RESEARCH ARTICLE SUMMARY

NEUROGENETICS

Chromatin accessibility dynamics in a model of human forebrain development

Alexandro E. Trevino*, Nasa Sinnott-Armstrong*, Jimena Andersen*, Se-Jin Yoon, Nina Huber, Jonathan K. Pritchard, Howard Y. Chang, William J. Greenleaf†, Sergiu P. Paşca†

INTRODUCTION: The cerebral cortex is responsible for higher-order functions in the nervous system and has undergone substantial expansion in size in primates. The development of the forebrain, including the assembly of the expanded human cerebral cortex, is a lengthy process that involves the diversification and expansion of neural progenitors, the generation and positioning of layer-specific glutamatergic neurons, the cellular migration of γ-aminobutyric acid (GABA)–ergic neurons, and the formation and maturation of glial cells. Disruption of these cellular events by either genetic or environmental factors can lead to neurodevelopmental disease, including autism spectrum disorders and intellectual disability.

RATIONALE: Human forebrain development is, to a large extent, inaccessible for cellular-level study, direct functional investigation, or manipulation. The lack of availability of primary brain tissue samples—in particular, at later stages—as well as the limitations of conventional in vitro cellular models have precluded a detailed mechanistic understanding of corticogenesis in healthy and disease states. Therefore, tracking epigenetic changes in specific forebrain cell lineages over long time periods, has the potential to unravel the molecular programs that underlie cell specification in the human cerebral cortex and, by temporally mapping disease risk onto these changes, to identify cell types and periods of increased disease susceptibility.

RESULTS: We used three-dimensional (3D) directed differentiation of human pluripotent stem cells into dorsal and ventral forebrain domains and applied the assay for transposase-accessible chromatin with high-throughput sequencing (ATAC-seq) in combination with RNA-sequencing (RNA-seq) to map the epigenetic and gene expression signatures of neuronal and glial cell lineages over 20 months in vitro. We show, through direct comparison with primary brain tissue from our study and several epigenetic datasets, that human stem cell–derived 3D forebrain organoids recapitulated in vivo chromatin accessibility patterns over time. We then integrated these data to discover putative enhancer-gene linkages and lineage-specific transcription factor regulators, including a diverse repertoire of factors that may control cortical specification. We validated protein expression of some of these transcription factors using immunofluorescence, confirming cellular and temporal dynamics in both primary tissue and forebrain organoids. Next, we used this resource to map genes and genetic variants associated with schizophrenia and autism spectrum disorders to distinct accessibility patterns to reveal cell types and periods of susceptibility. Last, we identified a wave of chromatin remodeling during cortical neurogenesis, during which a quarter of regulatory regions are active, and then highlighted transcription factors that may drive these developmental changes.

CONCLUSION: Using long-term 3D neural differentiation of stem cells as well as primary brain tissue samples, we found that organoids intrinsically undergo chromatin state transitions in vitro that are closely related to human forebrain development in vivo. Leveraging this platform, we identified epigenetic alterations putatively driven by specific transcription factors and discovered a dynamic period of chromatin remodeling during human cortical neurogenesis. Finally, we nominated several key transcription factors that may coordinate over time to drive these changes and mapped cell type–specific disease-associated variation over time and in specific cell types. ■
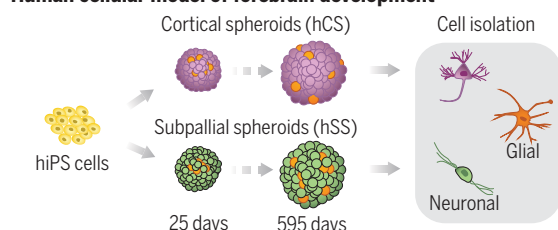

Developing a human cellular model of forebrain development to study chromatin dynamics. ATAC-seq and RNA-seq studies over long-term differentiation of human pluripotent stem cells into forebrain organoids and in primary brain tissue samples reveal dynamic changes during human corticogenesis, including a wave associated with neurogenesis, and identify disease-susceptible cell types and stages.

RESEARCH ARTICLE

NEUROGENETICS

Chromatin accessibility dynamics in a model of human forebrain development

Alexandro E. Trevino[1,2]*, Nasa Sinnott-Armstrong[3]*, Jimena Andersen[2,4]*, Se-Jin Yoon[2,4], Nina Huber[4], Jonathan K. Pritchard[3,5,6], Howard Y. Chang[3,6,7,8], William J. Greenleaf[3,9,10]†, Sergiu P. Pașca[4]†

Forebrain development is characterized by highly synchronized cellular processes, which, if perturbed, can cause disease. To chart the regulatory activity underlying these events, we generated a map of accessible chromatin in human three-dimensional forebrain organoids. To capture corticogenesis, we sampled glial and neuronal lineages from dorsal or ventral forebrain organoids over 20 months in vitro. Active chromatin regions identified in human primary brain tissue were observed in organoids at different developmental stages. We used this resource to map genetic risk for disease and to explore evolutionary conservation. Moreover, we integrated chromatin accessibility with transcriptomics to identify putative enhancer-gene linkages and transcription factors that regulate human corticogenesis. Overall, this platform brings insights into gene-regulatory dynamics at previously inaccessible stages of human forebrain development, including signatures of neuropsychiatric disorders.

T he assembly of the human cerebral cortex is a dynamic and lengthy process that begins during early gestation and continues postnatally (*1*). It is characterized by the inside-out generation of neurons in the pallium from progenitors, followed by the generation and maturation of astrocytes and other glial cells. γ-aminobutyric acid (GABA)–ergic interneurons, born in the subpallium, migrate into the cerebral cortex, where they integrate to form circuits (*2*, *3*). Although genetic and environmental perturbations of corticogenesis can cause severe neurodevelopmental disease (*4*), a detailed understanding of the molecular programs that govern these cellular events remains elusive (*1*).

Epigenetic gene regulation plays a crucial role in controlling developmental transitions and cellular differentiation (*5*, *6*). Gene-distal enhancer elements are dynamic throughout development, exhibiting only brief activity in restricted cell populations, yet are enriched for disease-associated genetic variants (*7*, *8*). Moreover, evolutionary divergence of regulatory elements can affect phenotypes in the absence of protein coding changes (*9*). Chromatin accessibility has emerged as an accurate proxy for regulatory potential (*10*). Therefore, defining accessibility across human brain cell lineages and time could provide understanding of gene-regulatory principles and disease signatures in the human forebrain.

Previous studies have begun to define the transcriptome and epigenome of the developing human forebrain, including characterization of spatiotemporal gene expression in the cortex (*11–14*), the molecular signature of cortical progenitors (*15*, *16*), epigenetics of early brain development (*17*), and the impact of recently evolved enhancers on gene expression (*18–21*). However, brain tissue remains difficult to study, precluding the tracking of epigenetic dynamics in specific cellular lineages over long time periods and, consequently, a mechanistic understanding of forebrain development.

Human induced pluripotent stem cell (hiPS) cell–derived three-dimensional (3D) organoids provide a distinct opportunity to study brain development in vitro and have provided insight into early stages of brain development (*17*, *22*). To capture later stages of development, we developed methods to generate 3D brain region–specific organoids that resemble the dorsal forebrain, called cortical spheroids (hCSs) (*23*, *24*) and the ventral forebrain, called human subpallial spheroids (hSSs) (*25*). Differentiations into regionalized organoids are highly reliable (*26*) and demonstrate a transition from fetal to early postnatal stages at ~280 days in vitro (*24*).

Results

Characterization of accessibility in hCSs and hSSs

We differentiated hCSs and hSSs from hiPS cells for ~20 months in vitro (Fig. 1A). As described (*23*, *24*, *26*, *27*), hCSs recapitulate key features of cortical development in vitro, including the emergence of radial glial progenitors (Fig. 1B and fig. S1A), followed by the progressive generation of deep and superficial glutamatergic neurons (Fig. 1C and fig. S1B), and last, the generation and maturation of astrocytes (Fig. 1D and fig. S1, C and D). Ventral forebrain hSSs generate glutamic acid decarboxylase 67$^+$ (GAD67$^+$) and GABA$^+$ interneurons (Fig. 1E and fig. S1E) (*25*).

We collected 117 ATAC-seq (assay for transposase-accessible chromatin with high-throughput sequencing) samples and 54 RNA-sequencing (RNA-seq) samples from hCSs and hSSs derived from 7 hiPS cell lines (tables S1 and S2). At early stages, we isolated purified nuclei from whole, undissociated spheroids (*28*). After day 79, we purified glial and neuronal cell lineages by either surface marker immuno-panning or fluorescence-activated cell sorting (FACS). For immunopanning, glial lineage cells were isolated by using an antibody against HepaCAM, and neuronal lineage cells were isolated by using an antibody against Thy1 (CD90) (*24*, *29*). For FACS, cells were labeled with viral reporters driven by the human glial fibrillary acidic protein (GFAP) promoter or the synapsin 1 (SYN1) promoter [pLV-GFAP:: enhanced green fluorescent protein (eGFP) and adeno-associated virus (AAV)–SYN1:: mCherry, respectively] (Fig. 1A) (*30*). We confirmed the specificity of these markers using a single-cell RNA-seq dataset (fig. S2). We refer to GFAP$^+$ or HepaCAM$^+$ cells as glial lineage cells, which may encompass radial glia, outer radial glia, and mature astrocytes, depending on the differentiation stage. We refer to SYN1$^+$ and Thy1$^+$ cells as neuronal lineage cells, which may encompass hCS (glutamatergic) or hSS (GABAergic) neurons. We also performed ATAC-seq in cerebral cortex isolated from human fetal tissue (HFT) at post-conception week (PCW) 20 and PCW21, including samples isolated by immunopanning (fig. S1, F and G).

