Single-cell regulatory landscape and disease vulnerability map of adult Macaque cortex

Ying Lei[1,13,†], Mengnan Cheng[1,2,†], Zihao Li[1,2,†], Zhenkun Zhuang[1,3,†], Liang Wu[1,2,†], Lei Han[1,13], Shang Liu[1,2], Zhihao Huang[1], Jingkuan Wei[4], Yuzhou Wang[1], Zifei Wang[1,2], Liqin Xu[1], Taotao Pan[1,2], Jiarui Xie[1,3], Chuanyu Liu[1], Giacomo Volpe[5], Carl Ward[5], Yiwei Lai[5], Jiangshan Xu[1,2], Yue Yuan[1,2], Mingyue Wang[1,2], Hao Yu[1], Haixi Sun[1], Qichao Yu[1], Liang Wu[1,2,14], Dandan Chen[6], Chunqing Wang[1,2], Chi Wai Wong[7], Wei Liu[7], Liangzhi Xu[7], Zhouchun Shang[1,8], Guibo Li[1,14], Kun Ma[1], Le Cheng[9], Fei Ling[3], Tao Tan[4], Kai Chen[4], Ao Chen[1], Bosiljka Tasic[10], Michael Dean[11], Weizhi Ji[4], Huanming Yang[1,15], Ying Gu[1], Miguel A. Esteban[5,1,17], Xun Xu[1,16], Hongkui Zeng[10,*], Longqi Liu[1,17,*], Yuyu Niu[4,12,*], Yong Hou[1,14,*], Shiping Liu[1,14,*]

Affiliations:

[1] BGI-Shenzhen, Shenzhen 518083, China

[2] BGI Education Center, University of Chinese Academy of Sciences, Shenzhen 518083, China.

[3] School of Biology and Biological Engineering, South China University of Technology, Guangzhou 510006, China

[4] Yunnan Key Laboratory of Primate Biomedical Research, Institute of Primate Translational Medicine, Kunming University of Science and Technology, Kunming 650500, China

[5] Laboratory of Integrative biology, Guangzhou Institutes of Biomedicine and Health, CAS, Guangzhou 510530, China

[6] BGI-GenoImmune, BGI-Shenzhen, Wuhan 430074, China

[7] Huazhen Biosciences, Guangzhou 510900, China

[8] MGI, BGI-Shenzhen, Shenzhen 518083, China

[9] BGI-Yunnan, BGI-Shenzhen, Kunming 650106, China

[10] Allen Institute for Brain Science, Seattle, WA 98109, USA

[11] Cancer and Inflammation Program, National Cancer Institute at Frederick, Building 560, Frederick, MD 21702, USA

[12] Faculty of Life Science and Technology, Kunming University of Science and Technology, Kunming, Yunnan 650500, China

[13] Shenzhen Bay Laboratory, Shenzhen 51800, China

[14] Shenzhen Key Laboratory of Single-Cell Omics, BGI-Shenzhen, Shenzhen 518120，China

[15] Guangdong Provincial Academician Workstation of BGI Synthetic Genomics，BGI-Shenzhen, Shenzhen 518120, China

[16] Guangdong Provincial Key Laboratory of Genome Read and Write, BGI-Shenzhen, Shenzhen 518120, China

[17] Institute for Stem Cells and Regeneration, Chinese Academy of Sciences, Beijing 100101, China

[†]These authors contributed equally

*Correspondence: liushiping@genomics.cn (S. L.), houyong@genomics.cn (Y. H.), niuyy@lpbr.cn (Y. N.), liulongqi@genomics.cn (L. L.), hongkuiz@alleninstitute.org (H. Z.)

Abstract:

Non-human primates (NHP) provide a unique opportunity to study human neurological diseases, yet detailed characterization of the cell types and transcriptional regulatory features in the NHP brain is lacking. We applied a combinatorial indexing assay, sci-ATAC-seq, as well as single-nuclei RNA-seq, to profile chromatin accessibility in 43,793 single cells and transcriptomics in 11,477 cells, respectively, from prefrontal cortex, primary motor cortex and the primary visual cortex of adult cynomolgus monkey *Macaca fascularis*. Integrative analysis of these two datasets, resolved regulatory elements and transcription factors that specify cell type distinctions, and discovered area-specific diversity in chromatin accessibility and gene expression within excitatory neurons. We also constructed the dynamic landscape of chromatin accessibility and gene expression of oligodendrocyte maturation to characterize adult remyelination. Furthermore, we identified cell type-specific enrichment of differentially spliced gene isoforms and disease-associated single nucleotide polymorphisms. Our datasets permit integrative exploration of complex regulatory dynamics in macaque brain tissue at single-cell resolution.

Main Text:

The cortical organization of macaque brain is similar to human in many aspects, such as specialized primary visual cortex and dorsal and ventral visual streams. The similar neocortex ratio of the monkey and human brain offers a unique model to study features of human neurodevelopment and neuropsychiatric diseases (*1-3*).

Single-cell genomic sequencing allows to study the underlying diversity and regulatory mechanisms of cortical cells at unprecedented resolution. Profiling of macaque brain by single-cell sequencing can provide deep understanding of the human brain and identify markers and molecular signatures of neuropsychiatric diseases. Previous single-cell RNA-sequencing studies of prenatal and adult macaque brain have shown functionally distinct cortical cell types as well as subtypes across multiple cortical brain areas, and variations in gene expression across these cell types (*2*). The state of chromatin, open or closed, exerts a fundamental role in regulating gene expression, and together with gene expression profile could provide complementary understanding of the molecular properties of brain cells. Single-cell-based chromatin states assays can identify cell-type-specific transcriptional regulatory elements and predict potential master transcriptional regulators (*4*). A survey of chromatin states in bulk of several cerebral cortical regions and hippocampus of the macaque brain revealed region-related chromatin accessibility patterns (*5*). However, a systematic characterization of single-cell-based chromatin accessibility of the macaque cortex with region-specific and single-cell resolution is an unmet requirement for advancing the field.

The cynomolgus monkey (*Macaca fascicularis*) is one of the most studied non-human primates (NHP) in neuroscience and medicine (*6, 7*). Recent advances in transgenesis and genome-editing technologies have led to successful development of

new cynomolgus monkey models to study genetic human neurological disorders (*8-10*), making this species an excellent experimental NHP model for brain researchers.

In the present study, we sought to explore the single-cell chromatin state of prefrontal cortex (PFC), primary motor cortex (M1) and primary visual cortex (V1) of cynomolgus monkey by single-cell combinatorial indexing assay for transposase-accessible chromatin using sequencing (sci-ATAC-seq), as well as full length single-nucleus RNA-seq (snRNA-seq). Through this massive parallel and integrative analysis, we defined cell type-specific and regional-specific regulatory elements, delineated the dynamic regulatory landscape of remyelination and linked macaque cortical cell types to human neurological disease risk.

Results

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

We applied modified combinatorial barcoding assisted single-cell ATAC-seq (*4*) to three tissues from cynomolgus monkey brain cortex: dorsolateral PFC, M1 and V1 (**Figure 1A**). We used 10,000 unique nuclear fragments per cell and 30% reads enriched in peak regions as cut-off for sci-ATAC-seq data to exclude low-quality cells (**Figure S1A and S1B**). A total of 43,793 qualified cells were used for further analysis: 14,480 from the PFC, 9,853 from M1 and 19,460 from V1 (**Table S1**). In our analysis, we achieved a median read depth per nucleus of 58,072, and a median fraction of reads in peak regions at 45.6% (**Figure S1B and S1C**). We used sci-ATAC-seq reads to generate a master list of 130,525 accessible sites, the median peak number was 4,973 (**Figure S1D**). These accessible sites were subsequently used to identify a total

of 27 cell clusters and visualized them with uniform manifold approximation and projection (UMAP) (**Figure 1B and Figure S1F**). The identity of these cell clusters was annotated manually by promoter accessibility of brain cell marker genes (**Figure 1C**). Using this approach, we characterized eight major cortical cell populations, including excitatory neurons (hereafter labelled as EX, accessible at *NEFH* and *SLC17A7*), inhibitory neurons (IN, accessible at *NEFH* and *GAD1*), oligodendrocytes (OLI, accessible at *MOBP*), oligodendrocyte precursor cells (OPC, accessible at *PDGFRA*), microglia (MIC, accessible at *AIF1*), astrocytes (AST, accessible at *SLC7A10*), endothelial cells (ENDO, accessible at *CLDN5*) and pericytes (PERI, accessible at *KCNJ8*) (**Figure 1C**). EX could be divided into upper layer (L1 to L4) and lower layer (L5 and L6) by gene activity score of known layer markers (**Figure S1E)**: *CUX2* for L2 to L4, *RORB* for L3 to L5 and *SEMA3E* for L5/6. The four sub clusters of IN could be further assigned to VIP, LAMP5, SST and PVALB, respectively, due to the distinct open peaks at the promoter of these genes (**Figure 1C**). It is well known that cortical IN roughly fall into two major branches corresponding to their differential origins in the caudal ganglionic eminence (CGE) and medial ganglionic eminence (MGE), respectively. *ADARB2*, a marker gene of CGE-derived IN, had accessible chromatin at its promoter in both VIP and LAMP5 sub clusters, while *LHX6*, a marker gene of MGE-derived IN, revealed higher accessibility of its promoter site in SST and PVALB sub clusters, indicating that VIP and LAMP5 IN are derived from CGE while SST and PVALB are from MGE during corticogenesis.

To identify the key transcriptional regulators that determine the specificity of cortical cell types, we performed transcription factor (TF) motif enrichment analysis at differentially accessible open chromatin regions in each cell type or subtypes (**Figure 1D, Table S2**). This showed that cell type-enriched TF binding motif in

mouse and human brains (*11, 12*) can also be found in the corresponding macaque cortical cell types. NEUROD1, NEUROG2 and MEF2A, MEF2B and MEF2C motifs were enriched in EX clusters, consistent with their roles in EX specification and activity-dependent regulation on synapse numbers (*13, 14*). Likewise, SOX-family motifs (SOX TFs contribute to OLI migration) were enriched in OLI (*15*), while PU.1 (relevant for microglial viability and function) motifs were enriched in MIC (*16*). Finally, NF1 and LHX2 (known regulators of astrogliogenesis) motifs were enriched in AST (*17, 18*).

These findings demonstrate species conservation of regulatory signatures for both neuronal and non-neuronal cortical cell types.

Linking chromatin accessibility to transcriptome in different cell types of the monkey cortex

We next assessed how the single-cell epigenetic profiles correlate with single-cell transcriptomic profiles by applying Smart-seq2 based snRNA-seq (*19*) to PFC, M1 and V1 of the macaque neocortex (**Figure 1A and Figure S2A**). After quality filtering and cell clustering, we profiled the transcriptome from 11,477 single-nuclei (6,949 from PFC, 1,971 from M1 and 2,557 from V1) and identified 17 distinct clusters which were displayed by UMAP visualization (**Figure 2A**, **Figure S2B and S2C**, **Table S1**). Based on the known marker genes for cortical cell types, all clusters were annotated as EX, IN and non-neuronal cells, the latter being OLI, OPC, AST, MIC, ENDO and PERI (**Figure 2B and Table S3**). We re-clustered EX into 11 sub clusters and arranged them into upper- and lower-cortical layers by known layer markers (**Figure S2E**). For IN types, they could be assigned to CGE- and MGE-derived subtypes with mutually exclusive markers **(Figure S2F).** Gene ontology (GO)

analysis of differentially expressed genes (DEG) for the major cell types were in agreement with the expected corresponding biological processes (**Figure S2D**), further confirming our cell type assignments.

In order to understand the link between transcriptome and open chromatin, we co-embedded the snRNA-seq and sci-ATAC-seq data of cortical cells with Seurat V3 (*20*). We converted the peak by cells matrix to the promoter region (TSS ±3 kb) and then anchored to the gene expression matrices. Using this anchor method, cells from our sci-ATAC-seq were positioned proximal to cells with matching assignments of the snRNA-seq data (**Figure 2C**). Notably, our high-resolution sci-ATAC-seq data contained similar sub clusters corresponding to the same major brain cell types as the snRNA-seq. The EX epigenetic subtypes corresponding to the upper and lower layers were well paired with the corresponding transcriptomic EX subtypes (**Figure 2D**). For the IN type, the epigenetic subtypes of VIP, LAMP5, SST and PVALB were also well paired with the corresponding transcriptomic subtypes. The epigenetic LAMP5 subtype of IN fell into two sub clusters, one that co-clustered with the transcriptomic LAMP5 subtype, while the other co-clustered with the transcriptomic subtype that was characterized by the expression of RELN (**Figure 2D and Figure S2F**). To further support the concordance of cell type identification between snRNA-seq and sci-ATAC-seq, we looked at key marker gene expression in the identified transcriptomic cell types and the gene activity score in the corresponding epigenetic cell types (**Figure 2E**).

In summary, our results demonstrate good consistency of the cortical cell classification between epigenetic and transcriptomic data, enabling further integrative analysis.

Regional excitatory neuron heterogeneity revealed by single-cell chromatin state in monkey cortex

Given that differences in the relative proportions of cell types across the cortical area suggest differential circuit functions, we proceeded to examine the proportion of sampled cell types across PFC, M1 and V1 in our sci-ATAC-seq dataset. The proportion of EX in V1 was 70.8%, while similar and lower proportions were observed in PFC and M1 (50.7% and 49.5%, respectively). The proportion of IN in V1 was 9.2%, while a higher proportion was found in PFC and M1 (20.0% and 11.7%, respectively). The proportions of non-neuronal cell types were similar across PFC, M1 and V1, except for OLI in M1 (16.7%), which was roughly three times higher compared with PFC (5.6%) and V1 (6.2%) (**Figure S3A**).

Interestingly, we found area-specific epigenetic clusters of EX in frontal cortex (FC: referring to PFC and M1) and V1, while no such brain region-separated clusters were identified in IN and non-neuronal cells (**Figure 3A**). Recent single-cell studies have resolved area-specific transcriptomic features of cortical EX in mice (*21*), and the molecular and epigenetic signatures in the developing human brain (*22, 23*). However, the epigenetic features that correspond to area-specific EX in adult cortex (monkey or human) are yet to be clarified. Therefore, we performed the inter-regional comparison of the epigenetic profile of EX between FC and V1. Given the clear laminar organization in the neocortex (Bernard et al., 2012; Tasic et al., 2018), we generated the differentially accessible peak maps of FC and V1 in upper- and lower-layer EX (**Figure 3B**). We obtained 20,291 and 6,235 differentially accessible peaks from FC and V1 of upper layer EX, and 10,984 and 4,226 differentially accessible peaks from FC and V1 of lower layer EX (**Table S4**). We assigned the peaks within TSS ± 3 kb to the nearest genes, then analyzed the GO enrichment of

differentially accessible genes in FC and V1 regions of upper and lower layers, respectively. Interestingly, we found that many differentially accessible genes of FC and V1 were involved in actin filament-based processes (**Figure 3B and Figure S3B**). Actin-related cytoskeletal proteins provide structural support to the specialized dendritic spines of EX and play a key role in functional organization of neurotransmitter receptors (*24*), though the role of actin filaments processes involved in the epigenetic control of spatial organization of excitatory neurons has not been fully investigated. Particularly, actin-binding proteins tropomodulin (Tmod) family TMOD1 and TMOD2 exhibit distinct roles in regulating dendritic spine development (*25*); we found that cis-regulatory elements for these two genes are differentially enriched in FC (TMOD2) and V1 (TMOD1), respectively. We also observed that two actin nucleators, the Arp2/3 complex and the formin Diaph1 that both play crucial roles in corticogenesis (*26-28*), are differentially accessed in their promoter regions in FC (ARPC2, ARPC5 and ACTR3B for Arp2/3 complex) and V1 (DIAPH1), respectively. Our findings provide a direct link between the epigenetic regulation of actin filaments and spatial diversity of cortical excitatory neurons. In addition, given the involvement of the pathological lipid pathway in human primary visual cortex in Parkinson's Disease (*29*), we surveyed this aspect and observed enriched lipid transport and metabolic pathways in excitatory neuron of V1 (**Figure S3B**), which underscores the pathogenetic deregulation of visual cortical excitatory neuron in neurodegenerative diseases.

