern 5 are gradient expression for PFC-preferred and V1-preferred genes, respectively. Notably, the gradient expression pattern of the axon guidance molecule *SLITI* (pattern 1) is highly consistent with previous in situ hybridization analysis in macaque cortex; highest in the prefrontal cortex, faint in the primary motor cortex and lowest in the primary visual cortex⁴⁵, in congruence with the gradient expression pattern of macaque cortical genes predicted by our dataset. When we examined the consensus gene patterns in different EX subtypes, for example, genes in pattern 1 of L2/3 IT (163 genes), L4/5 IT (192 genes), L5 IT (203 genes) and L6 IT (177 genes), we found that only 6 genes were shared by all four subtypes and that 126 genes were shared by more than one subtype, including the *TCF4*⁴⁹, a TF implicated in schizophrenia, which was shared by L2/3, L4/5 and L6. Moreover, genes in pattern 1 are involved in divergent signaling pathways, for instance GTPase activity regulating genes are enriched in L2/3, genes involved in regulation of macroautophagy are enriched in L4/5 and genes involved in membrane docking are enriched in L6 (Supplementary Fig. 10f). These findings

demonstrated the cell subtype/laminar-biased pattern for different cortical functions.

Spatial gene expression characterized by Stereo-seq

To spatially resolve the macaque cortical cell transcriptome, we performed the recently developed technology Stereo-seq¹⁴, in which DNA nanoball (DNB) patterned array chips are combined with in situ RNA capture to enable nanoscale transcriptome analysis of tissue sections. With the DNB approach, Stereo-seq achieves a resolution of 500 or 715 nm and captures a significantly higher number of spots per 100 μm^2 compared to other related techniques¹⁴. We used 10 μm paired adjacent tissue sections from macaque PFC, M1 and V1 tissues for Nissl staining or Stereo-seq, respectively (Fig. 4a). Stereo-seq captured an average number of 31,916 of 37.5 μm bins (bin 50, 50 \times 50 DNB) per section with 1,748 genes and 4,754 transcripts per 37.5 μm bin for analyzed sections (Supplementary Fig. 11). We performed unsupervised clustering with gene matrices of 37.5 μm bin with Seurat V3 and found that the clusters defined by the Stereo-seq data shows high similarity with cortical layers defined by Nissl staining in adjacent tissue sections, as exemplified by recognition of the L4 multi-sub-layers in V1 (Supplementary Fig. 12a). Thus, the high resolution achieved with Stereo-seq data enabling us to distinguish heterogenous features of macaque cortex tissue.

To register the spatial localization of macaque cortical cell types, we assigned the cell type from our snRNA-seq data to Stereo-seq data using SPOTlight⁵⁰. We found that the cell type annotations were highly concordant with spatial expression of known layer markers from macaque brain¹ and the cortical layers defined from histology staining of adjacent tissue section (Fig. 4b–d, Supplementary Fig. 12b). We then examined the area-differentially expressed genes of excitatory neuron subtypes between PFC, M1 and V1 of snRNA-seq in the corresponding cell subtypes annotated in Stereo-seq sections of PFC, M1 and V1. We found a significant number of genes were also exhibit area-specificity in Stereo-seq data (42 genes in PFC, $P=5.13 \times 10^{-9}$; 51 genes in M1, $P=3.10 \times 10^{-5}$; 81 genes in V1, $P=0.025$, by hypergeometric test), including known and newly identified area-enriched genes in EX (Fig. 4e, Supplementary Data 10), demonstrating the consistent finding for area-specific transcriptomics in cell subtype resolution between single cell transcriptomic and Stereo-seq.

Next, to explore the gradient gene expression across cortical layers in situ, we examined the spatial distribution of 119 genes with consensus gradient gene expression patterns between PFC, M1 and V1 of snRNA-seq (see methods). We found 33 genes (28%) showed consistent patterns in snRNA-seq and spatial transcriptomics across three cortical regions (Supplementary Fig. 13), including schizophrenia risk gene *ARHGAP10*⁵¹ (pattern 1), important transcription factor of central nerve system *RFX3*⁵² (pattern 1) and AD-related genes *LRPIB*⁵³ (pattern 4) and *PTK2*⁵⁴ (pattern 5); while other 86 genes showed consistent patterns between snRNA-seq and spatial transcriptomics mostly in two of the three cortical regions. This discrepancy of snRNA and Stereo-seq may be caused by the relatively lower gene capture in the spatial transcriptomics than snRNA-seq.

Collectively, we demonstrated that integrative analysis of single cell genomics and spatial resolved transcriptomics can identify consistent areal diversity. Additionally, we captured consensus gradient expression pattern in different cortical regions in situ.

Dynamic single cell regulatory landscape of oligodendrocyte trajectory

OLI wrap neuronal axons with myelin to support neuronal function⁵⁵. Previous work has reported on transcriptional and epigenetic regulation pathways of OLI maturation and myelination in mouse models and the human brain⁵⁶, but due to insufficient data coverage it wasn't possible to assess how these two layers of regulation are dynamically correlated. Given that macaque models are widely used to model

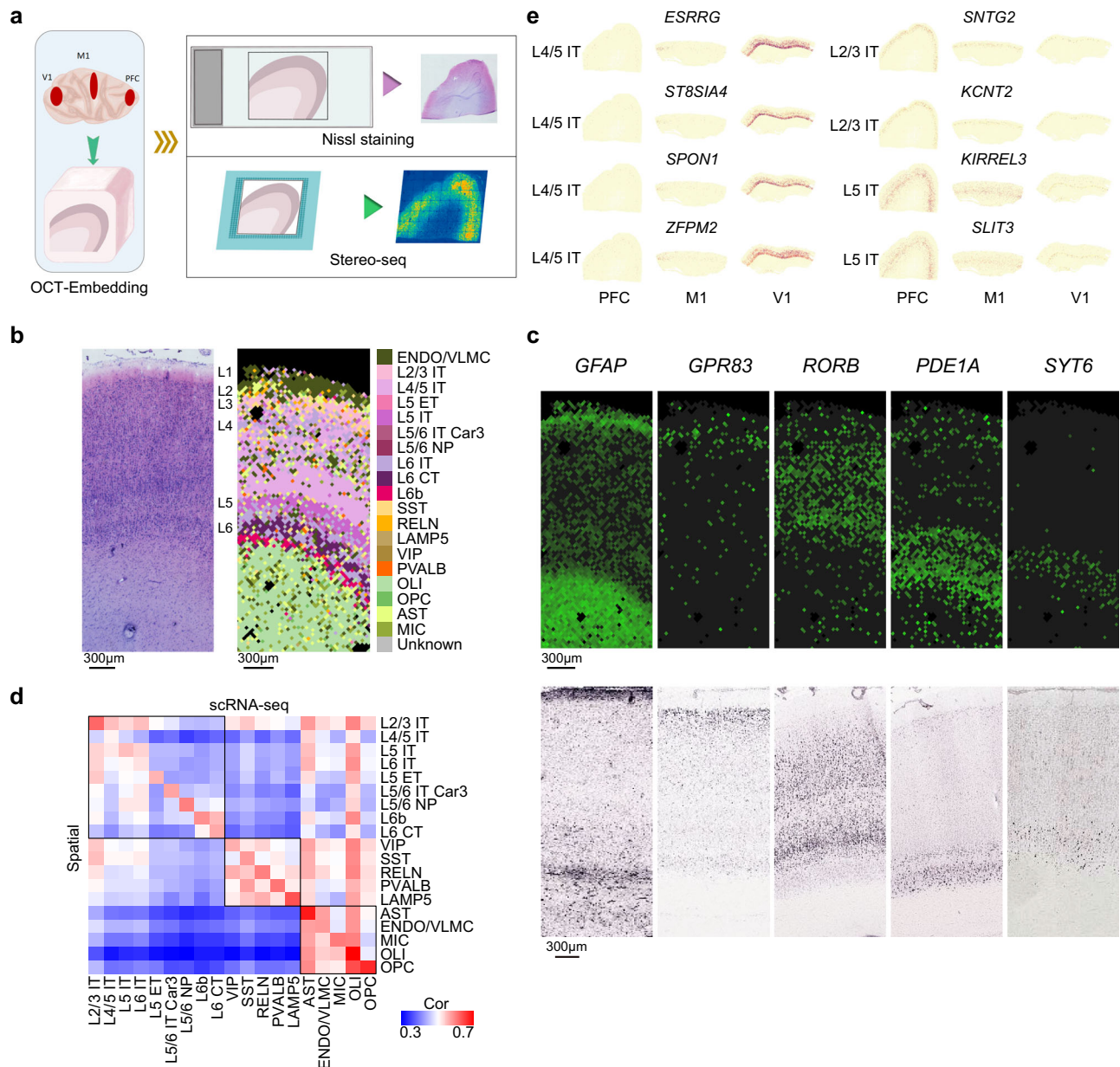


Fig. 4 | Characterization of cortical gene expression by Stereo-seq (Source data are provided as a Source Data file). **a** Schematic workflow of tissue sampling and Stereo-seq from the prefrontal cortex (PFC), primary motor cortex (M1) and primary visual cortex (V1). **b** Nissl staining of adjacent section to stereo-seq section of primary visual cortex (left), and cell type annotation for 37.5 μm bins of stereo-seq section of primary visual cortex by snRNA-seq data (right). **c** Representative micrographs of known layer-marker genes expression in the stereo-seq section of (b) (top), experiments have been performed independently in 2 slices, and RNA ISH images for corresponding layer markers in primary visual

cortex of macaque brain from the NIH Blueprint Non-Human Primate (NHP) Atlas.¹ *GFAP* (<http://www.blueprintnpatlas.org/ish/experiment/show/100140483>), *GPR83* (<http://www.blueprintnpatlas.org/ish/gene/show/183031>), *RORB* (<http://www.blueprintnpatlas.org/ish/gene/show/183109>), *PDE1A* (<http://www.blueprintnpatlas.org/ish/gene/show/183138>) and *SYT6* (<http://www.blueprintnpatlas.org/ish/experiment/show/100091672>) (bottom). **d** Heatmap showing Spearman's rank correlation between clusters identified in Stereo-seq versus snRNA-seq data. **e** Spatial expression maps of selected differentially expressed genes between PFC, M1 and V1 in EX subtypes.

disease, such as demyelination-related diseases autoimmune encephalomyelitis and multiple sclerosis^{57–59}, we first sought to investigate whether our data could recapture the dynamic signature of OPC differentiation and OLI development in those studies and extend these to human biology, and second whether we could deepen our understanding of the OLI regulatory landscape (Fig. 5a).