To evaluate the quality of ATAC-seq libraries generated, we used several metrics, including enrichment at transcription start sites (TSSs) and insert size distributions (table S3) (*30*). hCS and hSS TSS enrichments averaged 25.3-fold. We identified 703,306 reproducible peaks across the entire dataset, representing 244,044 contiguous open chromatin regions (*30*). ATAC-seq technical and biological replicates were highly correlated within peak regions [mean Pearson's correlation coefficient (r) = 0.97 and 0.94, respectively]. To confirm reproducibility, we compared longitudinal ATAC-seq data from hCSs and hSSs derived from two human hiPS cell lines, generated from two individuals, and found an overall mean correlation between lines of r = 0.94 (fig. S3A) (*30*). We did not detect changes in apoptosis, necrosis, or unfolded protein response–related gene sets over time [Spearman's rank correlation

[1]Department of Bioengineering, Stanford University, Stanford, CA, USA. [2]Stanford Human Brain Organogenesis Program, Stanford, CA, USA. [3]Department of Genetics, Stanford University, Stanford, CA, USA. [4]Department of Psychiatry and Behavioral Sciences, Stanford University, Stanford, CA, USA. [5]Department of Biology, Stanford University, Stanford, CA, USA. [6]Howard Hughes Medical Institute, Chevy Chase, MD, USA. [7]Department of Dermatology, Stanford University, Stanford, CA, USA. [8]Center for Personal Dynamic Regulomes, Stanford University, Stanford, CA, USA. [9]Department of Applied Physics, Stanford University, Stanford, CA, USA. [10]Chan Zuckerberg Biohub, San Francisco, CA, USA.
*These authors contributed equally to this work.
†Corresponding author. Email: wjg@stanford.edu (W.J.G.); spasca@stanford.edu (S.P.P.)
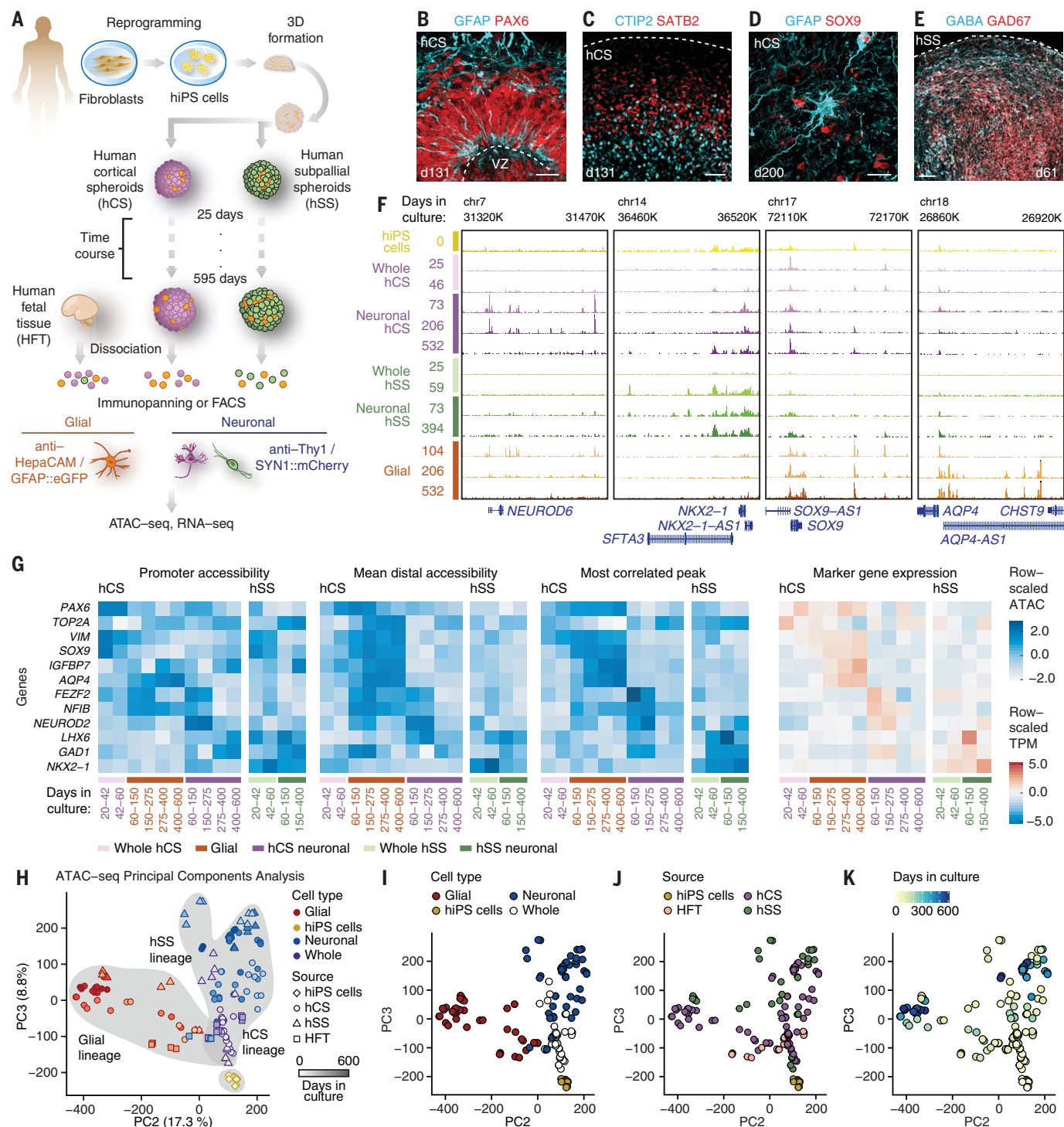
Fig. 1. Forebrain lineage markers and ATAC-seq clustering in hCSs and hSSs.
(**A**) Generation of hCSs and hSSs from hiPS cells. Neuronal and glial cells were collected by immunopanning or FACS for ATAC-seq and RNA-seq. Day 20 to 60 spheroids were collected intact. Cells from HFT at PCW20 and PCW21 were collected. (**B** and **C**) Immunohistochemistry in hCSs at day 131 (d131) showing expression of GFAP and PAX6 in a ventricular zone (VZ)–like region, and the layer-specific markers CTIP2 and SATB2. (**D** and **E**) Immunohistochemistry for the astrocyte markers GFAP and SOX9 in hCS day 200 (d200) and GABA and GAD67 in hSS day 61 (d61). (**F**) Genome browser plots of marker gene accessibility for hiPS cells, hCS neurons, hSS neurons, and glial cells. (**G**) Promoter and distal accessibility levels and expression. For the left three panels, heatmap color indicates the scaled accessibility level at the promoter (left), all distal elements averaged (middle), or the distal element with activity most correlated to gene expression (right). The rightmost panel shows RNA expression over time [in transcripts per million (TPM)]. Expression of genes with multiple RefSeq annotations was averaged. (**H**) PCA of all ATAC-seq samples. hiPS cells (yellow), whole hCSs or hSSs (green), glial cells (red), and neuronal cells (blue) are shown. Shapes indicate hCSs (circles), hSSs (triangles), hiPS cells (diamonds), and HFT (squares). Stage is represented by a gradient. hiPS cells represent day 0. (**I**) PCA of all ATAC-seq samples. Whole samples refer to both hCSs and hSSs. (**J**) PCA of all ATAC-seq samples by source. (**K**) PCA of all ATAC-seq samples showing differentiation timing. HFT at PCW20 to PCW21 were labeled as d140. Scale bars, 20 μm (B) and (D) and 50 μm (C) and (E).
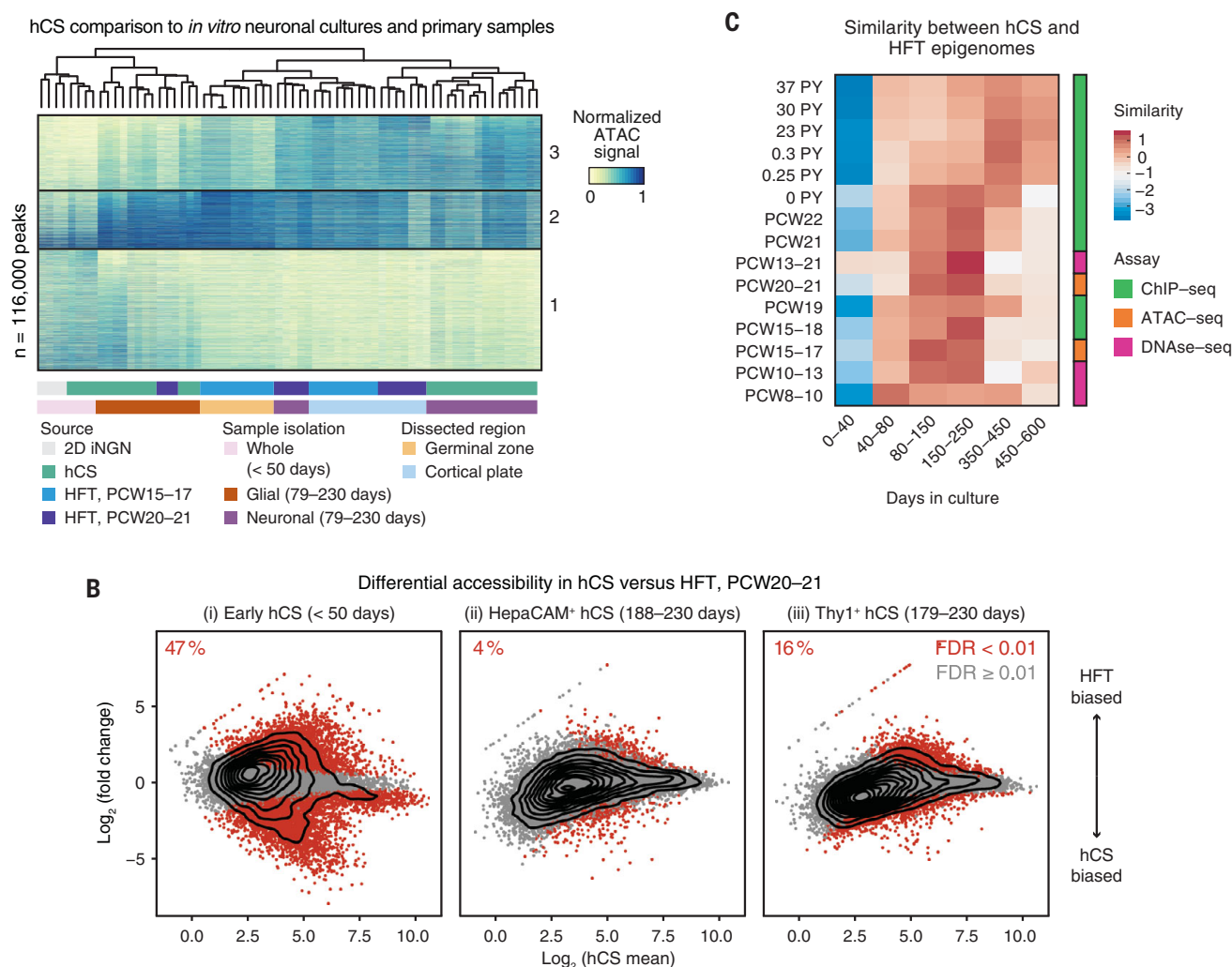
Fig. 2. Comparison of hCS and HFT chromatin accessibility landscapes. (**A**) Comparison of hCSs to HFT at the peak level by hierarchical and *K*-means clustering. Sample set includes iNGN, early hCSs, isolated hCS lineages (79 to 230 days), and PCW20 HFT and microdissected HFT from GZ and CP [from (*20*)]. (**B**) Differentially accessible peaks between early spheroids and glial and neuronal lineages (DESeq2). Differential peaks at FDR < 0.01 are in red; the three panels correspond to three differential tests. Panels compare (i) hCSs (<50 days) to HFT; (ii) HepaCAM⁺ cells from hCSs (179 to 230 days) and HFT; and (iii) Thy1⁺ cells from hCSs (179 to 230 days) and HFT. (**C**) Similarity of HFT samples to hCSs over time. Scaled Jaccard indices are plotted. Each row denotes the merged neuronal and glial hCS samples that are most similar to the HFT sample.

coefficient (ρ) = −0.08, 0.05, -0.01 and *P* = 0.96, 0.57, 0.74, respectively] (fig. S3, B to D, and table S4) (*30*).