Thus, our findings underline previously unappreciated regional heterogeneity of EX in monkey cortex and the key processes underlying this, which provides new candidates for future research on specialized cortical functions.

Single-cell analysis resolves the area-specific transcriptional profiles of monkey cortex

To uncover the transcriptional regulatory program underlying chromatin accessibility in area-specific EX of FC and V1, we analyzed TF binding site enrichment of EXs and determined 104 TF and 55 TF motifs that were differentially enriched between FC and V1 in upper and lower layer, respectively (**Figure 3C**). For example, TBX family members binding motif were enriched in upper layer FC and motif for HOX family members in upper layer V1. Instead, motifs for ELK family members and E2F family members were enriched in lower-layer FC and V1, respectively. Motifs for the human developing V1-enriched TF motifs ZEB1 (*23*) were enriched in the upper layer V1. Likewise, we found enrichment for the motifs of ETV family members, which are known for specifying excitatory neurotransmitter fates (*30*) in upper and lower layer FC. Axonal CREB is critical for long-term memory (*31, 32*) and its motif appeared enriched in upper layer FC. We also found enrichment of motifs for the regionalization regulator EMX2 (*33*) in upper layer V1. These findings indicate that area-enriched TF have direct linkage with area-specialized functions.

In addition, we sought to investigate the transcriptomic heterogeneity of area-specific EX in FC and V1 by looking at DEG. We found 37 and 46 genes specifically enriched in EX from upper layer FC and V1, while 69 and 55 genes were specifically enriched in EX from lower layer FC and V1 (**Figure 3D and Table S5**) ($\log_2$(FC)>1, *P* value<0.01). We performed GO analysis for these DEG (**Figure S3C**). Notably, we noticed that a series of candidates identified as area-specific genes (NDST3, GPC5, GPC6, HS6ST2, HPSE2, SULF2 and LRRTM4) are involved in heparan-sulfate proteoglycan (HSPG) pathways. HSPG family plays an important role in regulating axon guidance, synaptic organization and synaptic specificity (*34-36*).

GPC1-6 belongs to the subfamily of HPSG, and *NDST*, *HS2ST*, *HS3ST*, *HS6ST* and *SULF* encode for enzymes involved in HPSG biosynthesis (*36*, *37*). Leucine-rich repeat transmembrane (*LRRTM*) gene family produce synaptic organizing proteins and binding partners of HPSG (*38*, *39*). These findings emphasize the role of the HPSG pathway in area-specific EX function. We then compared our epigenetic data with transcriptomic data and found that 22% to 55% of the DEG corresponded to differentially accessible cis-regulatory elements. Moreover, the differentially accessible cis-regulatory elements at genes of HPSG pathways (*NDST3*, *HS6ST2*, *HPSE2*, *SULF2* and *LRRTM4*) were all located in enhancer regions (beyond TSS ± 3 kb) and not promoter regions (**Figure 3E and 3F**, **Figure S4**).

In conclusion, our data illustrate the distinct transcriptional regulatory profiles between different cortical areas and suggest a role for HPSGs pathway in the inter-regional diversity of excitatory neurons.

Dynamic regulatory landscape of remyelination profiled by integrated single-cell analysis

OLI sheath neuronal axons with myelin in the central nervous system (CNS) to support neuronal function in of action potentials (*40*). Dysfunction of remyelination can be seen in several neurodegenerative diseases and neurodevelopmental disorders, including multiple sclerosis (*41*). Studies have been performed to uncover the transcriptional and epigenetic regulation pathways of OLI maturation and myelination in mouse models and human brain (*42*), although how the two layers of regulation are dynamically correlated could not be properly assessed due to insufficient data coverage. Given the wide use of macaque models in neuropathological studies including the demyelination-related diseases autoimmune encephalomyelitis and

multiple sclerosis (*43-45*), we first sought to investigate whether our data could recapture the dynamic signature of OPC differentiation and OLI development seen in those studies in human, and second whether we could explore the regulatory landscape of myelination (**Figure 4A**).

We performed pseudotime ordering of cells by Monocle 2 (see Methods) to construct the lineage trajectories of OPC differentiation and OLI maturation. Pseudotime analysis of gene expression and gene score were both able to faithfully recapitulate the differentiation of OPC to OLI (**Figure 4B and 4C**). We found that both the gene expression and gene score of marker genes reported in mouse and human OPC and mature OLI (*46-48*) were also enriched in our macaque OPC (*NEU4*, *PDGFRA* and *CSPG4*) and OLI (*MBP* and *PLP1*) (**Figure 4D and 4E**). Despite recent studies highlighting the importance for glutamatergic, GABAergic and calcium signaling, their roles in both development and function of OLI cell lineage still remains controversial due insufficient knowledge about tissue heterogeneity and receptor activation during the maturation process. Using our data, we were able to confirm the GO enrichment of glutamate receptors in human OPC and pre-mature OLI during lineage maturation (*48*). In agreement with the findings in human, we also found enriched gene expression and high epigenetic score of ionotropic glutamate receptors in immature OLI, including AMPA (α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid) receptors (*GRIA1*, *GRIA3* and *GRIA4*), and kainate receptors (*GRIK1*, *GRIK2* and *GRIK3*) in OPC and immature OLI and NMDA (N-methyl-D-aspartate) receptor (*GRIN2A* and *GRIN2B*) in immature OLI. Furthermore, we found enriched gene expression and high epigenetic score of GABA (gamma-aminobutyric acid) receptors (*GABRA1* and *GABRB1*) and voltage-gated calcium channels (*CACNA1A*, *CACNA1C*, *CACNA1D*,

CACNA1G, *CACNG2*, *CACNG3*, *CACNG4* and *CACNG7*) in OPC and immature OLI. It has been proposed that the Wnt signaling pathway plays an important role as a stage-specific multiple functional regulator of OLI development (*49*). Consistent with this, our data showed that genes involved in Wnt signaling are enriched in OPC and immature OLI (**Figure S5B**). GO terms for gene sets also supported stage-enriched pathways and dynamic changes of neuronal activity along the pseudotime trajectory (**Figure S5A and Table S6**).

We then characterized the TF dynamics across OPC differentiation and OLI development by mapping chromVAR TF deviation scores to cells along the pseudotime trajectory (**Figure 4F**). We found that accessibility in regions characterized by SOX9 and SOX10 binding motifs was higher in OPC; these TF are related to the survival and migration of OPC (*50*). The binding motifs for OLIG1 and OLIG2 were enriched in OPC and immature OLI; these TF have been shown to initiate OLI differentiation and mediate OLI development (*51-53*). Moreover, in agreement with our finding of enriched signaling pathways along OLI development, we found high accessibility at regions enriched for binding motifs of NRF1, SP family and EGR1, which have been reported as regulators of glutamate receptors for neuronal activity and energy production (*54-56*) in OPC. Of note, TCF7L2 is a major transducer of β-catenin activity following activation of Wnt signaling. We found transient enrichment of the binding motif for TCF7L2 in OPC, immature OLI and mature OLI, which supports the recent finding that stage-specific TCF7L2-regulated transcriptional circuitry regulates initiation and maintenance of OLI differentiation (*49*). When inspecting specific TF such as TGIF1 and the SOX family, we also observed a good correlation between expression, gene score and motif accessibility over the pseudotime (**Figure 4G**).

Taken together, our analyses delineate the dynamic changes of chromatin state along the OLI lineage maturation responsible for changes in gene expression and cell fate decisions.

Mapping human disease risk loci to macaque cortex cell types

Linking cell-type specific regulatory elements with disease risk variants of genome-wide association studies (GWAS) can help to identify cell types that may be related to the diseases. The latter could be instrumental in the development of therapeutic approaches. Single-cell epigenetic data of mouse and human brain have been used and proved very useful in determining the cell type enrichment of neuropsychiatric disease risks(*48, 57*). To demonstrate the robustness of our dataset and to provide a further link to diseases for NHP- and human-related brain research, we addressed the cell type-specific enrichments of human neurological and neuropsychiatric disease risk factors in macaque cortical cell types. We mapped the differentially accessible peaks of each cluster and all accessible peaks to orthologous coordinates in the human hg19 genome, then performed linkage disequilibrium score regression (LDSC) to measure the enrichment of SNP heritability for human brain disorders and neurobehavioral traits within differential accessible peaks for each of our 27 epigenetic clusters. We adopted the GWAS summary for neurological and psychiatric disorders and neurobehavioral traits from recent studies (**Table S7**)

We found a highly significant enrichment of heritability in neurons for neuropsychiatric traits such as major depressive disorders (MDD) and schizophrenia (SCZ) (**Figure 5A**), consistent with cell type mapping studies of neurological trait risk in human and mouse (*48, 57*). In agreement with previous reports, we found the strongest enrichment of Alzheimer's disease (AD) SNP-heritability in MIC (**Figure**

5A and 5B left panel), and also in EX, AST and OLI, which is in line with microglia activation in AD and early AD-associated transcriptional changes occurring in other brain cell types (*58, 59*). Consistent with the recently reported correlation of common-variant genetic findings for SCZ with specific brain cell types determined by single cells profiling (*60*), we also confirmed the significant enrichment in SNP-heritability of MDD in VIP type IN (**Figure 5A and 5B middle panel**), importantly, our analysis also demonstrated the enrichment of SCZ heritability in frontal cortical EX and IN (**Figure 5A and 5B right panel**).

The high resolution of our full-length snRNA-seq data allowed us to examine the alternatively spliced (AS) genes between cell types and their association with specific neurological diseases. By using the differential expression of splicing junctions (DESJ) detection (see Methods), we were able to detect several DESJ across all cell clusters (**Figure S6A and S6B**). We found cell type-specific exon usage in neurological disease-associated risk genes, such as *AKAP9* which was associated with AD in African Americans and *NCAM1* has been associated with neurocognition alterations in SCZ patients (*61, 62*) (**Figure S6C**). We also combined the DESJ analysis with the GWAS SNP to locate cell types whose abnormal splicing due to change of sequence could be important for disease (**Figure 5C**). Interestingly, we saw similar sub clusters being enriched between the DESJ analysis and the GWAS SNP analysis. We also observed some differences, such as a stark enrichment for oligodendrocytes in AD and SST type inhibitory neuron in MDD (**Figure 5D**).

Thus, our open chromatin and transcriptomic single-cell maps provide cell type-specific datasets to evaluate genomic loci implicated in human neurological disorders in specific macaque brain cell types.

Discussion

Developing therapeutic approaches for disorders of the central nervous system has been hampered by the lack of adequate models of the human disease. While murine models have been instrumental to study developmental trajectories and disease evolution, gene therapies in those models do not always recapitulate the responses seen in human, thus leading to only modest outcomes. This is further stressing the importance of using larger animal models that are phylogenetically close to humans, such as pigs and monkeys. In particular monkeys, despite evolutionary differences in cognitive functions and behaviors, have social complexity, brain structure and neuronal circuitries that are more closely related to humans. As such, monkeys have proved to be excellent models to faithfully recapitulate several features of many neurodegenerative diseases such as Parkinson's (*63*) and Huntington's (*64, 65*). Therefore, dissecting cellular composition and regulatory circuitries at single cell resolution as well as determining molecular signatures associated with specific diseases in species phylogenetically closer to human is paramount to develop efficient treatments.

In the present study, we have generated an integrated large-scale single-cell open chromatin and transcriptomic mapping of adult primate cortex. We applied sci-ATAC-seq to profile 43,793 single cells and the full-length snRNA-seq to profile 11,477 single cells from three major cortical regions, PFC, M1 and V1. Our high sensitivity single-cell chromatin accessibility data showed high power to resolve cortical cell types as well as distinct excitatory neuron subtypes between the FC (including PFC and M1) and V1. Generating multi-omics data and applying combined

analysis can increase the resolution and the confidence of cell type annotation compared to individual methods. To demonstrate the robustness of our data we have performed co-embedding of single-cell ATAC-seq and snRNA-seq data that confirms the high consistency of our cell type assignments and linked the cell type-specific expression profile to corresponding regulatory programs. From this, we have been able to define regulatory elements and pathways that have been previously suggested in brain cells and to gain novel insights in to cell type specific regulatory networks (*12*). The integration of the two layers shows that many signaling pathways are shared by the area-specific epigenetic and transcriptomic profiles of EX neurons, such as the ion transport and synaptic transmission pathways. This analysis also highlights some unique pathways, such as the actin filaments pathway which display different area-specific promoter accessibility in FC (such as Apr2/3) and V1 (such as formin DIAPH1). Thus, our combined analysis of epigenetic and transcriptomic profiling clearly illustrated synergistic power for understanding spatial organization and cellular heterogeneity in cerebral cortex.

Dysregulations in myelination of adult brain is crucial for neurodegenerative diseases. Our data resource allowed us to perform an integrative analysis on single cell transcriptomic and epigenetic profiling of OPC and OLIs, providing a parallel dynamic landscape of gene expression, open chromatin states and transcription factor enrichments related to myelination. By applying pseudotime analysis, we have been able to map how specific changes in gene expression, chromatin openness and regulatory circuitries influence cell fate decisions along the linage maturation from OPCs to OLIs, mark the potential targets for demyelination diseases.

A major advantage of generating single-cell genomics data in macaque is to provide direct data resources for pre-clinical studies on NHP models of

neuropsychiatric and neurological diseases. We have also used our data to predict the enrichment of neurological and neuropsychiatric disorders risk using the cell-type specific epigenetic data; we have used disease associated genes identified by GWAS to determine a significant enrichment in SNP-heritability for SCZ in frontal cortical excitatory and inhibitory neurons as well as for MDD in the VIP inhibitory neuron types(*60*). In addition, we analyzed the colocalization of SNP from complex traits with cell type-specific exon use. The results implicate the correlation between genetically driven alternative splicing and neurological and neuropsychiatric diseases (such as AD and SCZ), which may predict the potentially novel loci in disease susceptibility. Notably, mapping the disease risk loci with epigenetic and alternative splicing data can elucidate different aspects of the disease susceptibility that recapitulate in different cell type.

Future work on brain tissues from cortical and sub-cortical areas and increased time-points will allow the integrative analysis of single cell transcriptomic, epigenetic, proteomics and spatial genomics to uncover the systemic molecular mechanisms of primate brain in health and disease conditions. Our dataset is the first at this high resolution and it provides valuable insights into the regulatory elements and expression profiles that shape the primate cortical cells organization, as well as their relevance to human disease.

