Single-cell spatial map of cis-regulatory elements for disease-related genes in the macaque cortex

Juan Meng[1,2,#], Cheng Chen[2,3,4,#], Zhiyong Zhu[3,4,#], Yongkang Sun[1,#], Yiming Huang[5,6,#], Kaijie Hu[1,#], Jiqiang Fu[1], Luyan Wu[1], Ling Li[1,2], Yiqin Bai[1], Tianyi Fei[6], Zhen Liu[1,7], Chao Li[1], Zhiming Shen[1,7], Longqi Liu[3,4], Chengyu Li[5], Tao Song[1], Cirong Liu[1], Muming Poo[1], Shiping Liu[3,4,*], Ying Lei[3,4*], and Yidi Sun[1,*]

[1]Institute of Neuroscience, CAS Center for Excellence in Brain Science and Intelligence Technology, State Key Laboratory of Genetic Evolution & Animal Models, Institute of Neuroscience, Shanghai Key Laboratory of Precision Gene Editing and Clinical Translation, Chinese Academy of Sciences, Shanghai 200031, China.

[2]University of Chinese Academy of Sciences, Beijing 100049, China

[3]State Key Laboratory of Genome and Multi-omics Technologies, BGI Research, Hangzhou 310030, China

[4]Key Laboratory of Spatial Omics of Zhejiang Province, BGI Research, Hangzhou 310030, China.

[5]Lingang Laboratory, Shanghai 200031, China

[6]School of Life Sciences and Technology, ShanghaiTech University, Shanghai 201210, China.

[7]Shanghai Center for Brain Science and Brain-Inspired Technology, Shanghai 201602, China.

[#] These authors contributed equally to this work

*Correspondence: Shiping Liu: liushiping@genomics.cn, Ying Lei: leiying1@genomics.cn and Yidi Sun: ydsun@ion.ac.cn

Abstract

Single-cell spatial transcriptomes have demonstrated molecular and cellular diversity in the brain[1, 2, 3, 4, 5, 6, 7, 8], but gene regulatory mechanisms underlying transcriptomic profiles and disease pathogenesis remain largely unknown in primates. Here we performed single-nucleus Assay for Transposase-Accessible Chromatin followed by sequencing (snATAC-seq) for ~1.6 million cells from 142 cortical regions of two male cynomolgus monkeys (Macaca fascicularis), and identified distinct chromatin accessibility profiles of cis-regulatory elements (CREs) for various cell types. By integrative analysis with large-scale spatial transcriptome data, we found that these CREs showed laminar and regional preferences, with their regional accessibility exhibiting striking dependence on the region's hierarchical level. Cross-species comparison of snATAC-seq data revealed human/macaque-enriched layer-4 glutamatergic neurons and LAMP5/LHX6-expressing GABAergic neurons as well as human/macaque-biased CREs for genes related to neurodevelopment and psychiatric diseases. Importantly, risk single-nucleotide polymorphisms for many brain disorders strongly associated with human/macaque-biased CREs in glutamatergic

neuronal types and those for Alzhemer's disease strongly associated with CREs exclusively in microglia. Our results provided the basis for understanding the spatial gene regulatory mechanisms underlying cellular diversity and disease pathogenesis in the primate cortex.

Introduction

Single-cell transcriptome analyses have facilitated our understanding of the diversity of cell types and their spatial distributions in the brain. Hundreds of glutamatergic, GABAergic, and glial cell types have now been identified[1, 2, 3, 4, 5, 6, 7, 8]. However, the gene regulatory mechanisms underlying their identity and function remain largely unclear. The spatiotemporal pattern of gene expression depends on the chromatin accessibility of cis-regulatory elements (CREs), which are binding sites for sequence-specific transcription factors (TFs) that recruit proteins for transcriptional initiation of their target genes. A comprehensive characterization of CREs at the single-cell level will help to elucidate gene regulatory mechanisms in diverse cell types and the molecular basis of brain development, function and diseases.

Recent studies using the snATAC-seq method have enabled the analysis of cell type-specific CREs in the adult mouse brain[9, 10, 11, 12]. Previous bulk sequencing assays have revealed that enhancer elements, the main component of CREs, exhibit rapid changes during mammalian evolution[13] and primate-specific enhancer elements were found to be involved in brain diseases[14]. Thus, new CREs could appear during primate evolution and specifically responsible for human diseases. Indeed, comparative studies of single-cell chromatin accessibility datasets have shown conserved chromatin accessibility and CREs in human, monkey and mouse brains[15, 16]. However, whether primate-specific CREs underlie transcriptomic profiles of primate-enriched cell types and human disease pathogenesis remains largely unknown. Moreover, chromatin accessibility in bulk tissues was found to exhibit region-selective patterns in the primate brain[17] and single-cell chromatin accessibility has been examined for a limited number of primate brain regions[15, 18, 19]. Thus, a comprehensive characterization of single-cell chromatin accessibility covering the entire macaque cortex will help to advance our understanding of CREs underlying the diversity of cell types and their regional distribution in the primate brain.

In this study, we generated a single-nucleus chromatin accessibility dataset for ~1.6 million

single nuclei from 142 cortical regions of the adult macaque brain using the snATAC-seq method. We defined 230 clusters of cells based on this chromatin accessibility atlas and identified 615,873 candidate CREs associated with various cell types in the macaque brain. Mapping of CREs with the spatial transcriptome showed their laminar and regional preferences. Further cross-species comparison of single-cell chromatin accessibility data from macaques, mice, and humans revealed human/macaque-biased CREs for psychiatric disorder-associated genes in layer-4 (L4) glutamatergic cell types and for neurogenesis-related genes in GABAergic neurons. Notably, we also found a strong association between human/macaque-biased CREs in specific cell types and risk single nucleotide polymorphisms (SNPs) for various brain diseases. These results provide a comprehensive database for studying gene regulatory programs at the single-cell level and for understanding pathogenic mechanisms underlying brain disorders. An interactive website for accessing our dataset is available through https://macaque.digital-brain.cn/chromatin-accessibility.

Results

Characterization of chromatin accessibility in various macaque cortical cell types

Processing of snATAC-seq data. To understand gene regulatory programs underlying cellular diversity in the primate cortex, we performed chromatin accessibility profiling using a previously reported droplet-based snATAC-seq method[20] for tissue samples from 142 cortical regions (128 from Macaque #1 and 132 from Macaque #2) covering the entire cortex of two adult male cynomolgus monkeys (*Macaca fascicularis*) (Fig. 1a and Supplementary Table 1). The data reliability was confirmed by the high numbers of uniquely mapped reads, high proportion of properly paired reads, high numbers of unique fragments, low doublet ratio, and consistent quality metrics between biological replicates (Extended Data Fig. 1a-e). A total of 1,615,430 nuclei passed quality control criteria including the number of unique fragments per cell, TSS enrichment score, promoter ratio and doublet removal (Methods and Extended Data Fig. 1f, g). The remaining cells exhibited an average of 20,534 chromatin fragments in each nucleus and 57.6% of reads in peak regions. These data were used for unsupervised clustering analysis, and assisted by previously reported spatial transcriptome dataset[21], our chromatin accessibility data were used for further spatial mapping of CREs and gene regulatory networks (GRNs; Fig. 1a).

Clustering analysis of chromatin accessibility profiles. Unsupervised clustering based on the representations after iterative latent semantic indexing (LSI) of chromatin accessibility in each cell resulted in three major types, glutamatergic (726,611; 45%), GABAergic (275,073; 17%) and non-neuronal (613,746; 38%) cells. Further iterative clustering of each major type revealed 91 subclusters (14 clusters) of glutamatergic neurons, 105 subclusters (8 clusters) of GABAergic neurons, and 34 subclusters (7 clusters) of non-neuronal cells (Fig. 1b and Supplementary Table 2). The major clusters and subclusters were annotated by the peak accessibility at the promoter region of their canonical marker genes[6, 22, 23]. For examples, high chromatin accessibility was observed at the promoter region of *SLC17A7* for all glutamatergic neurons and *GAD2* for GABAergic neurons. Clusters associated with astrocytes (ASCs), oligodendrocyte progenitor cells (OPCs), oligodendrocytes (OLGs) and microglia cells (MGs) showed open chromatin peaks at the promoter of *GFAP*, *PDGFRA*, *PLP1* and *PLAC8*, respectively (Fig. 1c). Cells from multiple cortical areas of two monkeys were present in the majority of cell types (Fig. 1b), and high consistency in the proportions of various clusters was found between biological replicates (Extended Data Fig. 1h-k), demonstrating the reliability of the batch correction. Hierarchical clustering of the averaged LSI representations in each cluster resulted in a taxonomy tree (Extended Data Fig. 1l). Integrative analysis of snATAC-seq and snRNA-seq data showed that the clusters of chromatin accessibility profiles identified by snATAC-seq corresponded well with snRNA-seq-defined cell types (Extended Data Fig. 2a, b and Supplementary Table 2).

Identification of open chromatin regions in each snATAC-seq clusters. We further applied the MACS2 method[24] to identify open chromatin regions for each cell type by aggregating chromatin accessibility profiles (Methods). After filtering out those with low confidence, we detected an average of 138,477 open chromatin regions per cell type and a total of 615,873 open chromatin regions in all cell types (Extended Data Fig. 2c). Among these open chromatin regions, more than 90% were located ±1 kb away from known transcriptional start site (TSS) of protein-coding and long noncoding RNA genes, consistent with previous data for human and mouse brains[11, 15, 18, 19] (Extended Data Fig. 2d). Interestingly, with comparable numbers of chromatin fragments among different cell types, the proportions of open chromatin peaks enriched in the TSS or promoter regions (more adjacent cis-regions) in non-neuronal cells and GABAergic cells were much higher than those in glutamatergic cells (Extended Data Fig. 2e). Our analysis of previously

published datasets of human, macaques and mice[11, 15, 18, 19] showed the same phenomenon (Extended Data Fig. 2e). The lower TSS enrichment in glutamatergic neurons might be induced by higher activity of glutamatergic neurons or more distant regulatory elements from the TSS in glutamatergic cells. As a previous study[25] has shown that neuronal activity induces a major reconfiguration of the chromatin state, with most of the activity induced open regions located in the intergenic and intronic regions.

Identification of CREs and their linked genes in various cell types. We thus defined these distal open chromatin regions as cis-regulatory elements (CREs) and linked them with putative target genes by calculating the co-accessibility of each distal CRE and open regions within ±1 kb from the TSS of targeted genes in each cell type using the Cicero method[26], resulting in 1,872,039 CRE-gene pairs (Supplementary Table 3). Over 70% of CREs (436,615) showed differential chromatin accessibility across different cell subclasses, and a large proportion of CRE-linked genes in each cell subclass overlapped with marker genes in the corresponding snRNA-seq-defined cell subclasses (Fig. 1e, Extended Data Fig. 2f and Supplementary Table 3). For examples, L2/3 glutamatergic cell subclass showed high chromatin accessibility in CREs linked to *GPR83* and *PDZD2* genes, which were also highly expressed in snRNA-seq-defined L2/3 neurons. Moreover, these cell subclass-restricted CREs were also enriched for distinct sets of DNA motifs recognized by known transcriptional regulators (Fig. 1d). For instances, the motif binding activity of the nuclear receptor gene *NR1D2,* which encodes the TF involved in several brain disorders, showed the highest level in L4_2 glutamatergic neurons. The activity of binding motifs by SOX family TFs *SOX9* and *SOX10*, which are involved in oligodendrocyte specification and myelination[27], were found to be highest in OPCs (Fig. 1d). These findings establish a basis for understanding gene regulatory mechanisms underlying various transcriptome-defined cell types in the macaque cortex.

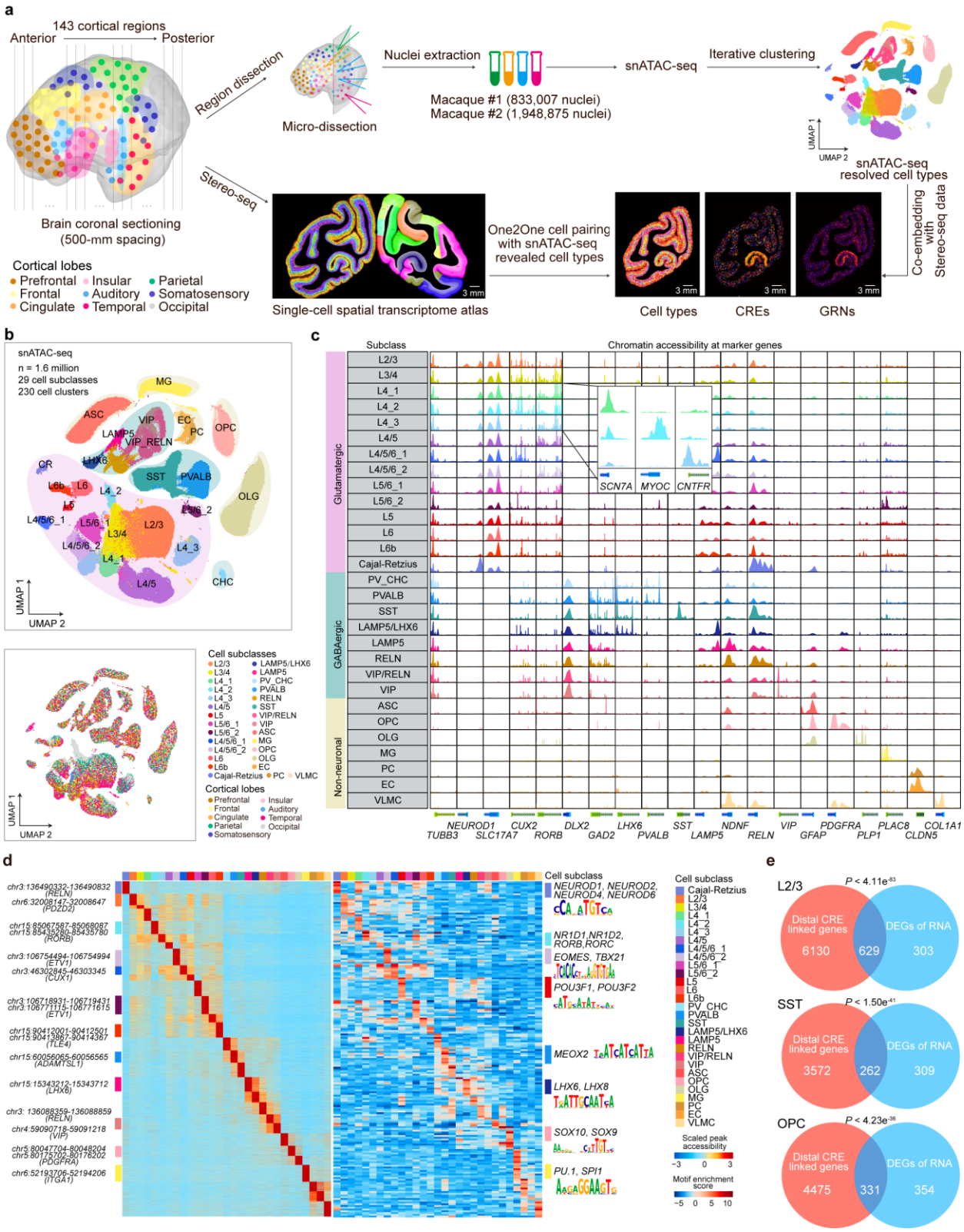


Fig. 1. Chromatin accessibility profiles of various cell types in the macaque cortex. a, Schematics of experimental procedure and integration of snATAC-seq with snRNA-seq and spatial transcriptome data.

Single nuclei from 142 cortical regions of two monkeys were isolated for snATAC-seq and snRNA-seq analysis, and co-embedded with spatial transcriptome data from a previous study[21]. **b**, Upper, UMAP embedding and clustering analysis based on chromatin accessibility profiles of snATAC-seq data. Below, UMAP layout showing single nuclei from different cortical lobes. Single nuclei are colored by cell clusters or cortical lobes. A full list and description of cell clusters and subclusters are provided in Supplementary Table 2. L2/3, layer 2/3, CR, Cajal-Retzius, PV, Parvalbumin, CHC, chandelier cell, SST, Somatostatin, ASC, astrocyte, OPC, oligodendrocyte precursor cell, OLG, oligodendrocyte, MG, microglia, EC, endothelial cells, PC, pericytes, VLMC, vascular leptomeningeal cells. **c**, Marker peaks of aggregated chromatin accessibility profiles for each snATAC-seq-based cell cluster at selected marker gene loci that were used for cell annotation. **d**, Left, Heatmap showing the chromatin accessibility of CREs across various cell clusters. Right, Heatmap showing the enriched TF motifs for various cell clusters. Representative TFs and their corresponding motif logos were shown on the right. **e**, Venn plots showing the overlap between distal CRE linked genes and differentially expressed genes in the same cell type. *P* values were calculated by hypergeometric test.

Spatial mapping of chromatin accessibility revealed by snATAC-seq data

Mapping chromatin accessibility of CREs to single cells on the spatial transcriptome map. To further examine the spatial accessibility of cell type-specific CREs, we integrated the snATAC-seq data with single-cell spatial transcriptome map of the macaque cortex[21] for one-to-one pairing. Specifically, we down-sampled the "gene activity score" matrix of snATAC-seq data to mimic the gene capture rate of Stereo-seq data, and then used autoencoder and generative adversarial network to transfer snATAC-seq-identified clusters onto the spatial map (Fig. 2a). The optimal transport method was used at the same time to achieve one-to-one mapping, in which chromatin accessibility profiles from snATAC-seq data were assigned to single cells on the spatial transcriptome map (Methods). The accuracy of this cell mapping was evidenced by the high consistency in the relative percentage of different cell types and marker genes between the snATAC-seq and spatial transcriptome data for the same cell types (Extended Data Fig. 3a, b).

Cortical layer preference of chromatin accessibility of CREs. To explore the spatial distribution pattern of chromatin accessibility for cell type-specific CREs, we then calculated the relative percentage of each cell type in each of the six cortical layers and found that glutamatergic subclusters showed clear laminar preference (Extended Data Fig. 3c). The cell type-specific CREs also showed laminar-specific chromatin accessibility in the spatial map, consistent with the

193 preferential laminar expression of their targeted genes. For example, the CRE linked to the marker
194 gene *GPR83* of L2/3 glutamatergic neurons showed open chromatin accessibility peaks only in
195 L2/3 (Fig. 2b). Spatial mapping of this CRE also showed significantly higher chromatin
196 accessibility in L2 and L3, consistent with the spatial expression preference of *GPR83* (Fig. 2c and
197 Extended Data Fig. 3d). In addition, distal CREs linked to *ADCYAP1*, *RORB*, and *TLE4* also
198 showed consistent preferences in spatial chromatin accessibility with the expression pattern of
199 corresponding marker genes for L2/3, L4 and L6 glutamatergic neurons, respectively (Fig. 2b, 2c
200 and Extended Data Fig. 3d).
201
202 ***Cortical region preference of chromatin accessibility of CREs.*** In addition to
203 layer-specific distribution, we also found cortical region-specific distribution of some cell clusters
204 based on their relative density among nine cortical regions (Extended Data Fig. 3c). For example,
205 the L4_2 glutamatergic subclass exhibiting high chromatin accessibility at the promoter of *MYOC*
206 was enriched in the V1 region of the occipital lobe quantified from both snATAC-seq and spatial
207 transcriptome data (Fig. 1c and Extended Data Fig. 3e, f). This cluster corresponded well to
208 snRNA-seq-defined L4.4 and L4.8 cell types that were also specifically distributed in V1[21]
209 (Extended Data Fig. 3g). The DNA sequences of CREs enriched in L4_2 glutamatergic cluster
210 were then scanned for TF-binding motifs with the CellOracle method[28], generating a gene
211 regulatory network (GRN) for all potential regulatory interactions (Fig. 2d). One marker CRE
212 linked to the *CPA6* gene was significantly enriched with the binding motif for TF MEF2C (Fig.
213 2e), which is involved in normal neuronal development, cell distribution, and electrical activity of
214 the neocortex as well as neuropsychiatric disorders[29]. Spatial mapping of this CRE also showed
215 that its chromatin accessibility was V1-specific (Fig. 2f). Furthermore, the RNAscope analysis
216 validated the V1-specific expression of *CPA6* and higher expression of *MEF2C* in V1 than V2 (Fig.
217 2g and Extended Data Fig. 3h). In addition, co-localization of the TF MEF2C and its target gene
218 *CPA6* were found in the V1 region (Fig. 2h), suggesting their V1-specific regulation. Similarly,
219 both the chromatin accessibility of CRE linked to *LAMA4* and the expression of *LAMA4* showed
220 V1 specificity (Extended Data Fig. 3i, j). These results indicate that the laminar- and regional-
221 specific accessibility of CREs might underlie the specific expression patterns of their targeted
222 genes.
223

Characterization of CREs in the piriform cortex. As the evolutionarily ancient part of the cerebral cortex in the brain, the allocortex has simpler structure than the more common neocortex[30]. To understand the regulatory differences between the allocortex and neocortex, we next performed differential CRE analysis between the allocortical piriform cortex and all the other neocortical regions. A total of 19,376 CREs showed significantly differential chromatin accessibility between the two regions (Extended Data Fig. 4a). The up-regulated CREs in the piriform cortex were enriched in oligodendrocytes (OLGs) and other glial cells, whereas the down-regulated CREs were mainly contributed by deep-layer glutamatergic neurons (Extended Data Fig. 4b). Moreover, the piriform cortex included a higher proportion of glial cells (especially OLGs) and lacked L4 neuronal subtypes (L4_1, L4_2 and L4_3) (Extended Data Fig. 4c), in line with the anatomical absence of layer 4 in the piriform cortex. In addition, GO functional enrichment analysis revealed that genes linked by up-regulated CREs (*CSF1R*, *P2RY1*, *CX3CR1*) were significantly enriched in the "gliogenesis" term, whereas genes linked by down-regulated CREs (*KCNIP4*, *RGS7*, *SLC24A2*) were enriched in "potassium ion transport" term (Extended Data Fig. 4d, e). Spatial mapping of these pathways showed that the "gliogenesis" pathway was significantly enriched in the piriform cortex, whereas the "potassium ion transport" showed a lower abundance than other neocortical lobes (Extended Data Fig. 4f, g). As examples, the CRE linked to *CSF1R* showed high chromatin accessibility in the piriform cortex (Extended Data Fig. 4h). In contrast, the CREs linked to *KCNIP4* showed high chromatin accessibility in the other neocortex (Extended Data Fig. 4i). Therefore, these results revealed different regulatory programs between the piriform and neocortical regions.

Gradients of CRE accessibility across the neocortex. Given the laminar and regional preferences of CRE accessibility, we next wondered whether the accessibility of cell type-specific CREs showed intracortical gradients along the neocortex. As an example, the *CBLN2* is known to exhibit an anterior-posterior gradient in expression enriched in the prefrontal cortex during both development and adulthood in macaques and humans, and its enhancer has been identified to be active in the prefrontal cortex[31]. In order to assess the potential gradients of CREs associated with *CBLN2* along the macaque cortical hierarchy, we mapped the CREs linked to *CBLN2* gene in our spatial transcriptome map and found their higher chromatin accessibility in L2 and L3 (Extended Data Fig. 4j, k). We further performed Spearman correlation analysis of

255 chromatin accessibility and the hierarchy levels of various brain regions in the visual system and
256 the somatosensory system. The results showed that the accessibility of most of these CREs in L3
257 exhibited a significantly positive correlation with the hierarchy level of brain regions in both the
258 visual system and the somatosensory system (Fig. 2i). Two example CREs (chr18:7561002-
259 7561502 and chr18:7597050-7597550) showed open chromatin peaks in L2/3 and L3/4 and
260 exhibited increasing accessibility along the hierarchy of both visual and somatosensory systems
261 in L3 (Fig. 2j, k and Extended Data Fig. 4l). Taken together, these results revealed differences in
262 CRE accessibility across cortical layers and regions, supporting the role of different regulatory
263 programs in the spatial organization of various cell types in the primate neocortex.

