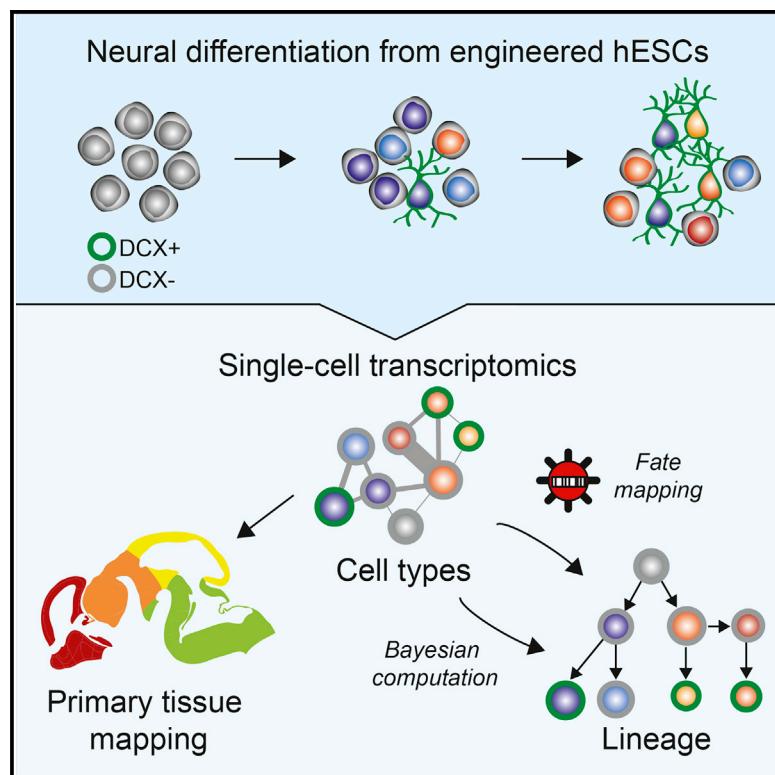


# A Single-Cell Roadmap of Lineage Bifurcation in Human ESC Models of Embryonic Brain Development

## Graphical Abstract



## Authors

Zizhen Yao, John K. Mich,  
Sherman Ku, ..., Boaz P. Levi,  
Yanling Wang, Sharad Ramanathan

## Correspondence

grimley@gmail.com (J.S.G.),  
boazl@alleninstitute.org (B.P.L.),  
ywseattle@gmail.com (Y.W.),  
sharad@post.harvard.edu (S.R.)

## In Brief

Yao et al. perform single-cell RNA-seq during neural differentiation of hESCs. They identify many classes of neural progenitors and neurons that map to early human brain cells, computationally infer and experimentally confirm lineage relationships between them, and show that Wnt signaling influences the bifurcation between forebrain and mid/hindbrain lineages in vitro.

## Highlights

- Single-cell RNA-seq of differentiating hESCs reveals a multitude of neural cell types
- hESC-derived cells exhibit a range of early brain regional identities
- Bayesian analyses reconstruct, and clonal analyses confirm, lineage relationships
- Wnt/β-catenin signaling controls bifurcations between forebrain and mid/hindbrain lineages

## Accession Numbers

GSE86894  
phs001205.v1.p1

# A Single-Cell Roadmap of Lineage Bifurcation in Human ESC Models of Embryonic Brain Development

Zizhen Yao,<sup>1,5</sup> John K. Mich,<sup>1,5</sup> Sherman Ku,<sup>1,5</sup> Vilas Menon,<sup>1,5</sup> Anne-Rachel Krostag,<sup>1</sup> Refugio A. Martinez,<sup>1</sup> Leon Furchtgott,<sup>2</sup> Heather Mulholland,<sup>1</sup> Susan Bort,<sup>1</sup> Margaret A. Fuqua,<sup>1</sup> Ben W. Gregor,<sup>1</sup> Rebecca D. Hodge,<sup>1</sup> Anu Jayabalu,<sup>1</sup> Ryan C. May,<sup>1</sup> Samuel Melton,<sup>3</sup> Angelique M. Nelson,<sup>1</sup> N. Kiet Ngo,<sup>1</sup> Nadiya V. Shapovalova,<sup>1</sup> Soraya I. Shehata,<sup>1</sup> Michael W. Smith,<sup>1</sup> Leah J. Tait,<sup>1</sup> Carol L. Thompson,<sup>1</sup> Elliot R. Thomsen,<sup>1</sup> Chaoyang Ye,<sup>1</sup> Ian A. Glass,<sup>4</sup> Ajamete Kaykas,<sup>1</sup> Shuyuan Yao,<sup>1</sup> John W. Phillips,<sup>1</sup> Joshua S. Grimley,<sup>1,\*</sup> Boaz P. Levi,<sup>1,6,\*</sup> Yanling Wang,<sup>1,\*</sup> and Sharad Ramanathan<sup>1,2,3,\*</sup>

<sup>1</sup>Allen Institute for Brain Science, Seattle, WA 98109, USA

<sup>2</sup>Molecular and Cellular Biology Department, Harvard University, Cambridge, MA 02138, USA

<sup>3</sup>School of Engineering and Applied Science, Harvard University, Cambridge, MA 02138, USA

<sup>4</sup>Division of Genetic Medicine, University of Washington, Seattle Children's Hospital, Seattle, WA 98105, USA

<sup>5</sup>Co-first author

<sup>6</sup>Lead Contact

\*Correspondence: [grimley@gmail.com](mailto:grimley@gmail.com) (J.S.G.), [boazl@alleninstitute.org](mailto:boazl@alleninstitute.org) (B.P.L.), [ywseattle@gmail.com](mailto:ywseattle@gmail.com) (Y.W.), [sharad@post.harvard.edu](mailto:sharad@post.harvard.edu) (S.R.)  
<http://dx.doi.org/10.1016/j.stem.2016.09.011>

## SUMMARY

During human brain development, multiple signaling pathways generate diverse cell types with varied regional identities. Here, we integrate single-cell RNA sequencing and clonal analyses to reveal lineage trees and molecular signals underlying early forebrain and mid/hindbrain cell differentiation from human embryonic stem cells (hESCs). Clustering single-cell transcriptomic data identified 41 distinct populations of progenitor, neuronal, and non-neuronal cells across our differentiation time course. Comparisons with primary mouse and human gene expression data demonstrated rostral and caudal progenitor and neuronal identities from early brain development. Bayesian analyses inferred a unified cell-type lineage tree that bifurcates between cortical and mid/hindbrain cell types. Two methods of clonal analyses confirmed these findings and further revealed the importance of Wnt/β-catenin signaling in controlling this lineage decision. Together, these findings provide a rich transcriptome-based lineage map for studying human brain development and modeling developmental disorders.

## INTRODUCTION

The human brain is a complex and highly evolved structure. Mouse models do not fully recapitulate cell-type diversity or lineage trajectories of the human brain (Florio et al., 2015; Konopka et al., 2012; Pollen et al., 2015; Reilly et al., 2015; Silbereis et al., 2016; Thomsen et al., 2016). Furthermore, human neurodevelopmental diseases such as autism spectrum disorders and schizophrenia are incompletely modeled in mouse. Stem cell-based models of human brain development have been pursued to un-

derstand and combat these disorders (Hook et al., 2014; Ricciardi et al., 2012).

The molecular networks that drive fate decisions and development of neurons and glia are not fully understood, and some may be unique in humans (Lui et al., 2014). Single-cell transcriptomics can reveal these networks with high-dimensional molecular characterization at an increasing scale (Klein et al., 2015; Macosko et al., 2015) and has already defined transcriptomic cell types from the blood (Paul et al., 2015), lung (Treutlein et al., 2014), and mouse brain (Tasic et al., 2016; Zeisel et al., 2015). In parallel, the recent progress in modeling human brain development from pluripotent stem cells (Chambers et al., 2009; Espuny-Camacho et al., 2013; Lancaster et al., 2013; Qian et al., 2016; Shi et al., 2012) promises to supply human neural tissue at developmental stages that are typically unavailable. Although several studies have characterized differentiated cells by gene expression (Edri et al., 2015; van de Leemput et al., 2014), only one *in vitro* differentiation study has carried out single-cell transcriptomics (Camp et al., 2015). As these cultures presumably contain a mixture of cell types, single-cell resolution studies are essential to characterize the cell types produced in culture and to determine how they compare to primary developing tissue.