Acknowledgments: We sincerely thank the support provided by China National GeneBank. We are grateful to M. Jan for helpful feedback on the manuscript. This work is supported by National Natural Science Foundation of China (No. 31900582), National Natural Science Foundation of China (No. 31900466), the Strategic Priority Research Program of the Chinese Academy of Sciences (Grant No. XDA16010114) ,

the National Key Research and Development Program (2016YFA0101401), Natural Science Foundation of Guangdong Province (No. 2018A030313379), Guangdong Provincial Key Laboratory of Genome Read and Write (No. 2017B03030101), Guangdong Province Science and Technology Program (2014A030312001), Yunnan Fundamental Research Projects (2018FA020), Shenzhen Key Laboratory of Single-Cell Omics (ZDSYS20190902093613831), Shenzhen Bay Laboratory (SZBL2019062801012), and Guangzhou Science and Technology Program (201807010066). G.V. is supported by Chinese Academy of Sciences President's International Fellowship for Foreign Experts (2020FSB0002). C.W. is supported by Chinese Academy of Sciences President's International Fellowship Initiative for Postdoctoral Researchers (2019PB0177) and Research Fund for International Young Scientists grant (31950410553). S. L., Y. H. and Y. N. conceived and supervised the study. H. Z., L, L. and Y. L. designed the experiments; Samples were collected by J. W., L. X., C. W. and W. L; Single nuclear suspension were prepared by Y. W.; Flow cytometry and FACS experiments were performed by M. C and Z. W.; Libraries were prepared by M. C., J. X., Y. Y., M. W., J. X., Y. W., C. L., C. W., D. C. and L. W.; L. X. preformed quality control of the libraries. L. W., Z. L. and L. H. performed the computational analysis of sci-ATAC-seq data; Z. Z., Z. H. and S. L. performed the computational analysis of single nucleus RNA-seq data. H. Y., H. S., and Q. Y. built the data website; Y. L., S. L., L. L., H. Z., L. H., B. T., H. Y., X. X., W. J., F. L., Z. S., G. L. ,T. T., K. C., M. D., A. C, Y. G., K. M. and L. C. interpreted the data. Y. L., MA. E., M. C, Z. L, Z. Z., L. W. and L. H. wrote the manuscript with input from all authors, G. V., C. W. and Y. L. revised the manuscript and provided critical review. All sequencing data will be available from CNGB Nucleotide Sequence Archive (CNSA: https://db.cngb.org/cnsa).

Reference and notes

1. A. Bernard *et al.*, Transcriptional architecture of the primate neocortex. *Neuron* **73**, 1083-1099 (2012).
2. Y. Zhu *et al.*, Spatiotemporal transcriptomic divergence across human and macaque brain development. *Science* **362**, (2018).
3. K. D. Hunt, The single species hypothesis: truly dead and pushing up bushes, or still twitching and ripe for resuscitation? *Hum Biol* **75**, 485-502 (2003).
4. S. Preissl *et al.*, Single-nucleus analysis of accessible chromatin in developing mouse forebrain reveals cell-type-specific transcriptional regulation. *Nat Neurosci* **21**, 432-439 (2018).
5. S. Yin *et al.*, Transcriptomic and open chromatin atlas of high-resolution anatomical regions in the rhesus macaque brain. *Nat Commun* **11**, 474 (2020).
6. Y. Kang, C. Chu, F. Wang, Y. Niu, CRISPR/Cas9-mediated genome editing in nonhuman primates. *Dis Model Mech* **12**, (2019).
7. L. Chansel-Debordeaux, E. Bezard, Local transgene expression and whole-body transgenesis to model brain diseases in nonhuman primate. *Animal Model Exp Med* **2**, 9-17 (2019).
8. Y. Chen *et al.*, Modeling Rett Syndrome Using TALEN-Edited MECP2 Mutant Cynomolgus Monkeys. *Cell* **169**, 945-955 e910 (2017).
9. W. Zhang *et al.*, SIRT6 deficiency results in developmental retardation in cynomolgus monkeys. *Nature* **560**, 661-665 (2018).
10. J. B. Koprich, T. H. Johnston, G. Reyes, V. Omana, J. M. Brotchie, Towards a Non-Human Primate Model of Alpha-Synucleinopathy for Development of Therapeutics for Parkinson's Disease: Optimization of AAV1/2 Delivery Parameters to Drive Sustained Expression of Alpha Synuclein and Dopaminergic Degeneration in Macaque. *PLoS One* **11**, e0167235 (2016).
11. J. W. Y. Roman Spektor, Seoyeon Lee, Paul D. Soloway, Single cell ATAC-seq identifies broad changes in neuronal abundance and chromatin accessibility in Down Syndrome. *https://www.biorxiv.org/content/10.1101/561191v1*, (2019).
12. A. Nott *et al.*, Brain cell type-specific enhancer-promoter interactome maps and disease-risk association. *Science* **366**, 1134-1139 (2019).
13. S. W. Flavell *et al.*, Activity-dependent regulation of MEF2 transcription factors suppresses excitatory synapse number. *Science* **311**, 1008-1012 (2006).
14. Y. Zhang *et al.*, Rapid single-step induction of functional neurons from human pluripotent stem cells. *Neuron* **78**, 785-798 (2013).
15. T. Baroti *et al.*, Transcription factors Sox5 and Sox6 exert direct and indirect influences on oligodendroglial migration in spinal cord and forebrain. *Glia* **64**, 122-138 (2016).
16. A. M. Smith *et al.*, The transcription factor PU.1 is critical for viability and function of human brain microglia. *Glia* **61**, 929-942 (2013).
17. Y. Zhu *et al.*, Inactivation of NF1 in CNS causes increased glial progenitor proliferation and optic glioma formation. *Development* **132**, 5577-5588 (2005).
18. L. Subramanian *et al.*, Transcription factor Lhx2 is necessary and sufficient to suppress astrogliogenesis and promote neurogenesis in the developing hippocampus. *Proc Natl Acad*

Sci U S A **108**, E265-274 (2011).

19. Y. Niu *et al.*, Dissecting primate early post-implantation development using long-term in vitro embryo culture. *Science* **366**, (2019).
20. T. Stuart *et al.*, Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888-1902 e1821 (2019).
21. B. Tasic *et al.*, Shared and distinct transcriptomic cell types across neocortical areas. *Nature* **563**, 72-78 (2018).
22. T. J. Nowakowski *et al.*, Spatiotemporal gene expression trajectories reveal developmental hierarchies of the human cortex. *Science* **358**, 1318-1323 (2017).
23. C. N. K. Ryan S. Ziffra, Amy Wilfert, Maximilian Haeussler, Alex M. Casella, Pawel F. Przytycki, Anat Kreimer, Katherine S. Pollard, Seth A. Ament, Evan E. Eichler, Nadav Ahituv, Tomasz J. Nowakowski, Single cell epigenomic atlas of the developing human brain and organoids. *https://www.biorxiv.org/content/10.1101/2019.12.30.891549v1*, (2019).
24. H. Stefen, C. Chaichim, J. Power, T. Fath, Regulation of the Postsynaptic Compartment of Excitatory Synapses by the Actin Cytoskeleton in Health and Its Disruption in Disease. *Neural Plast* **2016**, 2371970 (2016).
25. O. F. Omotade *et al.*, Tropomodulin Isoform-Specific Regulation of Dendrite Development and Synapse Formation. *J Neurosci* **38**, 10271-10285 (2018).
26. P. S. Wang *et al.*, Crucial roles of the Arp2/3 complex during mammalian corticogenesis. *Development* **143**, 2741-2752 (2016).
27. R. Shinohara *et al.*, A role for mDia, a Rho-regulated actin nucleator, in tangential migration of interneuron precursors. *Nat Neurosci* **15**, 373-380, S371-372 (2012).
28. M. Fritzsche, C. Erlenkamper, E. Moeendarbary, G. Charras, K. Kruse, Actin kinetics shapes cortical network structure and mechanics. *Sci Adv* **2**, e1501337 (2016).
29. D. Cheng *et al.*, Lipid pathway alterations in Parkinson's disease primary visual cortex. *PLoS One* **6**, e17299 (2011).
30. J. L. Juarez-Morales *et al.*, Evx1 and Evx2 specify excitatory neurotransmitter fates and suppress inhibitory fates through a Pax2-independent mechanism. *Neural Dev* **11**, 5 (2016).
31. T. Miyashita, E. Kikuchi, J. Horiuchi, M. Saitoe, Long-Term Memory Engram Cells Are Established by c-Fos/CREB Transcriptional Cycling. *Cell Rep* **25**, 2716-2728 e2713 (2018).
32. M. R. Matos *et al.*, Memory strength gates the involvement of a CREB-dependent cortical fear engram in remote memory. *Nat Commun* **10**, 2315 (2019).
33. A. Zembrzycki *et al.*, Genetic mechanisms control the linear scaling between related cortical primary and higher order sensory areas. *Elife* **4**, (2015).
34. F. E. Poulain, Analyzing the role of heparan sulfate proteoglycans in axon guidance in vivo in zebrafish. *Methods Mol Biol* **1229**, 469-482 (2015).
35. M. I. Lazaro-Pena, C. A. Diaz-Balzac, H. E. Bulow, S. W. Emmons, Synaptogenesis Is Modulated by Heparan Sulfate in Caenorhabditis elegans. *Genetics* **209**, 195-208 (2018).
36. G. Condomitti, J. de Wit, Heparan Sulfate Proteoglycans as Emerging Players in Synaptic Specificity. *Front Mol Neurosci* **11**, 14 (2018).
37. F. E. Poulain, H. J. Yost, Heparan sulfate proteoglycans: a sugar code for vertebrate development? *Development* **142**, 3456-3467 (2015).
38. R. T. Roppongi, B. Karimi, T. J. Siddiqui, Role of LRRTMs in synapse development and plasticity. *Neurosci Res* **116**, 18-28 (2017).

39. P. Zhang *et al.*, Heparan Sulfate Organizes Neuronal Synapses through Neurexin Partnerships. *Cell* **174**, 1450-1464 e1423 (2018).
40. M. Simons, K. A. Nave, Oligodendrocytes: Myelination and Axonal Support. *Cold Spring Harb Perspect Biol* **8**, a020479 (2015).
41. J. M. Williamson, D. A. Lyons, Myelin Dynamics Throughout Life: An Ever-Changing Landscape? *Front Cell Neurosci* **12**, 424 (2018).
42. B. Emery, Q. R. Lu, Transcriptional and Epigenetic Regulation of Oligodendrocyte Development and Myelination in the Central Nervous System. *Cold Spring Harb Perspect Biol* **7**, a020461 (2015).
43. Z. Peng *et al.*, Experimental Autoimmune Encephalomyelitis (EAE) Model of Cynomolgus Macaques Induced by Recombinant Human MOG1-125 (rhMOG1-125) Protein and MOG34-56 Peptide. *Protein Pept Lett* **24**, 1166-1178 (2018).
44. K. G. Haanstra *et al.*, Induction of experimental autoimmune encephalomyelitis with recombinant human myelin oligodendrocyte glycoprotein in incomplete Freund's adjuvant in three non-human primate species. *J Neuroimmune Pharmacol* **8**, 1251-1264 (2013).
45. H. I. McFarland *et al.*, Determinant spreading associated with demyelination in a nonhuman primate model of multiple sclerosis. *J Immunol* **162**, 2384-2390 (1999).
46. S. Marques *et al.*, Oligodendrocyte heterogeneity in the mouse juvenile and adult central nervous system. *Science* **352**, 1326-1329 (2016).
47. S. Jakel *et al.*, Altered human oligodendrocyte heterogeneity in multiple sclerosis. *Nature* **566**, 543-547 (2019).
48. B. B. Lake *et al.*, Integrative single-cell analysis of transcriptional and epigenetic states in the human adult brain. *Nat Biotechnol* **36**, 70-80 (2018).
49. Z. M. Dai *et al.*, Stage-specific regulation of oligodendrocyte development by Wnt/beta-catenin signaling. *J Neurosci* **34**, 8467-8473 (2014).
50. M. Finzsch, C. C. Stolt, P. Lommes, M. Wegner, Sox9 and Sox10 influence survival and migration of oligodendrocyte precursors in the spinal cord by regulating PDGF receptor alpha expression. *Development* **135**, 637-646 (2008).
51. J. Dai, K. K. Bercury, W. Jin, W. B. Macklin, Olig1 Acetylation and Nuclear Export Mediate Oligodendrocyte Development. *J Neurosci* **35**, 15875-15893 (2015).
52. A. Wegener *et al.*, Gain of Olig2 function in oligodendrocyte progenitors promotes remyelination. *Brain* **138**, 120-135 (2015).
53. Y. Yu *et al.*, Olig2 targets chromatin remodelers to enhancers to initiate oligodendrocyte differentiation. *Cell* **152**, 248-261 (2013).
54. S. S. Dhar, H. L. Liang, M. T. Wong-Riley, Nuclear respiratory factor 1 co-regulates AMPA glutamate receptor subunit 2 and cytochrome c oxidase: tight coupling of glutamatergic transmission and energy metabolism in neurons. *J Neurochem* **108**, 1595-1606 (2009).
55. A. Priya, K. Johar, B. Nair, M. T. Wong-Riley, Specificity protein 4 (Sp4) regulates the transcription of AMPA receptor subunit GluA2 (Gria2). *Biochim Biophys Acta* **1843**, 1196-1206 (2014).
56. X. Qin *et al.*, Early Growth Response 1 (Egr-1) Regulates N-Methyl-d-aspartate Receptor (NMDAR)-dependent Transcription of PSD-95 and alpha-Amino-3-hydroxy-5-methyl-4-isoxazole Propionic Acid Receptor (AMPAR) Trafficking in Hippocampal Primary Neurons. *J Biol Chem* **290**, 29603-29616 (2015).

