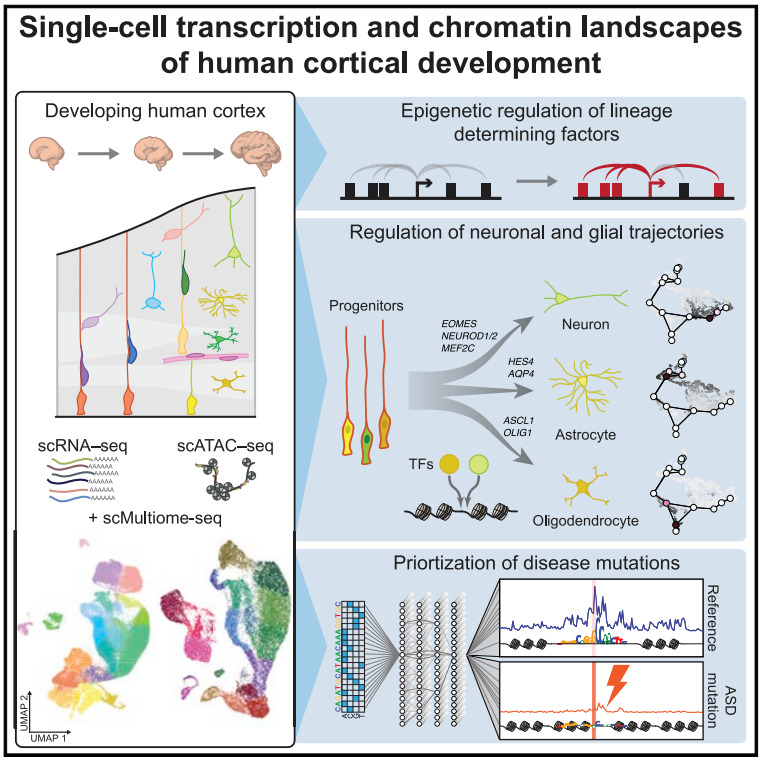
Cell

Resource

Chromatin and gene-regulatory dynamics of the developing human cerebral cortex at single-cell resolution

Graphical abstract



HIGHLIGHTS

- Single-cell RNA and chromatin profiling charts human corticogenesis
- Distinct TFs underlie neurogenesis and gliogenesis regulatory programs
- Lineage-determining TFs adopt an active chromatin state early in differentiation
- Neural networks prioritize noncoding *de novo* mutations in autism spectrum disorder

Authors

Alexandro E. Trevino, Fabian Müller, Jimena Andersen, ..., Anshul Kundaje, Sergiu P. Pașca, William J. Greenleaf

Correspondence

spasca@stanford.edu (S.P.P.), wjg@stanford.edu (W.J.G.)

In brief

A single-cell atlas of gene expression and chromatin accessibility of human developing cortex during mid-gestation reveals lineage-determining transcription factors for human corticogenesis and identifies prioritized mutations for autism spectrum disorder.

Trevino et al., 2021, Cell *184*, 1–17
September 16, 2021
https://doi.org/10.1016/j.cell.2021.07.039

CellPress

Please cite this article in press as: Trevino et al., Chromatin and gene-regulatory dynamics of the developing human cerebral cortex at single-cell resolution, Cell (2021), https://doi.org/10.1016/j.cell.2021.07.039

Cell

CellPress

Resource

Chromatin and gene-regulatory dynamics of the developing human cerebral cortex at single-cell resolution

Alexandro E. Trevino,[1,13] Fabian Müller,[1,2,13] Jimena Andersen,[3,4,13] Laksshman Sundaram,[5,13] Arwa Kathiria,[1] Anna Shcherbina,[6] Kyle Farh,[7] Howard Y. Chang,[1,8,9] Anca M. Pașca,[10] Anshul Kundaje,[1,5] Sergiu P. Pașca,[3,4,14,*] and William J. Greenleaf[1,11,12,*]

[1]Department of Genetics, Stanford University, Stanford, CA, USA
[2]Center for Bioinformatics, Saarland University, Saarbrücken, Germany
[3]Department of Psychiatry and Behavioral Sciences, Stanford University, Stanford, CA, USA
[4]Stanford Brain Organogenesis Program, Wu Tsai Neuroscience Institute Stanford University, Stanford, CA, USA
[5]Department of Computer Science, Stanford University, Stanford, CA, USA
[6]Biomedical Data Science Program, Stanford University, Stanford CA, USA
[7]Illumina Artificial Intelligence Laboratory, Illumina Inc, San Diego, CA, USA
[8]Center for Personal Dynamic Regulomes, Stanford University, Stanford, CA, USA
[9]Howard Hughes Medical Institute, Stanford University, Stanford, CA, USA
[10]Department of Pediatrics, Division of Neonatology, Stanford University, Stanford, CA, USA
[11]Department of Applied Physics, Stanford University, Stanford, CA, USA
[12]Chan–Zuckerberg Biohub, San Francisco, CA, USA
[13]These authors contributed equally
[14]Lead contact
*Correspondence: spasca@stanford.edu (S.P.P.), wjg@stanford.edu (W.J.G.)
https://doi.org/10.1016/j.cell.2021.07.039

SUMMARY

Genetic perturbations of cortical development can lead to neurodevelopmental disease, including autism spectrum disorder (ASD). To identify genomic regions crucial to corticogenesis, we mapped the activity of gene-regulatory elements generating a single-cell atlas of gene expression and chromatin accessibility both independently and jointly. This revealed waves of gene regulation by key transcription factors (TFs) across a nearly continuous differentiation trajectory, distinguished the expression programs of glial lineages, and identified lineage-determining TFs that exhibited strong correlation between linked gene-regulatory elements and expression levels. These highly connected genes adopted an active chromatin state in early differentiating cells, consistent with lineage commitment. Base-pair-resolution neural network models identified strong cell-type-specific enrichment of noncoding mutations predicted to be disruptive in a cohort of ASD individuals and identified frequently disrupted TF binding sites. This approach illustrates how cell-type-specific mapping can provide insights into the programs governing human development and disease.

INTRODUCTION

Dynamic changes in activity of *cis*-regulatory DNA elements, driven by changes in transcription factor (TF) binding, underlie phenotypic transformations during development (Buenrostro et al., 2018; Stergachis et al., 2013). Single-cell methods for measuring chromatin accessibility have emerged as a sensitive probe for this activity and, combined with tools to measure single-cell transcriptomes, have the potential to decipher how combinations of TFs drive gene expression programs (Kelsey et al., 2017; Klemm et al., 2019). Quantifying the dynamic activity of regulatory elements also enables the inference of the time point or cell type wherein disease-associated genetic variation impacts development. For instance, it is still unknown how genetic variants associated with autism spectrum disorder (ASD) interfere with the genetic programs underlying the development of the cerebral cortex (Rubenstein, 2011; Zhou et al., 2019).

Corticogenesis is a dynamic, highly regulated process characterized by the expansion of apical and basal radial glia (RG) and intermediate progenitors in the ventricular and subventricular zones (VZ, SVZ), the inside-out generation of glutamatergic neurons, and the differentiation of astrocytes and oligodendrocytes (Greig et al., 2013; Molnár et al., 2019; Silbereis et al., 2016). Cell types derived outside of the dorsal forebrain, including GABAergic neurons, microglia, and some oligodendrocytes, also migrate and integrate into the cortex (Wonders and

Please cite this article in press as: Trevino et al., Chromatin and gene-regulatory dynamics of the developing human cerebral cortex at single-cell resolution, Cell (2021), https://doi.org/10.1016/j.cell.2021.07.039

Anderson, 2006). Resolving gene-regulatory dynamics associated with these developmental trajectories requires investigation of both chromatin and gene expression states at single-cell resolution.

To map the gene-regulatory logic of human corticogenesis, we generated single-cell chromatin accessibility and RNA expression profiles from human fetal cortical samples spanning 8 weeks during mid-gestation. These paired maps revealed a class of genes with comparatively large numbers of nearby putative enhancers whose accessibility was strongly predictive of gene expression. These genes with predictive chromatin (GPCs) are frequently TFs, and we observed that their local accessibility precedes lineage-specific gene expression in cycling progenitors. We validated these findings using single-cell accessibility and expression profiles derived from the same cell (multiomics). We defined a developmental trajectory for cortical glutamatergic neurons, revealing a continuous progression of TF motif activities associated with neuronal specification and migration, and explored the co-dependencies in TF motif accessibility along this trajectory. In addition, we characterized the lineage potential of glial progenitors and provided evidence for two distinct astrocyte precursor subtypes. Finally, we trained a deep-learning model to infer base-pair-resolved, cell-type-specific chromatin accessibility profiles from DNA sequence. These models allowed prediction of the potential impact of genetic variants on the cell-type-specific chromatin landscape and prioritized rare *de novo* genetic variants associated with ASD, demonstrating the ability to map disease risk with single-cell and single-base resolution during cortical development.

RESULTS

A single-cell regulatory atlas of the developing human cerebral cortex

To capture cellular heterogeneity in the cerebral cortex, we created a gene-regulatory atlas using the Chromium platform (10x Genomics) to generate single-cell assay for transposase-accessible chromatin with sequencing (scATAC-seq) and single-cell RNA sequencing (scRNA-seq) libraries from four primary samples at post-conceptional week (PCW) 16, PCW20, PCW21, and PCW24 (Figure 1A). Overall, we obtained 57,868 single-cell transcriptomes and 31,304 single-cell epigenomes after quality control and filtering (Table S1; Figures S1A–S1H). Consistent with previous studies (Fietz et al., 2010; Hansen et al., 2010; Kang et al., 2011; Pollen et al., 2015; Trevino et al., 2020), CTIP2$^+$ cells were present in the cortical plate (CP) and SOX9$^+$ cells in the VZ, SVZ, and outer SVZ (oSVZ; Figures 1B and S1I), while the GFAP$^+$ scaffolding spanned the neocortex at PCW17 and PCW21 (Figures 1C and S1J). The proliferation marker KI67 colocalized with both GFAP$^+$ cells and with PPP1R17$^+$ intermediate progenitor cells (IPCs) in the SVZ and oSVZ (Figures 1C and S1J).

To assess global similarities and differences between individual cells, we performed unsupervised analyses, including dimension reduction using uniform manifold approximation and projection (UMAP) and clustering. For scATAC-seq, we employed an iterative approach (Granja et al., 2019) to obtain a low-dimensional embedding, cell clustering, and a consensus set of 657,930 accessible peaks representing potential *cis*-regulatory elements (CREs; STAR Methods). The structure of the RNA and chromatin representations were similar, with variation related to gestational time (Figure 1D) and cell types. Performing both assays on the same samples enabled us to dissect complementary aspects of gene regulation, including the relationship between gene expression (scRNA-seq) and chromatin accessibility-based gene activity score (scATAC-seq)—a metric defined by the aggregate local chromatin accessibility of genes (hereafter "gene activity score"; STAR Methods) (Pliner et al., 2018) as well as aggregate TF motif activity scores (Schep et al., 2017). Corticogenesis TFs such as *SOX9*, *EOMES*, *NEUROD2*, and *DLX2* showed strong cluster-specific enrichments in these three metrics (Figure 1E) consistent with their ascribed roles in RG, IPCs, cortical glutamatergic neurons (GluN), and GABAergic neurons (interneuron; IN), respectively.

