Temporal transcriptomic dynamics in developing macaque neocortex

Longjiang Xu[1,5] , Zan Yuan[2,3,5], Jiafeng Zhou[2,5], Yuan Zhao[1,5], Wei Liu[2] , Shuaiyao Lu[1], Zhanlong He[1], Boqin Qiang[2], Pengcheng Shu[2,4], Yang Chen[2]*, Xiaozhong Peng[1,2]**

1 Institute of Medical Biology Chinese Academy of Medical Sciences, Chinese Academy of Medical Science and Peking Union Medical College, Kunming 650118, China

2 The State Key Laboratory of Medical Molecular Biology, Institute of Basic Medical Sciences, Chinese Academy of Medical Sciences and School of Basic Medicine, Peking Union Medical College, 5 Dong Dan San Tiao, 100005 Beijing, China.

3 Agricultural Bioinformatics Key Laboratory of Hubei Province, Hubei Engineering Technology Research Center of Agricultural Big Data, College of Informatics, Huazhong Agricultural University, Wuhan, 430070, China

4 Chinese Institute for Brain Research, Beijing, 102206, China

5 These authors contributed equally to this work.

E-mail: Xiaozhong Peng (pengxiaozhong@pumc.edu.cn); Yang Chen (yc@ibms.pumc.edu.cn); Pengcheng Shu (pengcheng_shu@ibms.pumc.edu.cn).

Keywords: (scRNA-Seq, neural progenitor , macaque cortex , embryonic neurogenesis)

Abstract

Despite intense research focus in mice, the transcriptional regulation of neocortical neurogenesis remains limited in humans and non-human primates. Cortical development in rhesus macaque is known to recapitulate multiple facets cortical development in humans, including the complex composition of neural stem cells and thicker upper-layer neurons. To characterize temporal shifts in transcriptomic programming responsible for differentiation from stem cells to neurons, we sampled parietal lobes of rhesus macaque at E40, E50, E70, E80, and E90, spanning the full period of embryonic neurogenesis. Single cell RNA sequencing produced a transcriptomic atlas of the developing rhesus macaque neocortex parietal lobe. Identification of distinct cell types and neural stem cells emerging in different developmental stages revealed a terminally bifurcating trajectory from stem cells to neurons. Notably, deep-layer neurons appear in early stages of neurogenesis while upper-layer neurons appear later. While these different lineages show overlap in their differentiation program, cell fates are determined post-mitotically. Pseudotime trajectories from vRGs to oRGs revealed differences in dynamic gene expression profiles, and identified divergence in their activation of BMP, FGF, and WNT signaling pathways. These results provide a comprehensive picture of the temporal patterns of gene expression leading to different fates of radial glial progenitors during neocortex layer formation.

1. Introduction

The neocortex is the center for higher brain functions, such as perception and decision-making, and therefore the dissection of its developmental processes can be informative of the mechanisms responsible for these functions. Several studies have advanced our understanding of the principles underlying neocortical development in different species, especially in mice. Generally, the dorsal neocortex can be anatomically divided into six layers of cells, occupied by distinct neuronal cell types. The deep-layer neurons project to the thalamus (layer VI neurons) and subcortical areas (layer V neurons), while neurons occupying more superficial layers (upper-layer neurons) preferentially form intracortical projections[1]. The generation of distinct excitatory neuron cell types follows a temporal pattern in which early-born neurons migrate to deep layers (i.e., layers V and VI), while the later-born neurons migrate and surpass early-born neurons to occupy the upper layers (layer II-IV) [2]. In Drosophila, several transcription factors are sequentially expressed specifically in neural stem cells to control specification of daughter neuron fates, while very few such transcription factors have been identified in mammals thus far. Using single-cell RNA sequencing (scRNA-seq), Telley and colleagues found that daughter neurons exhibit the same transcriptional profiles of their respective progenitor radial glia, although these apparently heritable expression patterns fade as neurons mature[3]. However, the temporal expression profiles of neural stem cells and the contribution of these specific temporal expression patterns in determining neuronal fate have not been wholly clarified in humans and non-human primates.

In rodents, radial glia cells are found in the ventricular zone (VZ), where they undergo proliferation and differentiation. Although ventricular radial glia (vRG) is also found in humans and non-human primates, the vast majority of radial glia in these higher species occupy the outer subventricular zone (OSVZ), and are therefore termed outer radial glia (oRG). VRG and oRG are both accompanied by the expression of stem cell markers such as PAX6 and exhibit extensive self-renewal and proliferative capacities [4]. However, despite functional similarities, they have distinct molecular phenotypes. Previous scRNA-seq analyses have identified several molecular markers, including MOXD1 and HOPX for oRGs, and CRYAB and FBXO32 for vRGs[5]. Furthermore, oRGs are derived from vRGs and vRGs exhibit obvious differences in numerous cell-extrinsic mechanisms, including activation of the FGF-MAPK cascade, SHH, PTEN/AKT, and PDGF pathways, and oxygen (O2) levels. These pathways and factors are involved in three broad cellular processes that include, vRG maintenance, spindle orientation, and cell adhesion/extracellular matrix production[6].Some transcription factors have been

shown to participate in vRG generation, such as INSM and TRNP1. Moreover, the cell-intrinsic patterns of transcriptional regulation responsible for generation of vRGs have not been characterized.

ScRNA-seq is a powerful tool for investigating developmental trajectories, defining cellular heterogeneity, and identifying novel cell subgroups[7]. Several groups have sampled embryonic mouse neocortex tissue for scRNA-seq [8] [9], as well as discrete, discontinuous embryonic developmental stages in human and non-human primates, while studies spanning the full embryonic neurogenic stage in the neocortex of humans and other primates are lacking[5, 10] [11-12]. Rhesus macaque and humans share multiple aspects of neurogenesis, and more importantly, the rhesus monkey and human brains share more similar gene expression patterns than brains of mice and humans[13-15]. To establish a comprehensive, global picture of the neurogenic processes in the rhesus macaque neocortex, which can be informative of neocortex evolution in humans, we sampled neocortical tissue at five developmental stages (E40, E50, E70, E80 and E90) in rhesus macaque embryos, spanning the full neurogenesis period. Through strict quality control, cell type annotation, and lineage trajectory inference, we identified two broad transcriptomic programs responsible for the differentiation of deep-layer and upper-layer neurons. We also defined the temporal expression patterns of neural stem cells, including oRGs, vRGs and IPs, and identified novel transcription factors involved in oRG generation. These findings can substantially enhance our understanding of neocortical development and evolution in primates.

2. Results

2.1. scRNA-Seq Analysis of Cell Types in the Developing Macaque Neocortex

In order to establish a comprehensive view of the cellular composition of the rhesus macaque brain at different stages in embryonic period development, we conducted single-cell RNA sequencing (scRNA-seq) on the dissected parietal lobes of eight total rhesus macaque embryos, spanning five developmental stages of embryonic neurogenesis, including E40 (stage of peak neurogenesis) and E50 (Layer 6 formation); as well as at E70 (Layer 5 formation), E80 (Layer 4 formation), and E90 (Layer2-3 formation) (**Figure 1A**)[2]. We obtained a transcriptome from 53,295 cells after filtering out low quality cells and removing potential doublets. Each embedding was visualized using uniform manifold approximation and projection (UMAP) of dimension reduction using Seurat, which identified 28 distinct clusters. All cell clusters present in the samples were then annotated (**Figure 1B**) and the respective cell types were identified (**Figure 1C**) based on their expression of molecular markers (**Figure 1D** and **figure S2, B to C**). Each of the 28 identified clusters could be assigned to a cell type identity of neurocytes

(including radial glia (RG), outer radial glia (oRG), intermediate progenitor cells (IPCs), ventral precursor cells (VP), excitatory neurons (EN), inhibitory neurons (IN), oligodendrocyte progenitor cells (OPC), oligodendrocytes, astrocytes, microglia, Cajal-Retzius cells and meningeal cells, or non-neuronal cell types (including endothelial, vascular/pericyte, and blood cells). Each cell cluster was composed of multiple embryo samples, and the samples from similar stages generally harbored similar distributions of cell types.

In general, ventricular radial glia (vRG) (cluster 10) showed characteristic VIM and PAX6 expression; outer radial glia (oRG) (cluster 12) highly expressed HOPX and MOXD1, two previously verified markers[5]; two clusters (cluster 8 and cluster 22) of intermediate progenitor cells (IPCs) that strongly expressed EOMES, and one cluster of IPCs (cluster 22) were all topologically close to radial glia, while the other IPC cluster (cluster 8) was closer to neurons, indicating the presence of two stages of IPCs[16]; excitatory neurons were identified by the expression of well-established markers, such as NEUROD2 and NEUROD6; and the astrocyte and oligodendrocyte lineages were identified by AQP4 and SOX10 expression, respectively[17]. In addition, we identified DLX1+ and GAD1+ cells, which suggested the presence of inhibitory neurons (Figure 1C and figure S2).

Collectively, these results suggested that cortical neural progenitors undergo neurogenesis processes during the early stages of macaque embryonic cortical development, while gliogenic differentiation, including oligodendrocytes and astrocytes, occurs in later stages (Figure 1A and 1B).

2.2. Distinct Excitatory Neuronal Types Sequentially Emerge in Developing Cortex

To better understand the temporal dynamics of excitatory neuron (EN) development and differentiation, we compiled a subset of all ENs and re-clustered them into 11 subclusters (EN1-11) (**Figure 2A**). Accordingly, we identified a subcluster of neurons (EN 8) from the midbrain, which highly expressed TCF7L2 [18], which were not included in following analysis. We then calculated the relative expression levels of cellular markers for each group and annotated the EN subclusters based on published descriptions of marker function (**Figure S3B**). We found that all EN subclusters could be well-distinguished by differential expression of either deep-layer neuron markers (BCL11B, FEZF2, and SOX5)[19] or upper-layer neuron markers (CUX1 and SATB2)[20-21](**Figure 2, B to C**). In previous seminal studies, 3H-thymidine (3H-dT) tracing in macaque rhesus unraveled the sequential generation of cortical neurons, with deep-layer neurons appearing prior to upper-layer neurons[2]. In agreement with this early work, we found that early-born neurons (E40 and E50) predominantly outnumbered later-born neurons

in the deep-layer neuron subclusters (EN5 and 11), while upper-layer neuron subclusters contained the inverse proportions (**Figure 2, C and D**).

We then characterized temporal changes in the composition of each EN subcluster. While the EN 5 and EN 11 (deep-layer neurons) subclusters emerged at E40 and E50 and disappeared in later stages, EN subclusters 1, 2, 3, and 4 gradually increased in population size from E50 to E80 (**Figure 2D**). Notably, EN 6 was exclusively found in the cortex at E70 (**Figure S3A**). Since the brain is relatively small at E40 and E50, the cells collected at these stages inevitably included cells from adjacent regions.

2.3. Specification of Different Progenitor Fates Controlled by Regulator Genes

To focus on the differentiation of neural progenitors, we generated subsets of cell clusters 8, 10, 12, 14, 15, 16, 20, and 22, then annotated them as ventricular radial glia (RG_C10), outer radial glia (oRG_C12 and RG_C14), intermediate progenitor cells (IPC_C22 and IPC_C8), ventral precursor cells (VP_C15), or oligodendrocyte precursor cells (OPC_C16) (**Figure 3A**). Subclustering analysis revealed that these neural progenitors differentiated into diverse cell types with distinct lineages across the macaque embryonic neocortex (**Figure 3B**). We next investigated trajectories within the neural stem cell pool prior to branching as upper layer or deep layer excitatory neuron (EN). Notably, the C20 cluster did not clearly belong to any identifiable lineage. The oRG and IPC precursor cell groups exhibited characteristically high specific expression of HOPX and EOMES respectively.

Although the presence of oRGs is a well-established feature of primate neurogenesis, and their molecular markers are widely used, the genetic basis and molecular processes leading to their emergence are still poorly understood. To construct gene expression profiles that illustrate the progression from RG_C10 cells to oRGs, we specifically examined changes in expression among RG_C10, oRG_C12, and RG_C14, cells, then calculated pseudotime trajectories (**Figure 3, C to D**). Based on these trajectories, we then profiled the temporal shifts in expression of each gene and selected the 150 most significant genes (**Figure 3E**). We found a set of genes that were previously reported as highly expressed in outer radial glia that also showed high expression in the oRG_C12 differentiated pseudotime terminal [5, 22-23] (such as SFRP1, HOPX, FAM107A, TNC, PTN, and MOXD1), while high ASPM expression was typical of cells in the earlier, RG, pseudotime terminal. In addition to these markers, we also found enrichment for some potential regulatory genes at the oRG_C12 terminal, such as Regulator of the cell cycle (RGCC) which controls mitotic spindle orientation [24], and TTYH1, which regulates cell adhesion [25]. We also found genes that regulate ion channel expression,

such as ATP1A2, ATP1B2, and SCN4B, which were not previously known to participate in oRG differentiation.

Analysis of DEGs specific to the RG pseudotime terminal identified cell cycle-related genes such as MKI67, TOP2A, CDC20 and CCNA2. Within the glia-like cells that largely comprised the oRG_C12 terminal, we also detected a population of true glial genes that exhibited high expression of SLC1A3, ZFP36L1 [26], GFAP, DBI, and EDNRB [27]. In addition, our analyses uncovered several DEGs that have not yet been investigated for a role in radial glia function and differentiation, such as DKK3, DDAH1, SAT1, and PEA15. Finally, we screened for significant DEGs associated with RG-to-IPC and OPC-to-astrocyte/oligodendrocyte differentiation trajectories (**Figure S5 and S6**). These data suggested that the cell fate determination by diverse neural progenitors occurs in the embryonic stages of macaque cortical development and is controlled by several key transcriptional regulators.