We performed pseudotime ordering of cells by Monocle 2 to construct OLI trajectories using gene expression and gene activity score (Fig. 5b, c). We found that both the marker gene expression and activity score reported in mouse and human OPC and mature OLI^{60–62} were also enriched in our macaque OPC (*NEU4* and *PDGFRA*)

and OLI (*MBP* and *PLP1*) (Fig. 5d, e). Despite recent studies highlighting the importance of glutamatergic, GABAergic and calcium signaling in the OLI lineage, these relationships remain controversial given limited insight into tissue heterogeneity and receptor activation during OLI development and maturation. In earlier work, GO enrichment of glutamate receptors were confirmed in human OPC during lineage maturation⁶². Here, we found enriched gene expression and high epigenetic score of ionotropic glutamate receptors in macaque OPC and immature OLI, including the NMDA (N-methyl-D-aspartate) receptors (*GRIN2A* and *GRIN2B*) and kainate receptors (*GRIK1*, *GRIK2*, *GRIK4* and *GRIK5*) in OPC and immature

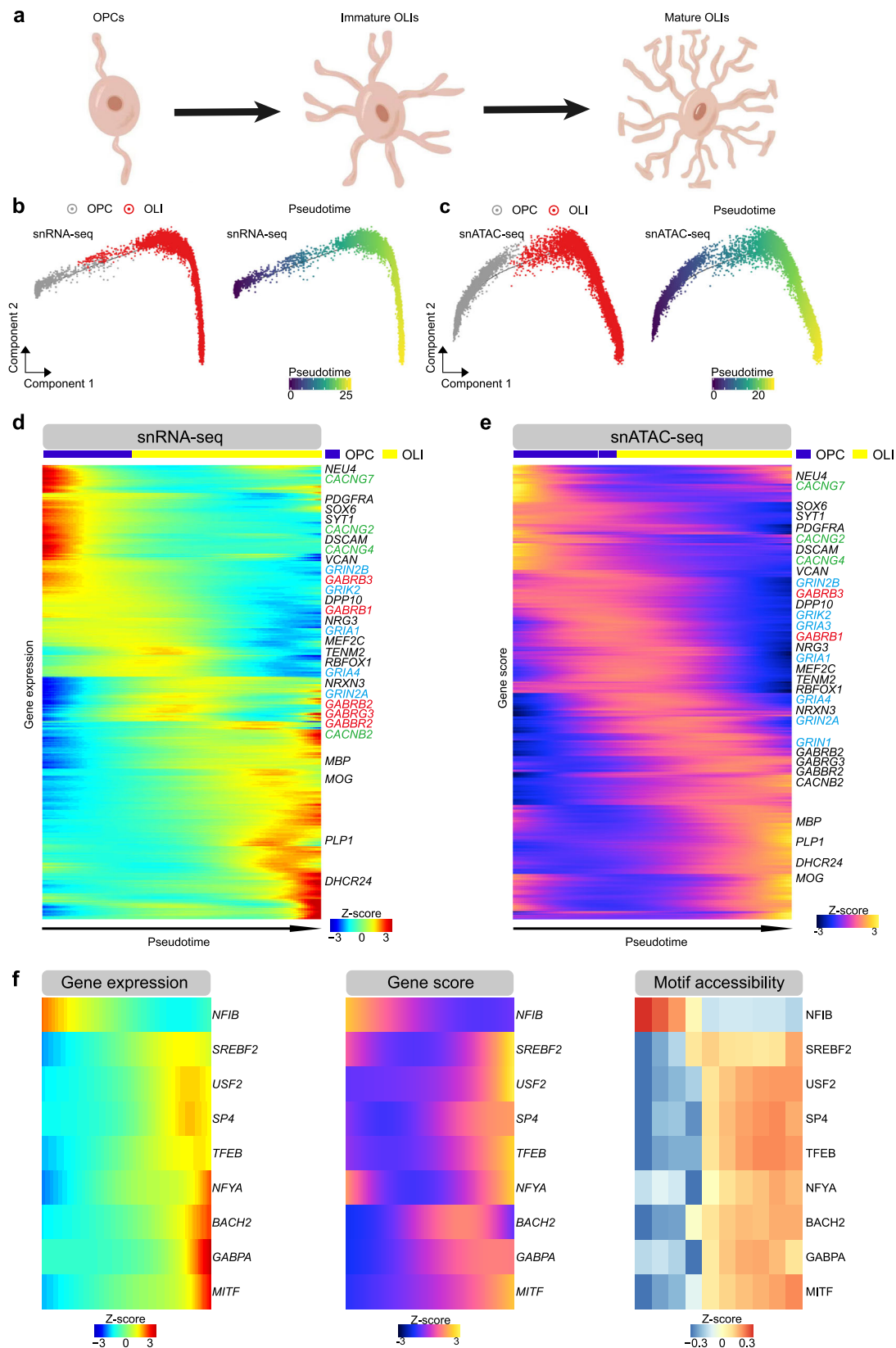


Fig. 5 | Integrated single cell regulatory landscape of oligodendrocyte trajectory (Source data are provided as a Source Data file). **a** Schematic illustration for oligodendrocyte maturation. **b** Pseudotime trajectory of OPC and OLI indicating the gene expression obtained from snRNA-seq data. **c** Pseudotime trajectory of OPC and OLI indicating the gene activity score obtained from snATAC-seq data. **d** Heatmap showing genes expression dynamics across oligodendrocyte lineage pseudotime trajectory indicated in (b). Genes of glutamate receptors, gamma-

aminobutyric acid (GABA) receptors and voltage-gated calcium channel (VGCCs) were blue, red, and green color-coded, respectively. **e** Heatmap showing accessibility dynamics (gene activity score) across oligodendrocyte lineage pseudotime trajectory in (c). Genes of glutamate receptors, gamma-aminobutyric acid (GABA) receptors and voltage-gated calcium channel (VGCCs) were blue, red, and green color coded, respectively. **f** Gene expression (left), gene activity score (middle) and motif enrichment (right) of 9 TFs across OLI pseudotime trajectory.

OLI, AMPA (α -amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid) receptors (*GRIA1*, *GRIA3* and *GRIA4*) in immature OLI. Furthermore, we found enriched gene expression and high epigenetic score of GABA (gamma-aminobutyric acid) receptors (*GABRB1*, *GABRB2*, *GABRB3*, *GABBR2* and *GABRG3*) and voltage-gated calcium channels (*CACNB2*, *CACNB4*, *CACNG2*, *CACNG4*, *CACNG7* and *CACNG8*) in OPC and immature OLI. When connecting the expression trajectories to accessibility dynamics, we defined a concordant set of 2211 genes in expression-activity score pairs (Pearson's correlation coefficient $R > 0.2$, adjusted $P < 0.01$, Supplementary Fig. 14a), including transcription factor *SOX5* (adjusted $P = 1.90 \times 10^{-55}$), *SOX6* (adjusted $P = 1.30 \times 10^{-106}$), *OLIG2* (adjusted $P = 2.40 \times 10^{-25}$) and *MYRF* (adjusted $P = 5.92 \times 10^{-72}$) known for roles in OPC differentiation and OLI development⁵⁶. GO terms for concordance gene sets supported stage-enriched pathways and dynamic changes of neuronal activity along the pseudotime trajectory (Supplementary Fig. 14a). Previous work proposed that the Wnt signaling pathway plays an important role as a stage-specific and multifunctional regulator of OLI development⁶³. Our data showed that genes involved in Wnt signaling are enriched in OPCs (Supplementary Fig. 14b).

To understand transcriptional regulation of OLI maturation, we characterized TF motif enrichment across OPC differentiation and OLI development by mapping chromVAR TF deviation scores to cells along the pseudo time trajectory. This approach identified 129 TFs that defined different stages of the oligodendrocyte maturation process (Supplementary Fig. 15). We also constructed the TFs regulatory network using peak-gene links and TF binding motif enrichment in OPC and OLI; the DEGs between OPC and OLI were identified among the TFs targeted genes (Supplementary Fig. 16). To link TF expression to enrichment of TF binding motifs across different stages of OLI maturation, we found 9 TFs that had congruent gene expression and gene activity scores/motif accessibility profiles (Fig. 5f), such as OLI differentiating activators and myelination regulators *TFEB*⁶⁴ in mature OLI.

To gain insight into the similarity and differences of dynamic trajectory of oligodendrocyte in adult macaque monkey and human, we adopted the snRNA-seq and snATAC-seq data of OPC and OLI of adult human brain from a recent study⁴³. After constructing the pseudotime trajectory of gene expression and gene activity score, we found a concordant set of 1756 genes in expression-activity score pairs across the human OLI trajectories (Pearson correlation coefficient $R > 0.2$, adjusted $P < 0.01$, Supplementary Fig. 17a). Similar with macaque OLI, these sequential activated genes were strongly enriched in trans-synaptic signaling and cellular morphogenesis pathways (Supplementary Fig. 17a). Among of them, 829 genes were overlapped with the concordant gene expression-activity score pairs along macaque monkey OLI trajectory (Supplementary Data 11), including myelin forming OLI signature *MOG*, *PLP1* and *OPALIN*, and mature OLI signature *KLK6* and *SLC5A11*⁴³. We found 5 transcription factors had congruent gene expression and gene activity scores /motif accessibility profiles along the human OLI trajectory (Supplementary Fig. 17b). Among of these, activation of transcription factor *EB* (*TFEB*) was observed along both human and macaque monkey OLI trajectory; and the sterol regulatory element-binding transcription factor (*SREBP*), involved in cholesterol and fatty acids biosynthesis of glia cell and associated with schizophrenia⁶⁵, were found activation in mature stage of macaque and human OLI trajectory.

Taken together, our analysis proposes master regulators for OLI maturation based on combinatorial analysis of single nucleus transcriptomic and chromatin accessibility, providing candidates for further functional studies in demyelination diseases and demonstrate the conserved regulatory landscape of oligodendrocyte trajectory in adult human and macaque monkey brains.

Mapping human traits and diseases to cortical cell types

Linking cell-type specific regulatory elements with disease risk variants identified in genome-wide association studies (GWAS) can help us understand whether specific cell types contribute to disease pathobiology, which in turn can become instrumental for developing targeted therapeutic approaches. Single-cell epigenetic data of mouse and human brain have already proven to be useful for mapping neuropsychiatric disease risks with specific cellular subtypes^{62,66}. However, the different enrichment of disease risks across cortical regions has not been illustrated before. To demonstrate the robustness of our dataset, we assessed enrichment of human neurological and neuropsychiatric disease risk factors in macaque cortical cell types. We mapped differentially accessible peaks of each cluster and all accessible peaks to orthologous coordinates in the human hg19 genome, then performed linkage disequilibrium score regression (LDSC) to measure SNP heritability enrichment for human traits within differential accessible peaks of each epigenetic cluster in PFC, M1 and V1, respectively. We adopted the GWAS summary for neurological and psychiatric disorders and neurobehavioral traits from recent studies (Supplementary Data 12) as well as non-brain-related diseases (from UK biobank).