We next called clusters in both datasets (Figure 1F; STAR Methods) and annotated these clusters using gene expression and gene activities of known markers (Lui et al., 2011; McConnell, 1995; Nowakowski et al., 2016, 2017; Polioudakis et al., 2019; Pollen et al., 2015; Thomsen et al., 2016) (Figures 1G–H, S2A, and S2B; Table S1; STAR Methods). In scRNA-seq, we observed a cluster of cycling cells (Cyc) expressing *TOP2A* and *KI67*. We also found that RG, expressing *SOX9* and *HES1*, included both ventricular radial glia (vRG: *FBXO32*, *CTGF*) and outer radial glia (oRG: *MOXD1*, *HOPX*), and these were separated according to time (early RG, PCW16: *NPY*, *FGFR3*; late RG, PCW20–24: *CD9*, *GPX3*). Cells in one scRNA-seq cluster expressed markers for truncated RG (tRG) and ependymal cells (tRG: *CRYAB*, *NR4A1*, *FOXJ1*). We also identified a cluster expressing genes associated with both RGs and oligodendrocyte lineage precursors (*ASCL1*, *OLIG2*, *PDGFRA*, *EGFR*). This cluster, which we named multipotent glial progenitor cells (mGPC), was different from the OPC and oligodendrocyte (OPC/Oligo) cluster that expressed *SOX10*, *NKX2.2*, and *MBP*. Genes associated with astrocyte identity (*AQP4*, *APOE*) were observed in the mGPC cluster as well as in the late RG cluster. A large domain was composed of neuronal IPC (*EOMES*, *PPP1R17*, *NEUROG1*) and GluN (*BCL11B*/*CTIP2*, *SATB2*, and *SLC17A7*/*VGLUT1*). Among the GluN clusters, we found cells expressing subplate markers (SP: *NR4A2*, *CRYM*). We also identified distinct clusters of IN expressing *DLX2* and *GAD2*—one of them expressed markers associated with medial ganglionic eminence (MGE: *LHX6*, *SST*) and the other expressed markers associated with both caudal ganglionic eminence and pallial-subpallial boundary (CGE: *SP8*, *NR2F2*; PSB: *MEIS2*, *ETV1*). In addition, we observed clusters of microglia (MG: *AIF1*, *CCL3*), endothelial cells (EC: *CLDN5*, *PECAM1*), pericytes (Peric: *FOXC2*, *PDGFRB*), leptomeningeal cells (VLMC: *COL1A1*, *LUM*), and red blood cells (RBC: *HEMGN*). Many of the above markers exhibited dynamic gene activity scores in corresponding clusters in scATAC-seq space (Figure 1H). While most clusters had cells representing all time points, some were strongly biased for earlier or later stages (e.g., mGPCs and tRGs; Figure S2C). To further corroborate cell-type identities and gestational time, we projected two previously published scRNA-seq datasets from human cortex into our scRNA-seq manifold (Bhaduri et al., 2020; Polioudakis et al., 2019). We computed Jaccard

Please cite this article in press as: Trevino et al., Chromatin and gene-regulatory dynamics of the developing human cerebral cortex at single-cell resolution, Cell (2021), https://doi.org/10.1016/j.cell.2021.07.039

CellPress

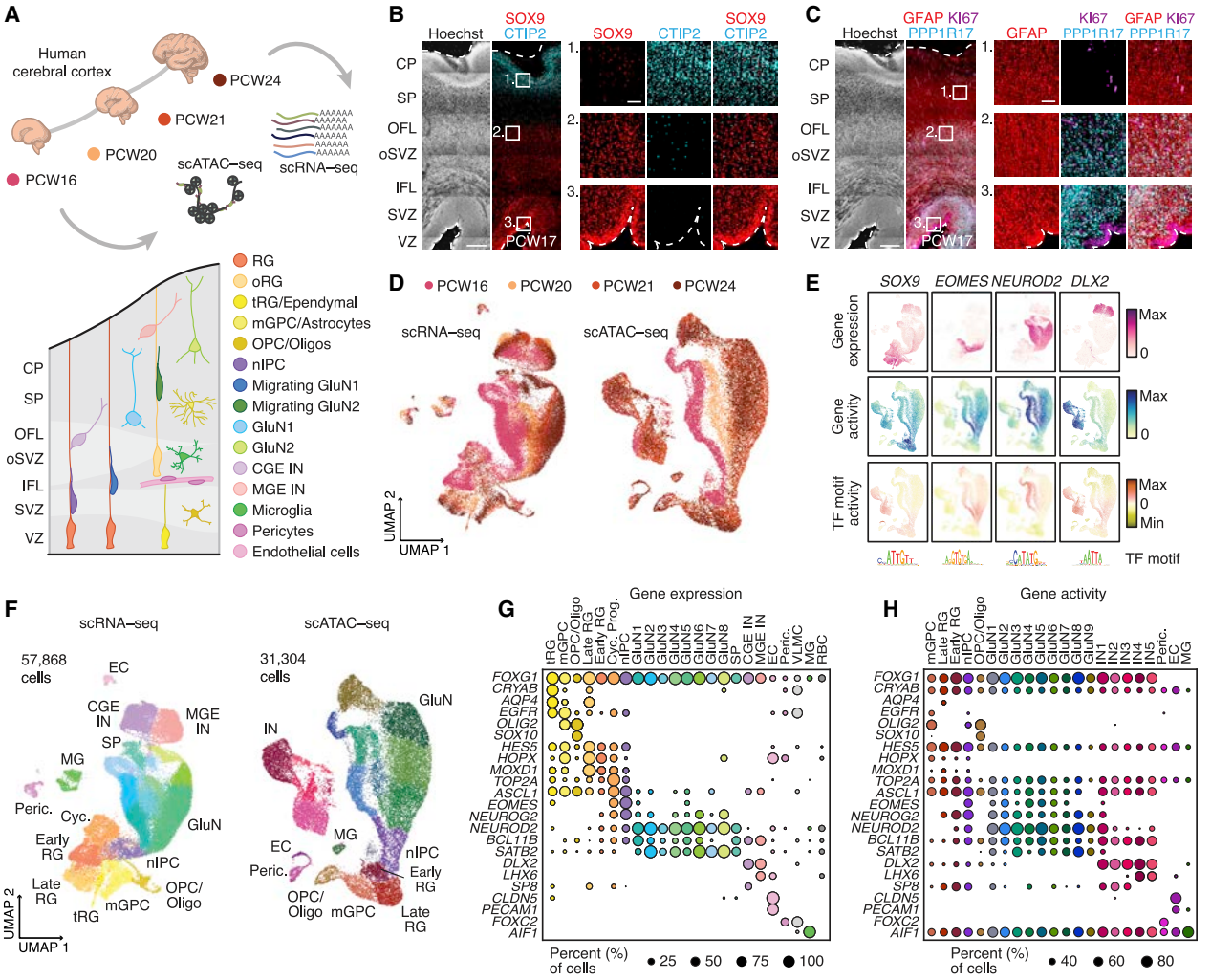


Figure 1. A single-cell epigenomic atlas of the human cerebral cortex

(A) Schematic of time, profiling methods, and cell types.

(B) Immunohistochemistry for SOX9 and CTIP2 at PCW17. VZ, ventricular zone; SVZ, subventricular zone; IFL, inner fiber layer; oSVZ, outer SVZ; OFL, outer fiber layer; SP, subplate; CP, cortical plate. This image was generated by automatic stitching of individual images.

(C) Immunohistochemistry for GFAP, KI67, and PPP1R17 at PCW17. This image was generated by automatic stitching of individual images.

(D) UMAP based on gene expression (left) and peak accessibility (right). Cells colored according to time.

(E) Multimodal profiling of *SOX9*, *EOMES*, *NEUROD2*, and *DLX2* including gene expression (scRNA-seq), gene activity scores, and TF motif activity (scATAC-seq).

(F) UMAP of cells colored by cluster. RG, radial glia; Cyc, cycling progenitors; tRG, truncated radial glia, mGPC, multipotent glial progenitor cell; OPC/Oligo, oligodendrocyte progenitor cell/oligodendrocyte; nIPC, neuronal intermediate progenitor cell; GluN, glutamatergic neuron; CGE IN, caudal ganglionic eminence interneuron; MGE IN, medial ganglionic eminence interneuron; EC, endothelial cell; MG, microglia; Peric., Pericytes

(G) Dotplot showing the cells expressing selected markers across scRNA-seq clusters.

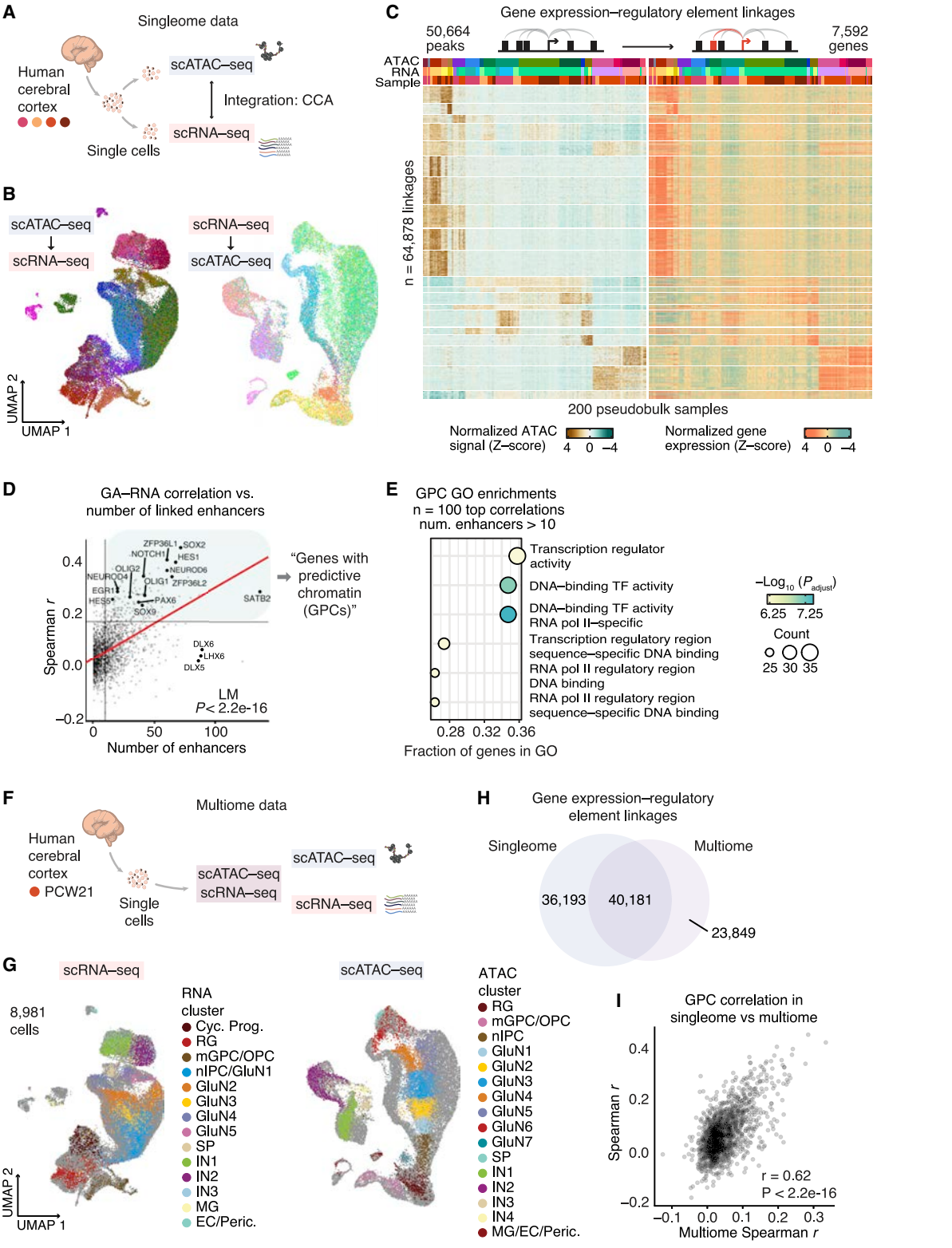
(H) Dotplot showing marker gene activity scores derived across scATAC-seq clusters.

Scale bars, 500 μm (B, C), 100 μm (insets B, C).

indices of correspondence and observed high agreement between cell types, cell-cycle phase, and gestational times in our data and the computationally matched independent annotation (Figures S3A–S3G).

We integrated the derived gene activity scores with gene expression levels using canonical correlation analysis (CCA) to match cell data from each modality to the nearest neighbors in the other data representation (Figure 2A) (Stuart et al., 2019). Cluster annotations of matched cells were consistent, except for the cycling progenitor cluster in scRNA-seq, which did not directly map to cells in the chromatin landscape (Figures 2B, S3H, and S3I). Using pseudo-bulk aggregates of these matched annotations, we applied a correlation-based approach that links gene-distal CRE accessibility to gene expression (Corces et al.,

Please cite this article in press as: Trevino et al., Chromatin and gene-regulatory dynamics of the developing human cerebral cortex at single-cell resolution, Cell (2021), https://doi.org/10.1016/j.cell.2021.07.039



(legend on next page)

Please cite this article in press as: Trevino et al., Chromatin and gene-regulatory dynamics of the developing human cerebral cortex at single-cell resolution, Cell (2021), https://doi.org/10.1016/j.cell.2021.07.039

2018; Ma et al., 2020; Trevino et al., 2020), identifying 64,878 CRE-gene pairs that represent potential enhancer-gene interactions (Table S2). In this analysis, a gene was linked to a median of five CREs, and linked CREs were more conserved than unlinked elements (Figure S3J, Wilcoxon rank-sum test $p < 2.2 \times 10^{-16}$) and more likely to be supported by cell-type-specific three-dimensional (3D) interactions from a recently published promoter-centric chromosome conformation capture dataset (Song et al., 2020) (Figures S3K–S3M). Co-variation of CRE accessibility and gene expression distinguished the identified cell types in both scRNA-seq and scATAC-seq (Figure 2C). Clustering the associated CRE accessibility revealed particularly high variability across clusters corresponding to glial cell populations, corroborated the distinctiveness of IN clusters, and indicated dynamic patterns of gene regulation across GluN clusters.