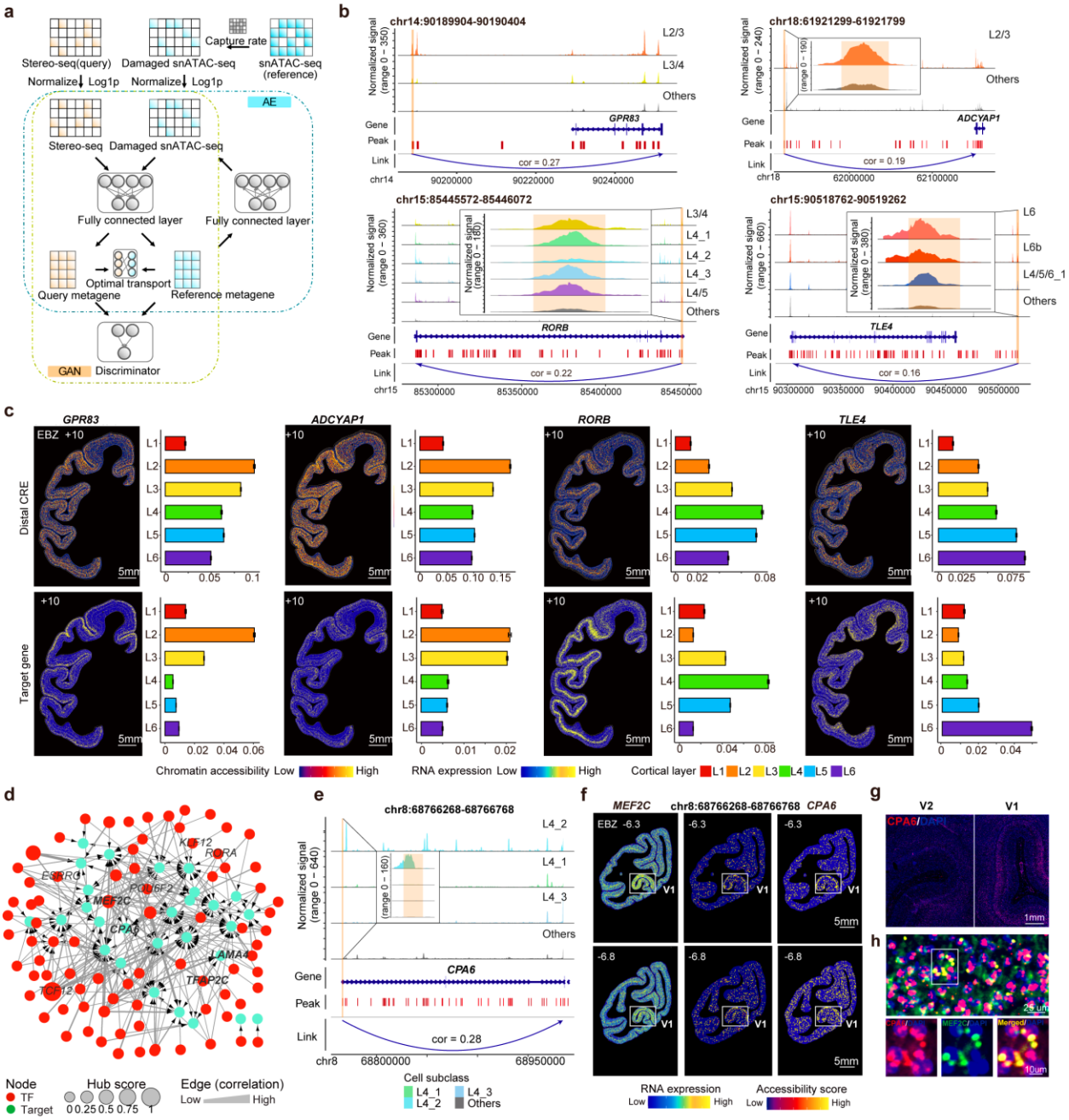

a
Stereo-seq(query)
Damaged snATAC-seq
Capture rate
snATAC-seq (reference)
Normalize Log1p
Normalize Log1p
AE
Stereo-seq
Damaged snATAC-seq
Fully connected layer
Fully connected layer
Optimal transport
Query metagene
Reference metagene
GAN
Discriminator
b
chr14:90189904-90190404
L2/3
L3/4
Others
GPR83
cor = 0.27
chr18:61921299-61921799
ADCYAP1
cor = 0.19
chr15:85445572-85446072
L4_1
L4_2
L4_3
L4/5
RORB
cor = 0.22
chr15:90518762-90519262
L6
L6b
L4/5/6_1
TLE4
cor = 0.16
c
GPR83
ADCYAP1
RORB
TLE4
Distal CRE
Target gene
EBZ +10
5mm
Chromatin accessibility Low High
RNA expression Low High
Cortical layer L1 L2 L3 L4 L5 L6
d
KLF12
RORA
ESRRG
FOXF2
MEF2C
CPA6
LAMA4
TFAP2C
TCF12
Node
TF
Target
Hub score
Edge (correlation)
Low High
e
chr8:68766268-68766768
L4_2
L4_1
L4_3
Others
CPA6
cor = 0.28
Cell subclass
L4_1
L4_2
L4_3
Others
f
MEF2C
CPA6
EBZ -6.3
-6.8
V1
5mm
RNA expression
Accessibility score
Low High
g
V2
V1
CPA6/DAPI
1mm
h
25 um
10um
i
Visual system
Somatosentory system
chr18:7561002–7561502
chr18:7568137–7568637
chr18:7568863–7569363
chr18:7591723–7592223
chr18:7592725–7593225
chr18:7593296–7593796
chr18:7594551–7595051
chr18:7595137–7595637
chr18:7597050–7597550
chr18:7597557–7598057
chr18:7598232–7598732
chr18:7600684–7601184
chr18:7606170–7606670
L2
L3
Correlation
j
chr18:7561002–7561502
chr18:7597050–7597550
Hierarchy levels
R=0.32
P=0.03
Chromatin accessibility
k
chr18:7561002–7561502
chr18:7597050–7597550
L2/3
L3/4
Others
CBLN2
Score=0.54
Score=0..27
Cell subclass
L2/3
L3/4
Other

Fig. 2. Characterization of CREs across various brain regions of the macaque cortex. a, Diagram showing the integration and one-to-one pairing of snATAC-seq and spatial transcriptome data. **b**, Genome browser track view showing differential accessible distal CREs linked to *GPR83*, *ADCYAP1*, *RORB* and *TLE4* in various glutamatergic neurons. **c**, Spatial maps and bar plots showing the chromatin accessibility of four example CREs and expression levels of these CREs-linked genes across cortical layers in a representative coronal section (EBZ +10). The quantification was calculated from chromatin accessibility in five representative coronal sections (Fig. 2c and Extended Data Fig. 3d). Error bars represent standard error of the mean (SEM). **d**, Gene regulatory networks of the V1-specific L4_2 glutamatergic cell cluster. Red and green colors represent TFs and their target genes, respectively. The directed edges indicate the regulation between TFs and their targeted genes. **e**, Genome browser track view showing a differential accessible distal CRE linked to *CPA6* across three L4 glutamatergic clusters and other glutamatergic clusters. **f**, Spatial maps of expression patterns of *MEF2C* and *CPA6*, and chromatin accessibility of the CRE linked to *CPA6* at two representative sections (EBZ -6.3 and -6.8). **g**, RNAscope assay showing the expression pattern of the *CPA6* gene in V1 and V2 regions of the macaque cortex. Probes against CPA6 (ACD, cat#1843691-C1) was list in Supplementary Table 8. DAPI was used for nuclear counterstaining. **h**, Co-localization of RNAscope labeling of the *CPA6* gene and the *MEF2C* gene in the V1 region. **i**, Heatmap showing the correlation of chromatin accessibility and the hierarchy levels of various brain regions in the visual and somatosensory system. Correlation coefficients and *P* values were calculated by Spearman correlation analysis. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$. **j**, Dot plots showing the correlation of chromatin accessibility and the hierarchy levels of various brain regions in the visual system. **k**, Genome browser tracks showing the chromatin accessibility of two CREs linked to *CBLN2* among L2/3, L3/4 and other cell subclasses.

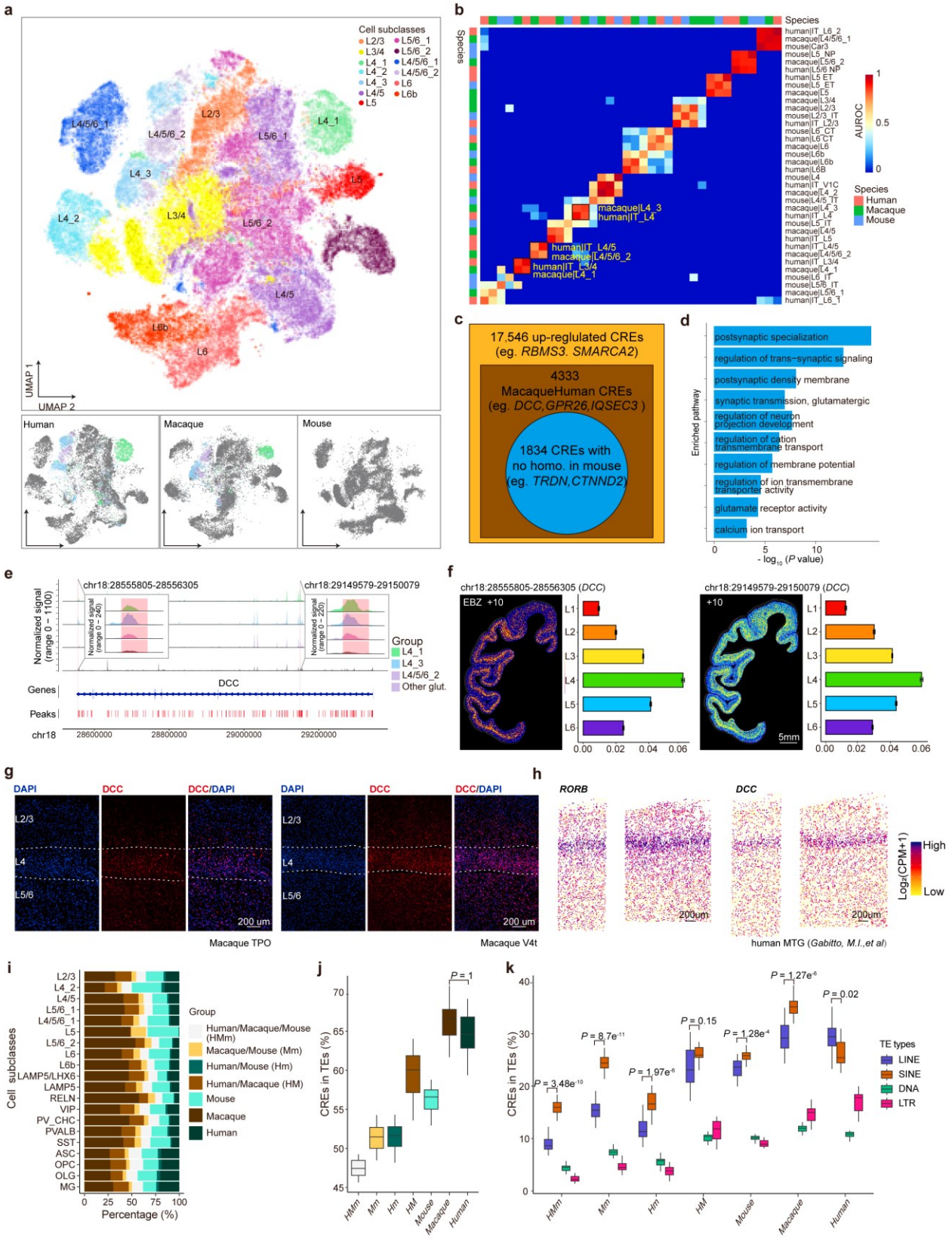
Human/macaque-biased CREs in glutamatergic neurons

Cross-species comparison of chromatin accessibility. To determine whether the gene regulation landscape in the cortex are conserved among humans, macaques and mice, we retrieved previously published snATAC-seq data for human and mouse cortices[11, 15]. Joint clustering analysis was performed using gene activity scores of homologous genes in snATAC-seq datasets for cortical tissues from all three species. We found generally conserved snATAC-seq-based clusters of GABAergic and non-neuronal cells across the three species, except for L4_1, L4_3 and L4/5/6_2 glutamatergic neurons that existed only in humans and macaques, as shown by the extracted maps of clusters for each species (Fig. 3a, Extended Data Fig. 5a-b and Supplementary Table 4). Further quantitative analysis among the three species using MetaNeighbor[32] showed that macaque L4_1, L4_3 and L4/5/6_2 neurons showed high similarity with human L3/4 IT, L4 IT and L4/5 IT cell

clusters, respectively, but no correspondence to any cluster in mice (Fig. 3b). Further comparison of all chromatin accessibility clusters revealed 12 clusters (belonging to L4_1, L4_3 and L4/5/6_2 subclasses) that were exclusively identified in primates but not in mice (Extended Data Fig. 5c). The human/macaque-biased clusters also corresponded well to snRNA-seq-based L4 and L4/5/6 glutamatergic cell types (Extended Data Fig. 3g), which were previously shown to be primate-specific[21].

Identification of human/macaque-biased CREs linked to disease associated genes. Further differential CRE analysis between human/macaque-biased cell types and all other glutamatergic cell types identified 17,546 CREs showing higher chromatin accessibility in the human/macaque-biased than other cell types (Extended Data Fig. 5d). Among these CREs, 4,333 CREs also showed high open chromatins in the corresponding human cell types, and 1,834 CREs had no homologous sequences in mice (Fig. 3c, Extended Data Fig. 5e and Supplementary Table 4). These human/macaque-biased CREs linked genes were significantly enriched in “glutamatergic synaptic transmission”, “postsynaptic specialization”, "regulation of trans-synaptic signaling" and “postsynaptic density membrane” (Fig. 3d). Specifically, we found that the chromatin accessibility of two CREs located in the intron of the neurodevelopmental disorder-associated gene *DCC* (netrin 1 receptor)[33] were much higher in these human/macaque-biased neurons than those in other cell types (Fig. 3e), consistent with the higher expression level of *DCC* in L4, L3/4, L3/4/5 and L4/5 glutamatergic neurons (Extended Data Fig. 5f). These CREs showed much higher accessibility in L4 for most coronal regions and in L3 and L5 for regions without L4 quantified by spatial mapping (Fig. 3f and Extended Data Fig. 5g), consistent with the expression pattern of *DCC*. RNAscope assay verified that the expression of *DCC* was enriched in layer 4 (Fig. 3g). Analysis using the previously published MERFISH dataset from the human MTG region[34] further validated the layer 4-specific expression of *DCC* gene, which is consistent with the expression pattern of *RORB*, a canonical marker gene of layer 4[34] (Fig. 3h). Another example is a distal CRE that linked to *TRDN* with higher chromatin accessibility in these human/macaque-biased glutamatergic neurons (Extended Data Fig. 5h). The decreased expression of *TRDN* was reported to be associated with apoptosis of dopaminergic cells in Parkinson’s disease[35]. Spatial mapping showed the higher chromatin accessibility of this CRE and higher expression level of *TRDN* in deeper layers (Extended Data Fig. 5i, j). These results suggested the unique roles of human/macaque-biased

332 CREs in regulating the spatial expression of disease-associated genes in the primate cortex.
333
334 Furthermore, we also examined whether CREs identified in macaque cell types that were
335 conserved across three species were present in human or mouse cortices. For this analysis, we
336 lifted CREs identified in each species to the human genome by sequence homology, and compared
337 the sequence similarity and chromatin accessibility of CREs in each conserved cell subclass. Only
338 CREs showing both sequence similarity and open chromatin between species were considered as
339 conserved (Methods). We found that in comparison with species-conserved CREs, a higher
340 proportion of CREs showed open chromatin in a species-specific pattern (Fig. 3i). Interestingly,
341 we found that ~60% of human/macaque-biased CREs were colocalized with known transposable
342 elements (TEs) (Fig. 3j), consistent with the notion that TEs as a driver of genomic diversity[36]. In
343 comparison with CREs conserved between primates and mice, human/macaque-biased or species-
344 specific CREs showed significantly higher frequency in co-localization with TE elements
345 including LINE, SINE, DNA and long terminal repeats. In addition, human/macaque-biased CREs
346 showed comparable enrichment with LINE and SINE, whereas conserved CREs between primates
347 and mice showed much higher co-localization with SINE element (Fig. 3k). Thus, we have
348 identified human/macaque-biased CREs that may determine layer- and region-specific
349 glutamatergic cell types and may be responsible for gene regulation underlying brain disorders.
350


a
Cell subclasses
L2/3
L3/4
L4_1
L4_2
L4_3
L4/5
L5
L5/6_1
L5/6_2
L4/5/6_1
L4/5/6_2
L6
L6b
UMAP 1
UMAP 2
Human
Macaque
Mouse
b
Species
AUROC
Human
Macaque
Mouse
macaque|L4_3
human|IT_L4
human|IT_L4/5
macaque|L4/5/6_2
human|IT_L3/4
macaque|L4_1
c
17,546 up-reglulated CREs
(eg. RBMS3. SMARCA2)
4333
MacaqueHuman CREs
(eg. DCC,GPR26,IQSEC3)
1834 CREs with
no homo. in mouse
(eg. TRDN,CTNND2)
d
Enriched pathway
postsynaptic specialization
regulation of trans−synaptic signaling
postsynaptic density membrane
synaptic transmission, glutamatergic
regulation of neuron projection development
regulation of cation transmembrane transport
regulation of membrane potential
regulation of ion transmembrane transporter activity
glutamate receptor activity
calcium ion transport
- log10 (P value)
e
chr18:28555805-28556305
chr18:29149579-29150079
Normalized signal (range 0 – 1100)
Group
L4_1
L4_3
L4/5/6_2
Other glut.
Genes
DCC
Peaks
chr18
28600000
28800000
29000000
29200000
f
chr18:28555805-28556305 (DCC)
chr18:29149579-29150079 (DCC)
EBZ +10
+10
5mm
L1
L2
L3
L4
L5
L6
0.00
0.02
0.04
0.06
g
DAPI
DCC
DCC/DAPI
L2/3
L4
L5/6
200 um
Macaque TPO
Macaque V4t
h
RORB
DCC
200um
Log2(CPM+1)
High
Low
human MTG (Gabitto, M.I.,et al)
i
Cell subclasses
L2/3
L4_2
L4/5
L5/6_1
L4/5/6_1
L5
L5/6_2
L6
L6b
LAMP5/LHX6
LAMP5
RELN
VIP
PV_CHC
PVALB
SST
ASC
OPC
OLG
MG
Percentage (%)
Group
Human/Macaque/Mouse (HMm)
Macaque/Mouse (Mm)
Human/Mouse (Hm)
Human/Macaque (HM)
Mouse
Macaque
Human
j
CREs in TEs (%)
P = 1
HMm
Mm
Hm
HM
Mouse
Macaque
Human
k
CREs in TEs (%)
P = 3.48e-10
P = 8.7e-11
P = 1.97e-6
P = 0.15
P = 1.28e-4
P = 1.27e-6
P = 0.02
TE types
LINE
SINE
DNA
LTR

351

352 **Fig. 3. Comparative analyses revealed human/macaque-biased CREs in glutamatergic neurons. a**,
353 UMAP co-embedding of snATAC-seq data from macaques, humans and mice. Colored dots at the bottom
354 indicate human/macaque-biased glutamatergic neurons derived from snATAC-seq data. **b**, Heatmap
355 showing the similarity of snATAC-seq derived glutamatergic cell clusters across macaques, humans and
356 mice using MetaNeighbor. **c**, The outer box represents the number of CREs showing higher chromatin
357 accessibility in human/macaque-biased than other cell types of the macaque cortex. The inner box represents
358 CREs with high chromatin accessibility in the corresponding human cell types. The inner circle indicates
359 CREs showing no homologous sequence in the mouse genome. **d**, GO enrichment for genes linked to
360 human/macaque-biased CREs (inner circle in **c**). **e**, Genome browser tracks showing the chromatin
361 accessibility of human/macaque-biased CREs located on the *DCC* gene in the human/macaque-biased and
362 other cell subclasses. **f**, Spatial map showing the chromatin accessibility of the human/macaque-biased
363 CREs shown in **e** in a representative Stereo-seq section at indicated coordinate. The right panel showing the
364 quantitative chromatin accessibility level across six cortical layers. Error bars represent standard error of the
365 mean (SEM). **g,** RNAscope results showing the layer 4-specific expression of *DCC* gene in TPO (left) and
366 V4t (right) regions of the macaque cortex. Probes against *DCC* (ACD, cat#1843701-C1) was listed in
367 Supplementary Table 8. DAPI was used for nuclear counterstaining**. h,** Scatterplots showing the expression
368 of layer 4 marker gene *RORB* and *DCC* gene in human MTG of the previously published MERFISH
369 dataset[34]. **i**, Bar plots showing 7 categories of CREs in various cell subclasses from macaques, humans and
370 mice. **j**, Boxplots showing the percentage of CREs colocalized with TE in different categories. **k**, Boxplots
371 showing the percentage of CREs located in different types of TEs. *P* values were calculated by Wilcoxon
372 rank-sum test.
373

374 ## Characterization of CREs in GABAergic cell types
375

376 ***Identification of a human/macaque-enriched LAMP5/LHX6 GABAergic cell type***. The
377 integrative clustering analysis also showed clear correspondence of snATAC-seq identified
378 clusters and snRNA-seq-defined GABAergic cell types (Extended Data Fig. 6a). However, based
379 on chromatin accessibility profiles, we found that transcriptome-defined LAMP5+ GABA cell
380 subclass could be further divided into LAMP5 and LAMP5/LHX6 cell types that exhibited low
381 and high chromatin accessibility of the *LHX6* gene, respectively. We identified 3,829 and 1,989
382 differentially accessible (DA) chromatin peaks in LAMP5/LHX6 and LAMP5 cells, respectively.
383 GO analysis of genes linked to these distal DA peaks revealed enrichment in trans-synaptic
384 signaling and postsynaptic membrane regulation for LAMP5/LHX6 cells, consistent with fast and
385 precise synaptic inhibition[37, 38, 39] In contrast, LAMP5 cell–specific peaks were associated with
386 pathways related to long-term synaptic depression, reflecting features of sustained inhibition in

387 CGE-derived neurons (Extended Data Fig. 6b). Interestingly, the enrichment of ZIC4, USF1, and
388 MEF2B motifs in primate LAMP5/LHX6 cells, but not in mice (Fig. 4b), suggesting an
389 evolutionary divergence in the regulatory programs governing interneuron specialization. These
390 transcription factors are known to regulate telencephalic regionalization, neuronal differentiation,
391 and activity-dependent plasticity[40, 41, 42]. Their primate-specific motif enrichment therefore implies
392 adaptive cis-regulatory innovations contributing to cortical expansion and refined inhibitory circuit
393 modulation in primates. Moreover, the pronounced regional variation in the ratio of LAMP5/LHX6
394 to LAMP5 cells, with the highest abundance in the cingulate cortex (Fig. 4c), indicates potential
395 region-specific functions. The preferential enrichment of LAMP5/LHX6 cells in association
396 cortices may reflect their role in higher-order integrative and cognitive processing[43, 44], whereas
397 the broader distribution of LAMP5 cells across cortical areas likely supports more generalized
398 inhibitory balance. By examining the shared CREs in LAMP5/LHX6 cells between macaques and
399 humans, we found enrichment of CREs for genes in the pathways of monoatomic and inorganic
400 cation transmembrane transport, such as *KCNQ3*, *SLC6A1* and *CACNA1G* (Extended Data Fig. 6c
401 and 6d), suggesting the distinct role of this enriched cell type in the primate brain.
402
403 ***Further characterization of LAMP5/LHX6 GABAergic cell subpopulations.***
404 Examination of chromatin accessibility in LAMP5/LHX6 cells showed that these cells could be
405 divided into two populations with differential accessibility of the *PROX1* gene (Fig. 4d), in line
406 with our finding based on a previous transcriptomic dataset on primate cortical cells[21] (Extended
407 Data Fig. 6e). Moreover, we also found differential chromatin accessibility for the marker gene
408 *NOS1* in macaque LAMP5/LHX6 cells (Fig. 4d), suggesting complementary roles of *PROX1* and
409 *NOS1*. To validate the presence of *PROX1*$^+$ and *NOS1*$^+$ subpopulations within LAMP5/LHX6 cells,
410 we analyzed a published MERFISH dataset from human dlPFC[45], which confirmed the existence
411 and spatial distribution of these two subgroups at single-cell resolution (Extended Data Fig. 6f-k).
412 We then identified the domains of regulatory chromatin (DORC) in each of the LAMP5/LHX6
413 cell populations and found that DORC of *SEC22A*, *ADARB2* and *PROX1* genes were upregulated
414 in *PROX1*$^+$ cells, and DORC of *ETS1, NOS1* and *EML6* genes were upregulated in *NOS1*$^+$ cells
415 (Fig. 4e and Extended Data Fig. 6l). This result further supports the importance of *PROX1* and
416 *NOS1* in determining the heterogeneity of primate cortical LAMP5/LHX6 cells. Among the CREs
417 associated with the *PROX1* and *NOS1* genes in LAMP5/LHX6 cells, 6 out of 21 CREs linked to

418 *PROX1* were human/macaque-biased (Fig. 4f), whereas 2 out of 5 distal CREs linked to *NOS1*
419 were human/macaque-biased (Fig. 4g). Differential spatial distributions were also observed, with
420 the *PROX1*$^{+}$ population being enriched in layers 3, 5, and 6, while the *NOS1*$^{+}$ population was
421 enriched in layers 2, 3, and 5 (Extended Data Fig. 6n). A previous study of human embryonic
422 ganglionic eminence development[46] identified an MGE-derived LHX6^{+}CRABP1^{-} population
423 proposed as progenitors of adult LAMP5/LHX6 neurons, which we found also co-express *PROX1*
424 and *NOS1*. Consistently, in macaque, adult LAMP5/LHX6 neurons from our previously snRNA-
425 seq dataset[21] clustered with an embryonic *LHX6*$^{+}$/*CCK*$^{+}$ population (*CRABP1*$^{-}$) reported as their
426 putative progenitors[42], likewise marked by *PROX1* and *NOS1* expression (Extended Data Fig.6n,
427 o). These findings support a developmental continuity from MGE-derived embryonic progenitors
428 to adult PROX1^{+} and NOS1^{+} LAMP5/LHX6 subpopulations.
429
430 ***Chromatin accessibility of CREs may underlie layer- and region-preference of SST***
431 ***neurons.*** SST cells were resolved into 27 transcriptionally distinct clusters characterized by
432 differential chromatin accessibility (Extended Data Fig. 7a), as shown by the top differentially
433 accessible (DA) peaks (Extended Data Fig. 7b). Genes linked to these DA peaks indicated cell
434 type-specific epigenomic regulation. Notably, clusters enriched for high *HTR2A/HTR2B* gene
435 activity (SST.1, SST.2 and SST.3) corresponded to serotonergic-responsive interneurons
436 (Extended Data Fig. 7c). To further characterize functional differences, we performed rGREAT
437 analysis on DA peaks from these three clusters and found that distinct SST subgroups were
438 enriched in different biological processes: SST.1 in neuropeptide signaling pathway, SST.2 in
439 energy homeostasis, and SST.3 in synapse organization (Extended Data Fig. 7d), underscoring the
440 biological relevance of this fine-grained classification.
441
442 Previous study has revealed the layer-specific distribution of SST neurons that contributed
443 to various anatomical and electrophysiological features of the human cortex[47]. Layer enrichment
444 of various SST cell types was also found in single-cell spatial transcriptome map of the macaque
445 cortex[21]. We then further focused on the layer-specific regulatory programs of various SST cell
446 types. We classified layer-specific SST neurons based on clustering analysis of their chromatin
447 accessibility profiles and annotated them based on their layer preferences (from L2/3 to L5/6, Fig.
448 4h), using gene activity scores of previously identified layer marker genes[21] (Extended Data Fig.

7e). Further GRN analysis revealed that different layer-enriched SST cell types showed layer-enriched activity scores for genes targeted by different TFs, such as *KLF7* and *FOXO3* for L2/3 SST cells, *ETS1* and *BCL11B* for L3 SST cells, *BATF* and *POU1F1* for L3/5 SST cells, *NR2F2* and *MITF* for L5 cells, and *BACH2* and *NFE2L2* for L5/6 SST cells (Fig. 4i and Supplementary Table 5). As an example, we demonstrated that the preferential accessibility of two CREs associated with *KIRREL3* for KLF7 binding in L2/3 on the spatial transcriptome map (Fig. 4j, k). Interestingly, we also noticed that the accessibility of the two CREs showed regional differences, with the highest levels in parietal and prefrontal lobes, respectively (Fig. 4l). In another example, we found the preferential L5/6 accessibility of CREs associated with *DCC* for BACH2 binding (Extended Data Fig. 7f, g) and the highest level of this CRE accessibility in the prefrontal and cingulate cortical regions (Extended Data Fig. 7h). These findings support the notion that binding of specific TFs to accessible CREs for genes expressed in SST cells may determine their layer-specific and region-enriched distribution.