Here, we present a study of early human brain cell-type development using single-cell transcriptomics and a two-dimensional *in vitro* brain development model. We demonstrate the biological relevance of our cell types by comparison to primary tissues from atlas data and cortical cells from mid-gestation human fetal embryos. We computationally identify cell types, predict their lineage relationships using a Bayesian lineage algorithm, and experimentally confirm the predictions using two independent clonal analysis techniques. Our lineage tree captures some of the earliest regional patterning events of the brain, including the canonical Wnt/β-catenin-signaling-dependent separation of cortical from mid/hindbrain cell types and the appearance of forebrain neurons that resemble some of the earliest formed neurons of the brain. These data constitute a deep and broad interrogation of human embryonic stem cell (hESC) neural differentiation and highlight key steps in regional patterning and lineage specification.

## RESULTS

### In Vitro Model of Human Brain Excitatory Cell Development

We developed and standardized an in vitro model of human cortical development based on the neuralization of hESCs, adapted from previous protocols (Chambers et al., 2009; Espuny-Camacho et al., 2013; Shi et al., 2012). The cortical induction (CI) phase utilizes SMAD inhibition (Chambers et al., 2009), the progenitor expansion (PE) phase includes EGF and FGF2, and the neural differentiation (ND) phase includes neurogenic and neurotrophic factors BDNF, GDNF, NT3, and cAMP (Hu et al., 2010) (Figure 1A). At the end of CI (day 12), pluripotency markers were completely lost, most cells expressed both PAX6 and FOXG1, and 92% ± 3% co-express PAX6 and SOX2, suggesting efficient telencephalic induction (Figures 1B and S1A–S1C). By the end of PE (day 26), 11% ± 2% of cells expressed the cortical intermediate progenitor marker EOMES (TBR2) (Figures 1B and S1D). Following ND (day 54), many cells expressed the neuronal marker MAP2 and subtype-specific markers TBR1, BCL11B (CTIP2), POU3F2 (BRN2), SATB2, and LHX2. In addition, we observed putative glial and outer radial glial cells marked by HOPX and GFAP (Figures 1B and S1E) (Pollien et al., 2015; Thomsen et al., 2016). Neuronal activity of day 54 cells was confirmed using calcium imaging and pharmacological blocking experiments with tetrodotoxin (TTX), an action potential generation inhibitor. Out of 1,148 recorded cells (three biological replicates), 17% ± 6% demonstrated calcium activity with a frequency of 2.2 ± 0.5 events/min. Calcium activity was blocked in 42.3% of those cells by TTX (Figure 1C). These observations are comparable to data obtained by similar methods in recent reports (Edri et al., 2015; Espuny-Camacho et al., 2013; Gaspard et al., 2008; Lancaster et al., 2013; Mariani et al., 2012; Shi et al., 2012).

We profiled the different stages of our in vitro differentiation protocol and established its reproducibility (Figures 1D and S1B–S1F) across replicate differentiations and cell lines (H1 and H9) by analyzing populations of cells from each time point (~1 × 10<sup>5</sup> to 1 × 10<sup>6</sup> cells per sample) using RNA sequencing (RNA-seq) analysis, flow cytometry, and immunostaining. Developing cortex markers *EMX2*, *PAX6*, and *LHX2* appeared between days 6 and 12 (Figures 1E, S1A, and S1G). Additionally, markers of ganglionic eminences (*DLX1*, *ASCL1*, and *GAD1*) as well as mid/hindbrain (*EN2*, *PAX7*, and *TFAP2B*) were observed. Analysis by flow cytometry showed 7% ± 3% of SOX2<sup>+</sup> cells lack PAX6 at day 26 (Figure S1A), suggesting non-cortical identity. Immunostaining corroborated the presence of SOX2<sup>+</sup>PAX6<sup>−</sup> cells at day 26 as well as the expression of TFAP2B, TH, GAD67, and PBX3 in cultured neurons at day 54 (Figures 1B and S1H). These results indicate a high degree of cellular heterogeneity and led us to transcriptionally profile single in vitro-derived cells through differentiation.

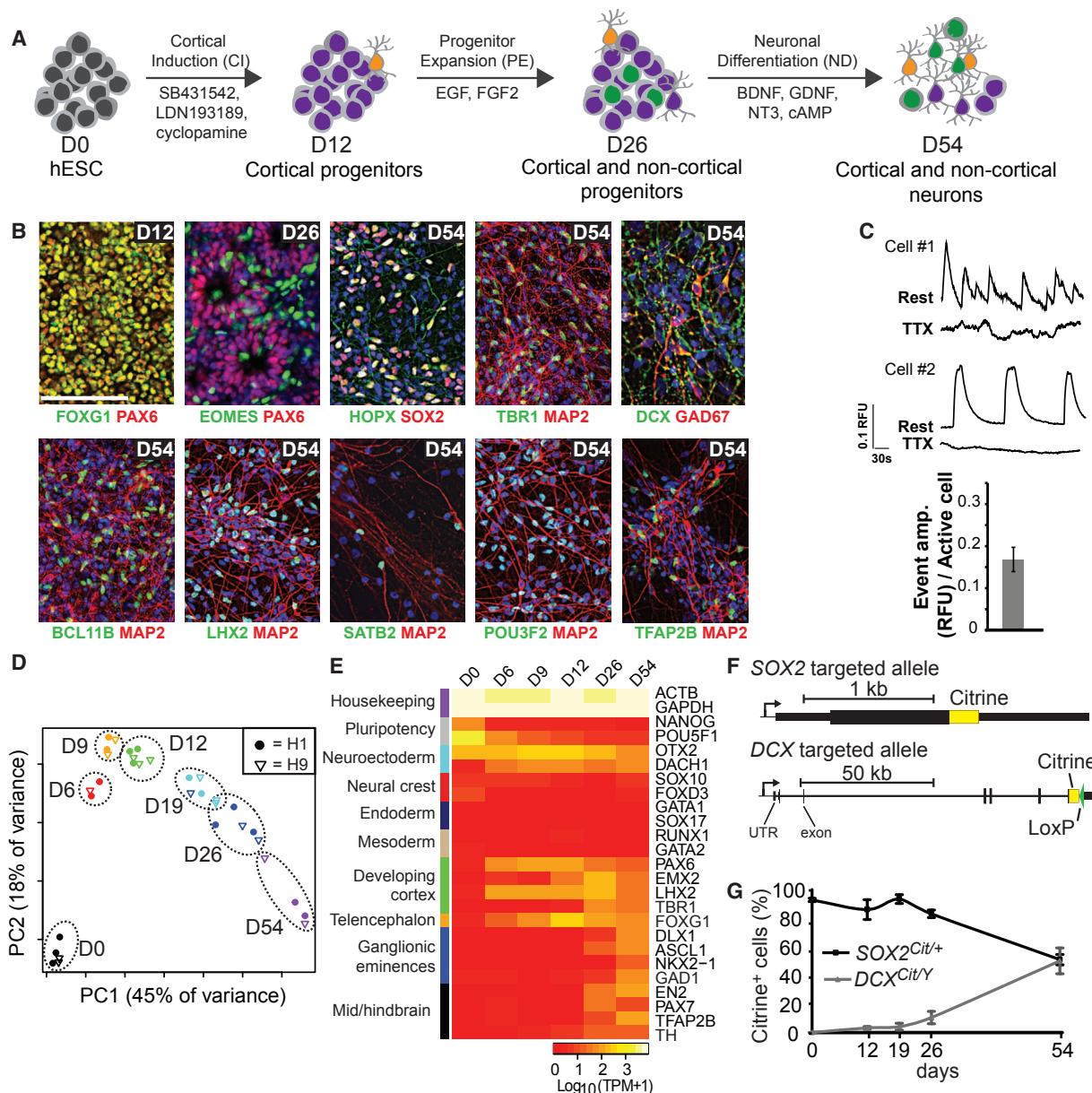
We engineered reporter lines using transcription activator-like effector nucleases (TALENs) (Miller et al., 2011) to fuse citrine fluorescent protein to endogenous SOX2 (a marker of progenitors) or DCX (a marker of immature neurons) in order to isolate both live progenitors and neurons (Figure 1F, S2A, and S2B). SOX2<sup>Cit/+</sup> cells exhibited near-uniform reporter expression in 93% ± 1% of hESCs (n = 3), which decreased to 51% ± 3%

(n = 3) citrine-positive cells by day 54 of differentiation (Figures 1G and S2C). In *DCX*<sup>Cit/Y</sup> cells, the citrine reporter was not detected in hESCs but increased to 50% ± 8% (n = 6) of cells at day 54. Both reporters mimicked expression of the endogenous protein, and produced neurons with the same markers as the parental H1 line (Figures S2D and S2E). Additionally, reporter lines were generated and validated for the early neurogenesis markers *OTX2* and *PAX6* (Figures S2F–S2K). These lines constitute an important set of tools for dissecting human neurogenesis.