We then identified genes whose expression could be well predicted from local single chromatin accessibility by ranking gene activity-expression correlations. Genes with the highest correlation included *SOX2* and *HES1*, and these genes were linked to greater numbers of putative enhancers. We hypothesized that these comprised a class of highly regulated genes that play a driving role in establishing cell identities in the developing cortex and defined a set of 185 genes with predictive chromatin (GPCs; genes in the top decile of gene activity-expression correlations, linked to >10 CREs) (Table S2; Figure 2D). This gene set was strongly enriched for transcription regulator activity and DNA-binding TF activity (Figure 2E).

To validate these inferences, we profiled scATAC-seq and scRNA-seq data from the same cells in PCW21 human cortex (multiome) (Figure 2F). Filtering across both data modalities resulted in 8,981 cells with high-quality transcriptome and epigenome profiles (Table S2; Figures S3N–S3T). We projected multiomic scATAC-seq and scRNA-seq profiles into the corresponding individually generated landscapes and confirmed that our cell-type annotations were well represented in the joint data (Figure 2G). Applying our CRE-gene linking approach to the true cell-to-cell matches, we found that 40,181 inferred peak-gene linkages (53%) were observed from this single time point measurement and an additional 23,849 were identified (Figure 2H; Table S2), demonstrating that most inferred CRE-gene interactions were observed in this joint dataset. Similarly, we applied CCA to multiome data, where the correct cell assignments are known. The inferences were generally validated by the true clusters, and this agreement was increased by assigning clusters based on 50 nearest neighbors in CCA space, rather than the single closest neighbor (Figures S4A and S4B). In addition, we found a strong concordance in GPC activity-expression correlations of *in silico*-linked singleome cells versus multiome cells (Figure 2I). GPCs are thus also readily apparent in this joint dataset, underlining the correspondence between their local accessibility and their transcription within the same cell.

Continuous trajectories of gene regulation across cortical neuron differentiation

GluN are born in a specific sequence during development. Although several key factors controlling their fate have been described (Greig et al., 2013), the logic that governs their specification, migration, and maturation has not been resolved in human development. To define a trajectory of GluN development, we annotated each cell in associated clusters with pseudotime values. This annotation was derived using an algorithm based on diffusion through cell-similarity networks derived from RNA velocity (Bergen et al., 2020; La Manno et al., 2018) (Figures 3A and S4C–S4F). Notably, the algorithm rooted the trajectory in the cluster of cycling cells (Figure S4D). To test how the architecture of the cortex mapped onto this trajectory, we next projected an independent scRNA-seq data comprising adult cortical neurons (Hodge et al., 2019) into the developmental landscape and identified the nearest neighbor cell for each adult scRNA-seq profile (Figure S4G). Adult GluN projected preferentially into the neighborhoods of cells annotated with later pseudotimes, and pseudotime was significantly associated with the annotated layer of adult cells (one-sided Wilcoxon rank-sum test $p < 9.6 \times 10^{-15}$; Figure S4H). As expected, we also observed significant association of earlier and later time points with deep and superficial adult cortical layers (one-sided Fisher's exact test $p < 2.7 \times 10^{-124}$; Figure S4I). When we compared the expression levels in migrating neurons from the early gestational time point (PCW16) to those from later time points (PCW20–PCW24), we observed increased expression of *LIMCH1*, *RUNX1*, *SNCB*, and *DOK5* and decreased expression of the AP-1 TF family (*JUN*, *FOS*), heat shock factors *HSPA1A/B* and *DUSP1* (Figures S4J and S4K; Table S3). Overall, we found surprisingly few differentially expressed genes previously implicated in neurogenesis, suggesting a considerable degree of gene expression and regulatory variability could be associated with pseudotime rather

Figure 2. Integrative and multiomic gene regulatory dynamics in the human cortex

(A) Generation and integration of singleome scATAC-seq and scRNA-seq data. CCA, canonical correlation analysis.

(B) UMAPs of scRNA-seq and scATAC-seq cells colored by cluster assignment of matched cells.

(C) Heatmap showing chromatin accessibility and gene expression of 64,878 significantly linked CRE-gene pairs. Shown are side-by-side heatmaps in which one row represents a pair of one CRE and one linked gene). Each CRE can be linked to multiple genes, and each gene can be linked to multiple CREs. Hence, each gene and each CRE can be represented by multiple rows in the corresponding heatmap. Pairs (rows) were clustered using k-means clustering (k = 20). For visualization, 10,000 rows were randomly sampled. Columns represent 200 pseudobulk samples, which have been annotated using the majority RNA cluster, ATAC cluster, and time of all cells in the pseudobulk.

(D) Correlation between single-cell gene expression and chromatin-derived gene activity scores (GA), and the number of linked CREs per gene. TFs are labeled.

(E) GO enrichment analysis of the 185 genes with GPCs in (D).

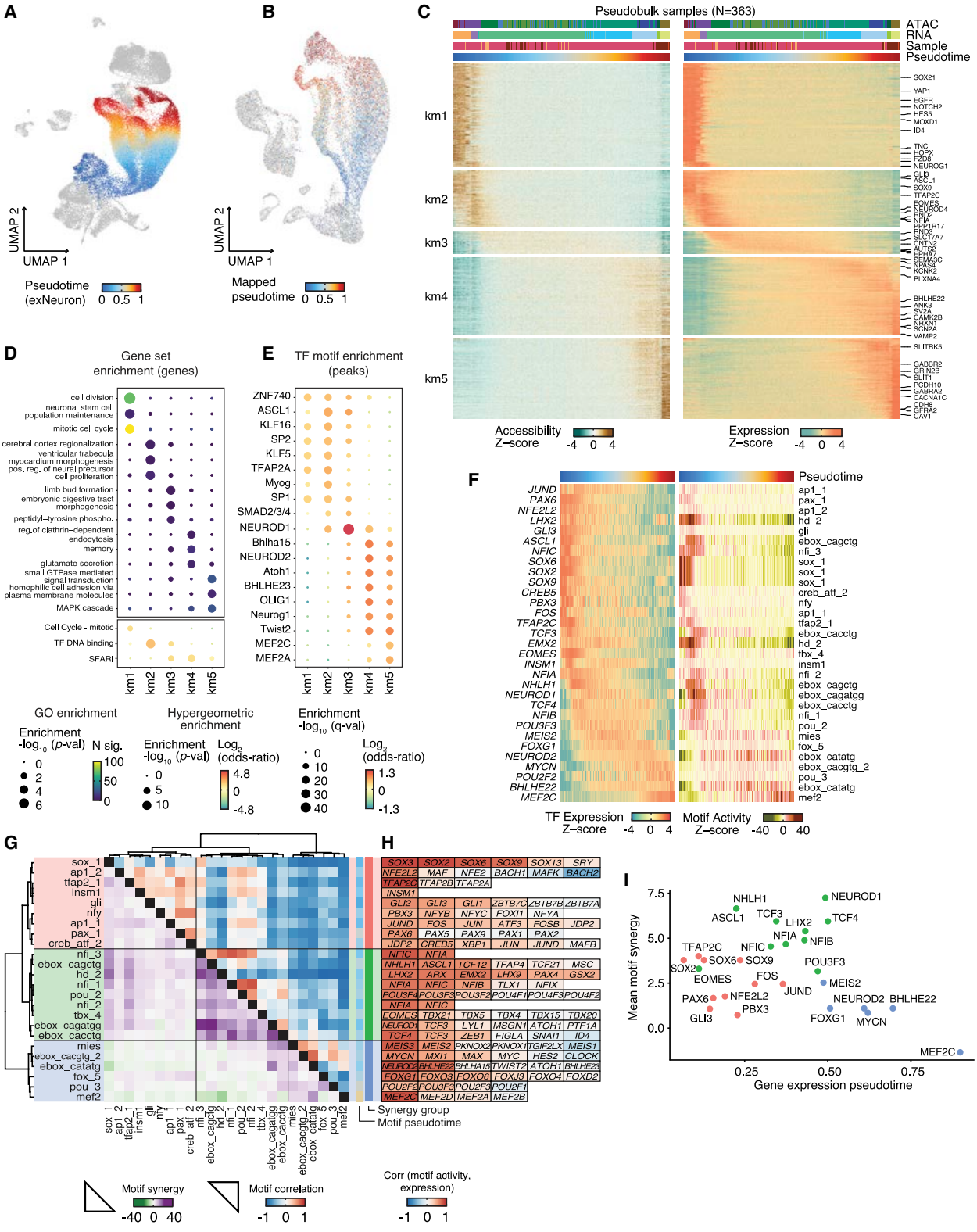
(F) Generation of scATAC-seq and scRNA-seq data from the same cells (multiome data).

(G) Projection of multiome scATAC-seq into singleome scATAC-seq UMAP space, and multiome scRNA-seq into singleome scRNA-seq UMAP space. Cell coloring corresponds to multiome cluster assignment, and gray cells show the singleome manifold onto which multiome cells have been projected.

(H) Venn diagram showing overlap of CRE-gene linkages identified in singleome versus multiome data.

(I) Correlation showing the correspondence between predictive chromatin in singleome versus multiome data.

Please cite this article in press as: Trevino et al., Chromatin and gene-regulatory dynamics of the developing human cerebral cortex at single-cell resolution, Cell (2021), https://doi.org/10.1016/j.cell.2021.07.039



(legend on next page)

Please cite this article in press as: Trevino et al., Chromatin and gene-regulatory dynamics of the developing human cerebral cortex at single-cell resolution, Cell (2021), https://doi.org/10.1016/j.cell.2021.07.039

than gestational time. We therefore decided to investigate the regulatory dynamics along the pseudotime axis.

To connect expression trajectories to the accessibility dynamics of regulatory elements, we transferred pseudotime values from RNA cells to their nearest ATAC cell neighbors, confirming this produced a smooth continuum of pseudotime in the chromatin manifold (Figure 3B). By applying our correlation-based CRE-to-gene linking approach to the glutamatergic neuronal lineage, we identified 13,989 dynamic interactions across pseudotime and grouped these into five clusters (Figure 3C; Table S3). Linked genes active early in pseudotime exhibited gene ontology (GO) enrichments for cell division and neural precursor proliferation, whereas later interactions were associated with morphogenesis, migration, and maturation (Figure 3D). Interestingly, genes encoding TFs and DNA-binding proteins were particularly enriched in intermediate interactions, while genes implicated in ASD susceptibility (Abrahams et al., 2013) were more likely to be linked later in pseudotime. To nominate TFs that may control these programs, we identified motifs that were enriched in the different clusters of linked regulatory elements. Motifs enriched in interactions early in the trajectory included ZNF740, KLF16, SP1/2, and ASCL1 (Figure 3E). Conversely, interaction clusters associated with intermediate and late pseudotime were associated with motifs of neuronal TFs (NEUROD1/2, NEUROG1, MEF2C).

We next characterized the TF-driven regulatory dynamics of neurogenesis over pseudotime. To mitigate correlation biases due to sequence similarity between motifs in this analysis, we utilized a resource of previously disambiguated clusters of TF motifs (Vierstra et al., 2020). We then linked specific TF genes to these motif clusters by correlating TF expression with the accessibility-derived motif activity scores, resulting in pairings of 31 TFs and 24 motif clusters (STAR Methods). We observed synchronized TF expression and motif activity for dynamic regulators along developmental pseudotime, starting with PAX6, SOX2/6/9, GLI3, and ASCL1 motifs, followed by intermediate stage factor motifs (EOMES, NFIA, NFIB, NEUROD1), and finally late-stage motifs (NEUROD2, BHLHE22, MEF2C; Figure 3F). Together, these data describe cohesive, sequential waves of motif activation during corticogenesis that are consistent across gestational time points.

To understand how TFs are coordinated during corticogenesis, we computed the genome-wide synergy and correlation patterns of motif family accessibility (Figures 3G and 3H; STAR Methods) (Schep et al., 2017). We found three broad classes of motifs that associated with accessibility and TF expression over pseudotime (Figures 3G–3I): (1) early-activity motifs exhibiting moderate synergies (SOX, GLI, PAX), (2) intermediate activity motifs (NFI, TBX/EOMES) that are highly synergetic within their class, and (3) late-activity motifs that are less cooperative and generally appear to operate more independently (NEUROD2/BHLHE22, MEF2). These findings suggest a higher degree of TF motif coordination early in neurogenesis and regulation of maturation by a smaller set of more independent TFs.