ATAC-seq data revealed lineage- and time-specific accessibility differences near marker genes, including the glial marker *SOX9*, the cortical neuron marker *NEUROD6*, the sub-pallial progenitor marker *NKX2-1*, and the mature astrocyte marker *AQP4* (Fig. 1F). Because enhancer activity and gene expression can be coordinated (*31*), we explored relationships between gene expression and local chromatin accessibility patterns (Fig. 1G and fig. S4, A and B) (*30*). We found that average distal enhancer accessibility, defined as the mean ATAC-seq signal in peaks within 500 kb of the TSS, correlated more strongly with gene expression than promoter-proximal chroma-

tin accessibility (Pearson's $r = 0.52$, $P = 9.9 \times 10^{-05}$ and $r = 0.69$, $P = 2.7 \times 10^{-6}$, respectively). We also visualized the best-correlated distal ATAC-seq peak for each gene across cell type and time (Pearson's $r = 0.88$, $P = 6.2 \times 10^{-15}$), emphasizing the concordance between the expression of well-established lineage markers and associated chromatin accessibility patterns. Conversely, nonforebrain markers were expressed at low levels and did not display lineage-specific patterns, and regulatory elements at these loci were poorly correlated to expression (Pearson's $r = 0.23$, $P = 0.10$; $r = 0.11$, $P = 0.56$, and $r = 0.55$, $P = 1.5 \times 10^{-3}$, respectively) (fig. S4, C and D) (*30*).

To examine the global structure of the ATAC-seq data, we performed principal components analysis (PCA) (Fig. 1, H to K). The first three

principal components explained 48% of the variance in accessibility (fig. S5A) (*30*). Samples grouped into co-accessible cell populations: hiPS cells, whole hCSs and hSSs at 25 to 59 days in culture, early hCS-derived neurons at 79 to 230 days in culture, hSS-derived neurons, glial progenitors, and mature glia. Generally, we saw that hSS-derived neurons grouped with late-stage hCS-derived neurons (after 230 days in culture). Overall, PCA captured the differentiation of hiPS cells into neuronal and glial lineages (Figs. S5, B to D, and S6, A to E) (*30*).

Direct comparison of hCS lineages to primary human tissue

To examine the fidelity of our in vitro chromatin landscapes to those in vivo, we performed *K*-means clustering on ATAC-seq data
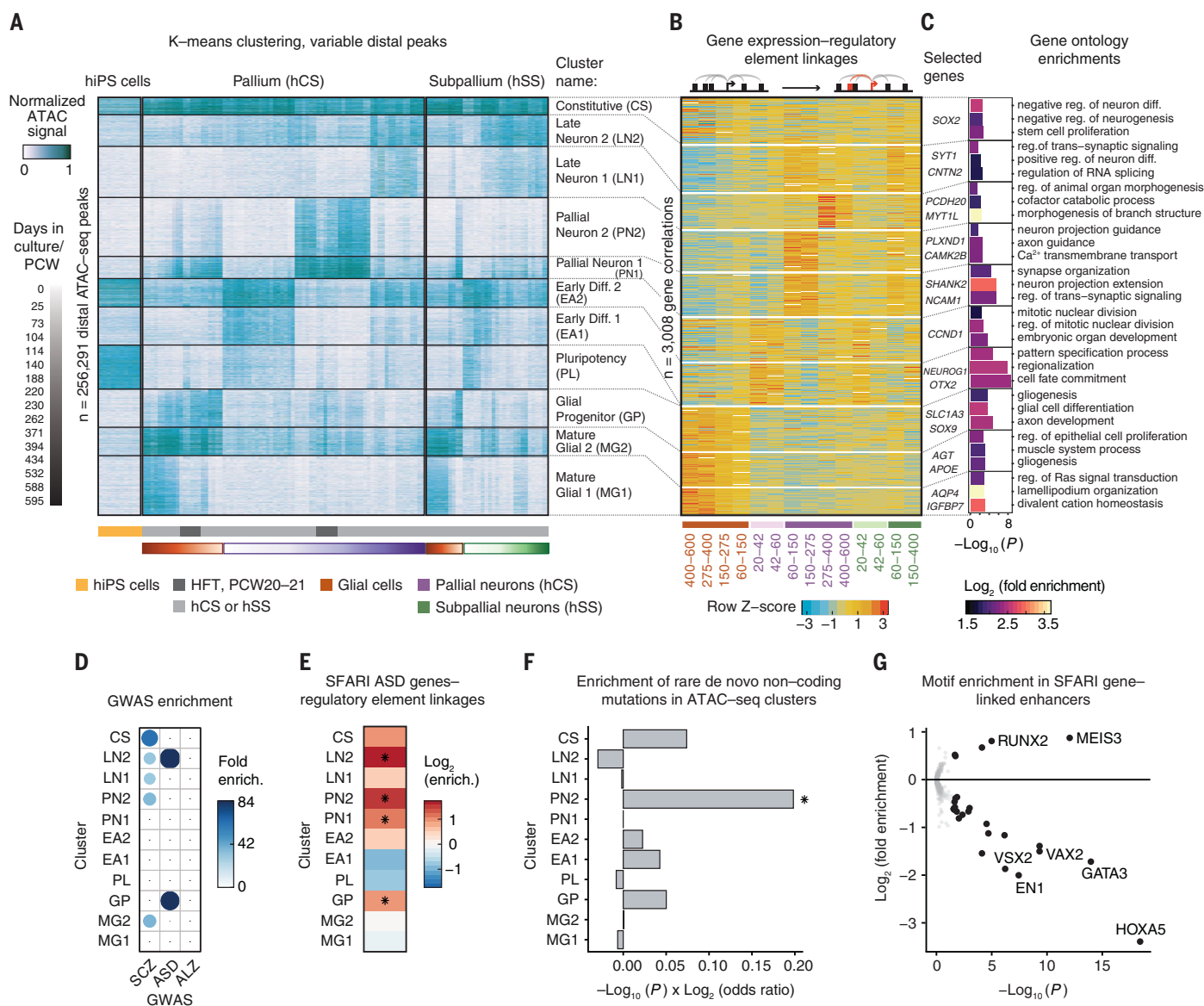
A
K–means clustering, variable distal peaks

B
Gene expression–regulatory element linkages

C
Selected genes

Gene ontology enrichments

Fig. 3. Gene-regulatory dynamics in hCSs, hSSs, and HFT. (**A**) *K*-means clustering of variable, distal accessible peaks. Activity is represented as normalized ATAC signal without additional row-scaling. Samples were sorted by stage and cell lineage. HFT samples were included in the *K*-means algorithm as d140. Stage or days in culture is represented by a gradient. (**B**) Heatmap showing expression of genes correlated to accessible elements from each *K*-means cluster within 500 kb of the TSS. RNA-seq clusters correspond to ATAC-seq clusters. Cluster PL was omitted because of lack of RNA-seq. (Top) Schematic of the correlation approach, displayed as row-standardized TPM. (**C**) Representative genes from each *K*-means cluster along with the GO enrichment within each cluster. *P* values derive from the hypergeometric test, and the color indicates the fold enrichment. (**D**) Enrichment of GWAS signal for schizophrenia, ASD, and Alzheimer's disease in clusters, estimated using LD score regression. The size and color indicate fold enrichment; data points without significance are shown with a black dot. (**E**) Enrichment of SFARI ASD genes in the linked gene expression clusters. Colors correspond to enrichment in the cluster, and stars indicate significance ($P < 0.05$). (**F**) Enrichment of de novo noncoding ASD mutations. Enrichment scores [$-\log_{10}(P) * \log_2(\text{enrichment})$]. (**G**) Motif enrichment in SFARI genes relative to enhancers of other linked genes displayed as a volcano plot [$\log_2(\text{enrichment})$ versus $-\log_{10}(P)$].

from early hCSs (25 to 46 days in vitro), 2D neurons derived by overexpression of NGN2 (iNGN; 27 to 41 days in vitro), hCS-derived glial and neuronal lineages (79 to 230 days in vitro), PCW20 HFT, and microdissected germinal zone (GZ) and cortical plate (CP) at PCW15 to PCW17 (*20*). As a feature selection step aimed at reducing the number of potentially noisy peaks with weak chromatin dynamics, we used the top 20% of peaks ranked by

standard deviation across this sample subset. We found three patterns of chromatin accessibility, which is consistent with hierarchical clustering of the samples ($K = 3$, $n = 116,000$ peaks) (Fig. 2A) (*30*). Accessible elements in the first cluster were most active in iNGN and hCSs at 25 to 46 days in vitro, whereas elements in the second cluster had the highest activity in hCSs and HFT glial lineage cells and in GZ, which is consistent with glial pro-