57. D. A. Cusanovich *et al.*, A Single-Cell Atlas of In Vivo Mammalian Chromatin Accessibility. *Cell* **174**, 1309-1324 e1318 (2018).
58. H. Mathys *et al.*, Single-cell transcriptomic analysis of Alzheimer's disease. *Nature* **570**, 332-337 (2019).
59. A. Grubman *et al.*, A single-cell atlas of entorhinal cortex from individuals with Alzheimer's disease reveals cell-type-specific gene expression regulation. *Nat Neurosci* **22**, 2087-2097 (2019).
60. N. G. Skene *et al.*, Genetic identification of brain cell types underlying schizophrenia. *Nat Genet* **50**, 825-833 (2018).
61. T. Ikezu *et al.*, Tau Phosphorylation is Impacted by Rare AKAP9 Mutations Associated with Alzheimer Disease in African Americans. *J Neuroimmune Pharmacol* **13**, 254-264 (2018).
62. P. F. Sullivan *et al.*, NCAM1 and neurocognition in schizophrenia. *Biol Psychiatry* **61**, 902-910 (2007).
63. A. Eslamboli *et al.*, Long-term consequences of human alpha-synuclein overexpression in the primate ventral midbrain. *Brain* **130**, 799-815 (2007).
64. L. H. Burns *et al.*, Selective putaminal excitotoxic lesions in non-human primates model the movement disorder of Huntington disease. *Neuroscience* **64**, 1007-1017 (1995).
65. R. J. Ferrante, N. W. Kowall, P. B. Cipolloni, E. Storey, M. F. Beal, Excitotoxin lesions in primates as a model for Huntington's disease: histopathologic and neurochemical characterization. *Exp Neurol* **119**, 46-71 (1993).
66. S. R. Krishnaswami *et al.*, Using single nuclei for RNA-seq to capture the transcriptome of postmortem neurons. *Nat Protoc* **11**, 499-524 (2016).
67. D. A. Cusanovich *et al.*, Multiplex single cell profiling of chromatin accessibility by combinatorial cellular indexing. *Science* **348**, 910-914 (2015).
68. S. Amini *et al.*, Haplotype-resolved whole-genome sequencing by contiguity-preserving transposition and combinatorial indexing. *Nat Genet* **46**, 1343-1349 (2014).
69. D. A. Cusanovich *et al.*, The cis-regulatory dynamics of embryonic development at single-cell resolution. *Nature* **555**, 538-542 (2018).
70. M. Martin, Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet* **17**, 10–12 (2011).
71. B. Langmead, C. Trapnell, M. Pop, S. L. Salzberg, Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**, R25 (2009).
72. C. Bravo Gonzalez-Blas *et al.*, cisTopic: cis-regulatory topic modeling on single-cell ATAC-seq data. *Nat Methods* **16**, 397-400 (2019).
73. J. H. Levine *et al.*, Data-Driven Phenotypic Dissection of AML Reveals Progenitor-like Cells that Correlate with Prognosis. *Cell* **162**, 184-197 (2015).
74. A. Dobin *et al.*, STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21 (2013).
75. B. Li, C. N. Dewey, RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323 (2011).
76. A. Butler, P. Hoffman, P. Smibert, E. Papalexi, R. Satija, Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol* **36**, 411-420 (2018).
77. T. Hayashi, R. L. Huganir, Tyrosine phosphorylation and regulation of the AMPA receptor by SRC family tyrosine kinases. *J Neurosci* **24**, 6152-6160 (2004).
78. A. N. Schep, B. Wu, J. D. Buenrostro, W. J. Greenleaf, chromVAR: inferring

transcription-factor-associated accessibility from single-cell epigenomic data. *Nat Methods* **14**, 975-978 (2017).

79. S. Heinz *et al.*, Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* **38**, 576-589 (2010).

80. X. Qiu *et al.*, Single-cell mRNA quantification and differential analysis with Census. *Nat Methods* **14**, 309-315 (2017).

Figure Legends

Main figures

Fig. 1. Single-cell chromatin accessibility of macaque cortex. (A) Schematic workflow of single-nucleus isolation from the visual cortex and frontal cortex for sci-ATAC-seq and snRNA-seq, sample processing, library generation and downstream analysis. **(B)** UMAP visualization of sci-ATAC data showing all clusters colored by cell types. A total of 43793 nuclei passed QC and were used for classifying 27 neuronal and non-neuronal cell types. The annotation of each cell type is indicated in the figure. **(C)** Integrative genomics viewer (IGV) plots showing color-coded aggregate read density for cells within each cell cluster at neuronal and non-neuronal cell-type-specific marker genes. The promoter region for cell type-specific marker gene and the corresponding gene names are highlighted. **(D)** (left) Heatmap showing the top 20 transcription factor (TF) binding motifs enriched at differential accessible peaks of all neuronal and non-neuronal cell types (indicated in 1B) by HOMER (in

total 99 TF motif); (right) UMAP of motif enrichment in the selected filter.

Fig. 2. Integration of single-nucleus transcriptome and chromatin accessibility. **(A)** UMAP projection of 17 clusters from the single-nucleus transcriptomic data. A total of 11,477 nuclei passed QC and were used for clustering macaque cortices into neuronal and non-neuronal cell types. Each cluster is color coded and associated with a specific cell type. **(B)** Violin plots of expression values for neuronal and non-neuronal cell type-specific marker genes. **(C)** Co-embedding of snRNA-seq (highlighted in blue) and sci-ATAC-seq (highlighted in red) clusters across PFC, M1 and V1 brain regions using Seurat V3. Annotation of each cluster is provided in the plot; the circles in the UMAP indicate IN and EX neurons. **(D)** UMAP projection of snRNA-seq and sci-ATAC-seq from (C). Cell clusters in each UMAP are color-coded and annotated according to cell identity. **(E)** SnRNA-seq expression levels and sci-ATAC-seq promoter enrichment score (calculated TSS ± 3 kb) for cell type-specific marker genes visualized by UMAP.

Fig. 3. Inter-regional diversity of excitatory neurons. (A) Stacked bar plot indicating the proportion of cells from PFC, M1 and V1 in each epigenetic cell cluster (left) and cell distribution of three brain regions are overlaid in UMAP plot of sci-ATAC-seq (right). **(B)** Volcano plot showing differential accessible peaks (beta value >1) between FC (including PFC and M1) and V1 excitatory neurons in the upper-layer and lower-layer of the cortex, respectively. Peaks within TSS±3 kb were annotated to the nearest genes by ChIPseeker. The genes enriched in actin filaments

signaling pathways were labeled to the corresponding peaks. **(C)** Row-normalized ChromVAR TF motif enrichment in FC (PFC and M1) and V1 EX neurons in upper (left) and lower layers (right). **(D)** Differentially expressed genes between FC (PFC and M1) and V1 excitatory neurons in upper-layer and low-layer of the cortex, respectively. The gene enriched in HPSGs pathway were labeled with blue or red colors. **(E)** Integrative genomics viewer plot showing single-cell chromatin accessibility at the *NDST3* locus in excitatory neurons. Differentially accessible peaks between FC (PFC and M1) and V1 in upper and lower layer are highlighted. **(F)** UMAP visualization highlighting the expression and gene score of *NDST3*.

Fig. 4. Integrated single cell regulatory landscape of myelination. (A) Schematic illustration for oligodendrocyte maturation. **(B)** Monocle 2 pseudotime trajectory of OPC and OLI indicating the gene expression obtained from snRNA-seq data. **(C)** Monocle 2 pseudotime trajectory of OPC and OLI indicating the gene score obtained from sci-ATAC-seq data **(D)** Heatmap showing selected genes expression dynamics across myelination pseudotime trajectory indicated in (B). Genes of glutamate receptors, gamma-aminobutyric acid (GABA) receptors and voltage-gated calcium channel (VGCCs) were blue, red and green color-coded, respectively. **(E)** Heatmap showing accessibility dynamics (gene score) across remyelination pseudotime trajectory in (C). Genes of glutamate receptors, gamma-aminobutyric acid (GABA) receptors and voltage-gated calcium channel (VGCCs) were blue, red and green color coded, respectively. **(F)** Heatmap ordering of average TF binding motif bias-corrected

deviations for 342 most variable TFs as determined by ChromVAR. Cells are binned by pseudotime into equally sized bins. **(G)** Gene expression, gene score and motif enrichment of selected TF across pseudotime trajectory.

Fig. 5. Mapping of disease genes to specific brain cell types. (A) Heatmap showing the enrichments of specific cell types in major 10 brain diseases among the 27 cell clusters (columns) identified from sci-ATAC-seq data. AD, Alzheimer's disease; ADHD, attention deficit hyperactivity disorder; ASD, autism spectrum disorder; BIP, bipolar disorder; ALS, amyotrophic lateral sclerosis; MDD, major depression disorders; SCZ, schizophrenia. **(B)** Enrichment of GWAS SNPs of AD, MDD and SCZ in the open chromatin of populations were overlaid on the epigenetic cell clusters. **(C)** Heatmap showing colocalization of GWAS SNPs of 10 brain diseases with cell type-specific alternative splicing sites. **(D)** Enrichment of GWAS SNPs of AD, MDD and SCZ in the cell type-specific alternative splicing sites were overlaid on the transcriptomic cell clusters.

Materials and Methods

Collection of monkey cerebral cortex tissue

All tissues were extracted from a female 72-month-old cynomolgus monkey (*Macaca fasicularis*). All relevant procedures involving animals were approved by in advance by the Institutional Animal Care and Use Committee of Yunnan Key Laboratory of Primate Biomedical Research.

Nuclei isolation

Single-nucleus preparations were performed as previous description(*66*), frozen monkey brain tissue pieces were placed in 1 ml homogenization buffer (pre-chilled) in 1 ml Dounce homogenizer (TIANDZ). Tissue was homogenized by 10 strokes of the loose pestle and 10 strokes of the tight pestle, Dounce homogenizer was submerged in ice during grinding. 2 ml homogenization buffer was added to the Dounce homogenizer then the homogenate was filtered through 40 μm cell strainer (Miltenyi Biotech) into 15 ml conical tube and centrifuged at 900 g for 10 mins to pellet nuclei.

Sci-ATAC-seq library preparation and sequencing

Sci-ATAC-seq was performed as described previously with modifications (*67*). Nuclei were strained in 40 μm strainer and centrifuged for 5 min at 500 g. The nuclei were resuspended in cold PBS (1% BSA) and counted using hemocytometer. Nuclei were adjusted concentration to 360/μL. For transposition, added 7 μL cell suspension

(around 2500 cell), 2 μL 5xTAG buffer and 1 μL unique barcoded Tn5 transposome into each well of 96-well plate, mixed gently and had a short spin (*68*). The plate was incubated at 55□ for 60 min with shaking (300 rpm). To quench the reaction, 10 μL of 40 mM EDTA was added to each well and gently mixed, then the plate was incubated at room temperature for 5 min. After reaction, 5 μL sorting buffer (5% BSA; 5 mM EDTA) was applied to each well, mixed well and pooled into one tube. The suspension was filtered through 40 μm strainer. Then one drop of DAPI (4′, 6-diamidino- 2-phenylindole, ThermoFisher Scientific) was added to the suspension and 25 nuclei were sorted by Aria II (BD) into 96-well plate containing 7 μL buffer EB, shortly spun down. Next, 1 μL 10% SDS was added to each well, mixed well and incubated at 55□ for 7 min with shaking (500 rpm) to lyse the nuclei. After the reaction, 1 μL 10% Triton-X was added to each well, spun down and incubated at room temperature for 5 min. For amplification, we added 1 μL unique barcoded N5&N7 (0.5 μM final concentration) and 10 μL NEBNext High-Fidelity 2x PCR Master Mix (NEB). PCR cycling conditions were as follow: 72 °C 5 min, 98 °C 30 s, (98 °C 10 s, 63 °C 30 s, 72 °C 30 s) × 11, held at 4 °C. After that, we pooled each 96-well plate into tubes, and added 5 volume PB including pH-indicator (Qiagen) and 200 μL sodium-acetate (3M, pH = 5.2), and reversed blending. Next, we used 4 columns to purify product following the MinElute PCR Purification Kit manual (Qiagen). DNA from each column was eluate using 25 μL EB buffer, pooled the 4 elution, and added EB buffer to 100 μL. To filtrate the fragment, we used the Ampure XP Bead by 0.5x & 0.7x. First added 50 μL XP beads, after incubation, collected

supernatant and then added 70 μL XP beads. Finally, we used 100 μL EB buffer to elute the DNA. We quantified the libraries by Qubit fluorimeter (Life technologies) and detected fragment size using 2100 High Sensitivity (Agilent). To sequence, each library we used 330 ng for cyclizing, and then used 8 ng to make DNB. Each library was loaded into 2 lanes using BGISEQ500. The sequencing primer used as: Tn5 primer1, Tn5 primer2, SCIMDA primer. The read lengths: PE 100, including 4 indexes. Index 1 and index 2 was represented Tn5 barcode, index 3 and index 4 represented the PCR barcode. There are common bases between Tn5 barcode and PCR index, so we used cold reaction of BGISEQ500 to get out of the imbalance.

Sci-ATAC-seq data processing and quality control

For sci-ATAC-seq data processing, we used the modified in-home pipeline of previously study (*69*). Briefly, to perform errors correction in cell barcodes, we first extracted the segments of barcode from the 4 constituent parts (read 1: 1-10, 32-41; read 2: 1-10, 38-47) and calculated the levenshtein distance for each piece from all possible barcode addresses. For each component, the segment has one and only one was within 3 mismatch bases in the segments list, the segment will be corrected to its best match. And only the reads with all segments of barcode can be corrected will be retained in downstream analysis. We next trimmed reads with cutadapt (*70*), and then mapped the reads to the reference genome downloaded from NCBI using bowtie2 (*71*) with '-X 2000 --mm --local' as options. Reads with mapping quality less than 10 and reads mapped to the mitochondria or genome scaffold (chrAQ*, chrU*,

chr*_random*, and chrK*) were filtered out. Then, PCR duplicates for all reads were removed according to the cell barcode by custom python script. Next, we merged the peaks of the libraries measured by macs2 as the reference peaks, and counted the reads number of each cell barcode, and only the cell barcode with more than 10,000 reads and 30% reads in the peaks will be used in downstream analysis. To calculate the gene activity score, we assignment the peak to the nearest the gene within 25 kb around the peak, then aggregate the reads of each peak assignment to the same gene as the gene activity scores.

sci-ATAC-seq data clustering and analysis

We applied cisTopic (*72*) to identify the sub-populations within all data. We run several models ranging from 10 to 200 topics and 120 burn-in iterations and 150 recording iterations for evaluating the optimal number of topics. We used 180 topics and default setting of cisTopic to determined topic associated regions; and used runUmap function to project the cells into two-dimensional space. We found the cell sub-populations on the topics-cells matrix by Rphenograph (*73*) package.