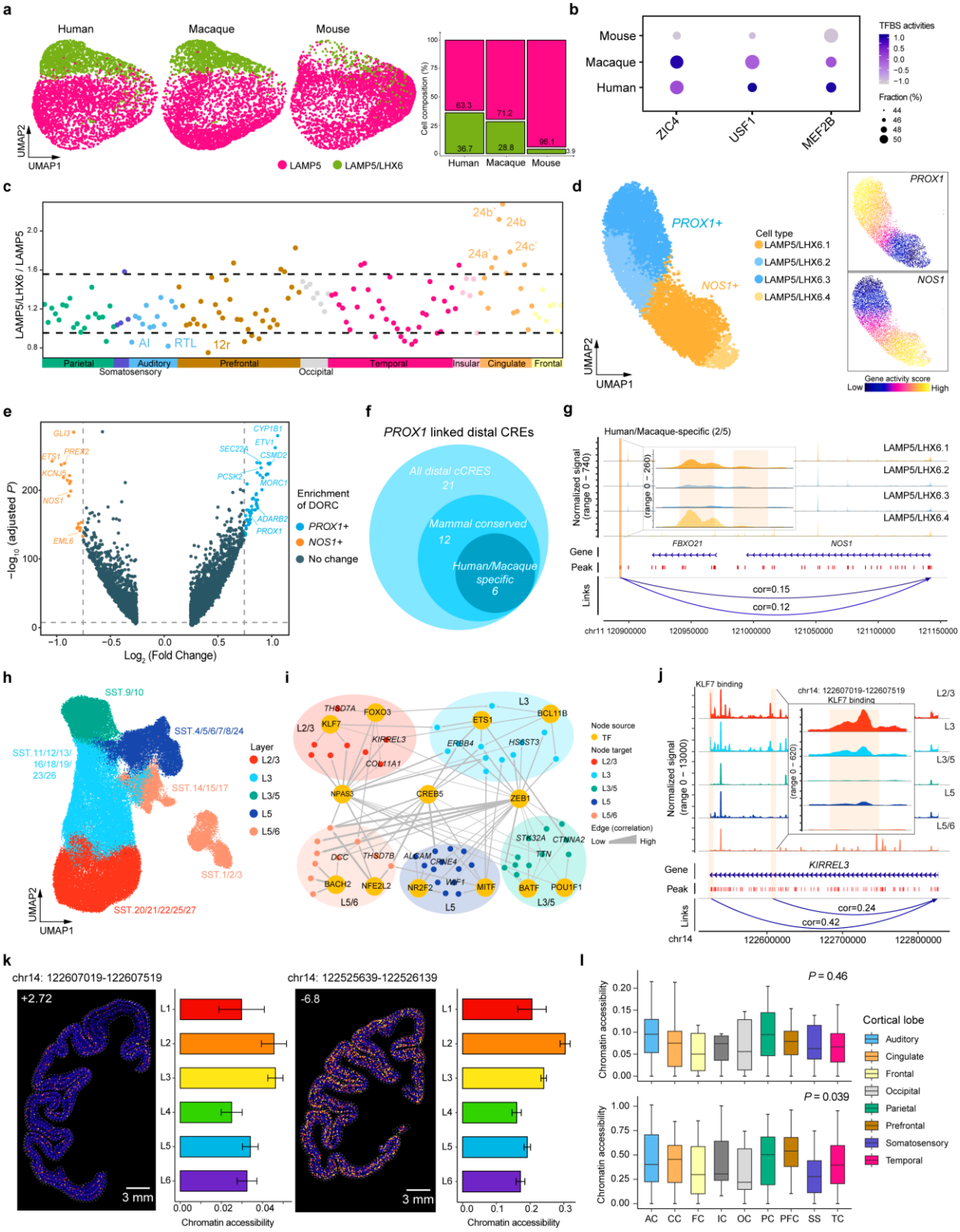


Fig. 4. Characterization of specialized regulatory landscape of GABAergic neurons. a, The UMAP visualization (left) and relative percentage (right) of GABAergic LAMP5 and LAMP5/LHX6 cell subclasses among humans, macaques and mice. **b**, Dot plot showing the transcription factor binding site (TFBS)

activities in LAMP5/LHX6 GABAergic neurons across species. **c**, Relative ratio of LAMP5/LHX6 vs. LAMP5 cell subclasses in various macaque cortical regions. **d**, UMAP embedding of LAMP5/LHX6 neurons (left) and normalized gene activity score of *PROX1* and *NOS1* (right). **e**, Volcano plot showing the differentially enriched DORCs in *PROX1*$^+$ and *NOS1*$^+$ populations of LAMP5/LHX6 neurons. Differential enrichment was defined as adjusted *P*-value $< 1\times10^{-5}$ and $|\log_2 FC| > 0.75$. **f**, The number of distal CREs linked to TSS region of *PROX1*. **g**, Genome browser track view showing two primate-specific distal CREs linked to TSS region of *NOS1* in LAMP5/LHX6 cell clusters. **h**, UMAP embedding of GABAergic SST neurons labeled with their layer specificity. **i**, Gene regulatory networks for layer-specific SST neuron groups. **j**, Genome browser track showing two differential accessible distal CREs linked to *KIRREL3* in L2/3 SST neurons. The two CREs harbor binding motifs of TF *KLF7*. **k**, Spatial visualization of chromatin accessibility of CREs linked to *KIRREL3* in SST neurons in representative sections at different coordinates in EBZ numbers (left). The bar plot illustrates the mean chromatin accessibility of CREs across different layers (right). Error bars represent standard error of the mean (SEM). **l**, Boxplot showing chromatin accessibility of two CREs linked to *KIRREL3* (upper/lower correspond to left/right in **k**) in SST neurons across cortical lobes. *P* values were calculated by Kruskal-Wallis test.

Characterization of CREs for cortical non-neuronal cells

Cross-species comparison of various non-neuronal cell types. We have further explored the diversity of gene regulatory mechanisms that may be responsible for generating distinct non-neuronal cell types and their distribution in various macaque brain regions, as revealed by the previous spatial transcriptome analysis[21]. Cross-species comparison of our macaque data with those from humans and mice[11, 15] showed that these non-neuronal cell types are all present in the cortices of the three species, with the percentage of various non-neuronal cells varied among the three species. While the astrocytes were found to be present at similar percentages, the percentages of oligodendrocytes (OLGs) increased gradually from mice to macaques and humans, and that of microglia in the human cortex was much higher than that found in both macaque and mouse cortices (Extended Data Fig. 8a, b). Notably, we found that the percentages of cortical oligodendrocyte precursor cells (OPCs) were similar between macaques and humans, but much higher than that in mice (Extended Data Fig. 8a, b). This finding is in line with the prolonged myelination and altered gene regulation in the oligodendrocyte lineage of the primate brain[2, 48, 49, 50].

Cortical hierarchy-dependent regulation of OPCs and OLGs. We next examined the abundance of OPCs vs. OLGs across the 143 cortical regions and found the opposite abundance

503 of OPCs and OLGs in both snATAC-seq and spatial transcriptome data (Fig. 5a and Extended Data
504 Fig. 8c). Interestingly, we observed a strong negative correlation (R = -0.9, $P < 2.2\text{x}10^{-16}$) between
505 the percentages of OPCs and OLGs among various cortical regions in the visual system[51] (Fig. 5b).
506 Specifically, higher abundances of OPCs and OLGs were found in cortical regions of high and low
507 hierarchy levels, respectively. The high potential for oligodendrocyte proliferation due to OPCs in
508 regions of higher hierarchy suggests higher plasticity in the regulation of axon myelination in these
509 regions. To further understand the differences in regulatory elements for OPCs and OLGs among
510 different brain regions, we performed Spearman correlation analysis of chromatin accessibility and
511 the hierarchy levels of various brain regions in the visual system. The results showed a consistent
512 higher number of CREs exhibiting chromatin accessibility negatively correlated with the hierarchy
513 level in OPCs regardless of thresholds for coefficients and P values (Extended Data Fig. 8d).
514 Interestingly, more than 30% of these negatively correlated CREs were human/macaque-biased
515 (Extended Data Fig. 8e, f), and their linked genes were significantly involved in GO terms
516 including glutamatergic synapse, visual system development and positive regulation of kinase
517 activity (Extended Data Fig. 8g and Supplementary Table 6). For examples, one mammal-
518 conserved CRE showing open chromatin peaks only in OPCs was linked to *PDGFRA*, a typical
519 marker gene for OPCs (Fig. 5c). Another example is a human/macaque-biased CRE with open
520 chromatin only in OPCs that linked to *FBLN2*, which is an extracellular matrix inhibitor of OPC
521 differentiation[52]. These results suggest the reciprocal regulation between OPCs and OLGs and the
522 importance of OPCs in regulating higher cortical functions.

523 ***CREs-determined GRNs between interlaminar and protoplasmic astrocytes***. Previous
524 spatial transcriptome studies have shown laminar preferences of astrocytes (ASCs)[21, 45]. Our
525 snATAC-seq-based cell clustering analysis further revealed differential chromatin accessibility of
526 CREs associated with interlaminar ASCs (enriched in L1) and protoplasmic ASCs (enriched in L2-
527 L6) (Extended Data Fig. 8h). For example, higher chromatin accessibility was found for CREs
528 linked to *ID3*, *FABP7*, *VCAN* and *MMP28* and those linked to *LUZP2*, *GRID1*, *KCNH5* and
529 *KCNMA1* in interlaminar and protoplasmic ASCs, respectively (Fig. 5e). Based on GO analysis,
530 CREs found in the interlaminar ASCs were enriched in genes associated with "cell adhesion" and
531 "cell projection" pathways, whereas those in the protoplasmic ASCs were associated with
532 "intracellular sodium ion homeostasis", "GABAergic synapse", and "positive regulation of
533 calcium ion transport" (Extended Data Fig. 8i). Moreover, we constructed GRNs for the two

534 groups of ASCs and found that interlaminar ASCs showed high enrichment of hub TF genes *ID3*
535 and *PBX3* (Fig. 5f). The targeted genes of *ID3* (including *GATA2*, *MSX1*, *TFAP2C* and *PBX3*) were
536 also involved in the central nervous system development (Fig. 5g). On the other hand, the
537 protoplasmic ASCs were enriched in the hub TF gene *NPAS3* (Fig. 5f), which contributes to
538 neurodevelopmental transcription factor networks and the regulation of brain glucose
539 metabolism[53] (Extended Data Fig. 8j). As examples, the CRE linked to *MMP28* showed high
540 chromatin accessibility in interlaminar but not protoplasmic ASCs (Fig. 5h). Spatial mapping of
541 this CRE also showed higher chromatin accessibility in L1, in line with the high expression of its
542 linked gene *MMP28* and its binding TF *ID3* in L1 (Fig. 5i, 5j and Extended Data Fig. 8k). In
543 contrast, one of the NPAS3-targeted CRE that linked to *LUZP2* showed significantly higher
544 chromatin accessibility in protoplasmic ASCs, and spatial mapping of this CRE also showed higher
545 chromatin accessibility across cortical layers, consistent with spatial expression pattern of *LUZP2*
546 (Extended Data Fig. 8l, m). Polymorphic variants in *LUZP2* have been reported to be associated
547 with Alzheimer's disease[54], schizophrenia[55], intelligence[56], and verbal memory[57]. Taken together,
548 these results provide new information on regulatory programs underlying the diversity and spatial
549 organization of non-neuronal cells in the primate brain.
550

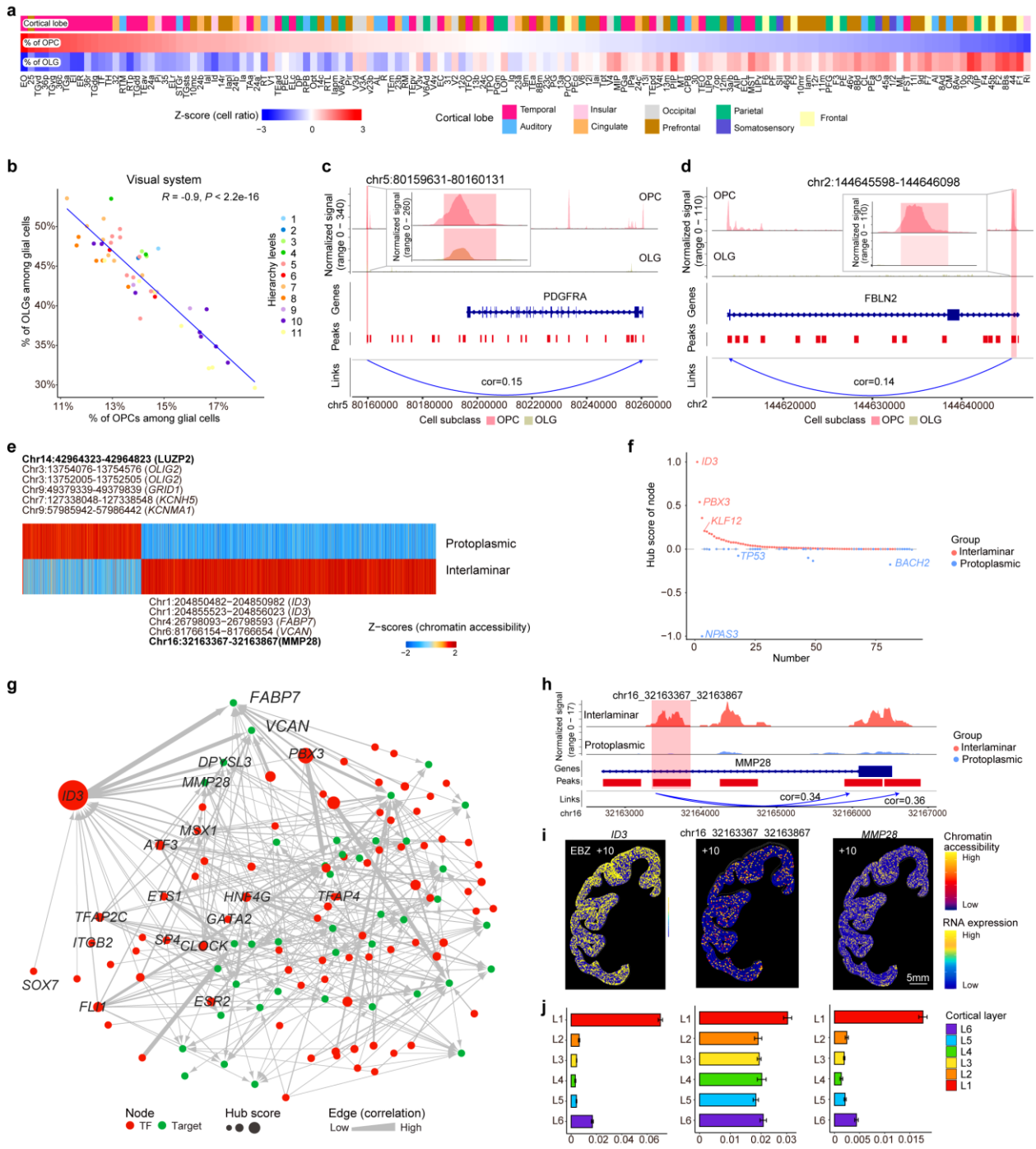


Fig. 5. Regional specificity of CREs in cortical non-neuronal cells. a, Heatmap showing the percentages of OPCs and OLGs among all glial cells in various brain regions of the macaque cortex after spatial mapping. The brain regions are ordered by the decreasing percentage of OPCs. **b**, Correlation between the percentage of OPCs and OLGs in various cortical regions in the visual system. Brain regions are colored by their hierarchical levels. Correlation coefficients and *P* values were calculated by Spearman correlation analysis. **c**, Genome browser track view showing one mammal-conserved CRE linked to the TSS region of *PDGFRA* in OPCs. **d**, Genome browser track view showing one human/macaque-biased CRE linked to the TSS region

559 of *FBLN2* in OPCs. **e**, Heatmap showing the chromatin accessibility of CREs in protoplasmic and
560 interlaminar astrocytes. **f**, Dot plot showing the hub score of enriched TFs in the GRNs of interlaminar and
561 protoplasmic astrocytes. The x-axis is ordered by the decreasing hub scores of TFs in the GRNs. **g**, Gene
562 regulatory network of TFs and targeted genes in interlaminar astrocytes. **h**, Genome browser tracks showing
563 the chromatin accessibility of CRE linked to *MMP28* between protoplasmic and interlaminar astrocytes. **i**,
564 Spatial maps of expression patterns of *ID3* and *MMP28*, and chromatin accessibility of the CRE linked to
565 *MMP28* at a representative section (EBZ +10). **j**, Bar plots showing quantitative levels of *ID3* and *MMP28*
566 expression as well as chromatin accessibility of the CRE linked to *MMP28* across six cortical layers. Error
567 bars represent standard error of the mean (SEM).
568

569 **Association of cell type-specific CREs with brain disease risk SNPs**
570

571 ***LDSC analysis of cell type-specific CREs and brain diseases***. To explore the contribution of cell
572 type-specific *cis*-regulatory elements (CREs) to the genetic architecture of brain diseases, we
573 applied linkage disequilibrium score regression (LDSC) to test the enrichment of GWAS-identified
574 risk SNPs from 28 brain disorders and human traits in our cell type-specific CREs (Supplementary
575 Table 7). Hierarchical clustering based on enrichment similarity revealed strong associations
576 between neuronal CREs, particularly those in glutamatergic neurons, and multiple brain disorders
577 (Fig. 6a, 6b and Extended Data Fig. 9a; Full list of cell types in Supplementary Table 7). For
578 examples, schizophrenia (SCZ) and bipolar disorder (BP) risk SNPs were enriched in CREs of
579 superficial-layer glutamatergic neurons, whereas BP SNPs were also present in PVALB and SST
580 neurons. Generalized epilepsy (GE) SNPs were enriched in RELN neurons, and Alzheimer's
581 disease (AD) SNPs were concentrated in microglia (Fig. 6a). Furthermore, SNPs related to
582 cognitive traits such as cognitive function (CGF), intelligence quotient (IQ), reaction time (RT),
583 and years of education (YED) showed the highest enrichment in L4 glutamatergic neurons (Fig.
584 6a). These results highlight the selective localization of disease-associated variants in CREs of
585 specific neuronal subtypes, underscoring their potential role in disease susceptibility.
586

587 ***CREs associated wi th SCZ and BP risk SNPs***. We next examined the genomic context
588 and evolutionary properties of CREs colocalized with disease-associated SNPs. Approximately 40%
589 of these CREs were intronic, and ~50% were located more than 1 kb from transcription start sites
590 (Extended Data Fig. 9b). Compared to background peaks, they exhibited significantly lower
591 phyloP scores[58, 59] (Extended Data Fig. 9c), suggesting rapid evolutionary turnover. Enrichment

592 patterns across diseases showed the highest correlation between SCZ and BP (Fig. 6c and Extended
593 Data Fig. 9d), consistent with their known genetic overlap[60]. Notably, we identified two shared
594 SNPs (rs8118905 and rs1040711) within a CRE linked to the *ADNP* promoter[61, 62], with
595 accessibility peaking in L2/3 glutamatergic neurons (Fig. 6d). Spatial mapping of this CRE also
596 showed the highest level of accessibility in L2 and L3 of the cortex (Fig. 6d and 6e). In addition,
597 we found two other SNPs rs2051307 and rs7753616 were linked to *SETBP1* gene in BP and
598 *MDGA1* gene in SCZ, respectively (Extended Data Fig. 9e). Both genes were related to
599 neurodevelopment[63, 64], and the spatial distribution of SNP-associated CREs for the two genes
600 were also enriched in the L2/3 glutamatergic neurons (Extended Data Fig. 9f). Together, these
601 findings indicate that SCZ- and BP-associated SNPs preferentially disrupt regulatory elements
602 active in L2/3 glutamatergic neurons, highlighting a shared neurodevelopmental basis for disease
603 risk.
604

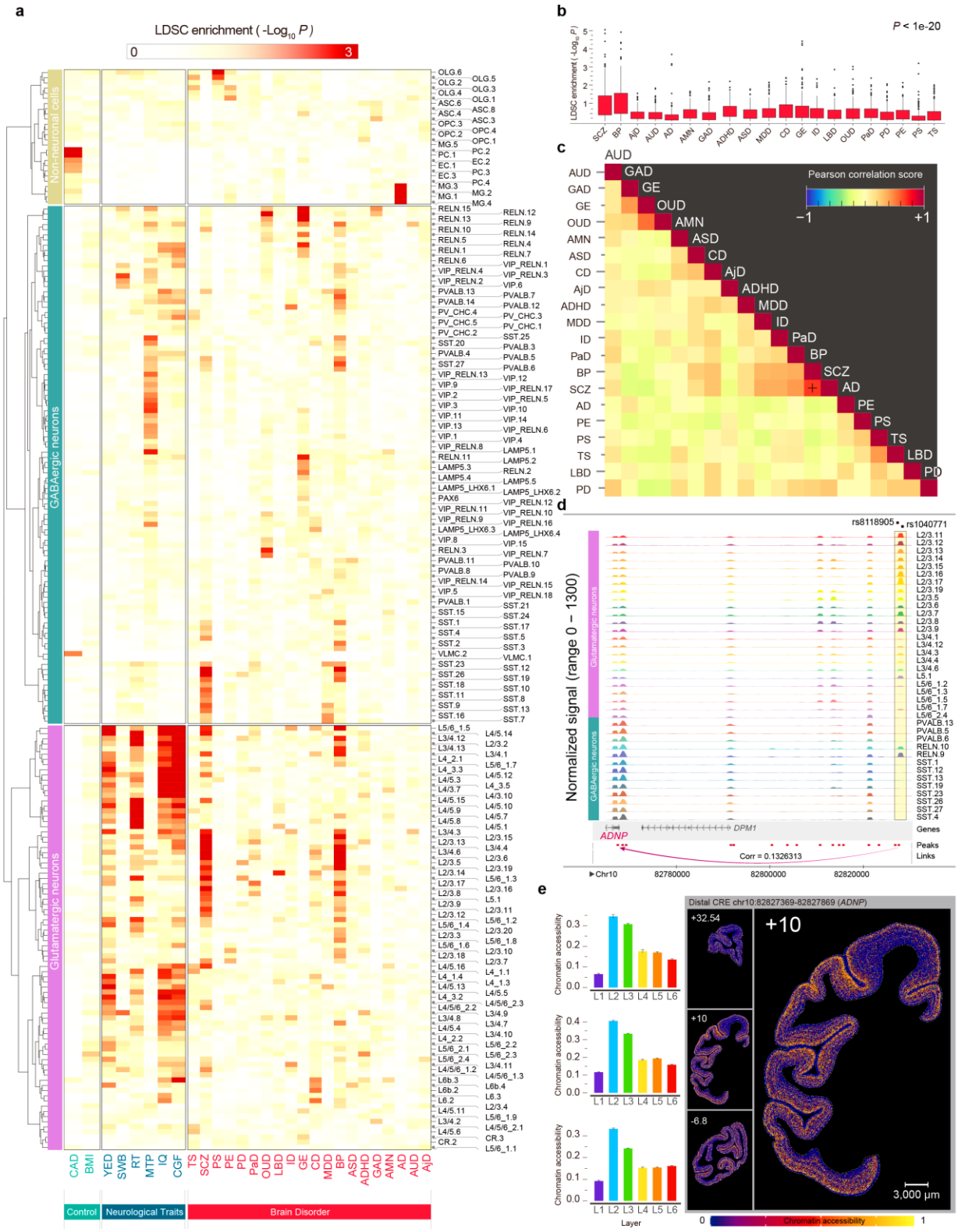


Fig. 6. Association of cell type-specific CREs with risk SNPs of various brain disorders and human traits. a, Heatmap showing the enrichment of cell type-specific CREs containing risk SNPs of various brain

608 disorders and human traits for all neuronal and non-neuronal cell types from snATAC-seq data of the
609 macaque cortex. Intensity of the red color codes LDSC enrichment score (defined as -$\log_{10}$ *P* values, with
610 *P* representing the probability of CREs containing risk SNPs). Black triangles mark the two brain disorders
611 with most prominent enrichment probabilities. AjD: adjustment disorder, AUD: alcohol dependence, AD:
612 Alzheimer disease, AMN: amnesia, GAD: anxiety disorder, ADHD: Attention deficit hyperactivity disorder,
613 ASD: Autism spectrum disorder, BP: bipolar disorder, MDD: Major depression disorder, CD: cognitive
614 disorder, GE: generalized epilepsy, ID: Insomnia, LBD: Lewy body dementia, OUD: opioid dependence,
615 PaD: panic disorder, PD: Parkinson Disease, PE: partial epilepsy, PS: psychosis, SCZ: Schizophrenia, TS:
616 Tourette syndrome, CGF, IQ: Intelligence, cognitive function, MTP: Morning Person, RT: reaction time,
617 SWB: subjective well-being, YED: Years of Education, BMI: Body mass index, CAD: coronary artery
618 disease. **b**, Boxplot showing the relative enrichment levels across different diseases. Significant difference
619 was observed among different diseases (P < 1e-20, Kruskal-Wallis test). **c**, Heatmap showing the Pearson
620 correlation coefficients for the association of cell type-specific CREs with disease risk SNPs loci for various
621 brain disorders. Star marked the high similarity in disease risk SNP-associated cell type-specific CREs
622 between schizophrenia and bipolar disorder. **d**, Genome browser tracks showing chromatin accessibility of
623 CREs for gene *ADNP* that contain risk SNPs (marked by dots) of both SCZ (schizophrenia) and BP (bipolar
624 disorder). The arrow line indicates the targeted gene *ADNP*. **e**, Spatial maps of cell type-enriched CREs
625 containing SCZ- and BP-associated SNPs loci shown in **d** at three representative sections along the coronal
626 coordinates, indicating the layer-specific distribution of the CRE for *ADNP.*
627

628 ***Human/macaque-biased CREs harboring disease risk SNPs***. To assess the evolutionary
629 specificity of disease-associated CREs, we classified SNP-containing CREs as human/macaque-
630 biased or mammal-conserved. Glutamatergic neurons, especially the L4_3 subclass, harbored
631 significantly more human/macaque-biased disease-associated CREs than other cell types (Fig. 7a).
632 In L4_3 subclass, human/macaque-biased CREs carrying SCZ SNPs displayed higher accessibility
633 than mammal-conserved CREs (Fig. 7b). Among them, genes linked to human/macaque-biased
634 CREs (181 gene; including *ABCA7*[65], *ADORA1*[66] and *CHGA*[67]) and mammal-conserved CREs
635 (293 genes, including *CHRM4*[68], *CXXC1*[69] and *OPCML*[70]) showed only minimal overlap (12 genes,
636 including *PDYN)*[71] (Fig. 7c and Extended Data Fig. 9g). Spatial mapping showed cell type- and
637 region-specific accessibility patterns of these CREs, with L4_3.11 neurons in the occipital lobe
638 exhibiting the highest accessibility (Fig. 7d). Thus, SCZ-associated regulatory mechanisms
639 appeared to be highly cell type- and brain region-specific. Gene regulatory network analysis of
640 L4_3 neurons identified transcription factors CTCF[72], NEUROD6[73], SREBF1[74], HIF1A[75] and
641 CXXC1[69] as central regulators (Fig. 7e). Furthermore, the human/macaque-biased CRE linked to

