Research article

A multi-omic single-cell landscape of cellular diversification in the developing human cerebral cortex

Yuhan Tian [a],[1], Xia Wu [a],[1], Songhao Luo [b],[1], Dan Xiong [a], Rong Liu [a], Lanqi Hu [a], Yuchen Yuan [a], Guowei Shi [a], Junjie Yao [a], Zhiwei Huang [b], Fang Fu [c], Xin Yang [c], Zhonghui Tang [a], Jiajun Zhang [b],[*], Kunhua Hu [d],[e],[**]

[a] Zhongshan School of Medicine, Sun Yat-sen University, Guangzhou 510275, China
[b] School of Mathematics, Sun Yat-sen University, Guangzhou 510275, China
[c] Guangzhou Women and Children's Medical Center, Guangzhou Medical University, Guangzhou 511436, China
[d] Guangdong Provincial Key Laboratory of Brain Function and Disease, Zhongshan School of Medicine, Sun Yat-sen University, Guangzhou 510275, China
[e] Public Platform Laboratory, The Third Affiliated Hospital of Sun Yat-Sen University, Guangzhou 510630, China

ARTICLE INFO

ABSTRACT

The vast neuronal diversity in the human neocortex is vital for high-order brain functions, necessitating elucidation of the regulatory mechanisms underlying such unparalleled diversity. However, recent studies have yet to comprehensively reveal the diversity of neurons and the molecular logic of neocortical origin in humans at single-cell resolution through profiling transcriptomic or epigenomic landscapes, owing to the application of unimodal data alone to depict exceedingly heterogeneous populations of neurons. In this study, we generated a comprehensive compendium of the developing human neocortex by simultaneously profiling gene expression and open chromatin from the same cell. We computationally reconstructed the differentiation trajectories of excitatory projection neurons of cortical origin and inferred the regulatory logic governing lineage bifurcation decisions for neuronal diversification. We demonstrated that neuronal diversity arises from progenitor cell lineage specificity and postmitotic differentiation at distinct stages. Our data paves the way for understanding the primarily coordinated regulatory logic for neuronal diversification in the neocortex.

1. Introduction

The highly derived mammalian brain region is the cerebral cortex, the neocortex, which exhibits an unparalleled diversity of neurons distributed in specific layers and is the basis of complex behavior in all mammals, especially humans [1–4]. Excitatory projection neurons, one of the two major classes of neocortical neurons, include callosal projection neurons (CPNs) and corticofugal projection neurons (CFuPNs). CFuPNs are classified into corticothalamic projection neurons (CThPNs) and subcerebral projection neurons (SCPNs). SCPNs include corticospinal motor neurons (CSMNs) and corticotectal projection neurons (CTPNs) [2,5–8]. Young neurons of these types are derived from radial glial cells (RGs) and neural intermediate progenitor cells (nIPCs) in the ventricular and subventricular zones. Subsequently, young neurons

migrate radially along the basal process of RGs into the cortical plate and detach from the radial fiber, forming corresponding inside-out layers [9–11]. After neurogenesis, CPNs, which enable information integration, are predominantly represented in layer 2; CSMNs and CTPNs, which exercise somatic and visual movements, respectively, are located in layer 5; and CThPNs are located in layer 6 [12]. However, the mechanisms that regulate excitatory projection neuronal diversity during human neurogenesis remain unclear. Specifically, there is controversy regarding whether neuronal diversity arises from progenitor cell lineage specificity or postmitotic differentiation [13–16].

Recent studies using single-cell gene expression or assays for transposase-accessible chromatin atlas of the human early developing cerebral cortex identified cell clusters and analyzed the molecular differentiation trajectories of these cells [17–19]. However, single-cell RNA

---

sequencing (scRNA-seq) data with only the RNA modality limits the analysis of the regulatory logic of neuronal diversity [17]. Similarly, individual single-cell assays for transposase-accessible chromatin sequencing (scATAC-seq) data are insufficient for identifying cell types and states [19]. To overcome these limitations, single-cell multi-omics techniques have recently emerged, which capture the combined epigenomic and gene expression profiles of the same individual cells. This technique has brought transformative advantages in identifying novel cell subsets and defining important epigenetic regulators (*cis*- or *trans*-regulatory elements) that regulate cellular states and cell differentiation [20]. Therefore, leveraging single-cell multi-omics techniques will deepen our understanding of the regulatory mechanisms underlying the unparalleled cellular diversity and dynamic cell-state transitions in the developing human neocortex.

In this study, we conducted scMultiome-seq, which combines single-cell Multiome ATAC + Gene Expression sequencing, to simultaneously profile chromatin accessibility and gene expression in the same individual cell derived from three early human fetal cortex tissues. We provided a comprehensive and precise annotation of cell subpopulations of the developing cerebral cortex in a combined chromatin accessibility and gene expression atlas at the single-cell level, especially for excitatory projection neurons. In addition, we reconstructed the differentiation trajectories of excitatory projection neurons, thereby revealing that neuronal diversity results from both progenitor cell lineage specificity and postmitotic differentiation, and identified lineage-specific cascades and lineage-bifurcation genes along these trajectories. Moreover, we trained a convolutional neural network (CNN) model using chromatin accessibility and differentiation trajectory to decode the function of essential transcription factors and the chromatin-accessible mode of their DNA binding site. Furthermore, we revealed the developmental differentiation regulatory logic of SCPNs and CPNs using 15,269 *cis*-regulatory elements (CREs) for gene pairs and prioritized functionally important transcription factors (TFs), such as TP73, NR2E1, BHLHE22, STAT5A, EOMES, CUX2, NFATC4, and MEF2C, for future functional studies. This study provides novel insights into the regulatory mechanisms of neurons in developing human cerebral cortex.

2. Materials and methods

2.1. Collection of human fetal brain tissues

Fetal brains at gestational weeks (GW) 11, 15, and 20 were collected from the Guangzhou Women and Children's Medical Center with informed written consent from pregnant women and with approval from the Research Ethics Board of the center (approval number: 392B00). Fresh fetal brains were dissected from the embryos and transferred to the laboratory in ice-cold Hanks' balanced salt solution (Ca^{2+}- and Mg^{2+}-free; Thermo Fisher Scientific, cat. no. 14170112) supplemented with 1% penicillin-streptomycin (10,000 U/mL; Thermo Fisher Scientific, cat. no. 15140163), and 1% 2-[4-(2-hydroxyethyl)piperazin-1-yl] ethanesulfonic acid (pH 7.4, 1 M; Gibco, cat. no. 15630106). Within one hour of sample collection, cortical plates were resected from fetal brains, snap-frozen in liquid nitrogen, and stored in liquid nitrogen for subsequent sample processing.

2.2. Isolation of single nuclei from cortical plates

Individual nuclei were isolated from each frozen cortical plate in liquid nitrogen, as described in [21], with minor modifications. Briefly, each frozen cortical specimen was transferred to a prechilled Dounce tube (Kimble, cat. no. D9063–1SET) on the ice. We added one mL prechilled homogenization buffer comprising 0.25 M sucrose, 25 mM KCl, 5 mM $MgCl_2$, 20 mM tricine-KOH pH 7.8, 1 mM dithiothreitol (DTT), 0.15 mM spermine, 0.5 mM spermidine, 5 µg/mL actinomycin, 0.32% Nonidet P-40 (NP40), 0.04% bovine serum albumin (BSA), 1 U/µL Protector RNase (Sigma Aldrich, cat. no. 3335402001), and 1 × protease

inhibitors in Dounce. After 25 strokes (20 strokes for the GW11 specimen) with a loose pestle, the homogenate was mixed with an equal volume of OptiPrep (Sigma, cat. no. D1556), and centrifuged at 10,000 ×g for 20 min at 4 °C. Subsequently, the pelleted nuclei were rinsed twice in ice-cold phosphate-buffered saline (PBS, pH 7.4) supplemented with 2% BSA and 0.2 U/µL Protector RNase and centrifuged at 800 ×g for 5 min at 4 °C. Afterwards, the nuclei were permeabilized as described in the 10 × Genomics protocol CG000366 (Nuclei Isolation from Embryonic Mouse Brain for Single Cell Multiome ATAC + Gene Expression Sequencing, Rev C) with minor modifications. Briefly, nuclei were resuspended in 100 µL ice-cold 0.1 × lysis buffer (10 mM Tris-HCl pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$, 1% BSA, 1 mM DTT, 1 U/µL Protector RNase, 0.01% Tween-20, 0.01% NP40 substitute, and 0.001% digitonin) for 10 s on ice. Permeabilization was stopped by adding 1 mL ice-cold wash buffer (10 mM Tris-HCl pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$, 0.1% Tween-20, 1% BSA, 1 mM DTT, and 1 U/µL Protector RNase). Subsequently, the nuclei were centrifuged at 800 ×g and 4 °C for 5 min and resuspended in 50 µL chilled 1 × nuclei buffer from the 10 × Single Cell Multiome ATAC + Gene Expression kit (10 × Genomics, cat. no. PN-1000285). Following the above process of nucleus isolation, we obtained individual nucleus suspensions from each cerebral cortical tissue with nuclear densities of 1700–4000/µL.

2.3. Generation of scMultiome-seq libraries

Each single-cell Multiome ATAC + Gene Expression sequencing (scMultiome-seq) library was prepared according to the manufacturer's manual of the Single Cell Multiome ATAC + Gene Expression kit to generate scMultiome-seq data. Briefly, five µL (for nuclei density < 3200/µL) or 16,000 nuclei were subjected to transposase treatment. After transposition, the nuclei were loaded onto Chip J to generate the droplets. After reverse transcription and pre-amplification, the cDNA and transposed DNA were cleaned using 1.6 × SPRIselect beads (Beckman, cat. no. B23318) and eluted with 160 µL elution buffer. Next, 40 µL of DNA was used for chromatin accessibility library construction through seven cycles of PCR amplification. In addition, 35 µL of DNA was used for cDNA amplification and gene expression library generation (10 cycles of PCR amplification). Finally, the size distribution of the cDNA and libraries (chromatin accessibility and gene expression) was assessed using Qsep (Bioptic, Qsep1), and the concentration was measured using a Qubit HS dsDNA kit (Thermo Fisher, cat. no. Q33230).

The single-cell assay for transposase-accessible chromatin sequencing (scATAC-seq) library was sequenced using Illumina MiSeq paired-end sequencing (50 + 8 + 24 + 50 bp). The single-cell RNA sequencing (scRNA-seq) library was sequenced using Illumina NovaSeq 6000 paired-end sequencing 150 + 150 bp. In this manner, we generated a set of scMultiome-seq data for each sample.

2.4. Processing of scMultiome-seq datasets and annotation of cell types

For each sample, scRNA-seq and scATAC-seq reads were first aligned to the GRCh38 (hg38) reference genome and quantified using 'cellranger-arc count' (10x Genomics, v2.0.1), in which genes were referenced using Gencode v.32. Next, count data of both modalities of each scMultiome-seq dataset were processed to generate the RNA and ATAC assays using Seurat (v4.2.0) [22] and Signac (v1.7.0) [23] in R (v4.1.0). Subsequently, cells in each scMultiome-seq dataset were filtered as follows, 1) we removed low-quality cells with < 1000 informative genes expressed, > 7500 informative genes expressed, > 30% counts corresponding to mitochondrial genes, < 1000 peaks, > 30,000 peaks, < 1 transcription start site (TSS) enrichment, and > 2 nucleosome signals. 2) We performed doublet analysis of scRNA-seq data using the 'DoubletFinder' R package (v2.0.3) [24].

After quality control and filtering across both data modalities, we performed unsupervised analyses to separately and jointly assess the global similarities and differences between individual cells in the RNA

and ATAC assays. For the scRNA-seq data, we performed data normalization, highly variable feature identification, data scaling, linear dimensional reduction (principal component analysis), nonlinear dimensional reduction using uniform manifold approximation and projection (UMAP), and clustering. For the scATAC-seq data, we performed data normalization, feature selection, linear dimension reduction (singular value decomposition), nonlinear dimension reduction using UMAP, and clustering. For comprehensive analysis of the RNA and ATAC modalities, we reconstructed the weighted nearest neighbor graph by 'FindMultiModalNeighbors' [25], generated the UMAP embedding, and done clustering. After the above manipulation, three cell atlases were generated for each sample: RNA, ATAC, and combined UMAP.

Finally, we manually inspected the expression of known marker genes and assigned cell types to the initial clusters for the three cell atlases of chromatin accessibility and gene expression independently and jointly (Table S1).

2.5. Data integration and cell type annotation

A reasonably strong batch effect was observed between the three scMultiome-seq datasets. Therefore, we integrated respectively the chromatin accessibility and gene expression information from the three scMultiome-seq datasets according to the following steps. For chromatin accessibility data, we called peaks using model-based analysis of ChIP-seq2 (MACS2) [26], in which the BAM files of the three scATAC-seq data were applied as input files. Peak regions with $-\log_{10}$ (p-value) < 20 were blocked to obtain a unified peak list. We used the MACS2 peak set to re-count the fragments in the filtered cells of each sample and generated a new chromatin accessibility assay. We then merged the scATAC-seq data from all time points and removed the batch effect from the merged ATAC data by finding integration anchors and integrating uncorrected latent semantic indexing embedding [22]. For the gene expression data, we merged the scRNA-seq data from all time points and eliminated the batch effect using Harmony [27]. Next, for the integrated ATAC and RNA data, we reduced the dimensions using UMAP, performed clustering, and independently generated an integrated ATAC and RNA UMAP. Subsequently, we obtained a combined UMAP by combining integrated chromatin accessibility and gene expression information, dimension reduction, and doing clustering.

Finally, we manually inspected the expression of known marker genes and assigned cell types to the initial clusters in the three integrated atlases of gene expression and chromatin accessibility independently and jointly.