genitors representing a larger proportion of the GZ. Elements in the third cluster demonstrated specific activity in neuronal lineage cells isolated from hCSs and HFT, and CP, but not in early hCSs or iNGN. Next, we computed differential accessibility between isolated hCS and HFT lineages directly. We found that 47% of regions were significantly differentially accessible between whole HFT samples and whole hCSs at 25 to 59 days [DEseq2 false
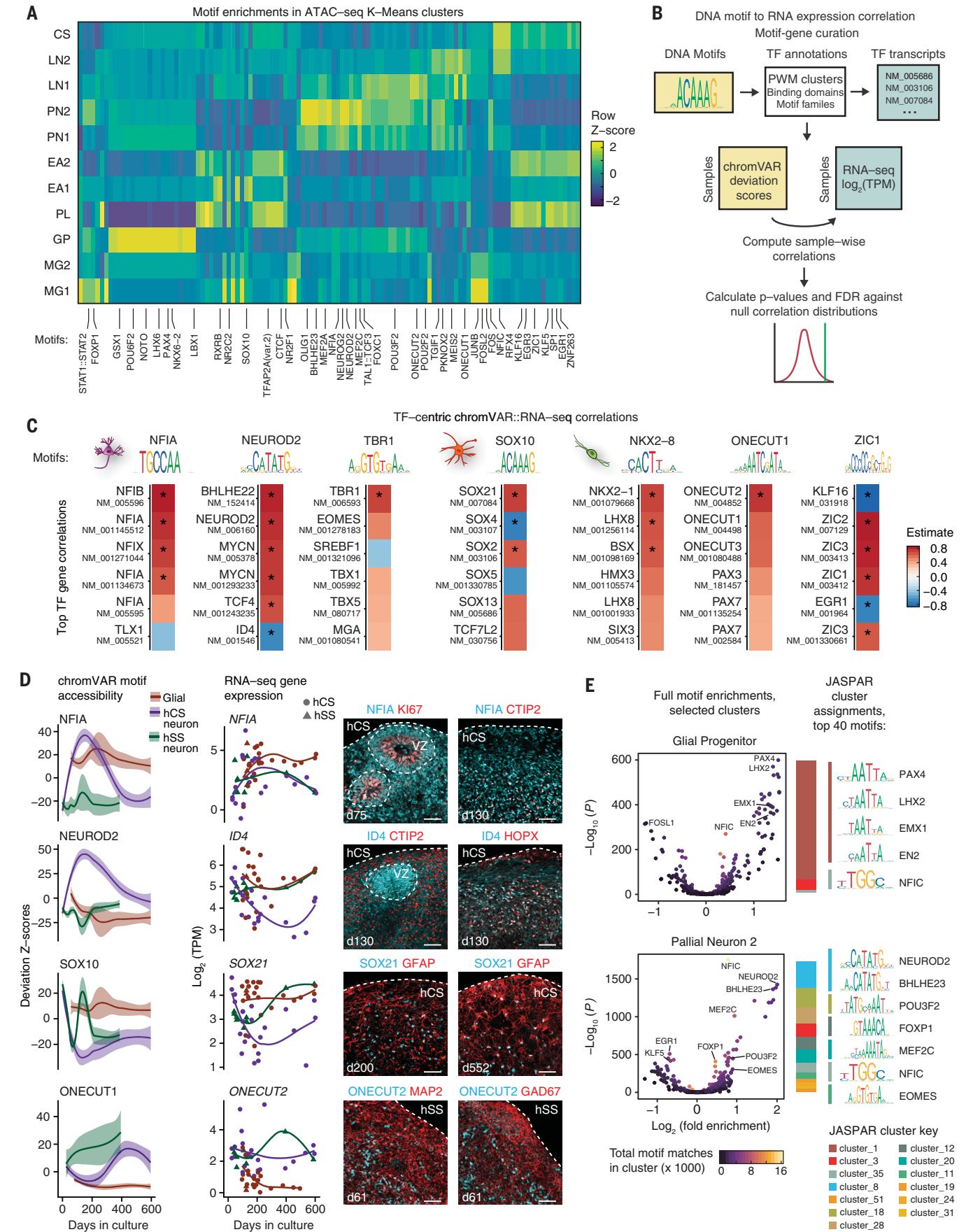
Fig. 4. Transcription factor activity in the forebrain. (**A**) Heatmap summarizing motif enrichments in ATAC-seq K-means clusters. The top 15 most-enriched motifs per cluster are displayed, and color indicates scaled $\log_2$(fold enrichment) versus other clusters. Select motifs are shown. (**B**) Schematic of chromVAR–RNA-seq expression correlation approach. Motifs are linked to TF genes that share a position weight matrix (PWM) cluster, family annotation, or binding domain by using available databases. For each motif, the correlation of chromVAR motif deviations to the expression values of each eligible TF gene is compared against a background set of correlations. A P value and FDR are computed against the null to determine significance. (**C**) ChromVAR-expression correlations for a subset of enriched motifs. Color indicates Pearson correlations for a TF motif–TF gene pair. The top six correlations are displayed. Stars indicate FDR-adjusted $P < 0.05$. Cell icons indicate the lineages with the greatest enrichments for a given motif. (**D**) ChromVAR motif accessibility (deviation Z scores), RNA expression ($\log_2$ TPM), and immunohistochemistry for candidate lineage-specific TFs identified by means of paired ATAC-seq and RNA-seq. For chromVAR, values are indicated with a smoothed line. RNA-seq values are shown by sample. Immunostainings show NFIA expression with KI67 and CTIP2 in hCSs (days 75 and 130); ID4 with HOPX and CTIP2 in hCS (day 130); SOX21 with GFAP in hCS (days 200 and 552); and ONECUT2 with MAP2 and GAD67 in hSS (day 61). (**E**) Motif volcano plots ($\log_2$ fold enrichment of motif in ATAC-seq cluster versus other clusters by $-\log_{10} P$ values) for GP and PN2. The color of each circle indicates the number of sequence matches in the foreground set for each motif. Cluster names are derived from the JASPAR database. The size of each colored box indicates the TF number from a given family that are enriched in the cluster. Scale bars, 50 μm (D).

discovery rate (FDR) < 0.01]. However, at later stages, only 4% of regions between HFT- and hCS-derived glia and 16% of regions between HFT- and hCS-derived neurons were differentially accessible (Fig. 2B) (*30*).

To explore the relationship of cortical development to our in vitro cultures over time, we integrated epigenetic datasets from PsychENCODE, ENCODE, and a recent study, representing human cortical tissue samples from early fetal to postnatal stages (*12, 17, 20, 32*). We computed the Jaccard similarity index (J) between peaks from each primary tissue sample and hCSs at different stages of in vitro differentiation. To account for differences between studies, we standardized the scores for each primary sample (Fig. 2C). We found that HFT at PCW8 to PCW10 was most similar to hCSs at 40 to 80 days. Mid– to late–fetal developmental stages up to birth [PCW10 to 0 postnatal years (PY)] resembled cultures at 80 to 250 days, whereas postnatal samples were most similar to hCSs after 350 days. We observed similar trends when displaying unscaled values of J (fig. S7, A to C) (*30*). The mapping of early hCS samples (<50 days in vitro) to primary tissue samples was weaker likely because they represent earlier stages of development for which no in vivo data are currently available. Nonetheless, we chose to keep these samples in downstream analyses because they may inform future in vitro models of development and disease. Although changes in tissue cell composition and limitations of our in vitro model make comparisons to later stages challenging, these data suggest that forebrain organoids undergo progressive chromatin remodeling commensurate with that observed in primary tissue. Despite limited samples for comparison, these trends were consistent across multiple epigenomic data types and studies and across long time scales, from early fetal to postnatal development.

Gene-regulatory landscapes in forebrain differentiation

To explore patterns of chromatin accessibility in cell lineages isolated from HFT and 3D forebrain cultures, we performed K-means

clustering on the top 40% of peaks ranked by standard deviation (hereafter, dynamic peaks; n = 256,291) (*30*). Inspection of these clusters revealed that they corresponded to six groups: (i) peaks with maximal accessibility in hiPS cells ["pluripotency" (PL)]; (ii) peaks specific to hCSs or hSSs between 25 and 59 days of differentiation ["early 1 and 2" (EA1 and EA2)]; (iii) glial lineage peaks from hCS, hSS, and HFT samples, which were subdivided into early and late-stages ["glial progenitor" (GP) and "mature glial 1 and 2" (MG1 and MG2)]; (iv) peaks specific to cortical neurogenesis between days 79 and 230, including HFT ["pallial neuron 1 and 2" (PN1 and PN2)]; (v) peaks specific to neurons from hCS after 230 days of differentiation and hSS ["late neuron 1 and 2" (LN1 and LN2)]; and (vi) peaks that were highly active across samples ["constitutive" (CS)] (Fig. 3A and table S1). We found that although hSS neurons were distinct from hCS neurons using a supervised approach (fig. S8A) (*30*), they did not form a separate cluster. Differential hSS-specific peaks comprised 41, 46, and 34% of the LN1, LN2, and EA2 clusters, respectively (fig. S8B) (*30*).

To verify the overlap of K-means clusters with epigenomes from various primary brain tissues and in vitro models, we computed the enrichment of reference datasets within the clusters (fig. S9, A to C) (*30*). Across studies and assays [including chromatin immunoprecipitation–sequencing (ChIP-seq), deoxyribonuclease-sequencing (DNase-seq), and ATAC-seq], we found that in vitro models were enriched in the early stage clusters EA1 and EA2, as well as the LN2 and CS clusters. Mid-fetal HFT (PCW14 to PCW22) datasets were strongly enriched in clusters PN1 and PN2, whereas glial maturation was well captured by a transition of enrichment from the GP to MG2 to MG1 clusters. Last, late fetal and early postnatal samples were enriched in the late neuronal LN2 cluster. Adult samples were not enriched in LN2, which could reflect time- or activity-dependent maturation not currently captured by our in vitro model. Using this collection of primary brain tissue– and cell culture–derived epigenomes, we computed

the fraction of peaks from each dataset overlapping our clusters, revealing that enrichments were driven by large fractions of overlaps (fig. S9D) (*30*).

We next aimed to define gene expression programs associated with this developmental accessibility landscape. We adapted an approach to discover putative enhancer-gene linkages on the basis of the coordination of accessibility and gene expression across cell lineages (*33*). For each gene, we then computed correlations across samples between gene expression and accessible peaks within a 500-kb window, taking into account the distribution of spurious correlations. This method identified 28,940 links corresponding to 8294 genes. For each cluster, we selected the top 400 strongest enhancer-gene correlations in the cluster for further analysis (omitting PL) (Fig. 3B). These groups included established lineage-specific markers, such as *AQP4* and *IGFBP7* in MG1 and *NFIA* and *SHANK2* in PN1. Each group was enriched for specific gene ontology (GO) terms, including cell fate commitment (EA1), synaptic signaling (PN1), neuron projection and axon guidance (PN2), and cation homeostasis and lamellipodium organization (MG1) (Fig. 3C and table S5).