Consistent with cell type mapping studies of neurological trait risk in human and mouse, we found a highly significant enrichment of heritability in neurons for neuropsychiatric traits such as major depressive disorders (MDD) and schizophrenia (SCZ) (Fig. 6a)^{62,66}. In line with studies reporting on microglial activation in AD and early AD-associated transcriptional changes occurring in other brain cell types^{67,68}, we found strong enrichment of Alzheimer's disease (AD) SNP-heritability in MIC ($P = 1.67 \times 10^{-4}$ in PFC, $P = 0.0095$ in M1 and $P = 0.0019$ in V1). Extending from earlier single cell transcriptomic analysis reporting on enrichment of SCZ genetic variants in cortical pyramidal neuron and interneurons⁶⁹, we here mapped significant enrichment of SCZ genetic variant to excitatory neuron of the L2/3 IT type ($P = 1.75 \times 10^{-4}$ in PFC, $P = 2.62 \times 10^{-6}$ in M1 and $P = 0.021$ in V1) and SST inhibitory neuron ($P = 3.83 \times 10^{-4}$ in PFC, $P = 0.020$ in M1 and $P = 0.0031$ in V1) across PFC, M1 and V1, while only enriched in L5 IT and L6 IT of M1. The MDD heritability enrichment in prefrontal cortex has been highlighted⁷⁰, and we discovered here that genetic variants of MDD were enriched in both PFC and M1 ($P = 0.028$ in L2/3 IT of PFC, $P = 0.0097$ in L2/3 IT and $P = 0.013$ in L5 IT of M1), but not in V1. Conversely, the SNP-heritability of autism disorder (ASD) was significantly enriched in excitatory neuron of M1 ($P = 0.013$ in L6 CT) and V1 ($P = 0.030$ in L6b), but not PFC. Consistent with cell types known to be affected in non-brain-related autoimmune diseases, we found SNP-heritability enrichment for asthma, hypothyroidism, and eczema in microglia. Further, to determine how similar of disease heritability loci enrichment between normal human and macaque in same cortical region, we performed the LDSC analysis on previous human PFC snATAC-seq data⁴³, the neuron cell subtypes were assigned by human brain single nuclei RNA-seq dataset^{18,19} (Allen Cell Types Database-Human Multiple Cortical Areas, <https://portal.brain-map.org/atlas-and-data/rnaseq/human-multiple-cortical-areas-smart-seq>). We found that the heritability of neurological diseases was enriched in broader neuron subtypes of human PFC than their enrichment in macaque neuron subtypes. At the same time, the strongly linked cell types to disease genetic risk loci were consistent in human PFC and macaque PFC, such as the microglia enrichment for AD and L2/3 IT enrichment for BD and SCZ. We also noticed that although the SNP-heritability of ASD was not enriched in any cell type of macaque PFC, they were enriched in three EX subtype of human PFC (Supplementary Fig. 18). Thus, our macaque cortex single-cell open chromatin landscape provides cell type-specific datasets as a resource for evaluating genomic loci implicated in human neurological traits in specific macaque brain cell types and cortical areas.

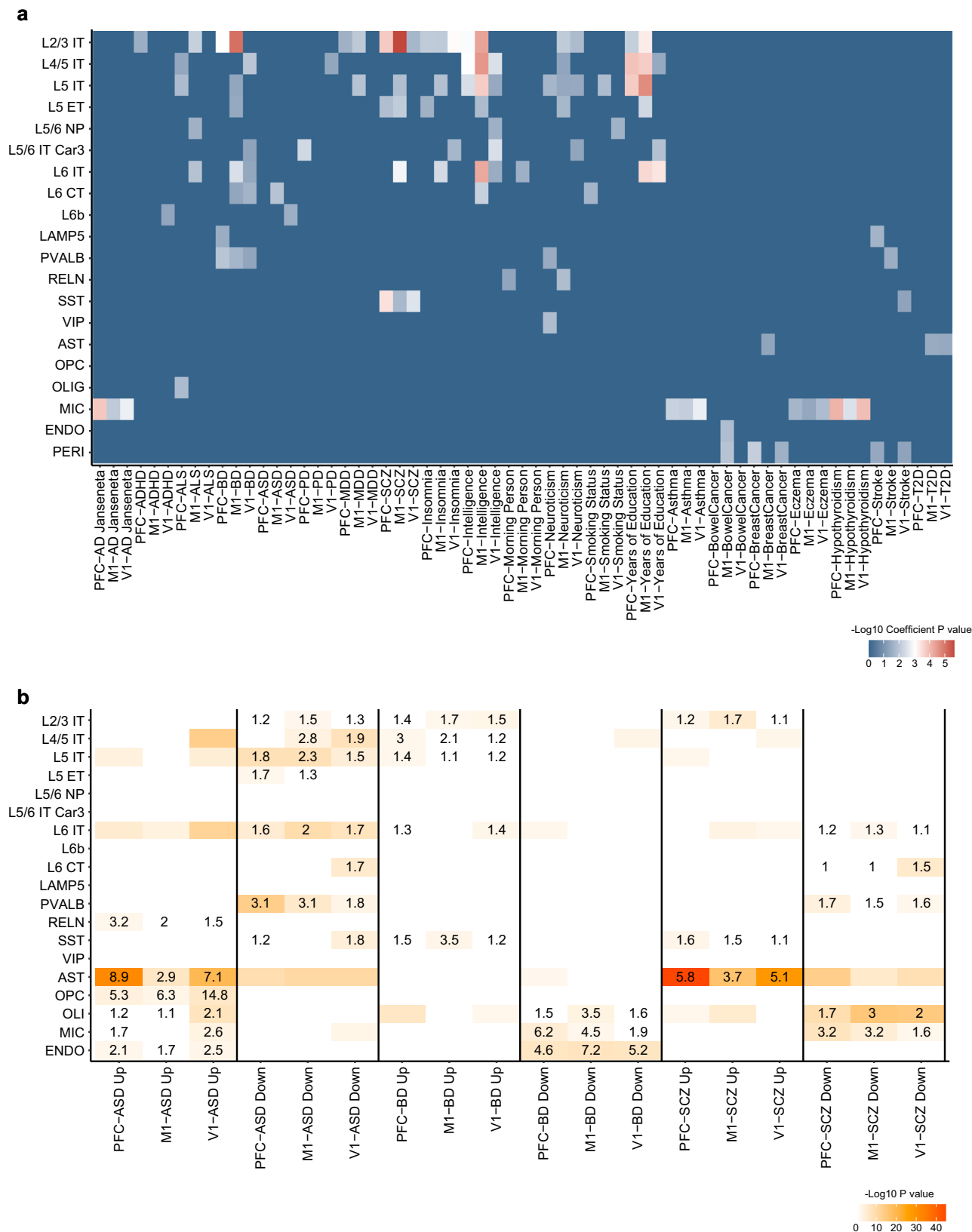


Fig. 6 | Cell type enrichment of heritability and dysregulated genes (Source data are provided as a Source Data file). **a** Heatmap showing the enrichments of heritability for human traits among the neuronal and non-neuronal cell types from snATAC-seq data of PFC, M1 and V1. *P* values were derived from LDSC enrichment tests. AD, Alzheimer’s disease; ADHD, attention deficit hyperactivity disorder; ASD, autism spectrum disorder; BD, bipolar disorder; ALS, amyotrophic lateral sclerosis; MDD, major depression disorders; SCZ, schizophrenia, T2D, type 2 diabetes. **b** Cell

type-enrichment of dysregulated genes for neurological diseases⁷¹ in Stereo-seq sections of PFC, M1 and V1. Differentially expressed genes (Up, up-regulated genes; Down, down-regulated genes) in brains of individuals with ASD, BD or SCZ versus controls. Numbers within heatmap cells indicate odd ratio (OR) value for the enrichments. A one-sided hypergeometric test was performed for gene set overlap significance.

To map spatial enrichment of disease-affected genes in specific cell types, we integrated large scale postmortem brain tissue level transcriptomics surveying differentially expressed gene in patients of ASD, bipolar disorder (BD) and SCZ⁷¹ with our Stereo-seq macaque cortical tissue dataset. Differentially expressed genes of cell types in each cortical region were calculated, then the significance of disease associated gene set enrichment in each cell type were calculated by hypergeometric test in PFC, M1 and V1, respectively (Fig. 6b). Previous study identified L2/3 and L5/6 cortico-cortical projection neurons as recurrently affected cell types across ASD patients⁷². We found ASD DEGs were enriched in L5 IT and L6 IT of all three regions; while the enrichments in L2/3 IT were only present in M1 ($P = 9.19 \times 10^{-3}$), but not PFC or V1 (Fig. 6b). Consistent with a recent single cell transcriptomics analysis of PFC samples from SCZ patient⁷³, we found a broad enrichment of SCZ DEGs in excitatory neuron subtypes, and the PVALB-type as the most affected inhibitory subtype ($P = 2.24 \times 10^{-4}$ of PFC and $P = 0.0070$ of V1). Notable, the ASD-and SCZ-associated perturbation were strongly enriched in astrocyte of all three regions. These findings highlight the value of our data for mapping and predicting cell types affected in disease across different cortical regions.

A recent single cell study on schizophrenia suggests a strong correlation between disease genetic risk loci and disease transcription perturbations within neuronal cell types⁷³. Our results revealed co-localization of SCZ heritability and SCZ-altered transcriptomics within neuron cell types in PFC and M1. Specifically, both SCZ heritability (Fig. 6a) and SCZ DEGs (Fig. 6b) are enriched in L2/3 IT and L6 IT and SST neuronal cell types in PFC or M1. However, this is not the case for ASD; although ASD DEGs is broadly enriched in neuronal subpopulation in all three regions (Fig. 6b), ASD heritability is not co-localized with ASD DEGs in any particular cell type of PFC, M1 or V1 (Fig. 6a). By combining epigenetic and spatial transcriptomics analysis, we have generated a resource that may provide clinical insights in cell type and cortical region-specific programs underlying genetic and disease states.

Discussion

Developing therapeutic approaches for central nervous system disorders has been hampered by the lack of models that can adequately mimic human disease pathogenesis. While murine models have been instrumental to study developmental trajectories and the genetic basis of disease phenotypes, they poorly mirror human physiology, behavior, and progressive disease mechanism. Using larger animal models that are phylogenetically close to humans, such as pigs and monkeys is therefore warranted for certain conditions. Despite evolutionary differences in cognitive functions and behaviors, monkeys have the social complexity, brain structure and neuronal circuitries that are more closely related to humans⁷⁴. As such, monkey models faithfully recapitulate several features of many neurodegenerative diseases such as Parkinson's⁷⁵ and Huntington's^{76,77}. Therefore, dissecting brain cellular composition and regulatory circuitries at a single cell resolution as well as determining disease-specific molecular signatures in species phylogenetically closer to human is an important goal.

In the present study, we generated a spatially resolved large-scale single-cell open chromatin and transcriptomic map of the adult primate cortex. We applied snATAC-seq and snRNA-seq to profile 358,237 single cells from three major cortical regions; prefrontal cortex, motor cortex and primary visual cortex, and performed the Stereo-seq on these cortical regions to obtain more than 30,000 spots (bins) per region. Relative to single gene or more qualitative approaches, multi-omics data applied in integrated analyses frameworks can increase both resolution and confidence of cell type annotations. To demonstrate the robustness of our data resource, we performed co-embedding of single-cell ATAC-seq and snRNA-seq data that confirmed the high consistency of our cell type assignments and linked the

cell type-specific expression profiles to their corresponding regulatory programs.

To map cell types and their molecular properties in their spatial contexts and to investigate cortical area heterogeneity in situ, we transferred the snRNA-seq data to Stereo-seq data. In addition, we integrate transcriptomic with spatial transcriptomics to profile the gene expression map on primate cortical areas. Through this approach, we were able to define the area specific expression and gradient gene expression pattern of excitatory neurons in situ. Thus, our combined epigenetic and transcriptomic analysis, spatially resolved with Stereo-seq profiling, synergized to reveal hidden aspects of spatial organization and cellular heterogeneity in cerebral cortex.

Dysfunctional myelination is a feature in several neurodegenerative diseases and neurodevelopmental disorders, including multiple sclerosis⁷⁸. Our data resource allowed us to perform an integrative analysis on single cell transcriptomic and epigenetic profiling of OPC and OLI, resolving a parallel dynamic gene expression landscape, open chromatin states and transcription factor enrichments related to myelination. By applying pseudotime analysis, we were able to map how specific gene expression changes, chromatin states and regulatory circuitries influence cell fate decisions throughout lineage maturation, define roles of master regulators, and identify potential targets for demyelination diseases. In addition, comparative analysis of the OLI lineage trajectories revealed the shared and divergent gene activation and TFs regulations between macaque and human.