2.6. Integration of public scMultiome-seq data

We projected a previously published scMultiome-seq dataset from the human fetal cortical plates at GW 18, 19, 23, and 24 onto our own scMultiome-seq data and predicted cell type annotations using Seurat (v4.2.0) [22,28].

2.7. Inference of RNA velocity

For each gene expression data of the three samples, a loom file, including spliced and unspliced matrices, was calculated using human Ensembl annotation version 98 using Velocyto (v0.17.17) [29]. Furthermore, these loom files were concatenated and processed based on the combined UMAP coordinates using Scanpy (v1.9.1) and scVelo (v0.2.4) [30]. Finally, the velocity vector field was displayed as streamlines using the scv.pl.velocity_embedding_stream function.

2.8. Inference of branching trajectory tree

We converted the Seurat object of gene expression data in the combined UMAP into a URD object and reconstructed the branching

trajectory tree using URD (v1.1.1) [31]. Specifically, we calculated the variable genes for each cell type, combined the genes of each cell type into a single list of variable genes, and loaded them into the URD object. We then calculated a diffusion map using the calcDM function with knn = 200 and sigma.use = 10. Subsequently, we assigned cycling radial glia (RG_Cyc) as the root and calculated the pseudotime using the floodPseudotime and floodPseudotimeProcess functions. Based on the pseudotime distribution of each cell type, the CSMN, CTPN, CThPN, superficial, layer 4, and deep CPNs were defined as the tips of the tree. Next, we determined the logistic parameters used to bias the transition probabilities using the pseudotimeWeightTransitionMatrix and biased the transition matrix according to the pseudotime. We then simulated 100,000 biased random walks from each tip to the root and processed them into visitation frequencies using simulateRandomWalksFromTips and processRandomWalksFromTips. Finally, we built a branching trajectory tree using the buildTree function.

2.9. Lineage-specific gene cascades

We identified lineage-specific genes for each branching trajectory based on our reconstructed tree using the aucprTestAlongTree function in the URD package. Specifically, for each trajectory, we focused on the segment from the tip to the root along each trajectory and the segment's opposition (the segment's sibling and the siblings' children). A pairwise comparison was performed between the cells in the segment and the segment's opposition to identify differentially expressed genes. A differentially expressed gene was considered a cascade on the trajectory if it was differentially highly expressed against > 60% of the cells in the segment's opposition and was not upregulated in the trajectory downstream of the segment's opposition (must.beat.sibs = 0.6).

To better demonstrate the expression cascade of lineage-specific genes for each trajectory, we first determined the "on and off" timing of lineage-specific gene expression and ordered genes by cascading using the geneCascadeProcess function from URD. Specifically, the average expression of each gene was calculated and smoothed using a sliding window along the trajectory. Genes were sorted based on the pseudotime value of "start and leave" gene expression. Next, we plotted the expression cascade heatmap of lineage-specific genes for each trajectory using the geneCascadeHeatmap.

2.10. Identification of the cell-type decision genes

To define genes associated with lineage bifurcations for each branch of all branch points, we first selected cells that appear at a 0.04 pseudotime units before and after the branch point, and identified differentially expressed genes in the branch compared with parent cells using the FindMarkers function in the Seurat package (min.pct = 0.1, logfc.threshold = 0.2). Next, to identify the top genes that distinguish cells on one branch from those on the other branch and the parent, we trained a Gradient Boosting Classifier using scikit-learn (v1.1.1) in each branch, calculated the importance score of differentially expressed genes, and selected the top 10 genes based on importance scores [32].

2.11. Architecture and training of deep-learning model

The AI-TAC model [33] is a convolutional neural network (CNN) that predicts chromatin accessibility across 18 cell types in the human cerebral cortex from DNA sequences using pseudobulk ATAC-seq derived from scATAC-seq data. The input of the AI-TAC was a one-hot-encoded DNA sequence (A = [1,0,0,0], C = [0,1,0,0], G = [0,0,1,0], T = [0,0,0,1]) of 251 bp around the summit of the ATAC-seq peak. The output was the accessibility profile of the corresponding peak across 18 cell types, quantified as its quantile normalized log count value. The model architecture comprised three convolutional layers and two fully connected layers. The first convolutional layer used 300 filters with a width of 19 bp, followed by rectified linear unit (ReLU) activation function, max

pooling with a width of 3 bp, and batch normalization. The remaining two convolutional layers used a similar architecture with different parameters (200 filters of width 11 bp and max pooling with a width of 4 bp for the second layer; 200 filters with a width of 7 bp and max pooling with a width of 4 bp for the third layer). Furthermore, the two fully connected layers comprised 1000 neurons, followed by the ReLU activation function and dropout at a rate of 0.03. Finally, the training used the Pearson correlation as the loss function and the Adam optimizer [34] for 10 epochs with a learning rate of 0.001 and a mini-batch size of 100. All model implementations were based on PyTorch version 1.4.0.

2.12. Learning motifs from the first layer filters of the CNN model

We calculated the position weight matrices (PWMs) of 300 filters with 19 bp in the first layer to explain the essential features that the CNN model learned from DNA sequences. First, we identified 19 bp sequences, which activated the filter by at least half of the maximum activation across the peaks with good predictions (Pearson correlation > 0.75). Next, for the 19 bp sequence, we designed a position frequency matrix based on each nucleotide and computed the PWM using a background uniform nucleotide frequency of 0.25. These operations created the first layer of 300 filters to represent the motifs learned from DNA sequences.

To explore the robustness of the motifs learned by each filter in the model, we calculated the reproducibility of the motifs. We retrained 10 additional CNN models to determine the PWMs, each using random sampling of 90% of the total data. Next, we used the TomTom algorithm [35] to compare whether the motifs in the original model were similar to at least one of the motifs in each additional model (q-value < 0.05).

Three metrics are calculated to measure the importance of each motif. The first metric is the overall influence, which quantifies motif effects on model predictions. This was performed by replacing the activation value of this motif with the average activation value across all peaks to obtain the predicted value of the model with this motif missing. Thus, the overall influence was computed as the average (across all samples) of the squared difference between the correlation loss of the original and missing motif models. In addition, the cell-type-specific influence of the motif was defined as the mean of the prediction difference between the original and missing motif models across all peaks. The second metric, information content, was computed as the sum of (across all 19 bp positions) the entropy differences between the PWM and background uniform nucleotide frequency of 0.25. The third metric, the number of peaks influenced by the motif, was defined as the number of samples for which the activation value of the motif was greater than half of the maximum activation value.

Last, we assigned known transcription factor (TF) to the motifs using TomTom to compare the PWM of each filter with the Catalog of Inferred Sequence Binding Preferences (CIS-BP) database (from chromVAR motifs 'human_pwms_v2') [36]. We further calculated the correlation between the expression level and accessibility of these alternative TFs with q-values < 0.05 and p-values within two orders of magnitude of the smallest p-value generated by TomTom. The results are shown in Fig. S7D.

2.13. Identification and validation of peak-gene links

We linked genes with their nearby peaks among all cell types using the LinkPeaks function from Signac with default parameters, and identified 15,269 significant peak-gene links.

To display the dynamics of gene-regulatory interactions across all cell types, we drew a cluster heat map of the peak-gene links as follows: 1) We defined 1000 seed cells by random sampling from the filtered 17,436 cells. 2) Each of the 1000 seed cells was combined with its 49 nearest neighbor cells in the combined UMAP space to generate a pseudobulk sample. 3) The chromatin accessibility and gene expression of each pseudobulk sample were obtained by summing the peak and

RNA signals of the 50 cells. 4) Each pseudobulk sample was annotated with the cell type and age in the combined UMAP. 5) These linkages were clustered using k-means (k = 20) clustering based on the z-score-scaled accessibility levels of associated peaks. 6) Twenty interaction clusters were manually ordered according to trends in chromatin accessibility and gene expression.

2.14. Enrichment of linked peaks in genomic elements

Enrichment analysis of the linked peaks in genomic elements was performed in defined regions and epigenetically annotated regions of the genome. Fold enrichment was calculated using the ratio between the (#bases in state AND overlap feature)/(#bases in genome) and [(#bases overlap feature)/(#bases in genome) × (#bases in state)/(#bases in genome)] [37,38].

For fold enrichment of the linked peaks in defined genome regions, TSS was defined as 1 kb upstream and 1 kb downstream of the true TSS, the promoter was defined as 2 kb upstream and 1 kb downstream of the TSS, and the transcription end site (TES) was defined as 1 kb upstream and 1 kb downstream of the true TES.

For fold enrichment of the linked peaks in epigenetically annotated genome regions, the chromatin state model of fetal brain tissue (E081) with 25-states was obtained from the Roadmap Epigenomics Project [37,39]. Validated human forebrain and non-forebrain enhancers were downloaded from the VISTA Enhancer browser [40].

2.15. Chromatin and gene-regulatory dynamics in the SCPN or superficial CPN trajectory

To study the dynamics of gene-regulatory interactions in the SCPN or superficial CPN differentiation trajectory, we first identified cluster-specific genes in the corresponding trajectory using the FindAllMarkers function from Seurat with an adjusted p-value < 0.01 and a log2-fold change > 0.25. Next, we ordered the cells based on their pseudotime scores in the differentiation trajectory and merged bins of 50 cells to obtain the pseudobulk samples. Last, we clustered the *cis*-regulatory elements (CREs) for gene pairs of cluster-specific genes across pseudobulk samples using k-means clustering based on the z-score-scaled accessibility levels of the linked CREs. Each pseudobulk sample was annotated using the cell type and the age of the combined UMAP.