642 *GFRA2*[76] and the mammal-conserved CRE linked to *NEFH*[77], containing binding sites for
643 regulators of SREBF1 and HIF1A, respectively, both exhibited the highest chromatin accessibility
644 in L4 of the spatial maps (Fig. 7f and Extended Data Fig. 9h). Collectively, these results indicate
645 that primate-evolved regulatory networks may play a pivotal role in SCZ pathology.
646
647 We also investigated the distribution and functional properties of human/macaque-biased
648 and mammal-conserved CREs carrying AD risk SNPs in microglia. AD SNP–associated CREs
649 were exclusively localized to microglia (Fig. 6a), with the highest ratio of human/macaque-biased
650 to conserved CREs observed among all brain diseases examined (Fig. 7g). These two CRE classes
651 were largely distinct in their target genes (127 vs. 207 genes with 17 shared) (Fig. 7h).
652 Human/macaque-biased CREs carrying AD risk SNPs were linked to genes involved in pathways
653 such as amyloid-beta clearance, LDL response, and small molecule metabolism—absent from
654 those associated with conserved CREs (Fig. 7i). Notable AD-related genes linked to
655 human/macaque-biased CREs included *C3*[78], *TREM2*[79], *NFYA*[80], *CDNF*[81] and *IDE*[82]. For instance,
656 one human/macaque-biased CRE in microglia was linked to *TREM2* and *NFYA.* On the other hand,
657 both mammal-conserved and human/macaque-biased CREs were linked to AD-related genes[83, 84,
658 85] *APOE, TOMM40* and *TRAPPC6A* in microglia (Fig. 7j). Spatial accessibility analysis revealed
659 that human/macaque-biased CREs were most accessible in occipital and somatosensory cortices,
660 whereas conserved CREs peaked in insular and auditory cortices (Fig. 7k). These results suggest
661 that human/macaque-biased microglial CREs regulate distinct gene programs relevant to AD
662 pathogenesis, highlighting their potential as species-specific therapeutic targets.
663
664 ***Differential enrichment of disease SNPs in human- and macaque-specific CREs of***
665 ***LAMP5/LHX6 GABA neurons.*** Given that LAMP5/LHX6 neurons are enriched in both primates
666 (Fig. 4a), we next assessed whether disease risk SNPs are differentially enriched between human-
667 specific and macaque-specific CREs in this cell type. We focused on three brain disorders—
668 attention deficit hyperactivity disorder (ADHD), cognitive disorder (CD), and generalized epilepsy
669 (GE)—which exhibited significant heritability enrichment in our LDSC analysis (Fig. 6a). For
670 each trait, genes linked to CREs harboring risk SNPs showed limited overlap between the two
671 species (human vs. macaque, ADHD: 919 vs. 1246 genes, 65 shared; CD: 771 vs. 1055 genes, 70
672 shared; GE: 470 vs. 634 genes, 33 shared). Notably, within conserved genomic regions, disease-

673 associated SNPs mapped to distinct CREs that were either human- or macaque-specific (Extended
674 Data Fig. 10a-c). For examples, an ADHD risk SNP (rs34376065) was co-localized with a human-
675 biased CRE linked to the *NTNG1* gene. In contrast, open chromatin peaks were only found in the
676 macaque genome for CREs harboring several ADHD risk SNPs (Extended Data Fig. 10b). These
677 results highlight species-specific divergence in regulatory architectures underlying disease
678 heritability, even within consensus primate cell types such as LAMP5/LHX6 neurons.
679

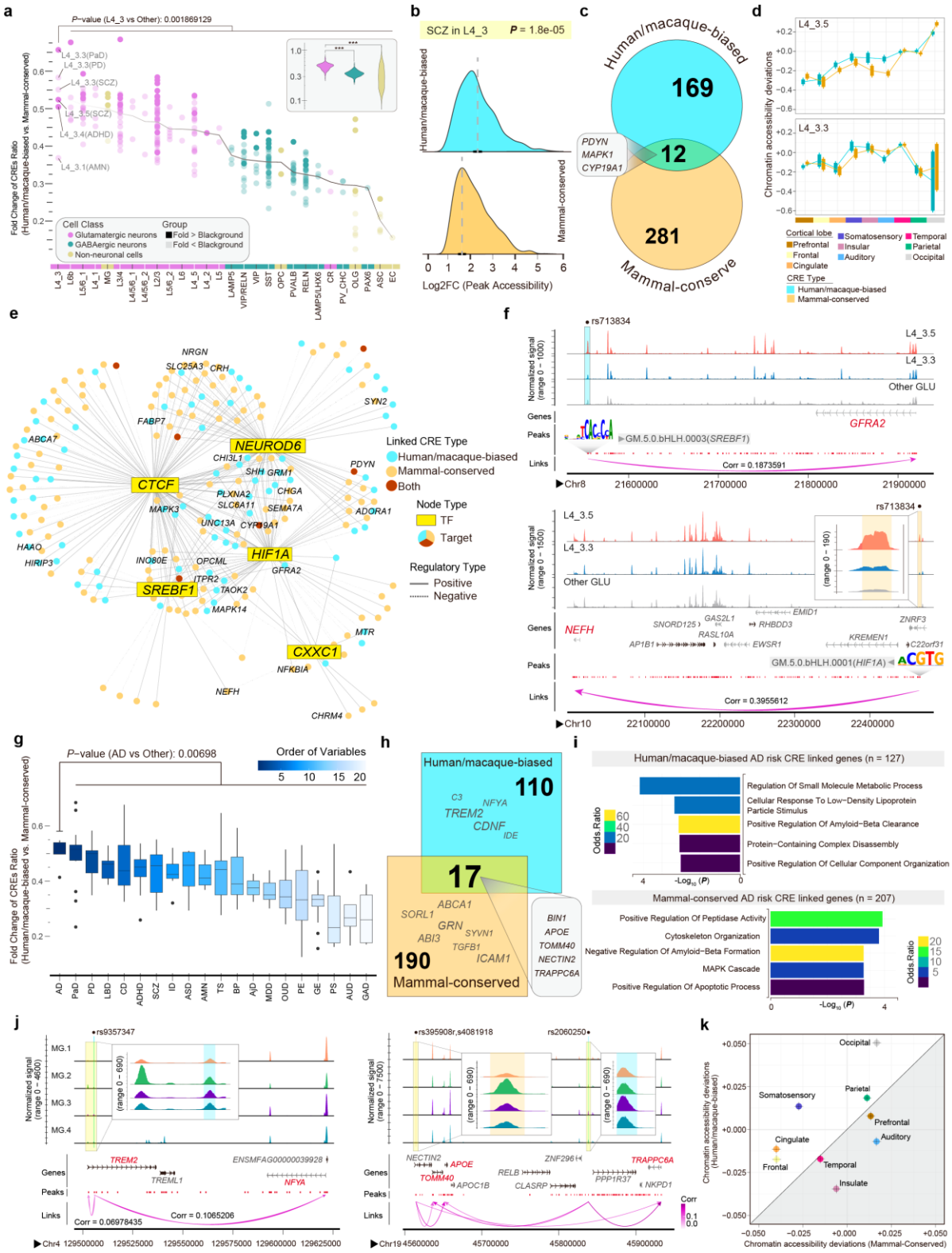


Fig. 7. Prominent association of human/macaque-biased CREs with risk SNPs of various brain disorders. **a**, Dot plot showing the ranked distribution of human/macaque-biased vs. mammal-conserved

CRE ratios (with disease-risk SNPs) across cell subclasses. Cell types with fold changes below the background are shown with reduced transparency. Fold changes for glutamatergic, GABAergic, and non-neuronal cell types are showed in the top-right corner. Significant difference was observed between L4_3 glutamatergic neuron and other cell subclasses (P = 0.001869, Wilcoxon rank-sum test). The violin plot (upper left) illustrates comparisons between glutamatergic neurons, GABAergic neurons, and non-neuronal cells, with asterisks indicating nominally significant differences (Wilcoxon rank-sum test, *** $P < 0.001$). **b**, Ridge plot depicting the chromatin accessibility of human/macaque-biased and mammal-conserved CREs (colocalized with SCZ risk SNPs) in the L4_3 glutamatergic cluster. Significant difference was observed between these two groups (P = 1.8e-05, T-test). **c**, Venn Diagram showing the overlap of genes associated with human/macaque-biased and mammal-conserved CREs (colocalized with SCZ risk SNPs) in the L4_3 glutamatergic cluster. **d**, Bar plot (Mean ± SE) illustrating the relative chromVAR deviations of SCZ risk SNPs associated CREs across cortical lobes in the L4_3 glutamatergic cluster. **e**, Gene regulatory network (GRN) showing the key transcription factors (TFs) for targeted genes linked to CREs associated with SCZ risk SNPs in L4_3 neuron. SCZ ontology genes were labeled. **f**, Genome browser tracks showing chromatin accessibility in L4_3 neurons at SCZ risk SNP-associated CREs linking to *GFRA2* and *NEFH*, respectively. Highlighted regions are marked with color bands. **g**, Box plot showing the ranked distribution of the fold change in the relative composition of human/macaque-biased versus mammal-conserved CREs (colocalized with disease risk SNPs) across various brain disorders. Significant difference was observed between Alzheimer's disease (AD) and other brain disorders (P = 0.00698, Wilcoxon rank-sum test). **h**, Venn Diagram illustrating the overlap of genes associated with human/macaque-biased and mammal-conserved CREs (colocalized with AD risk SNPs) in microglia cells. **i**, Pathways enriched by genes linking to AD risk SNP associated CREs in microglia. **j**, Genome browser tracks plot showing chromatin accessibility in microglia at the distal regions of genes linking to the AD risk SNP associated CREs: *TREM2* and *NFYA* (left, primate-specific), as well as *APOE*, *TOMM40*, and *TRAPPC6A* (right, human/macaque-biased and mammal-conserved). Highlighted regions are marked with color bands. **k**, Dot plot illustrating the relative chromVAR deviations of human/macaque-biased and mammal-conserved CREs (colocalized with AD risk SNPs) across brain lobes in microglia.

Discussion

We have performed a comprehensive study of chromatin accessibility at the single-cell level across 142 brain regions of the macaque cortex and identified specific CREs for various cell types. Leveraging the newly developed method for integrating chromatin accessibility and spatial transcriptome data, we found that many cell type-specific CREs showed prominent laminar and regional preferences in their chromatin accessibility, thus providing gene regulatory mechanisms underlying the laminar and regional diversification of cell types. Cross-species comparison with human and mouse snATAC-seq data further identified CREs associated with brain disease-related

genes in human/macaque-enriched L4 glutamatergic and GABAergic cell types. Further comprehensive association analysis between cell type-specific CREs and brain disease risk SNPs revealed potential contribution of human/macaque-biased CREs in various brain diseases. Notably, human/macaque-biased CREs in L4 glutamatergic neurons are much more prominently associated with risk SNPs for schizophrenia. Furthermore, we also identified human/macaque-biased CREs that regulate AD pathogenesis-related genes (*C3*, *TREM2*, *NFYA*, *CDNF* and *IDE)* exclusively in microglia. This study offers a comprehensive database for further investigating gene regulatory mechanisms underlying cell type diversity in the primate brain as well as brain disease pathogenesis.

Our analysis of the chromatin accessibility landscape in primate and mouse cortices revealed the role of human/macaque-biased CREs in the emergence of human/macaque-biased cell types as well as their potential involvement in psychiatric diseases. For example, in human/macaque-biased L4 glutamatergic neurons, we identified CREs localized in the regulatory region of *DCC*, a gene involved in neurodevelopment and psychiatric disorders[86, 87, 88]. We also found that human trait-related SNPs largely resided in CREs of human/macaque-biased glutamatergic cell types, suggesting potential contributions of these neuronal types to cortical development underlying human intellectual traits. The discovery of distinct presence of human/macaque-biased CREs related to some brain disorders such as schizophrenia and AD supports the notion that these brain diseases are due to cortical pathology that are largely primate-specific.

Previous studies have shown that cortical $LHX6^{+}LAMP5^{+}$ GABAergic neurons are more abundant in primates[4], but their molecular diversity and regulatory features remain poorly characterized. Here, we further subdivided the $LAMP5^{+}LHX6^{+}$ population into two transcriptionally and epigenetically distinct subtypes—*PROX1*$^{+}$ and *NOS1*$^{+}$—characterized by mutually exclusive chromatin accessibility at their respective lineage-defining genes (*PROX1* for CGE origin[46], *NOS1* for MGE origin[89]), as well as distinct laminar distributions within the cortex. Beyond previous transcriptomic observations, our analysis revealed subtype-specific DORCs and identified multiple human/macaque-biased CREs associated with *PROX1* and *NOS1*, suggesting evolutionary innovation in the regulatory architecture of these interneurons. These findings

751 highlight previously unrecognized epigenetic and spatial heterogeneity within LAMP5^{+}LHX6^{+}
752 neurons and provide a resource for understanding their developmental origin and primate-specific
753 specialization.
754
755 Our comprehensive analysis of chromatin accessibility and spatial transcriptome allowed
756 us to identify differences in chromatin accessibility across various cortical regions. This
757 information is critical for understanding the gene regulation underlying regional differentiation of
758 cell types. Our finding of the reciprocal relationship between the abundance of OPCs and OLGs
759 across all brain regions suggest that OLGs may provide negative regulation of OPC proliferation.
760 The opposite trend in the abundance of OPCs and OLGs at different hierarchy levels of the visual
761 system[51] further suggests the functional role of axon myelination in neural properties at various
762 hierarchy levels, and OPC biogenesis and myelination is tightly regulated across cortical hierarchy.
763 The higher abundance of OPCs in cortices of higher hierarchy suggests their higher capability of
764 generating new OLGs, contributing to the plasticity of myelin formation and regeneration in these
765 cortices.
766
767 Elucidation of regulatory mechanisms underlying diversity of cortical cell types and their
768 distribution is important for understanding not only the structure and function of the brain but also
769 pathogenic mechanisms underlying brain disorders. Through association analysis between cell-
770 type specific CREs and brain disease risk SNPs, we uncovered hundreds of disease-related CREs
771 in the primate brain, including those regulating genes involved in SCZ, BP and AD. Our finding
772 on AD risk SNP-associated CREs exclusively in microglia provides strong support to the notion
773 that microglia play a critical role in the AD etiology[90, 91]. Notably, prominent AD risk SNP alleles
774 were found in CREs linked to *TREM2* and *APOE* genes in microglia, and CREs for *TREM2* were
775 found to be human/macaque-biased. In addition, SCZ- and BP-associated CREs were generally
776 enriched only in the primate L2/3 glutamatergic neurons. Importantly, we found a large increase
777 in the frequency of transposon element insertion in human/macaque-biased CREs. This suggests
778 that the high frequency of brain disorder risk SNPs that are linked to human/macaque-biased CREs
779 might be attributed to the extensive proliferation of cortical neurons during primate brain
780 development.

781 # Methods

782 **Ethics statement**

783 Animal protocol was reviewed and approved by the Biomedical Research Ethics Committee of CAS Center for
784 Excellence in Brain Science and Intelligence Technology, Chinese Academy of Sciences (ION-2019011).
785 Animal care complied with the guideline of this committee. Adult macaque monkeys (M. fascicularis; #1, 6-
786 year-old, 4.2 kg, male; #2, 4-year-old, 3.7 kg, male; #3, 12-year-old, 5.1kg, female) were used in this study.

787

788 **Tissue collection and nucleus extraction**

789 We sampled 128 cortical regions from Macaque #1 and 132 regions from Macaque #2 for snATAC-seq data
790 generation. For Macaque #1, samples were collected from thick adjacent sections to the ones used for Stereo-
791 seq. For Macaque #2, we dissected brain samples from block-face cortical regions of the hemisphere opposite
792 to that used for Stereo-seq. Using tissue punchers (2.5 - 4 mm in diameter), we segmented the cortical areas at a
793 depth of 1 – 2 mm in the cryostat. After dissection, the samples were promptly frozen in liquid nitrogen and then
794 stored in dry ice or a -80 °C refrigerator. Throughout the sampling process, great care was taken to transfer the
795 tissues to pre-chilled tubes without allowing them to thaw. The single nucleus suspension was prepared following
796 the method described previously[92]. The isolated single nuclei were evenly split for snRNA-seq and snATAC-seq
797 library preparation, respectively. In brief, frozen sections of monkey brain tissue were placed in a Dounce
798 homogenizer containing 2 ml of pre-chilled homogenization buffer (containing 20mM Tris pH8.0, 500mM
799 sucrose, 0.1% NP-40, 0.2U/μL RNase inhibitor, 1x protease inhibitor cocktail, 1% bovine serum albumin (BSA),
800 and 0.1mM DTT), and the Dounce homogenizer was kept on ice during the grinding process. The tissue was
801 homogenized with 10-15 strokes of pestle A, followed by 10-15 strokes of pestle B. Subsequently, 2 ml of
802 homogenization buffer was added to the Dounce homogenizer, and the homogenate was filtered through 30 μm
803 MACS SmartStrainers (Miltenyi Biotech, #130-110-915) into a 15 ml conical tube. The tube was then
804 centrifuged at 500 g for 5 minutes at 4°C to pellet the nuclei. The pellet was resuspended in 1.5 ml of blocking
805 buffer (containing 1% BSA and 0.2U/μL RNase inhibitor in 1x PBS) and centrifuged again at 500 g for 5 minutes
806 at 4°C to pellet the nuclei. The nuclei were finally resuspended in cell resuspension buffer (PBS containing 1%
807 BSA) for subsequent snATAC-seq library preparation.

808

809 **Library preparation for single-nucleus ATAC-seq**

810 For the preparation of single-nucleus ATAC-seq libraries, we employed the DNBelab C Series Single-Cell ATAC
811 Library Prep Set (BGI-research) with the procedure of step transposed, droplet encapsulation, pre-amplification,
812 emulsion breakage, then the captured beads were collected. After DNA amplification and purification, the
813 indexed sequencing snATAC library was already to sequence. Finally, we used the Qubit ssDNA Assay Kit
814 (Thermo Fisher Scientific) to measure the concentrations of the sequencing libraries. All libraries were
815 sequenced using a paired-end sequencing scheme by the ultra-high-throughput DIPSEQ T1 sequencer platform

using the following read lengths: 70 bp for read 1, 50 bp for read 2, and 20 bp for the sample index sequencing scheme of the China National GeneBank (CNGB).

Data processing and quality control of snATAC-seq data

We used the newest pipeline to process the snATAC-seq data of DNBelab C4 (https://github.com/MGI-tech-bioinformatics). First, the raw reads were separated into insertions and barcodes and then filtered and demultiplexed using PISA (v1.1) with a minimum sequencing quality of 20. Next, the filtered reads were aligned to the *Macaca fascicularis* genome (5.0.91) using Chromap (v0.2.3_r407) and mismatched barcode was not allowed. 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 resulting BAM files were processed using d2c (v1.4.4) to identify barcodes from the same cell. The retained fragments of each library were used for further analysis.

Basic processing of the snATAC-seq data was performed using ArchR[93] (v1.0.2) package in R (v4.2.2). Genome annotation file of *Macaca fascicularis* was download from Ensembl (release 91). Fragments files were loaded by "createArrowFiles" function, with chrM and chrY chromosomes excluded. Each sample was separately computed for counts for each tile per cell via "addTileMatrix" function with tileSize = 500. Doublets were identified by "addDoubletScores" function with Doublet score < 0.15. A total number of 1,615,430 cells from 405 snATAC-seq libraries was kept with threshold of TSS enrichment >= 5 and <=30 and Promoter ratio >= 0.05 and <= 0.6.

Dimension reduction and clustering of snATAC-seq data

First, we used "addGroupCoverages" function and "addReproduciblePeakSet" function in MACS2[24] (v2.2.7.1) to create a reproducible merged peak set for each library. A new matrix was created within the new merged peak set with "addPeakMatrix" function for downstream analyses. Next, we further performed an iterative LSI dimensionality reduction via the "addIterativeLSI" function with varFeatures = 50000. Harmony was applied to remove the batch effects across libraries. In order to better cluster the data, we applied a three-step clustering approach to refine cell type identification based on chromatin accessibility. This approach resulted in 3 classes, 29 subclasses and 230 cell types overall. We first carried out unsupervised clustering for all cells with resolution = 3, categorized the initial cell clusters into 3 categories based on classical marker genes, including glutamatergic neurons (SLC17A6, SLC17A7), GABAergic neurons (*GAD1*, *GAD2*) and non-neuronal cells (*GFAP*, *PDGFRA*, *PLP1*, *PLAC8*, *CLDN5* and *COL1A1*), then repeated the above analysis workflow with higher resolution for each category.

A total of 230 cell clusters were kept after removing clusters with the number of cells less than 100, or ambiguous clusters (expressing both glutamatergic (*SLC17A6*, *SLC17A7*) and GABAergic (*GAD1*, *GAD2*) marker genes or non-neuronal marker genes). We have provided a comprehensive list of marker genes used for the identification

of ambiguous clusters in Supplementary Table 2. Previous study showed that most doublets would either form separate clusters, or would tend to end up at the fringes of other clusters (in graph embeddings)[94]. Therefore, we also applied the same procedure to further remove doublets from the clustering results. Specifically, cells were marked as doublets with their distances to the cluster center lower or higher than 1.5*interquartile range (25th and 75th percentile).

Correspondence of cell types defined by snATAC-seq and snRNA-seq data

We calculated gene activity score for each gene by summing up the counts within the gene body + 2 kb upstream of our snATAC-seq data using ArchR “addGeneScoreMatrix” function, and compared with the snRNA-seq data derived from a recently published macaque cortex study[21]. We randomly selected 200 cells from each cell cluster for both snATAC-seq and snRNA-seq data, and then performed data normalization using “NormalizeData” function. Integration analysis was conducted with Seurat “IntegrateData” function with top 3000 repeatedly variable genes as features. The proportion of nuclei overlap between omics represents the number of cells clustered in the same clusters of the integrative clustering result divided by the total number of cells for specific cell types from the two omics datasets, using the "compare_cl" function of the published pipeline[6].

Identification of reproducible peak sets

We performed peak calling according to the ENCODE ATAC-seq pipeline (https://www.encodeproject.org/atac-seq/). Across each cell type, we pooled all appropriately paired reads to produce a pseudo-bulk ATAC-seq dataset for each biological replicate using “getFrags” function. In addition, we created two pseudo-replicates by dividing the reads from each biological replicate in half. We called peaks for each of the four datasets and a pool of both replicates independently. Peak calling was performed on the Tn5- corrected single-base insertions using MACS2 (v2.2.7.1) with these parameters: —shift -75 --extsize 150 --nomodel --call-summits --SPMR -q 0.01. Finally, we extended peak summits by 250 bp on each side, resulting in a final width of 501 bp, for merging and downstream analysis. To identify a list of reproducible peaks, we filtered peaks with the following two criteria as in previous studies[15], (1) detected in the pooled dataset and overlapped ≥50% of peak length with a peak in both individual replicates; or (2) detected in the pooled dataset and overlapped ≥50% of peak length with a peak in both pseudo-replicates.

Due to the nature of the Poisson distribution test in MACS2, the MACS2 score was influenced by the number of nuclei or the read depth. Thus, we utilized the adjusted normalization method which converted the MACS2 scores ($-\log_{10}$ (q-value)) to “score per million (SPM)”, and filtered reproducible peaks by choosing a SPM cut-off of 2. Finally, we only kept reproducible peaks on chromosome 1–19 and both sex chromosomes, and filtered blacklist regions. A union peak list for the whole dataset was obtained by merging peak sets from all cell types

using Bedtools (v2.29.1).

Chromatin co-accessibility identification

We used R package Cicero[26] (v1.20.0) to compute co-accessible regions for all open regions. In each cell type, we randomly sampled 5,000 nuclei (All cells were retained for cell types with less than 5,000 nuclei). We set parameters of Cicero as: k = 50, window size = 500 kb, distance constraint = 250 kb. To avoid using an empirical cut-off, we also generated a random shuffled CRE-by-cell matrix and computed co-accessible regions on this shuffled matrix. Then, we fitted the distribution of co-accessibility scores from randomly shuffled background into a normal distribution model by using the R package fitdistrplus. Next, we tested each co-accessibility pair and set the cut-off at co-accessibility score with significance threshold of FDR < 0.01 to filter out co-accessible pairs.

We defined the CREs outside of ±1 kb of the gene transcriptional start site (TSS) as distal and the others as proximal. Then, we divided all co-accessible pairs into three groups: proximal-to-proximal, distal-to-distal, and distal-to-proximal. We focused only on distal-to-proximal pairs because they represent potential gene expression regulatory relationships. We classified all distal CREs in the above distal-to-proximal co-accessible pairs as final CREs.

We annotated the genome distribution of a total of 615873 CREs by “annotatePeak” function from R package ChIPseeker[95] (v.1.34.0), setting 1kb upstream and 1kb downstream of the TSS as promoter.

Differentially accessible CREs among different cell types and transcription factor (TF) binding motif analysis

We applied Wilcoxon rank sum test using the ‘‘getMarkerFeatures’’ function for identification of marker peaks with useMatrix = "PeakMatrix" at the cluster level. The CREs with log2 fold-change > 1 and FDR < 0.05 were identified as significant marker features for each cell type. Then, we removed the duplicated CREs and calculated the differential motifs across cell types with gimme maelstrom[96]. The “gimme.vertebrate.v5.0” was used as reference database.

We assessed TF binding motif enrichment in accessible peaks using chromVAR[97] (version 1.24.0). Briefly, we initially employed the “addMotifAnnotations” function in ArchR to determine motif presence in the peak set. Subsequently, GC bias was corrected using “addBgdPeaks”, followed by calculation of deviation z-scores for each TF motif in each cell using the “addDeviationsMatrix” function. The cell-by-motif matrix was then stored in ArchR arrow files.