2.4. The Generation of Deep-layer and Upper-layer Neurons Follows Distinct Terminal Trajectories

Based on our above finding that the deep-layer neurons appear earlier than upper layer neurons, we next sought to characterize dynamic shifts in the expression of transcriptional regulators that potentially contribute to determining neuronal fates. To this end, we performed pseudotime trajectory analysis of EN lineages (including deep-layer and upper-layer neurons) (**Figure 4A**), then superimposed the cluster labels from RG_C10, oRG_C12, and RG_C14 stem cells in UMAP plots (Figure 3A) and early (E40 and E50) and late (E70, 80 and 90) emergence stages (Figure 2A) over the extracted transcriptomic data. We then calculated the RG-specific differentially expressed genes (rgDEGs), as well as the DEGs of early and late neuron-specific (nDEGs). The set of genes that overlapped between of rgDEGs and nDEGs (termed mapping genes) thus be used to map neuronal subtypes to different neural stem cell populations (**Figure. S4**). We hypothesized that mapping genes, such as FEZF2 and DOK5, may function in radial glia cells to specify neuronal progeny (**Figure 4B**). Our results provide single-cell transcriptome data to support that apical progenitors and daughter excitatory neurons share molecular temporal identities during neocortex embryonic corticogenesis [28-30].

Since a substantial body of evidence indicates that neuronal fate is determined post-mitotically in the mammalian neocortex, analysis of differentially expressed genes over the course of the differentiation process could reveal the genetic mechanisms responsible for deciding stem cell fate as different neuronal progeny. Using Dynverse R packages, we constructed pseudotime trajectories for the cluster data, which identified a bifurcating trajectory from neural stem cells (including RG_C10, oRG_C12, and RG_C14) that leads to either deep-layer neurons in one

branch or to upper-layer neurons in the alternative branch (**Figure 4C**). However, it should be noted that the two distinct fates share a common path from neural stem cells to immature neurons (Figure 4C, from **dot1** to **dot4**), supporting the likelihood that neuronal fate is primarily determined post-mitotically.

Based on this trajectory, we then categorized the DEGs that exhibit dynamic, temporal shifts in expression over pseudotime, resulting in broad clusters. The first of these clusters was enriched with neural stem cells, characterized by high expression of PAX6, VIM, and TOP2A, which likely participate in regulating stemness and proliferation. The three remaining DEG clusters were enriched in either deep-layer neurons, upper-layer neurons, or both (**Figure 4D**). More specifically, we found that most of the DEGs were shared between both branches, such as STMN2, TUBB3, NEUROD2 and NEUROD6, further illustrating the shared differentiation processes between deep and upper layer ENs. We also identified deep layer-specific DEGs (including FEZF2, TBR1 and LMO3) and upper layer-specific genes (including MEF2C and DOK5). We then organized the cell types into a branching tree based on their differential expression of these marker genes, which includes stem-cell genes PAX6, HOPX and EOMES, in addition to the above deep layer- and upper layer-specific genes (**Figure 4E**). The separation into distinct, branch-specific sets of DEGs suggested a role for these cell type-specific transcription factors and other genes in terminal neuron differentiation.

2.5. Conserved and Divergent Features of Human, Macaque and Mouse Neocortical Progenitor Cells

Next, we performed liger integration [31] to integrate the macaque single-cell dataset with published mouse [9] and human dataset, which are created by combing two published human scRNA-seq data of 7 to 21 postconceptional week (PCW) neocortex [32-33], to conduct a cross-species comparison. Species-integrated UMAP showed that all major cell type were well integrated between the three species (**Figure 5A**).

It is clearly observed that similar cellular processes with comparable temporal progression to sequential generate of deep layer and upper layer neurons in both primates and rodents (**Figure S7, A to C**). However, in primates, the number and relative development period of upper neurons are longer than that of rodents (**Figure S7D**). Besides, another important evolutionarily divergent feature in cortical development between the of primates and rodents is the abundance of oRG cells in the outer subventricular zone (oSVZ) of primates (human and macaque). Our human-macaque-mouse cross-species analysis shown that strong expression of oRG marker gene, such as HOPX, MOXD1, FAM107A and CLU (3) in both human and

macaque datasets (**Figure 5B**), while barely detectable expression in mouse dataset, which is consistent with previous reports [34].

Furthermore, We picked HOPX-positive and EOMES-positive cells (ligercluster5, 20, 13, 15) from human, mouse dataset and performed developmental trajectories analysis with slingshot (**Figure 5C**). Comparing vRG to IPC trajectory between human, macaque, and mouse, we found this biological process of vRG-to-IPC is very conserved across species, but the vRG to oRG trajectory is divergent between species. The latter process is almost invisible in mice, but it is very similar in primates and macaque.

Previous studies have shown that neural progenitors exhibit distinct temporal expression patterns during neurogenesis in mice. However, similar temporal profiling has not been thoroughly investigated in macaque neural stem cells. Based on its topological position and associated markers that suggest it functions as a developmental root, we examined temporal expression profile of RG_C10 cells (Figure 3A). We first identified marker genes to distinguish RG_C10 cells in different embryonic stages (i.e., E40-E90), and used these markers to generate dynamic expression profiles. We then clustered these dynamically expressed macaque genes into five types (**Figure 6A**), consisting of Type 1 and 2 genes that decreased in expression during neurogenesis, Type 3, 4 and 5 genes that gradually ramp up in expression over the course of neurogenesis. We then performed a similar profiling of dynamic gene expression using data obtained from mouse apical progenitors[8] and categorized the DEGs into the same five categories. A total of 72 homologous transcription factors were shared between human, macaque and mouse transcriptomes. Analysis of their temporal dynamics revealed that exhibited similar temporal patterns between species such as NEUROD1, NEUROG2, POU3F2, ETV1, ETV5 and FOS (**Figure 6B** and **Figure S8**), which indicated that the majority of transcription factors had conserved temporal dynamics between species, and thus evolutionarily conserved patterns of regulation in neural stem cell differentiation.

To identify the master regulators (MRs) related to cortical neurogenesis among macaque and mouse, we used the SCENIC workflow to analyze the gene regulatory networks of each transcription factor (**Figure 6, C** to **D** and **F** to **G**). Among all the regulatory pairs associated with each embryonic stage, we selected seven TFs with the highest regulon specificity scores (Figure 6C and F) and their top 5 target genes, as visualized by Cytoscape (**Figure 6E** and **H**). Analysis of the regulatory networks of macaque and mouse[9] embryonic neocortical neurogenesis indicated that despite commonalities in the roles of classical developmental TFs such as EOMES, NEUROD1, and EMX1, macaque had a relatively divergent set of regulatory networks than those involved in mouse cortical neurogenesis (**Figure 6D** and **G**).

3. Discussion

The cerebral cortex region of the brain responsible for extraordinary cognitive capacity, such as abstract thinking and language. The neocortex of non-human primate rhesus macaque resembles that of humans in many aspects. Thus, a comprehensive investigation of macaque neurogenesis can improve our understanding of neocortical development and evolution. For this purpose, we performed scRNA-seq in embryonic neocortical tissues of macaque. Our data support previous omics studies of embryonic macaque brain development [35] [36] [13]. We constructed a single-cell resolution transcriptomic atlas of the developing macaque neocortex, which we then used to identify dynamically expressed genes that likely contribute to the maturation of distinct neuron types, temporal expression patterns of neural stem cells, and the generation of oRGs.

The existence of oRGs has long been demonstrated in humans and primates through live imaging and immunostaining [37]. ORGs share similarities in their patterns of gene expression with vRGs (e.g., SOX2, PAX6, and NESTIN) but also specifically express several genes (such as MOXD1, HOPX, and FAM107A) [5, 38]. Recently, the molecular mechanisms associated with the generation and amplification of oRGs was shown to include signaling pathways such as the FGF-MAPK cascade, SHH, PTEN/AKT, and PDGF pathways, as well as proteins such as INSM, GPSM2, ASPM, TRNP1, ARHGAP11B, PAX6, and HIF1α. A number of these proteins were validated by genetic manipulation of their activities or expression in mouse, ferret, and marmoset [39-42]. The dynamically expressed genes and pseudotime trajectories from vRGs to oRGs presented in this work are in agreement with these previous reports. In addition to known regulators, we detected differential expression of ion channel regulators in oRGs, such as ATP1A2, ATP1B2, and SCN4B, suggesting that hyperpolarization and depolarization processes may participate in promoting or maintaining oRG populations. Indeed, hyperpolarization in mouse apical progenitors has been shown to promote the generation of IPCs and indirect neurogenesis [43], although it remains unclear if similar mechanisms contribute to development of the primate neocortex.

The sequential generation of cortical neurons has long been observed in both rodents and primates, and many transcription factors are known to drive post-mitotic specification of neuronal progeny, mainly in rodents. Here, we found genes that are conserved across species, such as FEZF2 and TBR1, that are involved in specifying deep-layer neurons in the macaque neocortex. From the slingshot pseudotime analysis, which reflected the expression pattern of genes during appropriate lineage trajectories, it suggests that some genes are lineage-restricted or even involved in lineage fate determination. For example, FEZF2, PPP1R1B (also known as

DARPP32), SOX5, LMO3, CELF4 may be involved in the fate specialization of deep neurons, while TMOD1, PID1, LINGO1, CRYM and MEF2C for upper layer neurons. Notably, FEZF2 [44] and SOX5 [45] have been confirmed associated with deep layer neuron specification in mice, while others, such as TMOD1, PID1 and LINGO1, have not been reported before, and their potential functions in the fate specification of cortical projection neurons need to be further explored. To date, only a very limited set of genes have been shown to function in specifying neuronal progeny in progenitor cells, which suggests that regulation in progenitors is more complex than in neurons, where a single transcription factor can trigger a transcriptional cascade to promote the maturation of neurons. In the current study, we found two types of genes (Dynamic type1 and 5) which have gradually altered expression patterns along with neurogenic stage, indicating a transition in cell state, in both vRGs and oRGs. In addition to these gradual transition genes, we also identified sets of genes (Dynamic type 2-4) with sharp changes in expression during neurogenesis which likely function in specifying different neuronal subtypes. These findings led us to postulate that a combination of biological processes and pathways change over time to coordinate gradual transitions from neural stem cell to daughter neuron. In previous studies, FOXG1 have been proved to play pivotal roles in controlling the cell cycle to meet the growing demands of the developing cerebral cortex[46-48]. Role of FOXG1 in the macaque vRG transcription factors regulatory network might contribute to primate stem cell pools enlargement and neocortex expansion.

This study also faces limitations. During the macaque embryonic neurogenesis, ventral radial glia cells generate excitatory neurons by direct and indirect neurogenesis. In general, the number of progenitor cells gradually decreases over the course of development, with progenitor cells generating deep neurons in the early stage of development and upper excitatory neurons in the later stages. We observed the anomalous disappearance of two subclusters of ENs (EN 5 and EN 11) from later stage samples (Figure S3). This disappearance may could also be potentially explained by the development of axons and dendrites in early-born neurons, which have higher morphological complexity and greater vulnerability to enzymatic digestion and lysis during sampling. Nonetheless, this transcriptomic atlas uncovers the molecular signatures of the main cell types and temporal shifts in gene expression of differentiating progenitor cells during neocortex layer formation, thus providing a global perspective into neocortical neurogenesis in macaque.

4. Materials and Methods

4.1. Animal

Animals and frozen tissue samples from prenatal rhesus macaque (Macaca mulatta) were provided by the National Medical Primate Center, Institute of Medical Biology, Chinese Academy of Medical Sciences. Timed pregnancy derived biological replicate specimens at each of prenatal developmental stages (E40, E50, E70, E80 and E90)[49] were profiled. These time points were selected to coincide with peak periods of neurogenesis for the different layers of cortex based on previous studies[49-50]. All animal procedures were conducted following the international standards and were approved in advance by the Ethics Committee on Laboratory Animals at IMBCAMS (Institute of Medical Biology, Chinese Academy of Medical Sciences).

4.2. Fetal brain sample details

We collected eight fetal brains of rhesus macaque (Macaca mulatta) at five timepoint (E40, E50, E70, E80, E90), and the parietal lobe cortex was dissected. Because of the different development time of rhesus monkey, embryonic cortex size and morphology are quite different. In order to ensure that the anatomical sites of each sample are roughly the same, we use the lateral groove as a reference to collect parietal lobe for single cell sequencing (as indicated by bright yellow in Figure S1A), and do not make a clear distinction between the different regional parts including primary somatosensory cortex and association cortices in the process of sampling.

4.3. Cell preparation

We followed the BD Cell Preparation Guide to wash, count, and concentrate cells in preparation for use in BD Rhapsody™ System Whole transcriptome analysis.

4.4. Single-cell RNA-seq data processing

Single-cell capture was achieved by random distribution of a single-cell suspension across >200,000 microwells through a limited dilution approach. Beads with oligonucleotide barcodes were added to saturation so that a bead was paired with a cell in a microwell. Cell-lysis buffer was added so that poly-adenylated RNA molecules hybridized to the beads. Beads were collected into a single tube for reverse transcription. Upon cDNA synthesis, each cDNA molecule was tagged on the 5′ end (that is, the 3′ end of a mRNA transcript) with a molecular index and cell label indicating its cell of origin. Whole transcriptome libraries were prepared using the BD Resolve single-cell whole-transcriptome amplification workflow. In brief, second strand cDNA was synthesized, followed by ligation of the adaptor for universal amplification. Eighteen cycles of PCR were used to amplify the adaptor- ligated cDNA products. Sequencing libraries were prepared using random priming PCR of the whole-transcriptome amplification products to enrich the 3′ end of the transcripts linked with the cell label and molecular indices. Sequencing was performed with Illumina NovaSeq 6000 according to the manufacturer's

instructions. The filtered reads were aligned to the macaque reference genome file with Bowtie2 (v2.2.9). Macaque reference genome was downloaded from Ensemble database (ftp://ftp.ensembl.org/pub/release100/fasta/macaca_mulatta/dna/Macaca_mulatta.Mmul_10.dna.toplevel.fa.gz,ftp://ftp.ensembl.org/pub/release-100/gtf/macaca_mulatta/Macaca_mulatta.Mmul_10.100.gtf.gz). We analyzed the sequencing data following BD official pipleline and obtained the cell-gene expression matrix file.