A single-cell macaque genomics data resource can be applied directly in pre-clinical studies using NHP models of neuropsychiatric and neurological diseases. Here, we used our cell-type specific epigenetic data and the spatially resolved Stereo-seq data to predict the risk enrichment of neurological and neuropsychiatric disorders. Notably, mapping heritability loci with chromatin accessibility data, and mapping cell type perturbation with spatial transcriptomic data can help us understand whether or not aspects of disease biology is recapitulated in different cell types or brain regions. For example, in the PFC region, we found a highly overlapping cell type enrichment for genetic risk and disease-perturbation genes for SCZ, but not for ASD. Even for SCZ, the overlapping cell type enrichment for these two features varied between PFC, M1 and V1 such that the overlap of cell types was in PFC and M1 but not in V1. In addition to previously reported heterochronic and heterotopic divergence of disease associated genes expression between humans and macaques², we further demonstrated the shared and divergent cell type enrichments for disease genetic risk loci between normal human and macaque. Thus, our multi-omics macaque cortical cell data provide a valuable resource that can readily be applied to map human heritable traits and compared with disease model data.

We used female monkeys through the present study, which might be a limitation of the findings without comparison between the female and male. Future studies on tissues from cortical and sub-cortical areas from developmental and adult stages on female and male monkeys, subjected to integrative analysis of single cell transcriptomic, epigenetic, proteomics and spatial genomics promise to uncover systemic molecular mechanisms of the primate brain in health and disease. With the dataset at this high resolution, we identified regulatory elements and spatial expression profiles that shape primate cortical organization and harnessed valuable insights with relevance to human disease.

Methods

Ethics statement

All relevant procedures involving animals were reviewed and approved by the Institutional Animal Care and Use Committee of Yunnan Key Laboratory of Primate Biomedical Research, the Institutional Animal Care and Use Committee of Huazhen Bioscience (permit no. HZ2019027) and the Institutional Review Board on Ethics Committee of Beijing Genomics Institute (BGI; permit no. BGI-IRB 19125-T3).

Tissue processing and nuclei isolation

Tissues were sampled from three female 72-month-old cynomolgus monkeys (*Macaca fascicularis*) and immediately frozen in liquid nitrogen. All three regions (PFC, M1 and V1) were sampled from all three female macaques, the samples from one monkey (MK1) were performed Smart-seq2 single nucleus RNA-seq and sciATAC-seq, and samples from the other two monkeys (MK2 and MK3) were performed droplet-based single nucleus RNA-seq and ATAC-seq (Supplementary Data 1).

Single-nucleus preparations were performed as previous description⁷⁹, frozen monkey brain tissue pieces were placed in 1 ml homogenization buffer (pre-chilled) in 1 ml Dounce homogenizer (TIANDZ). Tissue was homogenized by 10 strokes of the loose pestle and 10 strokes of the tight pestle, Dounce homogenizer was submerged in ice during grinding. 2 ml homogenization buffer was added to the Dounce homogenizer then the homogenate was filtered through 40 μ m cell strainer (Miltenyi Biotech) into 15 ml conical tube and centrifuged at 900 g for 10 mins to pellet nuclei.

snRNA-seq library preparation and sequencing

For Smart-seq2 processing, nuclei were resuspended in 500 μ L blocking buffer containing 1X PBS (GIBCO), 2% filtered sterilized BSA (SIGMA), and 0.2 U/ μ L RNasin Plus by pipetting up and down gently on ice. Nuclei were transferred to 1.5 ml tube then split up into three tubes: experimental, control-isotype and control-DAPI. For single-nuclei isolation by flow cytometry, 250 μ L DAPI (0.1 μ g/ml) was added into the experimental tube to make the final volume as 500 μ L, while nothing added to the negative control tube. We also sorted NeuN⁺ nuclei from M1 region. For this purpose, rabbit anti-NeuN (Abcam, ab190195, final dilution of 1:500) was added to experimental tube, rabbit IgG-mono-clonal-isotype control (Abcam, ab199091, final dilution of 1:500) was added to control-isotype tube. After 30 min incubation at 4 °C, samples were then centrifuged for 5 min at 400 g to pellet nuclei and pellets were resuspended in 250 μ L blocking buffer (including 50 μ L leaving buffer). 250 μ L DAPI (0.1 μ g/ml) was added to nuclei to make the final volume for each tube as 500 μ L. Stained nuclei were filtered through 40 μ m filter before FACS (BD FACSAria II instrument). DAPI⁺ or DAPI⁺ NeuN⁺ nuclei in the experimental tube were sorted at a speed of 5–7 and the single nucleus was sorted into every single well of 384-well plate filled with 1.2 μ L lysis buffer (10% Triton X-100 0.0125 μ L, 40 U/ μ L RNase Inhibitor 0.0625 μ L, 10 M Oligo-dT Primer 0.02 μ L, 10 mM dNTP Mix 0.45 μ L, 5 \times SuperScriptII First-Strand Buffer 0.25 μ L, nuclease-free water 0.305 μ L and ERCC spike-in 0.1 μ L) in advance. Sorted nuclei in 384-well plates were briefly centrifuged and stored at –80 °C for further analysis. Sorted nuclei transcriptome amplifications were prepared by a modified Smart-seq2 protocol⁸⁰. After nuclei lysis, 1.2 μ L reverse transcription mixed solution was added into each well to complete the reverse transcription reactions. Then 1.8 μ L PCR reaction buffer was added into every well to complete amplification. At last, the amplified cDNA products of each single nucleus were quantified by Agilent Bioanalyzer 2100. For those single nucleus samples with high quality after amplification, the products were extracted by an automatic extractor from the 384-well plate to 96-well plate then purified by MGIEASY DNA Purification Magnetic Bead Kit (MGI) for the library construction. The libraries were prepared by MGIEASY RNA Library Preparation Kit (MGI) and each single nucleus sample was barcoded. Finally, the libraries were cyclized into ssDNA libraries by MGIEASY Cyclization Kit (MGI). All the single nucleus samples were sequenced on the BGISEQ500 sequencer with 100-bp pair-end reads.

For droplet-based DNBelab C4 RNA-seq processing, snRNA-seq libraries were prepared using DNBelab C Series Single-Cell Library Prep Set (MGI, #1000021082). The procedure was utilized as previously described¹⁵. Briefly, we followed the manufacturer's protocol for droplet generation using the single nucleus suspensions, emulsion

breakage, beads collection, reverse transcription and cDNA amplification to generate barcoded libraries. The resulting libraries were sequenced using DIPSEQ T1 sequencers at the China National Gene-Bank (Shenzhen, China). The read length followed: read 1, 30 bp, including 10 bp cell barcode 1, 10 bp cell barcode 2 and 10 bp UMI, read 2, 100 bp for transcript and 10 bp for sample index.

snRNA-seq data processing and quality control

Smart-seq2 data were processed as previous described⁸⁰, first, one-hundred-base-pair paired-end reads were pre-processed to remove adapters and filter out reads with low quality using default parameter by Cutadapt (v1.15)⁸¹. Next, filtered reads were aligned to the *Macaca fascicularis* genome (5.0.91) by STAR (v2.5.3)⁸² using a modified GTF file from ensemble release-91. In order to map the pre-mRNA fragments which may cover both exonic and intronic regions, we create a modified GTF annotation file which only contain transcript regions, and the original annotation rows for exons were all deleted, then we replace the feature type name from 'transcript' to 'exon'. Finally, transcripts per million mapped reads (TPM) were calculated using rsem-calculate-expression (RSEM)⁸³ (v 1.3.1) with default parameters.

We applied PISA (version 0.7, <https://github.com/shiquan/PISA>) to filter and demultiplex DNBelab C4 RNA-seq raw sequencing reads, then used STAR (version 2.6.1a) to align the reads with the previously mentioned modified GTF file of *Macaca fascicularis*_5.0 genome containing introns and exons and sort by sambamba (version 0.7.0). For each library, we sorted the cells from high to low according to the number of UMIs and kept the cells whose UMI number is greater than the UMI of the tenth-ranked cell divided by 10, and then used PISA to generate the UMI count matrix of the cell to the gene. For Smart-seq2 data, cells with unique mapped ratios > 15% and more than 500 genes with transcripts per million mapped reads value >1 and removed the top 5% of cells with the highest number of genes were used for downstream analysis. For DNBelab C4 data, cells with more than 500 genes and removed the top 5% of cells with the highest number of genes were used for downstream analysis. For both Smart-seq2 and DNBelab C4 RNA-seq data, cells with mitochondrial gene expression greater than 1% were filtered, and then all mitochondrial genes were deleted for downstream analysis.

snRNA-seq data clustering and analysis

Clustering and differential gene expression analysis was performed by Seurat (v 3.1.1) R toolkit⁸⁴ (R version 3.6.1). NormalizeData, FindVariableGenes and ScaleData were performed respectively in each cortical region of brain samples. Genes expressed in less than three cells were filtered out and cells with expressed genes less than 500 were excluded. SCTransform (SCT) was performed to integrate cells from 3 monkeys using the union set of 2000 variable genes in 3 batches. We select the top 20 dimensions (dims = 1–20) for SNN network construction. Then, cell clusters were identified by graph-based clustering approach, louvain algorithm (resolution = 0.5). We used uniform manifold approximation and projection (UMAP) to visualize the distance between cells in 2D space. The edges between two cells were measured by Jaccard distance to construct SNN graph. Then graph-based clustering was performed (times = 50) and modularity was calculated until reaching the maximum. "FindAllMarkers" and "FindMarkers" functions were used for differential gene expression analysis between clusters.