### Single-Cell Profiling and Identification of Cell Types

A method based on multiplexed single-cell RNA-seq (CelSeq; Hashimshony et al., 2012) was used to profile cells at multiple time points (days 0, 12, 19, 26, 40, and 54) (Figure 2A). We isolated both progenitor-enriched (*DCX*<sup>Cit−</sup> and *SOX2*<sup>Cit−</sup>) and neuron-enriched cells (*DCX*<sup>Cit+</sup> and *SOX2*<sup>Cit−</sup>), resulting in 4,368 cells harvested from two to six independent differentiations per time point (Figures 2B and S3A). We recognized, as others have (Paul et al., 2015), some well-to-well mixing was present that could lead to spurious appearance of cell types or frequencies, and we implemented a strategy that substantially reduced such technical artifacts (Figure S3B; **Supplemental Experimental Procedures**). Following batch correction, we analyzed cells with >20,000 transcripts as assessed by unique molecular identifiers (UMIs) (n = 2,684) and normalized the data by subsampling to 20,000 UMIs (Figures S3C and S3D). Single-cell data were evaluated for mapping to genome and transcriptome, inclusion of UMIs represented by single reads, sensitivity, linearity of amplification, number of genes detected, and the 3' read bias by library (Figures S4A–S4F). We included UMIs represented by only one read, although others have removed them to reduce spurious UMIs (Zeisel et al., 2015), since we saw no increased fidelity upon their removal (Figure S4B). Principal-component analysis (PCA) separated progenitors and neurons by principal component 1 (PC1) and further separated differentiation phase by PC2 (Figure 2C). Genes with variance greater than technical noise (estimated by External RNA Control Consortium [ERCC] spike-in controls; Figure S4H; Table S1) were used to drive iterative cell clustering at each time point using gene modules derived from weighted gene co-expression network analysis (WGCNA; Tasic et al., 2016; Thomsen et al., 2016; Zhang and Horvath, 2005) (Figure 2D; Data S1). Robustness of cell clusters was determined through bootstrapping, where a random 80% of cells from each time point were re-clustered 100 times (Figures 2D and S4I). In total, 41 cell-type clusters were observed: 14 neural progenitor, 5 transitional, 19 neuronal, and 3 non-neural cell types (Figures 2E and 2F). Cell types are named by time point of origin; whether they are a progenitor (P), transitional (T), neuronal (N) or other tissue (O) cell type; and where necessary, the most specifically expressed gene for that cell type (e.g., D54\_N\_NEUROD6). Lastly, we generated a “constellation diagram” that shows cell types, relative frequency per time point, average co-clustering, and frequency of inter-cluster mixing during bootstrapping analysis (Figure 2F).

Within progenitor and neuronal cell types found at and after day 26, we observed co-expression of markers suggestive of telencephalic identity (*LHX2*, *FOGX1*, and *FEZF2*; Hanashima et al., 2004; Hirata et al., 2004; Porter et al., 1997). At day 26, a progenitor cell type emerged that expressed markers suggestive



**Figure 1. In Vitro Neural Differentiation Generates Cortical and Non-cortical Cells**

(A) Schematic representation of in vitro neural differentiation of hESCs.

(B) Representative images of immunostaining on day (D)12, D26, and D54 of H1 differentiated cells, with DAPI in blue. Scale bar, 100 μm.

(C) Representative traces of calcium activity as imaged with FURA2-AM (top traces) and after blockade by TTX (bottom traces). Data quantified from three representative experiments ( $n = 1,148$  cells at D54, from three biological replicates) (bottom); RFU, relative fluorescent units. Error bars represent mean ± SD.

(D) Principal-component analysis of population RNA-seq data demonstrates the reproducibility of differentiation methods across multiple experiments from both H1 and H9 stem cell lines.

(E) Population RNA-seq expression of genes that mark indicated brain regions.

(F) Schematic of targeted loci of the SOX2<sup>Citrine/+</sup> and DCX<sup>Citrine/Y</sup> reporter cell lines.

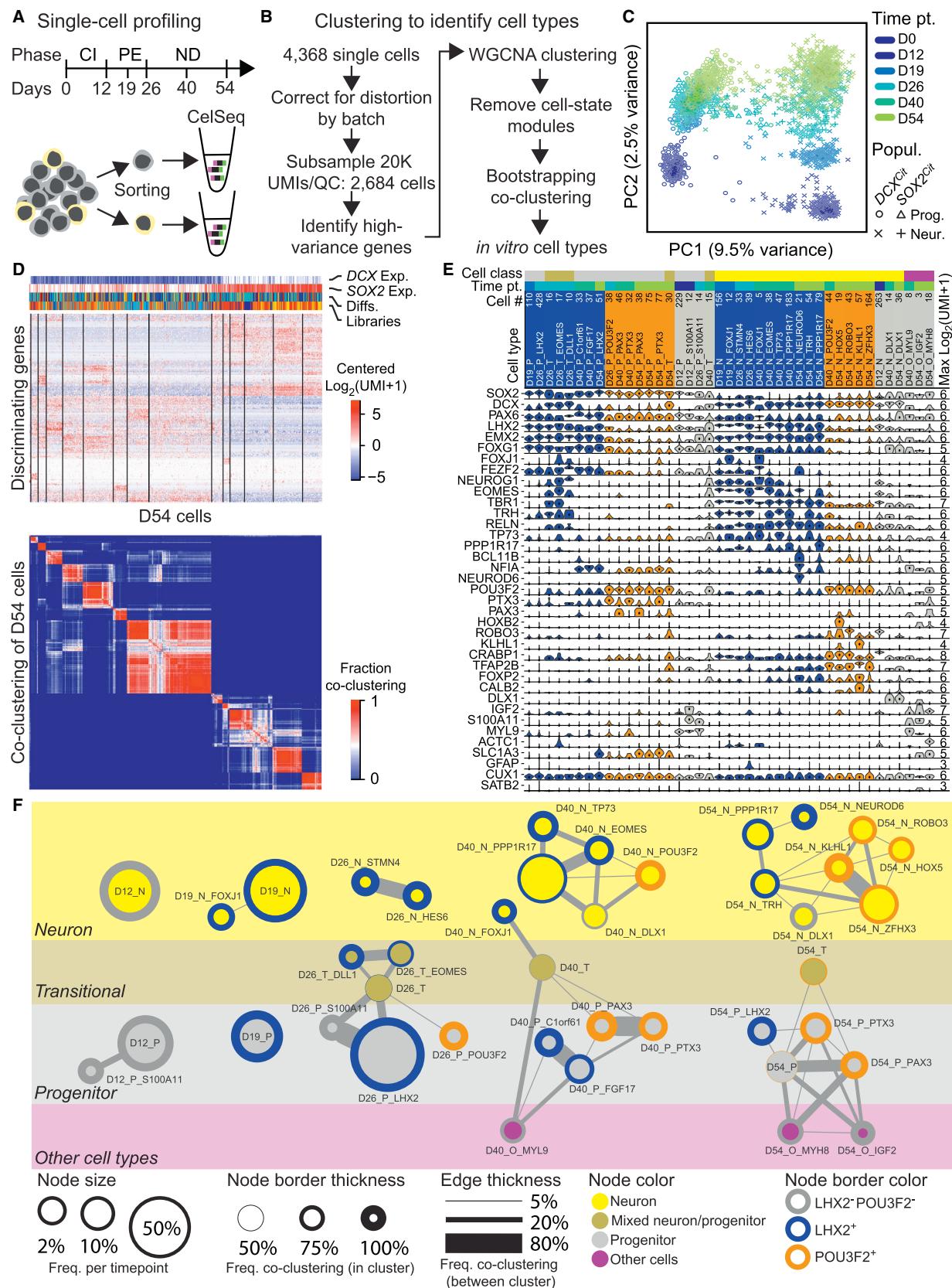
(G) Quantitation of the percentage of citrine-positive cells during differentiation by flow cytometry. Shown are mean ± SD values from three (SOX2<sup>Citrine/+</sup>) and six (DCX<sup>Citrine/Y</sup>) biological replicates.