Nuclei staining and sorting for snRNA-seq

Nuclei were resuspended in 500 μL blocking buffer containing 1X PBS (GIBCO), 2% filtered sterilized BSA (SIGMA), and 0.2 U/μL RNasin Plus by pipetting up and down gently on ice. Nuclei were transferred to 1.5 ml tube then split up into three tubes: experimental, control-isotype and control-DAPI. For single-nuclei isolation by flow

cytometry, 250 μL DAPI (0.1 μg/ml) was added into the experimental tube to make the final volume as 500 μl, while nothing added to the negative control tube. We also sorted $NeuN^+$ nuclei from M1 region. For this purpose, rabbit anti-NeuN (Abcam, final dilution of 1:500) was added to experimental tube, rabbit IgG-monoclonal-isotype control (Abcam, final dilution of 1:500) was added to control-isotype tube. After 30 min incubation at 4 °C, samples were then centrifuged for 5 min at 400 g to pellet nuclei and pellets were resuspended in 250 μL blocking buffer (including 50 μL leaving buffer). 250 μL DAPI (0.1 μg/ml) was added to nuclei to make the final volume for each tube as 500 μL. Stained nuclei were filtered through 40 μm filter before FACS (BD FACSAria II instrument). $DAPI^+$ or $DAPI^+$ $NeuN^+$ nuclei in the experimental tube were sorted at a speed of 5-7 and the single nucleus was sorted into every single well of 384 plate filled with 1.2 μL lysis buffer (10% Triton X-100 0.0125 μl, 40 U/μl RNase Inhibitor 0.0625 μl, 10 M Oligo-dT Primer 0.02 μL, 10 mM dNTP Mix 0.45 μL, 5× SuperScriptII First-Strand Buffer 0.25 μL, nuclease-free water 0.305 μL and ERCC spike-in 0.1 μL) in advance. Sorted nuclei in 384-well plates were briefly centrifuged and stored at -80 °C for further analysis.

snRNA-Seq library preparation and sequencing

Sorted nuclei transcriptome amplifications were prepared by a modified SMART-seq2 protocol (*19*). After nuclei lysis, 1.2 μL reverse transcription mixed solution was added into each well to complete the reverse transcription reactions. Then 1.8 μL PCR reaction buffer was added into every well to complete amplification. At last, the

amplified cDNA products of each single nucleus were quantified by Agilent Bioanalyzer 2100. For those single nucleus samples with high quality after amplification, the products were extracted by an automatic extractor from the 384 plate to 96-well plate then purified by MGIEASY DNA Purification Magnetic Bead Kit (MGI) for the library construction. The libraries were prepared by MGIEASY RNA Library Preparation Kit (MGI) and each single nucleus sample was barcoded. Finally, the libraries were cyclized into ssDNA libraries by MGIEASY Cyclization Kit (MGI). All the single nucleus samples were sequenced on the BGISEQ500 sequencer with 100-bp pair-end reads.

snRNA-seq data processing and quality control

Single-cell RNA-seq data were processed as previous described (*19*), first, one-hundred-base-pair paired-end reads were pre-processed to remove adapters and filter out reads with low quality using default parameter by Cutadapt (v1.15) (*70*). Next, filtered reads were aligned to the Macaca fascicularis genome (5.0.91) by STAR (v2.5.3) (*74*) using a modified GTF file from ensemble release-91. In order to map the pre-mRNA fragments which may cover both exonic and intronic regions, we create a modified GTF annotation file which only contain transcript regions, and the original annotation rows for exons were all deleted, then we replace the feature type name from ‘transcript’ to ‘exon’. Finally, transcripts per million mapped reads (TPM) were calculated using rsem-calculate-expression (RSEM) (*75*) with default parameters. Cells with unique mapped ratios > 15% and more than 500 and less than 9000 genes

with transcripts per million mapped reads value >1 were used for downstream analysis.

snRNA-seq data clustering and analysis

Clustering and differential gene expression analysis was performed by Seurat (v2.3.4) R toolkit (*76*). NormalizeData, FindVariableGenes and ScaleData were performed respectively in each regions of brain sample. Genes expressed in less than three cells were filtered out and cells with expressed genes less than 500 were excluded. Canonical correlation analysis (CCA) was performed to integrate cells from 4 regions using the union set of variable genes in 5 batches (NeuN+ sorted nuclei from M1 was treated as one additional batch). We select the top 19 dimensions (cc.use = 1-19) for SNN network construction. Then, cell clusters were identified by graph-based clustering approach, louvain algorithm. We used uniform manifold approximation and projection (UMAP) to visualize the distance between cells in 2D space. Briefly, the value of k. param is varied by input cell numbers. The edges between two cells were measured by Jaccard distance to construct SNN graph. Then graph-based clustering was performed (times = 50) and modularity was calculated until reaching the maximum. “FindAllMarkers” and “FindMarkers” functions were used for differential gene expression analysis between clusters.

Evaluation of differential junction reads (AS)

To assess the alternative splicing heterogeneity among brain cell systematically, we

performed DESJ-detection (https://github.com/liushang17/DESJ-detection) to detect the differential expression of splicing junctions (DESJ) between different functional cell type. First, we retained and annotated the splicing junctions that were detected in at least ten cells with an expression threshold of read count > 4. Then, we obtained the junction cell read count matrix and normalized it with uniquely mapped reads per cell. Finally, DESJ-detection was applied to detect the DESJ using the normalized matrix, the cell clusters information and junction annotation information with default parameters (mincell = 20, maxsd = 0.2, maxmean = 1). For clustering assay of junction reads coverage, average normalized feature count values of each cluster were applied, and data were row normalized. Junctions with NA values (no reads mapped to the genes) were excluded.

Co-embedding of snRNA-seq cells with sci-ATAC-seq cells

We performed co-clustering of sci-ATAC-seq data and snRNA-seq data with R package Seurat (v3.1) (*20, 78*) . Firstly, we converted the peak by cells matrix to the promoter regions (TSS ± 3 kb). For each of the major cell types (EX, IN, AST, MIC, OLI, OPC, ENDO and PERI), same number of cells were selected from both sci-ATAC-seq and RNA-seq data. Then, we used RunLSI function with a default dimension of 50 to reduce the dimension. The function of LSI is computed by the term frequency-inverse document frequency (TF-IDF) transformation. We had previously pre-processed and clustered a snRNA-seq dataset using default parameters without clustering cells by FindClusters function with a resolution of 0.6. Next, we

computed anchors between the sci-ATAC-seq dataset and the snRNA-seq dataset by FindTransferAnchors function; and used these anchors to transfer the cell-type labels from snRNA-seq data. Then we used predictions and confidence scores for each sci-ATAC-seq cell correspond to snRNA-seq cell to transfer the cluster IDs. Finally, we created snRNA-seq matrix by transfer cell type labels anchors with TransferData function. We then merged measured and imputed snRNA-seq data and run a standard UMAP analysis by reduce dimensionality to 50 to visualize all the cells together.

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

We determined TF binding motif enrichment in accessible peaks using chromVAR (v.1.4.1)(*79*). EXs was divided into upper-layer-EXs and lower-layer-EXs by layer makers. GC bias was corrected based on BSgenome.MfascicμLaris.NCBI.5.0 genome by addGCBias function. Then we fetched the human motifs in JASPAR database by getJasparMotifs function. The deviation z-scores for each TF motif in each cell was calculated by computeDeviations function. We obtained high variance TF motifs within upper-layer- and lower-layer-EXs by computeVariability function with cut-off at 1 (N=132 for upper layer, N=117 for lower layer). To determine TF binding motifs that are significantly different between FC and V1 region, two-sided Student's T-test were performed on TF motif deviation z-scores between FC (PFC and M1) and V1 EXs. To control the FDR (false discovery rate), we performed multiple-testing correction to calculate q-values with Benjamini-Hochberg procedure. Motifs with FDR ≤ 0.05 were determined as differentially enriched motifs (N=104 for

upper-layer-EXs, N=55 for lower-layer-EXs).

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

We used findMotifsGenome.pl tool in HOMER software(*80*) to perform TF enrichment analysis on differential accessible peaks of 27 cell clusters. Then we normalized the results of known motif by p-value and visualized with selected top 20 motifs for each cell cluster. Finally, we used min-max normalization for each TF motif across 27 cell clusters.

Pseudo time trajectory analysis of OLI lineage datasets

To uncover the dynamic change of both transcriptome and chromatin accessibility along the oligodendrocyte lineage. We constructed the development trajectory using oligodendrocytes and oligodendrocyte precursor cells. For RNA-seq data, monocle 2 (*81*) was used to order 2,857 Oligo cells and 689 OPCs from the single nuclei RNA-seq data set to developmental trajectory. TPM matrix of OLI cells and OPCs was input and then transformed into normalized mRNA counts using the “relative2abs” function. Genes used for ordering the cells along the trajectory were selected under the following criteria: mean_expression >= 0.5; dispersion_empirical >= 1. After that, the discriminative dimensionality reduction with trees (DDRTree) method was used to reduce data into two dimensions. The first 2000 genes significantly varied across

pseudotime were visualized using heatmap and gene ontologies were identified using metascape (https://metascape.org).

For ATAC-seq data, we used monocle 2 to plot a heatmap of gene activity over pseudo time in OPCs differentiation into OLI. We converted the peak by cells matrix to the gene activity score matrix by summing all peaks within the gene TSS ± 25 kb. To determine peaks that were differentially accessed between the different cell states, we applied differentialGeneTest function from monocle 2 to gene activity score matrix, then filtered genes with q-values less than 1e-10 to construct a trajectory. DDRTree is used to reduce dimensions and to reconstruct the temporal progression. Finally, we used genes with q-values less than 0.01 to plot heatmap of dynamics genes activity along trajectory.

To establish a heatmap of TF motifs enrichment over pseudotime, the cells were grouped into 10 bins based on their trajectory values.

Evaluation of enrichment of disease risk variants within cell-types

To predict the correlation for different complex traits with specific cortical cell type or subtype, we used the LDSC [https://github.com/bulik/ldsc] to predict enrichments of diseases SNP-heritability in differentially accessible peaks for each cell type or subtype. First, we used the liftOver [https://genome.ucsc.edu/cgi-bin/hgLiftOver] to lift the differentially accessible peaks of each cluster and all accessible peaks to human hg19 genome. Then, LD scores of SNPs of 1000 genomes phase 3 in those peaks were estimated according to recommended workflow. We generated the

summary statistics from previous GWAS studies (Table S7) and calculate the significant enrichment of each diseases following the "cell type specific analysis" workflow recommend by the LDSC authors. We also applied the LDSC to predict colocalization of diseases SNP-heritability and differentially used junctions for each cell type and subtype. We followed the similar procedure as above, while the differentially accessible peaks are replaced with differentially used junctions, then gained the enrichments score of each trait for each cluster.

Figure Legends

Supplemental Figures

Fig. S1. Quality assessment of sci-ATAC-seq data. (A & B) Histogram of usable reads (A) and read ratio (B) distribution in cells that passed quality control. (C & D) Box plot of read ratios in peak (C) and peak numbers (D) of each brain cell type either merged or separated by region. (E) Heatmap representing the average gene activity score for layer markers across 15 EX neuron clusters from sci-ATAC-seq. (F) Heatmap showing the enriched *cis*-regulatory elements in 27 clusters based on chromatin accessibility.

Fig. S2. Quality assessment of snRNA-seq data. (A) UMAP plot showing the distribution of brain region in each cluster of snRNA-seq. (B) UMAP plot showing the distribution of the number of genes detected in each cluster of snRNA-seq. (C)

UMAP plot showing the distribution of unique mapped reads in each cluster of snRNA-seq. (D) Heatmap displaying the DEGs of major cell classes in snRNA-seq data. Specific genes related to each cell type are highlighted with enriched gene ontology terms. (E) The snRNA-seq data of excitatory neuron were re-clustered into 11 clusters and visualized by UMAP, and the projection of original 4 cluster of EX neurons is showed in upper right corner (left). Heatmap showing selected layer marker expression for subpopulations of EX neuron identified in left UMAP (right). (F) Dot plot showing the expression values and proportion of cells for IN neuron marker genes and subtype enriched DEGs. Size of each dot represents the percentage of cells expressing the marker.

Fig. S3. Areal differences in chromatin state and gene expression within excitatory neurons. (A) The proportion of cells from epigenetic cell clusters in each of the brain region analyzed. (B) Differential accessible peaks between FC and V1 regions within TSS ± 3 kb were assigned to the nearest genes by ChIPseeker and enriched GO terms for differential accessible genes of FC (PFC and M1) and V1 in upper-layer and layer-layer EX neuron are indicated. (C) Enriched GO terms for differentially expressed genes of FC (PFC and M1) and V1 in upper-layer and lower-layer excitatory neuron are indicated.

Fig. S4. Single-cell chromatin accessibility at the HPSGs pathway-associated genes locus in excitatory neurons. Integrative genomics viewer plot showing

single-cell chromatin accessibility at *HPSE2, HS6ST2, LRRTM4* and *SULF2* locus in excitatory neurons. Differentially accessible peaks between FC (PFC and M1) and V1 in upper and lower layer are highlighted.

Fig. S5. Pseudotime heatmap of OLI lineage data. (A) Heatmap showing dynamic gene expression modules along remyelination pseudotime trajectory. Enriched gene ontologies terms for OPC (I), immature OLIs (II) and mature OLIs (III) are indicated. (B) Heatmap showing dynamic gene expression of genes enriched in regulation of Wnt pathway.

Fig. S6. Differential alternative splicing events across cortical cell types. (A) Heatmap of differentially spliced gene numbers for pairwise comparison of transcriptomic cell types. (B) Heatmap displaying variable enriched junction reads in major neuronal and non-neuronal cell types (left) and in excitatory neuron and inhibitory neuron cell subtypes (right) of snRNA-seq data. (C) Psi value for skip exon of *AKAP9* and *NCAM1* in neuronal and non-neuronal cell types.

Figure 1

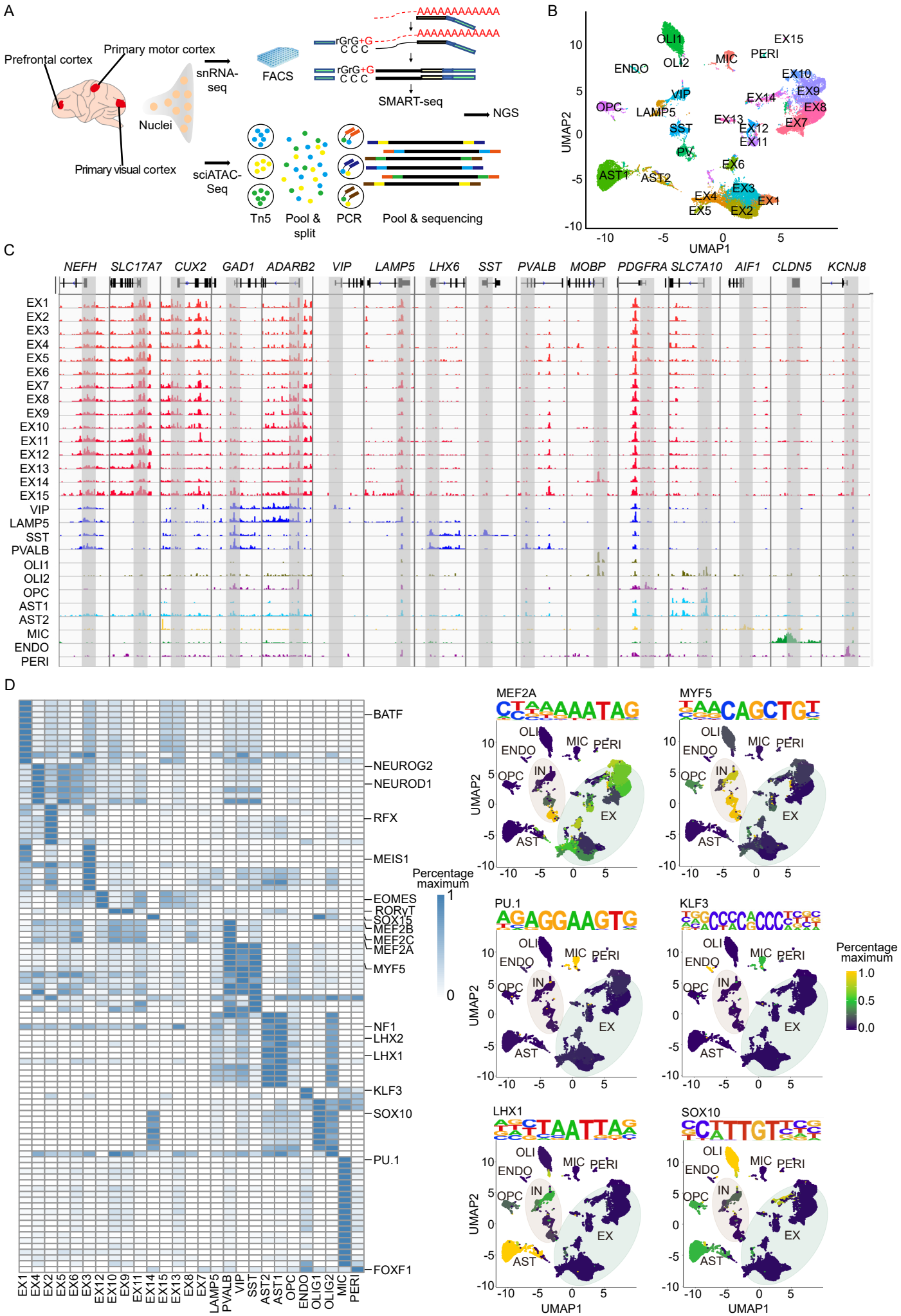

A
Prefrontal cortex
Primary motor cortex
Primary visual cortex
Nuclei
snRNA-seq
FACS
SMART-seq
NGS
sciATAC-Seq
Tn5
Pool & split
PCR
Pool & sequencing
B
UMAP1
UMAP2
OLI1
OLI2
ENDO
MIC
PERI
EX15
EX10
EX9
EX8
EX7
EX14
EX13
EX12
EX11
OPC
LAMP5
VIP
SST
PV
EX6
AST1
AST2
EX3
EX4
EX5
EX2
EX1
C
NEFH
SLC17A7
CUX2
GAD1
ADARB2
VIP
LAMP5
LHX6
SST
PVALB
MOBP
PDGFRA
SLC7A10
AIF1
CLDN5
KCNJ8
D
BATF
NEUROG2
NEUROD1
RFX
MEIS1
EOMES
RORγT
SOX15
MEF2B
MEF2C
MEF2A
MYF5
NF1
LHX2
LHX1
KLF3
SOX10
PU.1
FOXF1
Percentage maximum
MEF2A
MYF5
PU.1
KLF3
LHX1
SOX10
IN
EX
AST

