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

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

[1]Institute of Neuroscience, State Key Laboratory of Neuroscience, CAS Center for Excellence in Brain Science and Intelligence Technology, University of Chinese Academy of Sciences, Chinese Academy of Sciences, Shanghai 200031, China.

[2]BGI-Research, Hangzhou 310012, China.

[3]BGI-Shenzhen, Shenzhen 518103, China.

[4]Lingang Laboratory, Shanghai 200031, China

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

[6]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 and Yidi Sun: ydsun@ion.ac.cn

Abstract

Single-cell spatial transcriptomes have demonstrated molecular and cellular diversity in the brain[1-8], 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 macaque cortical cells, 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-specific CREs for genes related to neurodevelopment and psychiatric diseases. Importantly, risk single-nucleotide polymorphisms for many brain disorders strongly associated with human/macaque-specific 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[1-8]. 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 snATAC-seq method have enabled the analysis of cell type-specific CREs in the adult mouse brain[9-12]. 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 143 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-specific 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 strong association between human/macaque-specific 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 143 cortical regions covering the entire cortex of two adult 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.

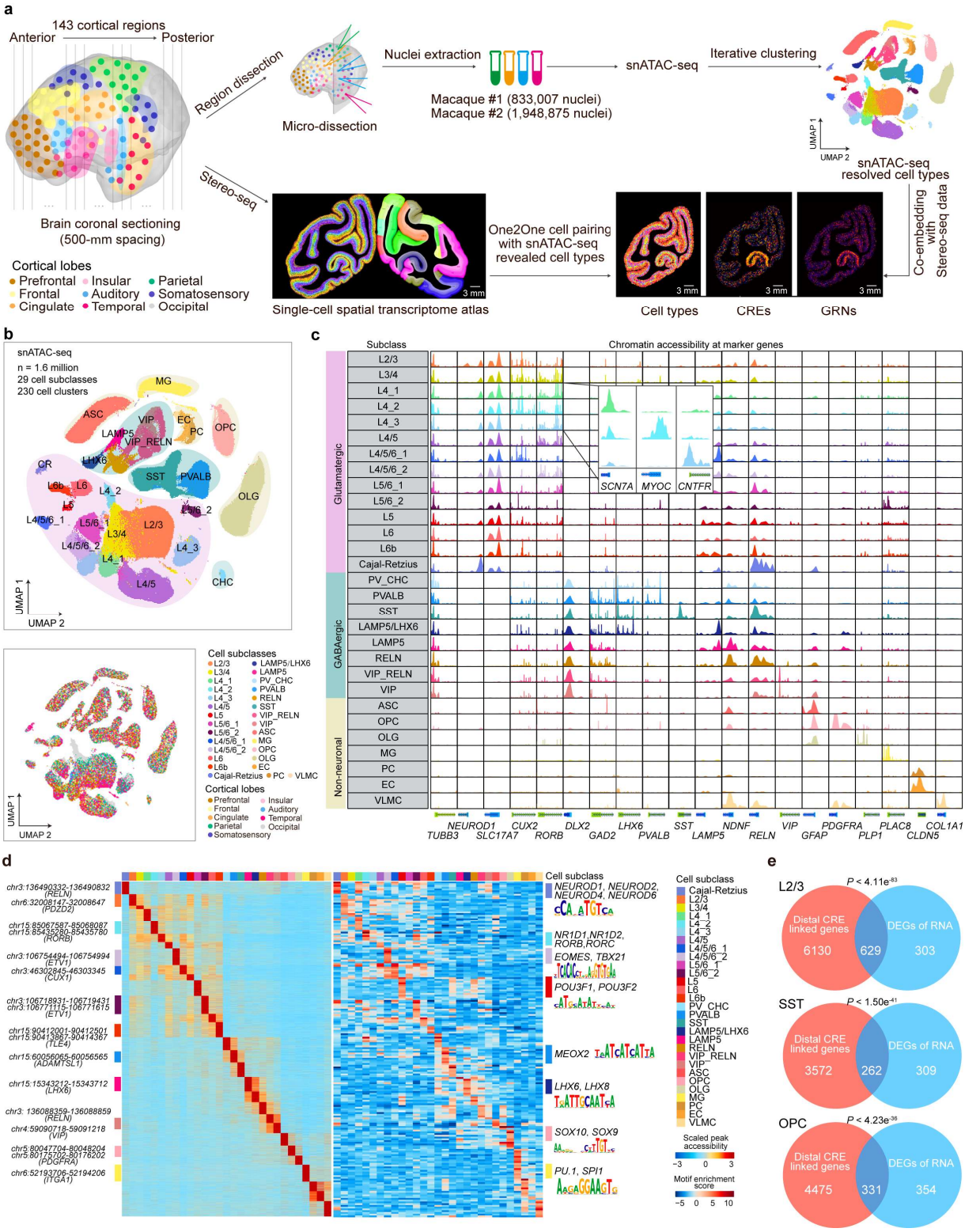


Fig. 1. Chromatin accessibility profiles of various cell types in the macaque cortex. a, Schematics of experimental procedure and integration of snATAC-seq with snRNA-seq and spatial transcriptome data.

Single nuclei from 143 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. We further 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 preferential laminar expression of their targeted genes. For example, the CRE linked to the marker

gene *GPR83* of L2/3 glutamatergic neurons showed open chromatin accessibility peaks only in L2/3 (Fig. 2b). Spatial mapping of this CRE also showed significantly higher chromatin accessibility in L2 and L3, consistent with the spatial expression preferences of *GPR83* (Fig. 2c Extended Data Fig. 3d). In addition, distal CREs linked to *ADCYAP1*, *RORB*, and *TLE4* also showed consistent preferences in spatial chromatin accessibility with the expression pattern of corresponding marker genes for L2/3, L4 and L6 glutamatergic neurons, respectively (Fig. 2b, 2c and Extended Data Fig. 3d).

Cortical region preference of chromatin accessibility of CREs. In addition to layer-specific distribution, we found cortical region-specific distribution of some cell clusters based on their relative density among nine cortical regions (Extended Data Fig. 3c). For example, the L4_2 glutamatergic subclass exhibiting high chromatin accessibility at the promoter of *MYOC* was enriched in the V1 region of the occipital lobe quantified from both snATAC-seq and spatial transcriptome data (Extended Data Fig. 3e, f). This cluster corresponded well to snRNA-seq-defined L4.4 and L4.8 cell types that were also specifically distributed in V1[21] (Extended Data Fig. 3g). The DNA sequences of CREs enriched in L4_2 glutamatergic cluster were then scanned for TF-binding motifs with the CellOracle method[28], generating a gene regulatory network (GRN) for all potential regulatory interactions (Fig. 2d). One marker CRE linked to the *CPA6* gene was significantly enriched with the binding motif for TF MEF2C (Fig. 2e), which is involved in normal neuronal development, cell distribution, and electrical activity of the neocortex as well as neuropsychiatric disorders[29]. Spatial mapping of this CRE also showed that its chromatin accessibility was V1-specific (Fig. 2f). Furthermore, the expression of *MEF2C* and its targeted gene *CPA6* also showed V1 preferences. Similarly, both the chromatin accessibility of CRE linked to *LAMA4* and the expression of *LAMA4* showed V1 specificity (Extended Data Fig. 3h, i). These results support the notion that laminar and regional specific expression patterns of cell types are regulated by specific corresponding CREs, which are controlled by specific TFs.

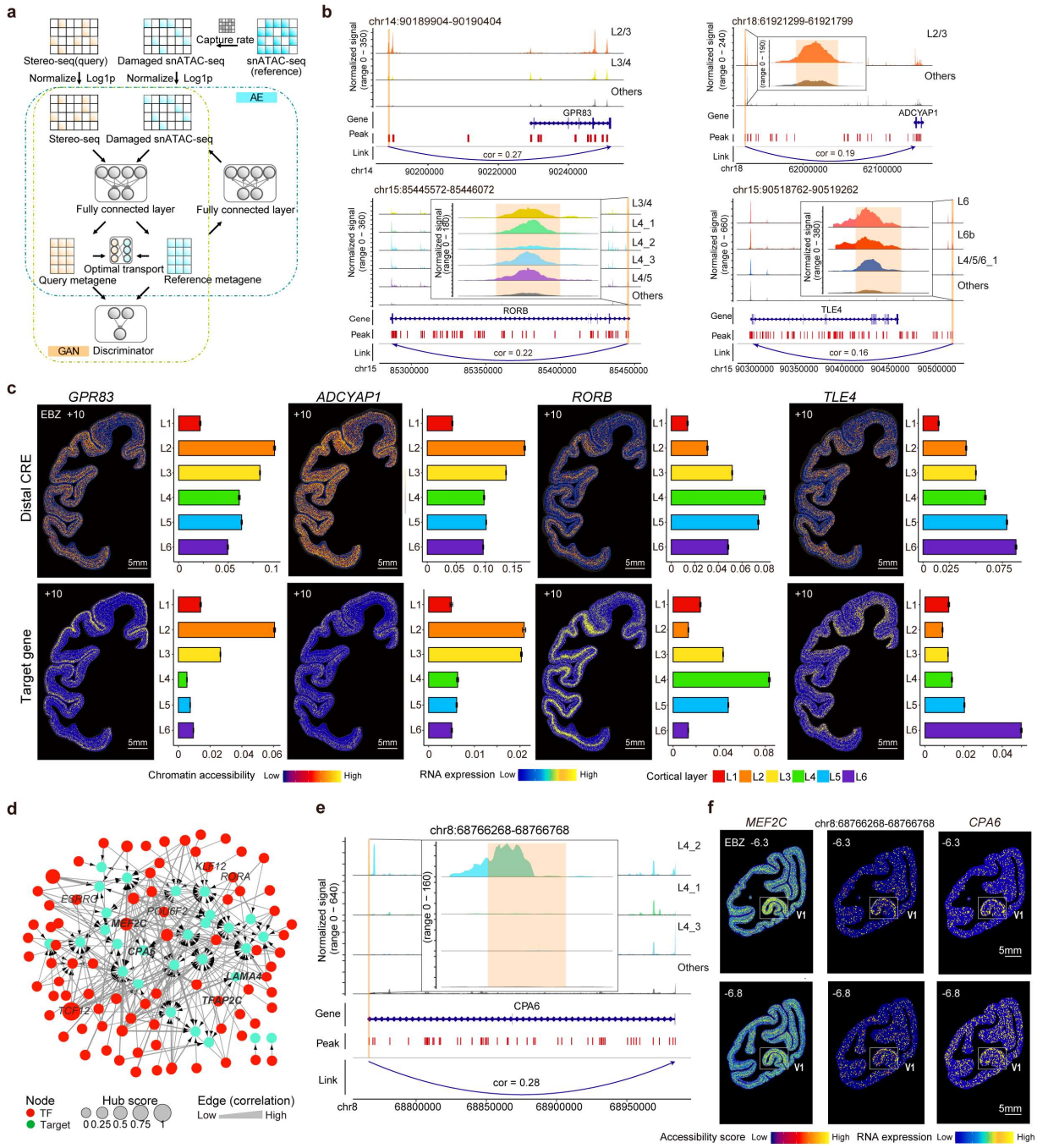


Fig. 2. Characterization of CREs across various brain regions of the macaque cortex. a, Diagram showing the integration and one-to-one pairing of snATAC-seq and spatial transcriptome data. **b**, Genome browser track view showing differential accessible distal CREs linked to *GPR83*, *ADCYAP1*, *RORB* 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).

Human/macaque-specific 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. 4a-c and Supplementary Table 4). Further quantitative analysis among the three species using MetaNeighbor[30] 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. 4d). The human/macaque-specific 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-specific CREs linked to disease associated genes. Further differential CRE analysis between human/macaque-specific cell types and all other glutamatergic cell types identified 17,546 CREs showing higher chromatin accessibility in the human/macaque-specific than other cell types (Extended Data Fig. 4e). 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. 4f and Supplementary Table 4). These human/macaque-specific 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)[31] were much higher in these human/macaque-specific 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 glutmatergic neurons (Extended Data Fig. 4g). 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. 4h), consistent with the expression pattern of *DCC*. Another example is a distal CRE that linked to *TRDN* with higher chromatin accessibility in these human/macaque-specific glutamatergic neurons (Fig. 3g). The decreased expression of *TRDN* was reported to be associated with apoptosis of dopaminergic cells in Parkinson’s disease[32]. Spatial mapping showed the higher chromatin accessibility of this CRE and higher expression level of *TRDN* in deeper layers (Fig. 3h and Extended Data Fig. 4i). These results suggested the unique roles of human/macaque-specific CREs in regulating the spatial expression of disease-associated genes in the primate cortex.

Furthermore, we also examined whether CREs identified in macaque cell types that were conserved across three species were present in human or mouse cortices. For this analysis, we lifted CREs identified in each species to the human genome by sequence homology, and compared the sequence similarity and chromatin accessibility of CREs in each conserved cell subclass. Only CREs showing both sequence similarity and open chromatin between species were considered as conserved (Methods). We found that in comparison with species-conserved CREs, a higher proportion of CREs showed open chromatin in a species-specific pattern (Fig. 3i). Interestingly, we found that ~60% of human/macaque-specific CREs were colocalized with known transposable elements (TEs) (Fig. 3j), consistent with the notion that TEs as a driver of genomic diversity[33]. In comparison with CREs conserved between primates and mice, human/macaque-specific or species-specific CREs showed significantly higher frequency in co-localization with TE elements including LINE, SINE, DNA and long terminal repeats. In addition, human/macaque-specific CREs showed comparable enrichment with LINE and SINE, whereas conserved CREs between primates and mice showed much higher co-localization with SINE element (Fig. 3k). Thus, we have identified human/macaque-specific CREs that may determine layer- and region-specific

glutamatergic cell types and may be responsible for gene regulation underlying brain disorders.