See also Figures S1 and S2.

of mid/hindbrain identity (PTX3 and POU3F2; Figures 2E, 2F, and S5A). At days 40 and 54, we observed additional progenitor and neuronal cell types expressing mid/hindbrain markers such as TFAP2B, PAX3, ROBO3, and HOXB2, but not LHX2 (Figures

2E, 2F, and S5A). Thus, our differentiation method furnishes a combination of forebrain and mid/hindbrain cell types.

To independently validate key features of the gene expression data, we generated a dataset using SmartSeq2 (that has no



multiplexing of samples; see below) and also conducted immunostaining. By both RNA-seq and immunostaining, LHX2 and POU3F2 showed little expression overlap within neural progenitors at day 26, and this segregation persists within neurons at day 54 (Figures 3A and 3B). In contrast *EOMES* and *TBR1* are frequently co-expressed in day 26 and day 54 neurons in all datasets (Figures 3C and 3D).

### Cell Types Show Forebrain and Mid/Hindbrain Regional Identities

To establish the nature of our cell types, we compared our cells to primary brain cell expression datasets. *EOMES* expression is frequently used as a marker of intermediate progenitor cells (Qian et al., 2016; Shi et al., 2012); however, many neurons co-express *EOMES* and *TBR1*, suggestive of preplate cells (Bulfone et al., 1999). Consistently, expression of *Eomes* and *Tbr1* is apparent in the same layer of early-borne mouse neurons at embryonic day 11.5 (E11.5) and E13.5, but not at E15.5 (Figure 3E). Among genes selectively expressed in LHX2<sup>+</sup> rather than POU3F2<sup>+</sup> neurons at days 40 and 54, 56% of those present in the Developing Mouse Brain Atlas (Thompson et al., 2014) at E13.5 and E15.5 showed expression in cortical preplate and the presumptive marginal zone (Table S5). These data suggest that some cultured neurons correspond to preplate or other early neuronal cell types.

We compared the single-cell transcriptomes of day 54 cells to the BrainSpan Atlas of the Developing Human Brain (Miller et al., 2014) and the Developing Mouse Brain Atlas. We focused on the conserved co-expressed gene modules identified by WGCNA (Langfelder and Horvath, 2007) that distinguished both the hESC-derived cell types and brain regions (gene modules shown in Table S1). Statistical significance of the assessed conserved gene modules was based on permutation analysis (Langfelder et al., 2011). We estimated the similarity between differentiated day 54 neurons and brain regions by comparing the Spearman correlation coefficient based on these conserved co-expressed gene modules (Table S1). Neuron clusters D54\_N\_NEUROD6 and D54\_N\_PPP1R17 (marked by *LHX2* and *TBR1*) showed strong correlation to cortex but weak correlation to mid/hindbrain regions in both mouse at E13.5 and human at post-conception week (PCW) 15–16 (Figures 4A and 4B). In contrast, *POU3F2*<sup>+</sup> neuron clusters D54\_N\_HOX5, D54\_N\_ROBO3, D54\_N\_KLHL1, and D54\_N\_ZFHX3 showed weak correlation to cortex but strong correlation to mid/hindbrain, while D54\_N\_TRH correlated best to diencephalon (forebrain) and midbrain. Finally, neuron cluster D54\_N\_DLX1 showed high cor-

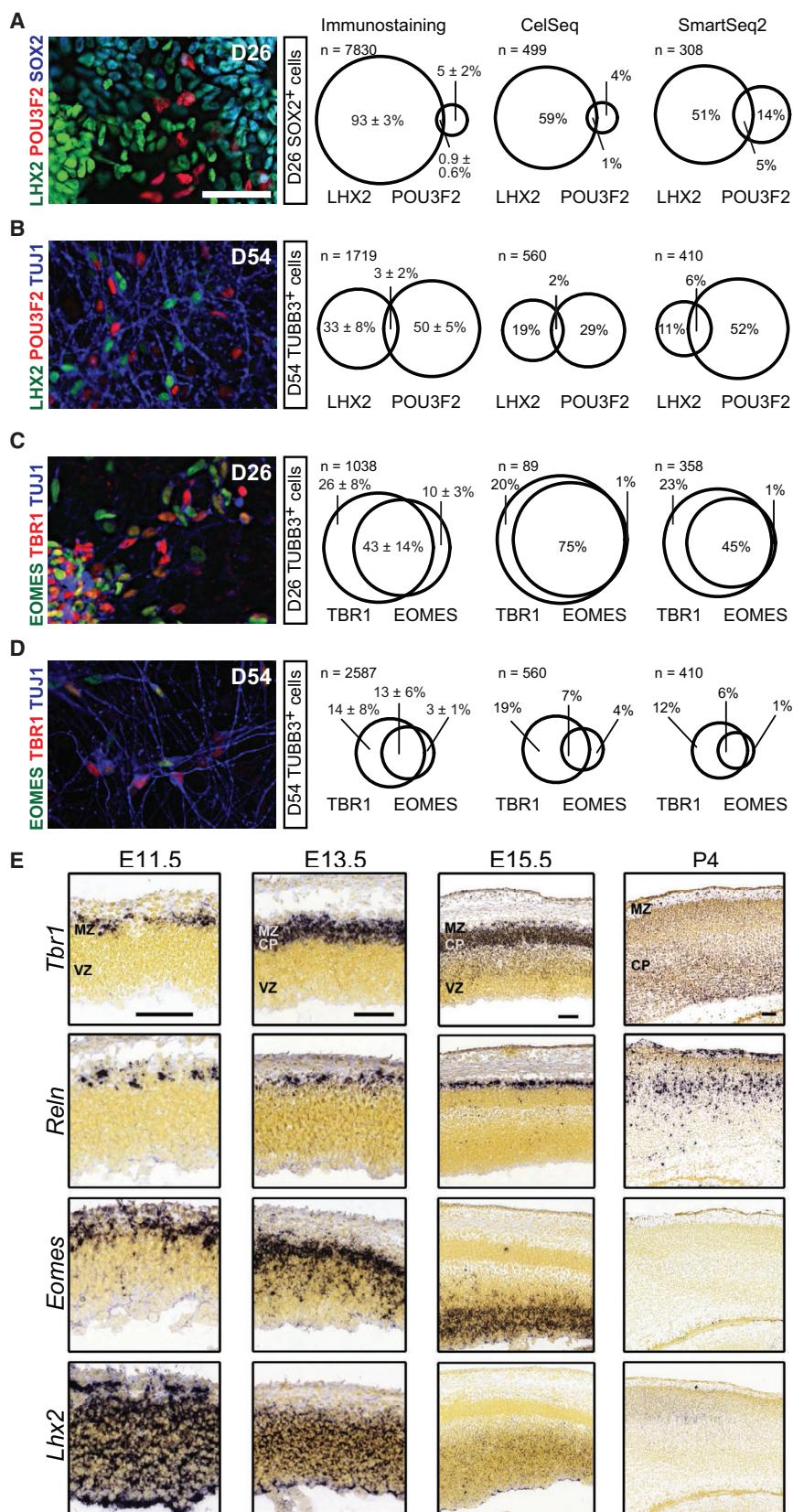
relation to the ganglionic eminences (Figures 4A and 4B). Antibody staining of human mid-gestational (122–132 days post-conception [dpc]) brain tissue showed that *LHX2* is a marker of human cortical, but not hindbrain, progenitors (Figure 4C). Markers associated with the LHX2-expressing neuronal cell types D54\_N\_NEUROD6, D54\_N\_PPP1R17, BCL11B<sup>+</sup>, and NFIA<sup>+</sup>NeuN<sup>+</sup> were highly expressed in the cortical plate and intermediate zone of the cortex, respectively, but not in hindbrain cells (Figure 4D). In contrast, markers of the POU3F2-expressing neuron clusters, TFAP2B and ROBO3, were present in hindbrain, but not cortex (Figure 4D). Thus, the POU3F2-expressing neurons likely correspond to posterior brain neuron types, while *LHX2*-expressing clusters largely correspond to forebrain cell types. These data demonstrate progenitors and neurons belonging to multiple human brain regions are generated in vitro despite the directed cortical induction.