2.16. Differential motif enrichment analysis

Underlying TF binding sites (TFBSs) in the human genome were acquired from JASPAR2022 (http://expdata.cmmt.ubc.ca/JASPAR/downloads/UCSC_tracks/2022/hg38/) [41]. TFBSs were then filtered using conserved regions (scores > 0.4 and width ≥ 20 bp) from phastCons (https://hgdownload.soe.ucsc.edu/goldenPath/hg38/phastCons100way/hg38.100way.phastCons/) [42]. To identify TF binding motifs that present more often in one of the CSMN and CTPN CREs, differential motif enrichment analysis was performed using a logistic regression model: glm (TFBS ~CSMNCTPN + CRE width + conservedbppercent, family = "binomial") in R, which explicitly controlled for the width and conservation differences between CSMN and CTPN CREs. TFBS was the dependent variable, indicating the presence or absence of a TF motif in the CRE. Moreover, CSMNCTPN was the independent variable of interest, in which a CRE belonged to CSMN (CSMNCTPN = 1) or CTPN (CSMNCTPN = 0). CRE width and conservedbppercent represent the width and percentage of CRE, respectively. csmnTFs were defined as significantly differentially enriched motifs present more often in CSMN than in CTPN CREs, and ctpnTFs were defined as significantly differentially enriched motifs present more often in CTPN than in CSMN CREs. The class of each TF was acquired using JASPAR2022, and Fisher's exact test was used to determine whether a significant enrichment or depletion of that class was present in significant csmnTFs or ctpnTFs compared to that in all TFs found in
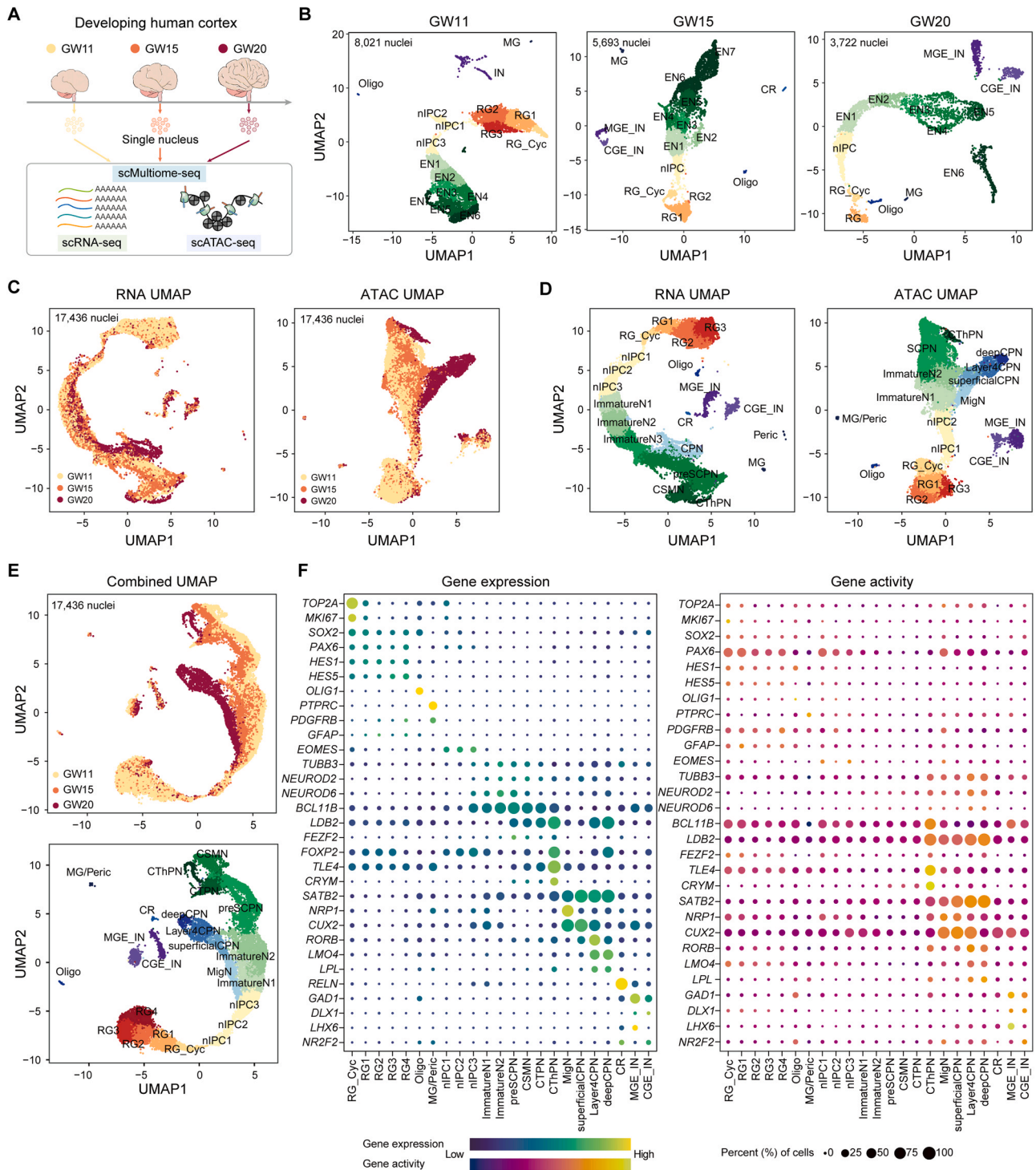
Fig. 1. A single-cell Multiome ATAC + Gene Expression atlas of the human developing cerebral cortex. A Generation of three scMultiome-seq datasets simultaneously profiling open chromatin and gene expression from the same cell. **B** Combined UMAP for each scMultiome-seq dataset derived from fetal cortical plates at gestational weeks (GW) 11, 15, and 20. Each dot represents a cell colored according to the cell type. RG, radial glia; RG_Cyc, radial glia with cycles; Oligo, oligodendrocyte; MG, microglia; Peric, Pericytes; nIPC, intermediate progenitor cell; EN, excitatory projection neuron; IN, interneuron; CGE_IN, caudal ganglionic eminence interneuron; MGE_IN, medial ganglionic eminence interneuron; CR, Cajal-Retzius. **C** Integrated UMAPs for gene expression (left) and chromatin accessibility (right). Each dot represents a cell colored according to fetal age. **D** ATAC and RNA UMAPs of cells are colored by cell type. Each dot represents a single cell. RG, radial glia; RG_Cyc, radial glia with cycles; Oligo, oligodendrocyte; MG, microglia; Peric, Pericytes; nIPC, neuronal intermediate progenitor cell; EN, excitatory projection neuron; IN, interneuron; CGE_IN, caudal ganglionic eminence interneuron; MGE_IN, medial ganglionic eminence interneuron; CR, Cajal-Retzius; MigN, migrating neuron; ImmatureN, immature neuron; CPN, callosal projection neurons; preSCPNs, pre-subcerebral projection neurons; SCPN, subcerebral neurons; CSMN, corticospinal motor neurons; CTPN, corticotectal projection neurons; CThPN, corticothalamic projection neurons. **E** Combined UMAP based on gene expression and chromatin accessibility information. Each dot represents a single cell. Cells are colored according to fetal age (top) and cell type (bottom), the full names of which are listed in Fig. 1D. **F** Dot plot showing gene expression (left) and gene activity (right) of selected markers across cell types, the full names of which are listed in Fig. 1D, in the combined UMAP. The color bar represents the average expression or activity level of a gene in a cell cluster. The size of the dot indicates the percentage of cells in cell clusters that express the gene.
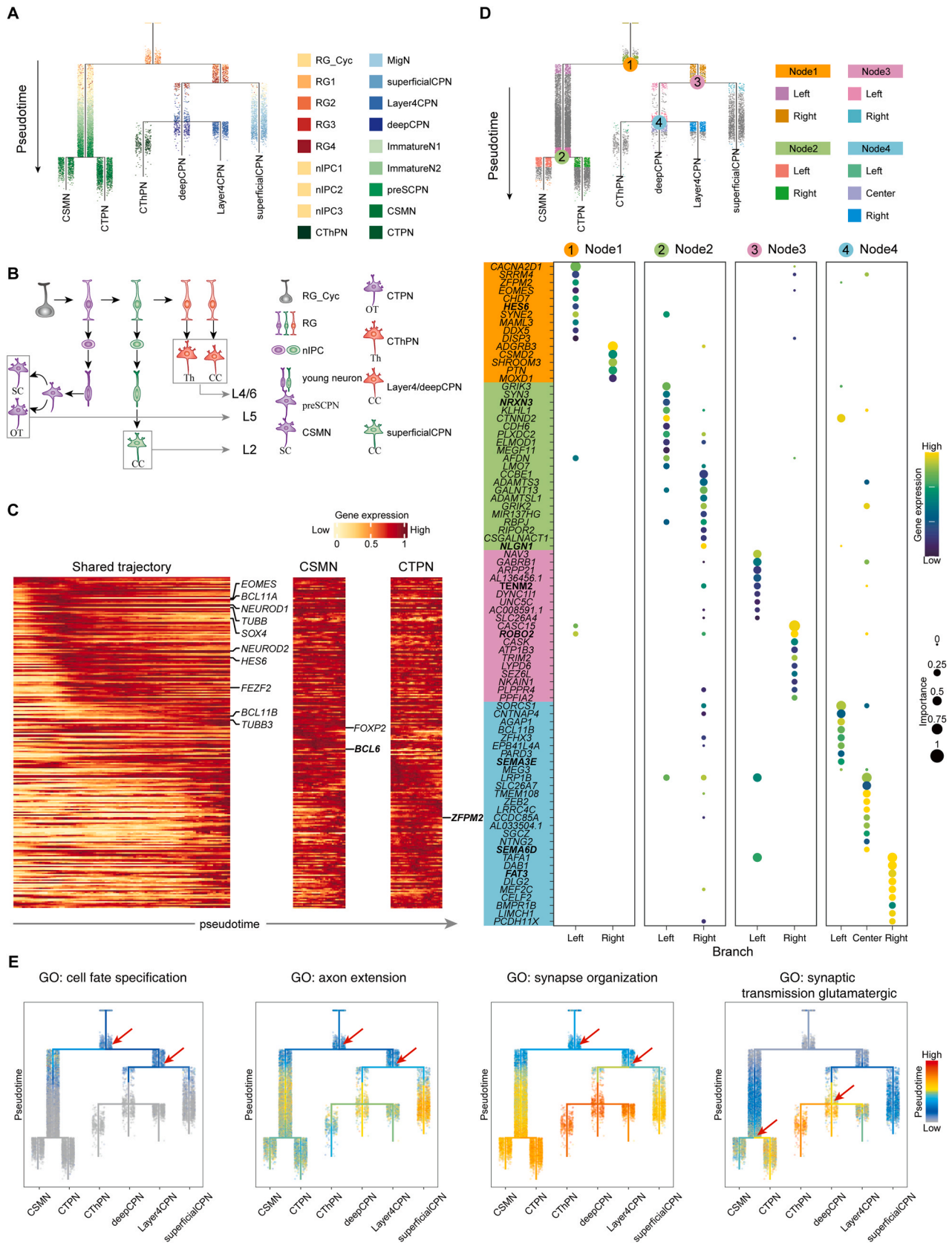
(caption on next page)

Fig. 2. Molecular differentiation trajectories of neocortical excitatory projection neurons. A Branched trajectory tree of excitatory projection neurons. The root of the tree is RG_Cyc; the tips are various excitatory projection neuron types. Each dot represents a single cell. Cells are colored according to the cell type, the full names of which are listed in Fig. 1D. **B** Schematic illustration of neurogenesis in excitatory projection neurons in the human neocortex. See Fig. 1D for abbreviations of the cell types. The early, intermediate, and late RG are labeled in purple, green, and orange, respectively. SCPN-related nIPC and young neurons are labeled in purple, and superficial CPN-related nIPC and young neurons are labeled in green. **C** Smoothened heatmap showing the cascade expression of lineage-specific genes over the shared trajectory and CSMN- and CTPN-specific trajectories. Genes in black and bold black represent cell-type specific markers and new genes, respectively. **D** Lineage-bifurcation genes that determine the direction of cell differentiation. Branched trajectory tree showing the nodes of cell differentiation direction determination and their corresponding cell differentiation directions. Bubble diagram showing the average expression and importance score of the top 10 lineage-bifurcation genes predicted to be involved in cell identity divergence. Each dot represents a gene, with its size indicating the importance score of the gene, and its color corresponding to the gene expression level. Bolded genes are shown in Fig. S6C. **E** Enrichment of gene sets associated with gene ontology (GO) items shown on branched trajectory trees. The gene features of the GO items are shown at the top of each graph. Red arrows represent branch node locations enriched for GO items.

JASPAR2022. A similar analysis was performed for other results.

2.17. TF binding motif enrichment analysis

To identify the TFs responsible for establishing and maintaining a class of cells, we first obtained TF-binding motif information using the getMatrixSet function in JASPAR2020 [43]. Next, we performed motif scanning for the sequences of all peaks using the AddMotifs function in Signac, using the genome information in the BSgenome.Hsapiens.UCSC. hg38 package. Finally, we identified enriched TF binding motifs in the provided CRE list using the FindMotifs function in Signac.

2.18. Gene ontology annotation

The Cluster Profiler (v4.2.2) package was used for gene ontology term enrichment [44]. Terms with a p-value < 0.05 were defined as significantly enriched.

3. Results

3.1. A single-cell comprehensive atlas of chromatin accessibility and gene expression in the developing human cerebral cortex

The emerging single-cell multi-omics technique that simultaneously profiles open chromatin and gene expression from the same individual cells offers the possibility of a deeper characterization of cell types and states in the early developing human cerebral cortex. Therefore, we collected fetal cortical plates at GW 11, 15, and 20 and conducted three scMultiome-seq assays to capture cellular heterogeneity (Fig. 1A; see Materials and methods). After quality control and filtering of both data modalities, we obtained 8021 cells with high-quality epigenome and transcriptome profiles from GW11, 5693 cells from GW15, and 3722 cells from GW20 (Fig. S1 and Table S2). For each scMultiome-seq dataset, we performed unsupervised clustering analysis to assess global similarities and differences between individual cells and generated single-cell atlases of chromatin accessibility and gene expression both independently and jointly. For the three UMAPs of each cerebral cortical tissue, such as ATAC, RNA, and combined UMAPs, each contained major cell types annotated with the expression of known marker genes, such as RG_Cyc, RG, oligodendrocyte, microglia, pericyte, nIPC, excitatory projection neuron, interneuron, caudal ganglionic eminence interneuron, medial ganglionic eminence interneuron, and Cajal-Retzius [2,3,18,45–48] (Fig. 1B and S2). In addition, in these UMAPs, the nIPC cluster was in proximity to RG cell populations, and various clusters of excitatory projection neurons appeared successively after the nIPC cluster (Fig. 1B and S2). These results indicate that the three scMultiome-seq datasets were of high quality.

The UMAPs of each cortical tissue showed that data of the epigenome and transcriptome may enable the recognition of different cell types (Fig. S2). Thus, we eliminated the batch effect of ATAC and RNA data from the three scMultiome-seq datasets and generated ATAC and RNA UMAPs to investigate whether both modalities have different recognition abilities for different cells in the developing human cerebral cortex. Specifically, for chromatin accessibility data, we created a new

chromatin accessibility assay for each sample using a consensus set of 103,566 accessible peaks, removed the technology-specific variation of the merged ATAC data, and produced an integrated ATAC UMAP [22]. Regarding the gene expression data, we merged the scRNA-seq data from all samples and generated an integrated RNA UMAP using Harmony [27]. UMAP indicates the differentiation trajectory of cells to some extent. We observed vertical parallelism in both ATAC and RNA UMAPs among cells at different gestational timepoints (Fig. 1C), suggesting that during gestation, cells undergo differentiation without affecting identity determination. The horizontal parallelism exhibited by cells of different clusters indicates cell differentiation pertaining to chromatin accessibility and gene expression may be associated with pseudotime rather than gestation time. Furthermore, we separately annotated clusters in ATAC and RNA UMAPs using the expression of known marker genes [5,18,47–49] (Fig. S3). Different cell types were observed for the two atlases. For example, CSMNs and Cajal-Retzius cells appeared only in the RNA UMAP, whereas subpopulations of CPNs, including superficial, layer 4, and deep CPNs, were annotated in ATAC UMAP (Fig. 1D). These results suggest that the two modalities have different discriminatory abilities in different cell types.