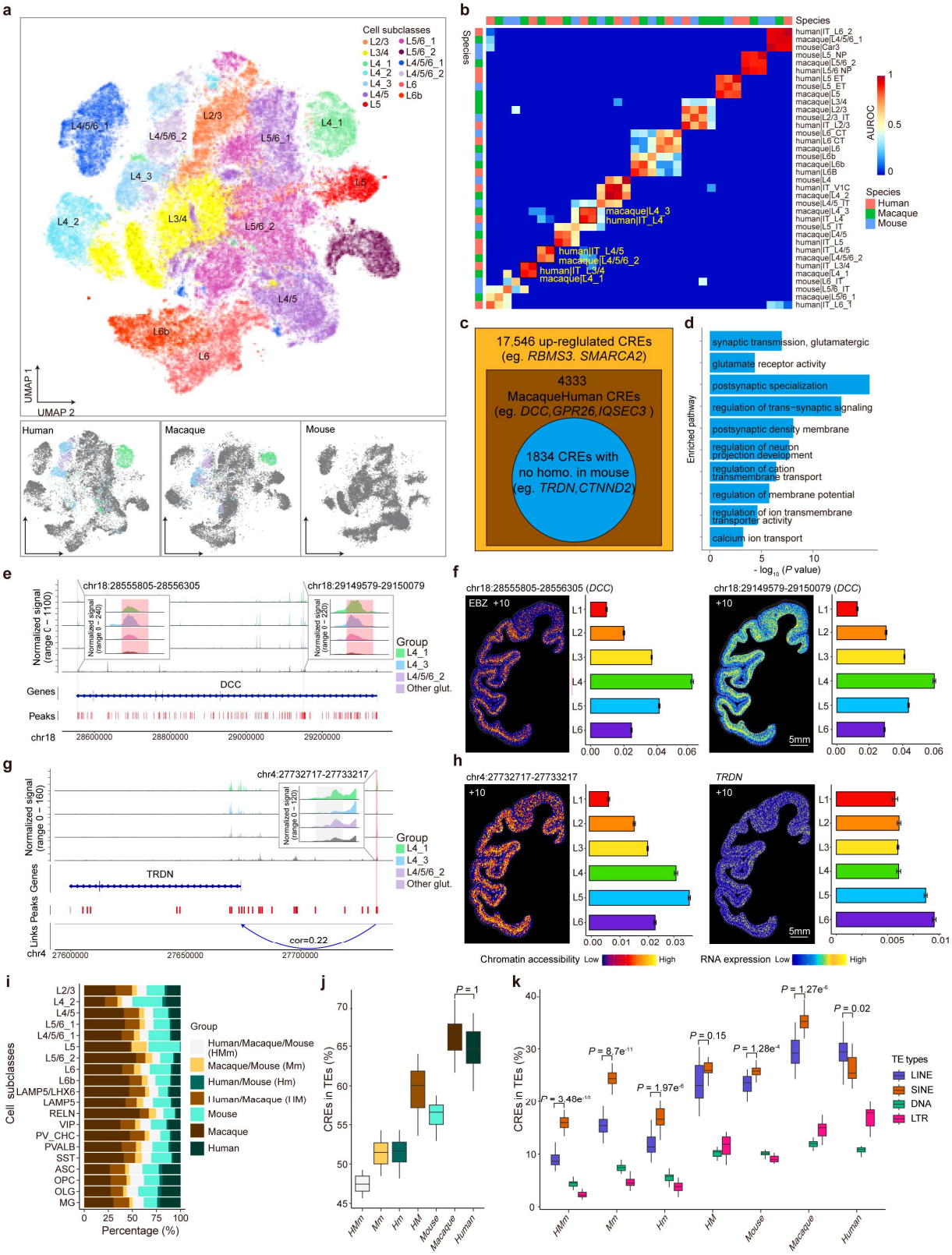


Fig. 3. Comparative analyses revealed human/macaque-specific CREs in glutamatergic neurons. a, UMAP co-embedding of snATAC-seq data from macaques, humans and mice. Colored dots at the bottom indicate human/macaque-specific glutamatergic neurons derived from snATAC-seq data. **b**, Heatmap showing the similarity of snATAC-seq derived glutamatergic cell clusters across macaques, humans and mice using MetaNeighbor. **c**, The outer box represents the number of CREs showing higher chromatin accessibility in human/macaque-specific than other cell types of the macaque cortex. The inner box represents CREs with high chromatin accessibility in the corresponding human cell types. The inner circle indicates CREs showing no homologous sequence in the mouse genome. **d**, GO enrichment for genes linked to human/macaque-specific CREs (inner circle in **c**). **e**, Genome browser tracks showing the chromatin accessibility of human/macaque-specific CREs located on the *DCC* gene in the human/macaque-specific and other cell subclasses. **f**, Spatial map showing the chromatin accessibility of the human/macaque-specific CREs shown in **e** in a representative Stereo-seq section at indicated coordinate. The right panel showing the quantitative chromatin accessibility level across six cortical layers. Error bars represent standard error of the mean (SEM). **g**, Genome browser tracks showing the chromatin accessibility of human/macaque-specific CRE linked to *TDRN* in the human/macaque-specific and other cell types in the macaque cortex. **h**, Spatial map showing the chromatin accessibility of the human/macaque-specific 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. **i**, Bar plots showing 7 categories of CREs in various cell subclasses from macaques, humans and mice. **j**, Boxplots showing the percentage of CREs colocalized with TE in different categories. **k**, Boxplots showing the percentage of CREs located in different types of TEs. *P* values were calculated by Wilcoxon rank-sum test.

Characterization of CREs in GABAergic cell types

Identification of a human/macaque-enriched LAMP5/LHX6 GABAergic cell type. The integrative clustering analysis also showed clear correspondence of snATAC-seq identified clusters and snRNA-seq-defined GABAergic cell types (Extended Data Fig. 5a and 5b). However, based on chromatin accessibility profiles, we found that transcriptome-defined LAMP5+ GABA cell subclass could be further divided into LAMP5 and LAMP5/LHX6 cell types that exhibited low and high chromatin accessibility of the *LHX6* gene, respectively. Comparison of snATAC-seq datasets for macaque, human and mouse cortices showed that the overall percentage of LAMP5/LHX6 cells among LAMP5+ cells in primates was much higher than that in mice (Fig. 4a). We found that the CRE-determined ZIC4, USF1 and MEF2B TF binding motifs were significantly enriched in LAMP5/LHX6 cells in primates (Fig. 4b). Furthermore, the ratio of LAMP5/LHX6 vs. LAMP5 cell types showed substantial variation in different cortical regions,

with the highest abundance in cingulate cortical lobe (Fig. 4c). By examining the shared CREs in LAMP5/LHX6 cells between macaques and humans, we found enrichment of CREs for genes in the pathways of monoatomic and inorganic cation transmembrane transport, such as *KCNQ3*, *SLC6A1* and *CACNA1G* (Extended Data Fig. 5c and 5d).

Further subdivision of LAMP5/LHX6 GABAergic cells with CGE/MGE origins. Further examination of chromatin accessibility in LAMP5/LHX6 cells showed that these cells could be further divided into two populations with differential accessibility of the *PROX1* gene (Fig. 4d), in line with our finding based on a previous transcriptomic dataset on primate cortical cells[21] (Extended Data Fig. 5e). Moreover, we also found differential chromatin accessibility for the marker gene *NOS1* in macaque LAMP5/LHX6 cells (Fig. 4d), suggesting complementary roles of *PROX1* and *NOS1*. Furthermore, we identified the domains of regulatory chromatin (DORC) in each of the LAMP5/LHX6 cell populations and found that DORC of *SEC22A*, *ADARB2* and *PROX1* genes were upregulated in *PROX1*$^+$ cells, and DORC of *ETS1, NOS1* and *EML6* genes were upregulated in *NOS1*$^+$ cells (Fig. 4e and Extended Data Fig. 5f). This result further supports the importance of *PROX1* and *NOS1* in determining the heterogeneity of primate cortical LAMP5/LHX6 cells. Among the CREs associated with the *PROX1* and *NOS1* genes in LAMP5/LHX6 cells, 6 out of 21 CREs linked to *PROX1* were human/macaque-specific (Fig. 4f), whereas 2 out of 5 distal CREs linked to *NOS1* were human/macaque-specific (Fig. 4g). In addition, we also found up-regulated activity scores for genes that marked caudal ganglionic eminence (CGE)-derived and medial ganglionic eminence (MGE)-derived GABAergic neurons in *PROX1*$^+$ and *NOS1*$^+$ (*PROX1*$^-$) cells, respectively (Extended Data Fig. 5g), consistent with the previous report on the transcriptional differences of these two populations[34]. Differential spatial distributions were also observed, with the *PROX1*$^+$ population being enriched in layers 3, 5, and 6, while the *NOS1*$^+$ population was enriched in layers 2, 3, and 5 (Extended Data Fig. 5h).

Chromatin accessibility of CREs may underlie layer- and region-preference of SST neurons. Previous study has revealed the layer-specific distribution of SST neurons that contributed to various anatomical and electrophysiological features of the human cortex[35]. Layer enrichment of various SST cell types was also found in single-cell spatial transcriptome map of the macaque cortex[21]. We then further focused on the layer-specific regulatory programs of various

SST cell types. We classified layer-specific SST neurons based on clustering analysis of their chromatin accessibility profiles and annotated them based on their layer preferences (from L2/3 to L5/6, Fig. 4h), using gene activity scores of previously identified layer marker genes[21] (Extended Data Fig. 5i). 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. 5j, k) and the highest level of this CRE accessibility in the prefrontal and cingulate cortical regions (Extended Data Fig. 5l). 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.

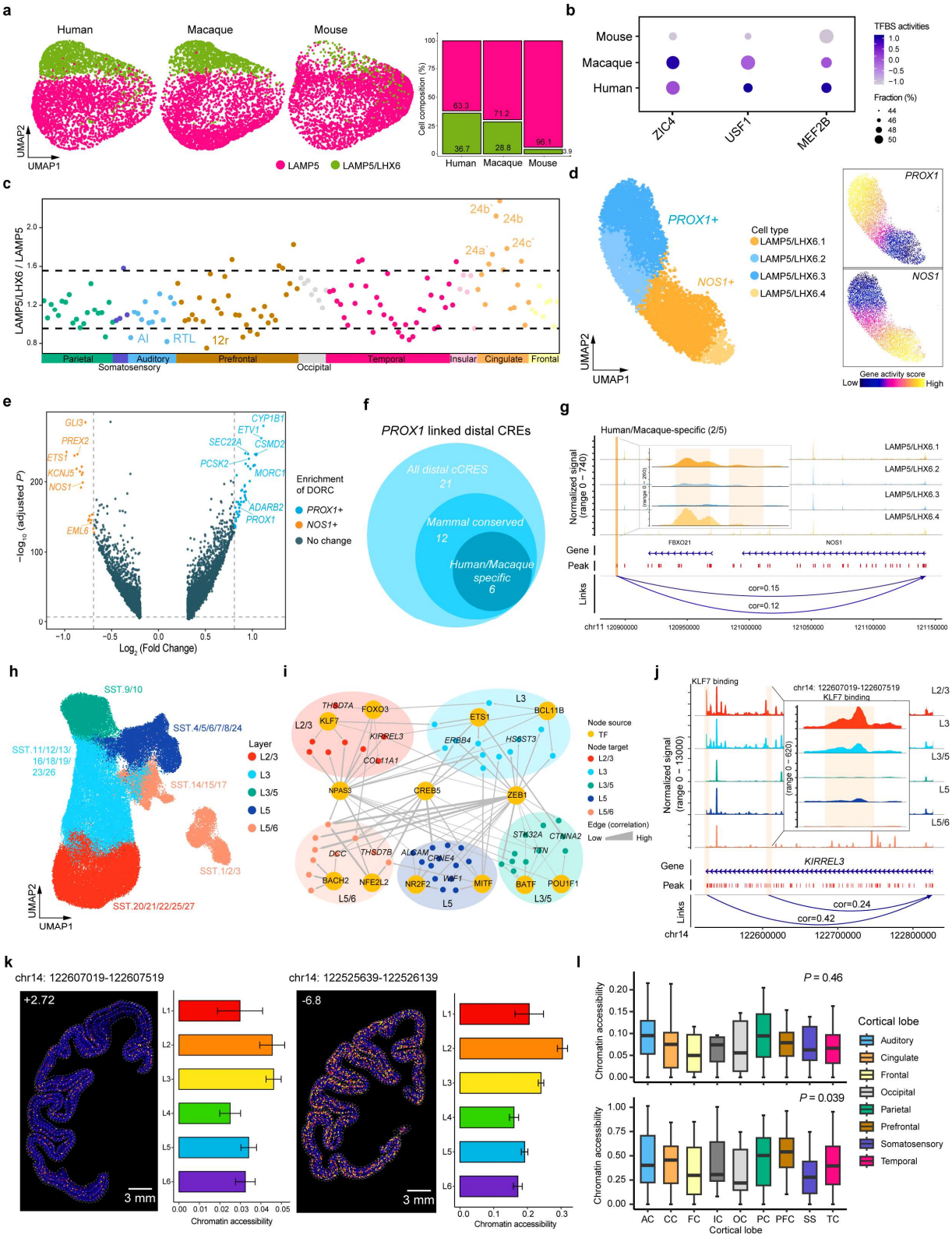


Fig. 4. Characterization of specialized regulatory landscape of GABAergic neurons. a, The UMAP visualization (left) and relative percentage (right) of GABAergic LAMP5 and LAMP5/LHX6 cell subclasses

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. **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. 6a, 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. 6a, b). This finding is in line with the prolonged myelination and altered gene regulation in the oligodendrocyte lineage of the primate brain[2,36-38].

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 of OPCs and OLGs in both snATAC-seq and spatial transcriptome data (Fig. 5a and Extended Data

Fig. 6c). Interestingly, we observed a strong negative correlation (R = -0.9, $P < 2.2\text{x}10^{-16}$) between the percentages of OPCs and OLGs among various cortical regions in the visual system[39] (Fig. 5b). Specifically, higher abundances of OPCs and OLGs were found in cortical regions of high and low hierarchy levels, respectively. The high potential for oligodendrocyte proliferation due to OPCs in regions of higher hierarchy suggests higher plasticity in the regulation of axon myelination in these regions. To further understand the differences in regulatory elements for OPCs and OLGs among different brain regions, we performed Spearman correlation analysis of chromatin accessibility and the hierarchy levels of various brain regions in the visual system. The results showed a consistent higher number of CREs exhibiting chromatin accessibility negatively correlated with the hierarchy level in OPCs regardless of thresholds for coefficients and P values (Extended Data Fig. 6d). Interestingly, more than 30% of these negatively correlated CREs were human/macaque-specific (Extended Data Fig. 6e, f), and their linked genes were significantly involved in GO terms including glutamatergic synapse, visual system development and positive regulation of kinase activity (Extended Data Fig. 6g and Supplementary Table 6). For examples, one mammal-conserved CRE showing open chromatin peaks only in OPCs was linked to *PDGFRA*, a typical marker gene for OPCs (Fig. 5c). Another example is a human/macaque-specific CRE with open chromatin only in OPCs that linked to *FBLN2*, which is an extracellular matrix inhibitor of OPC differentiation[40]. These results suggest the reciprocal regulation between OPCs and OLGs and the importance of OPCs in regulating higher cortical functions.