Clustering approach to link gene expression programs to cell-fate decisions

We observed extensive heterogeneity in glial populations, corresponding to distinct yet partially overlapping expression programs in the identified clusters (Figures S5A and S5B). We adopted an analysis to identify modules of co-expressed genes using fuzzy c-means clustering (STAR Methods; Figures 4A and S5C; Table S4), allowing cells to be annotated with module activities, and for genes to be shared between multiple modules (Figures S5C and S5D; Table S4), enabling analysis of how cells may progress from one module to another. We projected these cell loadings into a low-dimensional representation of differentiation (Figure 4A, bottom). The structure of this embedding and the underlying module assignments was stable to fuzzy clustering parameters (STAR Methods).

To understand the biological basis of these gene modules, we examined their expression across cell clusters, developmental stage, and pseudotime (Figure 4B), which was rooted in cycling (“Cyc”) cells and correlated with time (Figure S5E; STAR Methods). Glial maturation genes *FOXJ1*, *AQP4*, and *MBP*, which are markers for ciliated ependymal cells, astroglia, and oligodendrocytes, respectively (Barbarese et al., 1988; Jacquet et al., 2009; Zhang et al., 2016), were expressed in late-pseudotime cells and assigned to modules m5, m2, and m7. In contrast, the expression of genes associated with cell division and progenitors, such as *TOP2A*, *NR2F1*, and *NFIC*, peaked early in pseudotime and were assigned to modules m10, m6, and m3 (Figures 4C and S5F). Some modules (m6, m8) spanned many samples and stages, indicative of sustained expression programs, while others were restricted (m5, m14; Figures 4B, 4D, and S5G). Modules exhibited distinct GO enrichments, including “cation and metal ion binding” in m6, which may be related to the role of human astrocytes in metal homeostasis (Vasile et al., 2017; Zhang et al., 2016) and disease (Figures S5H and S5I). Module m5, comprising *FOXJ1*$^{+}$ cells, was enriched for dynein

Figure 3. Molecular signatures of cortical glutamatergic neurons

(A) UMAP of scRNA-seq cells highlighting the GluN trajectory and RNA-velocity derived pseudotime.

(B) UMAP of scATAC-seq with transferred pseudotime annotation.

(C) Heatmap showing chromatin accessibility and gene expression of 13,989 significantly linked CRE-gene pairs (columns: left, CRE accessibility; right, linked gene expression) across 363 pseudobulk samples aggregated along pseudotime bins in the GluN trajectory. Interactions (rows) were clustered using k-means clustering (k = 5).

(D) Gene set enrichment analysis of genes represented in the five interaction clusters.

(E) Enrichment of TF motifs in peaks represented in the five interaction clusters. Color represents odds ratios; size represents the $-\log_{10}$ (p value).

(F) Heatmaps showing *Z* score normalized expression (left) and motif activity (right) of TFs in 363 pseudobulk samples aggregated along pseudotime bins. Rows show 31 dynamic TFs associated with 24 motif clusters.

(G) TF motif correlation coefficients (upper) and synergy *Z* scores (lower) of the 24 motif clusters in (F).

(H) Correlation coefficients of TF motif cluster chromatin activity and annotated gene expression.

(I) Scatterplot showing aggregate gene expression pseudotime versus mean motif synergy. Point colors denote the cluster assignments in (G).

Please cite this article in press as: Trevino et al., Chromatin and gene-regulatory dynamics of the developing human cerebral cortex at single-cell resolution, Cell (2021), https://doi.org/10.1016/j.cell.2021.07.039

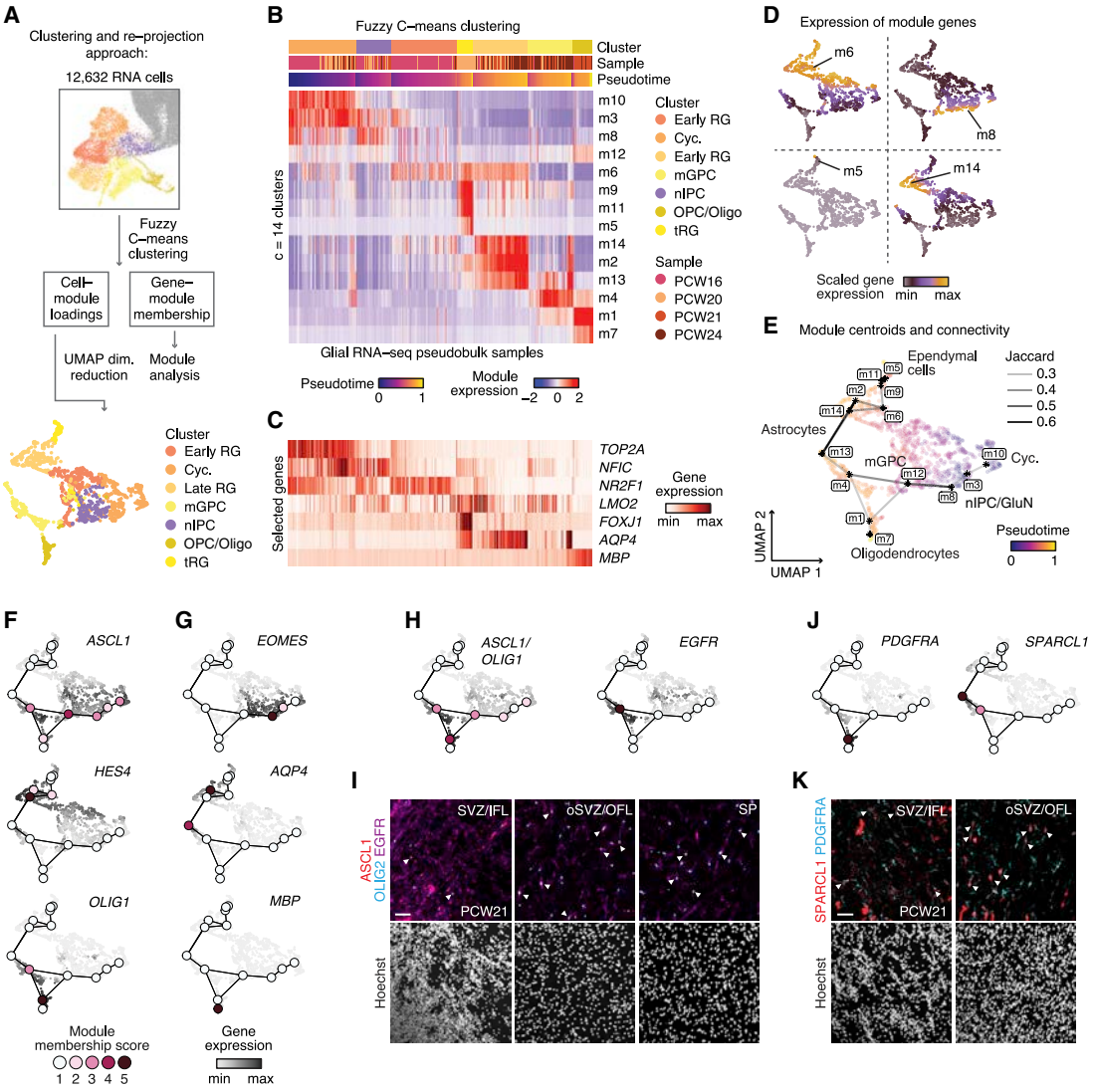


Figure 4. Regulatory logic of glial cell specification

(A) Schematic illustrating the expression-based clustering and reprojection of glial cells. Points in the bottom panel correspond to pseudobulk aggregates of 50 cells.

(B) Heatmap of module expression across pseudobulk aggregates, showing variation by cluster, sample age, and pseudotime.

(C) Heatmap showing the expression of selected genes across the same pseudobulks.

(D) Mean scaled expression in the low-dimensional UMAP embedding of selected gene modules. Figure S5G shows all modules.

(E) Projection of module centroids into UMAP space. Pseudobulk samples are colored by pseudotime. Module overlap is shown by links between centroids and was computed by thresholding the pairwise Jaccard index at >0.2.

(F) Module membership and expression values for *ASCL1*, *HES4*, and *OLIG1*.

(G) Module membership and expression values for *EOMES*, *AQP4*, and *MBP*.

(H) Module membership and expression values for *ASCL1/OLIG1* and *EGFR*.

(I) Immunohistochemistry showing expression and colocalization (white arrowheads) of ASCL1, OLIG2, and EGFR in cells of the SVZ, oSVZ, outer and inner fiber layers (OFL, IFL), and SP.

(J) Module membership and expression values for the oligodendrocyte progenitor marker *PDGFRA* and the astrocyte-associated gene *SPARCL1*.

(K) Immunohistochemistry showing expression and colocalization (white arrowheads) of SPARCL1 and PDGFRA in cells of the SVZ, oSVZ, and outer and inner fiber layers (OFL, IFL).

Scale bars, 50 μm (J and K).

Please cite this article in press as: Trevino et al., Chromatin and gene-regulatory dynamics of the developing human cerebral cortex at single-cell resolution, Cell (2021), https://doi.org/10.1016/j.cell.2021.07.039

binding and microtubule activity, consistent with the role in circulating cerebrospinal fluid (Ransom, 2012). Immunohistochemistry revealed that TFAP2C, which associated with module m6, was expressed in VZ and SVZ (Figures S6A and S6B). Similarly, PBXIP1, which was associated with m2, was expressed in RG in the VZ and SVZ but not in more mature CP astrocytes (Figures S6C and S6D). CRYAB, associated with m9, was expressed in tRG in the VZ, as described (Figures S6E and S6F) (Nowakowski et al., 2016).

Our clustering and reprojection approach enabled us to compute the degree of gene overlap between modules, which provided a measure of module similarity across our glial landscape (Figure S6G). To visualize these relationships, we computed the weighted average of module gene expression across pseudobulk aggregates and plotted these "module centroids" and their connectivity in the embedding, along with pseudobulks and their pseudotime values (Figure 4E). Investigation of module memberships in this representation revealed three broad programs emanating from the cycling cluster: (1) an *ASCL1*$^+$ program associated with m3 and m8 and terminating in *EOMES*$^+$ nIPCs; (2) a *HES4*$^+$ program associated with module m6 and terminating in astrocytes and ependymal cells; and (3) an *ASCL1*$^+$/*OLIG1*$^+$ program associated with m12, m1, and m4, branching into two endpoints (Figures 4F and 4G). The *ASCL1*$^+$/*OLIG1*$^+$ program was of particular interest, as it corresponded to the mGPC cluster of cells, which expressed markers associated with both astroglia (*GFAP*, *HOPX*, *EGFR*, *ASCL1*) and oligodendrocyte progenitors (*OLIG2*, *PDGFRA*), suggesting a common multipotent glial progenitor (Figures 4H and 4J). Immunohistochemistry revealed that ASCL1, OLIG2, and EGFR were often colocalized in the SVZ/IFL, oSVZ/OFL, and SP (Figures 4I, S7A, and S7B). If generated from a common glial progenitor, astrocyte and oligodendrocyte precursors might also share expression of markers associated with more differentiated states. We found that PDGFRA and OLIG2, markers associated with oligodendrocyte progenitors, and SPARCL1, which is a marker associated with mature astrocyte identity (Zhang et al., 2016), colocalized in the SVZ/IFL and oSVZ/OFL (Figures 4K and S7C–S7F). We speculate that a common multipotent glial progenitor, competent to differentiate into both astrocytes and oligodendrocytes, could explain this substantial overlap of expression programs.

Chromatin and gene expression profiles identify two astrocyte precursor populations

Human cortical astrocytes are larger, more morphologically complex (Oberheim et al., 2009; Zhang et al., 2016), and likely more diverse than those of other mammals (Vasile et al., 2017). However, the steps underlying the diversification of human astrocytes are unknown. We observed three interconnected fuzzy gene modules, largely derived from PCW24 tissue, expressing *AQP4*, *TNC*, *ALDH2*, and *APOE*, and other genes specifically expressed in astrocytes (m2, m13, m14) (Sloan et al., 2017; Wiese et al., 2012; Zhang et al., 2016) (Figures 5A, S8A, and S8B). To test whether these transcriptionally related yet distinct subpopulations associated with different regulatory factors, we computed differential motif enrichments between enhancers linked to genes in m13 versus m14. We found that the basic helix-loop-helix (bHLH) factor motifs ASCL1 and NHLH1 were enriched in module m13, while SOX21 was enriched in m14 (Figure 5B). In our glial cells, the accessibility of ASCL1 and NHLH1 motifs correlated best with the gene expression of bHLH factor *OLIG1* (Spearman rho = 0.34 and 0.36, respectively), and we have previously nominated SOX21 as a potential regulator of astrocyte maturation in cortical organoids (Trevino et al., 2020). Thus, two distinct astrocyte-like expression patterns could be distinguished by chromatin accessibility of *OLIG1* versus *SOX21* motifs.