Disease signatures in forebrain chromatin landscapes

Noncoding regions contain the majority of common variants that influence human disease (*7*), and we reasoned that genetic risk for brain disorders might localize to chromatin accessibility in specific cell lineages. We mapped single-nucleotide polymorphisms, shown to confer disease risk by genome-wide association studies (GWASs), onto our accessibility landscape. Using linkage disequilibrium (LD) score regression (*34, 35*), we measured enrichment of risk variants within each cluster (Fig. 3D). Variants associated with Alzheimer's disease, major depressive disorder, and epilepsy were not enriched in any clusters. Schizophrenia risk (*36*) was enriched across the MG2, PN2, LN1, LN2, and CS clusters, with PN2 cells exhibiting the most significant enrichment,

and considerable signal in the late neuronal clusters, which encompass hSS interneurons and late-stage hCS neurons (PN2 heritability enrichment = 37.7, block jackknife χ^2 test, $P = 4.5 \times 10^{-6}$; LN1 heritability enrichment = 30.8, block jackknife χ^2 test, $P = 5.2 \times 10^{-5}$). Meanwhile, the GP and LN2 clusters were enriched for autism spectrum disorder (ASD) risk (heritability enrichments = 84.7- and 99.3-fold, respectively; block jackknife χ^2 test, $P = 2.4 \times 10^{-3}$ and 7.1×10^{-3}). Despite a paucity of genome-wide significant ASD-associated variants to date, we found that GP was consistently and strongly enriched for risk across multiple GWASs. hSS peaks were enriched for genetic risk for schizophrenia, autism, and attention deficit hyperactivity disorder (fig. S10A) (*30*).

The Simons Foundation Autism Research Initiative (SFARI) curates a database of ASD-associated genes (http://gene.sfari.org). Although recent studies have examined expression patterns of these genes during fetal development or postnatally (*12, 37*), relatively little is known about the specific lineages that may be involved in ASD at different stages. Across our model, we found that 81% of all SFARI genes exhibited variable expression in our transcriptomic data ($n = 851$ out of 1054 genes, Fisher's exact test for enrichment $P = 4.19 \times 10^{-4}$, odds ratio = 1.18), and we identified significant enhancer-gene linkages for 54% of them ($n = 570$, $P = 3.51 \times 10^{-5}$, odds ratio = 1.25) (fig. S10B) (*30*). Consistent with our analysis of GWAS data, SFARI genes were significantly enriched for enhancer linkages in clusters GP and LN2 (hypergeometric test, $P = 2.24 \times 10^{-2}$ and 4.36×10^{-5}). In addition, we uncovered an enrichment in pallial neurons (PN1 and PN2 hypergeometric test, $P = 8.82 \times 10^{-3}$ and 4.36×10^{-4}) (Fig. 3E). The SFARI database includes genes with evidence from not only common variants but also de novo and rare syndromic and nonsyndromic mutations. This association with pallial neurons is further supported by the enrichment of recently identified de novo noncoding ASD mutations (*38*) within the PN2 cluster ($P = 0.027$) (Fig. 3F and fig. S10C). Last, we found that SFARI gene-linked enhancers were significantly enriched for MEIS3 homeobox transcription factor (TF) motifs, both relative to other gene linkages (Fig. 3G) and to unlinked enhancers at SFARI loci (fig. S10D).

Evolutionary conservation of epigenetic information in human forebrain

We investigated how evolutionary conservation might vary across our data, and we found that PhyloP conservation scores were not evenly distributed in our clusters, with PL and MG1 having the lowest PhyloP scores (indicating less conservation) and CS having the highest (fig. S11A) (*30*). Low conservation in MG1 could reflect morphological and phenotypic differ-

ences that have been observed between astrocytes in humans compared with other vertebrates (*39*). Peaks with an enhancer-gene linkage in the GP, PN1, and LN1 clusters demonstrated significantly reduced conservation scores relative to peaks without linkage, potentially reflecting a degree of evolutionary adaptation at these forebrain-specific loci contributing to gene expression (Mann-Whitney-Wilcoxon test with Bonferroni correction, $P_{adj} = 1.9 \times 10^{-2}$, 8.5×10^{-7}, and 1.6×10^{-3}, respectively). Although these differences suggest that a degree of selection may be acting on particular brain functions, they were quantitatively small. Therefore, we looked for evidence of process-specific selection, which might support the functional relevance of these observations. To do so, we defined gene-linked peaks in the top and bottom 5% of conservation scores as highly conserved and poorly conserved sets, respectively. We detected enrichment for forebrain differentiation, brain development, and transcriptional regulation GO terms among genes with highly conserved peak links relative to a background of genes with poorly conserved links (fig. S11B) (*30*). Last, we filtered genes to those with multiple highly conserved or poorly conserved linked enhancers and measured the distributions of these enhancers across our K-means clusters, relative to a permuted background (*30*). Poorly conserved enhancers were strongly enriched in cluster PL, and well-conserved enhancers were enriched in clusters PN1 and PN2 (fig. S11C).

Conservation can also be measured experimentally in vivo with enhancer-driven reporters. We quantified the overlap of our data with the VISTA mouse enhancer dataset (*40*) and observed that the VISTA enhancers that were accessible in our dynamic peaks were enriched for forebrain activity (odds ratio = 1.37, $P = 7.83 \times 10^{-6}$) (fig. S11D). Clusters MG2, GP, EA2, PN1, and CS were significantly enriched for forebrain-only activity (Benjamini-Hochberg adjusted $P = 7.13 \times 10^{-5}$, 6.20×10^{-4}, 4.78×10^{-2}, 1.64×10^{-3}, and 1.76×10^{-9}) (fig. S11E). Selected VISTA enhancers from PN1 and LN1—which include hSS enhancers—displayed activity in dorsal and ventral forebrain, respectively (fig. S11F) (*30*).

Last, we were interested in exploring how sequences that may have recently evolved in the human lineage are distributed during corticogenesis. To do this, we considered the overlap between our forebrain chromatin accessibility dataset and human accelerated regions (hARs), which are noncoding genomic loci, previously implicated in brain development, that exhibit high vertebrate conservation and increased substitution rates in humans compared with other great apes (*41, 42*). We observed that 35% of all hARs ($n = 965$ out of 2734) overlapped with our ATAC-seq data, corresponding to a rate of 3.8 hARs per thousand

dynamic peaks (fig. S11G and table S6) (*30*). These overlaps were broadly distributed across clusters, with the lowest enrichments in PL, and the highest in LN2, corresponding to late stages of differentiation (fig. S11H). Subpallial neuron peaks also exhibited overlap with hARs (4.05 to 4.34 hARs per megabase), including loci we linked to *GAD1*, *DLX2*, and the axon guidance ephrin receptor *EPHA5*. Recent work has provided evidence that a number of these loci may have functional consequences for developmental timing in humans (*19*). Together, our analyses uncovered lineage-specific patterns of functional conservation and acceleration across our chromatin accessibility landscape.

Transcription factor motifs in human forebrain development

We next explored the sequence features that underlie ATAC-seq clusters. We calculated the enrichment of DNA sequence motifs in each cluster versus all the other clusters (Fig. 4A and table S7) (*30*). We found that ZIC1, SP1, KLF5, and CTCF motifs were enriched in clusters PL and EA2, which is consistent with DNA binding domains shown to regulate pluripotency (*43*). Cluster GP was enriched for SOX10 and NFIC motifs, as well as a group of AT-rich sequence motifs bound by homeobox factors such as LHX and PAX, which play key roles in early cortical specification (*44*). The MG1 and MG2 clusters exhibited strong enrichment for STAT1::STAT2, SOX10, and the FOS family of motifs, which may play roles in astrocyte differentiation and maturation (*45*).

Sequence-based TF motif analysis does not provide a one-to-one correspondence between binding motifs and particular TF proteins (*46, 47*). To nominate specific TFs that may drive chromatin changes in an unbiased way, we used chromVAR to quantify the accessibility motifs across all samples and dynamic peaks (*48*) then computed sample-wise Pearson correlations between chromVAR motif deviations and log-transformed gene expression values of TFs from RNA-seq (Fig. 4B and table S8) (*30*). We highlighted TF genes with expression that were significantly correlated or anticorrelated to dynamic cluster-specific motif deviations (Fig. 4C). For each motif, we then plotted the chromVAR deviations alongside the RNA-seq profiles of each putative TF. Last, we verified the expression and localization of some of these factors in hCSs and HFT using immunohistochemistry (Fig. 4D).

In the glial lineage, we found that nuclear factor I A (NFIA) motif accessibility was strongly correlated with the expression of the *NFIA* gene (Pearson's $r = 0.77$). NFIA protein expression was widespread in hCSs (Fig. 4D, top row, and fig. S12, A to D) and HFT (fig. S12E), which is consistent with its RNA-seq expression and motif accessibility patterns. More specifically, NFIA was expressed in progenitors, including
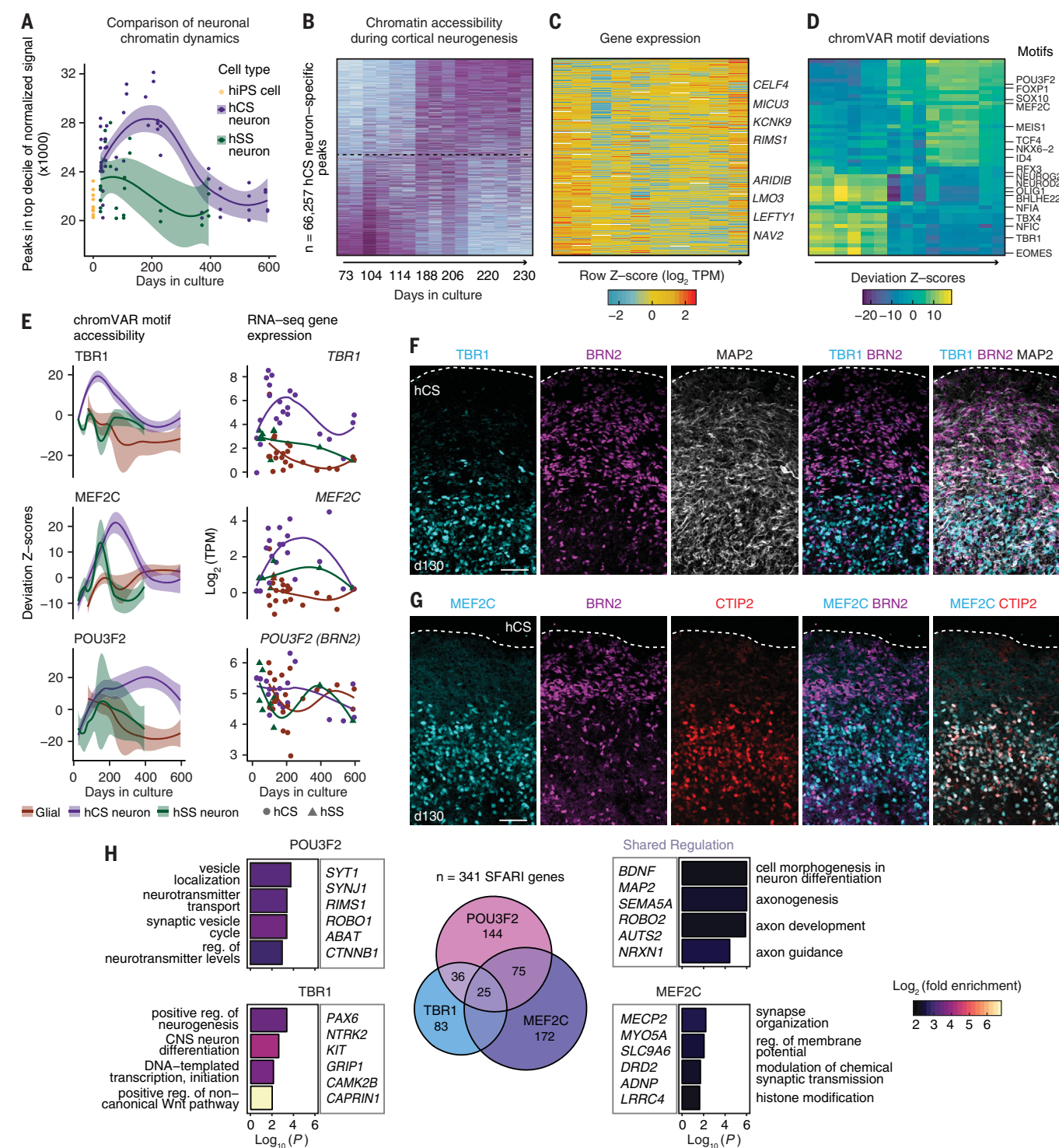
Fig. 5. A wave of chromatin accessibility and coordinated gene regulation during cortical neurogenesis. (**A**) Comparison of highly active open chromatin regions in neuronal lineages. The number of peaks in the top decile of ATAC-seq signal is shown. (**B**) Heatmap of hCS neuron peaks (DESeq2, FDR-adjusted $P < 0.01$) over 79 to 230 days. Peaks are sorted by the day-weighted mean of accessibility across these time points. Signal is displayed as row-scaled normalized accessibility. (**C**) Gene expression patterns of genes significantly linked to peaks in (B), displayed as row-scaled $\log_2$ (TPM). Select genes are shown. (**D**) ChromVAR deviations in hCSs. Row-scaled deviation Z-scores for each motif are plotted and sorted by the day-weighted mean of the deviations. (**E**) ChromVAR deviations and RNA-seq expression values for TBR1, MEF2C, and POU3F2 (BRN2). (**F**) Immunohistochemistry of hCS in cryosections at day 130 showing expression of the layer-specific markers TBR1 and BRN2 (or POU3F2) with MAP2. (**G**) Immunohistochemistry hCS at day 130 showing expression of MEF2C, BRN2, and CTIP2. (**H**) For each TF motif, the number of genes from the SFARI database with a linked enhancer having that motif is shown. Euler diagram indicates genes with binding motifs for more than one of the indicated TFs in linked enhancers. GO enrichments for groups of these SFARI genes are shown. "Shared regulation" indicates genes with two or more motifs. Scale bars, 50 µm (F) and (G).