Annotation of excitatory subtypes by snRNA-seq data of human cortex

The subtypes of excitatory neuron were annotated by co-clustering with human cortical data^{18,19} (Allen Cell Types Database-Human Multiple Cortical Areas <https://portal.brain-map.org/atlas-and-data/rnaseq/human-multiple-cortical-areas-smart-seq>) excitatory subtypes with R package Seurat (v3.1.1). We integrated the Smart-seq2 and

DNBelab C4 data of macaque cortex with human cortical data, respectively by following steps: Log normalization was performed to integrate cells from each region of human and monkey brain using the union set of 3000 variable genes. We identified anchors using the FindIntegrationAnchors function with a default dimension of 20. We then passed these anchors to the IntegrateData function, following by scaling the integrated data, running PCA, and visualizing the results with UMAP. After finding anchors, we used the TransferData function to classify the macaque cells based on human excitatory subtypes. Cells acquired the label of L2-3 (IT) and L3 (IT) were defined as L2/3 IT type neuron, cells acquired label of L3-4 (IT), L3-5 (IT), and L4 (IT) were defined as L4/5 IT type neuron, cells acquired labels of L4-5 (IT), L4-6 (IT) and L5 (IT) were defined as L5 IT type neuron, cells acquired labels of L5-6 (IT) and L6 (IT) were defined as L6 IT type neuron. Cells acquired labels of L5/6 IT Car3, L5-6 NP, and L6b were defined as the corresponding identities. Cells acquired labels of L3-5 ET and L5 ET were defined as L5 ET type neuron, and cells acquired labels of L5-6 CT and L6 CT were defined as L6 CT type neuron.

snATAC-seq library preparation and sequencing

Sci-ATAC-seq was performed as described previously with modifications⁸⁵. Nuclei were strained in 40 μm strainer and centrifuged for 5 min at 500 g. The nuclei were resuspended in cold PBS (1% BSA) and counted using hemocytometer. Nuclei were adjusted concentration to 360/ μL . For transposition, added 7 μL cell suspension (around 2500 cell), 2 μL 5xTAG buffer and 1 μL unique barcoded Tn5 transposome into each well of 96-well plate, mixed gently and had a short spin⁸⁶. The plate was incubated at 55°C for 60 min with shaking (300 rpm). To quench the reaction, 10 μL of 40 mM EDTA was added to each well and gently mixed, then the plate was incubated at room temperature for 5 min. After reaction, 5 μL sorting buffer (5% BSA; 5 mM EDTA) was applied to each well, mixed well and pooled into one tube. The suspension was filtered through 40 μm strainer. Then one drop of DAPI (4', 6-diamidino-2-phenylindole, ThermoFisher Scientific) was added to the suspension and 25 nuclei were sorted by Aria II (BD) into 96-well plate containing 7 μL buffer EB, shortly spun down. Next, 1 μL 10% SDS was added to each well, mixed well and incubated at 55°C for 7 min with shaking (500 rpm) to lyse the nuclei. After the reaction, 1 μL 10% Triton-X was added to each well, spun down and incubated at room temperature for 5 min. For amplification, we added 1 μL unique barcoded N5&N7 (0.5 μM final concentration) and 10 μL NEBNext High-Fidelity 2x PCR Master Mix (NEB). PCR cycling conditions were as follow: 72°C 5 min, 98°C 30 s, (98°C 10 s, 63°C 30 s, 72°C 30 s) \times 11, held at 4°C. After that, we pooled each 96-well plate into tubes, and added 5 volume PB including pH-indicator (Qiagen) and 200 μL sodium-acetate (3 M, pH = 5.2), and reversed blending. Next, we used 4 columns to purify product following the MinElute PCR Purification Kit manual (Qiagen). DNA from each column was eluate using 25 μL EB buffer, pooled the 4 elution, and added EB buffer to 100 μL . To filtrate the fragment, we used the Ampure XP Bead by 0.5x & 0.7x. First added 50 μL XP beads, after incubation, collected supernatant and then added 70 μL XP beads. Finally, we used 100 μL EB buffer to elute the DNA. We quantified the libraries by Qubit fluorimeter (Life technologies) and detected fragment size using 2100 High Sensitivity (Agilent). To sequence, each library we used 330 ng for cyclizing, and then used 8 ng to make DNB. Each library was loaded into 2 lanes using BGISEQ500. The sequencing primer used as: Tn5 primer1, Tn5 primer2, SCIMDA primer. The read lengths: PE 100, including 4 indexes. Index 1 and index 2 was represented Tn5 barcode, index 3 and index 4 represented the PCR barcode. There are common bases between Tn5 barcode and PCR index, so we used cold reaction of BGISEQ500 to get out of the imbalance.

Droplet-based DNBelab C4 ATAC-seq was performed using DNBelab C Series Single-Cell ATAC Library Prep Set (MGI, #1000021878) with the procedure of step transposed, droplet

encapsulation, pre-amplification, emulsion breakage, then the captured beads were collected. After DNA amplification and purification, the snATAC library is already to sequence. DNA nanoballs were loaded into the chips and sequenced on the DIPSEQ T1 sequencer at China National Gene Bank (CNGB).

snATAC-seq data processing and quality control

We used an in-house pipeline to process the sci-ATAC-seq data. Briefly, we extracted the segments of barcode from the 4 constituent parts (read 1: 1-10, 32-41; read 2: 1-10, 38-47), and retained the reads with all segments of barcode are fully matched. Then we mapped the reads to *macaca fascicularis* genome downloaded from NCBI using bowtie2⁸⁷ with '-X 2000 --mm --local' as options. Reads with mapping quality less than 10 and reads mapped to the mitochondria or genome scaffold (chrAQ*, chrU*, chr*_random*, and chrK*) were filtered out. Then, PCR duplicates were removed according to the cell barcode and mapping loci by custom python script. The retained fragments of each library were used for further analysis.

We used the open-source pipeline (v0.7, <https://github.com/shiquan/PISA>) to analyze the DNBelab C4-based snATAC-seq data. First, the raw reads were filtered and demultiplexed using PISA with 2 mismatches allowed in barcode. Then the retained reads were aligned to the *macaca fascicularis* genome using BWA-MEM () with the default parameters. Reads with mapping quality less than 10 and reads mapped to the mitochondria or genome scaffold (chrAQ*, chrU*, chr*_random*, and chrK*) were filtered out and PCR duplicates were removed according to the cell barcode and mapping loci. The retained fragments of each library were used for further analysis.

snATAC-seq data clustering and analysis

We applied ArchR v1.0.1 (<https://github.com/GreenleafLab/ArchR>) to identify the cell populations within all data. First, cells with fragments lower than 3000 and TSS enrichment lower than 4 were filtered, then doublet score were calculated via the addDoubletScores function and filtered the doublets by the filterDoublets function with parameter: filterRatio = 2. Before defining cell clusters, we first created 500 bp tiled matrix of the genome and used the addIterativeLSI function to reduce to 30-dimension base on the top 15000 most accessible site across all cells. Then we used the AddHarmony to correct the batch base on donors and different cortical regions and identified the clusters using Seurat's SNN graph clustering with a default resolution of 0.8. Then peaks of each cluster were calculated using the addReproduciblePeakSet function with macs2, and differential peaks of each cluster were identified by getMarkerFeatures function.

Co-embedding of snRNA-seq cells with snATAC-seq cells

We performed co-clustering of snATAC-seq data and snRNA-seq data with R package Seurat (v3.1.1)^{23,88}. To begin with, we converted the peak by cells matrix of snATAC-seq data to gene activity score. Following by reducing the dimension with a default dimension of 50. The function of LSI is computed by the term frequency-inverse document frequency (TF-IDF) transformation. Subsequently, we pre-processed and clustered snRNA-seq dataset using default parameters as described in 'snRNA-seq data clustering'. Next, we computed anchors between the snATAC-seq dataset and the snRNA-seq dataset by FindTransferAnchors function, and used these anchors to transfer the labels from snRNA-seq cells to snATAC-seq cells. we used predictions and confidence scores for each snATAC-seq cell correspond to snRNA-seq cell to transfer the cluster IDs that defined in "Annotation of excitatory subtypes by snRNA-seq data of human cortex" session. In addition, we created snRNA-seq matrix by transfer cell type labels anchors with TransferData function. We only kept cells with the consistent cell type defined by snATAC-seq and assigned by integration analysis. Finally, we merged measured and imputed snRNA-seq data

and snATAC-seq data and run a standard UMAP analysis by reduce dimensionality to 50 to visualize all the cells together.

Transcription factor (TF) binding motif analysis of snATAC-seq data

We determined TF binding motif enrichment in accessible peaks using chromVAR (v.1.4.1)⁸⁹ on cells of either all neuron and non-neuronal cell type or excitatory neuron sub-populations. GC bias was corrected based on BSgenome.Mfasciqlaris.NCBI.5.0 genome by addGCBias function. Then we get human motifs in JASPAR database (human 2020) by getJasparMotifs function. The deviation z-scores for each TF motif in each cell was calculated by computeDeviations function, and high variance TF motifs were obtained by computeVariability function with cut-off at 1.5. For excitatory neurons, Wilcoxon test were performed on TF motif deviation z-scores between 9 subtypes and motifs with $p \leq 0.01$ were determined as differentially enriched motifs.

To identify the TF binding motif enriched along OLI maturation, we proceeded all peaks of OPC and OLI to compute variability and TF z-scores using chromVAR as described above. Then, we ordered cells by pseudotime described below and kept motifs enriched in all cells.

Analysis of peak to gene correlations across all cell types

We integrate ATAC-seq data of each cell type with the RNA-seq data of corresponding cell type through addGeneIntegrationMatrix function of ArchR, and then use the peak accessibility of ATAC-seq data and the gene expression of RNA-seq data to establish a correlation through addPeak2GeneLinks function of ArchR in each cell type. To retain reliable correlations, we select correlations for downstream analysis by using correlation coefficients greater than 0.45 and $FDR < 0.01$. The overlap of ATAC-seq peak with promoter of genes were defined as peaks within $TSS \pm 1$ kb.

TF regulatory network construction

Candidate target genes of a certain TF in one cell type were defined as those with an accessible promoter ($TSS \pm 1$ kb) or an accessible *cis*-regulatory elements containing the TF binding motif (based on JASPAR 2020 motifs). We used cytoscape (v 3.9.1) to construct a cell type specific regulatory network of TF by above information, where each point represents a TF or targeted gene, and each edge represents a link.

Stereo-seq¹⁴ library preparation and sequencing

Tissue collection and processing. Tissues sampled from two female 60-month-old cynomolgus monkeys were embedded with pre-cooled Tissue-Tek OCT (Sakura, 4583) immediately and snap-frozen in pre-chilled isopentane using liquid nitrogen until the OCT was completely solid. Embedded tissues were transferred to a -80°C freezer for long time storage. The embedded tissues were cut to a thickness of $10\ \mu\text{m}$ using a Leica CMI950 cryostat and then placed either in glass slide for Nissl staining or pre-chilled Stereo-seq capture chips for Stereo-seq procedures. Stereo-seq experiments were performed as previously described¹⁴. First, Stereo-seq capture chips were washed with $\text{NF-H}_2\text{O}$ supplemented with $0.05\ \text{U}/\mu\text{L}$ RNase inhibitor (NEB, M0314L) and dried at room temperature. Then cryosections were adhered to the surface of the Stereo-seq capture chips and incubated at 37°C for 5 min. Chips with sections were fixed in pre-cooled methanol at -20°C for 40 min.

In situ reverse transcription. After fixation, chips with tissues were taken out and dry in the air. Then chips were washed with wash buffer ($0.1 \times$ SSC buffer [Thermo, AM9770] supplemented with $0.05\ \text{U}/\mu\text{L}$ RNase inhibitor [NEB, M0314L]), tissue sections placed on the chips were permeabilized using 0.1% pepsin (Sigma, P7000) in $0.01\ \text{M}$ HCl buffer, incubated at 37°C for 6 minutes. Permeabilization reagent was then removed, and chips were washed with wash buffer. RNA released

from the permeabilized tissues and captured by the DNB was reverse transcribed at 42°C for 2 h using SuperScript II (Invitrogen, 18064-014, $10\ \text{U}/\mu\text{L}$ reverse transcriptase, $1\ \text{mM}$ dNTPs, $1\ \text{M}$ betaine solution PCR reagent, $7.5\ \text{mM}$ MgCl_2 , $5\ \text{mM}$ DTT, $2\ \text{U}/\mu\text{L}$ RNase inhibitor, $2.5\ \mu\text{M}$ Stereo-seq-TSO [5-CTGCTGACGTACTGAGAGGC/rG//rG//iXNA_G/-3]) and $1 \times$ First-Strand buffer. After that, tissues were washed twice with wash buffer and removed from the chips with tissue removal buffer ($10\ \text{mM}$ Tris-HCl, $25\ \text{mM}$ EDTA, $100\ \text{mM}$ NaCl, and 0.5% SDS) at 37°C for 30 min. The chips were then treated with Exonuclease I (NEB, M0293L) at 37°C for 1 h and washed twice with the wash buffer. The resulting first strand cDNAs on the chips were amplified using KAPA HiFi Hotstart Ready Mix (Roche, KK2602) with a $0.8\ \mu\text{M}$ cDNA-PCR primer (5-CTGCTGACGTACTGAGAGGC-3), followed by incubation at 95°C for 5 min, 15 cycles at 98°C for 20 s, 58°C for 20 s, and 72°C for 3 min, and a final incubation at 72°C for 5 min.