To examine the differences between cells expressing these modules in more detail, we computed differential gene expression between the astrocytic cell clusters A1-HES and A2-OLIG, corresponding to expression of modules m2/14 and m13, respectively (Figures 5C and 5D; Table S5). Cluster A1-HES exhibited significantly higher expression of *HES4* and *CAV2*, while A2-OLIG was characterized by increased *SPARCL1*, *ID3*, and *IGFBP7* expression (Figures 5D and S8C). To determine whether these distinct astrocyte precursor subtypes were due to the sampling of different cortical areas, we used a recently published scRNA-seq dataset (Bhaduri et al., 2020) (Figures 5E and S8D). We found that gene sets attributed to our astrocytic classes were expressed in distinct cell populations in this independent dataset—an observation that could not be explained by differences in cortical area (Figure 5F). These developmental states may correspond to adult subtypes, such as protoplasmic astrocytes, found throughout the gray matter of the cortex, fibrous astrocytes found in the white matter, or primate-specific interlaminar astrocytes, which populate layer 1 (Hodge et al., 2019; Oberheim et al., 2009; Vasile et al., 2017).

Chromatin state links GPCs to lineage determination in cycling cells

We next examined how the chromatin state of progenitor cells could potentially affect the acquisition of expression programs characteristic of more differentiated cell states. We therefore focused on the heterogeneity among cells that expressed gene modules strongly associated with cell-cycle signatures (Figure 6A; Pearson r = 0.89, 0.91, respectively). To link chromatin accessibility to the glial-centric expression map, we projected pseudobulk aggregates of 13,378 glial scATAC-seq cells into our gene-module-derived manifold using accessibility-derived gene activity scores. Consistent with our CCA cluster matching analysis (Figures 2B, S3H, and S3I), pseudobulks comprised mainly of cells from ATAC cluster c15 (OPC/Oligo) projected into the oligodendrocyte endpoint of this map, cluster c10 (mGPC) data projected into the ASCL1$^+$/OLIG2$^+$ astrocyte compartment, and cluster c9 (late RG) data projected into both ependymal and HES4$^+$ astrocyte endpoints (Figure 6B). However, while we did not observe a distinct cycling cluster in our chromatin landscape, a subset of these ATAC-seq pseudobulk samples projected into the cycling, early-pseudotime compartment of the RNA-seq embedding. These samples partitioned into three distinct branches defined by their scATAC-seq cluster assignments (Figure 6C). We speculate that strong cell-cycle signatures in RNA-seq may have diminished these distinctions that are more evident in ATAC-seq data and that analyzing these

Please cite this article in press as: Trevino et al., Chromatin and gene-regulatory dynamics of the developing human cerebral cortex at single-cell resolution, Cell (2021), https://doi.org/10.1016/j.cell.2021.07.039

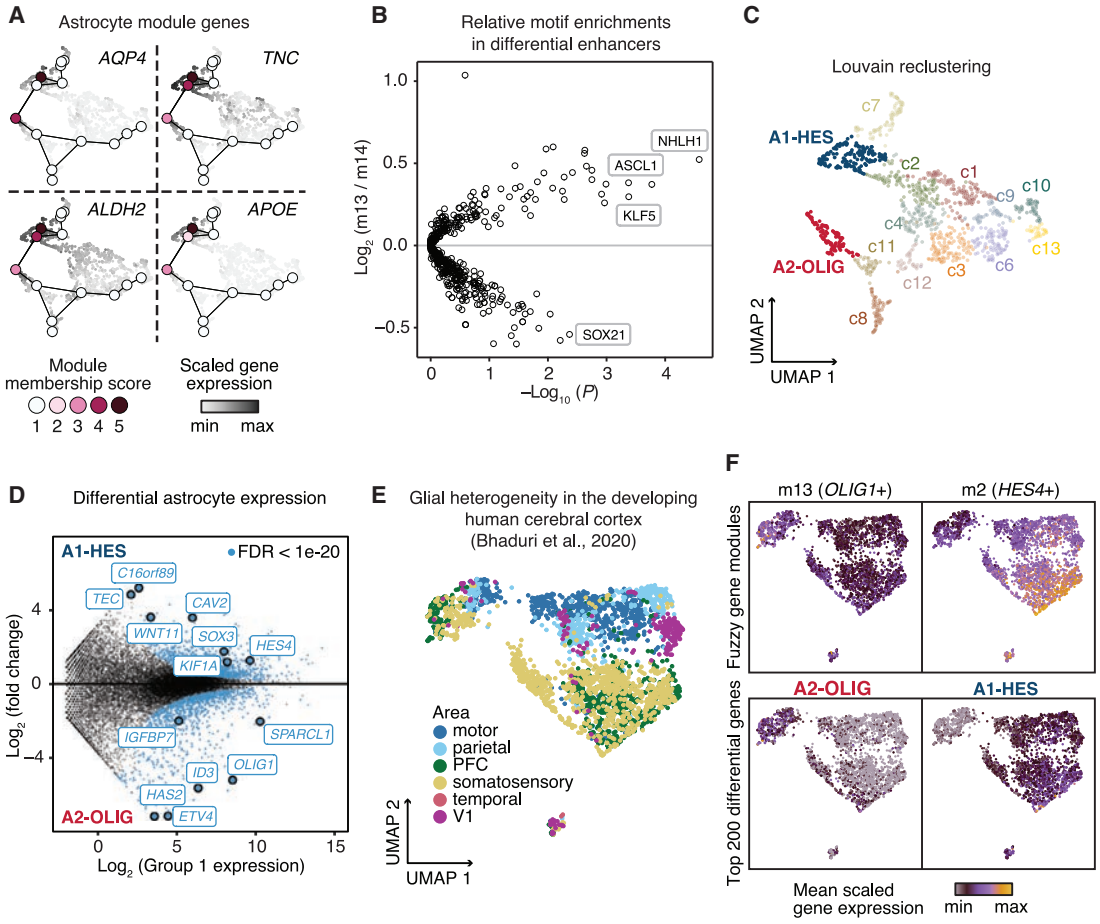


Figure 5. Astrocyte precursor heterogeneity

(A) Module membership and scaled gene expression of astrocyte-associated genes *AQP4*, *TNC*, *ALDH2*, and *APOE*

(B) Motif enrichments in GREs linked to module 13 genes relative to GREs linked to module 14.

(C) Reclustering of glial pseudobulk samples in fuzzy clustering embedding. *AQP4* positive clusters are highlighted and defined as A1-HES and A2-OLIG.

(D) Differential gene expression between A1-HES and A2-OLIG clusters, calculated using DESeq2. A threshold of Benjamini-Hochberg corrected FDR of 1e−20 was used for visualization (blue)

(E) UMAP of astrocytes from a human fetal scRNA-seq dataset (Bhaduri et al., 2020), colored by cortical area.

(F) Mean scaled expression of genes in modules m13 and m2 (top) and the top 200 differential genes from D (bottom) in Bhaduri et al. (2020).

separate branches might allow us to determine whether cycling progenitors are poised toward distinct postmitotic fates.

To explore factors that influence these fate decisions, we identified genes specific to each branch based on their gene activity scores (STAR Methods). We observed a strong overlap of these genes with the set of GPCs, including *HES1*, *RFX4*, *OLIG1*, *OLIG2*, *NEUROD6*, and *EOMES*. Overall, differential chromatin activity in all three branches of cycling cells was enriched for GPCs (Figure 6D). Each branch was enriched for at least one bHLH GPC TF in the top five most unique genes (*BHLHE40*, *OLIG1*, *OLIG2*, *NEUROD6*, *NEUROD4*) (Figure 6E). The similarity of annotated motifs for these factors is consistent with the hypothesis that they can compete for similar binding sites to drive multiple distinct cell fates (Imayoshi et al., 2013; Zhou and Anderson, 2002). Together, these results suggest that differential chromatin activity as well as gene expression of GPCs are prominent features that distinguish different types of cycling glial progenitor cells.

We next wondered whether these GPCs were both highly connected to dense collections of regulatory elements and highly enriched for lineage-defining transcription factors. To evaluate whether these links could be indicators of the eventual differentiation endpoint, and thus potentially drive differentiation, we re-projected ATAC-seq pseudobulk samples from A, B, and C cycling population branches by only using GPC-associated chromatin signals. We observed that samples moved forward in pseudotime to regions with distinct, more mature expression states (Figure 6F), whereas reprojections using random gene subsets or modules of genes moved non-specifically toward the center of the manifold (Figure S8E). This observation suggests that chromatin patterns linked to GPC genes in these cycling cells already exhibit a signature

Please cite this article in press as: Trevino et al., Chromatin and gene-regulatory dynamics of the developing human cerebral cortex at single-cell resolution, Cell (2021), https://doi.org/10.1016/j.cell.2021.07.039

CellPress

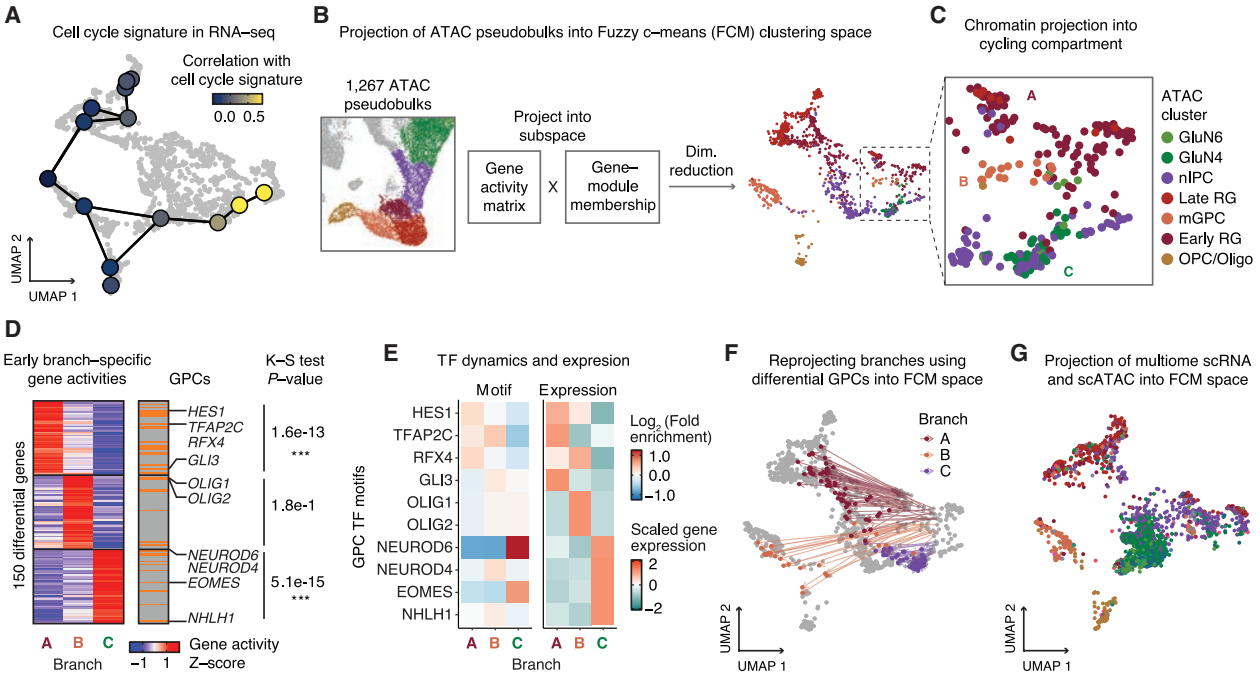


Figure 6. Chromatin state links GPCs to cell fates

(A) Pearson correlation of a cell-cycle signature (MSigDB) with module expression signature across pseudobulks.

(B) Schematic of ATAC-seq projection into fuzzy clustering embedding.

(C) Projection of ATAC-seq pseudobulks (each comprising 50 cells) into Cyc cluster and in the neighborhood of cycling-associated modules.

(D) Heatmap showing the 50 most uniquely active genes in branches A, B, and C. Gene activities are row scaled. Orange bars denote GPC labeling. The p value of a Kolmogorov-Smirnov test for enrichment of GPCs in differential, branch-specific genes are shown.

(E) Dynamics of GPC motifs and gene expression across three branches of cycling cells. Heatmaps represent enrichment of GPC TF motifs (left) and gene expression levels (right) in branch aggregates.

(F) Reprojection of branches A, B, and C using only chromatin accessibility associated with GPCs.