scMultiome-seq data, including the epigenome and transcriptome, makes it possible to resolve cellular heterogeneity more accurately and reliably. Thus, we combined chromatin accessibility and gene expression information, generated a combined UMAP, performed clustering, and annotated cell types using known marker gene expression or activity to better perform cell heterogeneity analysis. Structural variation in the combined UMAP was associated with sample time points and cell types in various directions, supporting our hypothesis that cell differentiation, involving chromatin accessibility and gene expression information, may be associated with pseudotime rather than sample gestation time (Fig. 1E). Moreover, the combined UMAP annotated more comprehensive cellular subtypes than ATAC or RNA UMAPs alone, particularly for excitatory projection neurons. For instance, among CPN clusters expressing *SATB2*, cells in superficial, layer 4, and deep CPNs expressed *CUX2*, *RORB*, and *LMO4* and *LPL*, respectively. Two subgroups of CFuPNs expressing *BCL11B*, $TUBB3^+NEUROD2^+NEUROD6^+BCL11$ $B^+FOXP2^+TLE4^+CRYM^+$ CThPNs and $TUBB3^+NEUROD2^+NEURO$ $D6^+BCL11B^+LDB2^+FEZF2^+$ SCPNs, appeared in the combined UMAP. *FOXP2* labels CSMNs but not CTPNs [49]. We also identified a migrating CPN cluster expressing *NRP1* and an immature $TUBB3^+NEURO$ $D2^+NEUROD6^+BCL11B^+LDB2^-FEZF2^-FOXP2^-TLE4^-CRYM^-$ SCPN cluster (Fig. 1F, Table S1). These results indicate that scMultiome-seq data are more advantageous for revealing cellular heterogeneity. To further corroborate cell-type identities, we integrated a previously published scMultiome-seq dataset from the human fetal cortical plates at GW 18, 19, 23, and 24 with our own scMultiome-seq data and observed a high degree of consistency between the predicted cell types and the original cell types, thereby validating the accuracy of our cell type annotation (Fig. S4; see Materials and methods) [28]. Altogether, we generated a comprehensive single-cell atlas of chromatin accessibility and gene expression in the developing human cerebral cortex and marked almost all known subpopulations of excitatory projection neurons, providing the possibility of resolving the regulatory logic of excitatory projection neuronal diversity in downstream analyses.
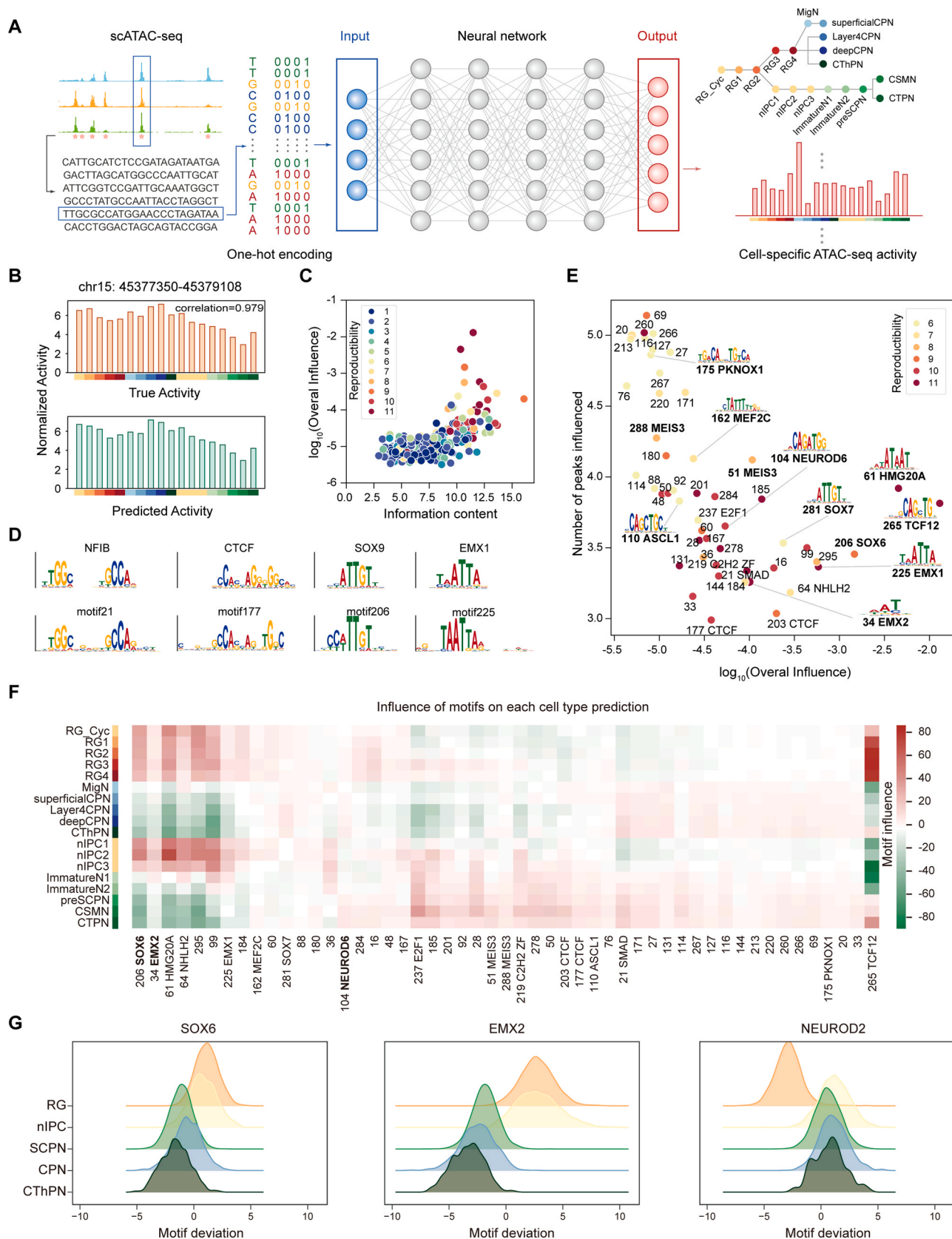
(caption on next page)

Fig. 3. Deep-learning model learns motifs associated with the human cerebral cortex. A Schematic of the deep-learning model for predicting cell-specific chromatin accessibility profiles from DNA sequences across neuron types, the full names of which are listed in Fig. 1D, during the development of neocortical excitatory projection neurons. The input was a 19-bp DNA sequence near the peaks of the pseudobulk ATAC-seq generated by scATAC-seq and converted to one-hot encoding (left panel). The output is the chromatin accessibility profile of 18 human cortical cells (right panel). Specifically, this deep-learning model uses a convolutional neural network. **B** An example comparing chromatin accessibility profiles of 18 human cortical cell types between true observations and deep-learning model predictions. **C** Scatter plot of the characteristic information of the 300 motifs (information content vs. overall influence), where the color of the dots corresponds to the reproducibility of each motif. **D** Position weight matrix of known TF motifs (top) and key motifs obtained from the first layer filters of convolutional neural network (bottom). **E** Scatter plot of motifs with high reproducibility (overall influence vs. the number of peaks influenced). **F** Influence of the learned motifs on chromatin accessibility in different cell types, the full names of which are listed in Fig. 1D. Red represents positive effects and green represents negative effects. Genes mentioned in this article are highlighted in bold. **G** Motif deviations of three known neural development-related TFs (SOX6, EMX2, and NEUROD2) in five cell types (RG, nIPC, SCPN, CPN, and CThPN).

3.2. Differentiation trajectories of neocortical excitatory projection neurons

Various types of excitatory projection neurons originate from mitotic divisions of progenitor cells in the ventricular or subventricular zones and are located in the corresponding layers of the cerebral cortex during inside-out formation of each layer [9,10]. However, the differentiation trajectories of various independently developing excitatory projection neurons remain unknown. Therefore, we calculated the RNA velocity based on gene expression data and displayed it in the combined UMAP space, excluding cells of non-cortical origin and gliogenic cells (Fig. S5A). A cluster of RG_Cyc cells was differentiated toward both nIPC and RG, indicating that proliferative RG cells are the origin of neuronal differentiation of cortical origin. Nevertheless, the RNA velocity streamplot was insufficient to reveal the differentiation trajectory of each type of independently developing excitatory projection neurons, possibly owing to the limitation of UMAP in showing the differentiation process. Here, we generated a branched trajectory tree by applying the URD algorithm to transcriptome data to describe the differentiation trajectories of neocortical excitatory projection neurons (Fig. 2A and S5C; see Materials and methods) [31]. The cluster of proliferative RG cells was designated as the root, and the tips were determined based on the pseudotime distribution of the cell types (Fig. S5B). The distribution of different RG cell clusters in each segment of the branched trajectory tree showed that the number of RG1 cells was highest in segment A at the early stage of differentiation, followed by a significant increase in the proportion of RG2, RG3, and RG4 cells in segments B-1 and B-2, and then by only RG4 cells in segment D (Fig. S5D). This result suggests that RG cells continuously transit from one state to another along their differentiation trajectory. RG cells were defined as early, intermediate, and late RG cells in segments A, B, and D, respectively. We found that early RGs were progenitors of SCPNs, intermediate RGs were progenitors of superficial CPNs, and late RGs were progenitors of layer 4 CPNs, deep CPNs, and CThPNs (Fig. 2A). GO enrichment analysis of genes in different RG cells revealed that genes active in early RGs were primarily enriched in chromosome segregation and mitotic nuclear division, those active in intermediate RGs were enriched in stem cell differentiation, and those active in late RGs were enriched in positive regulation of synapse assembly (Fig. S5E). These results suggest that progenitor cell lineage specificity determines neuronal cell type. In addition, immature neurons are present in the early differentiation phase of pre-subcerebral projection neurons (preSCPNs), and preSCPNs then diverge into CSMNs and CTPNs. Superficial CPNs undergo a migration phase before reaching terminal differentiation. These results suggest that postmitotic differentiation contributes to the diversity of excitatory projection neurons. We summarized the neurogenesis of the developing human cerebral cortex as follows (Fig. 2B): 1) RG_Cyc cells continually transitioned from one state to another, giving rise successively to early, intermediate, and late RG clusters along pseudotime; 2) early and intermediate RG cells differentiated into nIPCs, and late RG cells differentiated directly into layer 4 CPNs, deep CPNs, and CThPNs. 3) Layer 4 CPNs, deep CPNs, and CThPNs migrated to form the corresponding deeper layers (layers 4, 5, and 6, respectively); 4) early RG-generated nIPCs differentiated into young neurons, which migrated to layer 5 and developed into preSCPNs

that diverged into CSMNs and CTPNs; and 5) intermediate RG-generated nIPCs differentiated into migrating neurons in the sync, which then developed into superficial CPNs settling in layer 2. Progenitor cell lineage specificity and postmitotic differentiation contribute to the diversity of excitatory projection neurons. To confirm the veracity of the neurogenesis pattern, we expanded the sample size across these gestational weeks by incorporating four previously published scMultiome-seq datasets derived from fetal cortical plates at GW 18, 19, 23, and 24 [28]. A consistent branching trajectory tree was obtained (Fig. S5F).

The combined UMAP and branched trajectory tree showed that the migrating marker genes *NEUROD1* and *NRP1* [47] were highly expressed in young neurons of SCPNs and superficial CPNs, respectively (Fig. S5G), suggesting that the migration process of SCPNs and superficial CPNs may be controlled by different migrating genes. Therefore, we hypothesized that the differentiation trajectory of each excitatory projection neuron type is controlled by lineage-specific genes. To systematically resolve genes vital to the development of each excitatory projection neuron type, we identified lineage-specific genes for each differentiation trajectory and mapped their transcriptional changes over the corresponding differentiation trajectory (see Materials and methods; Table S3). Heat maps showed that the differentiation trajectory of each excitatory projection neuron type had its own specific gene (Fig. S6A). These lineage-specific genes are expressed in a cascade along the corresponding differentiation trajectory. These results suggest that the differentiation trajectory of each excitatory projection neuron type is controlled by lineage-specific gene cascades. More specifically, the lineage-specific genes on each differentiation trajectory contained known marker genes for excitatory projection neuron types (*CUX1* and *CUX2* for superficial CPNs; *LPL, LMO4,* and *LDB2* for deep CPNs; *LPL, LMO4,* and *RORB* for layer 4 CPNs; *FOXP2, TLE4,* and *CRYM* for CThPNs; and *FOXP2* for CSMNs) and young neurons (*NEUROD1* and *NRP1*) (Fig. S6A). These results indicate the accuracy of the lineage-specific gene identification. In addition, the expression of some lineage-specific genes was significantly higher in the corresponding cells than that in the opposing cells in the branched trajectory tree (Fig. S6B). The functions of these genes were consistent with those of the corresponding cell types, such as *BCL6, ZFPM2, GRIK4, BHLHE22, PTCHD1, PBX3,* and *PLXNA2* [50–55] (Fig. S6A). Among them, *BCL6, GRIK4,* and *BHLHE22* have been associated with neuron differentiation [5,56,57]. In particular, *BCL6* plays a role in the differentiation of CSMNs [5]. These results suggest that lineage-specific genes beyond the known marker genes provide a new gene list, referred to as new genes, for the developmental control of each cell type. Moreover, the molecular trajectories of the two SCPN subpopulations determined the genes that acted unanimously in the common differentiation trajectory and separately in the specific cell types (Fig. 2C). For example, *EOMES, NEUROD1, BCL11B,* and *FEZF2* appeared chronologically in the common differentiation trajectory, *FOXP2* appeared in CSMNs, and *ZFM2* appeared in CTPNs. We identified lineage-specific genes and described their stages of action in the differentiation trajectory.