Library construction and sequencing. The concentrations of the result cDNA products were quantified by the Qubit™ dsDNA Assay Kit (Thermo, Q32854) after purification using VAHTS DNA Clean Beads (Vazyme, N411-03, $0.6\times$). A total of 20 ng of products were fragmented using in-house Tn5 transposase at 55°C for 10 min, and then the reaction was stopped by the addition of 0.02% SDS. Fragmented products were amplified as described below: $25\ \mu\text{L}$ of fragmentation product, $1 \times$ KAPA HiFi Hotstart Ready Mix, $0.3\ \mu\text{M}$ Stereo-seq-Library-F primer (5phos/CTGCTGACGTACTGAGAGC*A-3), and $0.3\ \mu\text{M}$ Stereo-seq-Library-R primer (5-GAGACGTTCTCGACTCAGCAGA-3) in a total volume of $100\ \mu\text{L}$ with the addition of nuclease-free H_2O . The reaction was run as: 1 cycle at 95°C for 5 minutes, 13 cycles at 98°C for 20 seconds, 58°C for 20 seconds and 72°C for 30 seconds, and 1 cycle at 72°C for 5 minutes. PCR products were then purified using VAHTS DNA Clean Beads ($0.6\times$ and $0.15\times$). Finally, the library was used for DNB generation and sequenced using MGI DNBSEQ-Tx sequencer at China National Gene Bank (CNGB) as followed: 35 bp for read 1 and 100 bp for read 2.

Stereo-seq data processing

The raw data of Stereo-seq were processed as the previous study¹⁴. Briefly, the identity of coordinates was mapped the designed coordinates matrix of the in situ captured chip which allowing 1 base mismatch and UMIs with N base or more than 2 bases with quality lower than 10 were filtered out. The retained reads were then aligned to the *macaca fascicularis* genome using STAR, Mapped reads with $\text{MAPQ} > = 10$ were counted and annotated to their corresponding genes using the handleBam¹⁴ and used to generate a gene expression profile matrix. In this study, we divided the gene expression profile matrix into non-overlapping bins covering an area of 50×50 DNB, and the transcripts of the same gene aggregated within each bin. After this step, data were normalized using SCTransform function and unsupervised clustering by Seurat.

To deconvolution of cell types within each bin of Stereo-seq, we first down sampling the cell number of each cell type to 100 in the corresponding snRNA-seq data, and then applied SPOTlight to deconvolute and mapped 20 cell types to bins of Stereo-seq sections. We set probability lower than 0.2 of each cell type as noise, then chose the highest probability out of all cell type as the identified cell type for each bin.

Inter-regional comparative analysis across cell type/subtypes

We subset the cell type of DNBelab C4 RNA-seq data, and compared the differential expressed genes (DEGs) of three regions for each cell type with R package Seurat (v3.1.1). Cell type were subset, followed by Log normalization, and we used the FindAllMarkers function to find the differentially expressed genes in three different regions of each cell type. DEGs of three regions in each cell type were defined as those with a $\text{Foldchange} > 1.5$, positive ratio over 20% and Bonferroni-adjusted $P < 0.01$. Similarly, we compared the differential expressed genes

(DEGs) of PFC and M1 in each EX-subtype and DEGs were defined as a Foldchange > 1.5, positive ratio over 20% and Bonferroni-adjusted P value < 0.05. We use the subsetArchRProject function of the R package ArchR (v1.01) to subset the cell type of snATAC-seq data and compare the differences between the regions within each subtype separately and select DA peaks with Benjamini-Hochberg-adjusted P value < 0.01 and Foldchange > 1. To exclude the sampling/batch effects of the inter-regional comparison, we performed the inter-regional DEGs and DA peaks in each donor and confirmed the consistency between donors.

The Stereo-seq dataset was converted to a Seurat object, differentially expressed genes analysis of three cortical regions for each cell type were performed with R package Seurat (v3.1.1), followed by similar processing with RNA data. Differentially expressed genes of three regions in each cell type were defined as those with Bonferroni-adjusted $P < 0.01$.

Gradient gene expression pattern analysis by Mfuzz

We subset the subtypes of excitatory neuron of DNBelab C4 RNA-seq data, and then calculate the average expression value of all genes in each EX-subtype. Subsequently, the Mfuzz package⁴⁸ (V2.54.0) was used to characterize the genes with the same gradual expression pattern. For each gene in the average expression matrix, we added 0.000001 to avoid missing value. Then, we filtered genes and standardized the expression using 'filter.std(min.std=0)' and 'standardize()', and performed M-estimation using "mestimate" function. For the gradient expression pattern across L2/3 IT, L4/5 IT, L5 IT and L6 IT types in each cortical area, the number of clusters was set to 6. For the gradient expression pattern across PFC, M1 and V1 in each EX-subtype, the number of clusters was set to 5. Genes with probability of matching the pattern > 0.5 and maximum expression > 1 were retained for further analysis. Similar clustering analysis by Mfuzz is also applied to Stereo-seq data for the gradient expression pattern across L2/3 IT, L4/5 IT, L5 IT and L6 IT types in each cortical area, genes with probability of matching the pattern > 0.5 were retained for further analysis.

Pseudotime trajectory analysis of OLI lineage datasets

To uncover the dynamic change of both transcriptome and chromatin accessibility along the oligodendrocyte lineage. We constructed the development trajectory using oligodendrocytes and oligodendrocyte precursor cells. For RNA-seq data, monocle 2⁹⁰ (v2.14.0) was used to order 11,706 OLI cells and 5552 OPCs from the single nuclei RNA-seq data set to developmental trajectory. TPM matrix of OLI cells and OPCs was input and then transformed into normalized mRNA counts using the "relative2abs" function. Genes used for ordering the cells along the trajectory were selected under the following criteria: mean_expression >= 0.5; dispersion_empirical >= 1. After that, the discriminative dimensionality reduction with trees (DDRTree) method was used to reduce data into two dimensions. The first 2000 genes significantly varied across pseudotime were visualized using heatmap and gene ontologies were identified using metascap (https://metascap.org).

For ATAC-seq data, we used monocle 2 to plot a heatmap of gene activity over pseudo time in OPCs differentiation into OLI. We converted the peak by cells matrix to the gene activity score matrix by custom distance-weighted accessibility models in ArchR. To determine peaks that were differentially accessed between the different cell states, we applied differentialGeneTest function from monocle 2 to gene activity score matrix, then used top 1000 genes with the smallest p value to construct a trajectory. DDRTree is used to reduce dimensions and to reconstruct the temporal progression. Finally, we used genes with q -values less than 0.01 to plot heatmap of dynamics genes activity along trajectory. To establish a heatmap of TF motifs enrichment over pseudotime, the cells were grouped into 10 bins based on their trajectory values.

Evaluation of heritability and dysregulated genes enrichment within cell-types

We used the LDSC [https://github.com/bulik/ldsc] to predict enrichments of diseases SNP-heritability in differentially accessible peaks for each cell type or subtype in macaque and human. For macaque data, we used the liftOver tool (v1.6.0) [https://genome.ucsc.edu/cgi-bin/hgLiftOver] to lift the differentially accessible peaks of each cell type and all accessible peaks to human hg19 genome. For human dataset⁴³, peak matrix and fragments of snATAC-seq from control group was downloaded. To create chrom assay, we input peak matrix and fragments into Signac (v1.6.0) using CreateChromatinAssay function with default parameters (min.cells = 10, min.features = 200, genome = hg38), and then input the result data into Seurat and use the CreateSeuratObject function to create a Seurat object for downstream analysis. Neuron subtypes were assigned by co-embedded with human brain single nuclei RNA-seq dataset (https://portal.brain-map.org/atlas-and-data/rnaseq/human-multiple-cortical-areas-smart-seq) as described in 'Co-embedding of snRNA-seq cells with snATAC-seq cells' with parameter k.anchor = 150 in the FindTransferAnchors function. The maximum prediction score ≥ 0.4 in 84% of cells. We calculated the DARs (differentially accessible regions) for each cell type as previous described¹⁶, taking into account any technical artifacts related to the total accessibility of each cell. We converted DARs with P -values less than 0.01 and LogFC greater than 1 to hg19 format for LDSC analysis using the UCSC liftover tool [https://genome.ucsc.edu/cgi-bin/hgLiftOver]. Then, LD scores of SNPs of 1000 genomes phase 3 in differentially accessible peaks were estimated in human and macaque according to recommended workflow. We generated the summary statistics from previous GWAS studies reported in publications (Supplementary Data 12) and UK Biobank (https://www.ukbiobank.ac.uk/) and calculate the significant enrichment of each diseases following the "cell type specific analysis" workflow recommend by the LDSC authors.

To capture the perturbation of neurology disease in different cortical region of Stereo-seq section, we used a similar analysis as previous study¹². Briefly, we adapted the homologous in *Macaca fascicularis* genome for up-regulated genes and down-regulated genes patients of ASD, bipolar disorder (BD) and SCZ from recent large scale postmortem brain datasets⁷¹ (from DGE - Differential Gene Expression), and filtered out the genes with the FDR > 0.05 or not captured by the Stereo-seq. Differentially expressed genes of cell types in each cortical region were calculated by Seurat and using the cutoff of P values < 0.05, then the significance of each disease associated gene set in each cell type were calculated by hypergeometric test, and the P values < 0.05 were considered as significant enrichment in the corresponding cell type.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The raw data generated in this study have been deposited in CNGB Nucleotide Sequence Archive (CNSA: https://db.cngb.org/cnsa) under accession code CNP0000927. We have also provided the MBA website (https://db.cngb.org/mba), an open and interactive database for exploration. The public datasets used in this study can be accessed as described below: Allen Cell Types Database-Human Multiple Cortical Areas is available at https://portal.brain-map.org/atlas-and-data/rnaseq/human-multiple-cortical-areas-smart-seq. snATAC-seq data of adult human PFC is available at https://www.synapse.org/#!Synapse:syn22079621/. RNA ISH images for genes expressed in primary visual cortex of macaque brain is available at NIH Blueprint Non-Human Primate (NHP) Atlas: GFAP: http://www.blueprintnpatlas.org/ish/experiment/show/100140483, GPR83: http://www.blueprintnpatlas.org/ish/gene/show/183031, RORB: http://www.blueprintnpatlas.org/

ish/gene/show/183109, PDE1A: <http://www.blueprintnpatlas.org/ish/gene/show/183138> and SYT6 <http://www.blueprintnpatlas.org/ish/experiment/show/100091672>. Summary statistics files for each human trait were downloaded from the UK Biobank database or published studies (data links in Supplementary Data 12). The JASPAR database (2020) for human TF motif is available at <http://www.bioconductor.org/packages/release/data/annotation/html/JASPAR2020.html>. Source data are provided with this paper.

Code availability

All data were analyzed with standard programs and packages, as detailed above. Custom code using open-source software supporting the current study are available at <https://github.com/single-cell-BGI/MBA>.