(G) Projection of multiome scRNA-seq data into fuzzy clustering embedding. Cells (points) are colored by their mapped scATAC-seq cluster.

of an advanced transcriptional cell state. Similarly, when we projected the scRNA-seq data from the joint multiome dataset into the module-based manifold, a fraction of cells projecting to the cycling domain exhibited distinct accessibility signatures of more differentiated cells from each branch (Figure 6G). Based on these results, we propose that during corticogenesis, progenitors entering the cell cycle may be epigenetically primed toward future cell fates and that this information is encoded specifically in GPCs, a set of genes with large numbers of linked enhancers enriched for binding of lineage-defining TFs.

Deep-learning models prioritize disruptive noncoding mutations in ASD

We next used our atlas to interpret noncoding *de novo* mutations in ASD, using the Simons Simplex Collection catalog of over 200,000 such mutations in 1,902 families (An et al., 2018) (Table S6). Naive overlap of mutations with cluster-specific scATAC-seq peaks produced no enrichment for mutations in ASD individuals relative to those in unaffected siblings (odds ratio [OR] = 1.02 for GluN6 cluster, Fisher's exact test p = 1.0; Figure S8F), indicating that peak-level annotations alone are insufficient to resolve a sparse set of causal mutations.

Deep-learning models have proven useful for prioritizing disease-relevant noncoding genetic variants based on their predicted regulatory impact (Kelley et al., 2016, 2018; Zhou and Troyanskaya, 2015). We therefore trained convolutional neural networks, based on the recent BPNet architecture, to learn models that could predict base-resolution, pseudo-bulk chromatin accessibility profiles for each of our scATAC-seq-derived cell types from genomic sequence (Figure 7A; STAR Methods) (Avsec et al., 2020), using peak regions and genomic background, matched for GC content and motif density to correct for potential sequence composition biases (Figure S8G). The models showed high and stable correlation between total predicted and observed Tn5 insertion count coverage across all peak regions in held-out chromosomes across 5-folds of cross-validated models (e.g., GluN6, mean Spearman rho = 0.58; Figure S8H; Table S6). To predict cell-context-specific effects of a candidate mutation on chromatin accessibility, we used our cluster-specific BPNet models to compute local disruption score based on the allelic fold-change in predicted counts. For each cluster, we computed the enrichment of high-effect-size mutations in cases versus controls. We observed significant enrichment of ASD-related mutations for GluN2/3/4/6/9 (>1.2-fold), which is in line with previous studies (Gandal et al.,

Please cite this article in press as: Trevino et al., Chromatin and gene-regulatory dynamics of the developing human cerebral cortex at single-cell resolution, Cell (2021), https://doi.org/10.1016/j.cell.2021.07.039

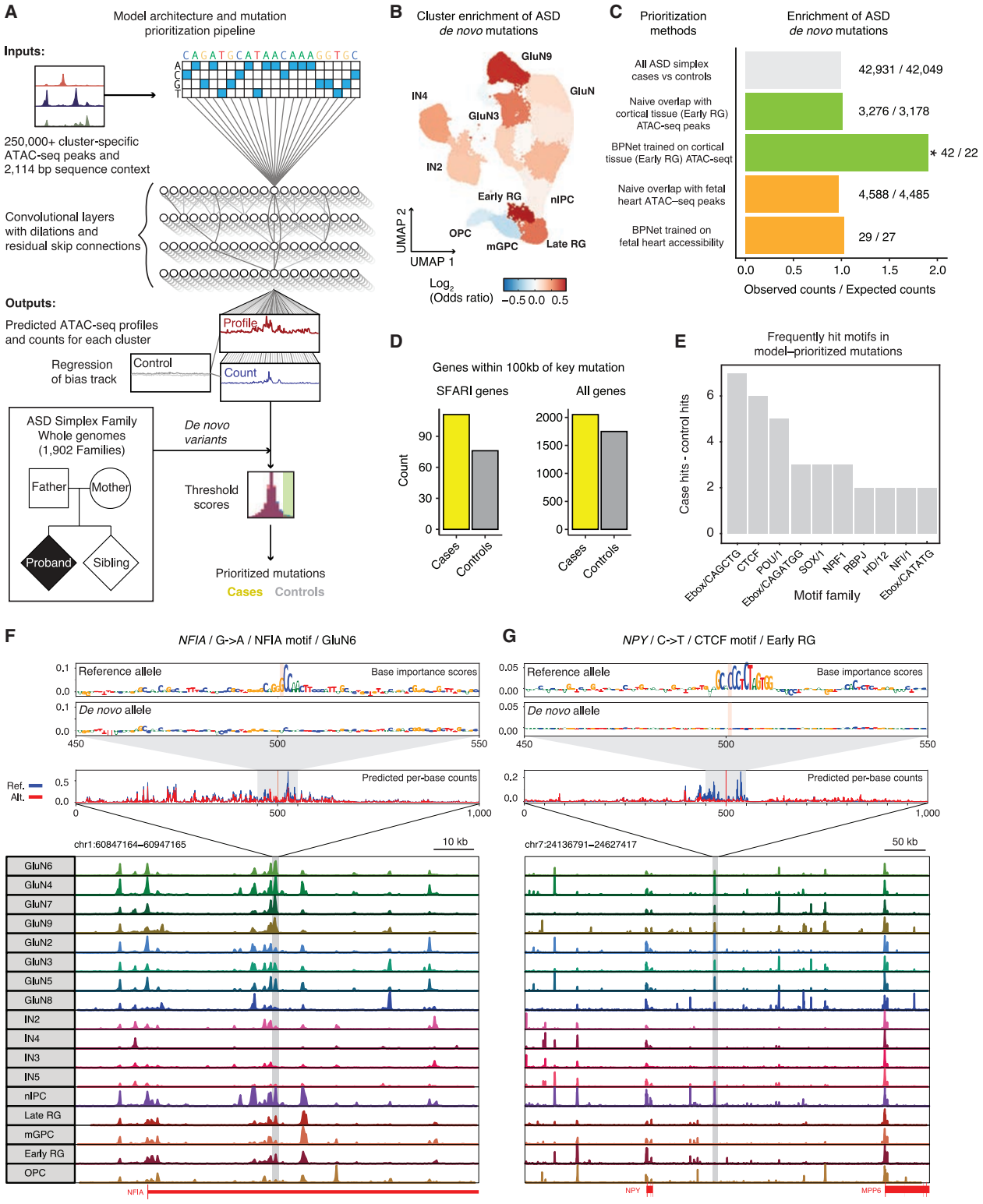


Figure 7. Disease association of gene regulatory elements

(A) Schematic of mutation prioritization pipeline.

(B) Cluster-specific BPNet enrichments visualized in scATAC UMAP.

(legend continued on next page)

Please cite this article in press as: Trevino et al., Chromatin and gene-regulatory dynamics of the developing human cerebral cortex at single-cell resolution, Cell (2021), https://doi.org/10.1016/j.cell.2021.07.039

2018; Li et al., 2018a; Parikshak et al., 2013; Trevino et al., 2020; Willsey et al., 2013). In addition, we found a strong association with IN2/3/4, nIPC, late RG, and early RG clusters. The early RG cluster showed the highest enrichment (OR = 1.909, excess of 20, Fisher's exact $p < 0.05$; Figure 7B; Table S6). We also observed this approach of prioritizing causal disruptive mutations was robust to threshold parameter selection (Figures S8I and S8J). In contrast, BPNet models trained on human fetal heart enhancers produced no enrichment (OR = 1.01, p = 1.0). Likewise, naive overlap enrichment with a set of fetal heart enhancers also produced no enrichment for case mutations (OR = 0.97, p = 1.0; Figure 7C). Together, these results suggest that the mutation effect scores from base-pair-resolution predictive models trained on chromatin accessibility landscapes in disease-relevant cell states are critical for prioritizing putative causal noncoding mutations.

The case and control mutations prioritized by the BPNet models had similar conservation scores and similar distances to the nearest transcription start site (TSS) (Figures S8K and S8L), highlighting the challenge of identifying these causal mutations by other means. Annotating the predicted high-effect-size mutations with their nearest genes, we observed a 1.4-fold enrichment for case mutations (n = 24) whose nearest gene was in the SFARI database compared with the control mutations (n = 17; Figure 7D). Next, we identified TF motifs that overlapped and were predicted to be disrupted by all the high-effect-size mutations from the BPNet models from all positively enriched clusters (Figure 7E, Table S6). We found that CTCF, which demarcates topological loop boundaries, was one of the most frequently disrupted motifs in cases versus controls. The NRF1 motif was another frequently disrupted motif. NRF regulates the GABA receptor subunit *GABRB1*, previously associated with disease (Li et al., 2018b). Other frequently disrupted motif families in cases relative to controls included E-box/bHLH family motifs (ASCL1, NEUROD6) and homeobox family (PAX5) motifs, with more lineage-specific effects. Homeobox proteins were also previously found to be disrupted by variants in ASD (Amiri et al., 2018; Trevino et al., 2020).

One highly disruptive mutation in our models was located in an intron of *NFIA* (Figures 7F and S8M). Loss-of-function mutations in this gene have previously been implicated in ASD (Iossifov et al., 2014). The mutation was in a linked intronic enhancer for the *NFIA* target gene. We observed that this enhancer was specifically accessible in different types of GluN clusters. The BPNet model for GluN6 predicts the mutation disrupting an NFIA motif, suggesting this mutation may dysregulate the NFIA gene expression via auto-regulatory feedback.

In the nIPC cluster, the BPNet model predicted a disruptive *de novo* mutation in an intergenic enhancer linked to the neuropeptide Y gene (*NPY*) whose TSS was 90 kb away from the mutation (Figure 7G). *NPY* is expressed in the subplate (Miller et al., 2014) and in early RG in the mid-gestation human cortex (Figure S8N), and genomic deletions of the NPY receptors have been associated with ASD (Ramanathan et al., 2004). The model further predicted this *de novo* mutation to disrupt a CTCF binding site at a chromatin loop anchor, suggesting a potential mechanistic impact on the chromatin architecture of this locus.

DISCUSSION

Here, we generate paired transcriptome and epigenome atlases of corticogenesis during a critical period of cortical development and describe how molecular interactions between DNA binding factors and *cis*-regulatory elements regulate gene expression programs. Furthermore, we describe how rare noncoding, *de novo* mutations may act to disrupt this logic.

We identified a set of genes (GPCs), enriched for lineage-defining TFs, whose local chromatin accessibility was predictive of expression levels using signals derived from single cells, possibly because of the large number of expression-linked enhancers. These linkages are evocative of other terms that have been used for similar phenomena, including "super enhancers" (Parker et al., 2013; Whyte et al., 2013) and "super-interactive promoters" (Song et al., 2020). Furthermore, chromatin accessibility of GPCs was consistent with a more differentiated cell state in some cycling progenitors. Recently, Ma et al. reported a phenomenon by which accessibility at similarly defined domains of regulatory chromatin delineate potential future cell states (Ma et al., 2020). We speculate that the coordinated effect of many enhancers on lineage-defining factors makes the expression of those factors more resistant to perturbation. Highly cooperative regulation of lineage-determining *trans*-acting factors may be a general principle of fate determination, acting as a positive feedback mechanism once a key differentiation gene has been expressed. Effectively, once activated, these enhancers might act as a ratchet, ensuring stable gene expression and preventing backtracking along a differentiation landscape when facing extrinsic or intrinsic perturbations.

Examining the trajectories of GluN migration and maturation, we found a molecular program that was consistent across 8 weeks of gestation and was defined by a sequence of motifs.

(C) Bar plot showing the enrichment of cases versus controls using different prioritization methods. Colors represent the baseline of all cleaned SSC mutations (gray), our scATAC-seq dataset (green), and a set of fetal heart enhancers (orange). * indicates a Fisher's exact test OR = 1.909, p = 0.004.
(D) Bar plot showing the number of prioritized mutations whose nearest gene is a SFARI gene. Cases (24) versus controls (17) are compared to the total number of prioritized mutations in cases (262) versus controls (232). Fisher's exact test OR = 1.24, p = 0.154.
(E) Bar plot showing the motifs that were most frequently disrupted in case mutations relative to control mutations. The y axis denotes the excess of overlaps with motifs by prioritized mutations in cases minus controls. The plot does not represent a statistical test.
(F) Example showing a disruptive case mutation in an intron of *NFIA*. The consensus logos show the importance of residues to predicted accessibility at the mutation. A 100 bp window flanking the mutation is shown. The genome tracks indicate predicted per-base counts for ref (blue) and alt (red) alleles in a 1,000 bp window flanking the mutation. The gene model around the mutation is shown along with tracks indicating the aggregate accessibility of scATAC-seq clusters at the locus.
(G) Example showing a disruptive case mutation at the *NPY* locus, as above.

Differences in neuronal regulatory activity across pseudotime were more pronounced than differences between developmental stages. We further found distinct patterns of co-accessibility and regulatory interactions between TFs early in pseudotime, whereas late TFs appeared to act more independently.