The reconstructed branched trajectory tree enabled us to identify lineage-bifurcation genes involved in cell-type divergence. To address this question, we screened the genes more differentially expressed in daughter branch cells than in parental cells and assigned an importance
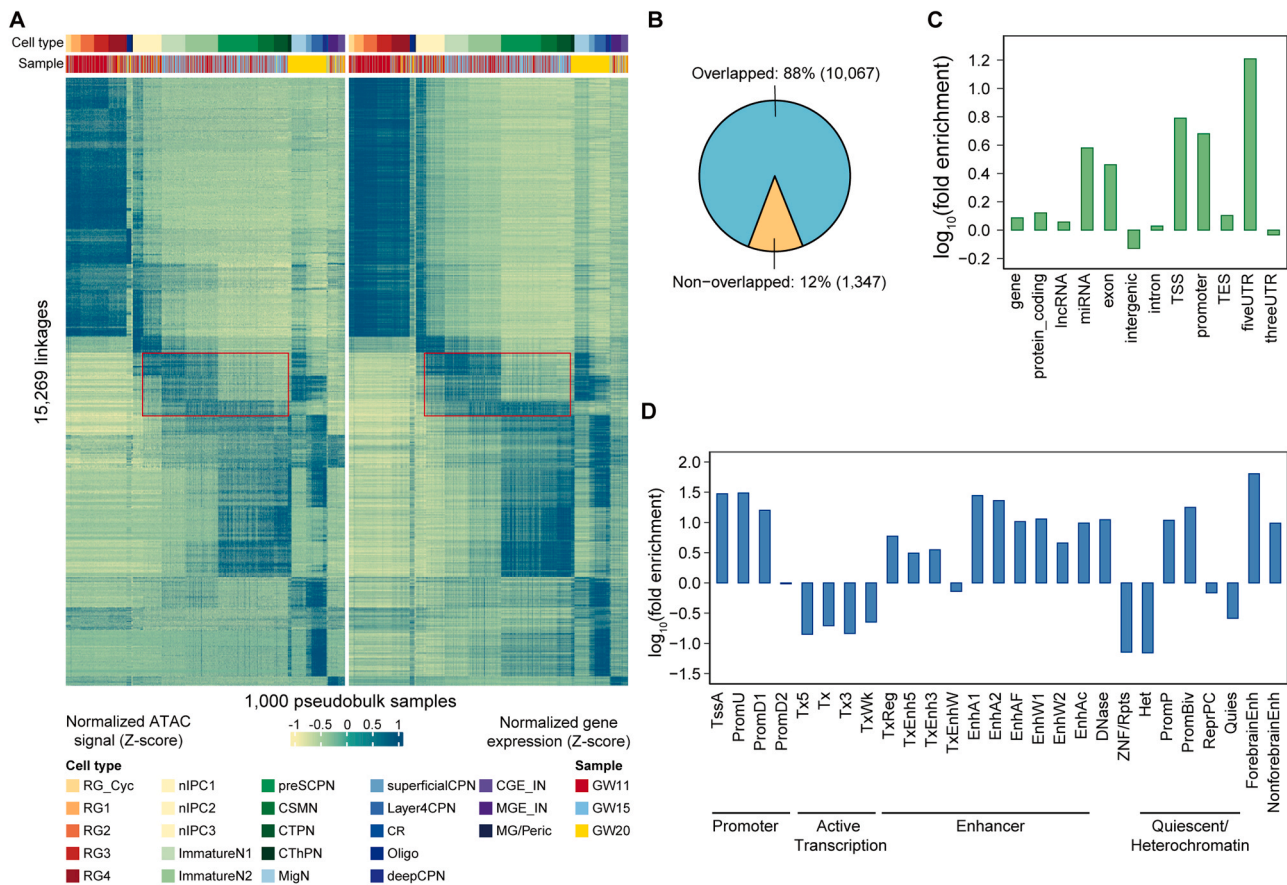
Fig. 4. Chromatin and gene-regulatory dynamics in the human cerebral cortex. A Heatmap showing chromatin accessibility (left) and gene expression (right) of 15,269 significant peak-gene links across 1000 pseudobulk samples. Each row represents a single peak and gene linkage. Columns represent 1000 pseudobulk samples. These pseudobulk samples were annotated using the combined UMAP cell types, the full names of which are listed in Fig. 1D, and fetal age and sorted according to the order of cell types in the differentiation trajectory. Yellow to dark green represents low-to-high chromatin accessibility activity (left) and gene expression levels (right). **B** Pie chart showing the percentage of linked peaks overlapped with roadmap enhancers. **C** Fold enrichment of linked peaks within defined genomic regions. **D** Fold enrichment of linked peaks in epigenetically annotated genomic regions. Fetal (E081) brain-derived chromatin 25-states were obtained from the Roadmap Epigenomics Project. In addition, validated human and non-forebrain enhancers were downloaded from the VISTA Enhancer browser.

score to each gene by training a gradient-boosting decision tree (see Materials and methods). The genes with the highest scores were assigned roles in establishing branch identity at branch points, and the top 10 genes for each daughter branch contained genes that governed the acquisition of cell identities, such as *EOMES*, *HES6*, and *BCL11B* (Fig. 2D), demonstrating the accuracy of lineage-bifurcation gene identification. Interestingly, the functions of *FAT3* and *ROBO2* [58,59] were consistent with those of their corresponding cell types (Fig. 2D), confirming the function of lineage-bifurcation genes in cell identity decisions. These results suggest that lineage-bifurcation genes other than the known genes can be novel candidate genes that play a role in controlling cell identity acquisition. We obtained a compendium of lineage-bifurcation genes associated with the decision of neuronal differentiation directions.

The function of lineage-bifurcation genes at each branch point in the branched trajectory tree was consistent with the stage of differentiation trajectory of excitatory projection neurons. For instance, *HES6* in node 1 is involved in neural cell differentiation [60], *TENM2* and *ROBO2* in node 3 are related to axon extension and synapse formation during neural development [61,62], and *NRXN3*, *NLGN1*, *SEMA3E*, *SEMA6D*, and *FAT3* in nodes 2 and 4 are associated with synaptic signal transmission [58,63–67] (Fig. S6C). This prompted us to plot gene ontology (GO) terms on the branched trajectory tree. GO terms related to cell fate specification, axon extension, and synapse organization were enriched in the daughter cells of nodes 1 and 3, and glutamatergic synaptic transmission was enriched in the daughter cells of nodes 2 and 4

(Fig. 2E). These results indicate that the bifurcation points of the branching trajectory tree match the important stages of neuronal differentiation well.

3.3. Using a deep-learning model to screen cell type specific chromatin-accessible motifs in the branching trajectory tree

Changes in the binding of TF to CREs alter CREs activity, which is the basis for phenotypic transformations during development [68,69]. Therefore, it is essential to decode the key TFs responsible for establishing and maintaining neuron-type identities during the development of excitatory projection neurons. We trained a CNN, a deep-learning model, using a recent AI-TAC framework with the peak sequences and chromatin accessibility of these peaks in each cell type, and the first-layer neurons could recognize significant sequences (or preferred motifs) representing TF-binding motifs [33,70] (Fig. 3A; see Materials and methods). The AI-TAC input was 90% of the 251-bp DNA sequences of the 100,818 peaks called out in at least one cell type, and the output was the chromatin accessibility across all cell types. The CNN model that predicted chromatin accessibility profiles across 18 cell types using genomic sequences showed good performance (Fig. 3B and S7C) and a high correlation (Fig. S7A) between the true and predicted chromatin accessibility profiles. Moreover, we trained 10 additional models with different 90% subsets of data using the same method. These models displayed a high correlation with the first CNN model, demonstrating the robustness of the first CNN model (Fig. S7B). Therefore, CNN can be
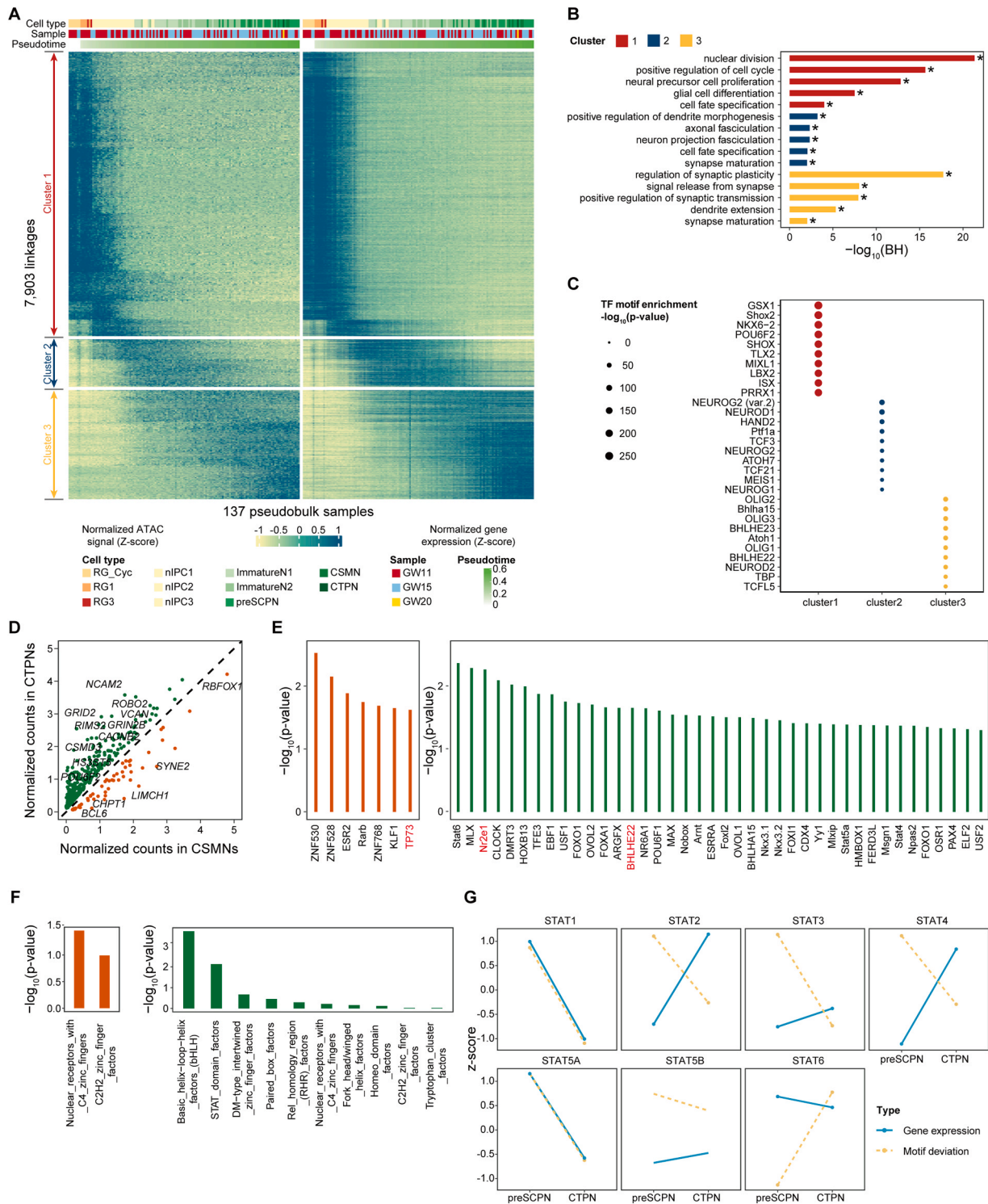
Fig. 5. Regulatory logic of SCPN specification. A Heatmap showing chromatin accessibility and gene expression of 7903 significant *cis*-regulatory elements (CREs) to cluster-specific gene pairs across 137 pseudobulk samples. These pseudobulk samples were aggregated along pseudotime bins in the preSCPN differentiation trajectory and annotated using the combined UMAP cell types, the full names of which are listed in Fig. 1D, and fetal age. Yellow to dark green represents low-to-high chromatin accessibility activity (left) and gene expression levels (right). Three interaction clusters were obtained by clustering CRE-gene pairs using k-means clustering. **B** GO enrichment analysis of genes represented in the three interaction clusters. **C** TF motif enrichment of CREs represented in three interaction clusters. **D** Normalized counts of differentially expressed genes in the CSMNs (x-axis) and CTPNs (y-axis). Differentially expressed genes were expressed higher in CSMNs or CTPNs than the other two clusters when comparing preSCPNs, CSMNs, and CTPNs. The genes mentioned in the text have been labeled. **E** TFs with significant differential enrichment of conserved motifs in CSMN and CTPN CREs. csmnTFs (left) are motifs that are found more often in CSMN CREs than in CTPN CREs. ctpnTFs (right) are motifs that are found more often in CTPN CREs than in CSMN CREs. The genes labeled in red are functionally consistent with the corresponding cell types. **F** Enrichment of TF classes for csmnTFs (left) and ctpnTFs (right). **G** Gene expression and TF motif deviation of the STAT domain factors were compared between the two cell types.
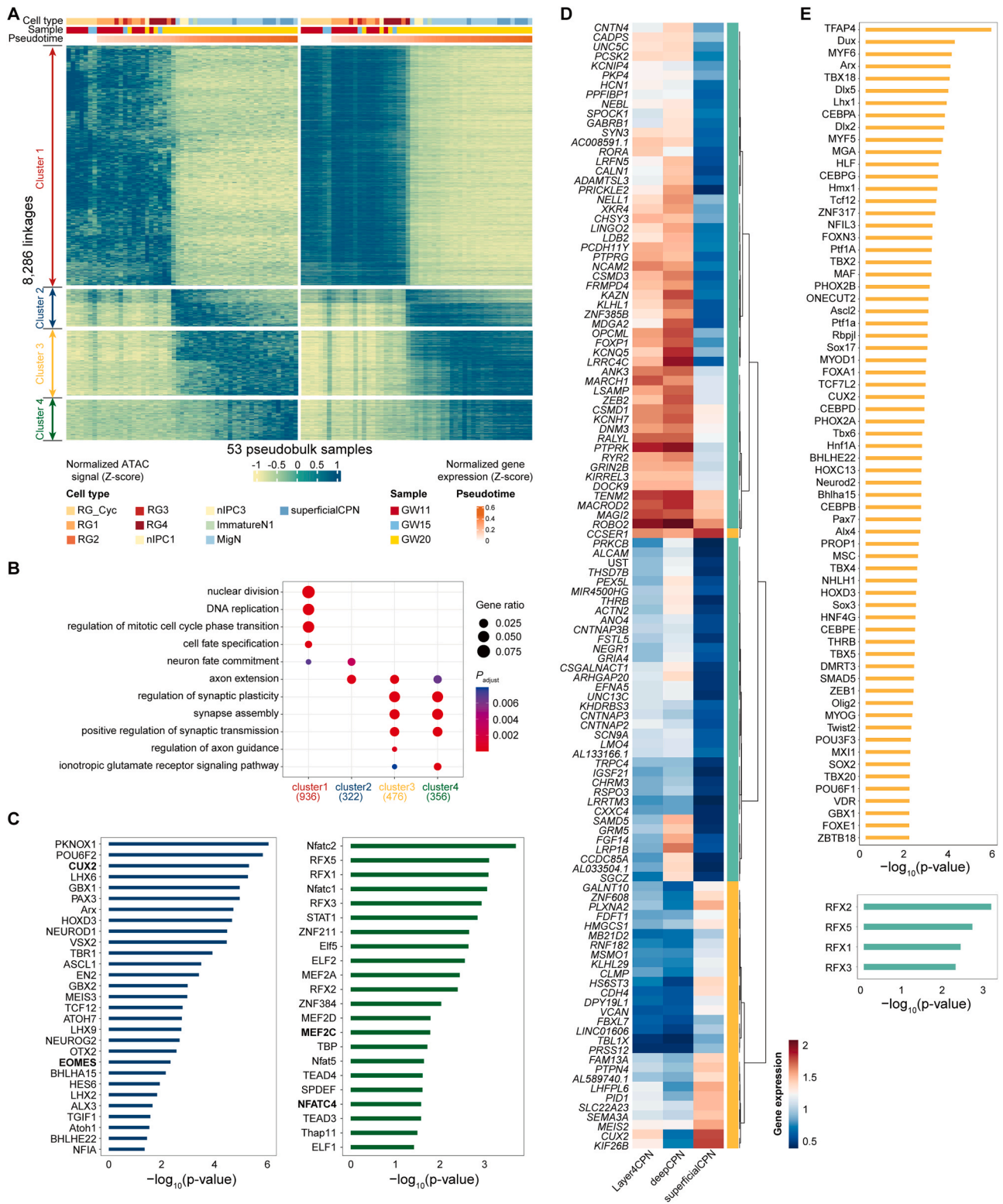
Fig. 6. Regulatory logic of CPN specification. A Heatmap showing chromatin accessibility and gene expression of 8286 significant CRE to cluster-specific gene pairs across 53 pseudobulk samples. These pseudobulk samples were aggregated along pseudotime bins in the superficial CPN differentiation trajectory and annotated using the combined UMAP cell types, the full names of which are listed in Fig. 1D, and fetal age. Yellow to dark green represents low-to-high chromatin accessibility activity (left) and gene expression levels (right). Four interaction clusters were obtained by clustering CRE-gene pairs using k-means clustering. **B** GO enrichment analysis of genes represented in the four interaction clusters. Each dot represents a gene's function. The color of the dot indicates the statistically significant level of enrichment analysis, and the size of the dot indicates the ratio of the number of genes associated with that ID in the interaction cluster to the total number of genes across that term. **C** Differentiation (left) and functional (right) TFs with significant differential enrichment in cluster2 and cluster4 CREs, respectively. **D** Heatmap showing the average expression of differentially expressed genes in superficial and deep CPNs (including layer 4 and deep CPNs). Genes are highly (yellow) and lowly (green) expressed in superficial CPNs compared to both populations of deep CPNs. **E** Superficial and deep TFs with significant differential enrichment in superficial and deep CREs, respectively.

used to obtain motifs representing TF binding motifs from the open chromatin regions of each cell type in the human cerebral cortex.