4.5. QC and Data analysis

To mitigate the effect of over-estimation of molecules from PCR and sequencing errors, we used the Unique Molecular Identifier (UMI) adjustment algorithms recursive substitution error correction (RSEC) and distribution-based error correction (DBEC) which were contained in the BD Rhapsody™ pipeline contains. Quality control was applied based on the detected gene number and the percentage of counts originating from mitochondrial RNA, ribosomal RNA, and hemoglobin gene per cell. Then cells were filtered to retain only higher quality (with less than 7.5% mitochondrial gene counts, less than 5.5% ribosomal gene counts, and detected genes above 400 and less than 6000). Additionally, we identified cell doublets using Scrublet v0.142 with default parameters and removed doublets before data analysis. After quality control, a total of 59,045 cells remained for subsequent analysis.

To integrate cells into a shared space from different datasets for unsupervised clustering, we used SCTransform workflow in R packages Seurat[51] V4.0.3 to normalize the single-cell RNA-seq data from different samples. We identified the variable features of each donor sequencing data using the “FindVariableFeatures” function. A consensus list of 2000 variable genes was then formed by detecting the greatest recovery rates genes across samples. Mitochondrial and ribosomal genes were not included in the variable gene list. Next, we used the SCTransform workflow in Seurat to normalize the single cell-seq data from different samples. During normalization, we also removed confounding sources of variation, including mitochondrial mapping and ribosomal mapping percentages. Standard RPCA workflow was used to perform the integration.

With a “resolution” of 0.5 upon running “FindClusters,” we distinguished major cell types of nerve cells and non-nerve cells in the umap according to known markers, including radial glia cell, outer radial glia cell, intermediate progenitor cell, excitatory neuron, inhibitory neuron, astrocyte, oligodendrocyte, Cajal-Retzius cell, microglia, endothelial cell, meningeal cells, and blood cells, which were distinguished on the first level. By following a similar pipeline, subclusters were identified. We finalized the resolution parameter on the FindClusters function once the cluster numbers did not increase when we increased the resolution. Then, we checked

the differentially expressed genes (DEGs) between each of the clusters using the "FindAllMarkers" and "FindMarkers" function with logfc.threshold=0.25.

4.6. Construction of single-cell developmental trajectory

Single-cell developmental lineage trajectories construction and discovering trajectory transitions were performed using Slingshot(V 2.2.0) [52], from the Dyno (V 0.1.2) [53] platform, with PCA dimensionality reduction plot results. The direction of the developmental trajectory was adjusted by reference to the verified relevant studies.

4.7. Cross-species transcriptome data integration and analysis

Mouse datasets were downloaded from the Gene Expression Omnibus (GEO SuperSeries GSE153164) and at the Single Cell Portal: https://singlecell.broadinstitute.org/single_cell/study/SCP1290/molecular-logic-of-cellular-diversification-in-the-mammalian-cerebral-cortex. Human datasets were download from the Gene Expression Omnibus (GEO) under the accession number GSE104276 and GSE104276. We created a human database by combing the two published human embryonic cortical development datasets of 7 to 21 postconceptional weeks (PCW) into one containing cell numbers comparable to our macaque and published mouse dataset using the single cell transcriptome analysis pipeline of R package Seurat. We used the biomaRt package to convert the gene symbols in the mouse and macaque expression matrices into their human homologs. To perform cross-species analysis, we used the LIGER [31] method to integrate the macaque single-cell dataset with the mouse and human datasets. We ran "optimizeALS" function in rliger to perform integrative non-negative matrix factorization(iNMF) on the scaled species-integrated dataset (k = 20, lambda = 5, max.iters = 30). Species-integrated UMAP showed that all major cell types were well integrated between the three species (Figure 5A). We used the "calcAlignment" function in rliger to quantify how well-aligned datasets are and got the metric of 0.891231545756746, which was very close to "1". This suggested LIGER well-integrated cross-species datasets.

4.8. Transcription Factor Regulatory Network Analysis

We inferred regulon activities for vRG cells of macaque and mouse using pySCENIC [56] workflow (https://pyscenic.readthedocs.io/en/latest/tutorial.html) with default parameters. Since there is no transcriptional regulator database for macaque, macaque's genome is similar to the human genome. When we analyzed the macaque dataset, we used the human transcriptional regulator database for pySCENIC analysis. GRNboost2 in scenic was used to infer gene regulatory networks based on co-expression patterns. RcisTarget[54] was used to analyze the TF-motif enrichment and direct targets among different species databases. The

activity of regulon for each developmental time point was identified using AUCell, and the top enriched activated TFs were ranked by -log10(p_value). Lastly, we evaluated the protein-protein interaction networks analysis of TFs and their target genes by the String[55] (V11.5) and visualized by Cytoscape[56] (version 3.8.1).

4.9. Mfuzz Clustering

Firstly we picked up vRG cells in each species dataset and screened the differentially expressed genes (DEGs) between adjacent development timepoints using "FindMarkers" function (with min.pct = 0.25, logfc.threshold = 0.25). After separate normalization of the DEG expression matrix from different species datasets, we use "standardise" function of mfuzz to perform standardization. The DEGs of vRG in each species were grouped into five different clusters using Mfuzz package in R with fuzzy c-means algorithm[57].

4.10. Transcription Factor and RNA binding protein analysis

Mouse, macaque, and human transcription factor(TF) gene lists were downloaded from AnimalTFDB (http://bioinfo.life.hust.edu.cn/AnimalTFDB/). Mouse, macaque, and human RNA binding protein(RBP) gene lists were downloaded from EuRBPDB (http://EuRBPDB.syshospital.org). Heatmaps of the TFs and RBPs expression profiles were generated with normalized and standardized DEGs expression matrixes from mfuzz.

Acknowledgments

Funding: This work was supported by CAMS Innovation Fund for Medical Sciences (CIFMS, 2021-I2M-1-024, 2021-I2M-1-019) and National Science and Technology Innovation 2030 Major Program 2021ZD0200900.Author Contributions: X.P. and B.Q. supervised the research. P.S. and Y.C. conceived of the experimental design and bioinformatics analysis. Y.Z., Z.H. and S.L. established the macaque embryonic developmental model. W.L. performed the anatomical analysis of monkeys and related tests. L.X. and J.Z. contributed to drafting or revising the manuscript and figures. Z.Y. and L.X. performed single-cell transcriptome data analysis. Competing interests: The authors declare that they have no competing interests.

Data availability: The raw sequence data reported in this paper have been deposited in the Genome Sequence Archive in National Genomics Data Center, China National Center for Bioinformation / Beijing Institute of Genomics, Chinese Academy of Sciences (GSA: CRA007718) that are publicly accessible at https://ngdc.cncb.ac.cn/gsa.

Author Contributions: X.P. and B.Q. supervised the research. P.S. and Y.C. conceived of the experimental design and bioinformatics analysis. Y.Z. , Z.H. and S.L. established the macaque embryonic developmental model. W.L. performed the anatomical analysis of monkeys and related tests. L.X. and J.Z. contributed to drafting or revising the manuscript and figures. Z.Y. and L.X. performed single-cell transcriptome data analysis.

Competing Interest Statement: The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

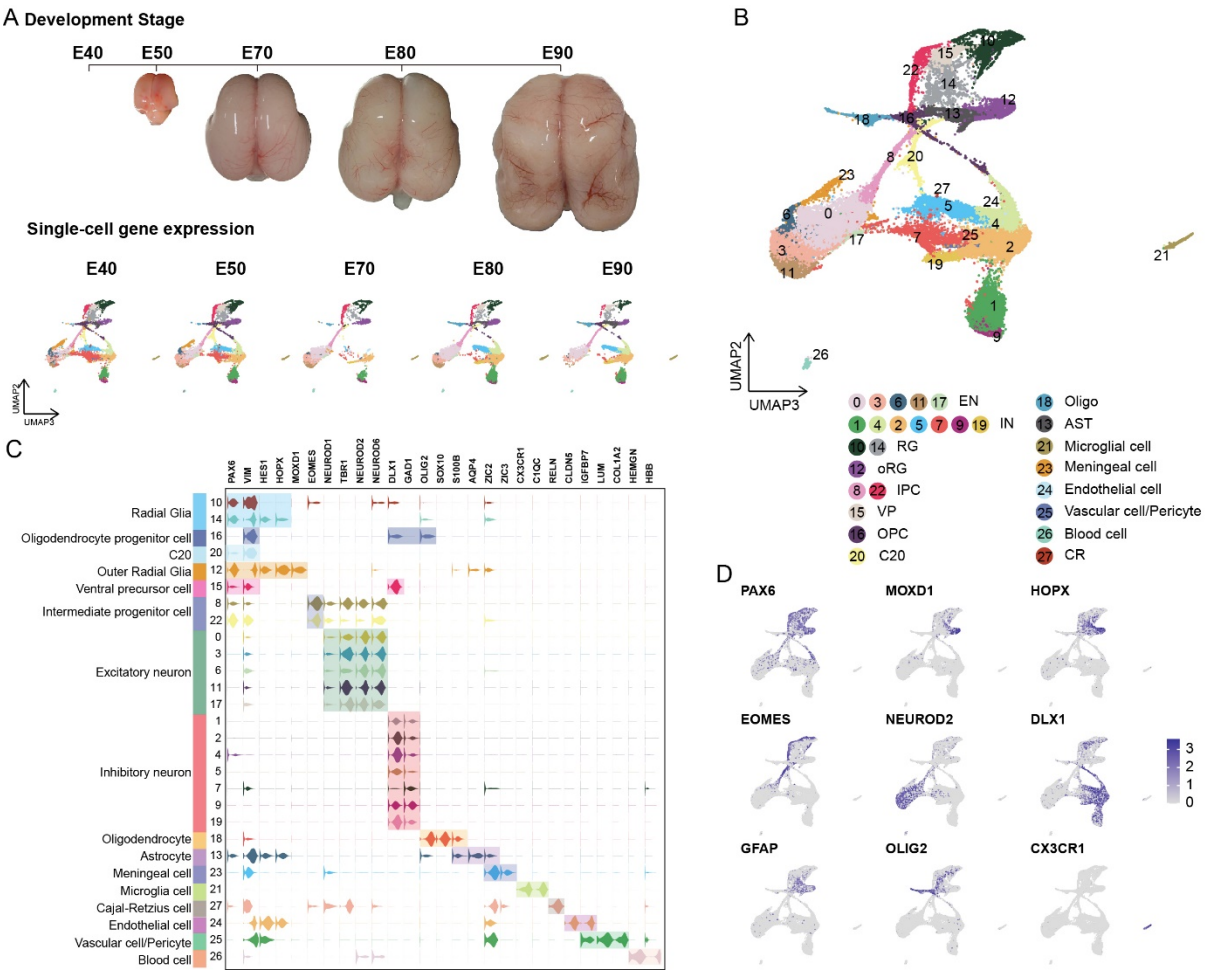

Figure 1. Cell types in macaque embryonic and fetal brain development. (A) Schematic diagram of sample collecting and data analyzing. We collected the parietal lobe from the embryos across developmental stages from E40 to E90. UMAP visualization of snRNA-seq data from individual time points. (B) The transcriptome data of single cells were collected and used to do clustering using Seurat. Visualization of major types of cells using UMAP. Dots, individual cells; color, clusters. (C) Violin plot of molecular markers for annotating cell types. (D) The expressions of the classic marker genes for each cell type were plotted to UMAP visualization. Light grey, no expression; Dark blue, relative expression.

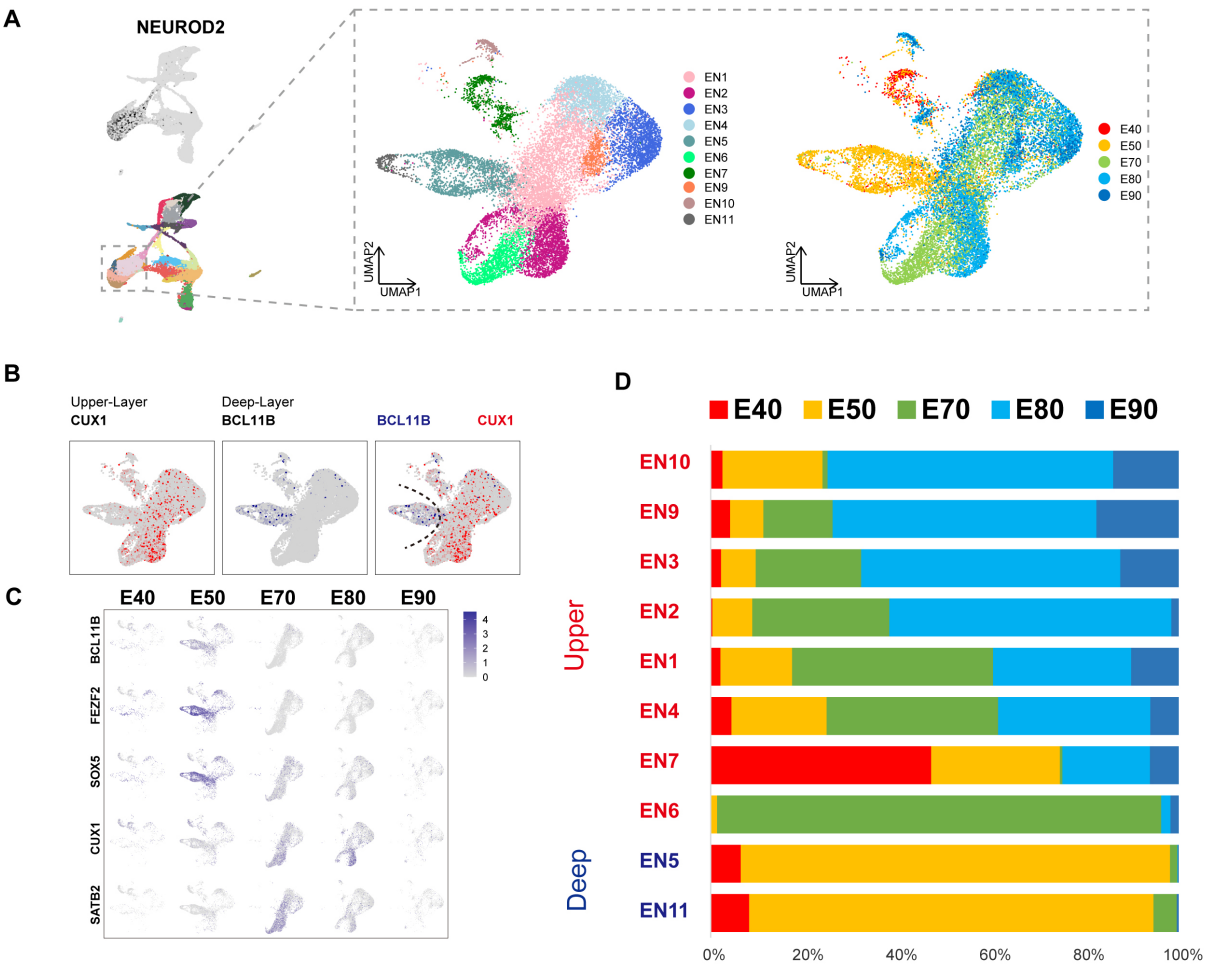

Figure 2. Excitatory neuron subclusters in the developing macaque cerebral cortex. (A) Left, Clustering of excitatory neuron subclusters collected at all time points, visualized via UMAP. Cells are colored according to subcluster identities (left) timepoint of collection(right). (B) Differentially expression of deep-layer marker BCL11B and upper layer marker CUX1 is highlighted, and the combined excitatory neuron populations (all time points). (C) Excitatory neuron subclusters UMAP plot shows the expression of classic markers for deep layers (BCL11B, FEZF2, SOX5) and upper layers (CUX1, SATB2) present each time point. (D) Proportion of different excitatory neuron subtclusters corresponding to excitatory neurons in each time point.