We also observed substantial sharing of TF-regulated gene expression programs among glial cells, with substantial overlap between gene modules containing canonical markers for astrocytes and oligodendrocytes. We validated the co-expression of several of these genes in human cerebral cortex. We also provided evidence for the existence of two lineages of astrocyte-like glial precursors (Vasile et al., 2017). Although glial modules were broadly interconnected, we found that the chromatin activity of GPCs in cycling cells was predictive of specific differentiated states, suggesting that progenitors entering the cell cycle are primed toward specific lineages.

Finally, our interpretable, cell-type-specific deep-learning models that link DNA sequence to chromatin accessibility can be used to assess the potential regulatory impacts of *de novo*, noncoding mutations. The modeling of the regulatory potential of individual base pairs was crucial to enable the identification of these putative causal mutations, as simple overlap with open chromatin regions did not provide the required specificity. We observed enrichments of mutations in ASD cases versus controls that approached levels observed for deleterious protein-coding mutations (An et al., 2018). We anticipate that as more large-scale ATAC-seq and RNA-seq datasets across development become available, similar approaches will allow accurate interpretation of gene-regulatory impacts of noncoding *de novo* mutations associated with other developmental disorders.

Limitations of the study

Although these data span 8 weeks of mid-gestation, an analysis at earlier and later time points would allow further study of gliogenesis and neuronal maturation and, for instance, connect astrocyte precursors to adult subtypes. Of particular interest would be to employ rapidly advancing lineage tracing methods to resolve developmental trajectories identified here. While the multiome data validate many key inferences, the use of data integration inferences to connect singleome ATAC-seq with RNA-seq and to infer lineage relationships between cells is a limitation of this study. Furthermore, our cell-specific models consider impacts of variants on peaks present only in that particular cell type. Therefore, these cell-type-specific models likely trade greater significance, afforded by scoring larger sets of overlapping mutations in pseudobulk peak calls, for a deeper understanding of the specific cell types affected by the variants. Finally, confirming the deleterious nature of noncoding *de novo* mutations prioritized in this study will require molecular validation in the cognate cell types.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cell.2021.07.039.

ACKNOWLEDGMENTS

We thank members of the Greenleaf, Pașca, Kundaje, and Chang labs for discussion and advice, especially X. Chen, B. Parks, F. Birey, J. Granja, and A. Banerjee. This work was supported by the Rita Allen Foundation (W.J.G.), S. Coates and the Vj Coates Foundation (S.P.P.), the Human Frontiers Science RGY006S (W.J.G), the Stanford Brain Organogenesis Program and the Big Idea Grant (S.P.P.), and the Kwan Fund (S.P.P). W.J.G. is a Chan Zuckerberg Biohub investigator and acknowledges grants 2017-174468 and 2018-182817 from the Chan Zuckerberg Initiative. S.P.P. is a New York Stem Cell Foundation Robertson Stem Cell investigator and a Chan Zuckerberg Ben Barres investigator. H.Y.C. is an investigator of the Howard Hughes Medical Institute. Fellowship support was provided by the NSF Graduate Research Fellowship Program, the Siebel Scholars, the Enhancing Diversity in Graduate Education Program, and the Weiland Family Fellowship (A.E.T.); the Idun Berry Postdoctoral Fellowship (J.A.); the Deutsche Forschungsgemeinschaft (DFG) postdoctoral fellowship (grant MU 4303/1-1, F.M.); and the BioX Bowes Fellowship (L.S.).

AUTHOR CONTRIBUTIONS

A.E.T., F.M., J.A., S.P.P., and W.J.G. conceived the project and designed experiments. A.E.T. and F.M. performed data analysis. J.A. guided the biological interpretation of the analysis and performed validations. A.S. developed the original code (ChromBPNet) and deep-learning models of base resolution chromatin accessibility. L.S. adapted and trained the deep-learning models on the primary tissue data and performed the analysis on disease relevance with assistance from A.E.T., A.S., K.F., W.J.G., and A. Kundaje. A.M.P. and J.A. processed the samples for single-cell experiments. A.E.T. and A. Kathiria performed single-cell experiments. A.E.T., F.M., J.A., L.S., S.P.P., and W.J.G. wrote the manuscript with input from all authors. S.P.P. and W.J.G supervised the work.

DECLARATION OF INTERESTS

W.J.G. was a consultant for 10x Genomics and is named as an inventor on patents describing ATAC-seq methods. H.Y.C. is a co-founder of Accent Therapeutics and Boundless Bio and an advisor of 10x Genomics, Arsenal Biosciences, and Spring Discovery. A.S. is an employee of Insitro, Inc, and receives consulting fees from Myokardia, Inc. K.F. is an employee of Illumina, Inc.

Received: December 29, 2020
Revised: May 18, 2021
Accepted: July 28, 2021
Published: August 13, 2021

REFERENCES

Abrahams, B.S., Arking, D.E., Campbell, D.B., Mefford, H.C., Morrow, E.M., Weiss, L.A., Menashe, I., Wadkins, T., Banerjee-Basu, S., and Packer, A. (2013). SFARI Gene 2.0: a community-driven knowledgebase for the autism spectrum disorders (ASDs). Mol. Autism *4*, 36.

Alexa, A., Rahnenführer, J., and Lengauer, T. (2006). Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. Bioinformatics *22*, 1600–1607.

Amiri, A., Coppola, G., Scuderi, S., Wu, F., Roychowdhury, T., Liu, F., Pochareddy, S., Shin, Y., Safi, A., Song, L., et al. (2018). Transcriptome and epigenome landscape of human cortical development modeled in organoids. Science *362*, eaat6720.

An, J.-Y., Lin, K., Zhu, L., Werling, D.M., Dong, S., Brand, H., Wang, H.Z., Zhao, X., Schwartz, G.B., Collins, R.L., et al. (2018). Genome-wide de novo risk score implicates promoter variation in autism spectrum disorder. Science *362*, eaat6576.

Avsec, Ž., Weilert, M., Shrikumar, A., Krueger, S., Alexandari, A., Dalal, K., Fropf, R., McAnany, C., Gagneur, J., Kundaje, A., et al. (2020). Deep learning at base-resolution reveals cis-regulatory motif syntax. bioRxiv. https://doi.org/10.1101/737981.

Barbarese, E., Barry, C., Chou, C.-H.J., Goldstein, D.J., Nakos, G.A., Hyde-DeRuyscher, R., Scheld, K., and Carson, J.H. (1988). Expression and localization of myelin basic protein in oligodendrocytes and transfected fibroblasts. J. Neurochem. *51*, 1737–1745.

Bentsen, M., Goymann, P., Schultheis, H., Klee, K., Petrova, A., Wiegandt, R., Fust, A., Preussner, J., Kuenne, C., Braun, T., et al. (2020). ATAC-seq footprinting unravels kinetics of transcription factor binding during zygotic genome activation. Nat. Commun. *11*, 4267.

Bergen, V., Lange, M., Peidli, S., Wolf, F.A., and Theis, F.J. (2020). Generalizing RNA velocity to transient cell states through dynamical modeling. Nat. Biotechnol. *38*, 1408–1414.

Bhaduri, A., Andrews, M.G., Mancia Leon, W., Jung, D., Shin, D., Allen, D., Jung, D., Schmunk, G., Haeussler, M., Salma, J., et al. (2020). Cell stress in cortical organoids impairs molecular subtype specification. Nature *578*, 142–148.

Buenrostro, J.D., Corces, M.R., Lareau, C.A., Wu, B., Schep, A.N., Aryee, M.J., Majeti, R., Chang, H.Y., and Greenleaf, W.J. (2018). Integrated Single-Cell Analysis Maps the Continuous Regulatory Landscape of Human Hematopoietic Differentiation. Cell *173*, 1535–1548.e16.

Corces, M.R., Granja, J.M., Shams, S., Louie, B.H., Seoane, J.A., Zhou, W., Silva, T.C., Groeneveld, C., Wong, C.K., Cho, S.W., et al. (2018). The chromatin accessibility landscape of primary human cancers. Science *362*, eaav1898.

Cusanovich, D.A., Hill, A.J., Aghamirzaie, D., Daza, R.M., Pliner, H.A., Berletch, J.B., Filippova, G.N., Huang, X., Christiansen, L., DeWitt, W.S., et al. (2018). A Single-Cell Atlas of In Vivo Mammalian Chromatin Accessibility. Cell *174*, 1309–1324.e18.

Farrell, J.A., Wang, Y., Riesenfeld, S.J., Shekhar, K., Regev, A., and Schier, A.F. (2018). Single-cell reconstruction of developmental trajectories during zebrafish embryogenesis. Science *360*, eaar3131.

Fietz, S.A., Kelava, I., Vogt, J., Wilsch-Bräuninger, M., Stenzel, D., Fish, J.L., Corbeil, D., Riehn, A., Distler, W., Nitsch, R., and Huttner, W.B. (2010). OSVZ progenitors of human and ferret neocortex are epithelial-like and expand by integrin signaling. Nat. Neurosci. *13*, 690–699.

Gandal, M.J., Zhang, P., Hadjimichael, E., Walker, R.L., Chen, C., Liu, S., Won, H., van Bakel, H., Varghese, M., Wang, Y., et al.; PsychENCODE Consortium (2018). Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder. Science *362*, eaat8127.

Granja, J.M., Klemm, S., McGinnis, L.M., Kathiria, A.S., Mezger, A., Corces, M.R., Parks, B., Gars, E., Liedtke, M., Zheng, G.X.Y., et al. (2019). Single-cell multiomic analysis identifies regulatory programs in mixed-phenotype acute leukemia. Nat. Biotechnol. *37*, 1458–1465.

Greig, L.C., Woodworth, M.B., Galazo, M.J., Padmanabhan, H., and Macklis, J.D. (2013). Molecular logic of neocortical projection neuron specification, development and diversity. Nat. Rev. Neurosci. *14*, 755–769.

Hansen, D.V., Lui, J.H., Parker, P.R.L., and Kriegstein, A.R. (2010). Neurogenic radial glia in the outer subventricular zone of human neocortex. Nature *464*, 554–561.

Hodge, R.D., Bakken, T.E., Miller, J.A., Smith, K.A., Barkan, E.R., Graybuck, L.T., Close, J.L., Long, B., Johansen, N., Penn, O., et al. (2019). Conserved cell types with divergent features in human versus mouse cortex. Nature *573*, 61–68.

Imayoshi, I., Isomura, A., Harima, Y., Kawaguchi, K., Kori, H., Miyachi, H., Fujiwara, T., Ishidate, F., and Kageyama, R. (2013). Oscillatory control of factors determining multipotency and fate in mouse neural progenitors. Science *342*, 1203–1208.

Iossifov, I., O'Roak, B.J., Sanders, S.J., Ronemus, M., Krumm, N., Levy, D., Stessman, H.A., Witherspoon, K.T., Vives, L., Patterson, K.E., et al. (2014). The contribution of de novo coding mutations to autism spectrum disorder. Nature *515*, 216–221.

Jacquet, B.V., Salinas-Mondragon, R., Liang, H., Therit, B., Buie, J.D., Dykstra, M., Campbell, K., Ostrowski, L.E., Brody, S.L., and Ghashghaei, H.T. (2009). FoxJ1-dependent gene expression is required for differentiation of radial glia into ependymal cells and a subset of astrocytes in the postnatal brain. Development *136*, 4021–4031.

Kang, H.J., Kawasawa, Y.I., Cheng, F., Zhu, Y., Xu, X., Li, M., Sousa, A.M.M., Pletikos, M., Meyer, K.A., Sedmak, G., et al. (2011). Spatio-temporal transcriptome of the human brain. Nature *478*, 483–489.

Karczewski, K.J., Francioli, L.C., Tiao, G., Cummings, B.B., Alföldi, J., Wang, Q., Collins, R.L., Laricchia, K.M., Ganna, A., Birnbaum, D.P., et al.; Genome Aggregation Database Consortium (2020). The mutational constraint spectrum quantified from variation in 141,456 humans. Nature *581*, 434–443.

Kelley, D.R., Snoek, J., and Rinn, J.L. (2016). Basset: learning the regulatory code of the accessible genome with deep convolutional neural networks. Genome Res. *26*, 990–999.

Kelley, D.R., Reshef, Y.A., Bileschi, M., Belanger, D., McLean, C.Y., and Snoek, J. (2018). Sequential regulatory activity prediction across chromosomes with convolutional neural networks. Genome Res. *28*, 739–750.

Kelsey, G., Stegle, O., and Reik, W. (2017). Single-cell epigenomics: Recording the past and predicting the future. Science *358*, 69–75.