To assess the similarity between the in vitro LHX2<sup>+</sup> cell types and primary cortical cell types at a single-cell level, we analyzed neurons and progenitors prospectively isolated from fixed primary human cortical samples using the fixed and recovered intact single cell RNA (FRISCR) technique (Thomsen et al., 2016). We directly compared 472 single cells from human cortical samples (115 and 96 dpc) to the in vitro differentiated cells. Progenitors (SOX2<sup>+</sup>PAX6<sup>+</sup>TuJ1<sup>-</sup>) and neurons (SOX2<sup>-</sup>PAX6<sup>-</sup>TuJ1<sup>+</sup>) were each further stratified by *EOMES* expression to isolate *EOMES*<sup>-</sup> radial glia, *EOMES*<sup>+</sup> intermediate progenitors (IPCs), and *EOMES*<sup>+</sup> and *EOMES*<sup>-</sup> neurons (Figure 5A). FRISCR data from fetal tissue were compared to SmartSeq2 single-cell RNA-seq data from *LHX2*<sup>+</sup> cultured cells at days 26 and 54 (Figure 5B; Data S1). Many canonical markers of progenitors, IPCs, and neurons were conserved in vitro (Figure 5B). In contrast, outer subventricular zone-localized radial glia (oRG) markers such as *HOPX* and *TNC* were reduced or absent in in-vitro-derived progenitors. Thus, although a few *HOPX*<sup>+</sup> cells were observed, most hESC-derived progenitors lacked oRG characteristics (Figure 5B; Data S1). Gene Ontology analysis showed significant enrichment of gliogenesis and oligodendrocyte differentiation genes in primary progenitors, while embryonic development, extracellular matrix, and stress response gene classes were enriched in hESC-derived progenitors (Figure 5C). Progenitors appear to mature as the ventricular zone-enriched radial glia (vRG) marker *ANXA1* and oRG markers *FAM107A* and *TNC* become more highly expressed at day 54 relative to day 26 (Figure 5D; Data S1), and gene upregulation at day 54 was predictive of its expression in primary progenitors. Thus, over time, the cultures increasingly resemble mid-gestation primary cortical cells.

### Figure 2. Identification of Cell Types through Single-Cell Transcriptomics

- (A and B) Single-cell profiling strategy (A) and methodology of cell type identification (B) from CelSeq single-cell RNA-seq data.
- (C) Principal-component analysis of all single cells used for analysis based on high variance genes (Table S1).
- (D) Representative heatmap showing most distinguishing genes at day (D)54 used to identify cell types (top, Data S1), and bootstrapping analysis of D54 cells.
- (E) Violin plots for most distinguishing and commonly used marker genes. Max Log<sub>2</sub>(UMI + 1) values are to the right, and the number of cells per cell type is listed above the cell-type name.
- (F) Constellation diagram of cell types identified at D12, D19, D26, D40, and D54. Cell-type name and relative number of cells per time point sampled are shown by circle size; the strength of intra- and inter-cell-type clustering from bootstrapping analysis is indicated by circle border with and edge width, and *POU3F2* and *LHX2* expression status is shown by circle border color. Neuronal cell types (yellow) were defined as cell types with strong DCX expression and progenitors with strong SOX2 expression (gray); transitional types express both SOX2 and DCX (gray-yellow transition), and other cell types (pink) express genes indicative of non-neuronal lineages.

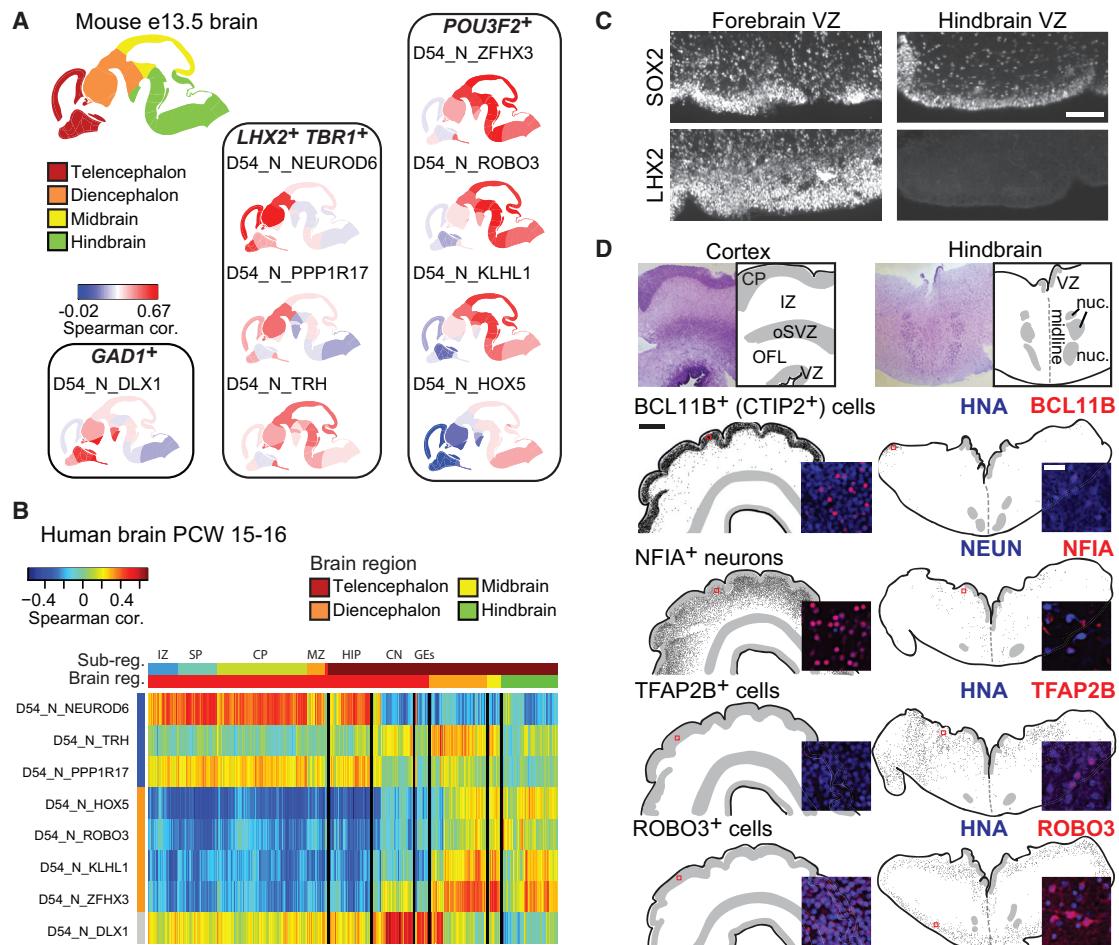
See also Figures S3 and S4, Table S1, and Data S1 and S2.