References

- Bernard, A. et al. Transcriptional architecture of the primate neocortex. *Neuron* **73**, 1083–1099 (2012).
- Zhu, Y. et al. Spatiotemporal transcriptomic divergence across human and macaque brain development. *Science* **362**, <https://doi.org/10.1126/science.aat8077> (2018).
- Hunt, K. D. The single species hypothesis: truly dead and pushing up bushes, or still twitching and ripe for resuscitation? *Hum. Biol.* **75**, 485–502 (2003).
- Kang, Y., Chu, C., Wang, F. & Niu, Y. CRISPR/Cas9-mediated genome editing in nonhuman primates. *Dis Model Mech* **12**, <https://doi.org/10.1242/dmm.039982> (2019).
- Chansel-Debordeaux, L. & Bezar, E. Local transgene expression and whole-body transgenesis to model brain diseases in nonhuman primate. *Anim. Model Exp. Med* **2**, 9–17 (2019).
- Chen, Y. et al. Modeling rett syndrome using TALEN-Edited MECP2 mutant cynomolgus monkeys. *Cell* **169**, 945–955.e910 (2017).
- Zhang, W. et al. SIRT6 deficiency results in developmental retardation in cynomolgus monkeys. *Nature* **560**, 661–665 (2018).
- Koprach, J. B., Johnston, T. H., Reyes, G., Omana, V. & Brotchie, J. M. Towards a non-human primate model of alpha-synucleinopathy for development of therapeutics for Parkinson's disease: optimization of AAV1/2 delivery parameters to drive sustained expression of alpha synuclein and dopaminergic degeneration in macaque. *PLoS One* **11**, e0167235 (2016).
- Khrameeva, E. et al. Single-cell-resolution transcriptome map of human, chimpanzee, bonobo, and macaque brains. *Genome Res* **30**, 776–789 (2020).
- Preissl, S. et al. Single-nucleus analysis of accessible chromatin in developing mouse forebrain reveals cell-type-specific transcriptional regulation. *Nat. Neurosci.* **21**, 432–439 (2018).
- Yin, S. et al. Transcriptomic and open chromatin atlas of high-resolution anatomical regions in the rhesus macaque brain. *Nat. Commun.* **11**, 474 (2020).
- Maynard, K. R. et al. Transcriptome-scale spatial gene expression in the human dorsolateral prefrontal cortex. *Nat. Neurosci.* **24**, 425–436 (2021).
- Ortiz, C. et al. Molecular atlas of the adult mouse brain. *Sci. Adv.* **6**, eabb3446 (2020).
- Chen, A. et al. Spatiotemporal transcriptomic atlas of mouse organogenesis using DNA nanoball-patterned arrays. *Cell* **185**, 1777–1792.e1721 (2022).
- Chuanyu L., et al. A portable and cost-effective microfluidic system for massively parallel single-cell transcriptome profiling. <https://www.biorxiv.org/content/10.1101/818450v3> (2019).
- Bakken, T. E. et al. Comparative cellular analysis of motor cortex in human, marmoset and mouse. *Nature* **598**, 111–119 (2021).
- Tasic, B. et al. Shared and distinct transcriptomic cell types across neocortical areas. *Nature* **563**, 72–78 (2018).
- Hodge, R. D. et al. Conserved cell types with divergent features in human versus mouse cortex. *Nature* **573**, 61–68 (2021).
- Network, B. I. C. C. A multimodal cell census and atlas of the mammalian primary motor cortex. *Nature* **598**, 86–102 (2021).
- Yao, Z. et al. A taxonomy of transcriptomic cell types across the isocortex and hippocampal formation. *Cell* <https://doi.org/10.1016/j.cell.2021.04.021> (2021).
- Yao, Z. et al. A transcriptomic and epigenomic cell atlas of the mouse primary motor cortex. *Nature* **598**, 103–110 (2021).
- Granja, J. M. et al. ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. *Nat. Genet* **53**, 403–411 (2021).
- Stuart, T. et al. Comprehensive integration of single-cell data. *Cell* **177**, 1888–1902.e1821 (2019).
- Roman Spektor, J. W. Y., Seoyeon L., & Soloway P. D. Single cell ATAC-seq identifies broad changes in neuronal abundance and chromatin accessibility in Down Syndrome. <https://www.biorxiv.org/content/10.1101/561191v1> (2019).
- Nott, A. et al. Brain cell type-specific enhancer-promoter interactome maps and disease-risk association. *Science* **366**, 1134–1139 (2019).
- Zhang, Y. et al. Rapid single-step induction of functional neurons from human pluripotent stem cells. *Neuron* **78**, 785–798 (2013).
- Baroti, T. et al. Transcription factors Sox5 and Sox6 exert direct and indirect influences on oligodendroglial migration in spinal cord and forebrain. *Glia* **64**, 122–138 (2016).
- Turnescu, T. et al. Sox8 and Sox10 jointly maintain myelin gene expression in oligodendrocytes. *Glia* **66**, 279–294 (2018).
- Hornig, J. et al. The transcription factors Sox10 and Myrf define an essential regulatory network module in differentiating oligodendrocytes. *PLoS Genet* **9**, e1003907 (2013).
- Smith, A. M. et al. The transcription factor PU.1 is critical for viability and function of human brain micro. *Glia. Glia* **61**, 929–942 (2013).
- Subramanian, L. et al. Transcription factor Lhx2 is necessary and sufficient to suppress astroglialogenesis and promote neurogenesis in the developing hippocampus. *Proc. Natl. Acad. Sci. USA* **108**, E265–E274 (2011).
- De Val, S. et al. Combinatorial regulation of endothelial gene expression by ets and forkhead transcription factors. *Cell* **135**, 1053–1064 (2008).
- Manuel, M. N., Mi, D., Mason, J. O. & Price, D. J. Regulation of cerebral cortical neurogenesis by the Pax6 transcription factor. *Front Cell Neurosci.* **9**, 70 (2015).
- Golonzhka, O. et al. Pbx regulates patterning of the cerebral cortex in progenitors and postmitotic neurons. *Neuron* **88**, 1192–1207 (2015).
- Arimatsu, Y., Ishida, M., Kaneko, T., Ichinose, S. & Omori, A. Organization and development of corticocortical associative neurons expressing the orphan nuclear receptor Nurr1. *J. Comp. Neurol.* **466**, 180–196 (2003).
- Zhang, K. et al. Imbalance of excitatory/inhibitory neuron differentiation in neurodevelopmental disorders with an NR2F1 point mutation. *Cell Rep.* **31**, 107521 (2020).
- Bunt, J. et al. Combined allelic dosage of Nfia and Nfib regulates cortical development. *Brain Neurosci. Adv.* **1**, 2398212817739433 (2017).
- Ziffra, R. S. et al. Single-cell epigenomics reveals mechanisms of human cortical development. *Nature* **598**, 205–213 (2021).
- Trevino, A. E. et al. Chromatin and gene-regulatory dynamics of the developing human cerebral cortex at single-cell resolution. *Cell* **184**, 5053–5069.e5023 (2021).
- Bakken, T. E. et al. A comprehensive transcriptional map of primate brain development. *Nature* **535**, 367–375 (2016).
- Lake, B. B. et al. Neuronal subtypes and diversity revealed by single-nucleus RNA sequencing of the human brain. *Science* **352**, 1586–1590 (2016).

42. Nowakowski, T. J. et al. Spatiotemporal gene expression trajectories reveal developmental hierarchies of the human cortex. *Science* **358**, 1318–1323 (2017).
43. Morabito, S. et al. Single-nucleus chromatin accessibility and transcriptomic characterization of Alzheimer's disease. *Nat. Genet* **53**, 1143–1155 (2021).
44. McColgan, P., Joubert, J., Tabrizi, S. J. & Rees, G. The human motor cortex microcircuit: insights for neurodegenerative disease. *Nat. Rev. Neurosci.* **21**, 401–415 (2020).
45. Sasaki, T., Komatsu, Y., Watakabe, A., Sawada, K. & Yamamori, T. Prefrontal-enriched SLIT1 expression in Old World monkey cortex established during the postnatal development. *Cereb. Cortex* **20**, 2496–2510 (2010).
46. Simonetti, M. et al. The impact of Semaphorin 4C/Plexin-B2 signaling on fear memory via remodeling of neuronal and synaptic morphology. *Mol. Psychiatry* **26**, 1376–1398 (2021).
47. Gilabert-Juan, J. et al. Semaphorin and plexin gene expression is altered in the prefrontal cortex of schizophrenia patients with and without auditory hallucinations. *Psychiatry Res.* **229**, 850–857 (2015).
48. Kumar, L. & M, E. F. Mfuzz: a software package for soft clustering of microarray data. *Bioinformatics* **2**, 5–7 (2007).
49. Doostparast Torshizi, A. et al. Deconvolution of transcriptional networks identifies TCF4 as a master regulator in schizophrenia. *Sci. Adv.* **5**, eaau4139 (2019).
50. Elosua-Bayes, M., Nieto, P., Mereu, E., Gut, I. & Heyn, H. SPOTlight: seeded NMF regression to deconvolute spatial transcriptomics spots with single-cell transcriptomes. *Nucleic Acids Res* <https://doi.org/10.1093/nar/gkab043> (2021).
51. Sekiguchi, M. et al. ARHGAP10, which encodes Rho GTPase-activating protein 10, is a novel gene for schizophrenia risk. *Transl. Psychiatry* **10**, 247 (2020).
52. Harris, H. K. et al. Disruption of RFX family transcription factors causes autism, attention-deficit/hyperactivity disorder, intellectual disability, and dysregulated behavior. *Genet Med.* **23**, 1028–1040 (2021).
53. Shinohara, M., Tachibana, M., Kanekiyo, T. & Bu, G. Role of LRP1 in the pathogenesis of Alzheimer's disease: evidence from clinical and preclinical studies. *J. Lipid Res.* **58**, 1267–1281 (2017).
54. de Pins, B., Mendes, T., Giralt, A. & Girault, J. A. The non-receptor tyrosine kinase Pyk2 in brain function and neurological and psychiatric diseases. *Front Synaptic Neurosci.* **13**, 749001 (2021).
55. Simons, M. & Nave, K. A. Oligodendrocytes: myelination and axonal support. *Cold Spring Harb. Perspect. Biol.* **8**, a020479 (2015).
56. Emery, B. & Lu, Q. R. Transcriptional and epigenetic regulation of oligodendrocyte development and myelination in the central nervous system. *Cold Spring Harb. Perspect. Biol.* **7**, a020461 (2015).
57. Peng, Z. et al. Experimental autoimmune encephalomyelitis (EAE) model of cynomolgus macaques induced by recombinant human MOG1-125 (rhMOG1-125) protein and MOG34-56 peptide. *Protein Pept. Lett.* **24**, 1166–1178 (2018).
58. Haanstra, K. G. et al. Induction of experimental autoimmune encephalomyelitis with recombinant human myelin oligodendrocyte glycoprotein in incomplete Freund's adjuvant in three non-human primate species. *J. Neuroimmune Pharm.* **8**, 1251–1264 (2013).
59. McFarland, H. I. et al. Determinant spreading associated with demyelination in a nonhuman primate model of multiple sclerosis. *J. Immunol.* **162**, 2384–2390 (1999).
60. Marques, S. et al. Oligodendrocyte heterogeneity in the mouse juvenile and adult central nervous system. *Science* **352**, 1326–1329 (2016).
61. Jakel, S. et al. Altered human oligodendrocyte heterogeneity in multiple sclerosis. *Nature* **566**, 543–547 (2019).
62. Lake, B. B. et al. Integrative single-cell analysis of transcriptional and epigenetic states in the human adult brain. *Nat. Biotechnol.* **36**, 70–80 (2018).
63. Dai, Z. M. et al. Stage-specific regulation of oligodendrocyte development by Wnt/beta-catenin signaling. *J. Neurosci.* **34**, 8467–8473 (2014).
64. Sun, L. O. et al. Spatiotemporal Control of CNS Myelination by Oligodendrocyte Programmed Cell Death through the TFEB-PUMA Axis. *Cell* **175**, 1811–1826.e1821 (2018).
65. Le Hellard, S. et al. Polymorphisms in SREBF1 and SREBF2, two antipsychotic-activated transcription factors controlling cellular lipogenesis, are associated with schizophrenia in German and Scandinavian samples. *Mol. Psychiatry* **15**, 463–472 (2010).
66. Cusanovich, D. A. et al. A Single-Cell Atlas of In Vivo Mammalian Chromatin Accessibility. *Cell* **174**, 1309–1324.e1318 (2018).
67. Mathys, H. et al. Single-cell transcriptomic analysis of Alzheimer's disease. *Nature* **570**, 332–337 (2019).
68. Grubman, A. et al. A single-cell atlas of entorhinal cortex from individuals with Alzheimer's disease reveals cell-type-specific gene expression regulation. *Nat. Neurosci.* **22**, 2087–2097 (2019).
69. Skene, N. G. et al. Genetic identification of brain cell types underlying schizophrenia. *Nat. Genet* **50**, 825–833 (2018).
70. Howard, D. M. et al. Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. *Nat. Neurosci.* **22**, 343–352 (2019).
71. Gandal, M. J. et al. Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder. *Science* **362**, <https://doi.org/10.1126/science.aat8127> (2018).
72. Velmeshev, D. et al. Single-cell genomics identifies cell type-specific molecular changes in autism. *Science* **364**, 685–689 (2019).
73. W. Brad Ruzicka, S. M., Davila-Velderrain J., Subburaju S., Reed Tso D., Hourihan M., & Kellis M. Single-cell dissection of schizophrenia reveals neurodevelopmental-synaptic axis and transcriptional resilience. <https://www.medrxiv.org/content/10.1101/2020.11.06.20225342v1> (2020).
74. Verdier, J. M. et al. Lessons from the analysis of nonhuman primates for understanding human aging and neurodegenerative diseases. *Front Neurosci.* **9**, 64 (2015).
75. Eslamboli, A. et al. Long-term consequences of human alpha-synuclein overexpression in the primate ventral midbrain. *Brain* **130**, 799–815 (2007).
76. Burns, L. H. et al. Selective putaminal excitotoxic lesions in non-human primates model the movement disorder of Huntington disease. *Neuroscience* **64**, 1007–1017 (1995).
77. Ferrante, R. J., Kowall, N. W., Cipolloni, P. B., Storey, E. & Beal, M. F. Excitotoxin lesions in primates as a model for Huntington's disease: histopathologic and neurochemical characterization. *Exp. Neurol.* **119**, 46–71 (1993).
78. Williamson, J. M. & Lyons, D. A. Myelin dynamics throughout life: an ever-changing landscape? *Front Cell Neurosci.* **12**, 424 (2018).
79. Krishnaswami, S. R. et al. Using single nuclei for RNA-seq to capture the transcriptome of postmortem neurons. *Nat. Protoc.* **11**, 499–524 (2016).
80. Niu, Y. et al. Dissecting primate early post-implantation development using long-term in vitro embryo culture. *Science* **366**, <https://doi.org/10.1126/science.aaw5754> (2019).
81. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet* **17**, 10–12 (2011).
82. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
83. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinforma.* **12**, 323 (2011).

84. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420 (2018).
85. Cusanovich, D. A. et al. Multiplex single cell profiling of chromatin accessibility by combinatorial cellular indexing. *Science* **348**, 910–914 (2015).
86. Amini, S. et al. Haplotype-resolved whole-genome sequencing by contiguity-preserving transposition and combinatorial indexing. *Nat. Genet.* **46**, 1343–1349 (2014).
87. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
88. Chung, C. Y. et al. Single-Cell Chromatin Analysis of Mammary Gland Development Reveals Cell-State Transcriptional Regulators and Lineage Relationships. *Cell Rep* **29**, 495–510 e496 (2019).
89. Schep, A. N., Wu, B., Buenrostro, J. D. & Greenleaf, W. J. chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. *Nat. Methods* **14**, 975–978 (2017).
90. Qiu, X. et al. Single-cell mRNA quantification and differential analysis with Census. *Nat. Methods* **14**, 309–315 (2017).
- M.C., Y.W., Z.W., J. Xie, Y.Y., M.W., J. Xu., C.L., C. Wang, D.C., Ms. L.W., and Liqin X. performed the experiments; Mr. L.W., Z.Z., Z.L., Y.S., Z.H., Shang L., T.P., C. Ward. and Y. Lai. performed the computational analysis and figure preparation; H. Yu, H.S., and Q.Y. conceived the data website; Y. Lei., Shiping L., L.L., H.Z., B.T., M.A.E., G.V. and L.H. interpreted the data, W. J., H. Yang, T.T., K.C., M.D., Y.G., K.M., A.C., Y.Li., Z.S., G.L., L.C. and F.L. provided relevant advice and review of the manuscript. Y. Lei. wrote the manuscript with input from all authors. All authors read and approved the manuscript for submission.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-022-34413-3>.

Correspondence and requests for materials should be addressed to Yuyu Niu, Hongkui Zeng, Yong Hou, Longqi Liu, Shiping Liu or Xun Xu.

Peer review information *Nature Communications* thanks the anonymous reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022

Acknowledgements

We sincerely thank the technical support provided by China National Gene Bank. This study was supported by National Science and Technology Innovation 2030 Major Program (2021ZD0200100), National Key Research and Development Program (2018YFA0801400 and 2021YFA0805700), National Natural Science Foundation of China (No. 31900582), Shenzhen Key Laboratory of Single Cell Omics (ZDSYS20190902093613831), Shenzhen Bay Laboratory (SZBL2019062801012), Guangdong Provincial Key Laboratory of Genome Read and Write (2017B030301011), Shenzhen Basic Research Project for Excellent Young Scholars (No. 2020251518), Shenzhen Science and Technology Innovation Fund (Grant KQJSCX20180329191008922) and Major Basic Research Project of Science and Technology of Yunnan (2018FA020 and 202001BC070001). Miguel A. Esteban's laboratory at the Guangzhou Institutes of Biomedicine and Health was supported by the Strategic Priority Research Program of the Chinese Academy of Sciences (XDA16030502) and the Innovative Team Program of Bioland Laboratory (Guangzhou Regenerative Medicine and Health Guangdong Laboratory) (2018GZR110103001).

Author contributions

X.X., Shiping L., L.L., Y.H., Y.N., and Y. Lei. conceived and supervised the study. H.Z., L.L. and Y. Lei. designed the experiments; Y. Lei., L.H., J.W., Liangzhi X., M.C., C. Wong, W. L. and Liqin X. collected the samples;

Ying Lei^{1,2,3,20}, Mengnan Cheng^{1,2,20}, Zihao Li^{1,4,20}, Zhenkun Zhuang^{1,5,20}, Mr. Liang Wu^{2,20}, Yunong sun^{1,4,20}, Lei Han^{1,2,3}, Zhihao Huang², Yuzhou Wang², Zifei Wang², Liqin Xu², Yue Yuan¹, Shang Liu², Taotao Pan², Jiarui Xie², Chuanyu Liu¹, Giacomo Volpe⁶, Carl Ward⁷, Yiwei Lai^{1,7}, Jiangshan Xu¹, Mingyue Wang¹, Hao Yu², Haixi Sun², Qichao Yu^{2,4}, Ms. Liang Wu², Chunqing Wang^{2,4}, Chi Wai Wong⁸, Wei Liu⁸, Liangzhi Xu⁸, Jingkuan Wei⁹, Dandan Chen¹⁰, Zhouchun Shang^{2,11}, Guibo Li², Kun Ma², Le Cheng¹², Fei Ling⁵, Tao Tan⁹, Kai Chen⁹, Bosiljka Tasic¹³, Michael Dean¹⁴, Weizhi Ji⁹, Huanming Yang^{2,15}, Ying Gu^{1,2,4}, Miguel A. Esteban^{1,7,16}, Yuxiang Li², Ao Chen², Yuyu Niu^{9,17,21} ✉, Hongkui Zeng^{13,21} ✉, Yong Hou^{2,3,18,21} ✉, Longqi Liu^{1,2,3,18,21} ✉, Shiping Liu^{1,2,3,18,21} ✉ & Xun Xu^{1,2,19,21} ✉

¹BGI-Hangzhou, Hangzhou 310012, China. ²BGI-Shenzhen, Shenzhen 518083, China. ³Shenzhen Bay Laboratory, Shenzhen 518000, China. ⁴College of Life Sciences, University of Chinese Academy of Sciences, Beijing 100049, China. ⁵School of Biology and Biological Engineering, South China University of Technology, Guangzhou 510006, China. ⁶Hematology and Cell Therapy Unit, IRCCS-Istituto Tumori "Giovanni Paolo II". Viale O Flacco 65, 70124 Bari, Italy. ⁷Laboratory of Integrative biology, Guangzhou Institutes of Biomedicine and Health, CAS, Guangzhou 510530, China. ⁸Huazhen Biosciences, Guangzhou

510900, China. ⁹Yunnan Key Laboratory of Primate Biomedical Research, Institute of Primate Translational Medicine, Kunming University of Science and Technology, Kunming 650500, China. ¹⁰BGI-GenImmune, BGI-Shenzhen, Wuhan 430074, China. ¹¹MGI, BGI-Shenzhen, Shenzhen 518083, China. ¹²BGI-Yunnan, BGI-Shenzhen, Kunming 650106, China. ¹³Allen Institute for Brain Science, Seattle, WA 98109, USA. ¹⁴Cancer and Inflammation Program, National Cancer Institute at Frederick, Building 560, Frederick, MD 21702, USA. ¹⁵James D. Watson Institute of Genome Sciences, Hangzhou 310058, China. ¹⁶Institute for Stem Cells and Regeneration, Chinese Academy of Sciences, 100101 Beijing, China. ¹⁷Faculty of Life Science and Technology, Kunming University of Science and Technology, Kunming, Yunnan 650500, China. ¹⁸Shenzhen Key Laboratory of Single-Cell Omics, BGI-Shenzhen, Shenzhen 518120, China. ¹⁹Guangdong Provincial Key Laboratory of Genome Read and Write, BGI-Shenzhen, Shenzhen 518120, China. ²⁰These authors contributed equally: Ying Lei, Mengnan Cheng, Zihao Li, Zhenkun Zhuang, Liang Wu, Yunong sun. ²¹These authors jointly supervised this work: Yuyu Niu, Hongkui Zeng, Yong Hou, Longqi Liu, Shiping Liu, Xun Xu. ✉ e-mail: niuyy@lpbr.cn; hongkuiz@alleninstitute.org; huyong@genomics.cn; liulongqi@genomics.cn; liushiping@genomics.cn; xuxun@genomics.cn