Khan, A., Fornes, O., Stigliani, A., Gheorghe, M., Castro-Mondragon, J.A., van der Lee, R., Bessy, A., Chèneby, J., Kulkarni, S.R., Tan, G., et al. (2018). JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework. Nucleic Acids Res. *46* (D1), D260–D266.

Klemm, S.L., Shipony, Z., and Greenleaf, W.J. (2019). Chromatin accessibility and the regulatory epigenome. Nat. Rev. Genet. *20*, 207–220.

La Manno, G., Soldatov, R., Zeisel, A., Braun, E., Hochgerner, H., Petukhov, V., Lidschreiber, K., Kastriti, M.E., Lönnerberg, P., Furlan, A., et al. (2018). RNA velocity of single cells. Nature *560*, 494–498.

Li, M., Santpere, G., Kawasawa, Y.I., Evgrafov, O.V., Gulden, F.O., Pochareddy, S., Sunkin, S.M., Li, Z., Shin, Y., Zhu, Y., et al. (2018a). Integrative functional genomic analysis of human brain development and neuropsychiatric risks. Science *362*, eaat7615.

Li, Z., Cogswell, M., Hixson, K., Brooks-Kayal, A.R., and Russek, S.J. (2018b). Nuclear Respiratory Factor 1 (NRF-1) Controls the Activity Dependent Transcription of the GABA-A Receptor Beta 1 Subunit Gene in Neurons. Front. Mol. Neurosci. *11*, 285.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550.

Lui, J.H., Hansen, D.V., and Kriegstein, A.R. (2011). Development and evolution of the human neocortex. Cell *146*, 18–36.

Lundberg, S., and Lee, S.I. (2017). A Unified Approach to Interpreting Model Predictions. arXiv, 1705.07874. https://arxiv.org/abs/1705.07874.

Ma, S., Zhang, B., LaFave, L., Earl, A.S., Chiang, Z., Hu, Y., Ding, J., Brack, A., Kartha, V.K., Tay, T., et al. (2020). Chromatin Potential Identified by Shared Single-Cell Profiling of RNA and Chromatin. Cell *183*, 1103–1116.e20.

McConnell, S.K. (1995). Constructing the cerebral cortex: neurogenesis and fate determination. Neuron *15*, 761–768.

McGinnis, C.S., Murrow, L.M., and Gartner, Z.J. (2019). DoubletFinder: Doublet Detection in Single-Cell RNA Sequencing Data Using Artificial Nearest Neighbors. Cell Syst. *8*, 329–337.e4.

McInnes, L., Healy, J., and Melville, J. (2020). UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. arXiv, 1802.03426. https://arxiv.org/abs/1802.03426.

Miller, J.A., Ding, S.-L., Sunkin, S.M., Smith, K.A., Ng, L., Szafer, A., Ebbert, A., Riley, Z.L., Royall, J.J., Aiona, K., et al. (2014). Transcriptional landscape of the prenatal human brain. Nature *508*, 199–206.

Molnár, Z., Clowry, G.J., Šestan, N., Alzu'bi, A., Bakken, T., Hevner, R.F., Hüppi, P.S., Kostović, I., Rakic, P., Anton, E.S., et al. (2019). New insights into the development of the human cerebral cortex. J. Anat. *235*, 432–451.

Nowakowski, T.J., Pollen, A.A., Sandoval-Espinosa, C., and Kriegstein, A.R. (2016). Transformation of the Radial Glia Scaffold Demarcates Two Stages of Human Cerebral Cortex Development. Neuron *91*, 1219–1227.

Nowakowski, T.J., Bhaduri, A., Pollen, A.A., Alvarado, B., Mostajo-Radji, M.A., Di Lullo, E., Haeussler, M., Sandoval-Espinosa, C., Liu, S.J., Velmeshev, D., et al. (2017). Spatiotemporal gene expression trajectories reveal developmental hierarchies of the human cortex. Science *358*, 1318–1323.

Oberheim, N.A., Takano, T., Han, X., He, W., Lin, J.H.C., Wang, F., Xu, Q., Wyatt, J.D., Pilcher, W., Ojemann, J.G., et al. (2009). Uniquely hominid features of adult human astrocytes. J. Neurosci. *29*, 3276–3287.

Parikshak, N.N., Luo, R., Zhang, A., Won, H., Lowe, J.K., Chandran, V., Horvath, S., and Geschwind, D.H. (2013). Integrative functional genomic analyses implicate specific molecular pathways and circuits in autism. Cell *155*, 1008–1021.

Parker, S.C.J., Stitzel, M.L., Taylor, D.L., Orozco, J.M., Erdos, M.R., Akiyama, J.A., van Bueren, K.L., Chines, P.S., Narisu, N., Black, B.L., et al.; NISC Comparative Sequencing Program; National Institutes of Health Intramural Sequencing Center Comparative Sequencing Program Authors; NISC Comparative Sequencing Program Authors (2013). Chromatin stretch enhancer states drive cell-specific gene regulation and harbor human disease risk variants. Proc. Natl. Acad. Sci. USA *110*, 17921–17926.

Pliner, H.A., Packer, J.S., McFaline-Figueroa, J.L., Cusanovich, D.A., Daza, R.M., Aghamirzaie, D., Srivatsan, S., Qiu, X., Jackson, D., Minkina, A., et al. (2018). Cicero Predicts cis-Regulatory DNA Interactions from Single-Cell Chromatin Accessibility Data. Mol. Cell *71*, 858–871.e8.

Polioudakis, D., de la Torre-Ubieta, L., Langerman, J., Elkins, A.G., Shi, X., Stein, J.L., Vuong, C.K., Nichterwitz, S., Gevorgian, M., Opland, C.K., et al. (2019). A Single-Cell Transcriptomic Atlas of Human Neocortical Development during Mid-gestation. Neuron *103*, 785–801.e8.

Pollen, A.A., Nowakowski, T.J., Chen, J., Retallack, H., Sandoval-Espinosa, C., Nicholas, C.R., Shuga, J., Liu, S.J., Oldham, M.C., Diaz, A., et al. (2015). Molecular identity of human outer radial glia during cortical development. Cell *163*, 55–67.

Ramanathan, S., Woodroffe, A., Flodman, P.L., Mays, L.Z., Hanouni, M., Modahl, C.B., Steinberg-Epstein, R., Bocian, M.E., Spence, M.A., and Smith, M. (2004). A case of autism with an interstitial deletion on 4q leading to hemizygosity for genes encoding for glutamine and glycine neurotransmitter receptor sub-units (AMPA 2, GLRA3, GLRB) and neuropeptide receptors NPY1R, NPY5R. BMC Med. Genet. *5*, 10.

Ransom, B.R. (2012). Neuroglia (Oxford University Press).

Rubenstein, J.L.R. (2011). Annual Research Review: Development of the cerebral cortex: implications for neurodevelopmental disorders. J. Child Psychol. Psychiatry *52*, 339–355.

Schep, A.N., Wu, B., Buenrostro, J.D., and Greenleaf, W.J. (2017). chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. Nat. Methods *14*, 975–978.

Sheffield, N.C., and Bock, C. (2016). LOLA: enrichment analysis for genomic region sets and regulatory elements in R and Bioconductor. Bioinformatics *32*, 587–589.

Shrikumar, A., Greenside, P., and Kundaje, A. (2019). Learning Important Features Through Propagating Activation Differences. arXiv, 1704.02685. https://arxiv.org/abs/1704.02685.

Silbereis, J.C., Pochareddy, S., Zhu, Y., Li, M., and Sestan, N. (2016). The Cellular and Molecular Landscapes of the Developing Human Central Nervous System. Neuron *89*, 248–268.

Sloan, S.A., Darmanis, S., Huber, N., Khan, T.A., Birey, F., Caneda, C., Reimer, R., Quake, S.R., Barres, B.A., and Paşca, S.P. (2017). Human Astrocyte Maturation Captured in 3D Cerebral Cortical Spheroids Derived from Pluripotent Stem Cells. Neuron *95*, 779–790.e6.

Smit, A.F.A., Hubley, R., and Green, P. (2010). RepeatMasker Open-3.0. http://www.repeatmasker.org.

Song, M., Pebworth, M.-P., Yang, X., Abnousi, A., Fan, C., Wen, J., Rosen, J.D., Choudhary, M.N.K., Cui, X., Jones, I.R., et al. (2020). Cell-type-specific 3D epigenomes in the developing human cortex. Nature *587*, 644–649.

Stergachis, A.B., Neph, S., Reynolds, A., Humbert, R., Miller, B., Paige, S.L., Vernot, B., Cheng, J.B., Thurman, R.E., Sandstrom, R., et al. (2013). Developmental fate and cellular maturity encoded in human regulatory DNA landscapes. Cell *154*, 888–903.

Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., 3rd, Hao, Y., Stoeckius, M., Smibert, P., and Satija, R. (2019). Comprehensive Integration of Single-Cell Data. Cell *177*, 1888–1902.e21.

Thomsen, E.R., Mich, J.K., Yao, Z., Hodge, R.D., Doyle, A.M., Jang, S., Shehata, S.I., Nelson, A.M., Shapovalova, N.V., Levi, B.P., and Ramanathan, S. (2016). Fixed single-cell transcriptomic characterization of human radial glial diversity. Nat. Methods *13*, 87–93.

Tjärnberg, A., Mahmood, O., Jackson, C.A., Saldi, G.-A., Cho, K., Christiaen, L.A., and Bonneau, R.A. (2021). Optimal tuning of weighted kNN- and diffusion-based methods for denoising single cell genomics data. PLoS Comput. Biol. *17*, e1008569.

Trevino, A.E., Sinnott-Armstrong, N., Andersen, J., Yoon, S.-J., Huber, N., Pritchard, J.K., Chang, H.Y., Greenleaf, W.J., and Paşca, S.P. (2020). Chromatin accessibility dynamics in a model of human forebrain development. Science *367*, eaay1645.

van Dijk, D., Sharma, R., Nainys, J., Yim, K., Kathail, P., Carr, A.J., Burdziak, C., Moon, K.R., Chaffer, C.L., Pattabiraman, D., et al. (2018). Recovering Gene Interactions from Single-Cell Data Using Data Diffusion. Cell *174*, 716–729.e27.

Vasile, F., Dossi, E., and Rouach, N. (2017). Human astrocytes: structure and functions in the healthy brain. Brain Struct. Funct. *222*, 2017–2029.

Vierstra, J., Lazar, J., Sandstrom, R., Halow, J., Lee, K., Bates, D., Diegel, M., Dunn, D., Neri, F., Haugen, E., et al. (2020). Global reference mapping of human transcription factor footprints. Nature *583*, 729–736.

Whyte, W.A., Orlando, D.A., Hnisz, D., Abraham, B.J., Lin, C.Y., Kagey, M.H., Rahl, P.B., Lee, T.I., and Young, R.A. (2013). Master transcription factors and mediator establish super-enhancers at key cell identity genes. Cell *153*, 307–319.

Wiese, S., Karus, M., and Faissner, A. (2012). Astrocytes as a source for extracellular matrix molecules and cytokines. Front. Pharmacol. *3*, 120.

Willsey, A.J., Sanders, S.J., Li, M., Dong, S., Tebbenkamp, A.T., Muhle, R.A., Reilly, S.K., Lin, L., Fertuzinhos, S., Miller, J.A., et al. (2013). Coexpression networks implicate human midfetal deep cortical projection neurons in the pathogenesis of autism. Cell *155*, 997–1007.

Wonders, C.P., and Anderson, S.A. (2006). The origin and specification of cortical interneurons. Nat. Rev. Neurosci. *7*, 687–696.

Yu, G., Wang, L.-G., Han, Y., and He, Q.-Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS *16*, 284–287.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., et al. (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol. *9*, R137.

Zhang, Y., Sloan, S.A., Clarke, L.E., Caneda, C., Plaza, C.A., Blumenthal, P.D., Vogel, H., Steinberg, G.K., Edwards, M.S.B., Li, G., et al. (2016). Purification and Characterization of Progenitor and Mature Human Astrocytes Reveals Transcriptional and Functional Differences with Mouse. Neuron *89*, 37–53.

Zhou, Q., and Anderson, D.J. (2002). The bHLH transcription factors OLIG2 and OLIG1 couple neuronal and glial subtype specification. Cell *109*, 61–73.

Zhou, J., and Troyanskaya, O.G. (2015). Predicting effects of noncoding variants with deep learning-based sequence model. Nat. Methods *12*, 931–934.

Zhou, J., Park, C.Y., Theesfeld, C.L., Wong, A.K., Yuan, Y., Scheckel, C., Fak, J.J., Funk, J., Yao, K., Tajima, Y., et al. (2019). Whole-genome deep-learning analysis identifies contribution of noncoding mutations to autism risk. Nat. Genet. *51*, 973–980.