**Figure 3. Validation of Single-Cell RNA-Seq Data and Creation of Preplate Cortical Neurons**

(A–D) Differentiating cultures were immunostained with the indicated cell type markers and types of progenitors (SOX2<sup>+</sup> cells) and neurons (TUJ1<sup>+</sup> cells) were quantified. n = 4–6 independent experiments per immunostaining condition, scale 50 µm. The quantification by immunostaining (left) is directly compared to single-cell gene expression data generated by CelSeq (middle) and SmartSeq2 (right). Cells were scored positive for CelSeq if their UMI was > 4 and for SmartSeq2 if TPM > 0.

(E) Chromogenic *in situ* hybridization data show early co-expression of *Tbr1*, *Reln*, *Eomes*, and *Lhx2* in E11.5–E13.5 mouse cortex but no overlap of *Tbr1* and *Eomes* after E15.5. Data are from Allen Developing Mouse Brain Atlas. Scale bar, 100 µm. See also Figure S5 and Table S5.



**Figure 4. Stem Cell-Derived Cell Types Resemble Forebrain and Mid/Hindbrain Cell Types**

(A) Spearman correlation of day (D)54 neuronal cell types to E13.5 Allen Brain Atlas of the Developing Mouse Brain based on genes differentially expressed between cell types and tissue regions (Table S1). Mouse regional gene expression levels are derived from *in situ* hybridization staining intensity.

(B) Correlation of single D54 neurons with regions of the human brain from the Brainspan Atlas of the Developing Human Brain. Spearman correlations are based on genes differentially expressed between cell types and tissue regions (Table S1). IZ, intermediate zone; SP, subplate zone; CP, cortical plate; MZ, marginal zone; HIP, hippocampus; CN, cerebral nuclei; GE, ganglionic eminences.

(C) Fluorescence micrographs of 122 dpc cortex and 132 dpc hindbrain. LHX2 marks human cortical, but not hindbrain, progenitors, while SOX2 marks progenitors in both regions. Scale bar, 100 μm.

(D) Immunohistochemistry of cortical and hindbrain cell-type markers. Top: Nissl stain and representation of tissue architecture are shown; below: tissue representation based on DAPI staining. VZ, ventricular zone; OFL, outer fiber layer; oSVZ, outer subventricular zone; IZ, intermediate zone; CP, cortical plate; nuc, medullary nuclei. The entire tissue section was scored, and each dot represents a positive cell. Scale bar, 1 mm. Inset: fluorescence micrograph showing a representative image (location indicated by red box); scale bar, 25 μm.

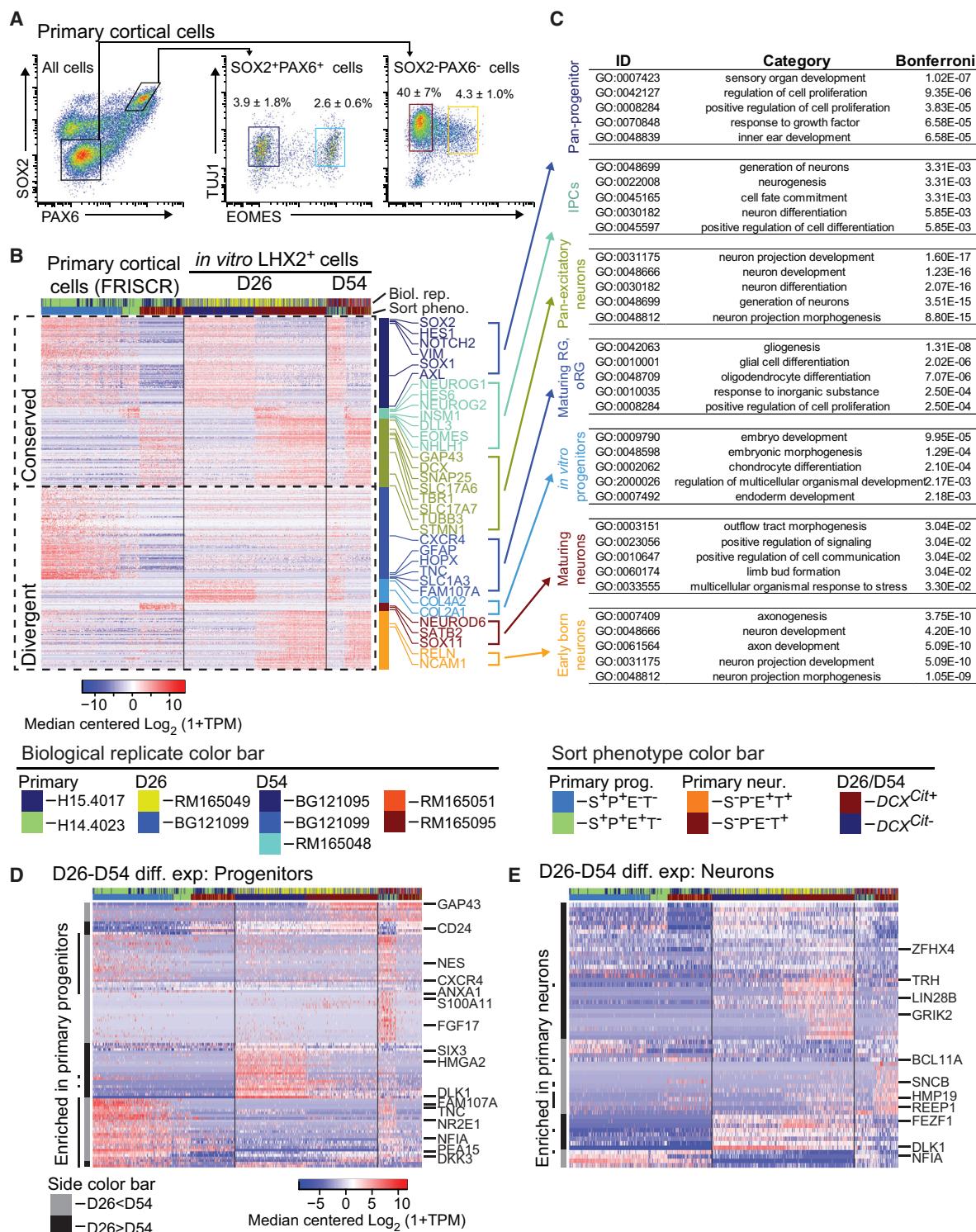
See also Table S1.

We then evaluated differences between primary and in-vitro-generated neurons. Although primary IPCs appear distinct from neurons or progenitors, cultured cells that express IPC genes (including *EOMES*) strongly resemble neurons (Figure 5B). Fetal *EOMES*<sup>+</sup> cells had higher expression of the progenitor eigengene (the average expression of the progenitor gene module) than hESC-derived cells ( $p = 3 \times 10^{-7}$ , t test), and hESC-derived cells exhibit higher expression of the neuron eigengene than fetal cells ( $p = 3 \times 10^{-20}$ , t test). hESC-derived neurons lacked substantial expression of *SATB2* (a marker of later-born callosal-projecting neurons) and expressed the Cajal-Retzius cell marker *RELN*. Lastly, genes upregulated at day 54 were more likely to be expressed in primary mid-gestation neurons (Figure 5E; Data

S1). These data show at a single-cell level that our hESC-derived cells are molecularly similar to early primary cortical cells and lack hallmarks of more mature neurons.

#### An Inferred Lineage Tree with Forebrain and Mid/Hindbrain Branches

The clustering of single-cell transcriptomes resulted in identification of different cell types, which can be linked as branches in a lineage tree or states along a differentiation trajectory. To resolve the hierarchy of these cell types with putative lineal relationships, typical techniques use distance metrics calculated from the high-dimensional molecular data (Shin et al., 2015; Trapnell et al., 2014); however, the number of master-like molecules



**Figure 5. Comparison of Single Stem Cell-Derived Forebrain Cells and Primary Human Single Cells**

(A) Flow cytometry plot showing primary cell populations that were sorted ( $n = 4$ ) and profiled using FRISCR ( $n = 2$ ). Values are mean  $\pm$  SD of population percentage derived from four brains.