HOPX$^+$ and SOX9$^+$ cells in hCSs and HFT, as well as in neurons, as shown by colocalization with CTIP2. The NFIA motif was also correlated to *NFIB*, which had similar protein expression patterns (fig. S12, F to H) (*30*). NEUROD2, which was strongly enriched in PN1 and PN2 clusters, was anticorrelated with *ID4* (r = –0.66), which also has a basic helix-loop-helix (bHLH) motif. Both *ID3* and *ID4* RNA expression increased over time in glial cells, and ID4 protein was localized to the nuclei of HOPX$^+$ cells but not CTIP2$^+$ neurons in hCSs (Fig. 4D, second row). In HFT, ID4 protein expression colocalized with HOPX$^+$ nuclei in the outer subventricular zone but not with CTIP2$^+$ cells (fig. S12I). ID-family TFs have been shown to dimerize with and repress other bHLH-family proteins (*49*), and Id4$^{-/-}$ mice displayed altered neural and glial differentiation (*50*). Thus, ID4 may act to maintain the radial glia population. The SOX10 motif was correlated with *SOX2*, *SOX9*, and most strongly *SOX21* expression (r = 0.73), and the latter increased in glial cells over time. We found protein expression of SOX21 in GFAP-expressing radial glia in both hCSs and HFT (fig. S13, A and B). Consistent with our genomic data, we found nuclear SOX21 expression in mature astrocytes at later stages in hCSs (200 and 552 days) (Fig. 4D, third row). In examining motifs specific to subpallial neurons (fig. S8A), we found that *ONECUT2*, *NKX2-1*, and *RFX4* were highly correlated to ONECUT1, NKX2-8, and RFX2 motif deviations (Fig. 4C). ONECUT2 was expressed in the nucleus of a subset of hSS cells but not hCS cells (Fig. 4D, bottom row, and fig. S14, A and B). ZIC1 was similarly expressed in the nuclei of hSS cells, localized with GABA, and was expressed to a much lesser degree in hCS cells (fig. S14, C to E). Last, RFX4 expression was also higher in hSSs and localized with GAD67 expression (fig. S14, F to I) (*30*). Together, these examples illustrate how this data can be interrogated to identify TFs that may regulate lineage specification in the forebrain.

To explore the diversity of sequence motifs in each of our dynamic clusters, we grouped enriched motifs (log$_2$ fold enrichment > 0.6) according to their JASPAR family assignments and computed the Shannon entropy of these labels by using the total number of enriched motifs as a prior estimate (fig. S15). The PL cluster had the highest diversity by this metric, followed by the PN2 cluster (H_{PL} = 2.61 and H_{PN2} = 2.39), whereas MG2 and GP had the lowest (H_{MG2} = 0.47 and H_{GP} = 0.66). As described above, GP accessibility was primarily associated with homeobox TFs. Meanwhile, the PN2 cluster exhibited a number of diverse yet highly enriched families, including bHLH, T-box, natural killer (NK)–related, and MADS box factors (Fig. 4E). Although this analysis does not capture accessibility contributions by weakly enriched factors, it suggests that the

period of cortical neurogenesis corresponding to 79 to 230 days is associated with a greater diversity of active TFs, prompting us to explore this stage further.

Gene-regulatory dynamics during human cortical neurogenesis

Upon closer inspection of the neuronal lineages, we observed more extensive chromatin remodeling in hCS neuronal lineages when compared with hSS interneurons, particularly between 79 and 230 days (Fig. 5A and table S9) (*30*). This could reflect the distinct trajectory of cortical interneurons, which require not just fate specification but also migration and integration into circuits (*51*). Migration and functional integration stages are not captured in our model without assembly. A differential test revealed that 26% of all dynamic peaks (n = 66,257) were specific to hCS and HFT neurons in this time window (FDR-adjusted P < 0.01) and that this wave of differential neuronal accessibility overlapped as expected with clusters PN1 and PN2 (J = 0.53). We then sorted hCS neuron–specific peaks by their ATAC-seq signal over time and observed dynamic changes (Fig. 5B).

Corticogenesis involves the sequential generation of layer-specific neurons over ~20 weeks in utero. To investigate the gene regulation underlying this process, we applied the analytical tools described above. Using our linked enhancer framework, we selected the strongest associations with the chromatin remodeling wave between 79 and 230 days and similarly sorted and plotted their expression in neurons over time (Fig. 5C); *NAV2*, *LEFTY1*, and the transcriptional cofactors *LMO3* and *ARID1B* were among the genes expressed at earlier stages, whereas *KCNK9*, *MICU3*, and *RIMS1* were found at later stages. We next selected TF motifs that were significantly enriched in clusters PN1 or PN2 and sorted their chromVAR deviations by the same metric, revealing concomitant progressions in motif accessibility over time (Fig. 5D); EOMES, TBR1, and NEUROD2 motifs represented the earliest phase; TCF4 and MEF2C represented a middle phase; and FOXP1, CUX2, and POU3F2 (BRN2) motifs represented the latest stages. The chromVAR deviations for many of these motifs also had strong correlations to the expression of at least one cognate TF (Fig. 5D).

We next identified chromatin accessibility linked to *TBR1* (TBR1 motif), which is known to be expressed in early-born deep-layer neurons, and accessibility linked to *POU3F2* (POU3F2 motif), which is expressed in late-born, superficial-layer neurons (*52*) and which has previously been shown to regulate gene expression networks related to the pathogenesis of schizophrenia and bipolar disorder (*53*). The expression of these transcripts peaked in hCS neurons between 93 and 114 days

for *TBR1* and 220 and 250 days for *BRN2*, and immunohistochemistry in hCSs and HFT confirmed this pattern of protein expression (Fig. 5, E and F). Our analysis highlighted MEF2C as a key TF motif associated with accessibility at the middle stage of the corticogenesis accessibility wave (with r = 0.83 correlation with the *MEF2C* gene). *MEF2C* has been previously associated with ASD (*54*), but its gene regulatory activity in the context of human cortical development has not been described. Consistent with our genomic measurements, we found that MEF2C was not expressed in progenitor cells or CTIP2$^+$ neurons at 75 days (fig. S16A). However, MEF2C partially colocalized with CTIP2 and RORB in hCSs after 130 days (Fig. 5G and fig. S16, B and C) as well as in HFT (fig. S16, D and E). These convergent lines of evidence suggest an important role for MEF2C in midcortical neurogenesis.

Because the *MEF2C* gene as well as the *TBR1* and *BRN2* genes have been linked to ASD (*55*, *56*), we next explored predicted downstream pathways and cell functions related to this association in human corticogenesis. Approximately 32% (n = 341 genes) of SFARI ASD genes had a linkage to an element containing one of these three motifs. More specifically, TBR1-specific gene linkages included the TF *PAX6* and calcium/calmodulin–dependent kinase *CAMK2B*, and overall, these genes were enriched for processes related to transcription regulation and cell signaling (relative to a background of variable genes in our forebrain data); MEF2C-specific gene targets included the Rett syndrome–related *MECP2* and the dopamine receptor *DRD2* and were enriched for synaptic organization as well as histone modification; BRN2-specific genes included *SYT1* and *SYTJ1* and were enriched for synaptic activity (Fig. 5H). Last, when we looked at the genes targeted by two or more of these three motifs, we discovered that they converged on a signature of axonogenesis and chemotaxis that included *BDNF*, *SEMA5A*, and *ROBO2* (Fig. 5H). Overall, these examples illustrate how this platform can be interrogated to identify factors that regulate development and disease risk during periods of human forebrain development that are inaccessible.