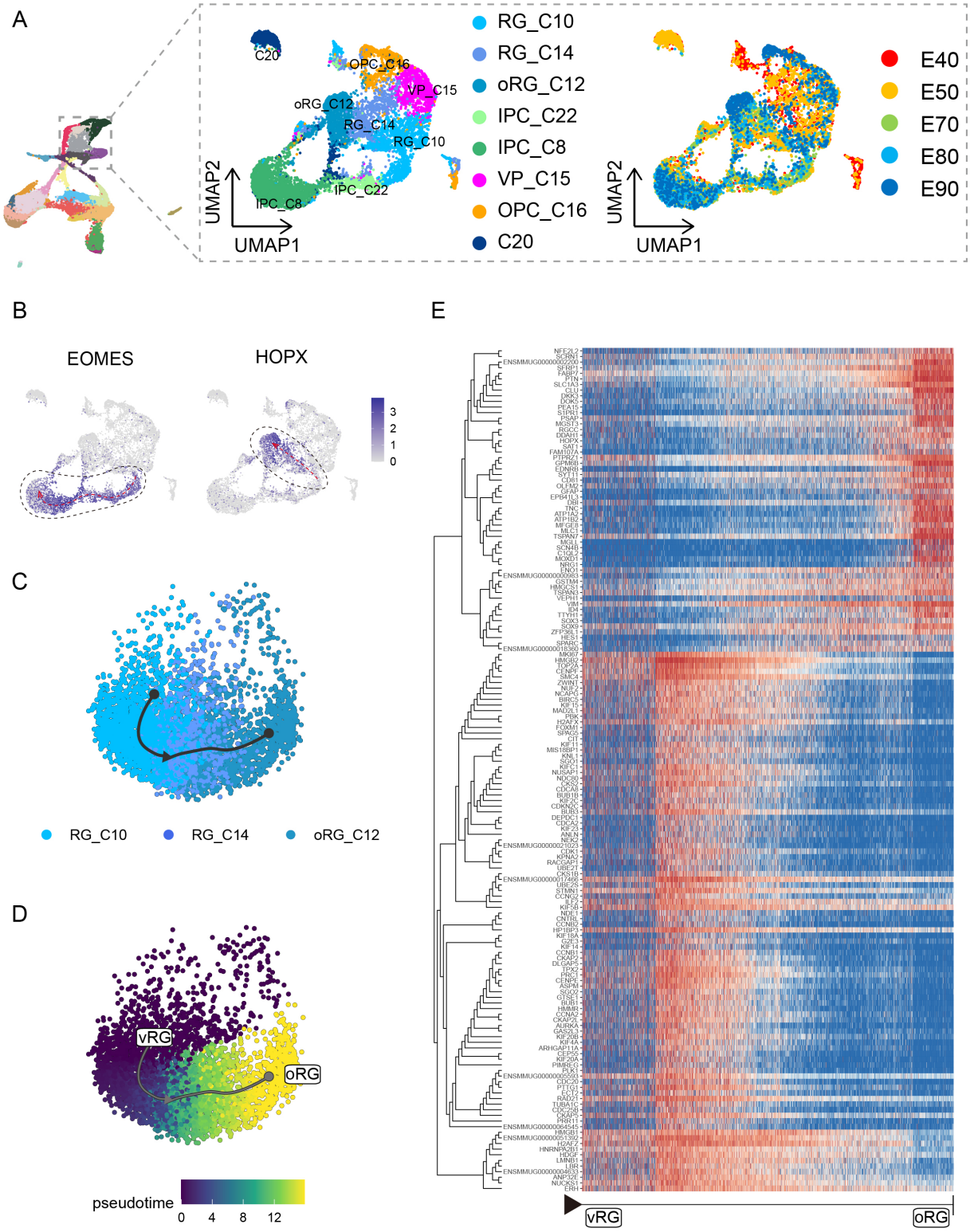

Figure 3. Cell diversity and regulation of progenitor cells in the macaque cortical neurogenesis.

(A) UMAP shows eight progenitor clusters and cell annotation. Left, cells are colored according to seurat clusters; right, cells are colored according to the collection time point. (B) Feature plot of outer radial glia marker genes HOPX shows higher expression in C10-C14-C12 (left). Feature plots of Intermediate progenitor cell marker gene EOMES show higher expression in C10-C22-C8 (right). (C) Pseudotime analysis by Slingshot of HOPX-positive cells (C10-C14-C12). The Slingshot result with the lines indicating the trajectories of lineages and the arrows indicating directions of the pseudotime. Cells are colored according to cell (C) and pesudotime (D). Dots: single cells; colors: cluster and subcluster identity. (E) Heatmap shows the relative expression of top150 genes displaying significant changes along the pseudotime axis of RG to oRG (C10-C14-C12). The columns represent the cells being ordered along the pseudotime axis.

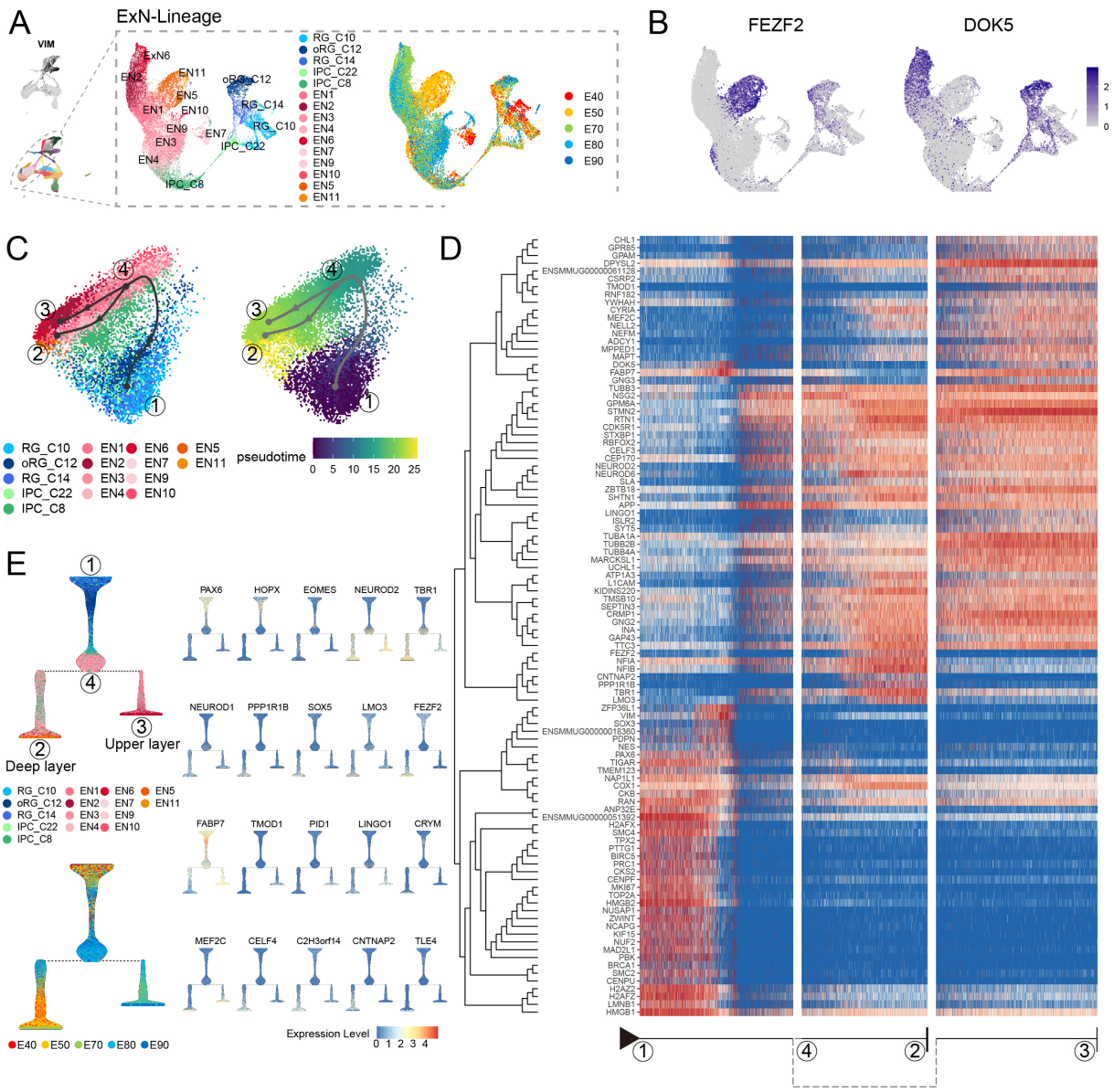

Figure 4. Transcriptional regulation of excitatory neuron lineage during embryonic cortical neurogenesis. (A) UMAP shows the alignment of macaque cortical NPCs, IPCs, and excitatory neurons. Left, cells are colored according to cell annotation; right, cells are colored according to the time point of collection. (B) Dot plot showing the marker genes for the deep layer excitatory neuron (FEZF2) and upper-layer excitatory neuron (DOK5). Light grey, no expression; Dark blue, relative expression. (C) Pseudotime analysis by Slingshot projected on PCA plot of RGCs, oRGCs, IPCs, and excitatory neuron subclusters. The Slingshot result with the lines indicating the trajectories of lineages and the arrows indicating directions of the pseudotime. Dots: single cells; colors: cluster and subcluster identity. Framed numbers marked

start point, endpoint, and essential nodes of Slingshot inference trajectory. Framed numbers"1" was excitatory neuron lineage trajectory start point (C10). Framed number"4" marked immature neurons. Framed numbers "2" and "3" marked deep-layer neurons and upper-layer neurons. Cells are colored according to cell annotation and pseudotime. (D) Heatmap shows the relative expression of top100 genes displaying significant changes along the pseudotime axis of each lineage branch. The columns represent the cells being ordered along the pseudotime axis. (E) Left, Slingshot branching tree related to Slingshot pseudotime analysis in C. The root is E40 earliest RG (C10), tips are deep layer excitatory neurons generated at the early stage (E40, E50), and upper-layer excitatory neurons generated at the later stage (E70, E80, E90). Right, Branching trees showing the expression of marker genes of apical progenitors (PAX6), outer radial glia cells (HOPX), intermediate progenitors (EOMES), and excitatory neurons (NEUROD2), as well as genes in Figure 3E, including callosal neurons (SATB2, CUX2), deeper layer neurons (SOX5, FEZF2), corticofugal neurons (FEZF2, TLE4). There is a sequential progression of radial glia cells, intermediate progenitors, and excitatory neurons.

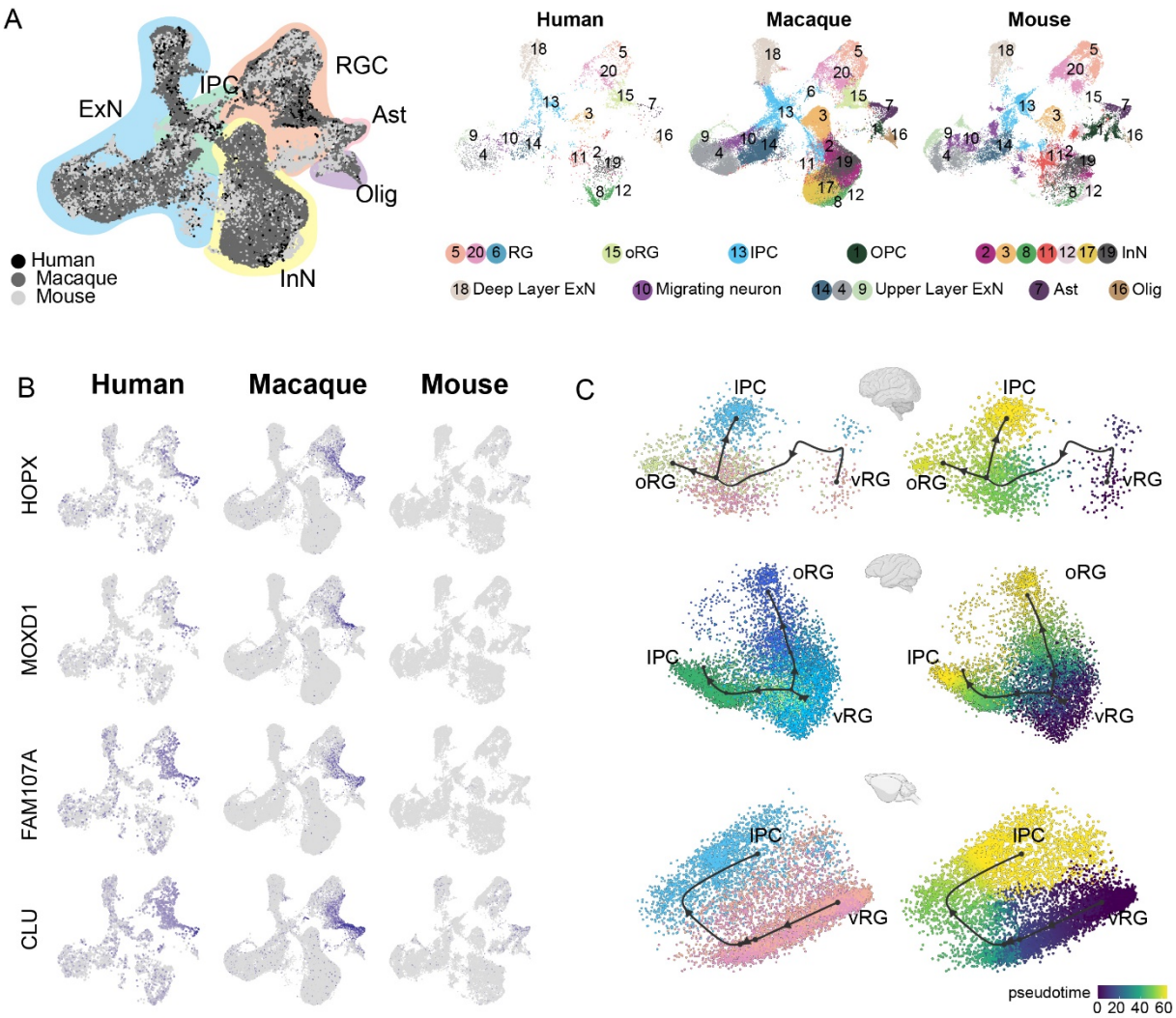

Figure 5. Integration of human, macaque and mouse single-cell dataset reveals conserved and divergent progenitor cell types. (A) Left: UMAP plot of cross-species integrated single cell transcriptome data with liger. Colors represent different major cell types (Black: Human dataset; Dark grey: macaque dataset; Liger grey: mouse). Right: The UMAP plot of each dataset, colored by ligercluster. (B) The expressions of the classic oRG marker genes were plotted to UMAP visualization. Light grey, no expression; Dark blue, relative expression. (C) Comparison of vRG→oRG and vRG→IPC developmental trajectories between human, macaque, and mouse.