To decode the key motifs important for cell identity determination, we assessed the repeatability of the motifs extracted from the first CNN model by comparing them with the motifs from 10 additional models using the TomTom algorithm [35]. The highly reproducible motifs, which occurred in over six of the 11 CNN models, displayed higher overall influence and information content than those with lower reproducibility (Fig. 3C). Therefore, these highly reproducible motifs matched well with TF patterns, which are vital for establishing and maintaining neuron-type identity. We identified key motifs that can be used to study important TFs.

To connect these key motifs to TFs, we compared the highly reproducible motifs with TF binding motifs from the Cis-BP database using the TomTom algorithm. Motifs 21, 177, 206, and 225 matched well with the patterns of NFIB, CTCF SOX9, and EMX1, respectively (Fig. 3D), suggesting reasonable correspondence between key motifs and matching TFs. As multiple TFs could share a single motif, we calculated the correlation between chromatin accessibility and gene expression of candidate TFs for each motif related to multiple TFs and specified a TF with a higher correlation to the motif (Fig. S7D; see Materials and methods). Among the assigned TFs, we found several typical regulatory factors related to neurogenesis, including ASCL1, TCF12, HMG20A, EMX1, NEUROD6, POU3F4, MEF2C, MEIS3, and SOX6 (Fig. 3E). Other motifs, such as CTCF, a TF critical for chromatin architecture, may have complementary effects on human cerebral cortex cells. In addition, examples of highly reproducible motifs that are not matched in Cis-BP, such as motifs 28 and 185, may correspond to unidentified TFBSs associated with human cerebral cortex development and differentiation. We decoded the key TFs vital in regulating neuron-type identities and/or neuron state transitions.

To directly assess the relationship between motifs and cell types, we quantified the predicted importance of each motif across the 18 cell types. SOX6 and EMX2 were essential for progenitor cells (RGs and nIPCs), but not for mature neurons (CPNs, CThPNs, and SCPNs), whereas NEUROD6 was required for mature neurons, but not for progenitor cells (Fig. 3F). TF motif deviation analysis also showed that SOX6 and EMX2 play regulatory roles in progenitor cells and that NEUROD2 (similar to NEUROD6) plays a regulatory role in mature neurons (Fig. 3G). These results are consistent with the known roles of the corresponding TFs [56,71], indicating that we determined the effect of the cell types of each motif. On the differentiation of SCPNs and superficial CPNs, SOX6 displayed a gradually increasing negative influence, whereas NEUROD6 showed a progressively deeper positive influence (Figs. 3A and 3F), consistent with the known functions of the two genes along the differentiation trajectories. Other TFs with similar negative effects on SCPN and superficial CPN differentiation, such as HMG20A and NHLH2, and positive effect gradients on SCPN developmental progression, such as SMAD, ASCL1, MEIS3, and E2F1, should be further investigated for their functions in modulating cell differentiation. Moreover, we noticed that some TFs exhibited opposite effects in similar cell types, such as TCF12 in RGs and nIPCs, and E2F1 in SCPNs and CPNs. Some Cis-BP-unmatched motifs exhibited this trend, providing a new TF pool that may be associated with human cerebral cortex development. We determined the cell type-specific influences of key motifs vital to regulating cell-type identities and/or cell-state transitions during the development of excitatory projection neurons.

3.4. Gene-regulatory dynamics across cell types in the developing cerebral cortex

During cellular differentiation from one state to another, regulatory element activity and gene expression levels are altered sequentially or synchronously. To explore the dynamic regulation of expression across cell types in the developing human cerebral cortex, we identified 15,269 positive peak-gene links that may represent potential enhancer-gene

interactions among all cell types using LinkPeaks in Signac [72] (see Materials and methods; Table S4). All linkages were clustered across 1000 pseudobulk samples, revealing a consistent trend of chromatin accessibility and gene expression across cell types, especially in RGs (Fig. 4A). Moreover, the state of chromatin accessibility and gene expression in cells, including nIPCs, immature neurons, and preSCPNs, depicted the trajectory of young neurons to preSCPNs (Fig. 4A).

Because peak-gene links represent potential enhancer-gene interactions, we performed the following enrichment analysis of linked peaks to resolve the characteristics of these regions. First, we merged the enhancer regions (represented by H3K27ac) of seven brain samples obtained from the Roadmap Epigenomics Project to produce a set of enhancer regions. The overlap between linked peaks and the roadmap enhancer set revealed that 88% of the linked peaks intersected with known enhancers (Fig. 4B), suggesting that the linked peaks are of enhancer function. To further investigate whether linked peaks were enriched in non-enhancer regions, we performed a fold enrichment analysis of the linked peaks in the defined genome regions. We found that linked peaks were significantly enriched in regions near promoters (transcription start site, promoter, 5' untranslated region) with known functionality, but nominally depleted in intergenic regions with poorly annotated functionality (Fig. 4C), suggesting the regulatory properties of the linked peaks. Furthermore, we performed a fold enrichment analysis of the linked peaks in the epigenetically annotated genome regions derived from the chromatin state model of fetal brain tissue (E081) with 25 states [37]. The results showed that the linked peaks were enriched in enhancers and promoters but depleted in actively transcribed regions, ZNF genes and repeats, heterochromatin, and quiescent regions (Fig. 4D). More importantly, these linked peaks were enriched in validated human forebrain enhancers [40] (Fig. 4D). These findings suggest that the linked peaks are strongly correlated with enhancers. Therefore, CRE-gene pairs can help further explore chromatin and gene regulation in neuronal differentiation trajectories.

3.5. Regulatory logic of SCPN cell type specification

CSMNs and CTPNs underwent a common differentiation path, where young neurons originated from the RG-sourced nIPC division and developed into preSCPNs, and diverged from the preSCPNs (Figs. 2A and 2B). However, it is unclear how cell fate is regulated in this process, such as the regulatory logic in the differentiation of RGs to preSCPNs, and the key factors regulating the differentiation of preSCPNs to CSMNs and CTPNs.

To study the dynamics of gene-regulatory interactions in the preSCPN differentiation trajectory, we clustered CRE-gene pairs of cluster-specific genes across 137 pseudobulk samples aggregated along the pseudotime of the preSCPN differentiation trajectory, and grouped these CRE-gene pairs into three clusters (Fig. 5A). GO enrichment analysis of genes in the different clusters revealed that genes active in early development were specifically enriched in nuclear division and neural precursor cell proliferation, axonal and synapse maturation in the intermediate stage, and glutamate receptor signaling pathway regulation in the late stage (Fig. 5B). TF motif enrichment analysis of linked CREs in the different clusters showed that POU6F2 was enriched in linked CREs that were open chromatin during early development, NEUROD1 in the intermediate stage, and NEUROG1, BHLHE23, Atoh1, and OLIG1 in the late stage (Fig. 5C). We resolved the TFs vital at each stage of the differentiation process of preSCPNs.

We then investigated the key factors that played a role in the bifurcation of preSCPNs into CSMNs and CTPNs. We compared gene expression levels in the three cell types and obtained genes that were specifically expressed at the highest levels in CSMNs and CTPNs, respectively (Fig. 5D and S8). We found that genes involved in somatic movement, such as *RBFOX1*, *SYNE2*, *LIMCH1*, and *CHPT1* [73,74], were significantly upregulated in CSMNs (Fig. S8A). Genes associated with vision, such as *VCAN* and *NCAM2* [75,76], were differentially expressed

in the CTPNs (Fig. S8B). The functions of these genes are consistent with the functions of the respective differentiated cell types, suggesting the successful identification of cell-type-specific genes. Within the cell-type-specific genes, many are known to exert regulating function during neuron differentiation, or even during the differentiation of the corresponding cell types. For example, *ROBO2* and *GRIN2B*, which are differentially expressed in CTPNs, are also involved in neuronal differentiation [77,78]. *BCL6* is differentially expressed in CSMNs involved in late CSMN development [5]. These results indicated that we identified genes involved in the differentiation of preSCPNs into CSMNs and CTPNs.

We then explored CREs separately associated with differentially upregulated genes in CSMNs and CTPNs using differential motif enrichment analysis. Seven TFs were more accessible in CSMNs (csmnTFs) and 40 TFs were more accessible in CTPNs (ctpnTFs). We found that csmnTFs included TP73, a TF associated with the risk of amyotrophic lateral sclerosis [79], whereas ctpnTFs contained the TFs BHLHE22 and NR2E1, both TFs related to vision [80,81] (Fig. 5E). These TFs separately reflected the functions of CSMNs and CTPNs, demonstrating the successful identification of cell type-specific regulatory TFs. Within the TFs, many are known to relate to the differentiation regulation. For instance, TP73 promotes terminal neuronal differentiation and BHLHE22 regulates cell fate determination [56,82]. Therefore, we consider that the rest TFs that have not been reported relating the differentiation of the two cell types may also function in regulating the divergent differentiation of preSCPNs into CSMNs and CTPNs. Comparing the results of identifying differentially enriched TFs between single-modal data and multiomic data, we have found that scMultiome-seq data exhibits significant advantages in revealing the regulatory factors underlying cell differentiation (Table S5). In addition, the *HECW2* gene associated with autism spectrum disorder (ASD) stabilizes TP73 and enhances its transcription activation function [83–85]. Therefore, it is worth investigating whether TP73's regulation of the differentiation process from preSCPNs to CSMNs is genetically linked to ASD. This could contribute to a deeper understanding of the molecular mechanisms underlying ASD and provide new directions for future treatments and interventions.

Furthermore, ctpnTFs were enriched in the pro-neural STAT domain and basic helix-loop-helix (bHLH) factors that promote neuronal differentiation [86,87] (Fig. 5F). However, the csmnTFs were not enriched in any specific TF class. These results suggest that daughter CTPNs undergo stronger regulation than CSMNs during the divergence of preSCPNs into two subgroups. The dysregulation of the Janus kinase-signal transducer and activator of transcription (JAK-STAT) signaling pathway affects neurodegenerative diseases, highlighting the importance of investigating the regulatory role of STAT family genes during the differentiation process from preSCPNs to CTPNs [87]. We calculated the gene expression and TF motif deviation for the STAT family genes and observed upregulation in the expression levels or TF motif deviation of five genes (STAT2, STAT3, STAT4, STAT5B, and STAT6) during CTPN differentiation, while STAT1 and STAT5A exhibited downregulation in both gene expression levels and TF motif deviation (Fig. 5G). These results indicate that STAT family genes play regulatory roles in the differentiation process from preSCPNs to CTPNs. In particular, STAT5A, which demonstrates coordinated changes in gene expression and TF motif deviation, was identified as playing a regulatory role in the differential motif enrichment analysis (Fig. 5E). In addition, The coordinated patterns of the STAT family genes merit further investigation.

3.6. Regulatory logic of CPN cell type determination

CPNs enable the bidirectional integration of modality-specific information and coordinate multiple higher brain functions by connecting the two cerebral hemispheres. The majority of CPNs were localized in layer 2, forming superficial CPNs. To resolve the regulatory dynamics of

gene expression in the differentiation trajectory of superficial CPNs, we clustered CRE-gene pairs of cluster-specific genes across 53 pseudobulk samples aggregated along pseudotime. These interactions formed four clusters representing the early, early-intermediate, late-intermediate, and late stages of superficial CPN differentiation (Fig. 6A). GO enrichment analysis of genes in different clusters revealed that nuclear division and cell fate specification were enriched in genes active in the early stage, axon extension and guidance in the intermediate stages, and synaptic transmission in the late stage (Fig. 6B). As CREs in clusters 2 and 4 separately contribute to cell differentiation and function maintenance of superficial CPNs, we performed differential motif enrichment analysis between CREs of the two stages to identify vital TFs functioning at the two important stages. We identified 256 TFs that were more accessible in cluster 2 (differentiatedTFs) and 22 in cluster 4 (functionalTFs) (Fig. 6C; Table S6). The differentiatedTFs included EOMES and CUX2 and were enriched in homeodomain factors implicated in neural stem cell patterning and neural progenitor fate specification [86] (Fig. 6C and S9A). The functionalTFs contained NFATC4 and MEF2C which have been reported to be associated with spatial memory and cognitive impairment [88,89], and were enriched in MADS-box and Rel homology helix factors (Fig. 6C and S9B). We identified TFs that play a role in cell differentiation and functional maintenance during the differentiation of superficial CPNs.