Discussion

The assembly of the human forebrain is a lengthy developmental process that, in primates, extends into postnatal life (*1*, *2*, *57*). Development can be disrupted by genetic and environmental factors, leading to disease. Our molecular understanding of forebrain development comes mostly from animal models. Recently, transcriptomic and epigenetic methods have been applied in human brain tissue (*18*, *20*, *21*, *58*). However, critical developmental time periods corresponding to mid- and

late gestation are still poorly characterized owing to tissue availability. We have used 3D organoids that can be maintained over long time periods to study chromatin accessibility dynamics during human forebrain development in specific cell lineages. Direct comparisons with human tissue validate that in vivo forebrain regulatory programs broadly map onto those in long-term cultures. Our accessibility landscape extends across cell lineages, forebrain domains, and time and reveals lineage-specificity of disease risk and conservation. We identified TFs associated with astrocyte maturation and interneuron specification and discovered a protracted period of chromatin remodeling during human cortical neurogenesis. Last, we have described how several key TFs may coordinate over time to regulate cellular functions as pallial neurons develop.

In contrast to expressed genes, regulatory elements are more numerous and cell lineage–restricted, and their sequences can provide clues about the gene programs that drive specification (6). By combining epigenetic information with transcriptomic and protein measurements, we nominate TFs with substantial regulatory potential in forebrain cell lineages. We have generated examples of how the data can be used to elucidate the regulation of forebrain lineages. We anticipate that future studies could leverage our data to experimentally demonstrate the detailed mechanisms that govern the production of pallial neurons versus glial cells, the specification of interneurons, and cortical cellular maturation.

Efforts to link disease risk to epigenetic or transcriptomic trajectories have primarily used dissected brain tissue at early stages of development or later postnatal stages (17, 20, 21, 37, 59, 60). By capturing both human neurogenesis and gliogenesis in vitro, we have further mapped genetic ASD risk to glial progenitor cells as well as mid- and late-stage neurons. Moreover, we have defined a set of ASD genes enriched for chemotaxis and axonogenesis that appear to be regulated by key TFs. Single-cell epigenomic assays, which have recently become tractable and scalable, will be especially suited to resolving how TFs coordinate across cells to achieve cortical specification and disease gene regulation.

The brain region–specific organoid platform holds the promise of flexibly modeling the development of other diverse brain regions, which are also currently inaccessible to molecular study, and modeling patient genetic backgrounds. Patient-derived organoids will allow for genetic and pharmacological manipulation of phenotypes to identify disease mechanisms, and we anticipate that comparison of epigenetic trajectories will provide further insights into pathophysiology.

Last, our data indicate that organoids intrinsically undergo chromatin state transitions in vitro that are closely related to human forebrain development in vivo. However, a full understanding of brain development will require the study of not only these intrinsic programs but also extrinsic cues that influence development and disease. The assembly of organoids from multiple brain regions, in combination with genetic or optogenetic tools, could be used to comprehensively assess the impact of connectivity and activity on development and maturation.

Methods

Detailed materials and methods can be found in the supplementary materials.

Generation of hCS and hSS from hiPS cells

Intact hiPS cell colonies maintained on mouse embryonic fibroblast feeders or feeder-free cells were enzymatically lifted before being transferred into ultralow-attachment plates in hiPS cell medium supplemented with the SMAD inhibitors dorsomorphin (5 μM) and SB-431542 (10 μM), as described (25–27). On the 6th day in suspension, the medium was switched to neural medium supplemented with epidermal growth factor (EGF) (20 ng/ml) and fibroblast growth factor 2 (FGF-2) (20 ng/ml). The neural medium was changed every day until day 17 and then every other day until day 24. For the generation of hSSs, the medium was additionally supplemented with inhibitor of Wnt production 2 (IWP-2) (5 μM) on days 4 to 24, smoothened agonist (SAG) (100 nM) on days 12 to 24, retinoic acid (RA) (100 nM) on days 12 to 15, and allopregnanolone (100 nM) on days 15 to 24. From day 25 to day 43, the neural medium for both hCSs and hSSs was changed every other day and supplemented with brain-derived neurotrophic factor (BDNF) (20 ng/ml) and neutrophin-3 (NT3) (20 ng/ml). From day 45 onwards, hCSs and hSSs were maintained in neural medium without growth factors with medium changes every 4 to 6 days.

Human tissue

Human brain tissue was obtained under a protocol approved by the Research Compliance Office at Stanford University. PCW20 and PCW21 forebrain tissue was delivered overnight on ice and immediately processed after arrival.

Dissociation, immunopanning, and cell sorting

Dissociation of neural spheroids and human tissue into single cells was performed as described (24–26, 29). Tissue was chopped and incubated in 40 U/ml papain enzyme solution at 37°C for 90 min then gently triturated to achieve a single-cell suspension. Single-cell suspensions were then either FACS-sorted or immunopanned to achieve astrocyte- or neuron-enriched populations. For FACS-sorting, cells separated on the basis of their expression of either GFP (pLV-GFAP::eGFP glia) or mCherry (AAV-DJ-hSYN1::mCherry neurons). For immunopanning, the single-cell suspension was added to a series of plastic petri dishes precoated with either antibody to against HepaCAM (glia) or antibody against Thy1 (neurons) and incubated for 10 to 30 min at room temperature. Bound cells were incubated in a trypsin solution (for glia) or Accutase (for neurons) at 37°C for 3 to 5 min then gently pipetted off the plates and processed for RNA-seq and ATAC-seq.

ATAC-seq processing

ATAC-seq was performed by resuspending cells in lysis buffer with Tween-20, followed by centrifugation at 4°C for 10 min and incubation with transposase at 37°C for 30 min. Sequencing was performed on an Illumina NextSeq. Data were processed against the GRCh38/hg38 reference genome. ATAC-seq data were aligned by using Bowtie2. Peaks were called by using MACS2, filtered to q value < 0.01, then merged within biological replicate groups and tiled to a width of 500 base pairs. Tiles were evaluated for variability and signal strength across samples and removed if they failed these metrics. Last, counts were generated for each peak set by using multiBamCov, cleaned with edgeR, and quantile normalized.

RNA-seq processing

RNA was isolated from single-cell suspensions and homogenized tissue by use of TRIzol, purified with direct-zol RNA MicroPrep kits, and processed with an Ovation NuGen RNA-seq kit. The quality of the libraries was assessed with BioAnalyzer before sequencing on an Illumina NextSeq. Reads were trimmed by using Skewer then pseudo-aligned to the RefSeq transcriptome with Kallisto to obtain per-transcript TPM values, as well as count estimates. Read depth, alignment rate, and RNA integrity number (RIN) score were evaluated, and samples were kept with RIN > 7.5. To correct for batch effects across samples collected at different times, we used the package "limma" on the matrix of log-transformed TPM values: "removeBatchEffect(log2(TPM + 1))".

REFERENCES AND NOTES

1. J. C. Silbereis, S. Pochareddy, Y. Zhu, M. Li, N. Sestan, The cellular and molecular landscapes of the developing human central nervous system. *Neuron* **89**, 248–268 (2016). doi: 10.1016/j.neuron.2015.12.008; pmid: 26796689
2. J. L. R. Rubenstein, Annual Research Review: Development of the cerebral cortex: implications for neurodevelopmental disorders. *J. Child Psychol. Psychiatry* **52**, 339–355 (2011). doi: 10.1111/j.1469-7610.2010.02307.x; pmid: 20735793
3. C. P. Wonders, S. A. Anderson, The origin and specification of cortical interneurons. *Nat. Rev. Neurosci.* **7**, 687–696 (2006). doi: 10.1038/nrn1954; pmid: 16883309
4. G. Juric-Sekhar, R. F. Hevner, Malformations of cerebral cortex development: Molecules and mechanisms. *Annu. Rev. Pathol.*

14, 293–318 (2019). doi: 10.1146/annurev-pathmechdis-012418-012927; pmid: 30677308

5. J. D. Buenrostro et al., Integrated single-cell analysis maps the continuous regulatory landscape of human hematopoietic differentiation. Cell 173, 1535–1548.e16 (2018). doi: 10.1016/j.cell.2018.03.074; pmid: 29706549

6. A. B. Stergachis et al., Developmental fate and cellular maturity encoded in human regulatory DNA landscapes. Cell 154, 888–903 (2013). doi: 10.1016/j.cell.2013.07.020; pmid: 23953118

7. ENCODE Project Consortium, An integrated encyclopedia of DNA elements in the human genome. Nature 489, 57–74 (2012). doi: 10.1038/nature11247; pmid: 22955616

8. F. Grubert et al., Genetic control of chromatin states in humans involves local and distal chromosomal interactions. Cell 162, 1051–1065 (2015). doi: 10.1016/j.cell.2015.07.048; pmid: 26300125

9. E. Z. Kvon et al., Progressive loss of function in a limb enhancer during snake evolution. Cell 167, 633–642.e11 (2016). doi: 10.1016/j.cell.2016.09.028; pmid: 27768887

10. S. L. Klemm, Z. Shipony, W. J. Greenleaf, Chromatin accessibility and the regulatory epigenome. Nat. Rev. Genet. 20, 207–220 (2019). doi: 10.1038/s41576-018-0089-8; pmid: 30675018

11. H. J. Kang et al., Spatio-temporal transcriptome of the human brain. Nature 478, 483–489 (2011). doi: 10.1038/nature10523; pmid: 22031440

12. M. Li et al., Integrative functional genomic analysis of human brain development and neuropsychiatric risks. Science 362, eaat7615 (2018). doi: 10.1126/science.aat7615; pmid: 30545854

13. J. A. Miller et al., Transcriptional landscape of the prenatal human brain. Nature 508, 199–206 (2014). doi: 10.1038/nature13185; pmid: 24695229

14. T. J. Nowakowski et al., Spatiotemporal gene expression trajectories reveal developmental hierarchies of the human cortex. Science 358, 1318–1323 (2017). doi: 10.1126/science.aap8809; pmid: 29217575

15. M. B. Johnson et al., Single-cell analysis reveals transcriptional heterogeneity of neural progenitors in human cortex. Nat. Neurosci. 18, 637–646 (2015). doi: 10.1038/nn.3980; pmid: 25734491

16. A. A. Pollen et al., Molecular identity of human outer radial glia during cortical development. Cell 163, 55–67 (2015). doi: 10.1016/j.cell.2015.09.004; pmid: 26406371

17. A. Amiri et al., Transcriptome and epigenome landscape of human cortical development modeled in organoids. Science 362, eaat6720 (2018). doi: 10.1126/science.aat6720; pmid: 30545853

18. A. E. Jaffe et al., Mapping DNA methylation across development, genotype and schizophrenia in the human frontal cortex. Nat. Neurosci. 19, 40–47 (2016). doi: 10.1038/nn.4181; pmid: 26619358

19. S. Kanton et al., Organoid single-cell genomic atlas uncovers human-specific features of brain development. Nature 574, 418–422 (2019). doi: 10.1038/s41586-019-1654-9; pmid: 31619793