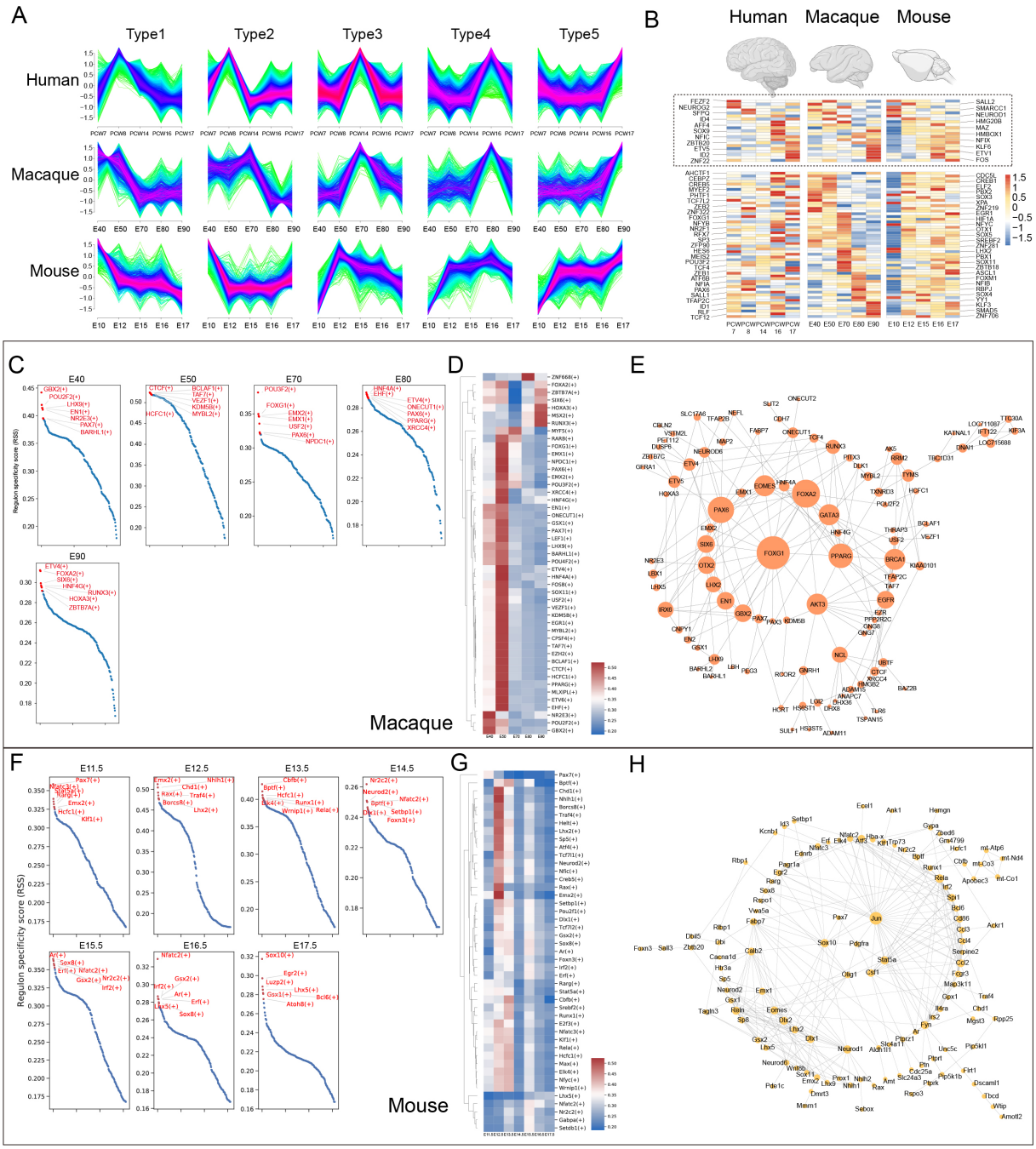

Figure 6. The cell-intrinsic patterns of transcriptional regulation comparative analysis responsible for generation of vRGs. (A) Normalized gene expressions of human, macaque and mouse are distributed within five sequential AP transcriptional states. (B) Temporal expression heatmap of homologous transcription factor genes of among human, macaque, and mouse temporal expression heatmap. (The TFs genes in the dashed boxes showed similar temporal expression patterns across species. (C) and (F) Top 7 regulon specificity score gene at macaque and mouse each embryonic cortex development timepoint. (D) and (G) Gene-expression heatmap of top 7 regulon specificity score gene in (C) and (F) Color scale: red, high expression;

blue, low expression. (E) and (H) Protein-protein interactions analysis of macaque and mouse potential transcriptional regulation in the macaque and mouse cortex neurogenesis using the STRING database (http://string-db.org). Nodes are the top 7 regulon specificity score TFs genes at each embryonic development timepoint and their top 5 target genes analysis by SCENIC. Node size is positively correlated with the number of directed edges.

References

[1] D. Jabaudon, *Nat Commun* **2017**, *8*, 16042, https://doi.org/10.1038/ncomms16042.

[2] P. Rakic, *Science (New York, N.Y.)* **1974**, *183* (4123), 425.

[3] L. Telley, G. Agirman, J. Prados, N. Amberg, S. Fièvre, P. Oberst, G. Bartolini, I. Vitali, C. Cadilhac, S. Hippenmeyer, L. Nguyen, A. Dayer, D. Jabaudon, *Science (New York, N.Y.)* **2019**, *364* (6440), https://doi.org/10.1126/science.aav2522.

[4] M. Betizeau, V. Cortay, D. Patti, S. Pfister, E. Gautier, A. Bellemin-Ménard, M. Afanassieff, C. Huissoud, R. J. Douglas, H. Kennedy, C. Dehay, *Neuron* **2013**, *80* (2), 442, https://doi.org/10.1016/j.neuron.2013.09.032.

[5] Alex A. Pollen, Tomasz J. Nowakowski, J. Chen, H. Retallack, C. Sandoval-Espinosa, Cory R. Nicholas, J. Shuga, Siyuan J. Liu, Michael C. Oldham, A. Diaz, Daniel A. Lim, Anne A. Leyrat, Jay A. West, Arnold R. Kriegstein, *Cell* **2015**, *163* (1), 55, https://doi.org/https://doi.org/10.1016/j.cell.2015.09.004.

[6] M. Penisson, J. Ladewig, R. Belvindrah, F. Francis, *Front Cell Neurosci* **2019**, *13*, 381, https://doi.org/10.3389/fncel.2019.00381.

[7] M. D. Luecken, F. J. Theis, *Mol Syst Biol* **2019**, *15* (6), e8746, https://doi.org/10.15252/msb.20188746.

[8] X. Ruan, B. Kang, C. Qi, W. Lin, J. Wang, X. Zhang, *Proc Natl Acad Sci U S A* **2021**, *118* (10), https://doi.org/10.1073/pnas.2018866118.

[9] D. J. Di Bella, E. Habibi, R. R. Stickels, G. Scalia, J. Brown, P. Yadollahpour, S. M. Yang, C. Abbate, T. Biancalani, E. Z. Macosko, F. Chen, A. Regev, P. Arlotta, *Nature* **2021**, *595* (7868), 554, https://doi.org/10.1038/s41586-021-03670-5.

[10] A. E. Trevino, F. Müller, J. Andersen, L. Sundaram, A. Kathiria, A. Shcherbina, K. Farh, H. Y. Chang, A. M. Pașca, A. Kundaje, S. P. Pașca, W. J. Greenleaf, *Cell* **2021**, *184* (19), https://doi.org/10.1016/j.cell.2021.07.039.

[11] T. J. Nowakowski, A. Bhaduri, A. A. Pollen, B. Alvarado, M. A. Mostajo-Radji, E. Di Lullo, M. Haeussler, C. Sandoval-Espinosa, S. J. Liu, D. Velmeshev, J. R. Ounadjela, J. Shuga, X. Wang, D. A. Lim, J. A. West, A. A. Leyrat, W. J. Kent, A. R. Kriegstein, *Science (New York, N.Y.)* **2017**, *358* (6368), 1318, https://doi.org/10.1126/science.aap8809.

[12] U. C. Eze, A. Bhaduri, M. Haeussler, T. J. Nowakowski, A. R. Kriegstein, *Nature Neuroscience* **2021**, *24* (4), 584, https://doi.org/10.1038/s41593-020-00794-1.

[13] T. E. Bakken, J. A. Miller, S.-L. Ding, S. M. Sunkin, K. A. Smith, L. Ng, A. Szafer, R. A. Dalley, J. J. Royall, T. Lemon, S. Shapouri, K. Aiona, J. Arnold, J. L. Bennett, D. Bertagnolli, K. Bickley, A. Boe, K. Brouner, S. Butler, E. Byrnes, S. Caldejon, A. Carey, S. Cate, M. Chapin, J. Chen, N. Dee, T. Desta, T. A. Dolbeare, N. Dotson, A. Ebbert, E. Fulfs, G. Gee, T. L. Gilbert, J. Goldy, L. Gourley, B. Gregor, G. Gu, J. Hall, Z. Haradon, D. R. Haynor, N. Hejazinia, A. Hoerder-Suabedissen, R. Howard, J. Jochim, M. Kinnunen, A. Kriedberg, C. L. Kuan, C. Lau, C.-K. Lee, F. Lee, L. Luong, N. Mastan, R. May, J. Melchor, N. Mosqueda, E. Mott, K. Ngo,

J. Nyhus, A. Oldre, E. Olson, J. Parente, P. D. Parker, S. Parry, J. Pendergraft, L. Potekhina, M. Reding, Z. L. Riley, T. Roberts, B. Rogers, K. Roll, D. Rosen, D. Sandman, M. Sarreal, N. Shapovalova, S. Shi, N. Sjoquist, A. J. Sodt, R. Townsend, L. Velasquez, U. Wagley, W. B. Wakeman, C. White, C. Bennett, J. Wu, R. Young, B. L. Youngstrom, P. Wohnoutka, R. A. Gibbs, J. Rogers, J. G. Hohmann, M. J. Hawrylycz, R. F. Hevner, Z. Molnár, J. W. Phillips, C. Dang, A. R. Jones, D. G. Amaral, A. Bernard, E. S. Lein, *Nature* **2016**, *535* (7612), 367, https://doi.org/10.1038/nature18637.

[14] A. Bernard, L. S. Lubbers, K. Q. Tanis, R. Luo, A. A. Podtelezhnikov, E. M. Finney, M. M. E. McWhorter, K. Serikawa, T. Lemon, R. Morgan, C. Copeland, K. Smith, V. Cullen, J. Davis-Turak, C.-K. Lee, S. M. Sunkin, A. P. Loboda, D. M. Levine, D. J. Stone, M. J. Hawrylycz, C. J. Roberts, A. R. Jones, D. H. Geschwind, E. S. Lein, *Neuron* **2012**, *73* (6), 1083, https://doi.org/10.1016/j.neuron.2012.03.002.

[15] H. Zeng, E. H. Shen, J. G. Hohmann, S. W. Oh, A. Bernard, J. J. Royall, K. J. Glattfelder, S. M. Sunkin, J. A. Morris, A. L. Guillozet-Bongaarts, K. A. Smith, A. J. Ebbert, B. Swanson, L. Kuan, D. T. Page, C. C. Overly, E. S. Lein, M. J. Hawrylycz, P. R. Hof, T. M. Hyde, J. E. Kleinman, A. R. Jones, *Cell* **2012**, *149* (2), 483, https://doi.org/10.1016/j.cell.2012.02.052.

[16] R. F. Hevner, *J Anat* **2019**, *235* (3), 616, https://doi.org/10.1111/joa.12939.

[17] J. C. Silbereis, E. J. Huang, S. A. Back, D. H. Rowitch, *Disease Models & Mechanisms* **2010**, *3* (11-12), 678, https://doi.org/10.1242/dmm.002915.

[18] M. Lee, J. Yoon, H. Song, B. Lee, D. T. Lam, J. Yoon, K. Baek, H. Clevers, Y. Jeong, *Dev Biol* **2017**, *424* (1), 62, https://doi.org/10.1016/j.ydbio.2017.02.010.