Gene regulatory network inference

We used Python package CellOracle[28] to infer GRNs by integrating with our previously published snRNA-seq data[28]. The snRNA-seq data were processed by Seurat with default parameters, using the 'FindVariableFeatures' to select the top 3,000 highly variable genes in each corresponding cell subclass, and randomly sampled 5,000 RNA nuclei for each cell subclass. After that, we predicted GRNs by three steps. First, we defined the co-accessible distal-to-proximal pairs as described above. Second, we mapped distal CREs to known motifs by the "scan" function of GimmeMotifs, using the macaque genome macFas5 and the "gimme.vertebrate.v5.0" motif database as references. Finally, we identified the regulatory relationships between TFs and the potential target genes by fitting a regularized linear regression model using processed snRNA-seq data. We evaluated key TFs potentially involved in regulatory functions across different cell types based on "eigenvector_centrality" values. Subsequently, we filtered downstream target genes highly regulated according to their "coef_abs" values.
To identify hub TFs in the putative gene regulatory networks, we then calculated the hub score to infer the importance of different TFs. The hub score was calculated by "hub_score" function from R package igraph (v 2.1.1), with "coef_abs" value as a weight edge attribute

Integration of snATAC-seq and Stereo-seq data

Data preprocessing

To enable the integration of the query modality (Stereo-seq) and the reference modality (snATAC-seq), a systematic data preprocessing pipeline was designed. First, gene names from both modalities were extracted, and their intersection was computed to generate a shared gene set. The data were subsequently trimmed to a unified gene feature space. For snATAC-seq data, the gene score matrix was subjected to normalization and logarithmic transformation for each cell. Subsequently, the top 3,000 highly variable genes were selected to construct a high-information gene subset, which reduced data dimensionality while retaining critical biological features.

To simulate the complexities and uncertainties inherent in cross-modal data integration and to create a more challenging and generalizable data distribution, the gene score matrix of snATAC-seq was intentionally perturbed with added noise. During the perturbation process, a disturbance matrix p_{ik} was constructed based on the proportional calculation and normalization of gene expression values. This matrix comprised two components: gene-level perturbation and cell-level perturbation.

Gene-level perturbation was calculated based on the ratio of average gene expression values between the Stereo-seq and snATAC-seq modalities. Let X_{ik} denote the gene activity score matrix of the snATAC-seq modality and Y_{jk} denote the gene expression matrix of the Stereo-seq modality. The average expression value for each gene k in the two modalities is defined as:

$$X_k = \frac{1}{n_x}\sum_{i=1}^{n_x} X_{ik}, Y_k = \frac{1}{n_y}\sum_{j=1}^{n_y} Y_{jk}$$

where n_x and n_y represent the number of cells in the snATAC-seq and Stereo-seq modalities, respectively.

The average capture rate, $\bar{p}$ (p_{mean}), is defined as the ratio of the average gene expression values between the two modalities and calculated as:

$$\bar{p} = \frac{\frac{1}{n_x * g}\sum_{i=1}^{n_x}\sum_{k=1}^{g} X_{ik}}{\frac{1}{n_y * g}\sum_{j=1}^{n_y}\sum_{k=1}^{g} Y_{jk}}$$

Based on this, the gene-specific perturbation factor r_k is defined as:

$$r_k = \frac{X_k}{Y_k \cdot \bar{p}}$$

The cell-specific average capture rate p_i is calculated by matching the i-th quantiles of the library sizes between the query modality ($ln_{RNA,j}$) and the reference modality ($ln_{ATAC,i}$). This ratio is used to adjust the gene expression distribution of the reference modality. The formula is as follows:：

$$p_i = \frac{Q_i^{ln_{RNA,j}}}{Q_i^{ln_{ATAC,i}}}$$

The cell-specific perturbation factor r_i is obtained by dividing by p_{mean}：

$$r_i = \frac{p_i}{\bar{p}}$$

Finally, by combining the cell-specific perturbation factor and the gene-specific perturbation factor, the gene expression matrix is adjusted to generate the gene-cell perturbation ratio：

$$p_{ik} = r_i \cdot r_k \cdot \bar{p}$$

Sampling of the original snATAC-seq data is performed through the perturbation matrix, simulating a distribution that is closer to the target modality.

$$X'_{ik} \sim \mathrm{Binomial}(n = X_{ik}, p = p_{ik})$$

Data perturbed X'_{ik} in this manner serves as a reference, enhancing the model's robustness to noise in real biological data and providing a complex scenario simulation for experimental design and biological hypothesis testing.

Next, we similarly trim the Stereo-seq data based on the set of highly variable genes derived from the reference modality, ensuring that its features are consistent with those of the snATAC-seq data for seamless cross-modal analysis. Additionally, the trimmed Stereo-seq data undergoes normalization and logarithmic transformation, using the same strategy applied to the snATAC-seq data.

Encoder and decoder

The autoencoder (AE) module consists of an encoder and a decoder, both utilizing fully connected layers to achieve dimensionality reduction and reconstruction of high-dimensional gene features. The encoder maps the input high-dimensional gene expression data (with dimensions corresponding to the number of genes) to a lower-dimensional latent space (with dimensions corresponding to the number of cell types) using a single fully connected layer, with the bias set to False. In the latent space, U_{ic} and V_{jc} represent cell-type-specific metagene expressions for snATAC-seq and Stereo-seq, which are weighted combinations of different genes, W_0 denotes the weight matrix of the encoder as follows:

$$U_{ic} = \mathrm{Encoder}(X'_{ik}) = W_0 \cdot X'_{ik}$$

$$V_{jc} = \mathrm{Encoder}(Y_{jk}) = W_0 \cdot Y_{jk}$$

As the key component for dimensionality reduction, the encoder's fully connected layer extracts cell-specific features related to cell types through linear mapping, while incorporating Dropout (dropout rate = 0.8) during the mapping process to enhance the model's generalization ability and reduce the risk of overfitting.

The decoder employs a symmetric structure, with a fully connected layer that maps the latent variables from the low-dimensional metagene features back to the original high-dimensional space. The decoder's task is to supervise the encoder by reconstructing the input reference modality gene expression matrix, ensuring that the latent variables extracted by the encoder preserve the major biological information. W_1 denotes the weight matrix of the decoder as,

$$X''_{ik} = \mathrm{Decoder}(U_{ic}) = W_1 \cdot U_{ic}$$

The use of fully connected layers not only simplifies the model design but also enhances the ability to adapt to the dimensionality reduction characteristics of different modalities, enabling the encoder and decoder to exhibit strong robustness in handling high-dimensional sparse data.

Discriminator

The design of the discriminator module is based on the low-dimensional metagene features generated by the encoder, serving as an integral part of the adversarial mechanism. Its task is to distinguish the modality source of the metagene features. The input to the discriminator is the latent variables generated by the encoder, and the output is the probability of modality labels. During adversarial training, the encoder attempts to generate modality-agnostic metagene features to "fool" the discriminator, making it difficult to accurately determine the modality source, thereby achieving cross-modal feature alignment in the latent variable space. W_2 denotes the weight matrix of the discriminator:

$$P_{ATAC} = \mathrm{Sigmoid}(W_2 \cdot U_{ic})$$

$$P_{RNA} = \mathrm{Sigmoid}(W_2 \cdot V_{jc})$$

Loss function design

1. Discriminator Loss

The discriminator loss is used to measure the discriminator's ability to distinguish between two modalities. The goal is to maximize the probability of correct classification while ensuring that the contributions of different modalities are balanced. The formula for calculating the discriminator loss is as follows:

$$L_{DISC} = W_{ATAC} \cdot BCELoss(P_i^{ATAC}, 1) + W_{RNA} \cdot BCELoss(P_j^{RNA}, 0)$$

$$= -\frac{W_{ATAC}}{n_x}\sum_{i=1}^{n_x}[\log(P_i^{ATAC})] - \frac{W_{RNA}}{n_y}\sum_{j=1}^{n_y}[\log(1 - P_j^{RNA})]$$

Where：

$$W_{ATAC} = \frac{n_x + n_y}{2n_x}, W_{RNA} = \frac{n_x + n_y}{2n_y}$$

2. Cross-Entropy Loss

This loss is used to evaluate the difference between the model's predicted categories and the true categories. The formula for calculating the cross-entropy loss is as follows:

$$L_{CE} = -\sum_{c=1}^{M} W_c \cdot T_{ic} \cdot \log(U_{ic})$$

Where T_{ic} denotes the true label indicating whether the observed sample j belongs to category c. To address the class imbalance issue, we utilize class weights W_c to adjust the contribution of each class to the loss. The class weights are calculated as the inverse of the frequency n_c of each class:

$$W_c = \frac{1}{n_c}$$

3. Reconstruction Loss

The reconstruction loss is used to ensure that the output of the autoencoder is as close as possible to the input reference data, thereby preserving key information and minimizing information loss. The basic form of the reconstruction loss is the mean squared error (MSE), and its formula is as follows:

$$L_{RC} = \frac{1}{n_x}\sum_{i=1}^{n_x}(X'_{ik} - X''_{ik})$$

4. Adversarial Loss

The adversarial loss is used to optimize the feature extractor in adversarial learning, making its generated features difficult for the discriminator to distinguish between different modalities. This is achieved by training the feature extractor to "fool" the discriminator, producing modality-agnostic features. The formula is as follows:

$$L_{ADV} = W_{ATAC} \cdot BCELoss(P_i^{ATAC}, 0.5) + W_{RNA} \cdot BCELoss(P_{RNA,j}, 0.5)$$

$$= -\frac{W_{ATAC}}{2n_x}\sum_{i=1}^{n_x}[\log(P_i^{ATAC}) + \log(1 - P_i^{ATAC})] - \frac{W_{RNA}}{2n_y}\sum_{j=1}^{n_y}[\log(P_j^{RNA}) + \log(1 - P_j^{RNA})]$$

5. Entropy Regularization Loss

Entropy regularization is used to encourage a more uniform distribution of the model's output categories, thereby preventing overly confident predictions. The formula for calculating the entropy regularization loss is as follows:

$$L_{ER} = -\frac{1}{n_x}\sum_{i=1}^{n_x}\sum_{c=1}^{C} U_{ic}\log U_{ic}$$

6. L2 Regularization Loss

L2 regularization is used to constrain the magnitude of the model parameters, preventing overfitting. The formula is as follows:

$$L_2 = \sum_{w\in\theta} w^2$$

Where w represents the model parameters, which are the training parameters of the autoencoder.

Model Training

In the training process, a two-step training approach is employed, optimizing the discriminator and the feature extractor separately. This training strategy aims to achieve effective integration of the feature spaces from the two modalities (snATAC-seq and Stereo-seq) through adversarial learning, while simultaneously performing cell type classification and the reconstruction of input gene expression data. The training parameters set for the entire process include 1000 epochs, the Adam optimizer, and a learning rate of 0.01. Regularization parameters such as the L2 regularization coefficient are set to 0.01, and the dropout rate is set to 0.8.

In the first step, the parameters of the feature extractor (autoencoder module) are fixed, and only the discriminator is optimized. The task of the discriminator is to learn to distinguish the features of data from different modalities, with the input being the features extracted by the autoencoder from both modalities (e.g., snATAC-seq and Stereo-seq). The discriminator's loss function is based on binary cross-entropy, aiming to maximize the accuracy of modality discrimination, while balancing the class imbalances caused by unequal sample sizes through a weighted loss. Through optimization, the discriminator gradually learns how to distinguish the features of the two modalities, providing a clear optimization target for the adversarial optimization of the feature extractor in the next step. The loss function for the first step is as follows:

$$L_{Stage1} = L_{DISC}$$

In the second step, the parameters of the discriminator are fixed, and only the feature extractor is optimized. The task of the feature extractor is to generate modality-agnostic shared features, while simultaneously performing cell type classification and high-fidelity reconstruction of the input data. To achieve this, the optimization of the feature extractor includes a weighted combination of multiple loss functions:

$$L_{Stage2} = L_{CE} + L_{RC} + L_{ADV} + L_{ER} + L_2$$

These two steps are carried out sequentially. In each training round, the discriminator is first updated to enhance its ability to distinguish between modalities, followed by the update of the feature extractor to make the modalities indistinguishable. After multiple rounds of alternative training, the model reaches a balance: the shared features generated by the feature extractor both preserve biological information and are modality-agnostic; the discriminator can no longer improve its classification accuracy, indicating that modality alignment has been achieved.

One-to-one pairing between omics data

The metagene generated U_{ic},V_{jc} by the encoder is used for cell type prediction and cell matching. We employ the optimal transport algorithm for cross-modal mapping. The goal of optimal transport is to minimize the distance between cells. The key difference from traditional optimal transport algorithms lies in the computation of the cost matrix. Based on U_{ic}, V_{jc}, we rank the different cell types for each cell according to their probabilities, with the cell type showing the highest probability first and the lowest last, resulting in a type ranking vector sU_{ic},sV_{jc} for each cell. For two cells, i and j, from different modalities, if their top B_{ij} most likely cell types align, then the cost matrix is defined as:

$$C_{ij} = C - B_{ij}$$

Next, to solve the optimal transport problem, the following objective function is constructed:

$$\min_{T} \sum_{i,j} C_{ij} \cdot T_{ij}$$

T_{ij} represents the transport amount from cell i to cell j. We use Earth Mover's Distance (EMD) to solve this. For each cell j in Stereo-seq, we search for the cell i in snATAC-seq with the highest transport amount. In this way, each cell on the spatial transcriptome map was paired with chromatin accessibility from the snATAC-seq data. The chromatin accessibility of CREs was thus quantified in different layers and regions based on this mapping result.

Layer and regional preference of various cell types and corresponding CREs

Based on one-to-one cell pairing result between snATAC-seq data and Stereo-seq data, we obtained the spatial distribution of 230 cell types among six cortical layers and 143 brain regions, classified into 9 cortical lobes, including prefrontal, frontal, cingulate, somatosensory, insular, auditory, temporal, parietal, and occipital lobes.

The cell density per cell cluster in six layers was calculated as the number of cells in the layer divided by the area size of the certain layer. The preference of each cell cluster in different brain regions was calculated as follows:

$$I = \frac{1}{\sum_{j=1}^{S} \left(-P_j * log_{10} P_j\right)}$$

Here, I represented specificity index among cortical lobes, P_j is the proportion in a single brain region j, S is the number of brain regions the cell cluster was distributed.

Cross-species analysis of various cell types

Human and mouse snATAC-seq data were retrieved from previously published studies[11, 15] and downloaded from the chromatin accessibility database (http://catlas.org/humanbrain and http://catlas.org/mousebrain) and the GEO database[11] (accession number: GSE246791). We applied the subsampling method to ensure similar numbers of nuclei across species at the subclass level based on previous studies[6]. Specifically, the species with the largest number of clusters under a given subclass was allowed a maximum of 200 nuclei per cluster. The remaining species then split this theoretical maximum (200 nuclei multiplied by the maximum number of clusters under the subclass) evenly across their clusters. The intersection of homologous genes among the three species according to the Ensemble ortholog database (version 91) were used for cross-species analysis, and gene activities of the 14,382 orthologous genes were quantified from chromatin accessibility of snATAC-seq data using "addGeneScoreMatrix" and "getMatrixFromArrow" function of ArchR. We then performed data normalization using "NormalizeData" function. Further integration analysis was conducted with Seurat "IntegrateData" function with top 2000 repeatedly variable genes as features. The proportion of nuclei overlap between species was calculated by the number of cells clustered in the same clusters of the integrative clustering result divided by the total number of cells for specific cell types from two species, using the "compare_cl" function from the published pipeline[6]. MetaNeighbor analysis was run to systematically assess the similarity between cell types within species. The data slot of integrated assay was reformatted and used as input with the "SummarizedExperiment" function of SummarizedExperiment (v.1.28.0) package. We calculated the AUROC score by the "MetaNeighborUS" function of MetaNeighbor (v.1.18.0) package with the parameter "one_vs_best=TRUE".

Cross-species analysis of CREs

The human and mouse CRE lists were download from http://catlas.org/humanbrain and http://catlas.org/renlab_downloads/wholemousebrain/sa2.subclassv3.final.peak.srt. First, we converted macaque and mouse CREs to human coordinates using the liftOver method with minMatch = 0.3 and UCSC chain files ("macFas5.hg38.all.chain", "hg38.MacFas5.all.chain", "mm10.hg38.all.chain", and

1152 “hg38.mm10.all.chain”). Next, we reciprocally lifted the accessible elements back to raw genomes and only kept
1153 the regions that mapped to their original loci. We further removed converted regions with length > 1kb or in
1154 chrM, chr_random, chr_Un, and chr_alt chromosomes. Finally, we compared the homologous CREs using
1155 intervene[98] (v.0.6.5) with f = 0.1. CREs showing both sequence similarity and open chromatin between two
1156 species were considered as conserved, whereas CREs with no homologous sequence or no open chromatin peaks
1157 in the other species were considered as species-specific. In this way, we classified CREs into 7 categories:
1158 Human/Macaque/Mouse-conserved (HMm, mammal-conserved), Human/Macaque-biased (HM,
1159 Human/Macaque-biased), Human/Mouse-specific (Hm), Macaque/Mouse-specific (Mm), Human-specific,
1160 Macaque-specific, Mouse-specific. We then calculated the proportion of these 7 categories of CREs among all
1161 CREs in each cell subclass.
1162 In addition, we download the known transposable elements annotation file from UCSC. Only 4 TE types were
1163 kept for further analysis, including "DNA", "SINE", "LINE", "LTR". We calculated the percentage of CREs
1164 located in TE in these 7 categories, and performed pairwise Wilcoxon rank sum test with "bonferroni" correction
1165 for multiple testing among various categories.
1166
1167 **Comparison of LAMP5 and LAMP5/LHX6 between species**
1168 Human and mouse snATAC-seq data were retrieved from previously published studies[11, 15], and we filtered to
1169 retain cells originating from the cortex. Despite some differences in annotation between datasets, we identified
1170 the most comparable subclass groups using gene activity scores of classical markers while preserving the original
1171 labels from the previous studies for visualization.
1172 In order to compare the cell composition of LAMP5 and LAMP5/LHX6, we merged these two subclasses to
1173 calculate their proportions. To facilitate downstream computations, 5,000 cells were randomly sampled from the
1174 merged data for each species. Gene activity scores were normalized using a log2 transformation, and we used
1175 Seurat and Signac R package to perform integration.
1176 To identify the primate-enriched motifs, we utilized Jaspar motifs (JASPAR2020, all vertebrate) to generate a
1177 motif matrix and motif object that was added to the Seurat object using Signac “AddMotifs” function. For motif
1178 activity scores, chromVAR was carried out according to default parameters with the Signac “RunChromVAR”
1179 function on the respective peak count matrices. The motif matrices of human, macaque, and mouse were merged,
1180 and differential enrichment was identified using the “FindMarkers” function with the mean.fxn parameter set to
1181 rowMeans.
1182
1183 **Characteristic of SST layer and GRN construction**
1184 We re-clustered the SST cell subclass as described above to identify cell groups with layer preferences. The
1185 previously published study had identified the different layer-specific markers of SST neurons[21], so we examined
1186 the gene activity score of these layer markers and assign different cell types to the respective layers.
1187

The GRN (gene regulatory network) construction was performed as described above. Briefly, we used CellOracle to calculate the transcription factor (TF)-target gene correlations in different SST groups with layer preferences. For visualization, we ranked the TF by their "eigenvector_centralit" and selected the top 5 TFs. At the same time, we also ranked the target genes by "coef_abs" and selected the top 5 genes. In the network, the target genes which were identified as layer markers in the previous study[21] was highlighted.

Correlation of OPC-specific and OLG-specific CREs with cortical hierarchy levels

Hierarchy levels of cortical regions in the visual system were derived from Felleman and Van Essen[51]. Based on one-to-one cell pairing result between snATAC-seq data and Stereo-seq data, we extracted the chromatin accessibility matrix of OPC-specific CREs in OPC cells and OLG-specific CREs in OLG cells. Then, we calculated the average chromatin accessibility for each OPC-specific CRE among all OPC cells and OLG-specific CRE among all OLG cells in various cortical regions. We applied the Spearman analysis to calculate the correlation between averaged chromatin accessibility and the hierarchy level for each CREs using the "cor.test" function. The CREs showing positive or negative correlation with the hierarchy levels were identified with cutoffs of P value < 0.01, $R > 0.4$ or $R < -0.4$. GO enrichment analysis was performed using these CREs associated genes by clusterProfiler (v.4.6.0).

LDSC cell type specific analysis

To enable comparison with GWAS data of human phenotypes, we utilized the `liftOver` with default parameters to convert cell type-specific CREs from "Macaca_fascicularis_5.0" to hg19 genomic coordinates via a chain file "macFas5ToHg19.over.chain" download from UCSC. We further filtered out converted regions exceeding 1 kb and retained only coordinates on each standard chromosome (chromosomes 1-22).

We used Linkage Disequilibrium Score Regression Analysis (LDSC) to assess the extent of genetic polymorphism enrichment associated with chromatin accessible regions differing among various cell types and traits. Summary statistics for GWAS of brain disorder, neurological traits, control traits were obtained from the GWAS Catalog and publications (Supplementary Table 7). We prepared the summary statistics in the standard format for linkage disequilibrium score regression and downloaded the relevant genotype resource from the pipline vignettes at https://github.com/bulik/ldsc. Using genotype data from the "1000G_Phase3_plinkfiles.tgz" and haplotype information from 'hapmap3_snps.tgz' as reference panels, we generated binary annotations for the homologous DNA sequences of each cell type and estimated LD scores on each standard chromosome. For each trait, we used the baseline model "1000G_Phase3_baseline_ldscores.tgz" along with the standardized regression weights "weights_hm3_no_hla.tgz" to partition phenotypic heritability into cell type-specific annotations, using the superset of all candidate regulatory peaks as background controls.

Variant to CRE mapping

To broadly map the genomic features of disease risk signals across species, we first extracted all potential pathogenic variants from the GWAS summary data that passed the statistical test (P-value ≤ 0.05). Next, we used the R package “SNPlocs.Hsapiens.dbSNP144.GRCh37” along with their rsid information to retrieve the coordinates of these signals in the hg19 reference genome from the "dbSNP144" database. Subsequently, we performed intersection analysis between these coordinates and Cis-Regulatory Elements (CREs) in the Macaca fascicularis genome to identify disease risk-related CREs that may have similar functions across different species.

Gene set related to brain disorder

The gene set related to brain disorder was obtained from Disease Ontology (DO). In R, we used the getHgDisease function from the geneset package to load the DO database by specifying source = "do". We then queried the DOID (Supplementary Table 7) and employed the transId method from the genekitr package, setting transTo = "symbol" and unique = TRUE, to retrieve the gene symbols.

Sequence evolutionary conservation analysis

The phyloP score is a conservation score that quantifies the evolutionary conservation of each nucleotide by comparing the aligned sequences across multiple species[58, 59]. By evaluating phyloP scores, we assessed the evolutionary conservation of CREs associated with disease risk loci and those not associated with disease, in the context of cell-type genetic identity across species, relative to a shuffle background.

To focus the nucleotide evolutionary conservation assessment on comparisons between the Macaca fascicularis and human and mouse genomes, we first downloaded whole-genome masked sequences for three species (macFas5, mm10, and hg38) from UCSC, retaining the standard chromosome assemblies (sequences prefixed with "chr") for alignment analysis. Next, we performed reference-based alignment using the `LASTZ` software, with parameter settings according to UCSC guidelines. For the closely related species macFas5 and hg38, we used the parameters “M=254 K=3000 L=3000 Y=3400 E=30 H=2000 O=400 T=2”; for the more distantly related macFas5 and mm10, we used the parameters “M=0 K=3000 L=3000 Y=9400 E=30 H=2000 O=400 T=1”. Subsequently, we applied `MULTIZ` to assemble the pairwise `LASTZ` alignment results into a multi-genome alignment output and set the phylogenetic tree for the three species to “((macfas5, hg38), mm10)” using the query from https://lifemap-ncbi.univ-lyon1.fr.

For evolutionary model generation, we used `phyloFit` with the parameters `-i MAF --tree "((macfas5, hg38), mm10)" --subst-mod REV`; then, we used `phyloP` to generate per-base phyloP score data, with the parameters `--mode CONACC --method LRT --wig-scores` to compare the evolutionary conservation of the macaque genome against human and mouse genomes. Finally, we obtained the average phyloP scores for all CREs using `bigWigAverageOverBed`.

In the statistical analysis, for each cell type's diff-CREs that co-localized with disease risk signals, we calculated its average phyloP score. To ensure a fair comparison of the evolutionary conservation of these Risk CREs, we performed 1000 bootstrap resampling on other diff-elements not belonging to Risk CREs and the supercluster of peaks, each time sampling an equal number of elements as the Risk CREs. We then aggregated the average scores from the 1000 re-samplings for statistical analysis.

Gene Ontology (GO) enrichment

Gene Ontology annotation of interested CREs was performed using R package rGREAT[99] (v.2.2.0). We downloaded AH111915 (Macaca fascicularis) database as reference using AnnotationHub (v. 3.8.0). For the pathway gene sets of Gene Ontology database, we extended the genes from its transcription start site with default parameters of "basal_upstream = 5000" and "basal_downstream = 1000". The total CREs of corresponding cell clusters were used as the background peak set for each. GREAT analysis was performed with "great" function of "min_gene_set_size = 5".

RNA Scope

Macaque #3 was transcardially perfused with diethyl pyrocarbonate-treated PBS (DEPC-PBS), and then followed by 4% DEPC-PFA. The brain was removed and immediately embedded in optimum cutting temperature compound (OCT, SAKURA, cat#4583). The brain samples were stored at -80 ° C. During samples processing, the OCT blocks were sectioned at 15 μm using a RWD freezing microtome, and brain sections were fixed onto adhesion microscope slides. The slides were baked at 60° C overnight. Then, RNA Scope Pretreatment Kit (Advanced Cell Diagnostics, cat#322381 and cat#322000) was used to unmask target RNA and permeabilize cells. Subsequently, RNA Scope Target Probe are hybridized to target RNA molecules. RNA Scope Detection Reagents (ACD, cat#322310) amplify the hybridization signals via sequential hybridization of amplifiers and color label probes. The RNA Scope procedure was performed according to the manufacturer's instructions (https://acdbio.com). The details about the probes employed in the experiments are as follows: Probes against CPA6 (ACD, cat#1843691-C1) and MEF2C (ACD, cat#1843711-C2) were used to identify V1-specific L4 neurons in macaque monkeys; Probes against DCC (ACD, cat#1843701-C1) were used to identify layer4 neurons in macaque monkeys;

Data visualization

For bigwig files, presented by Signac package "BigwigTrack" function. For the putative gene regulatory networks, presented with Cytoscape (v.3.10.1) tools. The residual graphs were plotted with R package ggplot2 (v.3.4.1) and pheatmap (v.1.0.12).