CREs-determined GRNs between interlaminar and protoplasmic astrocytes. Previous spatial transcriptome studies have shown laminar preferences of astrocytes (ASCs)[21,41]. Our snATAC-seq-based cell clustering analysis further revealed differential chromatin accessibility of CREs associated with interlaminar ASCs (enriched in L1) and protoplasmic ASCs (enriched in L2-L6) (Extended Data Fig. 6h). For example, higher chromatin accessibility was found for CREs linked to *ID3*, *FABP7*, *VCAN* and *MMP28* and those linked to *LUZP2*, *GRID1*, *KCNH5* and *KCNMA1* in interlaminar and protoplasmic ASCs, respectively (Fig. 5e). Based on GO analysis, CREs found in the interlaminar ASCs were enriched in genes associated with "cell adhesion" and "cell projection" pathways, whereas those in the protoplasmic ASCs were associated with "intracellular sodium ion homeostasis", "GABAergic synapse", and "positive regulation of calcium ion transport" (Extended Data Fig. 6i). Moreover, we constructed GRNs for the two

groups of ASCs and found that interlaminar ASCs showed high enrichment of hub TF genes *ID3* and *PBX3* (Fig. 5f). The targeted genes of *ID3* (including *GATA2*, *MSX1*, *TFAP2C* and *PBX3*) were also involved in the central nervous system development (Fig. 5g). On the other hand, the protoplasmic ASCs was enriched in the hub TF gene *NPAS3* (Fig. 5f), which contributes to neurodevelopmental transcription factor networks and the regulation of brain glucose metabolism[42]. As examples, the CRE linked to *MMP28* showed high chromatin accessibility in interlaminar but not protoplasmic ASCs (Fig. 5h). Spatial mapping of this CRE also showed higher chromatin accessibility in L1, in line with the high expression of its linked gene *MMP28* and its binding TF *ID3* in L1 (Fig. 5i, 5j and Extended Data Fig. 6j). In contrast, a CRE linked to *LUZP2* showed higher chromatin accessibility in protoplasmic ASCs, and spatial mapping of this CRE also showed higher chromatin accessibility across cortical layers, consistent with spatial expression pattern of *LUZP2* (Extended Data Fig. 6k, l). Taken together, these results provided new information on regulatory programs underlying the diversity and spatial organization of non-neuronal cells in the primate brain.

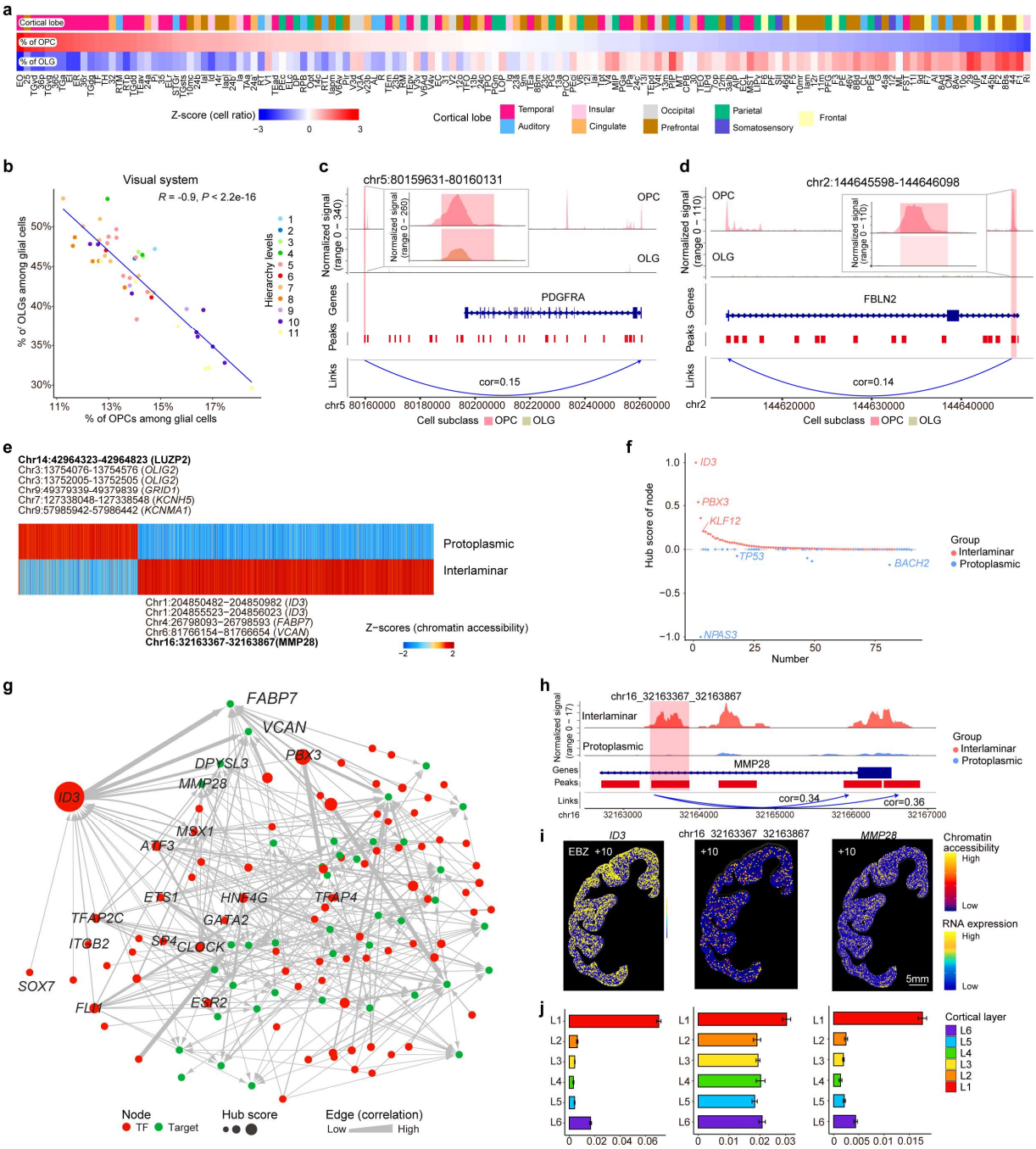


Fig. 5. Regional specificity of CREs in cortical non-neuronal cells. a, Heatmap showing the percentages of OPCs and OLGs among all glial cells in various brain regions of the macaque cortex after spatial mapping. The brain regions are ordered by the decreasing percentage of OPCs. **b**, Correlation between 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-specific CRE linked to the TSS

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

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

LDSC analysis of cell type-specific CREs and brain diseases. Most of the disease risk single nucleotide polymorphisms (SNPs) identified by genome-wide association studies (GWAS) arelocated in non-coding regions of the genome[43]. Previous studies in normal and disease individuals have shown enrichment of disease risk SNPs in cell type-specific CREs[44]. In this study, we further examined whether these CREs were primate-specific. We performed linkage disequilibrium score regression (LDSC) analysis of our identified cell type-specific CREs with available GWAS data for 28 types of brain disorders and human traits (Supplementary Table 7). Interestingly, by hierarchical clustering of similarity in human diseases and trait-associated SNPs, we found significant enrichments of risk SNPs for brain diseases in CREs in many neuronal cell types, especially glutamatergic cell types (Fig. 6a, 6b and Extended Data Fig. 7a; Full list of cell types in Supplementary Table 7). For examples, schizophrenia (SCZ) and bipolar (BP) risk SNPs were found in CREs of glutamatergic cells in superficial layers. Differential enrichment of risk SNP loci was also found in GABAergic neurons and non-neuronal cells. For examples, bipolar risk SNPs in PVALB and SST neurons, generalized epilepsy (GE) risk SNPs in RELN neurons, and AD risk SNPs in microglia (Fig. 6a). On the other hand, we found that SNPs related to human traits, such as cognitive function (CGF), intelligence quotient (IQ), reaction time (RT) and years of education (YED) were generally present at high levels in L4 glutamatergic neurons (Fig. 6a).

CREs associated with SCZ and BP risk SNPs. Further examination showed that around 40% and 50% of CREs enriched with disease-related SNPs in various cell types were located at intronic regions and sites distant (±1 kb) from the TSS, respectively (Extended Data Fig. 7b). To

further characterize these CREs, we observed that the CREs associated with disease risk SNPs exhibited significantly lower phyloP scores[45,46] compared to the background differentially accessible peaks (Extended Data Fig. 7c), highlighting the potential role of rapidly evolving regulatory elements as key contributors to disease mechanisms. Pearson correlation analysis of the enrichment scores of disease risk SNP for various cell type-specific CREs between two diseases showed that the strongest correlation occurred between SCZ and BP among all disease risk SNPs (Fig. 6c and Extended Data Fig. 7d), consistent with the previous finding of substantial overlap in genetic risk SNPs for these two diseases[47]. Among SCZ and BP shared risk SNPs, we found two SNP loci (rs8118905 and rs1040711) colocalized with one CRE that was linked to the promoter of *ADNP* gene (Fig. 6d), which is known to be associated with SCZ and BP[48,49]. The chromatin accessibility of this CRE was the highest in L2/3 glutamatergic neurons and spatial mapping of this CRE also showed the highest level of accessibility in L2 and L3 of the cortex (Fig. 6d and 6e). In addition, we found two other SNPs rs2051307 and rs7753616 were linked to *SETBP1* gene in BP and *MDGA1* gene in SCZ, respectively (Extended Data Fig. 7e). Both genes were related to neurodevelopment[50,51], and the spatial distribution of SNP-associated CREs for the two genes were also enriched in the L2/3 glutamatergic neurons (Extended Data Fig. 7f).

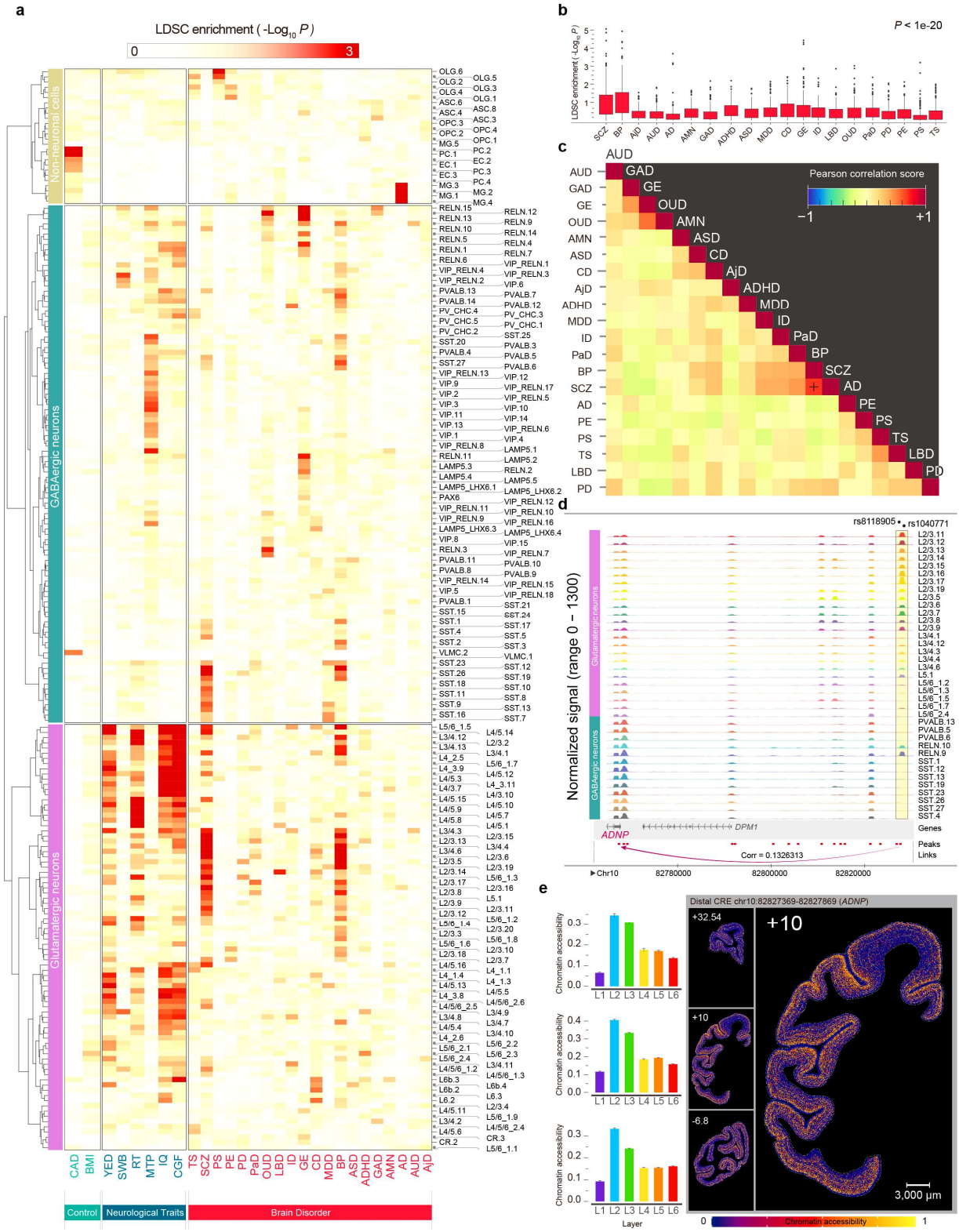


Fig. 6. Association of cell type-specific CREs with risk SNPs of various brain disorders and human traits. a, Heatmap showing the enrichment of cell type-specific CREs containing risk SNPs of various brain

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

Human/macaque-specific CREs harboring disease risk SNPs. We next examined the number of CREs associated with various disease risk SNPs that were human/macaque-specific and conserved across mammals for each cell type and found that many disease risk SNPs were co-localized with both types of CREs. The glutamatergic neurons exhibit significantly higher levels of human/macaque-specific CREs compared to GABAergic neurons and non-neuronal cells (Fig. 7a). In particular, we found that L4_3 glutamatergic neuron subclass, which was found to be human/macaque-specific (Fig. 3a-c), exhibited a striking enrichment of human/macaque-specific CREs that co-localized with various disease risk SNPs (Fig. 7a). For CREs associated with SCZ risk SNPs in these glutamatergic neurons, the chromatin accessibility of human/macaque-specific CREs was significantly higher than that of mammal-conserved CREs (Fig. 7b). Among genes linked to CREs that associated with SCZ risk SNPs in these cells, 181 were associated with primate-specific CREs, including *ABCA7*[52], *ADORA1*[53], and *CHGA*[54]. On the other hand, 293 genes were linked to mammal-conserved CREs, such as *CHRM4*[55], *CXXC1*[56], and *OPCML*[57], with 12 genes overlapping between these two groups, including *PDYN*[58] (Fig. 7c). Gene ontology

analysis revealed that pathways associated with human/macaque-specific CREs include the "stress-activated protein kinase signaling cascade", "regulation of mitotic cell cycle spindle assembly checkpoint", and "axon development", which were not observed in the pathways associated with mammal-conserved CREs (Extended Data Fig. 7g). Comparing the accessibility of SCZ risk SNP-associated CREs in these neurons across brain regions, we found that in L4_3.11 glutamatergic cell type, both human/macaque-specific and mammal-conserved CREs exhibited the highest chromatin accessibility in the occipital lobe. In contrast, chromatin accessibility of these CREs was lowest in L4_3.9 cell type in the occipital lobe (Fig. 7d). Thus, SCZ-associated regulatory mechanisms appeared to be highly cell type- and brain region-specific. Further examination of GRN of L4_3 neurons showed that transcription factors CTCF[59], NEUROD6[60], SREBF1[61], HIF1A[62] and CXXC1[56] were central regulators of genes associated with human/macaque-specific and mammal-conserved CREs (Fig. 7e). Furthermore, the human/macaque-specific CRE linked to *GFRA2*[63] and the mammal-conserved CRE linked to *NEFH*[64], containing binding sites for SREBF1 and HIF1A, respectively, both exhibited the highest chromatin accessibility in L4 of the spatial maps (Fig. 7f and Extended Data Fig. 7h). Thus, we have identified a set of human/macaque-specific CREs harboring SCZ risk SNPs and their gene regulatory networks in human/macaque-specific glutamatergic cell types. This information may help to elucidate the pathogenic mechanisms underlying SCZ.