(B) Expression of conserved and divergent gene modules between primary human cortical single cells and *in vitro*-differentiated progenitors and neurons at day D26 and D54 (Data S1). For each block, progenitors are to the left, intermediate progenitors are in the middle, and neurons are to the right.

(C) Results of Gene Ontology analysis of conserved and divergent gene expression modules. The top five most significant biological processes with a Bonferroni correction value  $<10^{-1}$  are shown.

(legend continued on next page)

instructing cell-fate decisions may be few (Colasante et al., 2015; Takahashi and Yamanaka, 2006; Vierbuchen et al., 2010). Furthermore, lineage algorithms based on transcriptomic data usually model a progression and not bifurcations (Shin et al., 2015; Trapnell et al., 2014). Indeed, a low-dimensional projection of the single-cell data like PCA (Figure 2C) does not readily suggest a biologically meaningful linkage based on spanning-tree methods.

To address this, we used a Bayesian computational approach to simultaneously infer cell clusters' lineal relationships and the key set of markers and transition genes that define these relationships (S.R., unpublished data). This method analyzes the relationships between all the clusters, three at a time, and assumes that good marker genes are uniquely expressed in a cell cluster, while genes establishing relationships are shared between two of the three clusters. Briefly, we determined relative relationships between all possible triplets of cell types at neighboring time points, assessed their putative lineage, and identified genes with expression patterns reflecting this relationship. In each case, the three types of cells were separated in a subset of transcription factor space, and one of the types is defined as the intermediate. The intermediate type could be the parent, leading to two daughters, or alternatively, could be a transitional cell type between the other two types. To demonstrate this, we show the transcription factor expression for a triplet with strong evidence for an intermediate state (Figure 6A).

After identifying 162 high-confidence triplets with evidence of a lineal relationship, we manually assembled these triplets into a putative in silico lineage tree (Figure 6B). The tree was rooted at day 12 and assembled iteratively using information from triplets containing successively more mature time points. Seven of the forty-one cell types could not be linked to other cell types with high confidence and were omitted from the lineage tree. Overall, our in silico tree suggests a major branch point separating the *POU3F2*-expressing and *LHX2*-expressing types and identifies potential transcription factor candidates involved at specific branch points (Figure 6C).

#### Molecular Regulation of the Lineage Tree Branches

Although not immediately apparent from the inferred lineage, there are known pathways driving brain regionalization, such as fibroblast growth factor (FGF), BMP (Bertacchi et al., 2013), retinoic acid (Maury et al., 2015), and Wnt/β-catenin (Maroof et al., 2013; Schüller and Rowitch, 2007) signaling pathways. PE phase containing FGF2 was removed entirely to determine if it was caudalizing the cells, but the chief markers of neuron types (forebrain excitatory, mid/hindbrain excitatory, and interneuron) were still present (Figure S5C). We also investigated whether the *LHX2*-*POU3F2* lineage tree bifurcation occurred due to incomplete SMAD inhibition during CI, but the appearance of day 26 *POU3F2*<sup>+</sup> progenitors was unaffected by higher concentrations of SMAD inhibitors (Figure S5D). Similarly, the pan-retinoic acid receptor inhibitor BMS-493 had no effect on

*POU3F2*<sup>+</sup> progenitors although it was able to block the demonstrated caudalizing effect of added retinoic acid (Figure S5E). In contrast, we observed that inhibiting and activating the canonical Wnt/β-catenin pathway prevented and promoted, respectively, the appearance of day 26 *POU3F2*<sup>+</sup> progenitors (control, 4.8% ± 1.0%; XAV-939, 0.6% ± 0.2% [p < 0.01, unpaired t test]; CHIR-99021, 94% ± 6% [p < 0.001, unpaired t test]; Figures 6D, 6E, and S5F). In addition, Wnt/β-catenin signaling promotes the *POU3F2* branch during CI, but not during PE phase (Figure S5F), and consistent with this, we observe the mid/hindbrain Wnt/β-catenin targets *AXIN2* and *TNFRSF19* (Figure S4G; Ha et al., 2012) selectively expressed in the minor day 12 progenitor cluster D12\_P\_S100A11 (Figure 6F, p < 0.01, Fisher's exact test); both of these insights together confirm the biological significance of our discovered cell types and suggest that the origin of the mid/hindbrain *POU3F2* branch can occur during early neurodevelopment prior to the bifurcation at day 26 (Figure 6B). These functional data agree with recent reports of in vitro neural regionalization (Maroof et al., 2013; Maury et al., 2015) and demonstrate direct concordance between our observed lineage branches and those regional cell types previously studied alone.

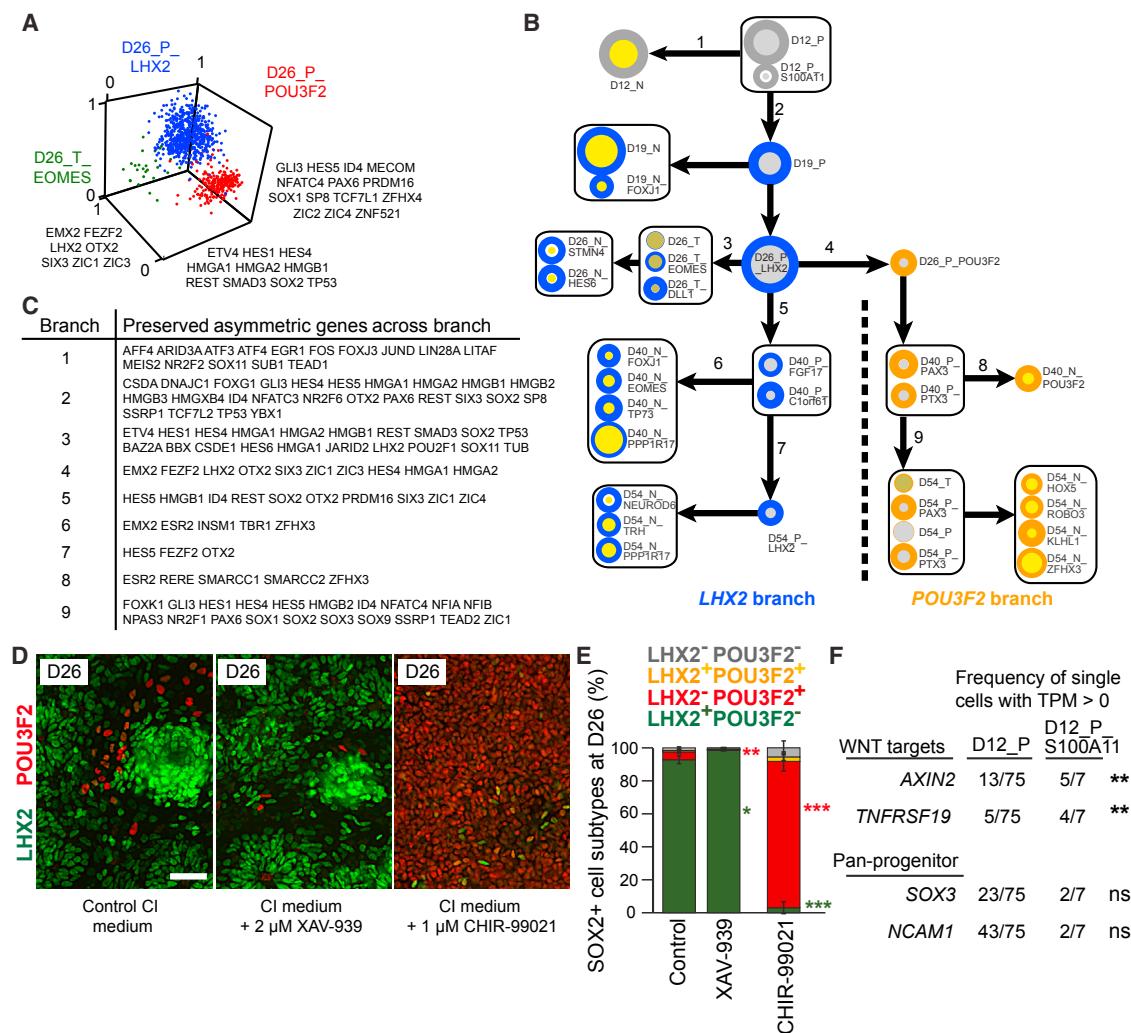
#### Clonal Analysis Confirms Forebrain Cell Types Segregating from Mid/Hindbrain Cell Types

The in silico lineage tree predicts that *POU3F2*-expressing and *LHX2*-expressing day 54 cell types arise from either distinct or common progenitors as both can be detected at day 26. To test this, we undertook clonal analysis of cell fate from day 26 progenitors using a viral barcoding strategy (Figure 7A). We cloned a 10-bp degenerate barcode library in the 3' UTR of a *tdTomato* expression cassette into a VSV-G-pseudotyped retrovirus library (Figure S6A), which showed no obvious tropism bias (Figure S6B). Clones tracked by daily time-lapse microscopy display neuronal and non-neuronal morphologies, and clones analyzed by immunostaining demonstrate subtype diversity (Figure 7B and S6C). In addition, strong transcriptomic correlation was observed between infected and uninfected cells (Figure S6D).