Figure 2

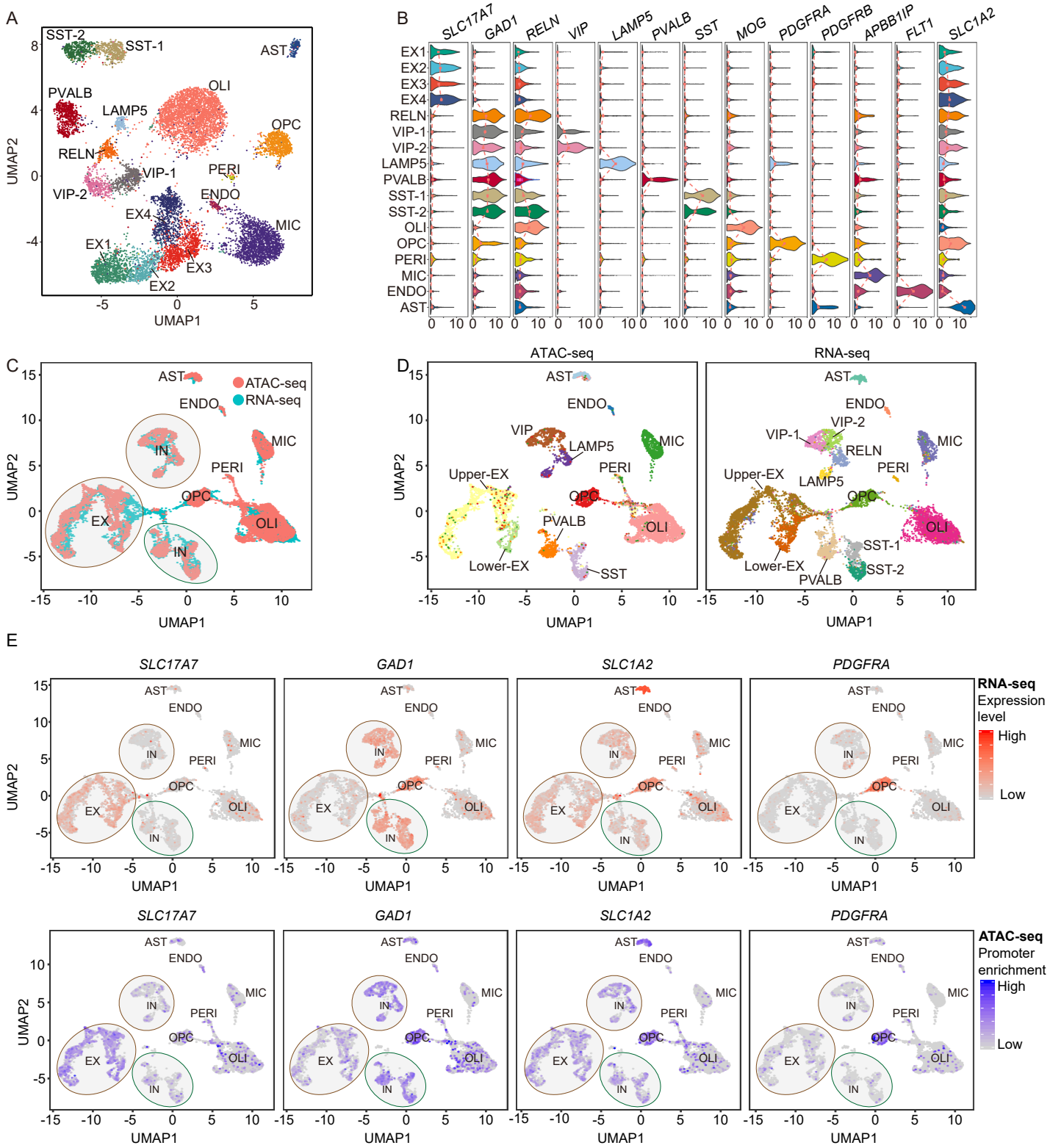

A
SST-2
SST-1
AST
OLI
PVALB
LAMP5
OPC
RELN
VIP-1
PERI
VIP-2
ENDO
EX4
MIC
EX1
EX3
EX2
UMAP2
UMAP1
B
SLC17A7
GAD1
RELN
VIP
LAMP5
PVALB
SST
MOG
PDGFRA
PDGFRB
APBB1IP
FLT1
SLC1A2
EX1
EX2
EX3
EX4
RELN
VIP-1
VIP-2
LAMP5
PVALB
SST-1
SST-2
OLI
OPC
PERI
MIC
ENDO
AST
C
AST
ATAC-seq
RNA-seq
ENDO
MIC
IN
PERI
OPC
EX
OLI
IN
UMAP2
UMAP1
D
ATAC-seq
RNA-seq
AST
ENDO
VIP
LAMP5
MIC
Upper-EX
PERI
OPC
PVALB
OLI
Lower-EX
SST
VIP-2
VIP-1
RELN
SST-1
SST-2
E
SLC17A7
GAD1
SLC1A2
PDGFRA
RNA-seq
Expression level
High
Low
ATAC-seq
Promoter enrichment
High
Low
AST
ENDO
MIC
IN
PERI
OPC
EX
OLI
UMAP2
UMAP1

Figure 3

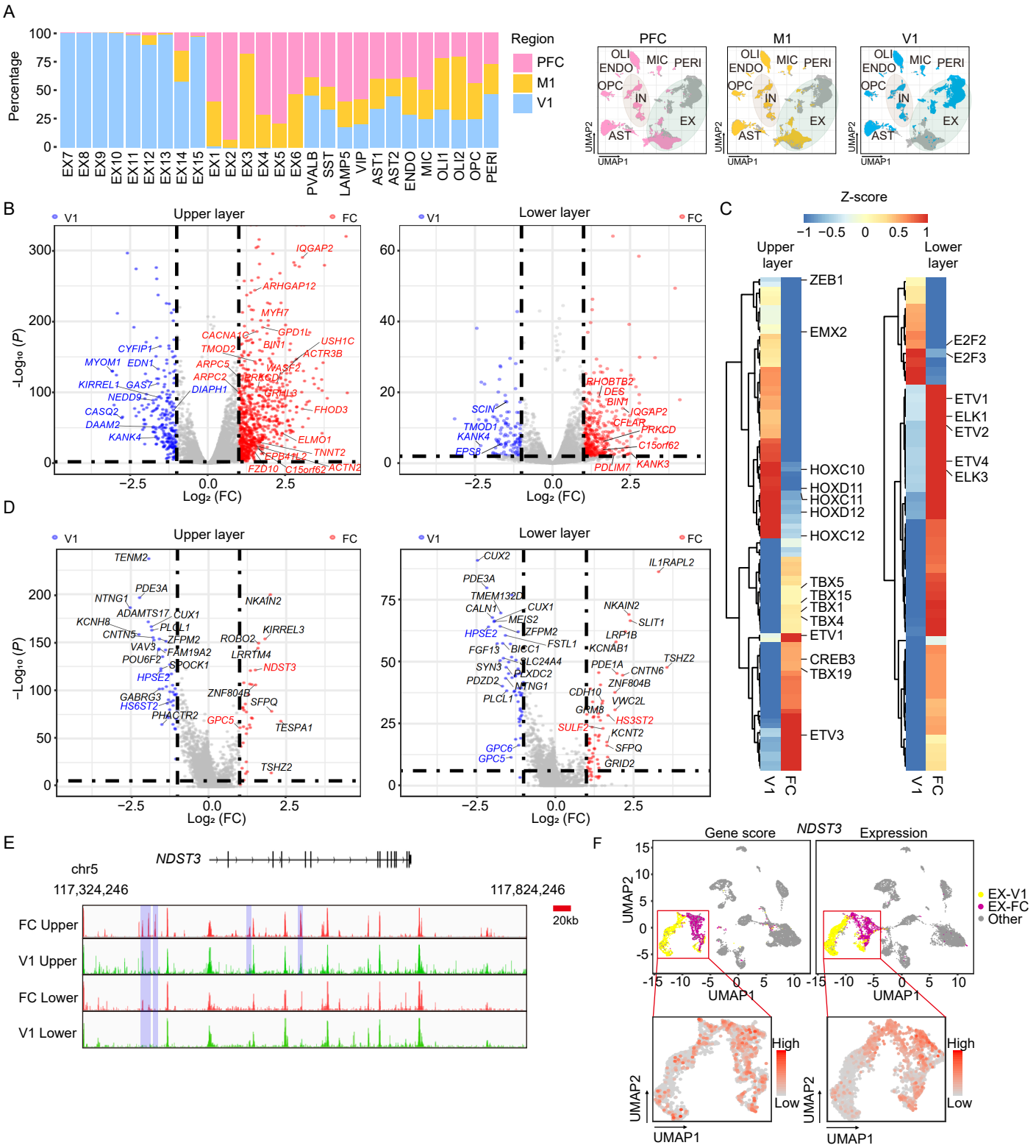

A
Percentage
Region
PFC
M1
V1
EX7 EX8 EX9 EX10 EX11 EX12 EX13 EX14 EX15 EX1 EX2 EX3 EX4 EX5 EX6 PVALB SST LAMP5 VIP AST1 AST2 ENDO MIC OLI1 OLI2 OPC PERI
PFC
M1
V1
OLI ENDO MIC PERI OPC IN EX AST
UMAP1
UMAP2
B
Upper layer
Lower layer
V1
FC
-Log10 (P)
Log2 (FC)
C
Z-score
Upper layer
Lower layer
ZEB1
EMX2
HOXC10
HOXD11
HOXC11
HOXD12
HOXC12
TBX5
TBX15
TBX1
TBX4
ETV1
CREB3
TBX19
ETV3
E2F2
E2F3
ETV1
ELK1
ETV2
ETV4
ELK3
D
Upper layer
Lower layer
E
chr5
NDST3
117,324,246
117,824,246
20kb
FC Upper
V1 Upper
FC Lower
V1 Lower
F
Gene score
NDST3
Expression
EX-V1
EX-FC
Other
High
Low
UMAP1
UMAP2