To explore how the key differentialTFs and functionTFs regulate the differentiation trajectory of superficial CPNs, we calculated the gene expression and TF motif deviation for both types of TFs and obtained 15 differentialTFs and 19 functionTFs with both scalars (Fig. S9C and S9D). Nine differentialTFs displayed a consistent sequential decrease in both gene expression and TF motif deviation along the cell differentiation trajectory, and five showed a continuous reduction in TF motif deviation and a partial reduction in gene expression (Fig. S9C). Eight function TFs exhibited a gradual increase in both gene expression and TF motif deviation along the cell differentiation trajectory, eight showed a continuous increase in TF motif deviation and a partial increase in gene expression, and three showed a continuous increase in TF motif deviation but a decrease in gene expression (Fig. S9D). TF chromatin accessibility was more consistent with the differentiation of superficial CPNs.

To compare the differences between superficial and deep CPNs (layer 4/deep CPNs), we identified genes with differential expression levels in superficial CPNs compared to both populations of deep CPNs (Fig. 6D). Moreover, we performed a differential motif enrichment analysis between superficial and deep CREs linked with differentially expressed genes in superficial and deep CPNs, revealing 67 TFs that were more accessible in superficial CPNs and four in deep CPNs (Fig. 6E). The enrichment of neurogenesis-related genes in superficial CPNs, but not in deep CPNs, suggests that superficial CPNs are tightly controlled during neurogenesis. We identified genes and TFs that maintain the function of superficial CPNs.

4. Discussion

The process of human cerebral cortex neurogenesis occurs during the early stages of fetal development, making the acquisition of early-stage cortical plates crucial for studying the regulatory mechanisms of neuronal diversity in the cerebral cortex. In this study, we simultaneously profiled the transcriptomic and epigenomic datasets of the same individual cells derived from three early human fetal cortex tissues at GW 11, 15, and 20, and built a comprehensive single-cell atlas of combined gene expression and chromatin accessibility. Various cell types, including almost all the currently known excitatory neuron subpopulations and their young counterparts, were captured and annotated in our comprehensive atlas. In comparison, a scMultiome-seq dataset previously published by Zhu et al. from the human fetal cortical plates at GW 18, 19, 23, and 24 failed to capture SCPNs and their associated young cells (Fig. S4) [28]. Therefore, a single framework comprising all excitatory neuron types and their young counterparts, transcriptome,

and epigenome in our study provides us with the opportunity to delineate the differentiation trajectories of various independently developing excitatory projection neuron types and to identify regulatory factors, thus elucidating the mechanistic principles underlying cellular diversity.

The comprehensive annotation of excitatory projection neurons allowed us to map the differentiation trajectory of the independently developing individual neuron types. Our analysis provides novel insights into the current debate on whether fate-restricted progenitors exist [13–16], supporting the conclusion that neuronal diversity originates from both fate-restricted progenitors and postmitotic differentiation programs (Fig. 2A and B), which is consistent with a similar observation in the developing neocortex in mice [90]. Specifically, during neurogenesis, radial glial progenitor cells continually transition from one state to another, giving rise to young neurons that reflect their respective neuronal state. Young neurons then undergo a postmitotic differentiation program following the program provided by progenitor cells, resulting in various specific excitatory projection neuron types in the neocortex.

TFs that regulate neuronal differentiation and development are essential for normal neuronal development. Abnormal neuronal development can lead to neurodegenerative diseases. Therefore, decoding key TFs that regulate neuronal development during cortical development is important and clinically significant. In the present study, we identified TFs with cell type-specific effects and resolved TFs that function at critical stages of cell fate determination in SCPNs and CPNs. By elucidating these TFs, we lay the groundwork for comprehending the predominantly coordinated regulatory logic governing neuronal diversification in the neocortex. This facilitates our understanding of the molecular-level regulatory mechanisms involved at various stages of human cortical development. Furthermore, we found that some TFs are genetically linked to ASD, such as, TFE3, ARX, and MEF2C (https://gene.sfari.org/). Given that the key TFs identified in this study are regulators that play a regulatory role during neuronal differentiation and development, key TFs that have not yet been reported provide a new pool of candidate TFs for ASD research. Besides, it is unknown how genetic variants associated with ASD interfere with the genetic programs underlying cerebral cortex development [91,92]. Specifically, we still do not know in which neuronal differentiation and developmental processes TFs disrupted by ASD-associated genetic variants play a regulatory role. At what stage of neuronal genesis are the regulatory roles exerted? To explore these questions, we can subsequently analyze the enrichment of neurodegenerative disease-associated genetic variants in the TFs identified in this study. Understanding which neuronal differentiation and developmental processes, such as CPNs or SCPNs, are regulated by the TFs interfered with by the genetic variants can help us to understand the cell types and pathogenesis associated with the disease and provide valuable insights into the therapeutic targets.

Understanding the regulatory mechanisms underlying cellular differentiation programs in complex heterogeneous populations is critical to answer numerous scientific questions. Theoretically, combining the dual modalities of transcriptomic and epigenomic profiles in the scMultiome-seq dataset would be more efficient and accurate in predicting the differentiation trajectory with complex and dynamic heterogeneous subpopulations. However, current trajectory inference algorithms such as RNA velocity, Monocle3, and URD [29,31,93] have been developed solely for single-cell gene expression or chromatin accessibility data. In this study, when performing the trajectory inference using the gene expression data alone, we found there were mixed between different cell populations on the tree and that some cells were not assigned to the trajectory branching tree. These existing limitations in the current algorithms undervalue the potential of scMultiome-seq data for mining more accurate cellular differentiation programs in the developing neocortex. To realize this potential, Li et al. developed MultiVelo to improve cell fate prediction using chromatin accessibility and gene expression profiles within the same cell [94]. However, a stream plot of velocity vectors from MultiVelo was displayed in the

UMAP space, limiting the demonstration of the differentiation trajectory of each type of independently developing excitatory projection neurons. Therefore, developing appropriate trajectory inference software for single-cell multi-omics datasets will help researchers better understand complex biological processes. Moreover, the availability of suitable software will facilitate the mining of single-cell gene expression and chromatin accessibility data derived from scarce samples and sequencing in large quantities [18,95].

5. Conclusion

This study sheds light on the biological processes and regulatory mechanisms that govern the diversity of excitatory projection neurons. These findings provide a research framework for elucidating the pathogenesis of neurodegenerative diseases related to neural development and the identification of therapeutic targets.

CRediT authorship contribution statement

Fang Fu: Resources. **Zhiwei Huang:** Software. **Zhonghui Tang:** Funding acquisition, Conceptualization. **Xia Wu:** Writing – review & editing, Investigation. **Xin Yang:** Resources. **Yuhan Tian:** Writing – review & editing, Writing – original draft, Visualization, Software, Project administration, Formal analysis, Data curation. **Kunhua Hu:** Writing – review & editing, Supervision, Resources, Project administration, Funding acquisition, Conceptualization. **Dan Xiong:** Software. **Jiajun Zhang:** Funding acquisition, Conceptualization. **Songhao Luo:** Writing – review & editing, Writing – original draft, Visualization, Software, Formal analysis, Data curation. **Lanqi Hu:** Investigation. **Rong Liu:** Investigation. **Guowei Shi:** Software. **Yuchen Yuan:** Software. **Junjie Yao:** Investigation.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability

Single-cell multiome-seq data have been deposited in the GSA (PRJCA017396) and are publicly available as of the date of publication. All original codes were made publicly available at GitHub (https://github.com/YuhanTian/corticalDevelopment). Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

Declaration of competing interest

The authors have declared no competing interests.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.csbj.2024.05.019.

References

[1] Greig LC, Woodworth MB, Galazo MJ, Padmanabhan H, Macklis JD. Molecular logic of neocortical projection neuron specification, development and diversity. Nat Rev Neurosci 2013;14:755–69.

[2] Lodato S, Arlotta P. Generating neuronal diversity in the mammalian cerebral cortex. Annu Rev Cell Dev Biol 2015;31:699–720.

[3] Lui JH, Hansen DV, Kriegstein AR. Development and evolution of the human neocortex. Cell 2011;146:18–36.

[4] Sousa AMM, Meyer KA, Santpere G, Gulden FO, Sestan N. Evolution of the Human Nervous System Function, Structure, and Development. Cell 2017;170:226–47.

[5] Arlotta P, Molyneaux BJ, Chen J, Inoue J, Kominami R, Macklis JD. Neuronal subtype-specific genes that control corticospinal motor neuron development in vivo. Neuron 2005;45:207–21.

[6] Brown SP, Hestrin S. Intracortical circuits of pyramidal neurons reflect their long-range axonal targets. Nature 2009;457:1133–6.

[7] Leone DP, Srinivasan K, Chen B, Alcamo E, McConnell SK. The determination of projection neuron identity in the developing cerebral cortex. Curr Opin Neurobiol 2008;18:28–35.

[8] Kwan KY, Sestan N, Anton ES. Transcriptional co-regulation of neuronal migration and laminar identity in the neocortex. Development 2012;139:1535–46.

[9] Molnár Z., Clowry G.J., N.Š., Alzu'Bi A., Bakken T., Hevner R.F., et al. New insights into the development of the human cerebral cortex2019.

[10] Fernandez V, Llinares-Benadero C, Borrell V. Cerebral cortex expansion and folding: what have we learned? EMBO J 2016;35:1021–44.

[11] Parnavelas JG. The origin and migration of cortical neurones: new vistas 2000;23:0–131.

[12] Molyneaux BJ, Arlotta P, Menezes JR, Macklis JD. Neuronal subtype specification in the cerebral cortex. Nat Rev Neurosci 2007;8:427–37.

[13] Guo C, Eckler MJ, McKenna WL, McKinsey GL, Rubenstein JL, Chen B. Fezf2 expression identifies a multipotent progenitor for neocortical projection neurons, astrocytes, and oligodendrocytes. Neuron 2013;80:1167–74.

[14] Gao P, Postiglione MP, Krieger TG, Hernandez L, Wang C, Han Z, et al. Deterministic progenitor behavior and unitary production of neurons in the neocortex. Cell 2014;159:775–88.

[15] Franco SJ, Gil-Sanz C, Martinez-Garay I, Espinosa A, Harkins-Perry SR, Ramos C, et al. Fate-restricted neural progenitors in the mammalian cerebral cortex. Science 2012;337:746–9.

[16] Llorca A, Ciceri G, Beattie R, Wong FK, Diana G, Serafeimidou-Pouliou E, et al. A stochastic framework of neurogenesis underlies the assembly of neocortical cytoarchitecture. Elife 2019;8.

[17] Zhong S, Zhang S, Fan X, Wu Q, Yan L, Dong J, et al. A single-cell RNA-seq survey of the developmental landscape of the human prefrontal cortex. Nature 2018;555:524–8.

[18] Trevino AE, Muller F, Andersen J, Sundaram L, Kathiria A, Shcherbina A, et al. Chromatin and gene-regulatory dynamics of the developing human cerebral cortex at single-cell resolution. Cell 2021;184:5053–69. e23.

[19] Ziffra RS, Kim CN, Ross JM, Wilfert A, Turner TN, Haeussler M, et al. Single-cell epigenomics reveals mechanisms of human cortical development. Nature 2021;598:205–13.

[20] Ma A, Xin G, Ma Q. The use of single-cell multi-omics in immuno-oncology. Nat Commun 2022;13:2728.

[21] Lau SF, Cao H, Fu AKY, Ip NY. Single-nucleus transcriptome analysis reveals dysregulation of angiogenic endothelial cells and neuroprotective glia in Alzheimer's disease. Proc Natl Acad Sci USA 2020;117:25800–9.

[22] Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck 3rd WM, et al. Comprehensive Integration of Single-Cell Data. Cell 2019;177:1888–902. e21.

[23] Stuart T, Srivastava A, Madad S, Lareau CA, Satija R. Single-cell chromatin state analysis with Signac. Nat Methods 2021;18:1333–41.

[24] McGinnis CS, Murrow LM, Gartner ZJ. DoubletFinder: Doublet Detection in Single-Cell RNA Sequencing Data Using Artificial Nearest Neighbors. Cell Syst 2019;8:329–37. e4.

[25] Hao Y, Hao S, Andersen-Nissen E, Mauck 3rd WM, Zheng S, Butler A, et al. Integrated analysis of multimodal single-cell data. Cell 2021;184:3573–87. e29.

[26] Feng J, Liu T, Qin B, Zhang Y, Liu XS. Identifying ChIP-seq enrichment using MACS. Nat Protoc 2012;7:1728–40.

[27] Korsunsky I, Millard N, Fan J, Slowikowski K, Zhang F, Wei K, et al. Fast, sensitive and accurate integration of single-cell data with Harmony. Nat Methods 2019;16:1289–96.

[28] Zhu X, Bendl J, Rahman S, Vicari JM, Coleman C, Clarence T, et al. Multi-omic profiling of the developing human cerebral cortex at the single-cell level. Sci Adv 2023;9:eadg3754.

[29] La Manno G, Soldatov R, Zeisel A, Braun E, Hochgerner H, Petukhov V, et al. RNA velocity of single cells. Nature 2018;560:494–8.

[30] Bergen V, Lange M, Peidli S, Wolf FA, Theis FJ. Generalizing RNA velocity to transient cell states through dynamical modeling. Nat Biotechnol 2020;38:1408–14.

[31] Farrell JA, Wang Y, Riesenfeld SJ, Shekhar K, Regev A, Schier AF. Single-cell reconstruction of developmental trajectories during zebrafish embryogenesis. Science 2018;360.