Data availability

The processed data ready for exploration could be accessed and downloaded via https://macaque.digital-brain.cn/chromatin-accessibility (reviewers can download the data using the following ids and passwords: All raw data have been deposited to CNGB Nucleotide Sequence Archive (accession code: CNP0005530) and are publicly available as of the date of publication.

Code availability

All data were analyzed with standard programs and packages, as detailed above, and the corresponding codes were accessible through https://github.com/JuneMeng/snATAC_MacaqueCTX_Atlas. Custom code of cell-cell pairing of snATAC-seq and spatial transcriptome data is available at https://github.com/sunyk740/RMGE/. Additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

Reference

1. Lein E, Borm LE, Linnarsson S. The promise of spatial transcriptomics for neuroscience in the era of molecular cell typing. *Science* **358**, 64-69 (2017).

2. Zhu Y*, et al.* Spatiotemporal transcriptomic divergence across human and macaque brain development. *Science* **362**, (2018).

3. Khrameeva E*, et al.* Single-cell-resolution transcriptome map of human, chimpanzee, bonobo, and macaque brains. *Genome Res* **30**, 776-789 (2020).

4. Krienen FM*, et al.* Innovations present in the primate interneuron repertoire. *Nature* **586**, 262-269 (2020).

5. Berg J*, et al.* Human neocortical expansion involves glutamatergic neuron diversification. *Nature* **598**, 151-158 (2021).

6. Bakken TE*, et al.* Comparative cellular analysis of motor cortex in human, marmoset and mouse. *Nature* **598**, 111-119 (2021).

7. Network BICC. A multimodal cell census and atlas of the mammalian primary motor cortex. *Nature* **598**, 86-102 (2021).

8. Yao Z*, et al.* A transcriptomic and epigenomic cell atlas of the mouse primary motor cortex. *Nature* **598**, 103-110 (2021).

9. 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).

10. Cusanovich DA*, et al.* A Single-Cell Atlas of In Vivo Mammalian Chromatin Accessibility. *Cell* **174**, 1309-1324 e1318 (2018).

11. Zu S*, et al.* Single-cell analysis of chromatin accessibility in the adult mouse brain. *Nature* **624**, 378-389 (2023).

12. Li YE*, et al.* An atlas of gene regulatory elements in adult mouse cerebrum. *Nature* **598**, 129-136 (2021).

13. Villar D*, et al.* Enhancer evolution across 20 mammalian species. *Cell* **160**, 554-566 (2015).

14. Won H*, et al.* Chromosome conformation elucidates regulatory relationships in developing human brain. *Nature* **538**, 523-527 (2016).

15. Li YE*, et al.* A comparative atlas of single-cell chromatin accessibility in the human brain. *Science* **382**, eadf7044 (2023).

16. Zemke NR*, et al.* Conserved and divergent gene regulatory programs of the mammalian neocortex. *Nature* **624**, 390-402 (2023).

17. Yin S*, et al.* Transcriptomic and open chromatin atlas of high-resolution anatomical regions in the rhesus macaque brain. *Nat Commun* **11**, 474 (2020).

18. Chiou KL*, et al.* A single-cell multi-omic atlas spanning the adult rhesus macaque brain. *Sci Adv* **9**, eadh1914 (2023).

19. Lei Y*, et al.* Spatially resolved gene regulatory and disease-related vulnerability map of the adult Macaque cortex. *Nat Commun* **13**, 6747 (2022).

20. Han L*, et al.* Cell transcriptomic atlas of the non-human primate Macaca fascicularis. *Nature* **604**, 723-731 (2022).

21. Chen A*, et al.* Single-cell spatial transcriptome reveals cell-type organization in the macaque cortex. *Cell* **186**, 3726-3743 e3724 (2023).

22. Hodge RD*, et al.* Conserved cell types with divergent features in human versus mouse cortex. *Nature* **573**, 61-68 (2019).

23. Yao Z*, et al.* A taxonomy of transcriptomic cell types across the isocortex and hippocampal formation. *Cell* **184**, 3222-3241 e3226 (2021).

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

25. Su Y*, et al.* Neuronal activity modifies the chromatin accessibility landscape in the adult brain. *Nat Neurosci* **20**, 476-483 (2017).

26. Pliner HA*, et al.* Cicero Predicts cis-Regulatory DNA Interactions from Single-Cell Chromatin Accessibility Data. *Mol Cell* **71**, 858-871 e858 (2018).

27. Reiprich S*, et al.* Transcription factor Sox10 regulates oligodendroglial Sox9 levels via microRNAs. *Glia* **65**, 1089-1102 (2017).

28. Kamimoto K, Stringa B, Hoffmann CM, Jindal K, Solnica-Krezel L, Morris SA. Dissecting cell identity via network inference and in silico gene perturbation. *Nature* **614**, 742-751 (2023).

29. Basu S*, et al.* The Mef2c Gene Dose-Dependently Controls Hippocampal Neurogenesis and the Expression of Autism-Like Behaviors. *J Neurosci* **44**, (2024).

30. Pandya D, Seltzer B, Petrides M, Cipolloni PB. *Cerebral Cortex: Architecture, Connections, and the Dual Origin Concept*. Oxford University Press (2015).

31. Shibata M*, et al.* Hominini-specific regulation of CBLN2 increases prefrontal spinogenesis. *Nature* **598**, 489-494 (2021).

32. Crow M, Paul A, Ballouz S, Huang ZJ, Gillis J. Characterizing the replicability of cell types defined by single cell RNA-sequencing data using MetaNeighbor. *Nat Commun* **9**, 884 (2018).

33. Manitt C*, et al.* dcc orchestrates the development of the prefrontal cortex during adolescence and is altered in psychiatric patients. *Transl Psychiatry* **3**, e338 (2013).

34. Gabitto MI*, et al.* Integrated multimodal cell atlas of Alzheimer's disease. *Nat Neurosci* **27**, 2366-2383 (2024).

35. Seo MH, Lim S, Yeo S. Association of decreased triadin expression level with apoptosis of dopaminergic cells in Parkinson's disease mouse model. *BMC Neurosci* **22**, 65 (2021).

36. Chuong EB, Elde NC, Feschotte C. Regulatory activities of transposable elements: from conflicts to benefits. *Nat Rev Genet* **18**, 71-86 (2017).

37. Tasic B*, et al.* Shared and distinct transcriptomic cell types across neocortical areas. *Nature* **563**, 72-78 (2018).

38. Gouwens NW, *et al.* Integrated Morphoelectric and Transcriptomic Classification of Cortical GABAergic Cells. *Cell* **183**, 935-953 e919 (2020).

39. Wei JR, *et al.* Identification of visual cortex cell types and species differences using single-cell RNA sequencing. *Nat Commun* **13**, 6902 (2022).

40. Chen WG, *et al.* Upstream stimulatory factors are mediators of Ca2+-responsive transcription in neurons. *J Neurosci* **23**, 2572-2581 (2003).

41. Flavell SW, *et al.* Genome-wide analysis of MEF2 transcriptional program reveals synaptic target genes and neuronal activity-dependent polyadenylation site selection. *Neuron* **60**, 1022-1038 (2008).

42. Micali N, *et al.* Molecular programs of regional specification and neural stem cell fate progression in macaque telencephalon. *Science* **382**, eadf3786 (2023).

43. Jahangir M, Shah SM, Zhou JS, Lang B, Wang XP. Parvalbumin interneurons in the anterior cingulate cortex exhibit distinct processing patterns for fear and memory in rats. *Heliyon* **11**, e41218 (2025).

44. Qi C, *et al.* Anterior cingulate cortex parvalbumin and somatostatin interneurons shape social behavior in male mice. *Nat Commun* **16**, 4156 (2025).

45. Fang R, *et al.* Conservation and divergence of cortical cell organization in human and mouse revealed by MERFISH. *Science* **377**, 56-62 (2022).

46. Shi Y, *et al.* Mouse and human share conserved transcriptional programs for interneuron development. *Science* **374**, eabj6641 (2021).

47. Urban-Ciecko J, Barth AL. Somatostatin-expressing neurons in cortical networks. *Nat Rev Neurosci* **17**, 401-409 (2016).

48. Berto S, *et al.* Accelerated evolution of oligodendrocytes in the human brain. *Proc Natl Acad Sci U S A* **116**, 24334-24342 (2019).

49. Miller DJ, *et al.* Prolonged myelination in human neocortical evolution. *Proc Natl Acad Sci U S A* **109**, 16480-16485 (2012).

50. Caglayan E, *et al.* Molecular features driving cellular complexity of human brain evolution. *Nature* **620**, 145-153 (2023).

51. Felleman DJ, Van Essen DC. Distributed hierarchical processing in the primate cerebral cortex. *Cereb Cortex* **1**, 1-47 (1991).

52. Ghorbani S, *et al.* Fibulin-2 is an extracellular matrix inhibitor of oligodendrocytes relevant to multiple sclerosis. *J Clin Invest* **134**, (2024).

53. Li Y, *et al.* Npas3 deficiency impairs cortical astrogenesis and induces autistic-like behaviors. *Cell Rep* **40**, 111289 (2022).

54. Cummings AC, *et al.* Genome-wide association and linkage study in the Amish detects a novel candidate late-onset Alzheimer disease gene. *Ann Hum Genet* **76**, 342-351 (2012).

55. Schizophrenia Working Group of the Psychiatric Genomics C. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421-427 (2014).

56. Loo SK, *et al.* Genome-wide association study of intelligence: additive effects of novel brain expressed genes. *J Am Acad Child Adolesc Psychiatry* **51**, 432-440 e432 (2012).

57. Seshadri S, *et al.* Genetic correlates of brain aging on MRI and cognitive test measures: a genome-wide association and linkage analysis in the Framingham Study. *BMC Med Genet* **8 Suppl 1**, S15 (2007).

58. Zhang G, *et al.* Construction of single-cell cross-species chromatin accessibility landscapes with combinatorial-hybridization-based ATAC-seq. *Dev Cell* **59**, 793-811 e798 (2024).

59. Pollard KS, Hubisz MJ, Rosenbloom KR, Siepel A. Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res* **20**, 110-121 (2010).

60. Bipolar D, Schizophrenia Working Group of the Psychiatric Genomics Consortium. Electronic address drve, Bipolar D, Schizophrenia Working Group of the Psychiatric Genomics C. Genomic Dissection of Bipolar Disorder and Schizophrenia, Including 28 Subphenotypes. *Cell* **173**, 1705-1715 e1716 (2018).

61. Merenlender-Wagner A, *et al.* Autophagy has a key role in the pathophysiology of schizophrenia. *Mol Psychiatry* **20**, 126-132 (2015).

62. Liu X, Bipolar Genome S, Kelsoe JR, Greenwood TA. A genome-wide association study of bipolar disorder with comorbid eating disorder replicates the SOX2-OT region. *J Affect Disord* **189**, 141-149 (2016).

63. Kahler AK, *et al.* Association analysis of schizophrenia on 18 genes involved in neuronal migration: MDGA1 as a new susceptibility gene. *Am J Med Genet B Neuropsychiatr Genet* **147B**, 1089-1100 (2008).

64. Banfi F, *et al.* SETBP1 accumulation induces P53 inhibition and genotoxic stress in neural progenitors underlying neurodegeneration in Schinzel-Giedion syndrome. *Nat Commun* **12**, 4050 (2021).

65. Yamazaki K, *et al.* ABCA7 Gene Expression and Genetic Association Study in Schizophrenia. *Neuropsychiatr Dis Treat* **16**, 441-446 (2020).

66. Gotoh L, *et al.* Association analysis of adenosine A1 receptor gene (ADORA1) polymorphisms with schizophrenia in a Japanese population. *Psychiatr Genet* **19**, 328-335 (2009).

67. Takahashi N, *et al.* Association between chromogranin A gene polymorphism and schizophrenia in the Japanese population. *Schizophr Res* **83**, 179-183 (2006).

68. Pozhidaev IV, *et al.* Association of Cholinergic Muscarinic M4 Receptor Gene Polymorphism with Schizophrenia. *Appl Clin Genet* **13**, 97-105 (2020).

69. Li M, *et al.* Novel genetic susceptibility loci identified by family based whole exome sequencing in Han Chinese schizophrenia patients. *Transl Psychiatry* **10**, 5 (2020).

70. Zhang Z, *et al.* The Schizophrenia Susceptibility Gene OPCML Regulates Spine Maturation and Cognitive Behaviors through Eph-Cofilin Signaling. *Cell Rep* **29**, 49-61 e47 (2019).

71. Ventriglia M, *et al.* Allelic variation in the human prodynorphin gene promoter and schizophrenia. *Neuropsychobiology* **46**, 17-21 (2002).

72. Juraeva D, *et al.* Integrated pathway-based approach identifies association between genomic regions at CTCF and CACNB2 and schizophrenia. *PLoS Genet* **10**, e1004345 (2014).

73. Tiihonen J*, et al.* Molecular signaling pathways underlying schizophrenia. *Schizophr Res* **232**, 33-41 (2021).

74. 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).

75. Aberg KA*, et al.* Methylome-wide association study of schizophrenia: identifying blood biomarker signatures of environmental insults. *JAMA Psychiatry* **71**, 255-264 (2014).

76. Souza RP*, et al.* Genetic association of the GDNF alpha-receptor genes with schizophrenia and clozapine response. *J Psychiatr Res* **44**, 700-706 (2010).

77. Pinacho R*, et al.* Altered CSNK1E, FABP4 and NEFH protein levels in the dorsolateral prefrontal cortex in schizophrenia. *Schizophr Res* **177**, 88-97 (2016).

78. Shim KH*, et al.* Subsequent correlated changes in complement component 3 and amyloid beta oligomers in the blood of patients with Alzheimer's disease. *Alzheimers Dement* **20**, 2731-2741 (2024).

79. Colonna M, Wang Y. TREM2 variants: new keys to decipher Alzheimer disease pathogenesis. *Nat Rev Neurosci* **17**, 201-207 (2016).

80. Wang Y*, et al.* Key variants via the Alzheimer's Disease Sequencing Project whole genome sequence data. *Alzheimers Dement* **20**, 3290-3304 (2024).

81. Joshi H*, et al.* Decreased Expression of Cerebral Dopamine Neurotrophic Factor in Platelets of Probable Alzheimer Patients. *Alzheimer Dis Assoc Disord* **36**, 269-271 (2022).

82. Corraliza-Gomez M*, et al.* Insulin-degrading enzyme (IDE) as a modulator of microglial phenotypes in the context of Alzheimer's disease and brain aging. *J Neuroinflammation* **20**, 233 (2023).

83. Chang JY, Chang NS. WWOX dysfunction induces sequential aggregation of TRAPPC6ADelta, TIAF1, tau and amyloid beta, and causes apoptosis. *Cell Death Discov* **1**, 15003 (2015).

84. Chiba-Falek O, Gottschalk WK, Lutz MW. The effects of the TOMM40 poly-T alleles on Alzheimer's disease phenotypes. *Alzheimers Dement* **14**, 692-698 (2018).

85. Jackson RJ, Hyman BT, Serrano-Pozo A. Multifaceted roles of APOE in Alzheimer disease. *Nat Rev Neurol* **20**, 457-474 (2024).

86. Reynolds LM*, et al.* DCC Receptors Drive Prefrontal Cortex Maturation by Determining Dopamine Axon Targeting in Adolescence. *Biol Psychiatry* **83**, 181-192 (2018).

87. Li HJ*, et al.* Further confirmation of netrin 1 receptor (DCC) as a depression risk gene via integrations of multi-omics data. *Transl Psychiatry* **10**, 98 (2020).

88. Vosberg DE, Leyton M, Flores C. The Netrin-1/DCC guidance system: dopamine pathway maturation and psychiatric disorders emerging in adolescence. *Mol Psychiatry* **25**, 297-307 (2020).

89. Tricoire L*, et al.* Common origins of hippocampal Ivy and nitric oxide synthase expressing neurogliaform cells. *J Neurosci* **30**, 2165-2176 (2010).

90. Kosoy R*, et al.* Genetics of the human microglia regulome refines Alzheimer's disease risk loci. *Nat Genet* **54**, 1145-1154 (2022).

91. Mathys H*, et al.* Single-cell atlas reveals correlates of high cognitive function, dementia, and resilience to Alzheimer's disease pathology. *Cell* **186**, 4365-4385 e4327 (2023).

92. Bakken TE*, et al.* Single-nucleus and single-cell transcriptomes compared in matched cortical cell types. *PLoS One* **13**, e0209648 (2018).

93. Granja JM*, et al.* ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. *Nat Genet* **53**, 403-411 (2021).

94. Zeisel A*, et al.* Molecular Architecture of the Mouse Nervous System. *Cell* **174**, 999-1014 e1022 (2018).

95. Yu G, Wang LG, He QY. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics* **31**, 2382-2383 (2015).

96. van Heeringen SJ, Veenstra GJ. GimmeMotifs: a de novo motif prediction pipeline for ChIP-sequencing experiments. *Bioinformatics* **27**, 270-271 (2011).

97. Schep AN, Wu B, Buenrostro JD, Greenleaf WJ. chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. *Nat Methods* **14**, 975-978 (2017).

98. McLean CY*, et al.* GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotechnol* **28**, 495-501 (2010).

99. Gu Z, Hubschmann D. rGREAT: an R/bioconductor package for functional enrichment on genomic regions. *Bioinformatics* **39**, (2023).

Acknowledgements

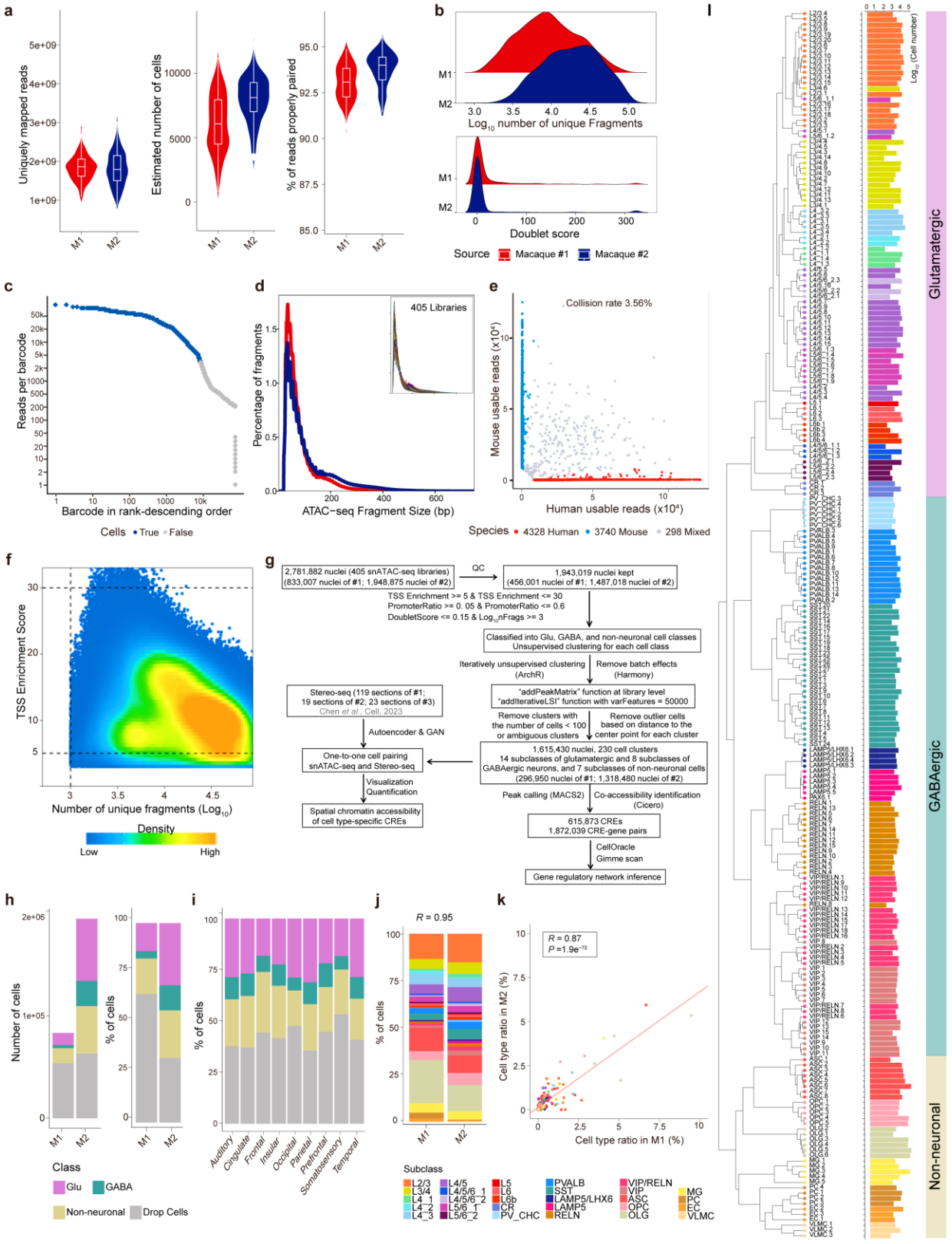
The project was supported by National Science and Technology Innovation 2030 Major Program (STI2030-2021ZD0200100), National Natural Science Foundation of China (NSFC) Outstanding Youth Foundation (No. T2422026), National Key R&D Program of China No.2022YEF0203200, National Natural Science Foundation of China (No. U23A6010, 32100487) and Shanghai Science and Technology Development Funds (23QA1410400).

Author contributions

Methodology, J.M., C.C., Z.Z., Y.S., K.H. and T.F.; Investigation, J.M., Y.L., C.C., Z.Z., Y.B.; Formal Analysis, J.M., C.C., Z.Z., Y.S. and Y.H.; Writing – Original Draft, Y.L. and Y-D.S.; Writing – Review & Editing, J.M., T.F., Z.L., Z.S., L.L., C.L., and M.P.; Conceptualization and supervision, S.L. and Y-D.S..

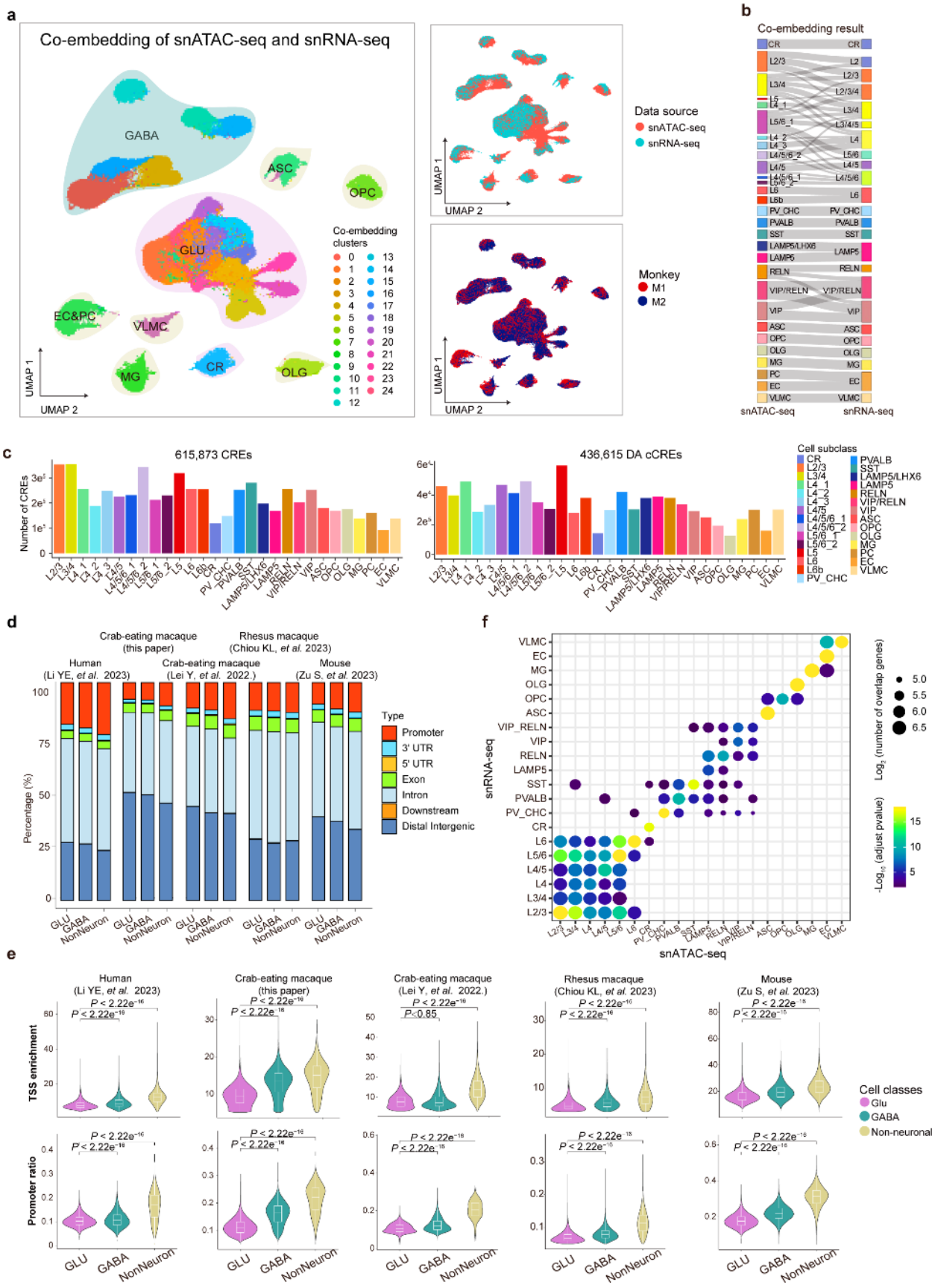
Competing interests

All authors declare no competing interests.