Figure 4

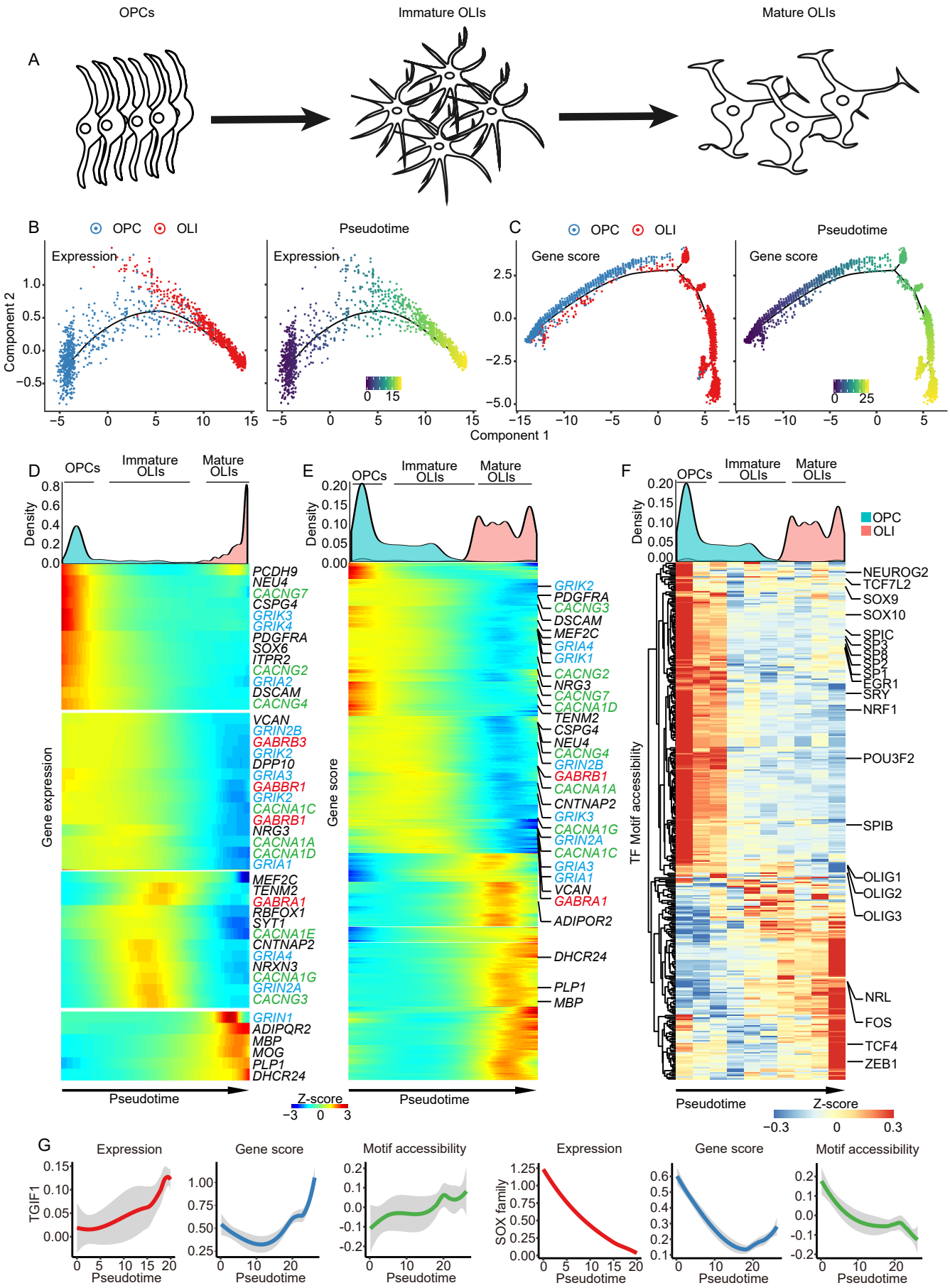
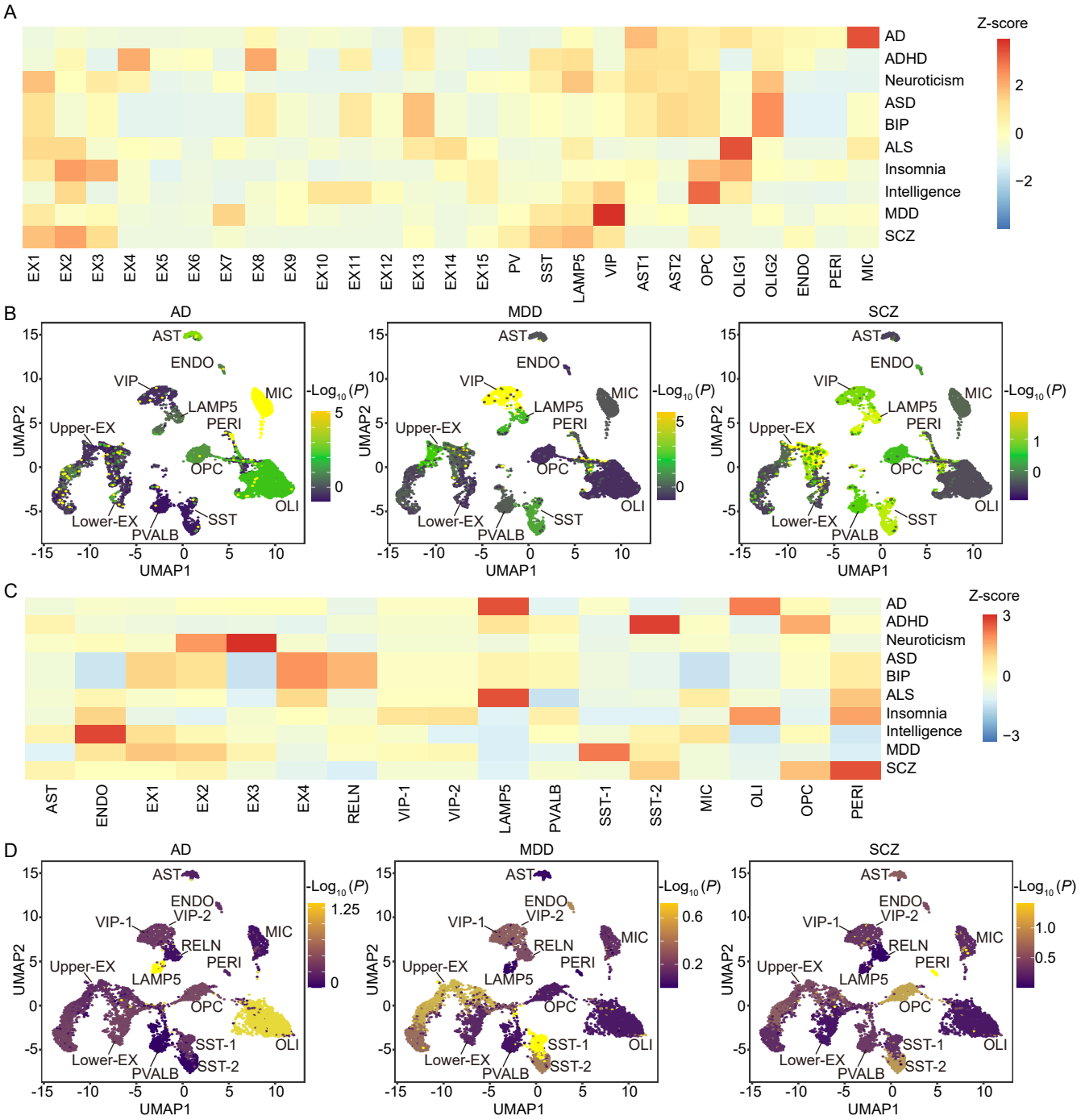
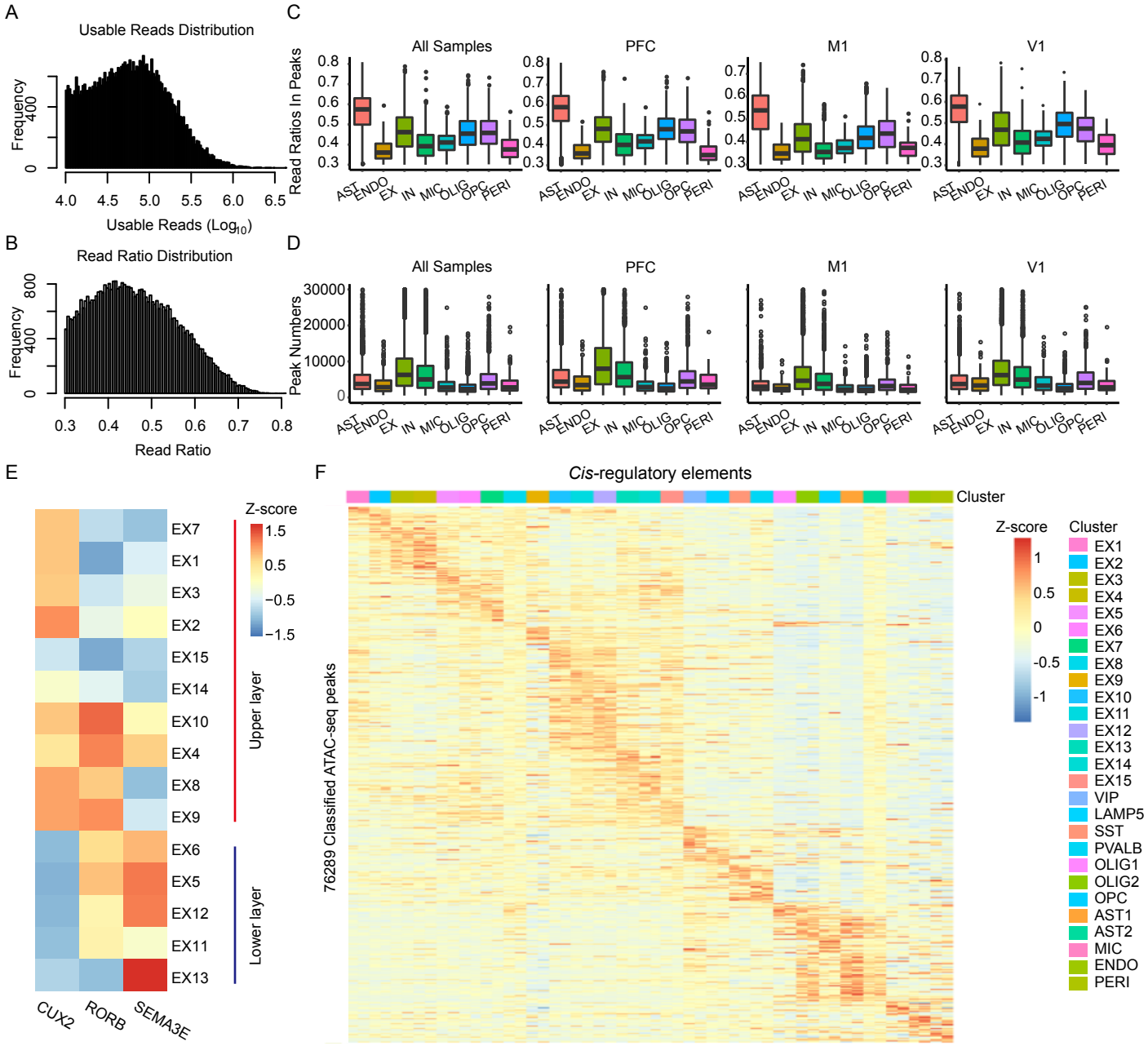

A
OPCs
Immature OLIs
Mature OLIs
B
OPC
OLI
Expression
Pseudotime
Component 2
Component 1
C
Gene score
D
E
F
Density
Gene expression
Gene score
TF Motif accessibility
Pseudotime
Z-score
PCDH9
NEU4
CACNG7
CSPG4
GRIK3
GRIK4
PDGFRA
SOX6
ITPR2
CACNG2
GRIA2
DSCAM
CACNG4
VCAN
GRIN2B
GABRB3
GRIK2
DPP10
GRIA3
GABBR1
GRIK2
CACNA1C
GABRB1
NRG3
CACNA1A
CACNA1D
GRIA1
MEF2C
TENM2
GABRA1
RBFOX1
SYT1
CACNA1E
CNTNAP2
GRIA4
NRXN3
CACNA1G
GRIN2A
CACNG3
GRIN1
ADIPQR2
MBP
MOG
PLP1
DHCR24
GRIK2
PDGFRA
CACNG3
DSCAM
MEF2C
GRIA4
GRIK1
CACNG2
NRG3
CACNG7
CACNA1D
TENM2
CSPG4
NEU4
CACNG4
GRIN2B
GABRB1
CACNA1A
CNTNAP2
GRIK3
CACNA1G
GRIN2A
CACNA1C
GRIA3
GRIA1
VCAN
GABRA1
ADIPOR2
DHCR24
PLP1
MBP
OPC
OLI
NEUROG2
TCF7L2
SOX9
SOX10
SPIC
SP3
SP8
SP2
SP1
EGR1
SRY
NRF1
POU3F2
SPIB
OLIG1
OLIG2
OLIG3
NRL
FOS
TCF4
ZEB1
G
Expression
Gene score
Motif accessibility
TGIF1
SOX family

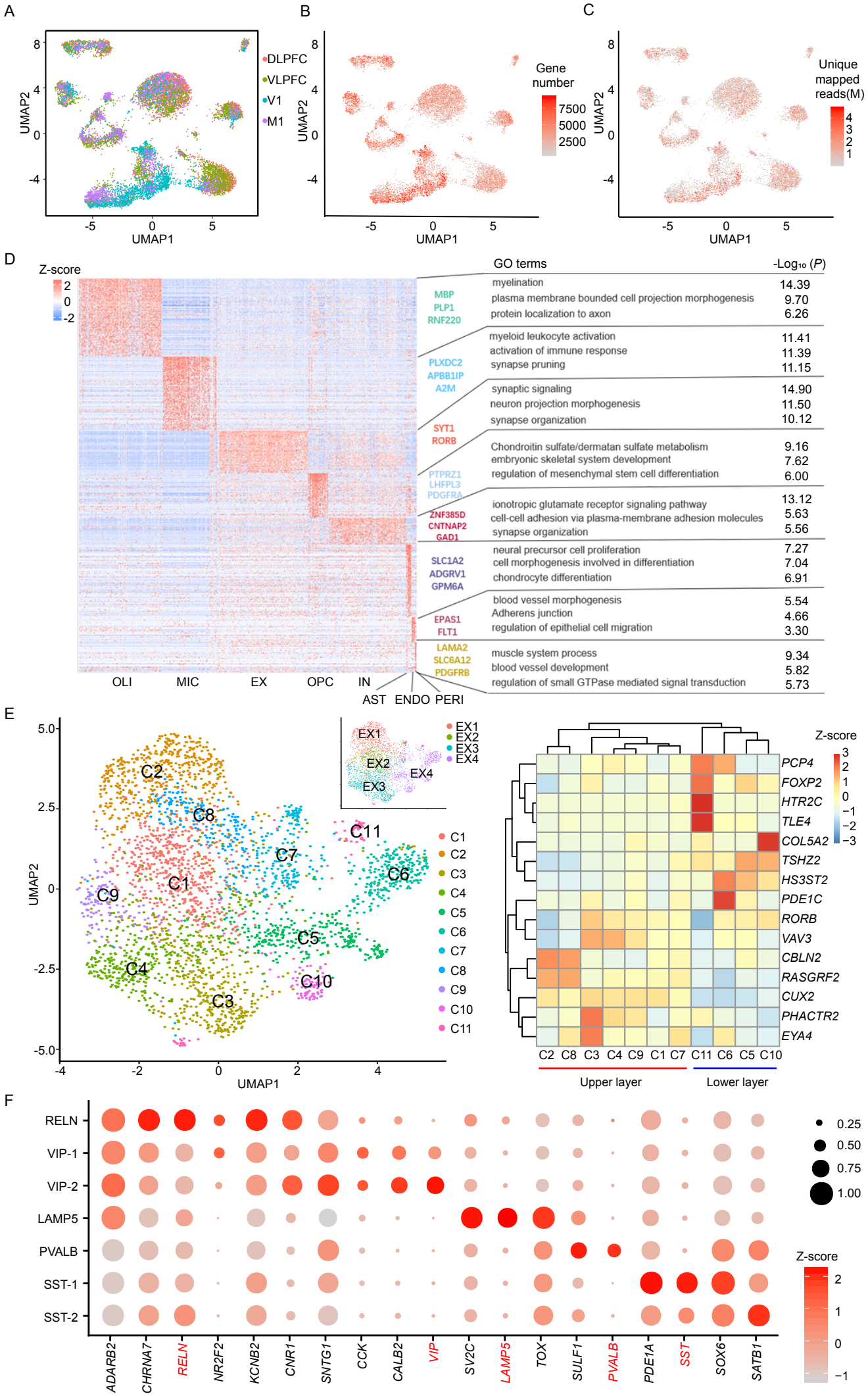
Figure 5


A
Z-score
AD
ADHD
Neuroticism
ASD
BIP
ALS
Insomnia
Intelligence
MDD
SCZ
EX1 EX2 EX3 EX4 EX5 EX6 EX7 EX8 EX9 EX10 EX11 EX12 EX13 EX14 EX15 PV SST LAMP5 VIP AST1 AST2 OPC OLIG1 OLIG2 ENDO PERI MIC
B
AD
MDD
SCZ
UMAP1
UMAP2
-Log10 (P)
AST
ENDO
VIP
MIC
LAMP5
PERI
Upper-EX
OPC
OLI
Lower-EX
PVALB
SST
C
Z-score
AST ENDO EX1 EX2 EX3 EX4 RELN VIP-1 VIP-2 LAMP5 PVALB SST-1 SST-2 MIC OLI OPC PERI
D
AD
MDD
SCZ
UMAP1
UMAP2
-Log10 (P)
AST
ENDO
VIP-1
VIP-2
MIC
RELN
PERI
Upper-EX
LAMP5
OPC
OLI
SST-1
SST-2
Lower-EX
PVALB