Human/macaque-specific and mammal-conserved CREs harboring AD risk SNPs. Most notably, we found that CREs associated with AD risk SNPs exclusively resided in microglia cell types (Fig. 6a), and the ratio of CREs harboring AD risk SNPs among human/macaque-specific CREs vs. mammal conserved ones was the highest in AD among all examined brain diseases (Fig. 7g). Further comparison of genes that linked to human/macaque-specific (127) vs. mammal-conserved (207) CREs associated with AD risk SNPs in microglia showed only 17 overlapped genes (out of 317), indicating largely different sets of genes regulated by two groups of CREs (Fig. 7h). Further analysis revealed that prominent GO pathways linked to human/macaque-specific CREs including “regulation of small molecule metabolic process”, “positive regulation of amyloid-beta clearance”, “cellular response to low-density lipoprotein”, which were not found for those linked to mammal-conserved CREs (Fig. 7i). Among the human/macaque-specific CREs linked to AD risk SNPs, we identified genes known to play roles in AD pathogenesis, including

C3[65], *TREM2*[66], *NFYA*[67], *CDNF*[68] and *IDE*[69]. For instance, one human/macaque-specific CRE in microglia was linked to *TREM2* and *NFYA*. On the other hand, both mammal-conserved and human/macaque-specific CREs were linked to AD-related genes[70-72] *APOE, TOMM40* and *TRAPPC6A* in microglia (Fig. 7j). Furthermore, among the AD risk SNP-associated CREs, human/macaque-specific CREs exhibited higher chromatin accessibility in the occipital and somatosensory cortices, whereas mammal-conserved CREs showed higher accessibility in the insular and auditory cortices (Fig. 7k). Thus, AD risk SNPs may function in microglia via human/macaque-specific CREs regulating distinct sets of genes that differ from those regulated by mammal-conserved CREs, pointing to importance of primate-specific microglia targeting for AD therapy.

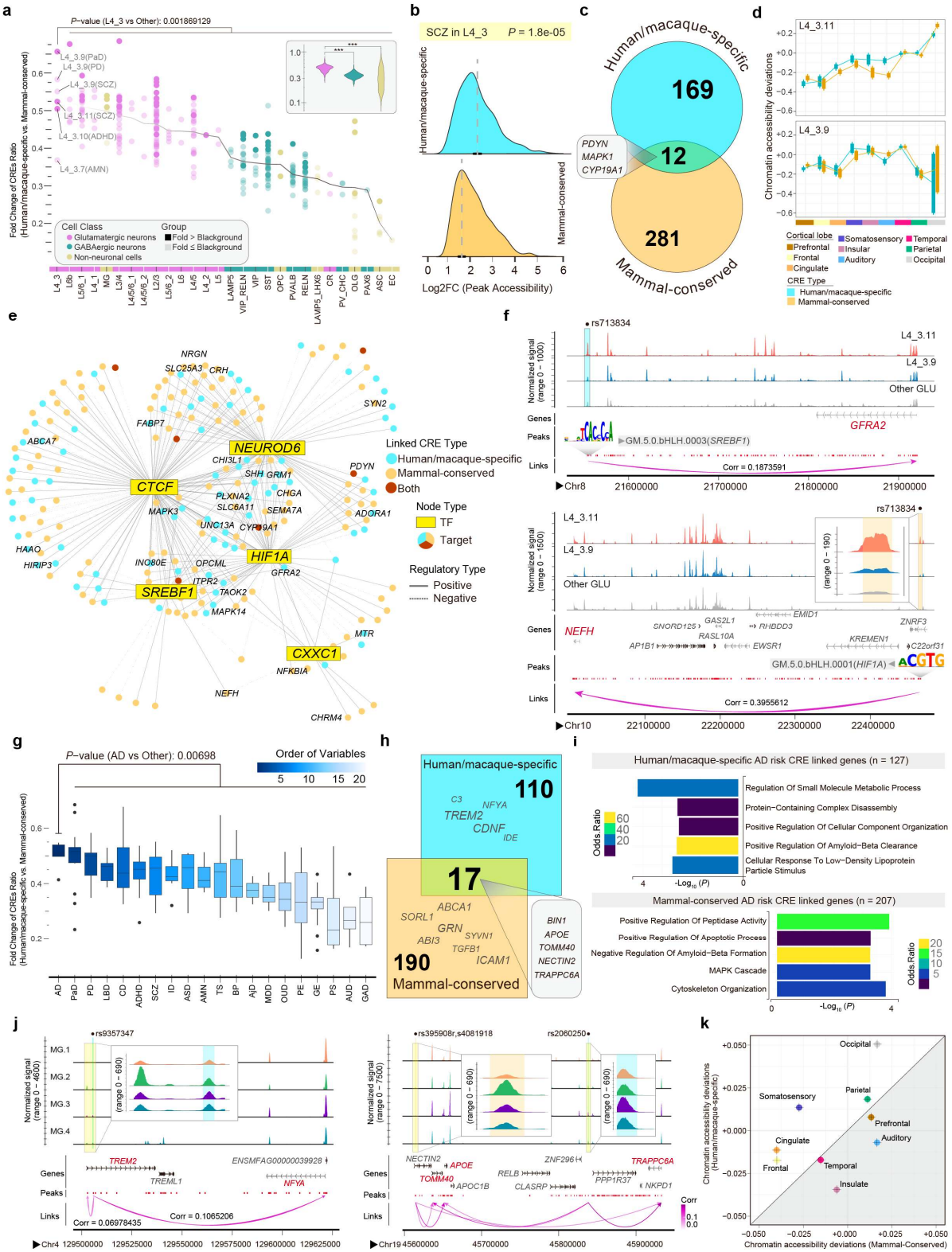


Fig. 7. Prominent association of primate-specific CREs with risk SNPs of various brain disorders. a, Dot plot showing the ranked distribution of primate-specific vs. mammal-conserved CRE ratios (with

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). **b**, Ridge plot depicting the chromatin accessibility of primate-specific 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 primate-specific 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 primate-specific 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 primate-specific 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, primate-specific and mammal-conserved). Highlighted regions are marked with color bands. **k**, Dot plot illustrating the relative chromVAR deviations of primate-specific 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 143 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-specific CREs in various brain diseases. Notably, human/macaque-specific CREs in L4 glutamatergic neurons are much more prominently associated with risk SNPs for schizophrenia. Furthermore, we also identified human/macaque-specific CREs that regulate AD pathogenesis-related genes (*C3*[73], *TREM2*[66], *NFYA*[74], *CDNF*[75] and *IDE*[76]) 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-specific CREs in the emergence of human/macaque-specific cell types as well as their potential involvement in psychiatric diseases. For example, in human/macaque-specific L4 glutamatergic neurons, we identified CREs localized in the regulatory region of *DCC*, a gene involved in neurodevelopment and psychiatric disorders[77-79]. We also found that human trait-related SNPs largely resided in CREs of human/macaque-specific 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-specific 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.

A previous study has shown that *LAMP5* and *LHX6* are expressed in GABAergic neurons derived from caudal ganglionic eminence (CGE) and medial ganglionic eminence (MGE), respectively[80]. However, we found that primate LAMP5-expressing cells included cells originated from both CGE and MGE, depending on whether they express *LHX6*. Furthermore, we found that the LAMP5/LHX6 subclass could be further divided into $PROX1^{+}$ and $NOS1^{+}$ cell types with complementary and opposite chromatin accessibility of *PROX1* and *NOS1* genes, which are two marker genes for cells derived from CGE and MGE, respectively. Previous studies have shown differential expression of the MGE marker gene *NOS1*[81] in LAMP5/LHX6-expressing human cortical cells[4], as well as differential expression of CGE marker gene *PROX1* in primate LAMP5/LHX6 cells[34]. Here, we have identified CREs specifically associated with the regulation of *NOS1* and *PROX1* genes, and these CREs are likely to play important roles in the differentiation of two subtypes of LAMP5/LHX6 GABAergic neurons. Thus, these CREs represent the best

candidates for designing specific probes for studying the regulation and function of GABAergic neurons of different developmental origins.

Our comprehensive analysis of chromatin accessibility and spatial transcriptome allowed us to identify differences in chromatin accessibility across various cortical regions. This information is critical for understanding the gene regulation underlying regional differentiation of cell types. Our finding of the reciprocal relationship between the abundance of OPCs and OLGs across all brain regions suggest that OLGs may provide negative regulation of OPC proliferation. The opposite trend in the abundance of OPCs and OLGs at different hierarchy levels of the visual system[39] further suggests the functional role of axon myelination in neural properties at various hierarchy levels, and OPC biogenesis and myelination is tightly regulated across cortical hierarchy. The higher abundance of OPCs in cortices of higher hierarchy suggests their higher capability of generating new OLGs, contributing to the plasticity of myelin formation and regeneration in these cortices.

Elucidation of regulatory mechanisms underlying diversity of cortical cell types and their distribution is important for understanding not only the structure and function of the brain but also pathogenic mechanisms underlying brain disorders. Through association analysis between cell-type specific CREs and brain disease risk SNPs, we uncovered hundreds of disease-related CREs in the primate brain, including those regulating genes involved in SCZ, BP and AD. Our finding on AD risk SNP-associated CREs exclusively in microglia provides strong support to the notion that microglia play a critical role in the AD etiology[82,83]. Notably, prominent AD risk SNP alleles were found in CREs linked to *TREM2* and *APOE* genes in microglia, and CREs for *TREM2* were found to be human/macaque-specific. In addition, SCZ- and BP-associated CREs were generally enriched only in the primate L2/3 glutamatergic neurons. Importantly, we found a large increase in the frequency of transposon element insertion in human/macaque-specific CREs. This suggests that the high frequency of brain disorder risk SNPs that were linked to human/macaque-specific CREs might be attributed to the extensive proliferation of cortical neurons during primate brain development.

Methods

Tissue collection and nucleus extraction

We dissected brain samples from block-face cortical regions of the hemisphere opposite to that used for Stereo-seq. Using tissue punchers (2.5 - 4 mm in diameter), we segmented the cortical areas at a depth of 1 – 2 mm in the cryostat. After dissection, the samples were promptly frozen in liquid nitrogen and then stored in dry ice or a -80 °C refrigerator. Throughout the sampling process, great care was taken to transfer the tissues to pre-chilled tubes without allowing them to thaw. The single nucleus suspension was prepared following the method described previously[84]. The isolated single nuclei were evenly split for snRNA-seq and snATAC-seq library preparation, respectively. In brief, frozen sections of monkey brain tissue were placed in a Dounce homogenizer containing 2 ml of pre-chilled homogenization buffer (containing 20mM Tris pH8.0, 500mM sucrose, 0.1% NP-40, 0.2U/μL RNase inhibitor, 1x protease inhibitor cocktail, 1% bovine serum albumin (BSA), and 0.1mM DTT), and the Dounce homogenizer was kept on ice during the grinding process. The tissue was homogenized with 10-15 strokes of pestle A, followed by 10-15 strokes of pestle B. Subsequently, 2 ml of homogenization buffer was added to the Dounce homogenizer, and the homogenate was filtered through 30 μm MACS SmartStrainers (Miltenyi Biotech, #130-110-915) into a 15 ml conical tube. The tube was then centrifuged at 500 g for 5 minutes at 4°C to pellet the nuclei. The pellet was resuspended in 1.5 ml of blocking buffer (containing 1% BSA and 0.2U/μL RNase inhibitor in 1x PBS) and centrifuged again at 500 g for 5 minutes at 4°C to pellet the nuclei. The nuclei were finally resuspended in cell resuspension buffer (PBS containing 1% BSA) for subsequent snATAC-seq library preparation.

Library preparation for single-nucleus ATAC-seq

For the preparation of single-nucleus ATAC-seq libraries, we employed the DNBelab C Series Single-Cell ATAC Library Prep Set (BGI-research) with the procedure of step transposed, droplet encapsulation, pre-amplification, emulsion breakage, then the captured beads were collected. After DNA amplification and purification, the indexed sequencing snATAC library was already to sequence. Finally, we used the Qubit ssDNA Assay Kit (Thermo Fisher Scientific) to measure the concentrations of the sequencing libraries. All libraries were 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[85] (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 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)[86]. 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[87] (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[88]. The "gimme.vertebrate.v5.0" was used as reference database.

We assessed TF binding motif enrichment in accessible peaks using chromVAR[89] (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:

$$L_{DISC} = W_{ATAC} \cdot \mathrm{BCELoss}(P_i^{ATAC}, 1) + W_{RNA} \cdot \mathrm{BCELoss}(P_j^{RNA}, 0)$$
$$= -\frac{W_{ATAC}}{n_x}\sum_{i=1}^{n_x}[\log(P_i^{ATAC})] - \frac{W_{RNA}}{n_y}\sum_{j=1}^{n_y}[\log(1 - P_j^{RNA})]$$

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:

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

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

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 14382 homologous genes according to the Ensemble ortholog database (version 91) were used for cross-species analysis, and gene activities of these 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 "hg38.mm10.all.chain"). Next, we reciprocally lifted the accessible elements back to raw genomes and only kept the regions that mapped to their original loci. We further removed converted regions with length > 1kb or in chrM, chr_random, chr_Un, and chr_alt chromosomes. Finally, we compared the homologous CREs using intervene[90] (v.0.6.5) with f = 0.1. CREs showing both sequence similarity and open chromatin between two species were considered as conserved, whereas CREs with no homologous sequence or no open chromatin peaks in the other species were considered as species-specific. In this way, we classified CREs into 7 categories: Human/Macaque/Mouse-conserved (HMm, mammal-conserved), Human/Macaque-specific (HM, human/macaque-specific), Human/Mouse-specific (Hm), Macaque/Mouse-specific (Mm), Human-specific, Macaque-specific, Mouse-specific. We then calculated the proportion of these 7 categories of CREs among all CREs in each cell subclass.