a
Uniquely mapped reads
Estimated number of cells
% of reads properly paired
M1
M2
b
Log10 number of unique Fragments
Doublet score
Source
Macaque #1
Macaque #2
c
Reads per barcode
Barcode in rank-descending order
Cells
True
False
d
405 Libraries
Percentage of fragments
ATAC−seq Fragment Size (bp)
e
Collision rate 3.56%
Mouse usable reads (x10⁴)
Human usable reads (x10⁴)
Species
4328 Human
3740 Mouse
298 Mixed
f
TSS Enrichment Score
Number of unique fragments (Log10)
Density
Low
High
g
2,781,882 nuclei (405 snATAC-seq libraries)
(833,007 nuclei of #1; 1,948,875 nuclei of #2)
QC
1,943,019 nuclei kept
(456,001 nuclei of #1; 1,487,018 nuclei of #2)
TSS Enrichment >= 5 & TSS Enrichment <= 30
PromoterRatio >= 0. 05 & PromoterRatio <= 0.6
DoubletScore <= 0.15 & Log10nFrags >= 3
Classified into Glu, GABA, and non-neuronal cell classes
Unsupervised clustering for each cell class
Iteratively unsupervised clustering
(ArchR)
Remove batch effects
(Harmony)
"addPeakMatrix" function at library level
"addIterativeLSI" function with varFeatures = 50000
Stereo-seq (119 sections of #1;
19 sections of #2; 23 sections of #3)
Chen et al., Cell, 2023
Remove clusters with
the number of cells < 100
or ambiguous clusters
Remove outlier cells
based on distance to the
center point for each cluster
Autoencoder & GAN
One-to-one cell pairing
snATAC-seq and Stereo-seq
1,615,430 nuclei, 230 cell clusters
14 subclasses of glutamatergic and 8 subclasses of
GABAergic neurons, and 7 subclasses of non-neuronal cells
(296,950 nuclei of #1; 1,318,480 nuclei of #2)
Visualization
Quantification
Spatial chromatin accessibility
of cell type-specific CREs
Peak calling (MACS2)
Co-accessibility identification
(Cicero)
615,873 CREs
1,872,039 CRE-gene pairs
CellOracle
Gimme scan
Gene regulatory network inference
h
Number of cells
% of cells
Class
Glu
GABA
Non-neuronal
Drop Cells
i
% of cells
Auditory
Cingulate
Frontal
Insular
Occipital
Parietal
Prefrontal
Somatosensory
Temporal
j
R = 0.95
% of cells
Subclass
L2/3
L3/4
L4_1
L4_2
L4_3
L4/5
L4/5/6_1
L4/5/6_2
L5/6_1
L5/6_2
L5
L6
L6b
CR
PV_CHC
PVALB
SST
LAMP5/LHX6
LAMP5
RELN
VIP/RELN
VIP
ASC
OPC
OLG
MG
PC
EC
VLMC
k
R = 0.87
P = 1.9e−72
Cell type ratio in M2 (%)
Cell type ratio in M1 (%)
l
Log10 (Cell number)
Glutamatergic
GABAergic
Non-neuronal

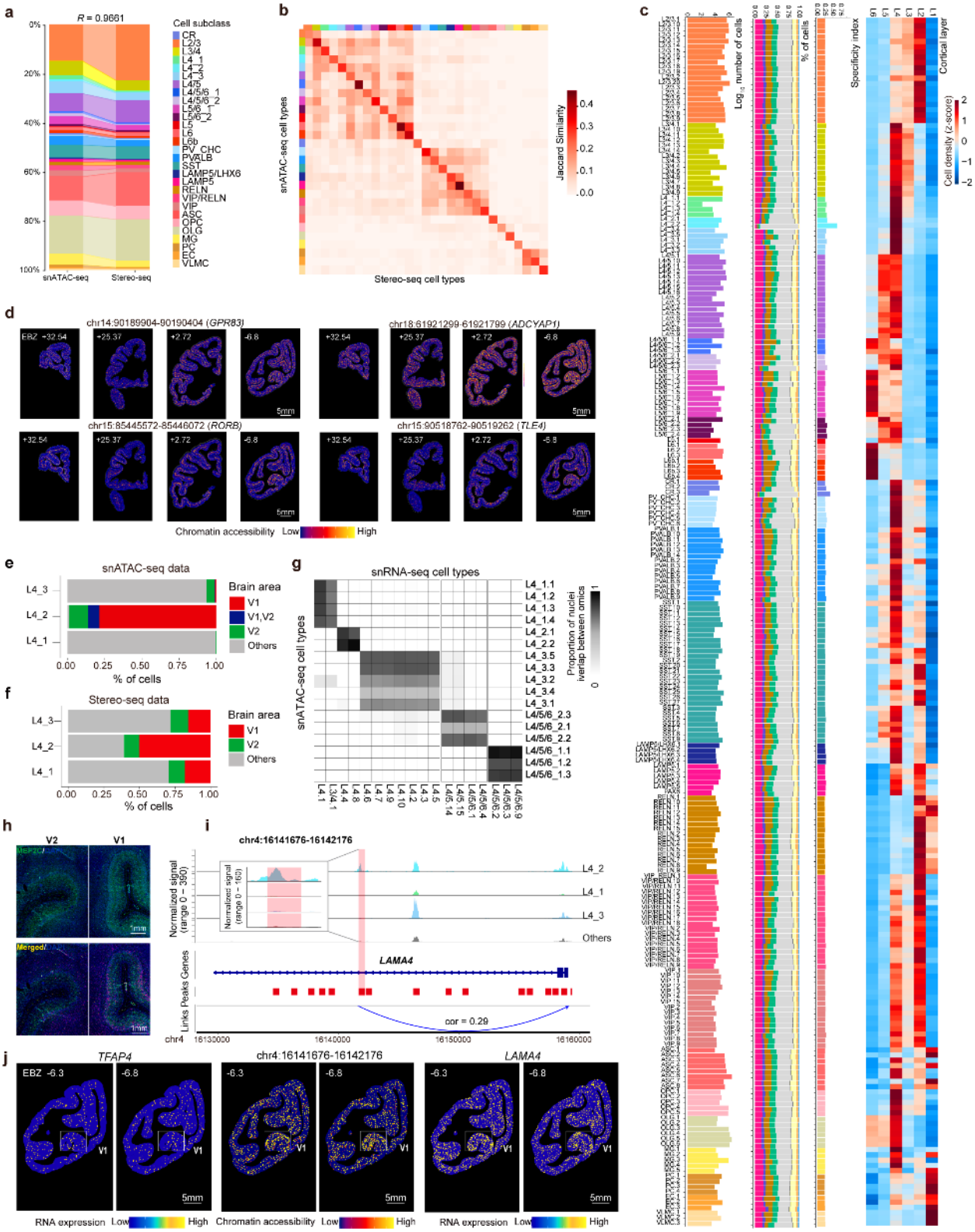
1633

Extended Data Fig. 1. Quality control metrics of snATAC-seq data. a, Box plots showing the distribution of uniquely mapped reads (left), estimated numbers of cells (middle) and percent of reads properly paired (right) in biological replicates Macaque #1 and #2 of snATAC-seq libraries from each tissue sample. **b**, Density plots showing the distribution of number of unique fragments (upper) and doublet score (lower) in biological replicates Macaque #1 and #2 of snATAC-seq libraries. **c**, Knee plot for the detection of cells by the relationship between reads per barcode vs. the number of barcodes. **d**, Frequency distribution plot showing the fragment size distribution of each snATAC-seq library. **e**, Fraction of cell collision estimated from species mix samples of human and mouse. **f**, Density plot illustrating fragments per nucleus and individual TSS enrichment. Nuclei in middle right quadrant were selected for analysis (5 < TSS enrichment < 30 and number of unique fragments > 1000). **g**, The quality control process for the snATAC-seq dataset. **h**, Bar plot showing the numbers (left) and percentages (right) of glutamatergic, GABAergic and non-neuronal cells in biological replicates Macaque #1 and #2. **i**, Bar plot showing percentages of glutamatergic, GABAergic and non-neuronal cells in 9 cortical lobes. **j**, Bar plot showing percentages of various cell clusters in biological replicates Macaque #1 and #2. **k**, Dot plot showing the relationship of cell type compositions between Macaque #1 and #2. Correlation coefficient and *P* value was correlated by Pearson correlation test. **l**, Hierarchical clustering of 230 cell subclusters based on chromatin accessibility and the number of nuclei in each subcluster.


a
Co-embedding of snATAC-seq and snRNA-seq
GABA
ASC
OPC
GLU
EC&PC
VLMC
MG
CR
OLG
UMAP 1
UMAP 2
Co-embedding clusters
Data source
snATAC-seq
snRNA-seq
Monkey
M1
M2
b
Co-embedding result
snATAC-seq
snRNA-seq
c
615,873 CREs
436,615 DA cCREs
Number of CREs
Cell subclass
d
Crab-eating macaque (this paper)
Rhesus macaque (Chiou KL, et al. 2023)
Human (Li YE, et al. 2023)
Crab-eating macaque (Lei Y, et al. 2022.)
Mouse (Zu S, et al. 2023)
Percentage (%)
Type
Promoter
3' UTR
5' UTR
Exon
Intron
Downstream
Distal Intergenic
GLU
GABA
NonNeuron
f
snRNA-seq
snATAC-seq
Log2 (number of overlap genes)
-Log10 (adjust pvalue)
e
Human (Li YE, et al. 2023)
Crab-eating macaque (this paper)
Crab-eating macaque (Lei Y, et al. 2022.)
Rhesus macaque (Chiou KL, et al. 2023)
Mouse (Zu S, et al. 2023)
TSS enrichment
Promoter ratio
Cell classes
Glu
GABA
Non-neuronal

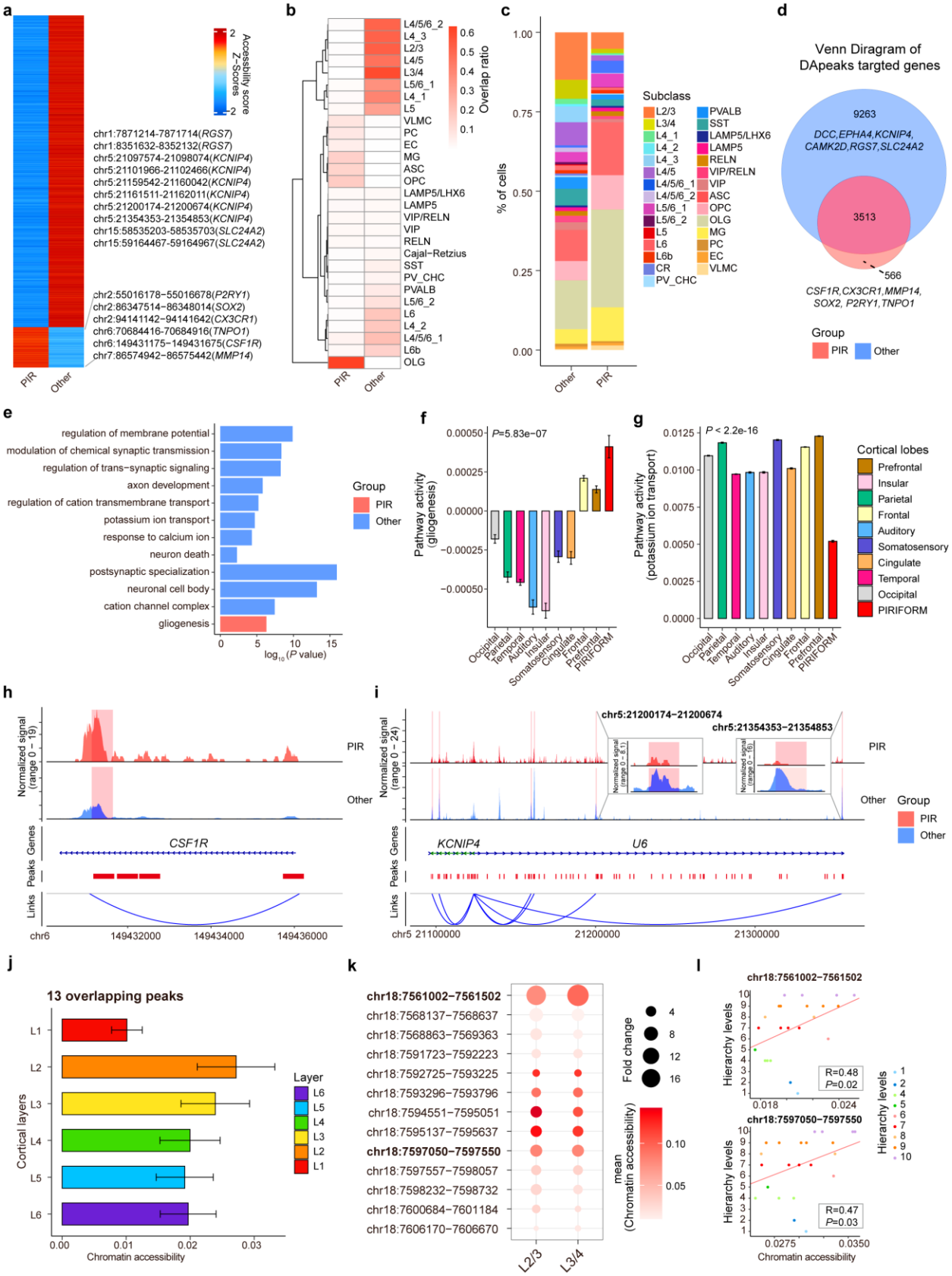
1650

1651 **Extended Data Fig. 2. Statistics of CREs across various cell types in the macaque cortex. a**, UMAP showing the
1652 co-embedding of snATAC-seq and snRNA-seq data from the entire macaque cortex (left), and UMAP layout showing
1653 single nuclei from different cortical lobes (top right) and different monkeys (bottom right). **b**, Correspondence of
1654 snATAC-seq-revealed cell clusters with snRNA-seq-identified cell types. **c**, The number of identified CREs and
1655 differential accessible peaks in various cell subclasses. **d**, Bar plots showing the fraction of CREs that overlaps with
1656 different classes of annotated sequences in glutamatergic, GABAergic and non-neuronal cells in our macaque datasets
1657 and published snATAC-seq datasets of human, macaque and mouse brain[11, 15, 18, 19]. **e**, Comparison of the TSS
1658 enrichment and promoter ratio of glutamatergic, GABAergic and non-neuronal cells in our macaque datasets and
1659 published snATAC-seq datasets of human, macaque and mouse brain[11, 15, 18, 19]. **f**, Dot plotshowing the overlap of
1660 distal CREs linked genes and differentially expressed genes between snATAC-seq and snRNA-seq identified cell types.
1661 *P* values were calculated by hypergeometric test.
1662



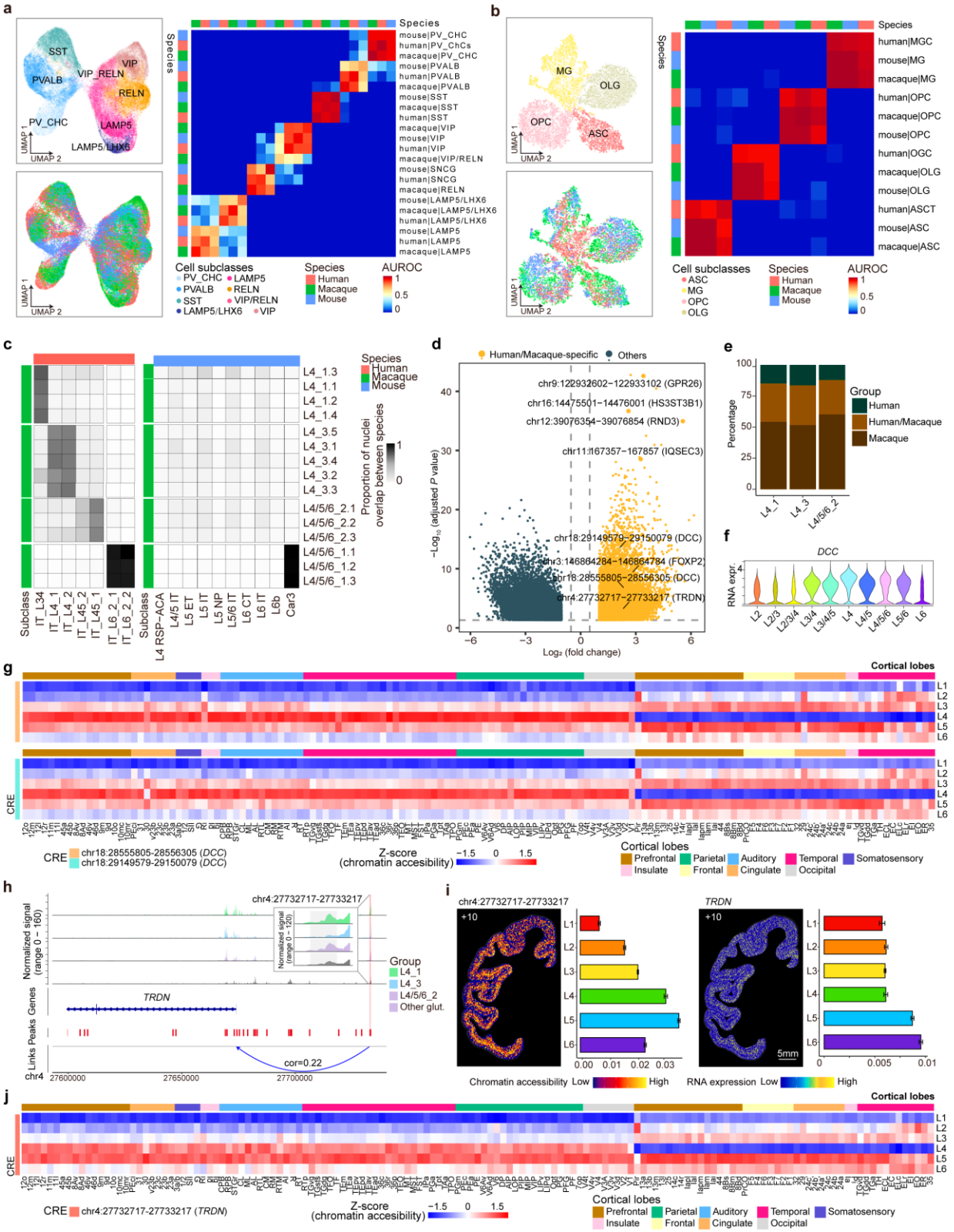
Extended Data Fig. 3. Spatial mapping of chromatin accessibility of CREs in various cell types. a, Correspondence of percentages of various cell clusters revealed by snATAC-seq clustering and spatial mapping result

on the Stereo-seq map. **b**, Heatmap showing the Jaccard similarity of marker genes identified from snATAC-seq data and spatial transcriptome data for the same cell cluster. **c**, Layer- and region-distribution of snATAC-seq-defined cell clusters by spatial mapping. The first to fourth columns from left to right represent the number of cells in each cluster, the regional distribution preference, percentage of cells in nine cortical lobes and the relative density of cells in six cortical layers for each cell type. **d**, Spatial maps showing the chromatin accessibility of four example CREs shown in **Fig. 2e** across cortical layers in representative coronal sections at indicated EBZ coordinates. **e**, Bar plots showing the relative distribution of three L4 clusters across cortical brain regions in the snATAC-seq dataset. **f**, Bar plots showing the relative distribution of three L4 clusters across cortical brain regions in the spatial transcriptome mapping result. **g**, Heatmap showing the correspondence of snATAC-seq-identified cell clusters and transcriptome-defined cell types. **h**, RNAscope experiment results showing the expression pattern of the *MEF2C* gene in V1 and V2 regions of the macaque cortex. Probes against *MEF2C* (ACD, cat#1843711-C2) was list in Supplementary Table 8. DAPI was used for nuclear counterstaining. **i**, Genome browser track view showing a differential accessible distal CRE linked to *LAMA4* across three L4 glutamatergic clusters and other glutamatergic clusters. **j**, Spatial maps of expression patterns of *TFAP4* and *LAMA4*, and chromatin accessibility of the CRE linked to *LAMA4* at two representative sections (EBZ -6.3 and -6.8).


a
Accessibility score
Z-Scores
chr1:7871214-7871714(RGS7)
chr1:8351632-8352132(RGS7)
chr5:21097574-21098074(KCNIP4)
chr5:21101966-21102466(KCNIP4)
chr5:21159542-21160042(KCNIP4)
chr5:21161511-21162011(KCNIP4)
chr5:21200174-21200674(KCNIP4)
chr5:21354353-21354853(KCNIP4)
chr15:58535203-58535703(SLC24A2)
chr15:59164467-59164967(SLC24A2)
chr2:55016178-55016678(P2RY1)
chr2:86347514-86348014(SOX2)
chr2:94141142-94141642(CX3CR1)
chr6:70684416-70684916(TNPO1)
chr6:149431175-149431675(CSF1R)
chr7:86574942-86575442(MMP14)
PIR
Other
b
Overlap ratio
L4/5/6_2
L4_3
L2/3
L4/5
L3/4
L5/6_1
L4_1
L5
VLMC
PC
EC
MG
ASC
OPC
LAMP5/LHX6
LAMP5
VIP/RELN
VIP
RELN
Cajal−Retzius
SST
PV_CHC
PVALB
L5/6_2
L6
L4_2
L4/5/6_1
L6b
OLG
c
% of cells
Subclass
d
Venn Diagram of DApeaks targted genes
9263
DCC,EPHA4,KCNIP4, CAMK2D,RGS7,SLC24A2
3513
566
CSF1R,CX3CR1,MMP14, SOX2, P2RY1,TNPO1
Group
PIR
Other
e
regulation of membrane potential
modulation of chemical synaptic transmission
regulation of trans−synaptic signaling
axon development
regulation of cation transmembrane transport
potassium ion transport
response to calcium ion
neuron death
postsynaptic specialization
neuronal cell body
cation channel complex
gliogenesis
log10 (P value)
f
P=5.83e−07
Pathway activity (gliogenesis)
g
P < 2.2e-16
Pathway activity (potassium ion transport)
Cortical lobes
Prefrontal
Insular
Parietal
Frontal
Auditory
Somatosensory
Cingulate
Temporal
Occipital
PIRIFORM
h
Normalized signal (range 0 − 19)
Genes
Peaks
Links
CSF1R
chr6
149432000
149434000
149436000
i
Normalized signal (range 0 − 24)
chr5:21200174−21200674
chr5:21354353−21354853
KCNIP4
U6
chr5 21100000
21200000
21300000
j
13 overlapping peaks
Cortical layers
Layer
Chromatin accessibility
k
chr18:7561002−7561502
chr18:7568137−7568637
chr18:7568863−7569363
chr18:7591723−7592223
chr18:7592725−7593225
chr18:7593296−7593796
chr18:7594551−7595051
chr18:7595137−7595637
chr18:7597050−7597550
chr18:7597557−7598057
chr18:7598232−7598732
chr18:7600684−7601184
chr18:7606170−7606670
L2/3
L3/4
Fold change
mean (Chromatin accessibility)
l
chr18:7561002−7561502
R=0.48
P=0.02
chr18:7597050−7597550
R=0.47
P=0.03
Hierarchy levels
Chromatin accessibility

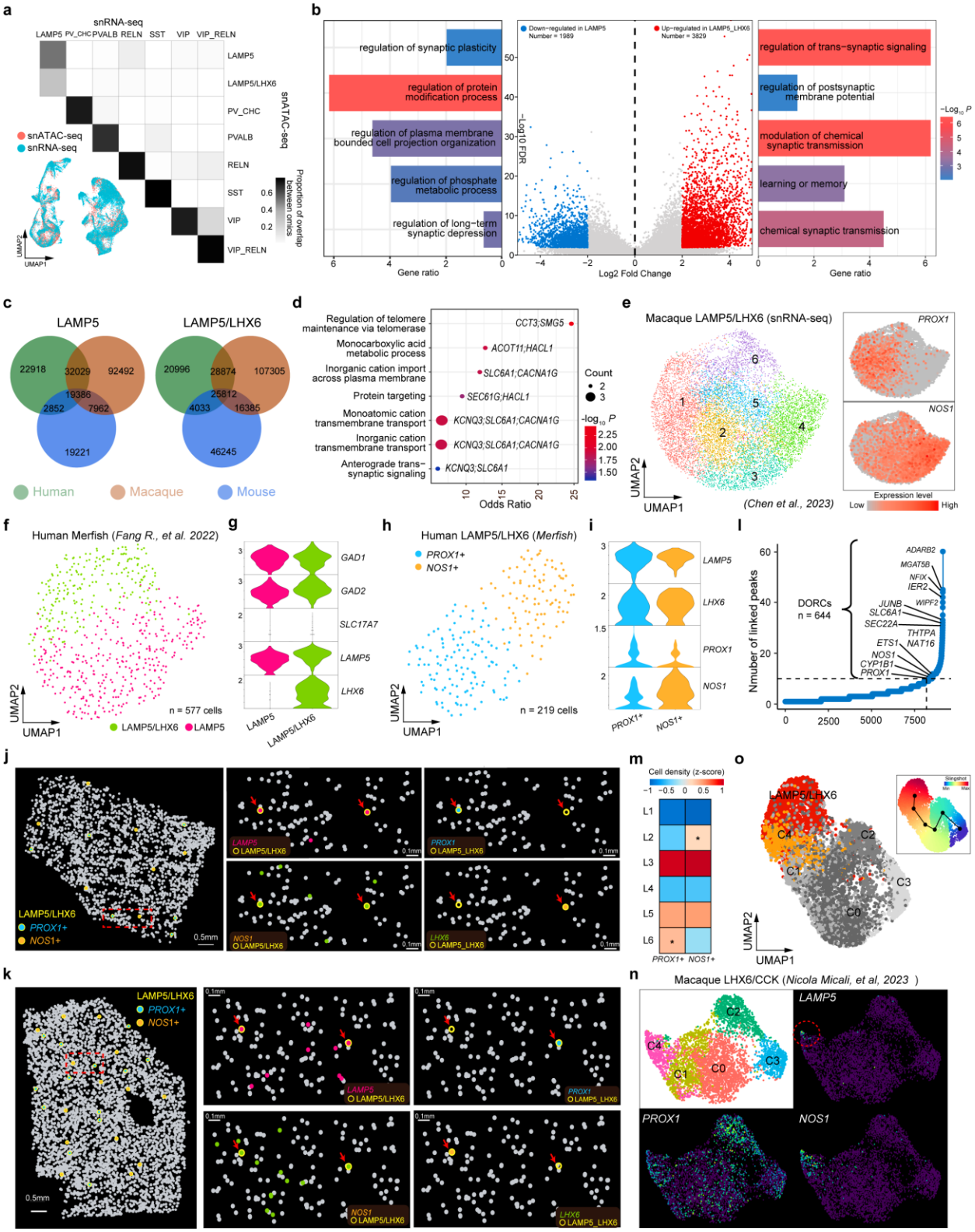
1683

1684 **Extended Data Fig. 4. Characterization of CREs for the piriform cortex and *CBLN2 gene*. a**, Heatmap showing
1685 the chromatin accessibility of differential CREs between piriform cortex and all the other neocortex regions. **b**,
1686 Heatmap showing the overlap between the up-regulated CREs in piriform cortex and the other neocortex regions and
1687 the differentially accessible CREs in various cell types. **c**, Bar plot showing percentages of various cell clusters in
1688 piriform cortex and the other neocortex regions. **d**, Venn plot showing the overlap of genes linked by up-regulated
1689 CREs in piriform cortex and the other neocortex regions. **e**, Pathway enrichment for genes linked to the up-regulated
1690 CREs in piriform cortex and the other neocortex regions shown in **d**. **f-g**, Barplot showing the pathway activity of
1691 "gliogenesis" (**f**) and "potassium ion transport" (**g**) across nine cortical lobes. The pathway activity was calculated by
1692 averaged expression level of enriched genes in the pathway subtracted by the aggregated expression of control feature
1693 sets. **h**, Genome browser tracks showing one up-regulated CRE in piriform cortex linked to the TSS region of *CSF1R*.
1694 **i**, Genome browser tracks showing the down-regulated CREs in piriform cortex linked to the TSS region of *KCNIP4*.
1695 **j**, Bar plots showing the quantitative chromatin accessibility level of the 13 overlapping peaks in our macaque dataset
1696 with the CREs of *CBLN2* in *Shibata M, et al. Nature. 2021* across six cortical layers. Error bars represent standard
1697 error of the mean (SEM). **k**, Dot plot showing the the fold change in the relative accessibility of L2/3, L3/4 subclass versus other
1698 cells. Color indicates the mean accessibility score. Size indicates the fold change. All comparisions are statistical significant
1699 (Wilcoxon rank-sum test). **l**, Dot plots showing the correlation of chromatin accessibility and the hierarchy levels of
1700 various brain regions in the somatosensory system.
1701