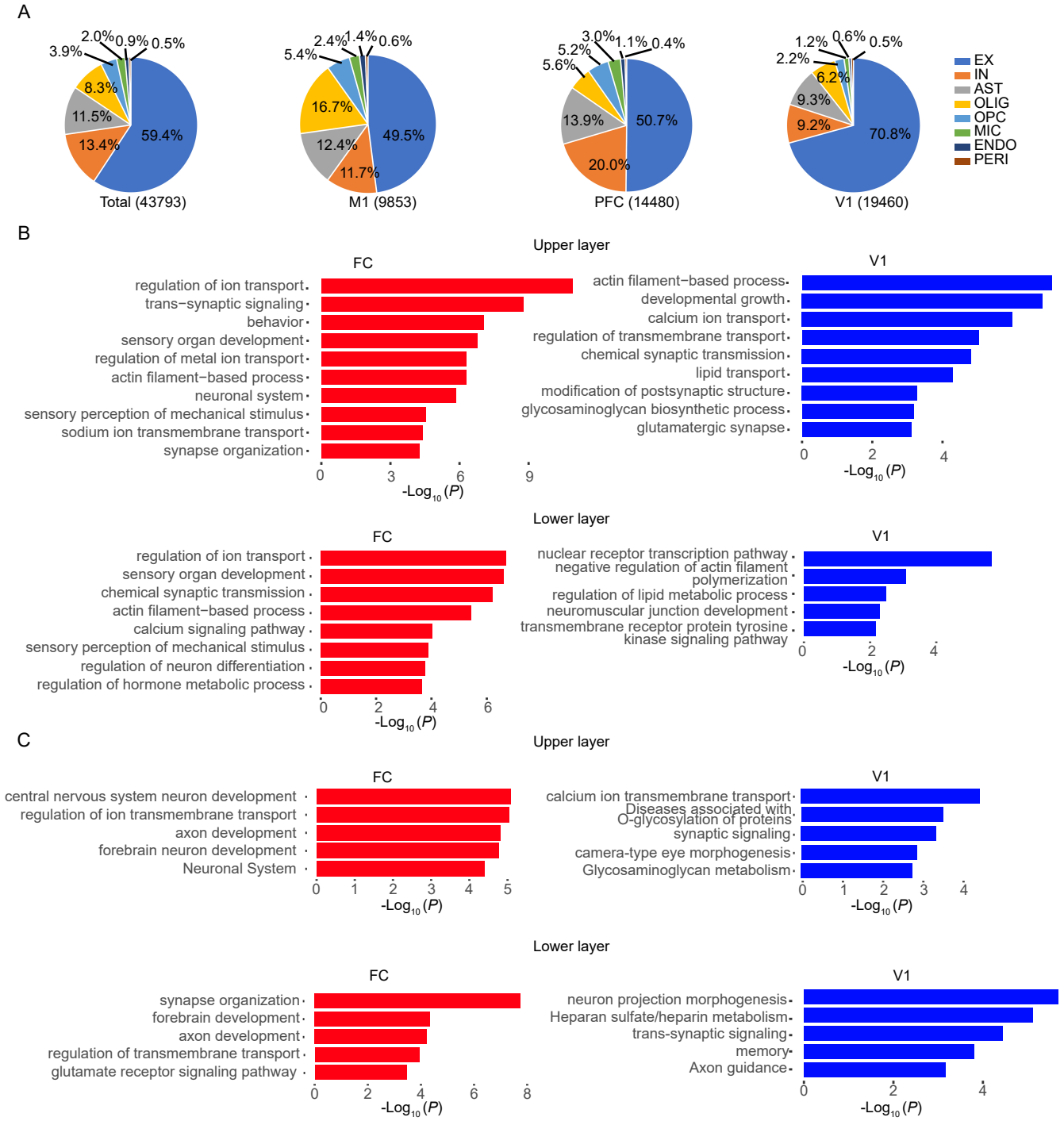
Supplemental Figure 1



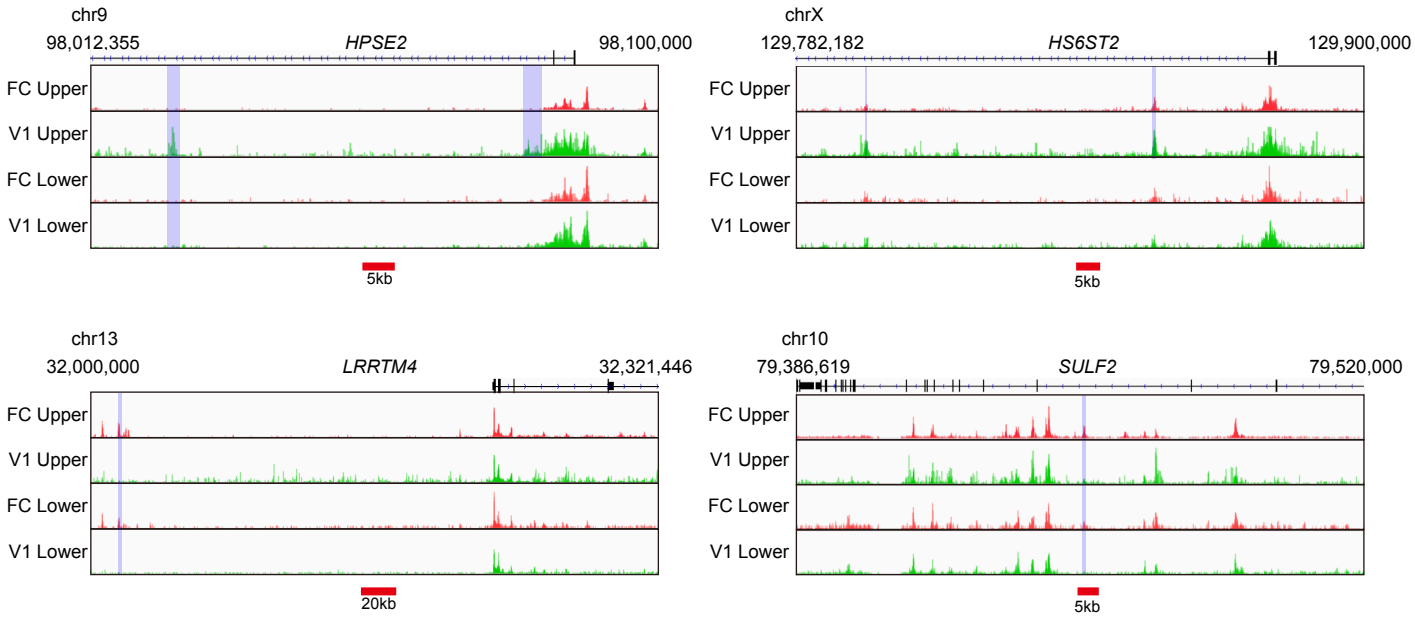
Supplemental Figure 2



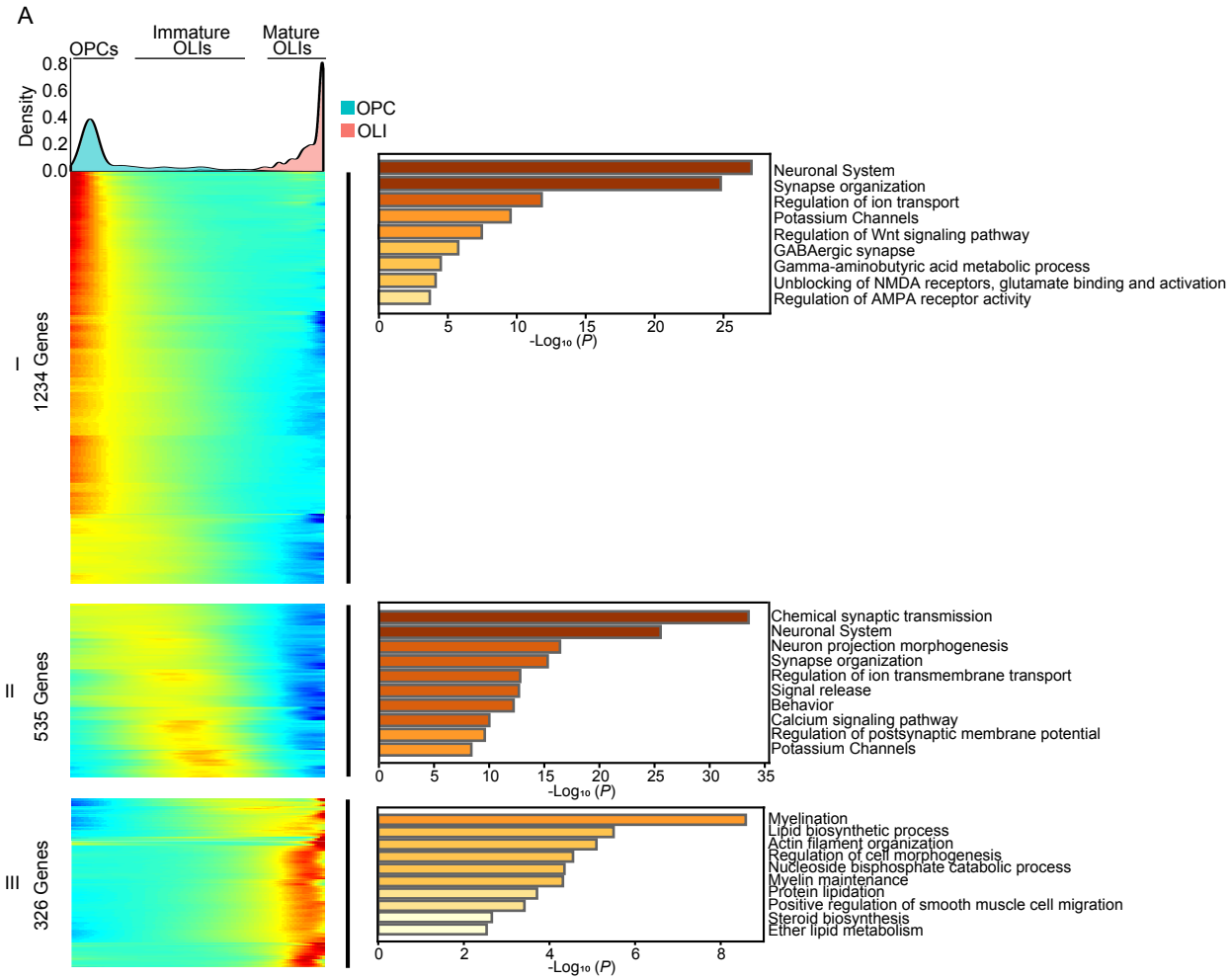
Supplemental Figure 3

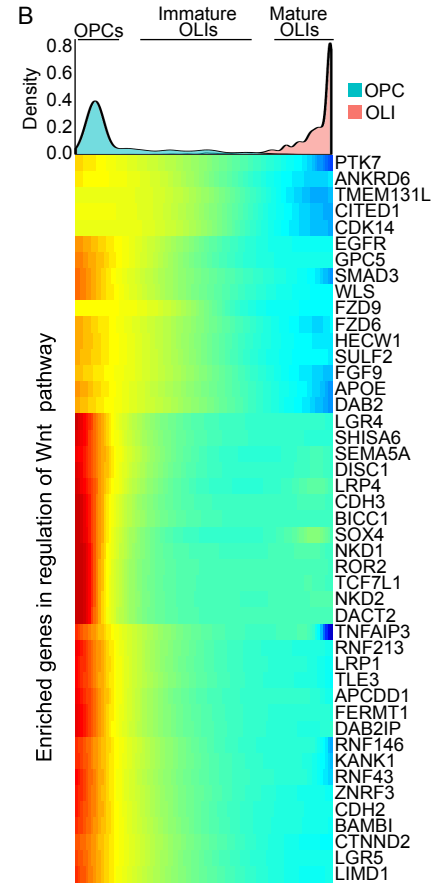
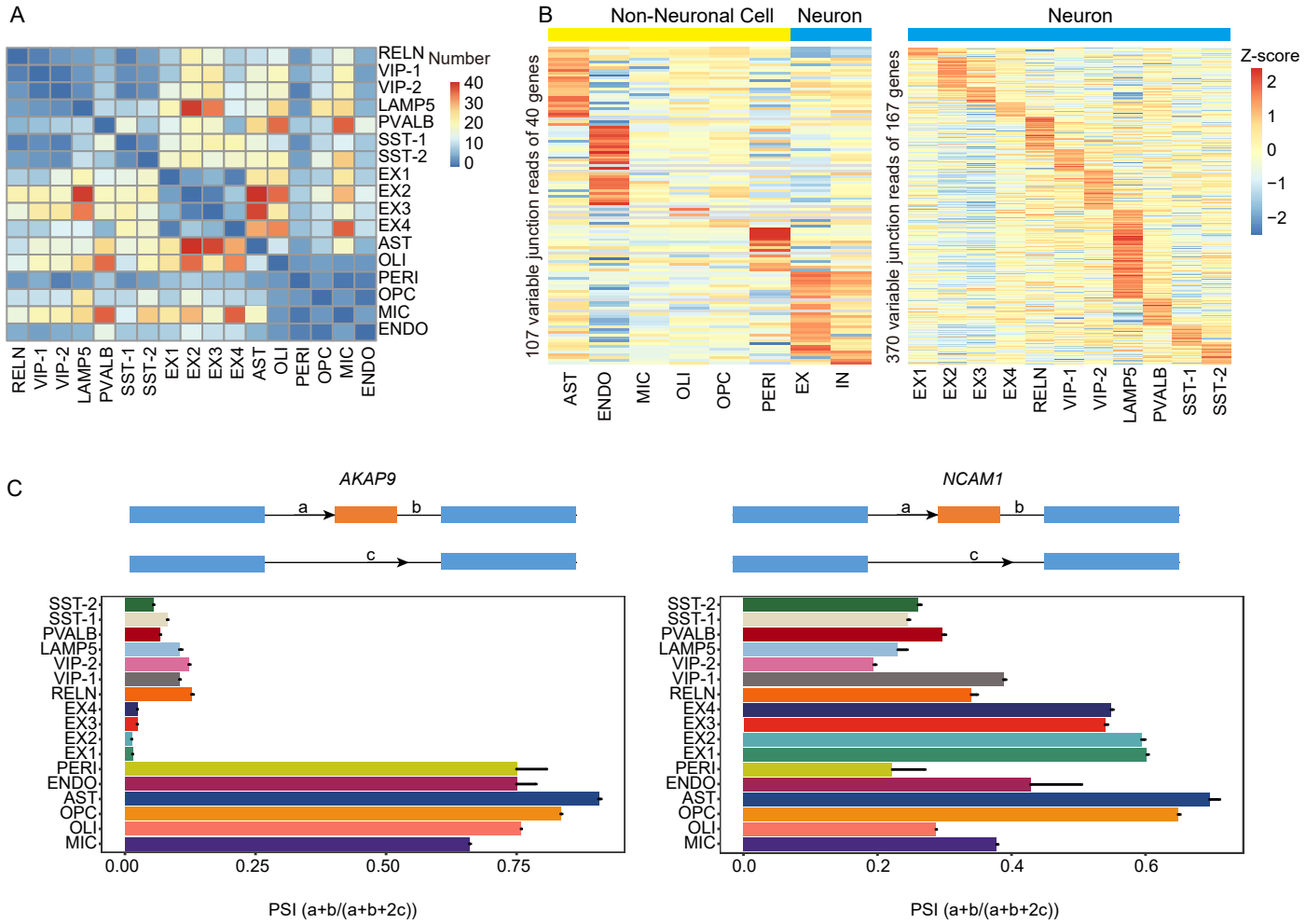


Supplemental Figure 4



Supplemental Figure 5





Supplemental Figure 6