In addition, we download the known transposable elements annotation file from UCSC. Only 4 TE types were kept for further analysis, including "DNA", "SINE", "LINE", "LTR". We calculated the percentage of CREs located in TE in these 7 categories, and performed pairwise Wilcoxon rank sum test with "bonferroni" correction for multiple testing among various categories.

Comparison of LAMP5 and LAMP5/LHX6 between species

Human and mouse snATAC-seq data were retrieved from previously published studies[11,15], and we filtered to retain cells originating from the cortex. Despite some differences in annotation between datasets, we identified the most comparable subclass groups using gene activity scores of classical markers while preserving the original labels from the previous studies for visualization.

In order to compare the cell composition of LAMP5 and LAMP5/LHX6, we merged these two subclasses to calculate their proportions. To facilitate downstream computations, 5,000 cells were randomly sampled from the merged data for each species. Gene activity scores were normalized using a log2 transformation, and we used Seurat and Signac R package to perform integration.

To identify the primate-enriched motifs, we utilized Jaspar motifs (JASPAR2020, all vertebrate) to generate a motif matrix and motif object that was added to the Seurat object using Signac "AddMotifs" function. For motif activity scores, chromVAR was carried out according to default parameters with the Signac "RunChromVAR" function on the respective peak count matrices. The motif matrices of human, macaque, and mouse were merged, and differential enrichment was identified using the "FindMarkers" function with the mean.fxn parameter set to rowMeans.

Characteristic of SST layer and GRN construction

We re-clustered the SST cell subclass as described above to identify cell groups with layer preferences. The previously published study had identified the different layer-specific markers of SST neurons[21], so we examined the gene activity score of these layer markers and assign different cell types to the respective layers.

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[39]. 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[45,46]. 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[91] (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".

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: User id1: juanmeng2021@outlook.com, Password: sunlab2021; User id2: huangyimingy@gmail.com, Password: sunlab2021; User id3: DeboreBerley243@gmail.com, Password: sunlab2021). All raw data have been deposited to CNGB Nucleotide Sequence Archive (accession code: CNP0005530, https://db.cngb.org/search/project/CNP0005530; reviewer's link: https://db.cngb.org/cnsa/project/CNP0005530_f28bb03b/reviewlink/) and are publicly available as of the date of publication.

Code availability

All data were analyzed with standard programs and packages, as detailed above. 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, L. E. & Linnarsson, S. The promise of spatial transcriptomics for neuroscience in the era of molecular cell typing. *Science* **358**, 64-69, doi:10.1126/science.aan6827 (2017).
2 Zhu, Y. *et al.* Spatiotemporal transcriptomic divergence across human and macaque brain development. *Science* **362**, doi:10.1126/science.aat8077 (2018).
3 Khrameeva, E. *et al.* Single-cell-resolution transcriptome map of human, chimpanzee, bonobo, and macaque brains. *Genome Res* **30**, 776-789, doi:10.1101/gr.256958.119 (2020).
4 Krienen, F. M. *et al.* Innovations present in the primate interneuron repertoire. *Nature* **586**, 262-269, doi:10.1038/s41586-020-2781-z (2020).
5 Berg, J. *et al.* Human neocortical expansion involves glutamatergic neuron diversification. *Nature* **598**, 151-158, doi:10.1038/s41586-021-03813-8 (2021).
6 Bakken, T. E. *et al.* Comparative cellular analysis of motor cortex in human, marmoset and mouse. *Nature* **598**, 111-119, doi:10.1038/s41586-021-03465-8 (2021).
7 Network, B. I. C. C. A multimodal cell census and atlas of the mammalian primary motor cortex. *Nature* **598**, 86-102, doi:10.1038/s41586-021-03950-0 (2021).
8 Yao, Z. *et al.* A transcriptomic and epigenomic cell atlas of the mouse primary motor cortex. *Nature* **598**, 103-110, doi:10.1038/s41586-021-03500-8 (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, doi:10.1038/s41593-018-0079-3 (2018).
10 Cusanovich, D. A. *et al.* A Single-Cell Atlas of In Vivo Mammalian Chromatin Accessibility. *Cell* **174**, 1309-1324 e1318, doi:10.1016/j.cell.2018.06.052 (2018).
11 Zu, S. *et al.* Single-cell analysis of chromatin accessibility in the adult mouse brain. *Nature* **624**, 378-389, doi:10.1038/s41586-023-06824-9 (2023).
12 Li, Y. E. *et al.* An atlas of gene regulatory elements in adult mouse cerebrum. *Nature* **598**, 129-136, doi:10.1038/s41586-021-03604-1 (2021).
13 Villar, D. *et al.* Enhancer evolution across 20 mammalian species. *Cell* **160**, 554-566, doi:10.1016/j.cell.2015.01.006 (2015).
14 Won, H. *et al.* Chromosome conformation elucidates regulatory relationships in developing human brain. *Nature* **538**, 523-527, doi:10.1038/nature19847 (2016).
15 Li, Y. E. *et al.* A comparative atlas of single-cell chromatin accessibility in the human brain. *Science* **382**, eadf7044, doi:10.1126/science.adf7044 (2023).
16 Zemke, N. R. *et al.* Conserved and divergent gene regulatory programs of the mammalian neocortex. *Nature* **624**, 390-402, doi:10.1038/s41586-023-06819-6 (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, doi:10.1038/s41467-020-14368-z (2020).
18 Chiou, K. L. *et al.* A single-cell multi-omic atlas spanning the adult rhesus macaque brain. *Sci Adv* **9**, eadh1914, doi:10.1126/sciadv.adh1914 (2023).
19 Lei, Y. *et al.* Spatially resolved gene regulatory and disease-related vulnerability map of the adult Macaque cortex. *Nat Commun* **13**, 6747, doi:10.1038/s41467-022-34413-3 (2022).
20 Han, L. *et al.* Cell transcriptomic atlas of the non-human primate Macaca fascicularis. *Nature* **604**, 723-731, doi:10.1038/s41586-022-04587-3 (2022).
21 Chen, A. *et al.* Single-cell spatial transcriptome reveals cell-type organization in the macaque cortex. *Cell* **186**, 3726-3743 e3724, doi:10.1016/j.cell.2023.06.009 (2023).
22 Hodge, R. D. *et al.* Conserved cell types with divergent features in human versus mouse cortex. *Nature* **573**, 61-68, doi:10.1038/s41586-019-1506-7 (2019).
23 Yao, Z. *et al.* A taxonomy of transcriptomic cell types across the isocortex and hippocampal formation. *Cell* **184**, 3222-3241 e3226, doi:10.1016/j.cell.2021.04.021 (2021).
24 Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**, R137, doi:10.1186/gb-2008-9-9-r137 (2008).
25 Su, Y. *et al.* Neuronal activity modifies the chromatin accessibility landscape in the adult brain. *Nat Neurosci* **20**, 476-483, doi:10.1038/nn.4494 (2017).
26 Pliner, H. A. *et al.* Cicero Predicts cis-Regulatory DNA Interactions from Single-Cell Chromatin Accessibility Data. *Mol Cell* **71**, 858-871 e858, doi:10.1016/j.molcel.2018.06.044 (2018).

27 Reiprich, S. *et al.* Transcription factor Sox10 regulates oligodendroglial Sox9 levels via microRNAs. *Glia* **65**, 1089-1102, doi:10.1002/glia.23146 (2017).
28 Kamimoto, K. *et al.* Dissecting cell identity via network inference and in silico gene perturbation. *Nature* **614**, 742-751, doi:10.1038/s41586-022-05688-9 (2023).
29 Basu, S. *et al.* The Mef2c Gene Dose-Dependently Controls Hippocampal Neurogenesis and the Expression of Autism-Like Behaviors. *J Neurosci* **44**, doi:10.1523/JNEUROSCI.1058-23.2023 (2024).
30 Crow, M., Paul, A., Ballouz, S., Huang, Z. J. & Gillis, J. Characterizing the replicability of cell types defined by single cell RNA-sequencing data using MetaNeighbor. *Nat Commun* **9**, 884, doi:10.1038/s41467-018-03282-0 (2018).
31 Manitt, C. *et al.* dcc orchestrates the development of the prefrontal cortex during adolescence and is altered in psychiatric patients. *Transl Psychiatry* **3**, e338, doi:10.1038/tp.2013.105 (2013).
32 Seo, M. H., 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, doi:10.1186/s12868-021-00668-7 (2021).
33 Chuong, E. B., Elde, N. C. & Feschotte, C. Regulatory activities of transposable elements: from conflicts to benefits. *Nat Rev Genet* **18**, 71-86, doi:10.1038/nrg.2016.139 (2017).
34 Ma, S. *et al.* Molecular and cellular evolution of the primate dorsolateral prefrontal cortex. *Science* **377**, eabo7257, doi:10.1126/science.abo7257 (2022).
35 Urban-Ciecko, J. & Barth, A. L. Somatostatin-expressing neurons in cortical networks. *Nat Rev Neurosci* **17**, 401-409, doi:10.1038/nrn.2016.53 (2016).
36 Berto, S. *et al.* Accelerated evolution of oligodendrocytes in the human brain. *Proc Natl Acad Sci U S A* **116**, 24334-24342, doi:10.1073/pnas.1907982116 (2019).
37 Miller, D. J. *et al.* Prolonged myelination in human neocortical evolution. *Proc Natl Acad Sci U S A* **109**, 16480-16485, doi:10.1073/pnas.1117943109 (2012).
38 Caglayan, E. *et al.* Molecular features driving cellular complexity of human brain evolution. *Nature* **620**, 145-153, doi:10.1038/s41586-023-06338-4 (2023).
39 Felleman, D. J. & Van Essen, D. C. Distributed hierarchical processing in the primate cerebral cortex. *Cereb Cortex* **1**, 1-47, doi:10.1093/cercor/1.1.1-a (1991).
40 Ghorbani, S. *et al.* Fibulin-2 is an extracellular matrix inhibitor of oligodendrocytes relevant to multiple sclerosis. *J Clin Invest* **134**, doi:10.1172/JCI176910 (2024).
41 Fang, R. *et al.* Conservation and divergence of cortical cell organization in human and mouse revealed by MERFISH. *Science* **377**, 56-62, doi:10.1126/science.abm1741 (2022).
42 Li, Y. *et al.* Npas3 deficiency impairs cortical astrogenesis and induces autistic-like behaviors. *Cell Rep* **40**, 111289, doi:10.1016/j.celrep.2022.111289 (2022).
43 Hindorff, L. A. *et al.* Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci U S A* **106**, 9362-9367, doi:10.1073/pnas.0903103106 (2009).
44 Morabito, S. *et al.* Single-nucleus chromatin accessibility and transcriptomic characterization of Alzheimer's disease. *Nat Genet* **53**, 1143-1155, doi:10.1038/s41588-021-00894-z (2021).
45 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, doi:10.1016/j.devcel.2024.01.015 (2024).
46 Pollard, K. S., Hubisz, M. J., Rosenbloom, K. R. & Siepel, A. Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res* **20**, 110-121, doi:10.1101/gr.097857.109 (2010).
47 Bipolar, D., Schizophrenia Working Group of the Psychiatric Genomics Consortium. Electronic address, d. r. v. e., 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, doi:10.1016/j.cell.2018.05.046 (2018).
48 Merenlender-Wagner, A. *et al.* Autophagy has a key role in the pathophysiology of schizophrenia. *Mol Psychiatry* **20**, 126-132, doi:10.1038/mp.2013.174 (2015).
49 Liu, X., Bipolar Genome, S., Kelsoe, J. R. & Greenwood, T. A. A genome-wide association study of bipolar disorder with comorbid eating disorder replicates the SOX2-OT region. *J Affect Disord* **189**, 141-149, doi:10.1016/j.jad.2015.09.029 (2016).
50 Kahler, A. K. *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, doi:10.1002/ajmg.b.30726 (2008).
51 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, doi:10.1038/s41467-021-24391-3 (2021).
52 Yamazaki, K. *et al.* ABCA7 Gene Expression and Genetic Association Study in Schizophrenia. *Neuropsychiatr Dis Treat* **16**, 441-446, doi:10.2147/NDT.S238471 (2020).
53 Gotoh, L. *et al.* Association analysis of adenosine A1 receptor gene (ADORA1) polymorphisms with schizophrenia in a Japanese population. *Psychiatr Genet* **19**, 328-335, doi:10.1097/YPG.0b013e3283328e26 (2009).
54 Takahashi, N. *et al.* Association between chromogranin A gene polymorphism and schizophrenia in the Japanese population. *Schizophr Res* **83**, 179-183, doi:10.1016/j.schres.2005.12.854 (2006).
55 Pozhidaev, I. V. *et al.* Association of Cholinergic Muscarinic M4 Receptor Gene Polymorphism with Schizophrenia. *Appl Clin Genet* **13**, 97-105, doi:10.2147/TACG.S247174 (2020).
56 Li, M. *et al.* Novel genetic susceptibility loci identified by family based whole exome sequencing in Han Chinese schizophrenia patients. *Transl Psychiatry* **10**, 5, doi:10.1038/s41398-020-0708-y (2020).
57 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, doi:10.1016/j.celrep.2019.08.091 (2019).
58 Ventriglia, M. *et al.* Allelic variation in the human prodynorphin gene promoter and schizophrenia. *Neuropsychobiology* **46**, 17-21, doi:10.1159/000063571 (2002).
59 Juraeva, D. *et al.* Integrated pathway-based approach identifies association between genomic regions at CTCF and CACNB2 and schizophrenia. *PLoS Genet* **10**, e1004345, doi:10.1371/journal.pgen.1004345 (2014).
60 Tiihonen, J. *et al.* Molecular signaling pathways underlying schizophrenia. *Schizophr Res* **232**, 33-41, doi:10.1016/j.schres.2021.05.011 (2021).
61 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, doi:10.1038/mp.2008.110 (2010).
62 Aberg, K. A. *et al.* Methylome-wide association study of schizophrenia: identifying blood biomarker signatures of environmental insults. *JAMA Psychiatry* **71**, 255-264, doi:10.1001/jamapsychiatry.2013.3730 (2014).
63 Souza, R. P. *et al.* Genetic association of the GDNF alpha-receptor genes with schizophrenia and clozapine response. *J Psychiatr Res* **44**, 700-706, doi:10.1016/j.jpsychires.2010.01.002 (2010).
64 Pinacho, R. *et al.* Altered CSNK1E, FABP4 and NEFH protein levels in the dorsolateral prefrontal cortex in schizophrenia. *Schizophr Res* **177**, 88-97, doi:10.1016/j.schres.2016.04.050 (2016).
65 Shim, K. H. *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, doi:10.1002/alz.13734 (2024).
66 Colonna, M. & Wang, Y. TREM2 variants: new keys to decipher Alzheimer disease pathogenesis. *Nat Rev Neurosci* **17**, 201-207, doi:10.1038/nrn.2016.7 (2016).
67 Wang, Y. *et al.* Key variants via the Alzheimer's Disease Sequencing Project whole genome sequence data. *Alzheimers Dement* **20**, 3290-3304, doi:10.1002/alz.13705 (2024).
68 Joshi, H. *et al.* Decreased Expression of Cerebral Dopamine Neurotrophic Factor in Platelets of Probable Alzheimer Patients. *Alzheimer Dis Assoc Disord* **36**, 269-271, doi:10.1097/WAD.0000000000000488 (2022).
69 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, doi:10.1186/s12974-023-02914-7 (2023).
70 Chang, J. Y. & Chang, N. S. WWOX dysfunction induces sequential aggregation of TRAPPC6ADelta, TIAF1, tau and amyloid beta, and causes apoptosis. *Cell Death Discov* **1**, 15003, doi:10.1038/cddiscovery.2015.3 (2015).
71 Chiba-Falek, O., Gottschalk, W. K. & Lutz, M. W. The effects of the TOMM40 poly-T alleles on Alzheimer's disease phenotypes. *Alzheimers Dement* **14**, 692-698, doi:10.1016/j.jalz.2018.01.015 (2018).