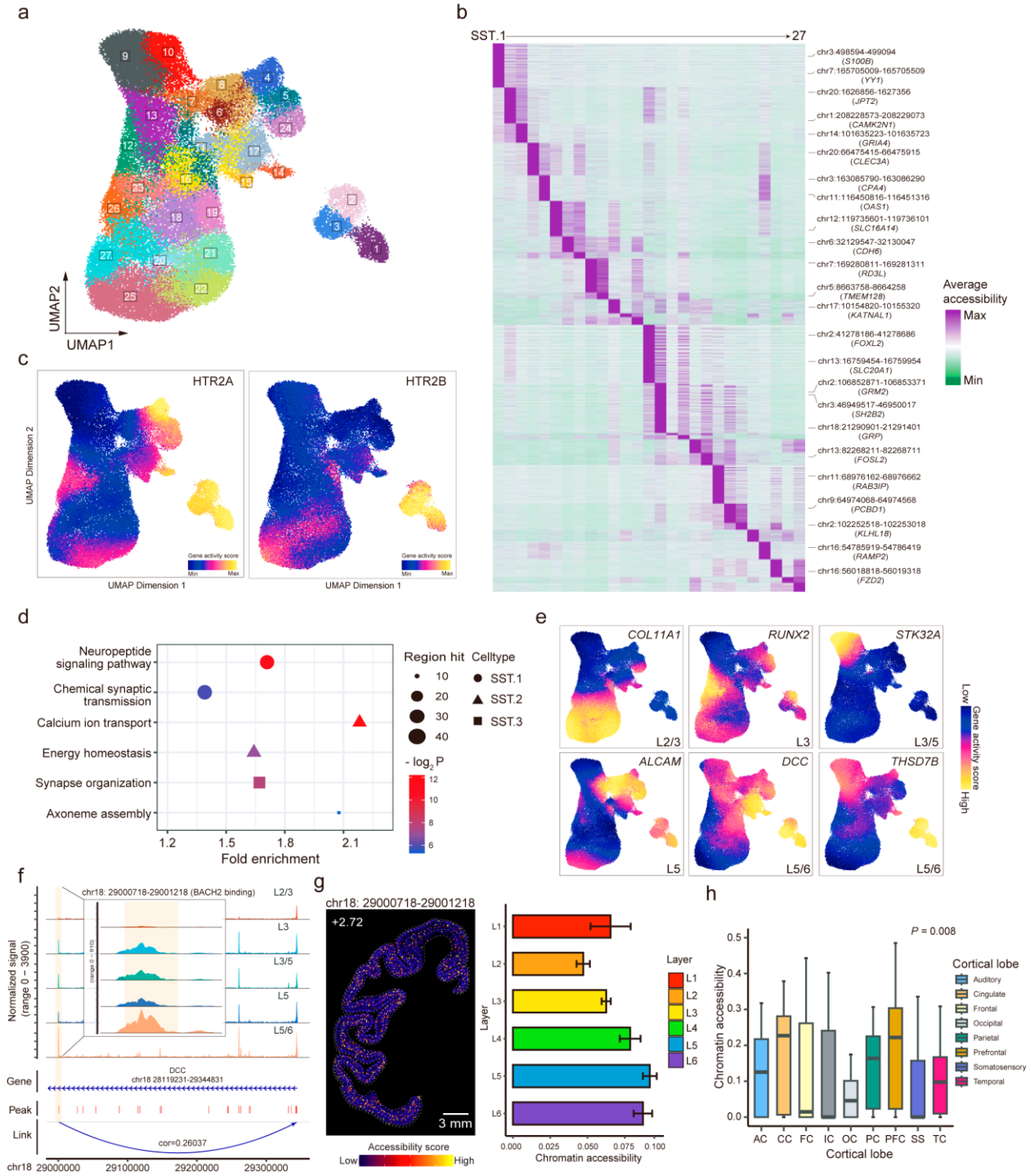
Extended Data Fig. 5. Cross-species comparison of CREs for glutamatergic neurons. a, **Left**, UMAP co-embedding of GABAergic neurons (upper) and non-neuronal cells (below) from snATAC-seq data of macaques,

humans and mice. Single nuclei were colored by their cell subclasses. **Right,** Heatmap showing the similarity of snATAC-seq derived GABAergic cell clusters across macaques, humans and mice using MetaNeighbor. **b**, Similar to (**a**), but for non-neuron cells. **c**, Heatmap depicting the pairwise comparisons of snATAC-seq derived glutamatergic cell types between macaques and humans or mice. **d**, Volcano plot showing the differentially accessible CREs between human/macaque-biased and other snATAC-seq-defined cell types. **e**, Bar plots showing the conservation between humans and macaques for CREs with high chromatin accessibility in human/macaque-biased cell types shown in **d**. **f**, Violin plots showing the expression level of *DCC* among various glutamatergic neuron subclasses of the macaque snRNA-seq data. **g**, Heatmap showing the average chromatin accessibility of CREs associated with *DCC* across cortical layers and regions. **h**, Genome browser tracks showing the chromatin accessibility of human/macaque-biased CRE linked to *TDRN* in the human/macaque-biased and other cell types in the macaque cortex. **i**, Spatial map showing the chromatin accessibility of the human/macaque-biased CRE shown in **g** and the expression level of *TDRN* in a representative Stereo-seq section at indicated coordinate. The right panel showing the quantitative chromatin accessibility or gene expression level across six cortical layers. **j**, Heatmap showing the average chromatin accessibility of CRE linked to *TRDN* across cortical layers and regions.



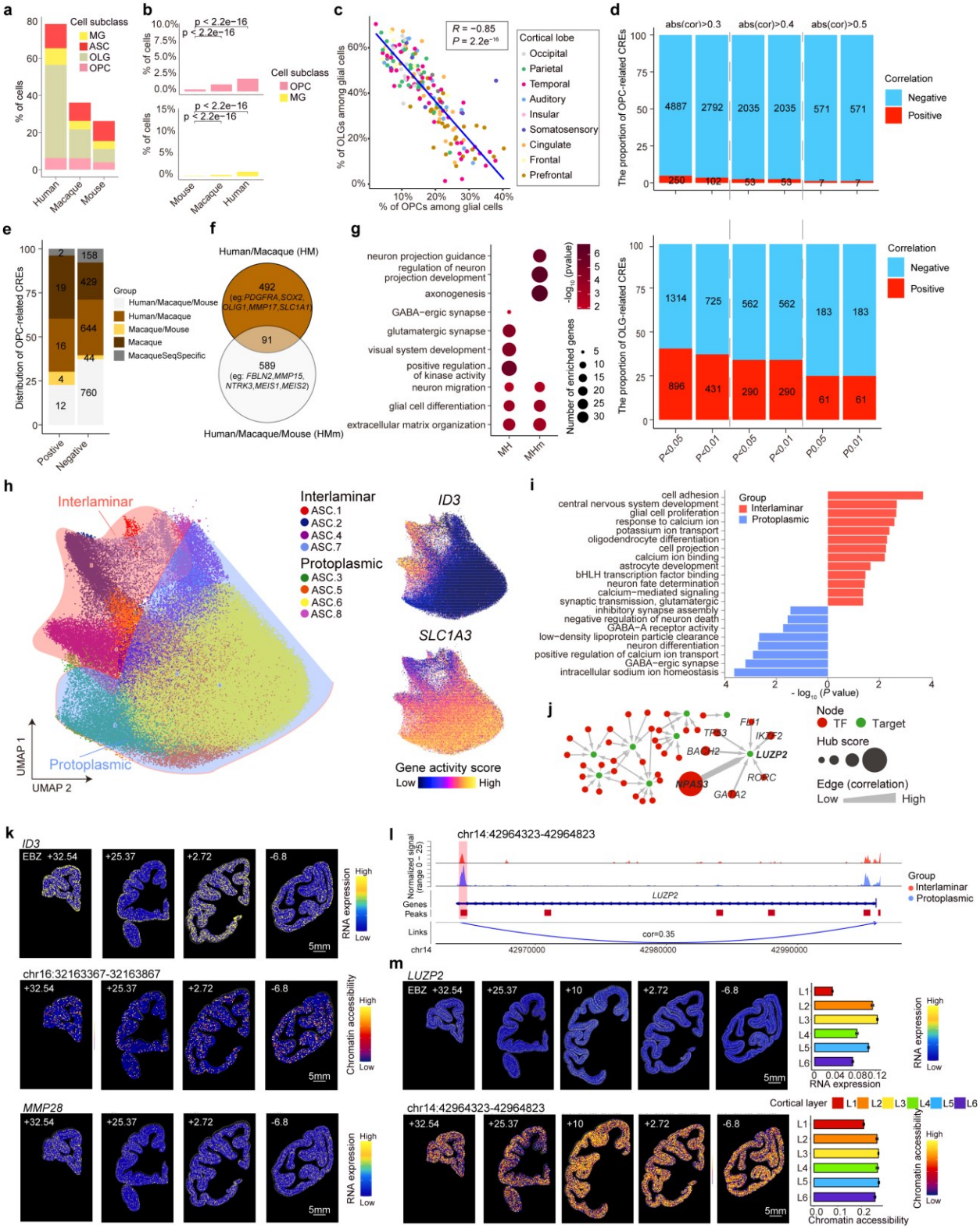
Extended Data Fig. 6. Characterization of CREs in LAMP5/LHX6 GABAergic neuron subgroups. a, UMAP embedding of snATAC-seq and snRNA-seq GABAergic cells from macaque cortex (lower left). Heatmap of overlap scores showing the correspondence between snRNA-seq and snATAC-seq derived macaque GABAergic cell

subclasses (upper right). **b**, Volcano plot showing differential accessible peaks between LAMP5 and LAMP5/LHX6 neurons (middle). GO pathway enrichment analysis of genes linked to distal DA peaks in LAMP5 (left bar plot) and in LAMP5/LHX6 (right bar plot). **c,** Venn plot showing the overlap of CREs among human, macaque and mouse LAMP5 (left), and LAMP5/LHX6 cells (right). **d,** Gene ontology enrichment of the shared CREs between human and macaque LAMP5/LHX6 neurons. **e,** UMAP embedding of snRNA-seq LAMP5/LHX6 neurons (left) and normalized gene expression of PROX1 and NOS1 (right). **f,** UMAP showing the LAMP5 and LAMP5/LHX6 cell clusters derived from MERFISH data of the human dorsolateral prefrontal cortex (dlPFC) (*Fang R., et al. Science. 2022*). **g,** Violin plots depicting the expression levels of marker genes for LAMP5 and LAMP5/LHX6 cell types. **h,** UMAP illustrating the PROX1+ and NOS1+ subpopulations within the LAMP5/LHX6 cell type (left), alongside their corresponding gene expression patterns (right). **i,** Violin plots depicting the expression levels of marker genes for PROX1+ and NOS1+ subclusters. **j-k**, Dot plots showing the spatial pattern of LAMP5/LHX6 cell types (left) and gene expression patterns for indicated genes (right) in the MERFISH tissue sections. The four insets on the right are magnified views of the region outlined by the red dashed box on the left, depicting the spatial expression patterns of the genes *LAMP5*, *LHX6*, *PROX1*, and *NOS1*, respectively. LAMP5/LHX6 cells were marked by yellow circles and red arrows. **l,** The number of significantly correlated peaks for each gene in LAMP5/LHX6 neurons. **m,** Layer distribution of the two distinct groups of LAMP5/LHX6. An asterisk (*) indicates a significant difference between the two groups in the given layer ($P < 0.01$). **n,** Re-clustering of MGE-derived LHX6/CCK neurons from development macaque brain identified five distinct subclusters (C0 to C4). Cluster C4 showed co-expression of *PROX1* and *NOS1* with relatively low levels of *LAMP5*. **o**, Integration of developmental clusters with adult LAMP5/LHX6 populations demonstrated preferential alignment of C4 with adult LAMP5/LHX6 neurons. The upper right panel shows the pseudotime trajectory inferred by Slingshot, ranging from early (blue) to late (red) stages.



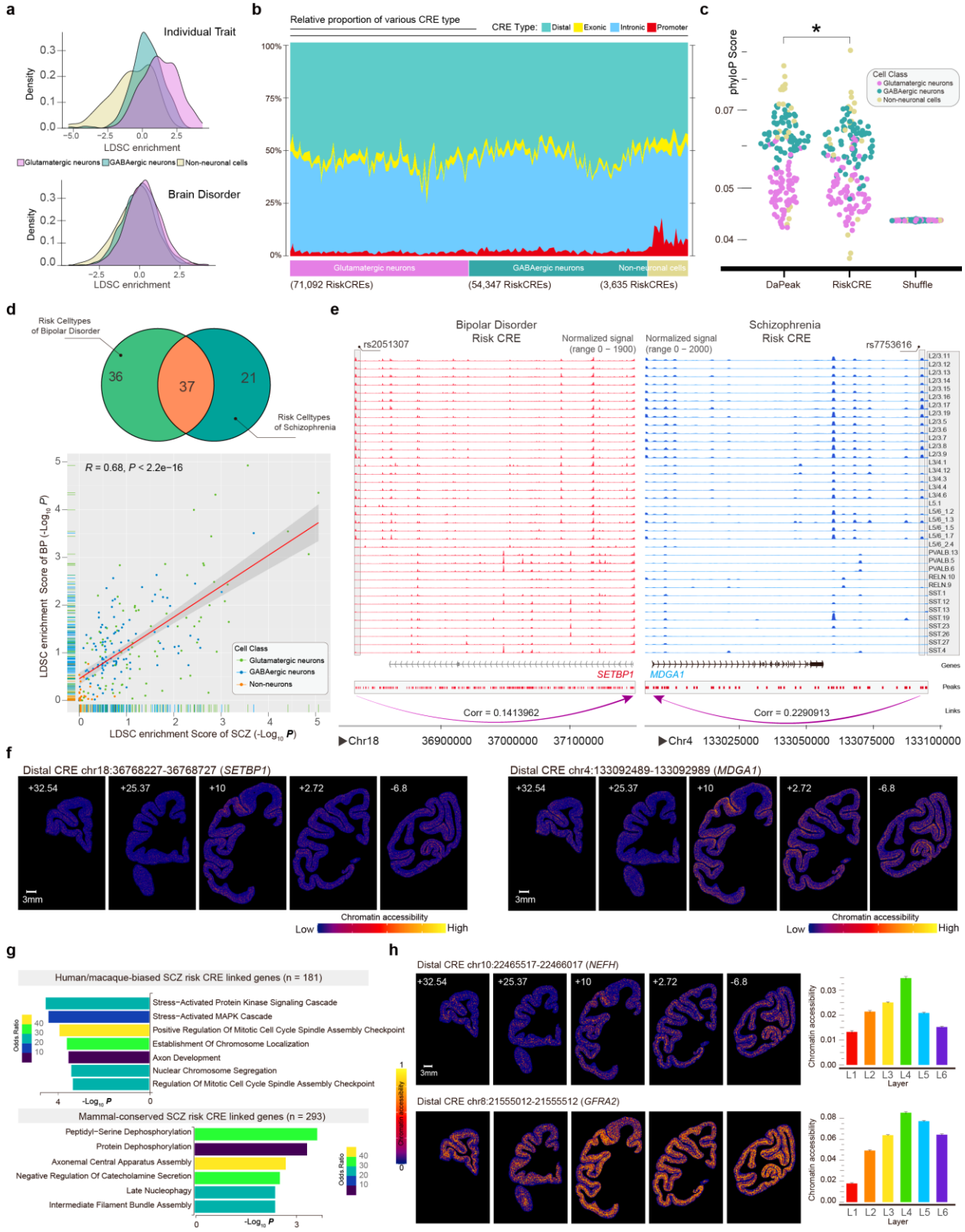
Extended Data Fig. 7. Characterization of CREs in layer-specific SST GABAergic neurons. a, UMAP embedding of snATAC-seq-based SST neurons. **b,** Heatmap of top differentially accessible (DA) peaks across 27 SST subclusters (1,000 cells randomly sampled per cluster) and distal linked genes are annotated. **c,** Normalized gene activity scores of *HTR2A* and *HTR2B* across SST subclusters. **d,** Representative Gene Ontology terms enriched by rGREAT analysis of DA peaks. **e,** Normalized gene activity scores of layer markers for SST neurons (right). **f,** Genome browser track view showing a differential accessible distal CRE linked to *DCC* across various SST cell types. **g,** Spatial visualization

1753 of chromatin accessibility for the CRE linked to *DCC* (shown in **j**) in representative section (left), and bar plots
1754 showing the average chromatin accessibility in six cortical layers for the CRE (right). **h,** Boxplot showing chromatin
1755 accessibility of the CRE associated with *DCC* (shown in **j**) in SST neurons across cortical lobes. *P*-values: Kruskal-
1756 Wallis test.
1757



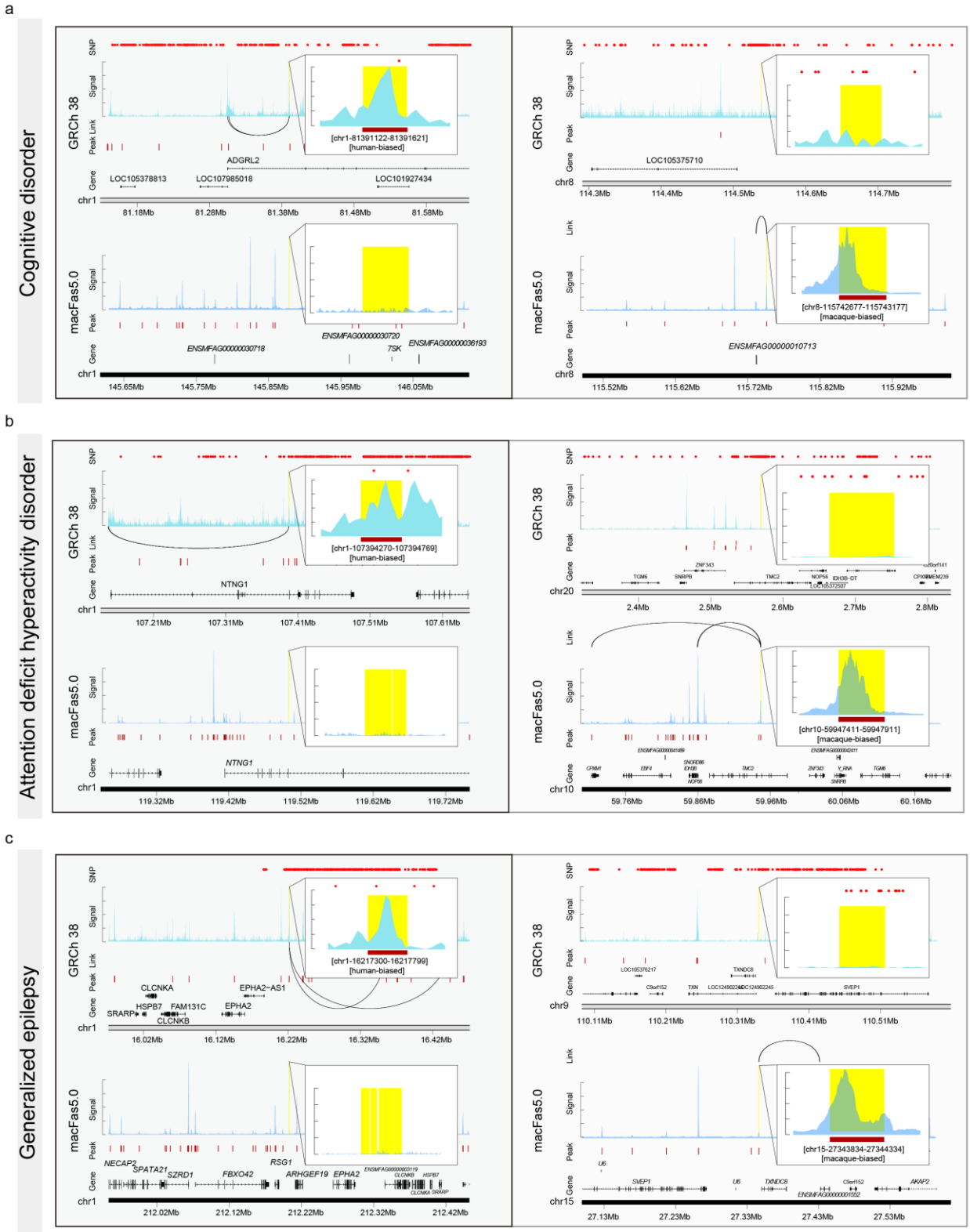
Extended Data Fig. 8. Characterization of CREs for cortical non-neuronal cells. a, Histogram showing the percentages of different glial cell subclasses based on snATAC-seq data across humans, macaques and mice. **b**, The percentages of OPCs and MGs based on snRNA-seq data across humans, macaques and mice. *P* values were calculated

by Z-test. **c**, Dot plot showing the relative percentages of OPCs and OLGs among various macaque cortical regions using the snATAC-seq data. Coefficient and *P* value were calculated by Spearman correlation test. **d**, Bar plots showing the numbers and proportions of CREs in OPCs (upper) and OLGs (lower) with significantly negative or positive correlation (Spearman correlation) with the hierarchy level of cortical regions in the visual system. **e**, Bar plots showing the distribution in species-specific or mammal-conserved groups for CREs in OPCs negatively or positively ($|R| > 0.4$ and $P < 0.01$) correlated with the hierarchy level of cortical regions in the visual system. **f**, Venn diagram showing the relationship between human/macaque-biased and mammal-conserved CREs in OPCs negatively ($R < -0.4$ and $P < 0.01$) correlated with the hierarchy level of cortical regions in the visual system. **g**, Pathway enrichment of genes linked with human/macaque-biased and mammal-conserved CREs in OPCs positively or negatively correlated with the hierarchy levels of brain regions in the visual system. **h**, UMAP embedding of astrocytes (left) and normalized gene activity score of *ID3* and *SLC1A3* (right). **i**, Pathway enrichment of CREs upregulated in the interlaminar and protoplasmic ASCs. **j,** Gene regulatory network of TFs and targeted genes in protoplasmic astrocytes. **k**, Spatial maps of expression patterns of *ID3* and *MMP28*, and chromatin accessibility of the CRE linked to *MMP28* at four representative sections. **l**, Genome browser tracks showing the chromatin accessibility of CRE linked to *LUZP2* between protoplasmic and interlaminar astrocytes. **m**, Spatial maps (left) and quantification (right) of expression pattern of *LUZP2* and chromatin accessibility of the CRE linked to *LUZP2* at five representative sections. Error bars represent SEM.



Extended Data Fig. 9. Cell type enrichment of CREs colocalized with disease-associated risk variants. a, The distribution of LDSC enrichment scores across glutamatergic neurons, GABAergic neurons and non-neuron for

individual trait (upper) and brain disorders (bottom). **b**, The composition of brain disorder-associated CREs. **c**, Comparison of phyloP scores among Risk CREs (DaPeaks with disease risk SNPs), DaPeaks without disease risk SNPs, and a shuffled background. Significant differences were observed between Risk CREs and the DaPeaks (*P < 0.05; Wilcoxon rank-sum test). **d**, Venn plot showing the overlap between SCZ and BP risk variants enriched cells types (upper) and Pearson correlation between LDSC enrichment score of SCZ and BP (bottom). **e**, Genome browser tracks showing chromatin accessibility for gene *NT5DC1* and *SEMA6A*, CREs colocalized with risk variants of bipolar (left) and schizophrenia (right) in neuron cell types were highlighted. **f**, Spatial visualization of cell type-enriched CREs that co-localized with disease-associated SNPs loci in representative sections at different coordinates in EBZ numbers. **g**, Pathways enriched by genes linking to SCZ risk SNP associated CREs in L4_3 glutamatergic neurons. **h**, Spatial visualization of *NEFH*- and *GFRA2*-linked CREs associated with AD SNPs in representative EBZ sections (left). The bar plot illustrates the mean chromatin accessibility of CREs across different layers (right). Error bars represent SEM.



Extended Data Fig. 10. Genome browser tracks showing chromatin accessibility of human- and macaque-biased CREs co-localized with disease risk SNPs in LAMP5/LHX6 neurons. **a-c**, Genome browser tracks showing chromatin accessibility of human- and macaque-specific CREs in LAMP5/LHX6 neurons co-localized with risk SNPs

1800 of cognitive disorder (**a**), attention deficit hyperactivity disorder (**b**) and generalized epilepsy (**c**). Highlighted regions
1801 are indicated by color bands. **Left,** a human-active CRE (top) with a conserved but inactive/weakly active counterpart
1802 in macaque (bottom). **Right,** a macaque-active CRE (bottom) with a conserved but inactive/weakly active counterpart
1803 in human (top). Red dots at the top of each panel represent the genomic loci of risk SNPs.
1804

1805 ## Supplemental tables

1806

1807 **Supplementary Table 1. Quality metrics of snATAC-seq libraries and brain region list.**

1808 **Supplementary Table 2. Correspondence of snATAC-seq-identified cell clusters with snRNA-seq-defined cell**

1809 **types.**

1810 **Supplementary Table 3. The cis-regulatory elements and their linked genes among various cell types.**

1811 **Supplementary Table 4. Cross-species comparison of snATAC-seq data of humans, macaques and mice.**

1812 **Supplementary Table 5. Putative gene regulatory networks across layer-specific SST neurons.**

1813 **Supplementary Table 6. List of CREs in OPCs and OLGs positively and negatively correlated with hierarchy**

1814 **levels of various brain regions in the visual system.**

1815 **Supplementary Table 7. The full list of linkage disequilibrium score regression analysis result.**

1816 **Supplementary Table 8. Probes design for RNAScope Multiplex Assays.**

1817