[19] J. Tsyporin, D. Tastad, X. Ma, A. Nehme, T. Finn, L. Huebner, G. Liu, D. Gallardo, A. Makhamreh, J. M. Roberts, S. Katzman, N. Sestan, S. K. McConnell, Z. Yang, S. Qiu, B. Chen, *Cell Rep* **2021**, *35* (12), 109269, https://doi.org/10.1016/j.celrep.2021.109269.

[20] Z. Li, W. A. Tyler, E. Zeldich, G. Santpere Baró, M. Okamoto, T. Gao, M. Li, N. Sestan, T. F. Haydar, *Science Advances* **2020**, *6* (45), https://doi.org/10.1126/sciadv.abd2068.

[21] Z. He, D. Han, O. Efimova, P. Guijarro, Q. Yu, A. Oleksiak, S. Jiang, K. Anokhin, B. Velichkovsky, S. Grunewald, P. Khaitovich, *Nat Neurosci* **2017**, *20* (6), 886, https://doi.org/10.1038/nn.4548.

[22] J. Liu, W. Liu, L. Yang, Q. Wu, H. Zhang, A. Fang, L. Li, X. Xu, L. Sun, J. Zhang, F. Tang, X. Wang, *Cell Stem Cell* **2017**, *21* (5), 635, https://doi.org/10.1016/j.stem.2017.08.013.

[23] M. B. Johnson, X. Sun, A. Kodani, R. Borges-Monroy, K. M. Girskis, S. C. Ryu, P. P. Wang, K. Patel, D. M. Gonzalez, Y. M. Woo, Z. Yan, B. Liang, R. S. Smith, M. Chatterjee, D. Coman, X. Papademetris, L. H. Staib, F. Hyder, J. B. Mandeville, P. E. Grant, K. Im, H. Kwak, J. F. Engelhardt, C. A. Walsh, B. I. Bae, *Nature* **2018**, *556* (7701), 370, https://doi.org/10.1038/s41586-018-0035-0.

[24] Z. Guo, M. Chen, Y. Chao, C. Cai, L. Liu, L. Zhao, L. Li, Q. R. Bai, Y. Xu, W. Niu, L. Shi, Y. Bi, D. Ren, F. Yuan, S. Shi, Q. Zeng, K. Han, Y. Shi, S. Bian, G. He, *EMBO Rep* **2021**, *22* (9), e51781, https://doi.org/10.15252/embr.202051781.

[25] X. Zhou, S. Zhong, H. Peng, J. Liu, W. Ding, L. Sun, Q. Ma, Z. Liu, R. Chen, Q. Wu, X. Wang, *Nat Commun* **2020**, *11* (1), 4063, https://doi.org/10.1038/s41467-020-17890-2.

[26] Q. Weng, J. Wang, J. Wang, D. He, Z. Cheng, F. Zhang, R. Verma, L. Xu, X. Dong, Y. Liao, X. He, A. Potter, L. Zhang, C. Zhao, M. Xin, Q. Zhou, B. J. Aronow, P. J. Blackshear, J. N. Rich, Q. He, W. Zhou, M. L. Suva, R. R. Waclaw, S. S. Potter, G. Yu, Q. R. Lu, *Cell Stem Cell* **2019**, *24* (5), 707, https://doi.org/10.1016/j.stem.2019.03.006.

[27] A. Gadea, S. Schinelli, V. Gallo, *J Neurosci* **2008**, *28* (10), 2394, https://doi.org/10.1523/JNEUROSCI.5652-07.2008.

[28] L. Telley, G. Agirman, J. Prados, N. Amberg, S. Fievre, P. Oberst, G. Bartolini, I. Vitali, C. Cadilhac, S. Hippenmeyer, L. Nguyen, A. Dayer, D. Jabaudon, *Science* **2019**, *364* (6440), https://doi.org/10.1126/science.aav2522.

[29] Y. Lin, J. Yang, Z. Shen, J. Ma, B. D. Simons, S. H. Shi, *Curr Opin Neurobiol* **2021**, *66*, 144, https://doi.org/10.1016/j.conb.2020.10.017.
[30] P. Shu, C. Wu, X. Ruan, W. Liu, L. Hou, H. Fu, M. Wang, C. Liu, Y. Zeng, P. Chen, B. Yin, J. Yuan, B. Qiang, X. Peng, W. Zhong, *Dev Cell* **2019**, *49* (5), 764, https://doi.org/10.1016/j.devcel.2019.04.017.
[31] J. D. Welch, V. Kozareva, A. Ferreira, C. Vanderburg, C. Martin, E. Z. Macosko, *Cell* **2019**, *177* (7), https://doi.org/10.1016/j.cell.2019.05.006.
[32] S. Zhong, S. Zhang, X. Fan, Q. Wu, L. Yan, J. Dong, H. Zhang, L. Li, L. Sun, N. Pan, X. Xu, F. Tang, J. Zhang, J. Qiao, X. Wang, *Nature* **2018**, *555* (7697), 524, https://doi.org/10.1038/nature25980.
[33] X. Fan, Y. Fu, X. Zhou, L. Sun, M. Yang, M. Wang, R. Chen, Q. Wu, J. Yong, J. Dong, L. Wen, J. Qiao, X. Wang, F. Tang, *Science Advances* **2020**, *6* (34), eaaz2978, https://doi.org/10.1126/sciadv.aaz2978.
[34] X. Wang, J.-W. Tsai, B. LaMonica, A. R. Kriegstein, *Nature Neuroscience* **2011**, *14* (5), 555, https://doi.org/10.1038/nn.2807.
[35] Y. Zhu, A. M. M. Sousa, T. Gao, M. Skarica, M. Li, G. Santpere, P. Esteller-Cucala, D. Juan, L. Ferrández-Peral, F. O. Gulden, M. Yang, D. J. Miller, T. Marques-Bonet, Y. Imamura Kawasawa, H. Zhao, N. Sestan, *Science (New York, N.Y.)* **2018**, *362* (6420), https://doi.org/10.1126/science.aat8077.
[36] X. Luo, Y. Liu, D. Dang, T. Hu, Y. Hou, X. Meng, F. Zhang, T. Li, C. Wang, M. Li, H. Wu, Q. Shen, Y. Hu, X. Zeng, X. He, L. Yan, S. Zhang, C. Li, B. Su, *Cell* **2021**, *184* (3), https://doi.org/10.1016/j.cell.2021.01.001.
[37] D. V. Hansen, J. H. Lui, P. R. Parker, A. R. Kriegstein, *Nature* **2010**, *464* (7288), 554, https://doi.org/10.1038/nature08845.
[38] T. J. Nowakowski, A. A. Pollen, C. Sandoval-Espinosa, A. R. Kriegstein, *Neuron* **2016**, *91* (6), 1219, https://doi.org/10.1016/j.neuron.2016.09.005.
[39] X. Wang, J. W. Tsai, B. LaMonica, A. R. Kriegstein, *Nat Neurosci* **2011**, *14* (5), 555, https://doi.org/10.1038/nn.2807.
[40] N. Kalebic, C. Gilardi, M. Albert, T. Namba, K. R. Long, M. Kostic, B. Langen, W. B. Huttner, *Elife* **2018**, *7*, https://doi.org/10.7554/eLife.41241.
[41] L. Xing, A. Kubik-Zahorodna, T. Namba, A. Pinson, M. Florio, J. Prochazka, M. Sarov, R. Sedlacek, W. B. Huttner, *EMBO J* **2021**, *40* (13), e107093, https://doi.org/10.15252/embj.2020107093.
[42] M. Heide, C. Haffner, A. Murayama, Y. Kurotaki, H. Shinohara, H. Okano, E. Sasaki, W. B. Huttner, *Science (New York, N.Y.)* **2020**, *369* (6503), 546, https://doi.org/10.1126/science.abb2401.
[43] I. Vitali, S. Fievre, L. Telley, P. Oberst, S. Bariselli, L. Frangeul, N. Baumann, J. J. McMahon, E. Klingler, R. Bocchi, J. Z. Kiss, C. Bellone, D. L. Silver, D. Jabaudon, *Cell* **2018**, *174* (5), 1264, https://doi.org/10.1016/j.cell.2018.06.036.
[44] B. J. Molyneaux, P. Arlotta, T. Hirata, M. Hibi, J. D. Macklis, *Neuron* **2005**, *47* (6), 817, https://doi.org/10.1016/j.neuron.2005.08.030.
[45] D. P. Leone, K. Srinivasan, B. Chen, E. Alcamo, S. K. McConnell, *Curr Opin Neurobiol* **2008**, *18* (1), 28, https://doi.org/10.1016/j.conb.2008.05.006.
[46] M. Brancaccio, C. Pivetta, M. Granzotto, C. Filippis, A. Mallamaci, *Stem Cells (Dayton, Ohio)* **2010**, *28* (7), 1206, https://doi.org/10.1002/stem.443.
[47] J. Seoane, H.-V. Le, L. Shen, S. A. Anderson, J. Massagué, *Cell* **2004**, *117* (2), 211.
[48] S. Xuan, C. A. Baptista, G. Balas, W. Tao, V. C. Soares, E. Lai, *Neuron* **1995**, *14* (6), 1141.
[49] T. E. Bakken, J. A. Miller, S. L. Ding, S. M. Sunkin, K. A. Smith, L. Ng, A. Szafer, R. A. Dalley, J. J. Royall, T. Lemon, S. Shapouri, K. Aiona, J. Arnold, J. L. Bennett, D. Bertagnolli, K. Bickley, A. Boe, K. Brouner, S. Butler, E. Byrnes, S. Caldejon, A. Carey, S. Cate, M. Chapin,

J. Chen, N. Dee, T. Desta, T. A. Dolbeare, N. Dotson, A. Ebbert, E. Fulfs, G. Gee, T. L. Gilbert, J. Goldy, L. Gourley, B. Gregor, G. Gu, J. Hall, Z. Haradon, D. R. Haynor, N. Hejazinia, A. Hoerder-Suabedissen, R. Howard, J. Jochim, M. Kinnunen, A. Kriedberg, C. L. Kuan, C. Lau, C. K. Lee, F. Lee, L. Luong, N. Mastan, R. May, J. Melchor, N. Mosqueda, E. Mott, K. Ngo, J. Nyhus, A. Oldre, E. Olson, J. Parente, P. D. Parker, S. Parry, J. Pendergraft, L. Potekhina, M. Reding, Z. L. Riley, T. Roberts, B. Rogers, K. Roll, D. Rosen, D. Sandman, M. Sarreal, N. Shapovalova, S. Shi, N. Sjoquist, A. J. Sodt, R. Townsend, L. Velasquez, U. Wagley, W. B. Wakeman, C. White, C. Bennett, J. Wu, R. Young, B. L. Youngstrom, P. Wohnoutka, R. A. Gibbs, J. Rogers, J. G. Hohmann, M. J. Hawrylycz, R. F. Hevner, Z. Molnar, J. W. Phillips, C. Dang, A. R. Jones, D. G. Amaral, A. Bernard, E. S. Lein, *Nature* **2016**, *535* (7612), 367, https://doi.org/10.1038/nature18637.

[50] E. S. Lein, T. G. Belgard, M. Hawrylycz, Z. Molnar, *Annu Rev Neurosci* **2017**, *40*, 629, https://doi.org/10.1146/annurev-neuro-070815-013858.

[51] R. Satija, J. A. Farrell, D. Gennert, A. F. Schier, A. Regev, *Nat Biotechnol* **2015**, *33* (5), 495, https://doi.org/10.1038/nbt.3192.

[52] K. Street, D. Risso, R. B. Fletcher, D. Das, J. Ngai, N. Yosef, E. Purdom, S. Dudoit, *BMC Genomics* **2018**, *19* (1), 477, https://doi.org/10.1186/s12864-018-4772-0.

[53] W. Saelens, R. Cannoodt, H. Todorov, Y. Saeys, *Nat Biotechnol* **2019**, *37* (5), 547, https://doi.org/10.1038/s41587-019-0071-9.

[54] A. Verfaillie, H. Imrichova, R. Janky, S. Aerts, *Curr Protoc Bioinformatics* **2015**, *52*, 2 16 1, https://doi.org/10.1002/0471250953.bi0216s52.

[55] D. Szklarczyk, A. L. Gable, D. Lyon, A. Junge, S. Wyder, J. Huerta-Cepas, M. Simonovic, N. T. Doncheva, J. H. Morris, P. Bork, L. J. Jensen, C. v. Mering, *Nucleic Acids Research* **2019**, *47* (D1), D607, https://doi.org/10.1093/nar/gky1131.

[56] P. Shannon, A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski, T. Ideker, *Genome Res* **2003**, *13* (11), 2498, https://doi.org/10.1101/gr.1239303.

[57] L. Kumar, M. E Futschik, *Bioinformation* **2007**, *2* (1), 5.