72 Jackson, R. J., Hyman, B. T. & Serrano-Pozo, A. Multifaceted roles of APOE in Alzheimer disease. *Nat Rev Neurol* **20**, 457-474, doi:10.1038/s41582-024-00988-2 (2024).
73 . (!!! INVALID CITATION !!! 63).
74 . (!!! INVALID CITATION !!! 65).
75 . (!!! INVALID CITATION !!! 66).
76 . (!!! INVALID CITATION !!! 67).
77 Reynolds, L. M. *et al.* DCC Receptors Drive Prefrontal Cortex Maturation by Determining Dopamine Axon Targeting in Adolescence. *Biol Psychiatry* **83**, 181-192, doi:10.1016/j.biopsych.2017.06.009 (2018).
78 Li, H. J. *et al.* Further confirmation of netrin 1 receptor (DCC) as a depression risk gene via integrations of multi-omics data. *Transl Psychiatry* **10**, 98, doi:10.1038/s41398-020-0777-y (2020).
79 Vosberg, D. E., Leyton, M. & Flores, C. The Netrin-1/DCC guidance system: dopamine pathway maturation and psychiatric disorders emerging in adolescence. *Mol Psychiatry* **25**, 297-307, doi:10.1038/s41380-019-0561-7 (2020).
80 Shi, Y. *et al.* Mouse and human share conserved transcriptional programs for interneuron development. *Science* **374**, eabj6641, doi:10.1126/science.abj6641 (2021).
81 Tricoire, L. *et al.* Common origins of hippocampal Ivy and nitric oxide synthase expressing neurogliaform cells. *J Neurosci* **30**, 2165-2176, doi:10.1523/JNEUROSCI.5123-09.2010 (2010).
82 Kosoy, R. *et al.* Genetics of the human microglia regulome refines Alzheimer's disease risk loci. *Nat Genet* **54**, 1145-1154, doi:10.1038/s41588-022-01149-1 (2022).
83 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, doi:10.1016/j.cell.2023.08.039 (2023).
84 Bakken, T. E. *et al.* Single-nucleus and single-cell transcriptomes compared in matched cortical cell types. *PLoS One* **13**, e0209648, doi:10.1371/journal.pone.0209648 (2018).
85 Granja, J. M. *et al.* ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. *Nat Genet* **53**, 403-411, doi:10.1038/s41588-021-00790-6 (2021).
86 Zeisel, A. *et al.* Molecular Architecture of the Mouse Nervous System. *Cell* **174**, 999-1014 e1022, doi:10.1016/j.cell.2018.06.021 (2018).
87 Yu, G., Wang, L. G. & He, Q. Y. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics* **31**, 2382-2383, doi:10.1093/bioinformatics/btv145 (2015).
88 van Heeringen, S. J. & Veenstra, G. J. GimmeMotifs: a de novo motif prediction pipeline for ChIP-sequencing experiments. *Bioinformatics* **27**, 270-271, doi:10.1093/bioinformatics/btq636 (2011).
89 Schep, A. N., Wu, B., Buenrostro, J. D. & Greenleaf, W. J. chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. *Nat Methods* **14**, 975-978, doi:10.1038/nmeth.4401 (2017).
90 McLean, C. Y. *et al.* GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotechnol* **28**, 495-501, doi:10.1038/nbt.1630 (2010).
91 Gu, Z. & Hubschmann, D. rGREAT: an R/bioconductor package for functional enrichment on genomic regions. *Bioinformatics* **39**, doi:10.1093/bioinformatics/btac745 (2023).

Acknowledgements

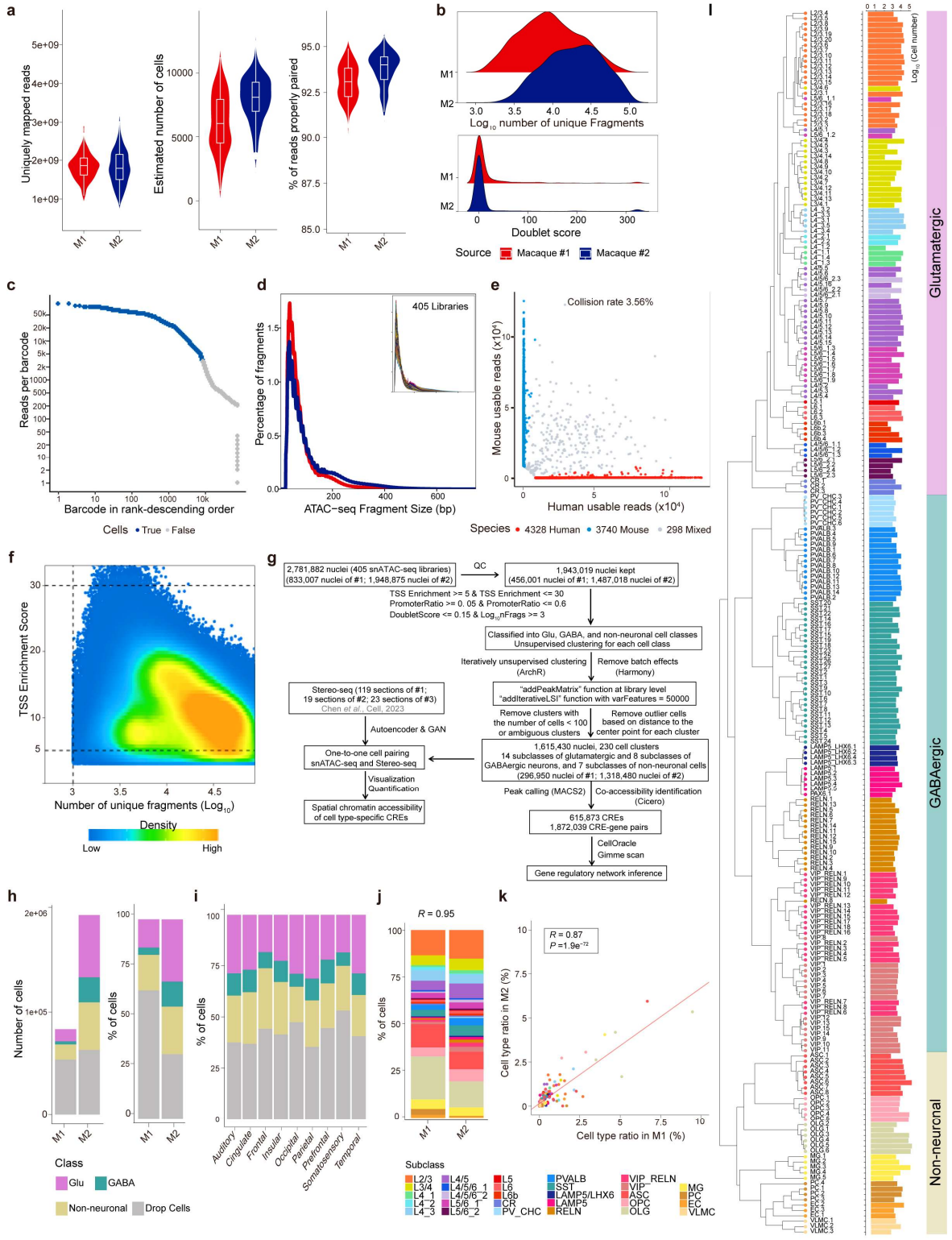
The project was supported by National Science and Technology Innovation 2030 Major Program (STI2030-2021ZD0200100), National Natural Science Foundation of China (NSFC) Outstanding Youth Foundation (No. T2422026), National Key R&D Program of China No.2022YEF0203200, National Natural Science Foundation of China (No. U23A6010, 32100487) and Shanghai Science and Technology Development Funds (23QA1410400). This paper is from the Mesoscopic Brain Mapping Consortium. This work is part of the "SpatioTemporal Omics Consortium" (STOC) paper package. A list of STOC members is available at: http://sto-consortium.org.

Author contributions

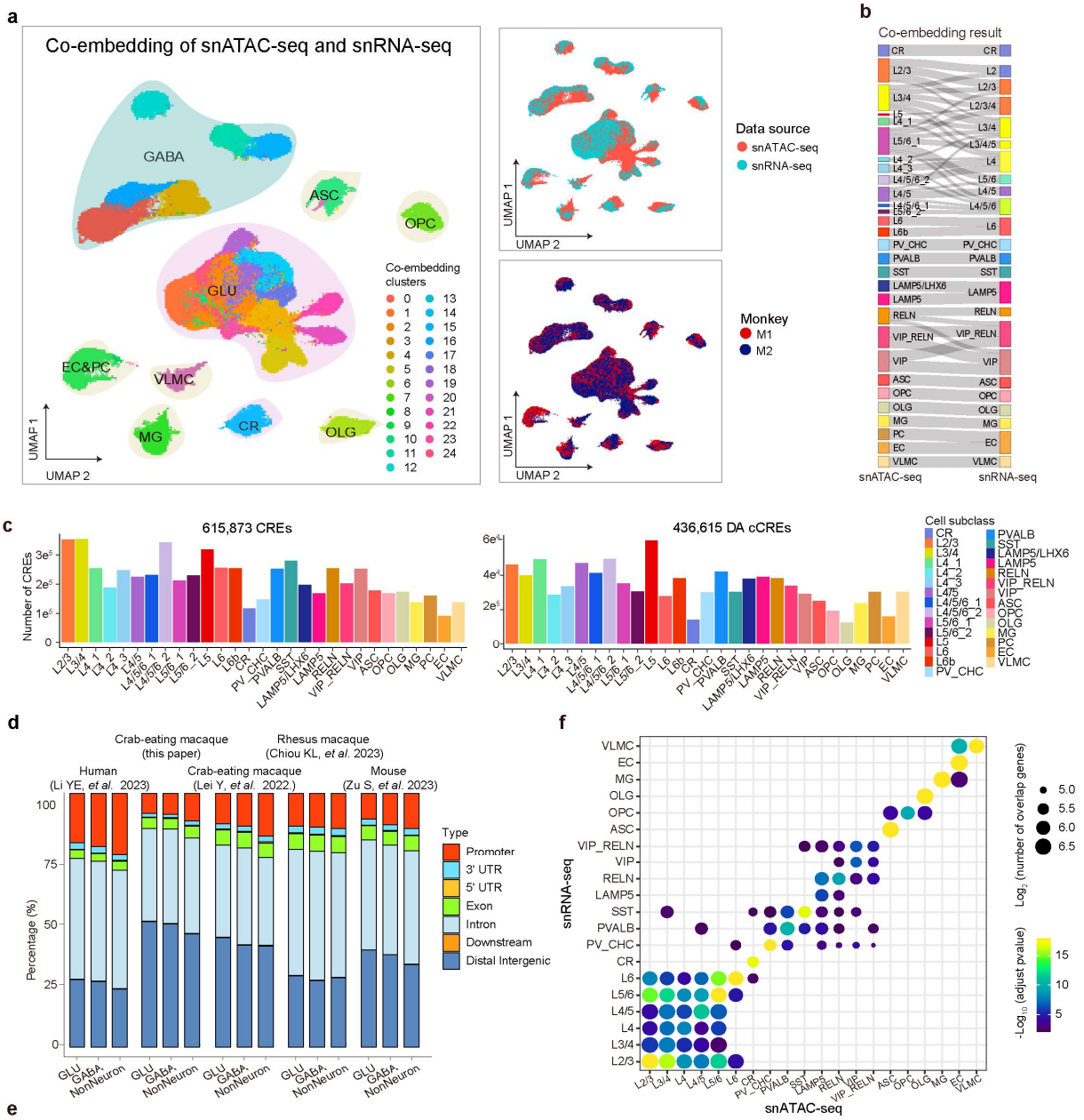
Methodology, J.M., C.C., Z.Z., Y.S., K.H. and T.F.; Investigation, J.M., Y.L., C.C., Z.Z., Y.B.; Formal Analysis, J.M., C.C., Z.Z., Y.S. and Y.H.; Writing – Original Draft, Y.L. and Y-D.S.; Writing – Review & Editing, J.M., T.F., Z.L., Z.S., L.L., C.L., and M.P.; Conceptualization and supervision, S.L.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

Competing interests

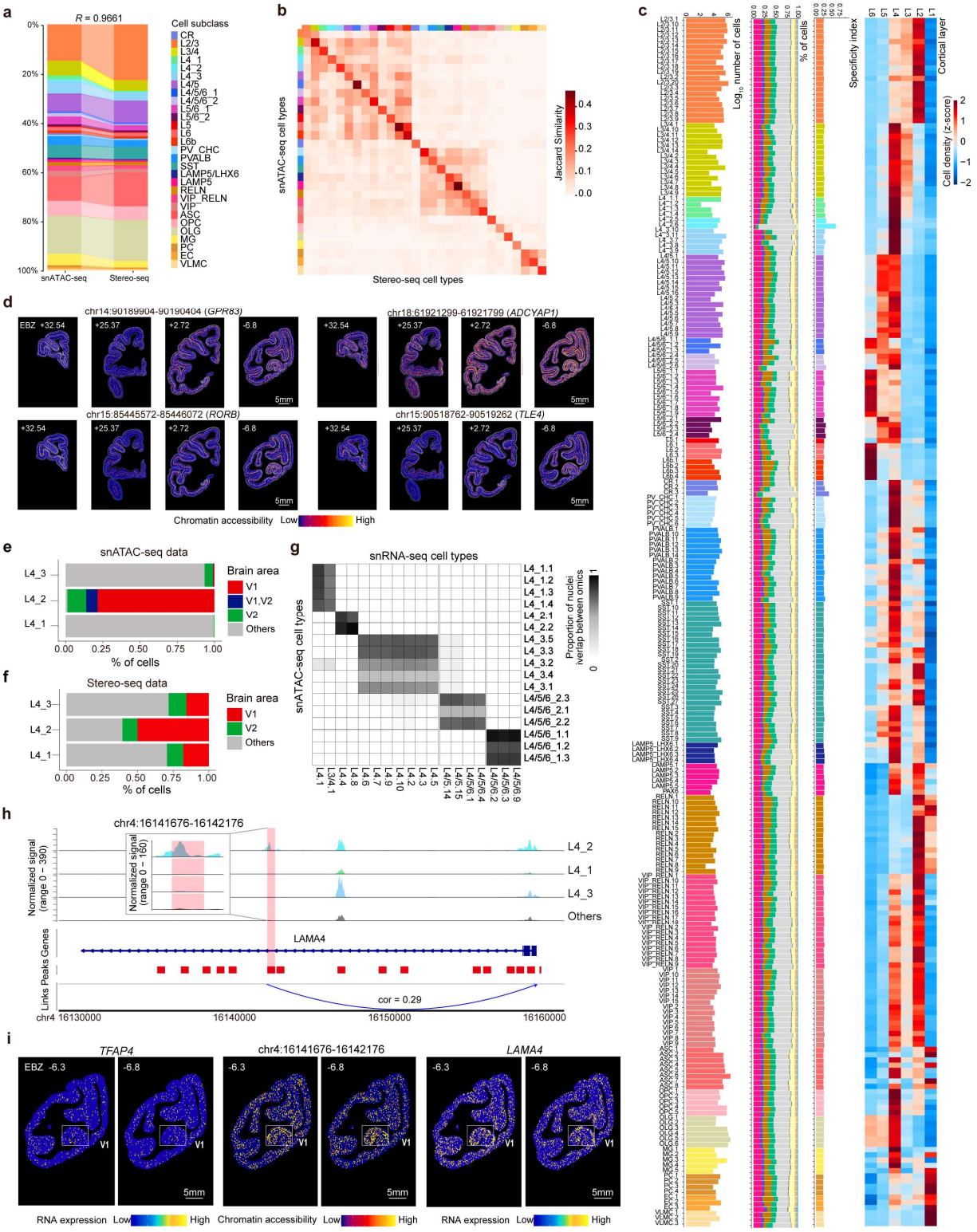
All authors declare no competing interests.