[32] Pedregosa A, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O. Scikit-learn: Machine Learning in Python. The Journal of Machine Learning Research 2011;12:2825–30.

[33] Maslova A, Ramirez RN, Ma K, Schmutz H, Wang C, Fox C, et al. Deep learning of immune cell differentiation. Proc Natl Acad Sci USA 2020;117:25655–66.

[34] Kingma D, Ba JJCS. Adam: A Method Stoch Optim 2014.

[35] Gupta S, Stamatoyannopoulos JA, Bailey TL, Noble WS. Quantifying similarity between motifs. Genome Biol 2007;8:R24.

[36] Weirauch MT, Yang A, Albu M, Cote AG, Montenegro-Montero A, Drewe P, et al. Determination and inference of eukaryotic transcription factor sequence specificity. Cell 2014;158:1431–43.

[37] Roadmap Epigenomics C, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, et al. Integrative analysis of 111 reference human epigenomes. Nature 2015;518:317–30.

[38] de la Torre-Ubieta L, Stein JL, Won H, Opland CK, Liang D, Lu D, et al. The Dynamic Landscape of Open Chromatin during Human Cortical Neurogenesis. Cell 2018;172:289–304. e18.

[39] Ernst J, Kellis M. Large-scale imputation of epigenomic datasets for systematic annotation of diverse human tissues. Nat Biotechnol 2015;33:364–76.

[40] Visel A, Minovitsky S, Dubchak I, Pennacchio LA. VISTA Enhancer Browser–a database of tissue-specific human enhancers. Nucleic Acids Res 2007;35:D88–92.

[41] Castro-Mondragon JA, Riudavets-Puig R, Rauluseviciute I, Lemma RB, Turchi L, Blanc-Mathieu R, et al. JASPAR 2022: the 9th release of the open-access database of transcription factor binding profiles. Nucleic Acids Res 2022;50:D165. D73.

[42] Pollard KS, Hubisz MJ, Rosenbloom KR, Siepel A. Detection of nonneutral substitution rates on mammalian phylogenies. Genome Res 2010;20:110–21.

[43] Fornes O, Castro-Mondragon JA, Khan A, van der Lee R, Zhang X, Richmond PA, et al. JASPAR 2020: update of the open-access database of transcription factor binding profiles. Nucleic Acids Res 2020;48:D87–92.

[44] Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS 2012;16:284–7.

[45] McConnell SK. Constructing the cerebral cortex: neurogenesis and fate determination. Neuron 1995;15:761–8.

[46] Polioudakis D, de la Torre-Ubieta L, Langerman J, Elkins AG, Shi X, Stein JL, et al. A Single-Cell Transcriptomic Atlas of Human Neocortical Development during Mid-gestation. Neuron 2019;103:785–801. e8.

[47] Di Bella DJ, Habibi E, Stickels RR, Scalia G, Brown J, Yadollahpour P, et al. Molecular logic of cellular diversification in the mouse cerebral cortex. Nature 2021;595:554–9.

[48] Martinez-Cerdeno V, Noctor SC. Cajal, Retzius, and Cajal-Retzius cells. Front Neuroanat 2014;8:48.

[49] Tomassy GS, De Leonibus E, Jabaudon D, Lodato S, Alfano C, Mele A, et al. Area-specific temporal control of corticospinal motor neuron differentiation by COUP-TFI. Proc Natl Acad Sci USA 2010;107:3576–81.

[50] Fujii T, Iijima Y, Kondo H, Shizuno T, Hori H, Nakabayashi T, et al. Failure to confirm an association between the PLXNA2 gene and schizophrenia in a Japanese population. Prog Neuropsychopharmacol Biol Psychiatry 2007;31:873–7.

[51] Knight HM, Walker R, James R, Porteous DJ, Muir WJ, Blackwood DH, et al. GRIK4/KA1 protein expression in human brain and correlation with bipolar disorder risk variant status. Am J Med Genet B Neuropsychiatr Genet 2012;159B:21–9.

[52] Leamey CA, Glendining KA, Kreiman G, Kang ND, Wang KH, Fassler R, et al. Differential gene expression between sensory neocortical areas: potential roles for Ten_m3 and Bcl6 in patterning visual and somatosensory pathways. Cereb Cortex 2008;18:53–66.

[53] Pastore SF, Ko SY, Frankland PW, Hamel PA, Vincent JB. PTCHD1: Identification and Neurodevelopmental Contributions of an Autism Spectrum Disorder and Intellectual Disability Susceptibility Gene. Genes (Basel) 2022;13.

[54] Ramirez M, Badayeva Y, Yeung J, Wu J, Abdalla-Wyse A, Yang E, et al. Temporal analysis of enhancers during mouse cerebellar development reveals dynamic and novel regulatory functions. Elife 2022;11.

[55] Wiggs JL, Yaspan BL, Hauser MA, Kang JH, Allingham RR, Olson LM, et al. Common variants at 9p21 and 8q22 are associated with increased susceptibility to optic nerve degeneration in glaucoma. PLoS Genet 2012;8:e1002654.

[56] Tutukova S, Tarabykin V, Hernandez-Miranda LR. The Role of Neurod Genes in Brain Development, Function, and Disease. Front Mol Neurosci 2021;14:662774.

[57] Li Y, Smith JJ, Marques F, Osuma A, Huang HC, Kratsios P. Cell context-dependent CFI-1/ARID3 functions control neuronal terminal differentiation. Cell Rep 2023;42:112220.

[58] Nagae S, Tanoue T, Takeichi M. Temporal and spatial expression profiles of the Fat3 protein, a giant cadherin molecule, during mouse development. Dev Dyn 2007;236:534–43.

[59] Gonda Y, Andrews WD, Tabata H, Namba T, Parnavelas JG, Nakajima K, et al. Robo1 regulates the migration and laminar distribution of upper-layer pyramidal neurons of the cerebral cortex. Cereb Cortex 2013;23:1495–508.

[60] Gratton MO, Torban E, Jasmin SB, Theriault FM, German MS, Stifani S. Hes6 promotes cortical neurogenesis and inhibits Hes1 transcription repression activity by multiple mechanisms. Mol Cell Biol 2003;23:6922–35.

[61] Vysokov NV, Silva JP, Lelianova VG, Ho C, Djamgoz MB, Tonevitsky AG, et al. The Mechanism of Regulated Release of Lasso/Teneurin-2. Front Mol Neurosci 2016;9:59.

[62] Blockus H, Rolotti SV, Szoboszlay M, Peze-Heidsieck E, Ming T, Schroeder A, et al. Synaptogenic activity of the axon guidance molecule Robo2 underlies hippocampal circuit function. Cell Rep 2021;37:109828.

[63] Deck M, Lokmane L, Chauvet S, Mailhes C, Keita M, Niquille M, et al. Pathfinding of corticothalamic axons relies on a rendezvous with thalamic projections. Neuron 2013;77:472–84.

[64] Hishimoto A, Pletnikova O, Lang DL, Troncoso JC, Egan JM, Liu QR. Neurexin 3 transmembrane and soluble isoform expression and splicing haplotype are associated with neuron inflammasome and Alzheimer's disease. Alzheimers Res Ther 2019;11:28.

[65] Feng P, Akladious AA, Hu Y. Hippocampal and motor fronto-cortical neuroligin1 is increased in an animal model of depression. Psychiatry Res 2016;243:210–8.

[66] Gu Z, Kalambogias J, Yoshioka S, Han W, Li Z, Kawasawa YI, et al. Control of species-dependent cortico-motoneuronal connections underlying manual dexterity. Science 2017;357:400–4.

[67] Song W, Lin GN, Yu S, Zhao M. Genome-wide identification of the shared genetic basis of cannabis and cigarette smoking and schizophrenia implicates NCAM1 and neuronal abnormality. Psychiatry Res 2022;310:114453.

[68] Stergachis A.B., Neph S., Reynolds A., Humbert R., Miller B., Paige S.L., et al. Developmental Fate and Cellular Maturity Encoded in Human Regulatory DNA Landscapes 2013;154:888–903.

[69] Buenrostro JD, Corces MR, Lareau CA, Wu B, Schep AN, Aryee MJ, et al. Integrated single-cell analysis maps the continuous regulatory landscape of human hematopoietic differentiation. Cell 2018;173.

[70] Novakovsky G, Dexter N, Libbrecht MW, Wasserman WW, Mostafavi S. Obtaining genetics insights from deep learning via explainable artificial intelligence. Nat Rev Genet 2023;24:125–37.

[71] Lim JW, Bunt J, Bridges CR, Moey C, Jacobs MT, Chen K-S, et al. 2021 EMX2 Transcr Regul Nfib Expr Neural Progenit Cells Early cortical Dev 2021;12(26): 474186.

[72] Ma S, Zhang B, LaFave LM, Earl AS, Chiang Z, Hu Y, et al. Chromatin Potential Identified by Shared Single-Cell Profiling of RNA and Chromatin. Cell 2020;183: 1103–16. e20.

[73] O'Leary A, Fernandez-Castillo N, Gan G, Yang Y, Yotova AY, Kranz TM, et al. Behavioural and functional evidence revealing the role of RBFOX1 variation in multiple psychiatric disorders and traits. Mol Psychiatry 2022;27:4464–73.

[74] Lin YH, Zhen YY, Chien KY, Lee IC, Lin WC, Chen MY, et al. LIMCH1 regulates nonmuscle myosin-II activity and suppresses cell migration. Mol Biol Cell 2017;28: 1054–65.

[75] Parcerisas A, Ortega-Gasco A, Pujadas L, Soriano E. The Hidden Side of NCAM Family: NCAM2, a Key Cytoskeleton Organization Molecule Regulating Multiple Neural Functions. Int J Mol Sci 2021;22.

[76] Li H, Li H, Yang L, Sun Z, Wu S, Sui R. Clinical and genetic study on two Chinese families with Wagner vitreoretinopathy. Ophthalmic Genet 2020;41:432–9.

[77] Couch JA, Chen J, Rieff HI, Uri EM, Condron BG. robo2 and robo3 interact with eagle to regulate serotonergic neuron differentiation. Development 2004;131: 997–1006.

[78] Bell S, Maussion G, Jefri M, Peng H, Theroux JF, Silveira H, et al. Disruption of GRIN2B Impairs Differentiation in Human Neurons. Stem Cell Rep 2018;11: 183–96.

[79] Russell KL, Downie JM, Gibson SB, Tsetsou S, Keefe MD, Duran JA, et al. Pathogenic effect of TP73 Gene Variants in People With Amyotrophic Lateral Sclerosis. Neurology 2021;97:e225–35.

[80] Homman-Ludiye J, Bourne JA. Mapping arealisation of the visual cortex of non-primate species: lessons for development and evolution. Front Neural Circuits 2014;8:79.

[81] Zhang CL, Zou Y, Yu RT, Gage FH, Evans RM. Nuclear receptor TLX prevents retinal dystrophy and recruits the corepressor atrophin1. Genes Dev 2006;20: 1308–20.

[82] Nemajerova A, Moll UM. Tissue-specific roles of p73 in development and homeostasis. J Cell Sci 2019;132.

[83] Berko ER, Cho MT, Eng C, Shao Y, Sweetser DA, Waxler J, et al. De novo missense variants in HECW2 are associated with neurodevelopmental delay and hypotonia. J Med Genet 2017;54:84–6.

[84] Iossifov I, Ronemus M, Levy D, Wang Z, Hakker I, Rosenbaum J, et al. De novo gene disruptions in children on the autistic spectrum. Neuron 2012;74:285–99.

[85] Krumm N, O'Roak BJ, Shendure J, Eichler EE. A de novo convergence of autism genetics and molecular neuroscience. Trends Neurosci 2014;37:95–105.

[86] Guillemot F. Spatial and temporal specification of neural fates by transcription factor codes. Development 2007;134:3771–80.

[87] Nicolas CS, Amici M, Bortolotto ZA, Doherty A, Csaba Z, Fafouri A, et al. The role of JAK-STAT signaling within the CNS. JAKSTAT 2013;2:e22925.

[88] Wang J, Zhang Q, Chen Y, Yu S, Wu X, Bao X, et al. Novel MEF2C point mutations in Chinese patients with Rett (-like) syndrome or non-syndromic intellectual disability: insights into genotype-phenotype correlation. BMC Med Genet 2018;19: 191.

[89] Groth RD, Mermelstein PG. Brain-derived neurotrophic factor activation of NFAT (nuclear factor of activated T-cells)-dependent transcription: a role for the transcription factor NFATc4 in neurotrophin-mediated gene expression. J Neurosci 2003;23:8125–34.

[90] Telley L, Agirman G, Prados J, Amberg N, Fievre S, Oberst P, et al. Temporal patterning of apical progenitors and their daughter neurons in the developing neocortex. Science 2019;364.

[91] Rubenstein JL. Annual Research Review: Development of the cerebral cortex: implications for neurodevelopmental disorders. J Child Psychol Psychiatry 2011; 52:339–55.

[92] Zhou J, Park CY, Theesfeld CL, Wong AK, Yuan Y, Scheckel C, et al. Whole-genome deep-learning analysis identifies contribution of noncoding mutations to autism risk. Nat Genet 2019;51:973–80.

[93] Cao J, Spielmann M, Qiu X, Huang X, Ibrahim DM, Hill AJ, et al. The single-cell transcriptional landscape of mammalian organogenesis. Nature 2019;566: 496–502.

[94] Li C, Virgilio MC, Collins KL, Welch JD. Multi-omic single-cell velocity models epigenome-transcriptome interactions and improves cell fate prediction. Nat Biotechnol 2023;41:387–98.

[95] Morabito S, Miyoshi E, Michael N, Shahin S, Martini AC, Head E, et al. Single-nucleus chromatin accessibility and transcriptomic characterization of Alzheimer's disease. Nat Genet 2021;53:1143–55.