20. L. de la Torre-Ubieta et al., The dynamic landscape of open chromatin during human cortical neurogenesis. Cell 172, 289–304.e18 (2018). doi: 10.1016/j.cell.2017.12.014; pmid: 29307494

21. H. Won et al., Chromosome conformation elucidates regulatory relationships in developing human brain. Nature 538, 523–527 (2016). doi: 10.1038/nature19847; pmid: 27760116

22. S. P. Paşca, The rise of three-dimensional human brain cultures. Nature 553, 437–445 (2018). doi: 10.1038/nature25032; pmid: 29364288

23. A. M. Paşca et al., Functional cortical neurons and astrocytes from human pluripotent stem cells in 3D culture. Nat. Methods 12, 671–678 (2015). doi: 10.1038/nmeth.3415; pmid: 26005811

24. S. A. Sloan et al., Human astrocyte maturation captured in 3D cerebral cortical spheroids derived from pluripotent stem cells. Neuron 95, 779–790.e6 (2017). doi: 10.1016/j.neuron.2017.07.035; pmid: 28817799

25. F. Birey et al., Assembly of functionally integrated human forebrain spheroids. Nature 545, 54–59 (2017). doi: 10.1038/nature22330; pmid: 28445465

26. S.-J. Yoon et al., Reliability of human cortical organoid generation. Nat. Methods 16, 75–78 (2019). doi: 10.1038/s41592-018-0255-0; pmid: 30573846

27. S. A. Sloan, J. Andersen, A. M. Paşca, F. Birey, S. P. Paşca, Generation and assembly of human brain region-specific three-dimensional cultures. Nat. Protoc. 13, 2062–2085 (2018). doi: 10.1038/s41596-018-0032-7; pmid: 30202107

28. M. R. Corces et al., An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. Nat. Methods 14, 959–962 (2017). doi: 10.1038/nmeth.4396; pmid: 28846090

29. Y. Zhang et al., Purification and characterization of progenitor and mature human astrocytes reveals transcriptional and functional differences with mouse. Neuron 89, 37–53 (2016). doi: 10.1016/j.neuron.2015.11.013; pmid: 26687838

30. Materials and methods are available as supplementary materials.

31. H. A. Pliner et al., Cicero predicts cis-regulatory dna interactions from single-cell chromatin accessibility data. Mol. Cell 71, 858–871.e8 (2018). doi: 10.1016/j.molcel.2018.06.044; pmid: 30078726

32. A. Kundaje et al., Integrative analysis of 111 reference human epigenomes. Nature 518, 317–330 (2015). doi: 10.1038/nature14248; pmid: 25693563

33. M. R. Corces et al., The chromatin accessibility landscape of primary human cancers. Science 362, eaav1898 (2018). doi: 10.1126/science.aav1898; pmid: 30361341

34. B. Bulik-Sullivan et al., An atlas of genetic correlations across human diseases and traits. Nat. Genet. 47, 1236–1241 (2015). doi: 10.1038/ng.3406; pmid: 26414676

35. H. K. Finucane et al., Partitioning heritability by functional annotation using genome-wide association summary statistics. Nat. Genet. 47, 1228–1235 (2015). doi: 10.1038/ng.3404; pmid: 26414678

36. Schizophrenia Working Group of the Psychiatric Genomics Consortium, Biological insights from 108 schizophrenia-associated genetic loci. Nature 511, 421–427 (2014). doi: 10.1038/nature13595; pmid: 25056061

37. N. N. Parikshak et al., Integrative functional genomic analyses implicate specific molecular pathways and circuits in autism. Cell 155, 1008–1021 (2013). doi: 10.1016/j.cell.2013.10.031; pmid: 24267887

38. J. Zhou et al., Whole-genome deep-learning analysis identifies contribution of noncoding mutations to autism risk. Nat. Genet. 51, 973–980 (2019). doi: 10.1038/s41588-019-0420-0; pmid: 31133750

39. N. A. Oberheim et al., Uniquely hominid features of adult human astrocytes. J. Neurosci. 29, 3276–3287 (2009). doi: 10.1523/JNEUROSCI.4707-08.2009; pmid: 19279265

40. A. Visel, S. Minovitsky, I. Dubchak, L. A. Pennacchio, VISTA Enhancer Browser—A database of tissue-specific human enhancers. Nucleic Acids Res. 35 (Database), D88–D92 (2007). doi: 10.1093/nar/gkl822; pmid: 17130149

41. J. A. Capra, G. D. Erwin, G. McKinsey, J. L. R. Rubenstein, K. S. Pollard, Many human accelerated regions are developmental enhancers. Philos. Trans. R. Soc. Lond. B Biol. Sci. 368, 20130025 (2013). doi: 10.1098/rstb.2013.0025; pmid: 24218637

42. K. S. Pollard et al., Forces shaping the fastest evolving regions in the human genome. PLOS Genet. 2, e168 (2006). doi: 10.1371/journal.pgen.0020168; pmid: 17040131

43. S. K. Balakrishnan, M. Witcher, T. W. Berggren, B. M. Emerson, Functional and molecular characterization of the role of CTCF in human embryonic stem cell biology. PLOS ONE 7, e42424 (2012). doi: 10.1371/journal.pone.0042424; pmid: 22879976

44. V. S. Mangale et al., Lhx2 selector activity specifies cortical identity and suppresses hippocampal organizer fate. Science 319, 304–309 (2008). doi: 10.1126/science.1151695; pmid: 18202285

45. K. Hisanaga, S. M. Sagar, K. J. Hicks, R. A. Swanson, F. R. Sharp, c-fos proto-oncogene expression in astrocytes associated with differentiation or proliferation but not depolarization. Brain Res. Mol. Brain Res. 8, 69–75 (1990). doi: 10.1016/0169-328X(90)90011-2; pmid: 2166202

46. A. Jolma et al., DNA-binding specificities of human transcription factors. Cell 152, 327–339 (2013). doi: 10.1016/j.cell.2012.12.009; pmid: 23332764

47. A. Jolma et al., DNA-dependent formation of transcription factor pairs alters their binding specificity. Nature 527, 384–388 (2015). doi: 10.1038/nature15518; pmid: 26550823

48. A. N. Schep, B. Wu, J. D. Buenrostro, W. J. Greenleaf, chromVAR: Inferring transcription-factor-associated accessibility from single-cell epigenomic data. Nat. Methods 14, 975–978 (2017). doi: 10.1038/nmeth.4401; pmid: 28825706

49. H. A. Sikder, M. K. Devlin, S. Dunlap, B. Ryu, R. M. Alani, Id proteins in cell growth and tumorigenesis. Cancer Cell 3, 525–530 (2003). doi: 10.1016/S1535-6108(03)00141-7; pmid: 12842081

50. L. Bedford et al., Id4 is required for the correct timing of neural differentiation. Dev. Biol. 280, 386–395 (2005). doi: 10.1016/j.ydbio.2005.02.001; pmid: 15882580

51. A. Kepecs, G. Fishell, Interneuron cell types are fit to function. Nature 505, 318–326 (2014). doi: 10.1038/nature12983; pmid: 24429630

52. H. Zeng et al., Large-scale cellular-resolution gene profiling in human neocortex reveals species-specific molecular signatures. Cell 149, 483–496 (2012). doi: 10.1016/j.cell.2012.02.052; pmid: 22500809

53. C. Chen et al., The transcription factor POU3F2 regulates a gene coexpression network in brain tissue from patients with psychiatric disorders. Sci. Transl. Med. 10, eaat8178 (2018). doi: 10.1126/scitranslmed.aat8178; pmid: 30545964

54. H. Li et al., Transcription factor MEF2C influences neural stem/progenitor cell differentiation and maturation in vivo. Proc. Natl. Acad. Sci. U.S.A. 105, 9397–9402 (2008). doi: 10.1073/pnas.0802876105; pmid: 18599437

55. J. H. Notwell et al., TBR1 regulates autism risk genes in the developing neocortex. Genome Res. 26, 1013–1022 (2016). doi: 10.1101/gr.203612.115; pmid: 27325115

56. N.-P. Tsai et al., Multiple autism-linked genes mediate synapse elimination via proteasomal degradation of a synaptic scaffold PSD-95. Cell 151, 1581–1594 (2012). doi: 10.1016/j.cell.2012.11.040; pmid: 23260144

57. B. Martynoga, D. Drechsel, F. Guillemot, Molecular control of neurogenesis: A view from the mammalian cerebral cortex. Cold Spring Harb. Perspect. Biol. 4, a008359 (2012). doi: 10.1101/cshperspect.a008359; pmid: 23028117

58. S. Preissl et al., Single-nucleus analysis of accessible chromatin in developing mouse forebrain reveals cell-type-specific transcriptional regulation. Nat. Neurosci. 21, 432–439 (2018). doi: 10.1038/s41593-018-0079-3; pmid: 29434377

59. D. Velmeshev et al., Single-cell genomics identifies cell type-specific molecular changes in autism. Science 364, 685–689 (2019). doi: 10.1126/science.aav8130; pmid: 31097668

60. A. J. Willsey et al., Coexpression networks implicate human midfetal deep cortical projection neurons in the pathogenesis of autism. Cell 155, 997–1007 (2013). doi: 10.1016/j.cell.2013.10.020; pmid: 24267886

Science

Chromatin accessibility dynamics in a model of human forebrain development

Alexandro E. Trevino, Nasa Sinnott-Armstrong, Jimena Andersen, Se-Jin Yoon, Nina Huber, Jonathan K. Pritchard, Howard Y. Chang, William J. Greenleaf and Sergiu P. Pasca

Organoids recapitulate brain development

Gene expression changes and their control by accessible chromatin in the human brain during development is of great interest but limited accessibility. Trevino *et al.* avoided this problem by developing three-dimensional organoid models of human forebrain development and examining chromatin accessibility and gene expression at the single-cell level. From this analysis, they matched developmental profiles between the organoid and fetal samples, identified transcription factor binding profiles, and predicted how transcription factors are linked to cortical development. The researchers were able to correlate the expression of neurodevelopmental disease risk loci and genes with specific cell types during development.

Science, this issue p. eaay1645

ARTICLE TOOLS	http://science.sciencemag.org/content/367/6476/eaay1645
SUPPLEMENTARY MATERIALS	http://science.sciencemag.org/content/suppl/2020/01/22/367.6476.eaay1645.DC1
REFERENCES	This article cites 76 articles, 12 of which you can access for free http://science.sciencemag.org/content/367/6476/eaay1645#BIBL
PERMISSIONS	http://www.sciencemag.org/help/reprints-and-permissions

Use of this article is subject to the Terms of Service