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
Percentage of fragments
ATAC−seq Fragment Size (bp)
405 Libraries
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
Remove clusters with the number of cells < 100 or ambiguous clusters
Remove outlier cells based on distance to the center point for each cluster
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)
Peak calling (MACS2)
Co-accessibility identification (Cicero)
615,873 CREs
1,872,039 CRE-gene pairs
CellOracle
Gimme scan
Gene regulatory network inference
Stereo-seq (119 sections of #1;
19 sections of #2; 23 sections of #3)
Chen et al., Cell, 2023
Autoencoder & GAN
One-to-one cell pairing
snATAC-seq and Stereo-seq
Visualization
Quantification
Spatial chromatin accessibility
of cell type-specific CREs
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

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.

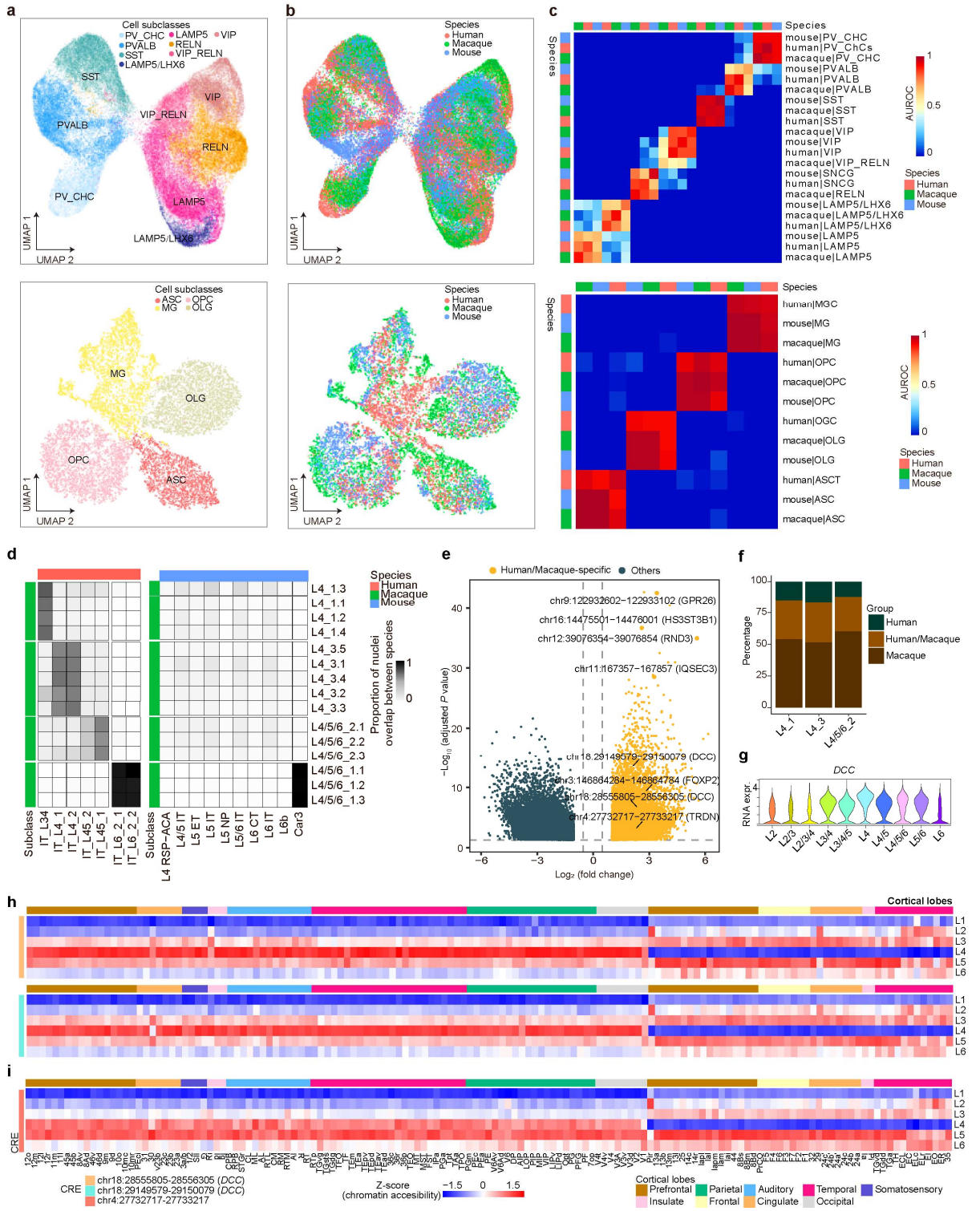

a
Co-embedding of snATAC-seq and snRNA-seq
GABA
ASC
OPC
GLU
EC&PC
VLMC
MG
CR
OLG
UMAP 1
UMAP 2
Co-embedding clusters
Data source
snATAC-seq
snRNA-seq
Monkey
M1
M2
b
Co-embedding result
snATAC-seq
snRNA-seq
c
615,873 CREs
436,615 DA cCREs
Number of CREs
Cell subclass
d
Crab-eating macaque (this paper)
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

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



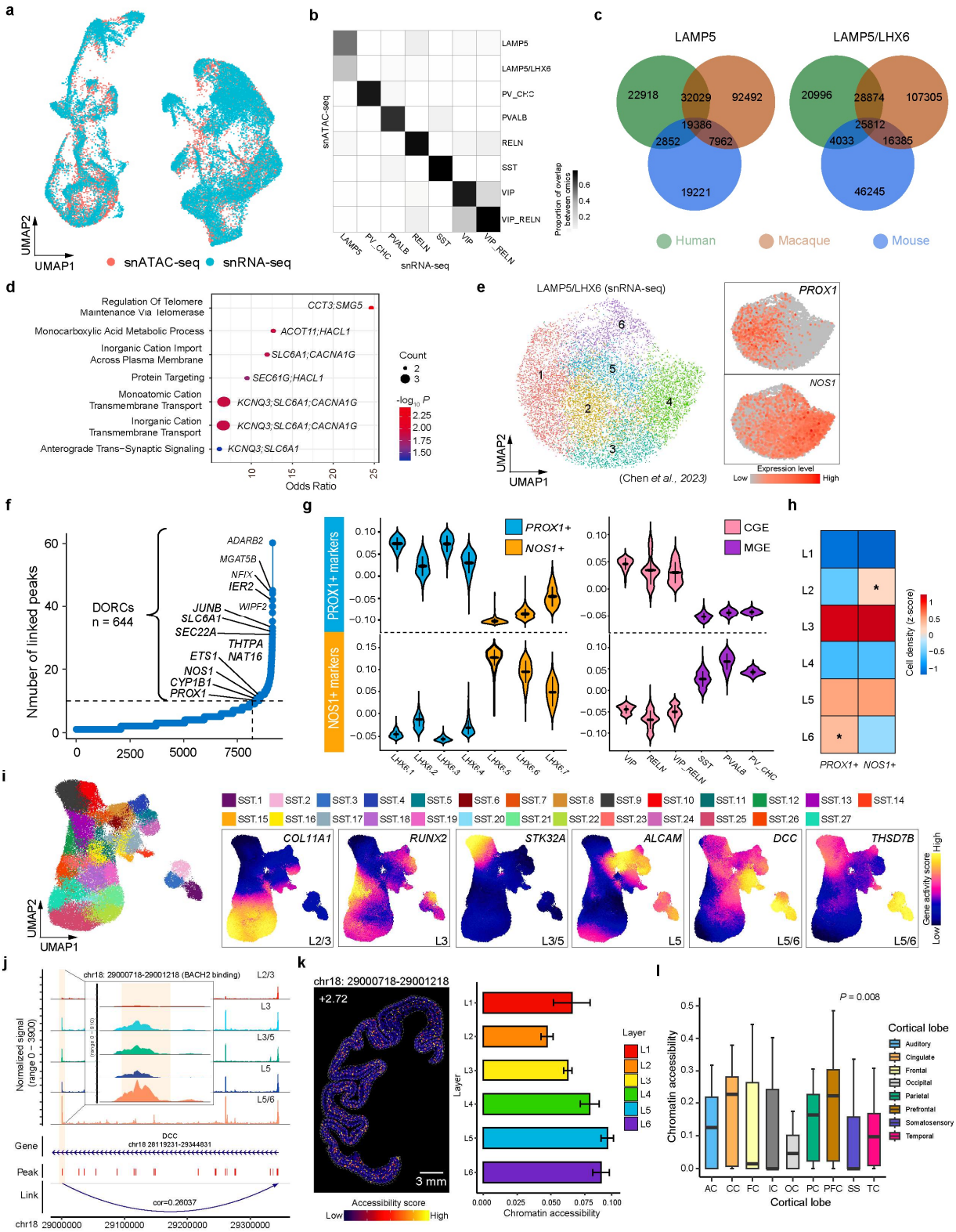
Extended Data Fig. 3. Spatial mapping of chromatin accessibility of CREs in various cell types. a, Correspondence of percentages of various cell clusters revealed by snATAC-seq clustering and spatial mapping result on the Stereo-seq map. **b**, Heatmap showing the Jaccard similarity of marker genes identified 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**, Genome browser track view showing a differential accessible distal CRE linked to *LAMA4* across three L4 glutamatergic clusters and other glutamatergic clusters. **i**, 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).



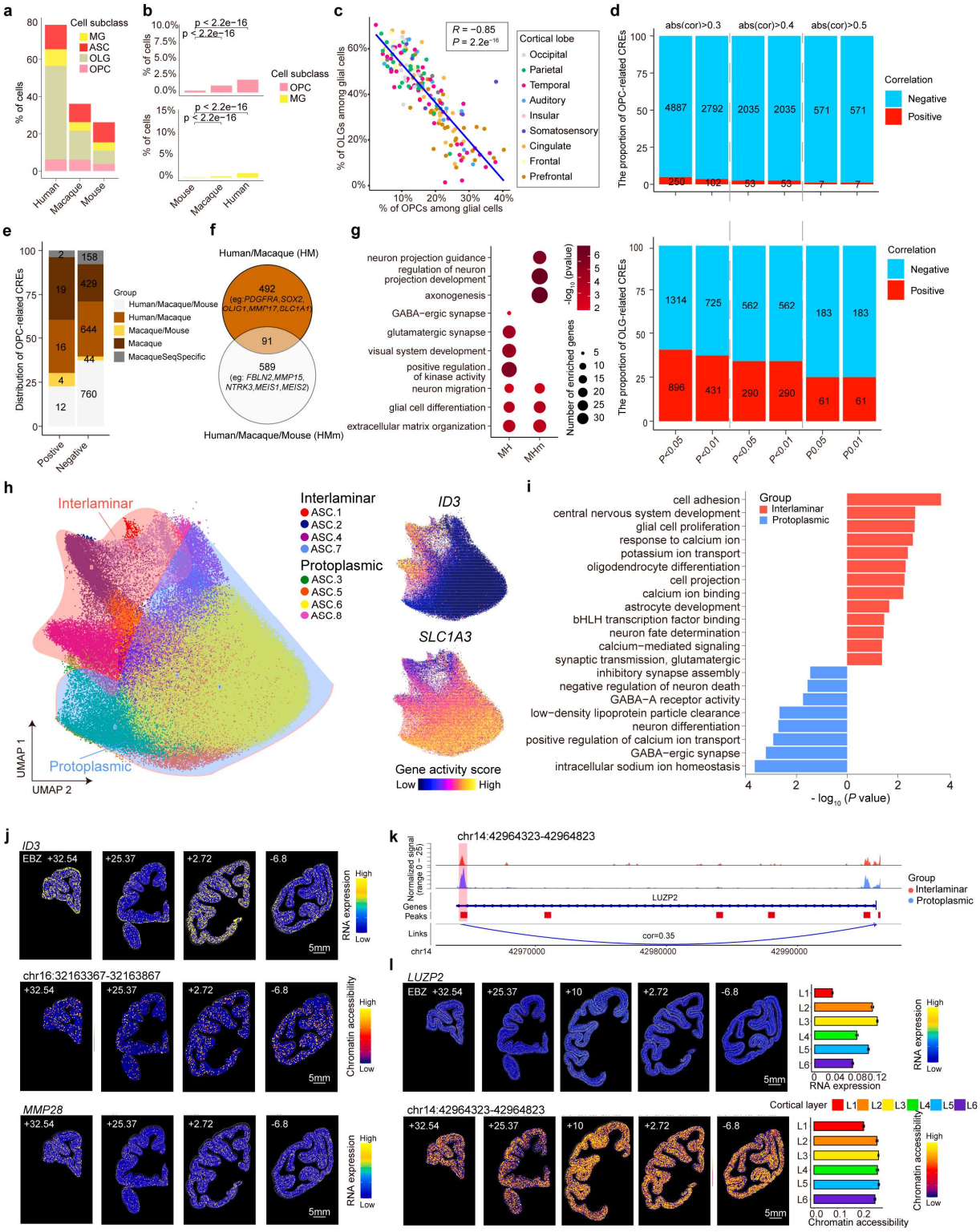
Extended Data Fig. 4. Cross-species comparison of CREs for glutamatergic neurons. a, UMAP co-embedding of GABAergic neurons (upper) and non-neuronal cells (below) from snATAC-seq data of macaques, humans and mice. Single nuclei were colored by their cell subclasses. **b**, UMAP layout of GABAergic neurons (upper) and non-neuronal

cells (below) from snATAC-seq data of macaques, humans and mice. Single nuclei were colored by different species. **c**, Heatmap showing the similarity of snATAC-seq derived GABAergic (upper) and non-neuronal (below) cell clusters across macaques, humans and mice using MetaNeighbor. **d**, Heatmap depicting the pairwise comparisons of snATAC-seq derived glutamatergic cell types between macaques and humans or mice. **e**, Volcano plot showing the differentially accessible CREs between human/macaque-specific and other snATAC-seq-defined cell types. **f**, Bar plots showing the conservation between humans and macaques for CREs with high chromatin accessibility in human/macaque-specific cell types shown in **e**. **g**, Violin plots showing the expression level of *DCC* among various glutamatergic neuron subclasses of the macaque snRNA-seq data. **h**, Heatmap showing the average chromatin accessibility of CREs associated with *DCC* across cortical layers and regions. **i**, Heatmap showing the average chromatin accessibility of CRE linked to *TRDN* across cortical layers and regions.