Supporting Information

Temporal transcriptomic dynamics in developing macaque neocortex

Longjiang Xu , Zan Yuan, Yuan Zhao, Jiafeng Zhou , Wei Liu， Shuaiyao Lu, Zhanlong He, Boqin Qiang, Pengcheng Shu, Yang Chen*, Xiaozhong Peng**

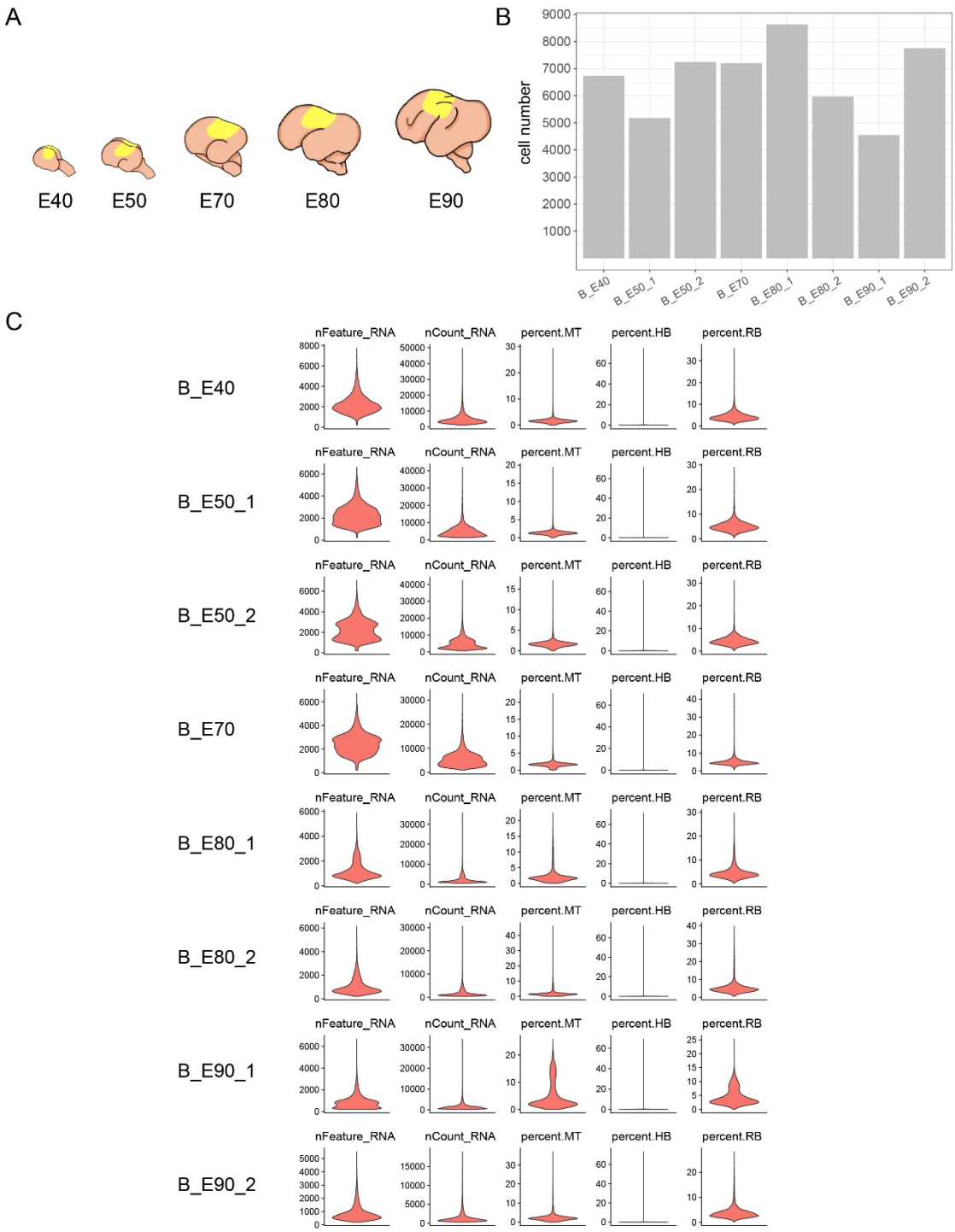

Figure S1. Sample collection and Quality Control.

(A) Schematic diagram of sample collecting anatomical area.

(B) The cell number per sample after quality control.

(C) Single-cell transcriptome library information for each sample.

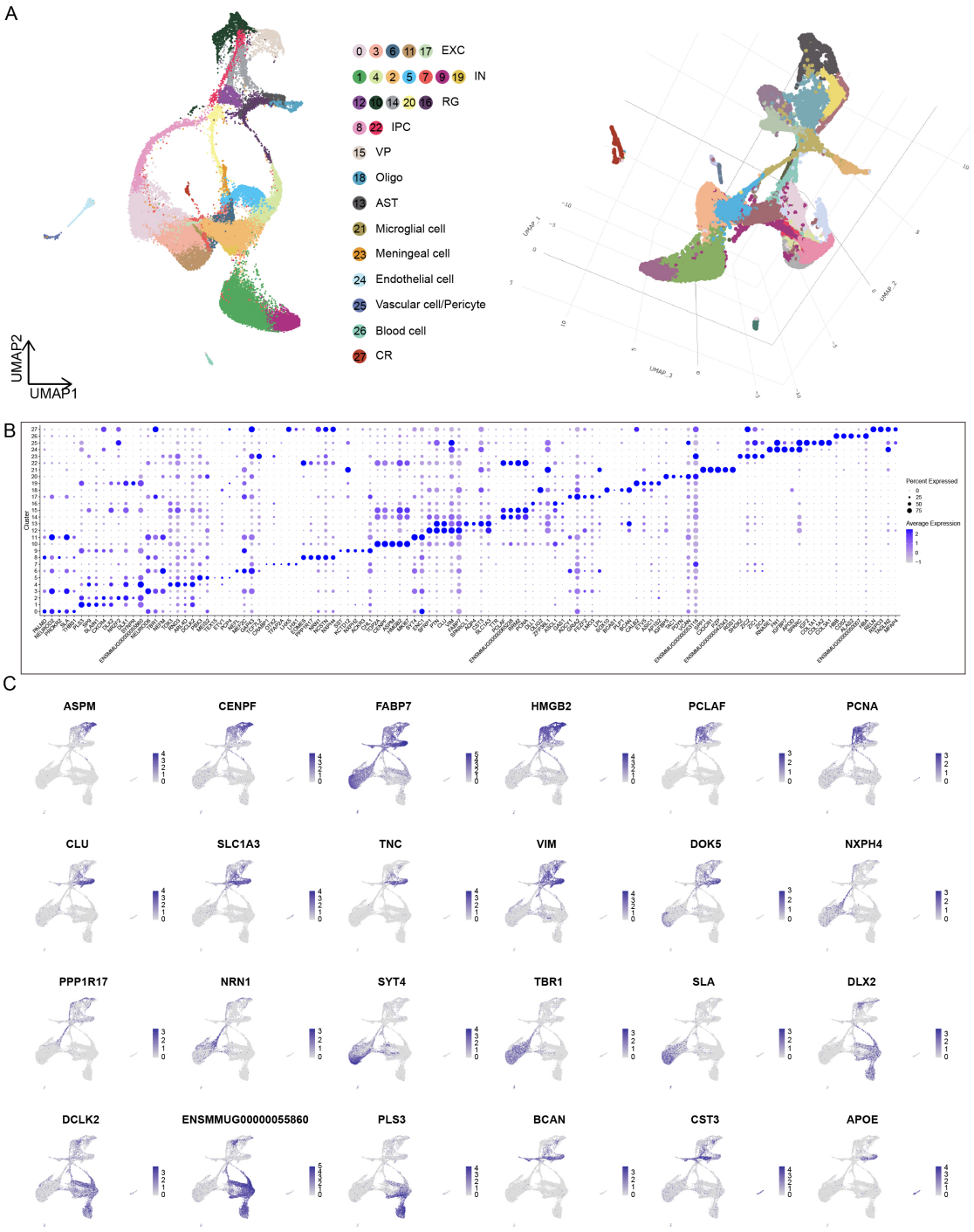

Figure S2. scRNA-Seq uncovers cell type in the developing macaque neocortex.

(A) Visualization of different dimensionality reduction of all cells. (Left, UMAP visualization with UMAP1 and UMAP2; right, 3D model of UMAP visualization with UMAP1, UMAP2, and UMAP3).

(B) Top marker genes for each of the 28 cell clusters shown in Figure 1B.

(C) Feature plots of marker gene expression. Colors represent scaled gene expression.

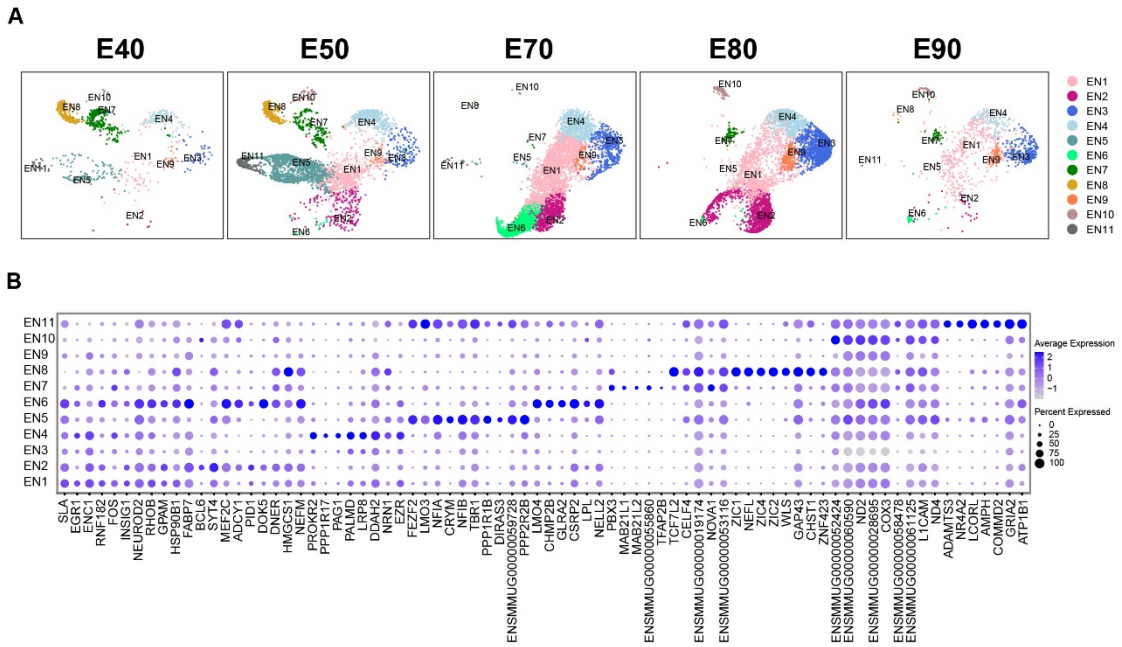

Figure S3. Additional information for excitatory neuron subcluster.

(A) UMAP visualization of excitatory neuron subcluster cell scRNA-seq data from individual time points. Cells are colored by excitatory neuron subcluster assignment.

(B) Top marker genes for each of the excitatory neuron subclusters.

Figure S4. Stem and Excitatory neuron subclusters mapping genes.

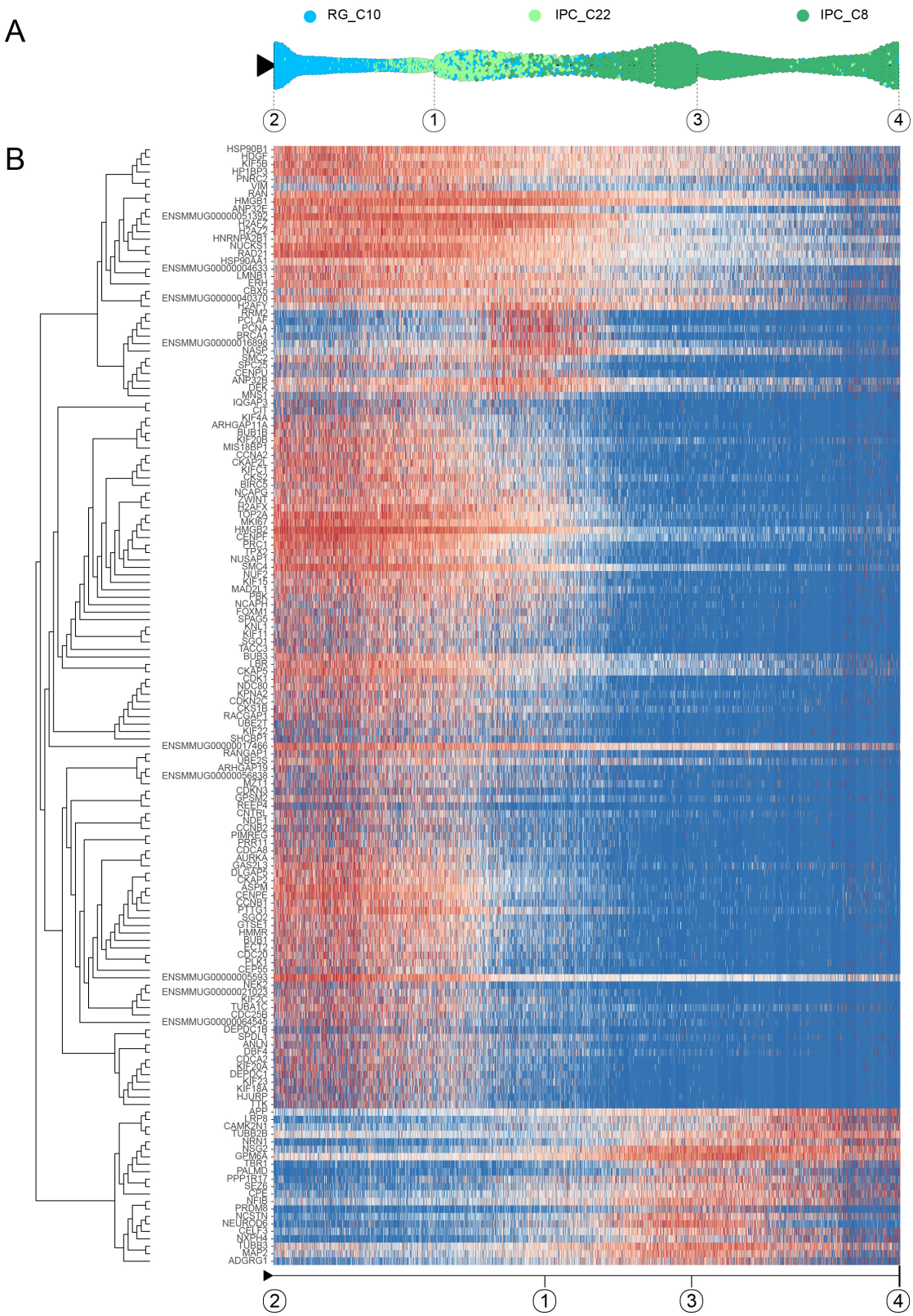

Figure S5. Developmental regulation of gene expression from RG to IPC.

(A) Pseudotime analysis by Slingshot of EOMES-positive cells (C10-C22-C8). Dots: single cells; Cells are colored by their identity.