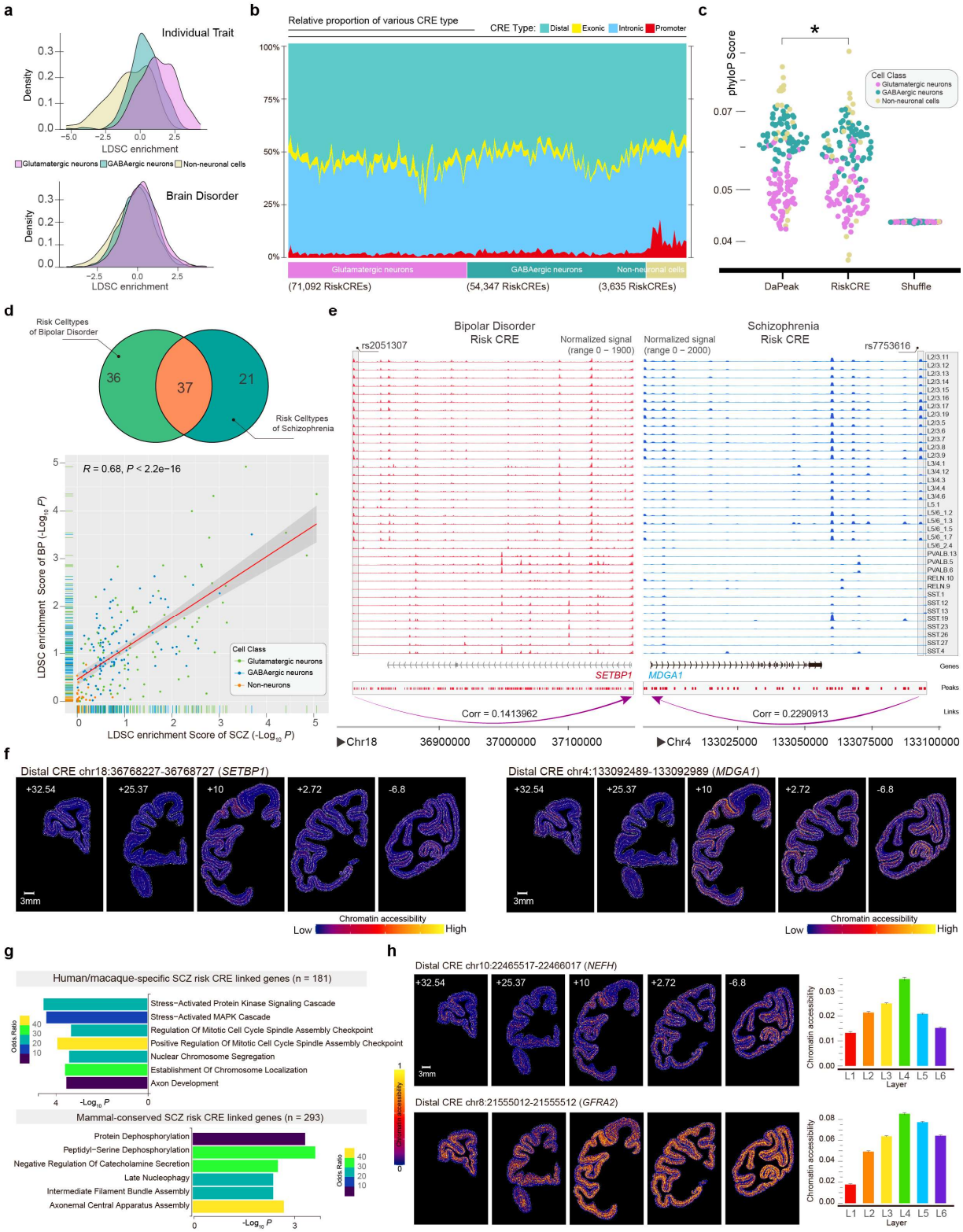
Extended Data Fig. 5. Characterization of CREs for GABAergic neurons. a, UMAP embedding of snATAC-seq and snRNA-seq GABAergic cells from macaque cortex. **b,** Heatmap of overlap scores showing the correspondence

between snRNA-seq and snATAC-seq derived macaque GABAergic cell subclasses. **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,** The number of significantly correlated peaks for each gene in LAMP5_LHX6 neurons. **g,** The gene activity scores of marker gene set for *PROX1*$^{+}$ and *NOS1*$^{+}$ in LAMP5_LHX6 cell clusters (left) and in MGE- and CGE-derived inhibitory neurons. **h,** Layer distribution of the two distinct groups of LAMP5/LHX6. An asterisk (*) indicates a significant difference between the two groups in the given layer ($P < 0.01$). **i,** UMAP embedding of snATAC-seq-based SST neurons (left) and normalized gene accessibility of layer markers for SST neurons (right). **j,** Genome browser track view showing a differential accessible distal CRE linked to *DCC* across various SST cell types. **k,** Spatial visualization of chromatin accessibility for the CRE linked to *DCC* (shown in **j**) in representative section (left), and bar plots showing the average chromatin accessibility in six cortical layers for the CRE (right). **i,** Boxplot showing chromatin accessibility of the CRE associated with *DCC* (shown in **j**) in SST neurons across cortical lobes. *P*-values: Kruskal-Wallis test.



Extended Data Fig. 6. 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-specific 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-specific 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**, Spatial maps of expression patterns of *ID3* and *MMP28*, and chromatin accessibility of the CRE linked to *MMP28* at four representative sections. **k**, Genome browser tracks showing the chromatin accessibility of CRE linked to *LUZP2* between protoplasmic and interlaminar astrocytes. **l**, 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. 7. 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.

Supplemental tables

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

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

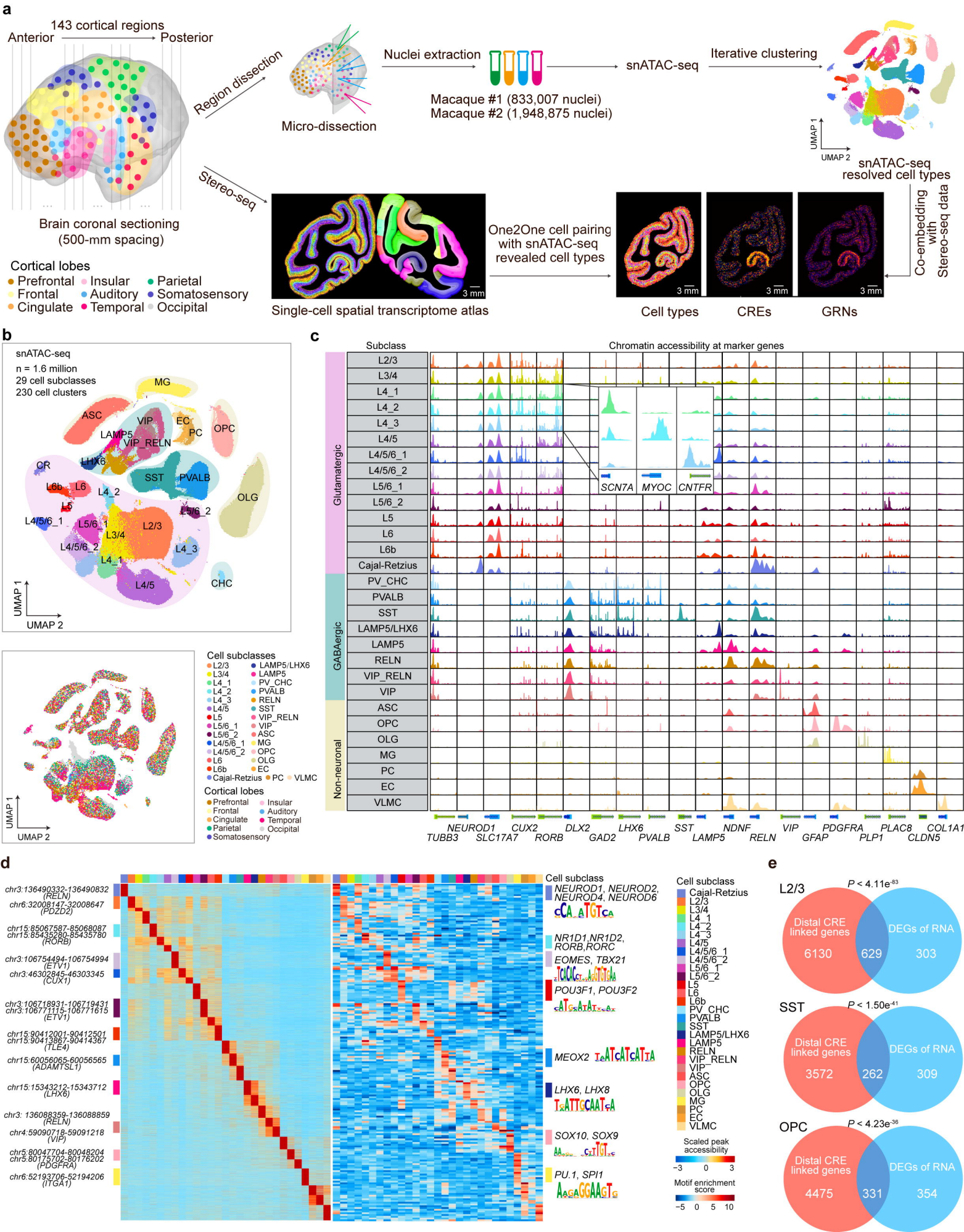
Supplementary Table 3. The cis-regulatory elements and their linked genes among various cell types.

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

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

Supplementary Table 6. List of CREs in OPCs and OLGs positively and negatively correlated with hierarchy levels of various brain regions in the visual system.

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


a
143 cortical regions
Anterior
Posterior
Brain coronal sectioning
(500-mm spacing)
Region dissection
Micro-dissection
Nuclei extraction
Macaque #1 (833,007 nuclei)
Macaque #2 (1,948,875 nuclei)
snATAC-seq
Iterative clustering
UMAP 1
UMAP 2
snATAC-seq
resolved cell types
Stereo-seq
Cortical lobes
Prefrontal
Frontal
Cingulate
Insular
Auditory
Temporal
Parietal
Somatosensory
Occipital
3 mm
Single-cell spatial transcriptome atlas
One2One cell pairing
with snATAC-seq
revealed cell types
Cell types
CREs
GRNs
Co-embedding
with
Stereo-seq data
b
snATAC-seq
n = 1.6 million
29 cell subclasses
230 cell clusters
MG
ASC
VIP
EC
PC
OPC
LAMP5
VIP_RELN
LHX6
OR
SST
PVALB
L6b
L6
L4_2
L5
L5/6_2
OLG
L4/5/6_1
L5/6_1
L2/3
L3/4
L4/5/6_2
L4_3
L4_1
L4/5
CHC
UMAP 1
UMAP 2
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
Cajal-Retzius
LAMP5/LHX6
LAMP5
PV_CHC
PVALB
RELN
SST
VIP_RELN
VIP
ASC
MG
OPC
OLG
EC
PC
VLMC
Cortical lobes
Prefrontal
Frontal
Cingulate
Parietal
Insular
Auditory
Temporal
Occipital
Somatosensory
c
Subclass
Chromatin accessibility at marker genes
Glutamatergic
GABAergic
Non-neuronal
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
Cajal-Retzius
PV_CHC
PVALB
SST
LAMP5/LHX6
LAMP5
RELN
VIP_RELN
VIP
ASC
OPC
OLG
MG
PC
EC
VLMC
SCN7A
MYOC
CNTFR
NEUROD1
TUBB3
CUX2
SLC17A7
DLX2
RORB
LHX6
GAD2
SST
PVALB
NDNF
LAMP5
VIP
RELN
PDGFRA
GFAP
PLAC8
PLP1
COL1A1
CLDN5
d
chr3:136490332-136490832 (RELN)
chr6:32008147-32008647 (PDZD2)
chr15:85067587-85068087
chr15:85435280-85435780 (RORB)
chr3:106754494-106754994 (ETV1)
chr3:46302845-46303345 (CUX1)
chr3:106718931-106719431
chr3:106771115-106771615 (ETV1)
chr15:90412001-90412501
chr15:90413867-90414367 (TLE4)
chr15:60056065-60056565 (ADAMTSL1)
chr15:15343212-15343712 (LHX6)
chr3: 136088359-136088859 (RELN)
chr4:59090718-59091218 (VIP)
chr5:80047704-80048204
chr5:80175702-80176202 (PDGFRA)
chr6:52193706-52194206 (ITGA1)
Cell subclass
NEUROD1, NEUROD2, NEUROD4, NEUROD6
NR1D1,NR1D2, RORB,RORC
EOMES, TBX21
POU3F1, POU3F2
MEOX2
LHX6, LHX8
SOX10, SOX9
PU.1, SPI1
Cell subclass
Cajal-Retzius
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
PV_CHC
PVALB
SST
LAMP5/LHX6
LAMP5
RELN
VIP_RELN
VIP
ASC
OPC
OLG
MG
PC
EC
VLMC
Scaled peak accessibility
−3
0
3
Motif enrichment score
−5
0
5
10
e
L2/3
P < 4.11e-83
Distal CRE linked genes
6130
629
DEGs of RNA
303
SST
P < 1.50e-41
Distal CRE linked genes
3572
262
DEGs of RNA
309
OPC
P < 4.23e-36
Distal CRE linked genes
4475
331
DEGs of RNA
354