(B) Heatmap shows the relative expression of top150 genes displaying significant changes along the pseudotime axis of RG to IPC (C10-C22-C8). The columns represent the cells being

ordered along the pseudotime axis. Start point, endpoint, and important nodes of Slingshot inference trajectory were marked by framed numbers. Framed numbers”2” were radial glia cells at the beginning of pseudo time (C10). Framed numbers”1” and “3” marked intermediate state nodes. Framed number “4” marked intermediate progenitor cell (C8).

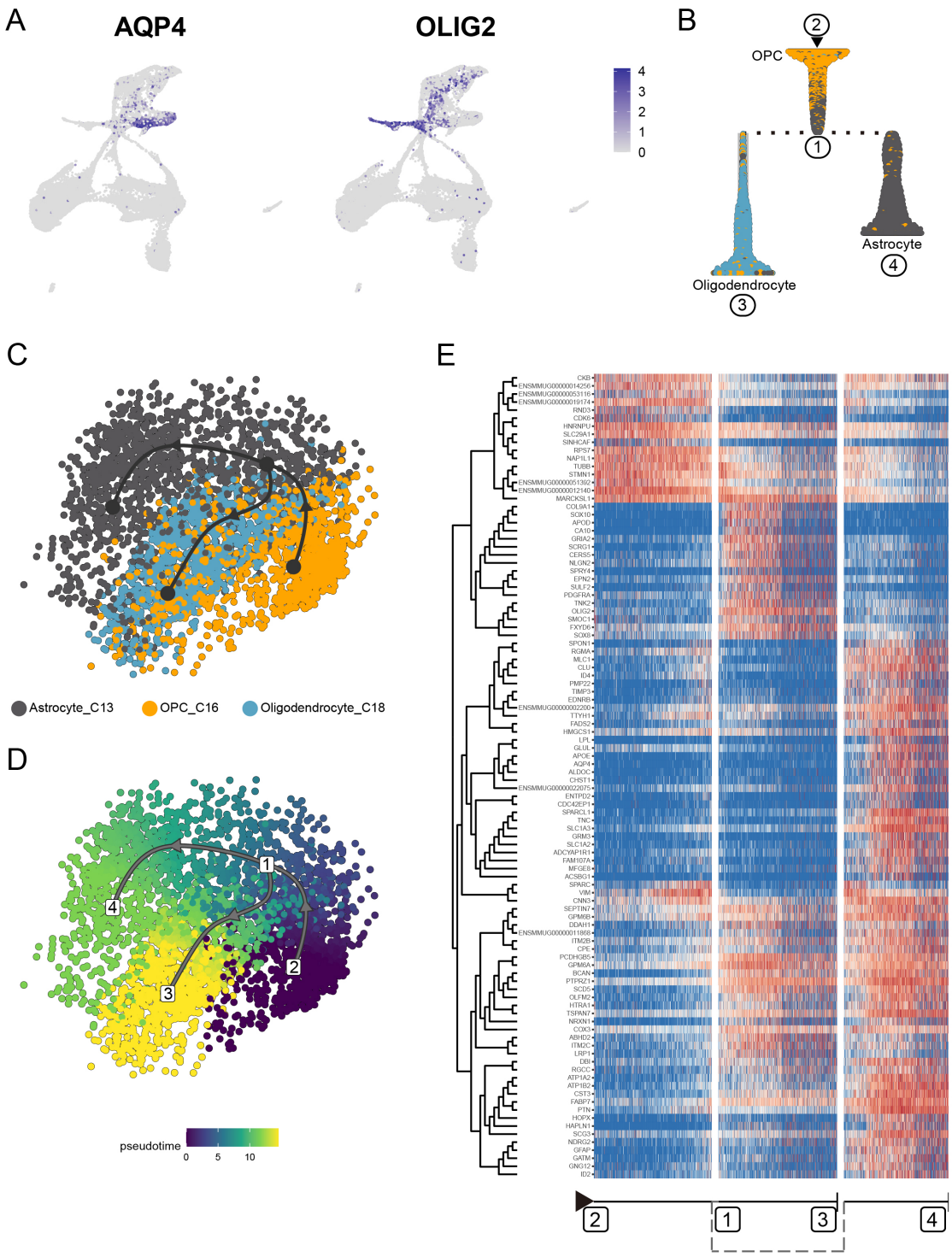

Figure S6. Transcriptional regulation of OPC differentiation into astrocytes and oligodendrocytes.

(A) Feature plot of classic marker for astrocytes (AQP4) and oligodendrocytes (OLIG2). **(B)** Slingshot branching tree related to Slingshot pseudotime analysis in **(C)** and **(D)**. Root is OPC (C16), tips astrocytes (C13), and oligodendrocytes (C18).

(C and **D)** Pseudotime analysis by Slingshot of glial cell lineage (C16→C13, C16→C12). The Slingshot result with the lines indicating the trajectories of lineages and the arrows indicating directions of the pseudotime. Cells are colored according to their identity **(C)** and pesudotime (**D**). Dots: single cells; colors: cluster and subcluster identity.

(E) Heatmap shows the relative expression of top150 genes displaying significant changes along the pseudotime axis of glial cell lineage (C16→C13, C16→C12). The columns represent the cells being ordered along the pseudotime axis.

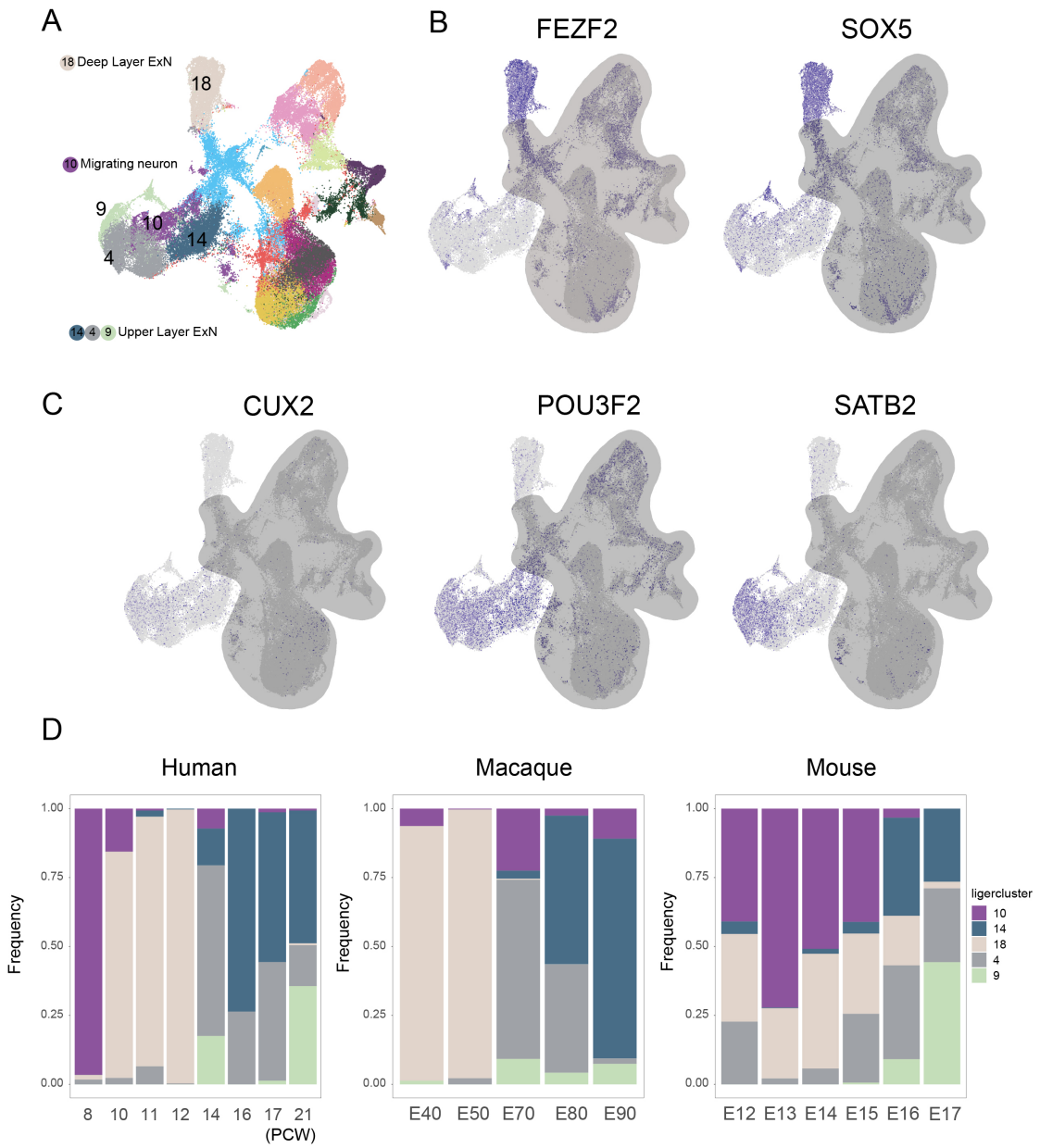

Figure S7. Upper layer and deep layer excitatory neuron proportion analysis among species

(A) Cell type annotation of integrated dataset.

(B and **C)** Feartureplot shows the express of upper layer marker genes CUX2, POU3F2, SATB2 and deep layer marker genes FEZF2, SOX5.

(A) Proportion analysis of excitatory neuron subclusters in human, macaque, and mouse datasets at different developmental timepoint.

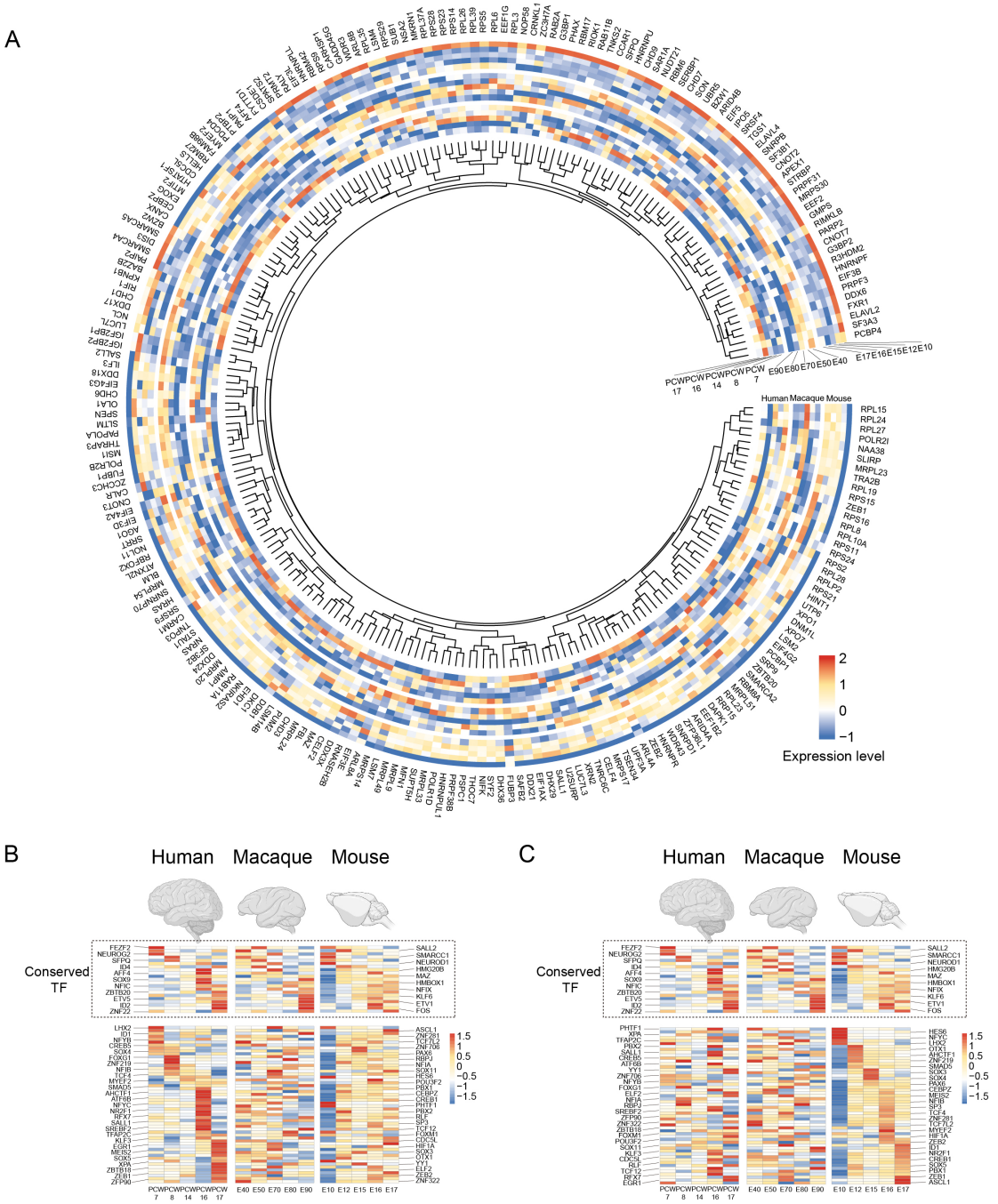

Figure S8. Temporal expression pattern of RNA binding protein and transcription factor genes in human, macaque, and mouse vRG.

(A) Temporal expression heatmap of RNA binding protein genes in human, macaque, and mouse ventricular radial glia.

(B and C) Temporal expression heatmap of Transcription factors genes in human, macaque and mouse ventricular radial glia sorted by temporal pattern of human **(B)** and mouse **(C)**. (Note: each line is the same homologous gene).

Dataset S1. (separate file)

Marker list of 28 cell clusters.

Dataset S2. (separate file)

Gene expression importance across the vRG to oRG (C10-C14-C12) lineage trajectory related to Fig. 3E.(vRG: milestone2; oRG: milestone4)

Dataset S3. (separate file)

Gene expression importance across the RG to deep-layer neurons and upper-layer.

Dataset S4. (separate file)

Gene expression importance across the vRG to IPC(C10-C22-C8) trajectory related to fig. S5B.

Dataset S5. (separate file)

Gene expression importance across the OPC differentiation into astrocytes and oligodendrocytes lineage trajectory related to fig. S6E.

Dataset S6. (separate file)

The original data of normalized gene expression matrix in human, macaque, and mouse vRG related to Fig.6A.