For barcoded lineage tracing, day 27 cells were infected following re-plating at day 26 with the barcoded retrovirus library and then allowed to differentiate until day 54 (Figures 7A and S6E) when cells were processed to simultaneously recover the retroviral barcode (clone association) and the transcriptome (cell type). To match day 54-infected cells profiled by SmartSeq2 with cell types identified by CelSeq, we developed a consensus predictive gene set consisting of genes from CelSeq modules that were also detected by SmartSeq2 (Figures S7A–S7D; Table S1) and then built a random forest model using these features. From two independent differentiation experiments containing in sum 81 total clones, we sorted single infected cells based on *tdTomato* fluorescence and DCX-citrine expression (Figure S6E). We obtained barcodes from 375 *tdTomato*<sup>+</sup> sorted single cells, and 315 high-quality cells could be used to evaluate cell type composition of the clones. 41 multicellular clones were

(D and E) Heatmaps show genes differentially expressed between D26 and D54 progenitors (D) and neurons (E) (Data S1). Column color bars and single-cell order are as in (B), and row color bars show if expression is elevated at D26 (black) or D54 (gray). Black line shows genes with expression enriched in primary progenitors relative to primary neurons (D) or primary neurons relative to primary progenitors (E).

See also Data S1.



**Figure 6. A Lineage Tree from Single-Cell Transcriptomics that Is Modulated by Canonical Wnt/β-Catenin Signaling**

(A) Example of one triplet of transcriptomic cell types showing strong evidence for an intermediate or parent state (in blue). The non-intermediate states (red and green) express genes only along one of the two horizontal axes, whereas the intermediate state expresses both sets of genes and also expresses a set of marker genes (vertical axis) that is not highly expressed in either of the other two states. Axis values represent means of normalized gene expression over all the genes on a given axis. Transcriptomic types are named as in Figure 2.

(B and C) In silico lineage tree assembled from triplets showing strong evidence of an intermediate state. Circles around groups of cell types indicate that they are not distinguishable in terms of lineage or progression using the tree-building algorithm. Transcriptomic types are named and colored as in Figure 2. Arrows indicate proposed lineage/progression links, and key asymmetrically regulated genes indicated by numbers next to the lineage arrow are listed in (C).

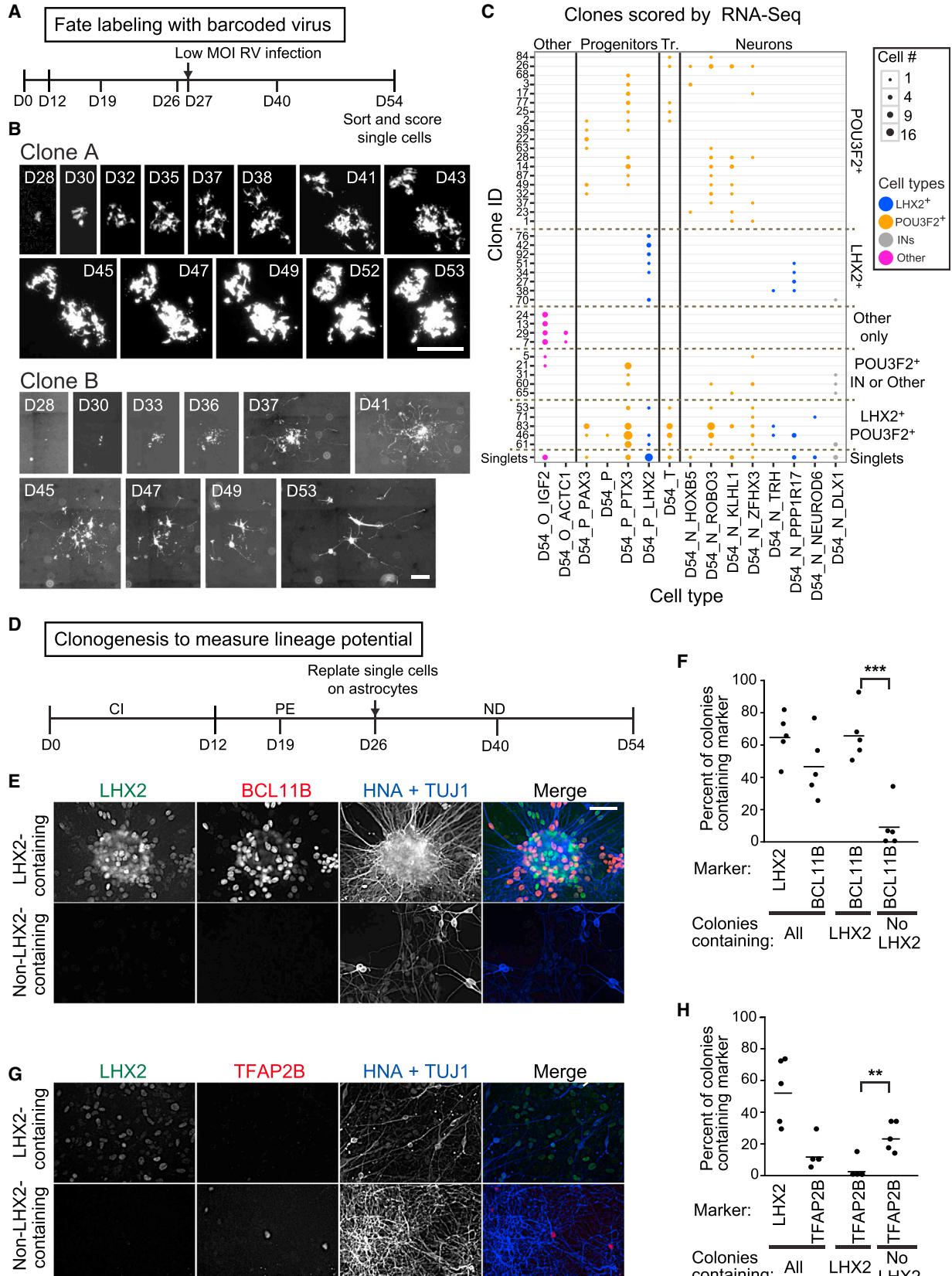
(D) Differentiating cultures were treated with 2 μM XAV-939 or 1 μM CHIR-99021 during CI phase and then fixed and immunostained at day (D)26 for LHX2 (green), POU3F2 (red), and SOX2 (not shown) to identify progenitors of each lineage branch. Scale bar, 50 μm.

(E) The proportion of SOX2+ progenitors containing for LHX2 and/or POU3F2 was quantified. Data are from n = 3 independent experiments, with >1,000 cells counted per condition. \*p < 0.05, \*\*p < 0.01, and \*\*\*p < 0.001 by unpaired t test relative to control. Error bars represent mean ± SD.

(F) Canonical Wnt/β-catenin target genes AXIN2 and TNFRSF19 (but not pan-progenitor genes SOX3 or NCAM1) are disproportionately expressed by the minor progenitor cluster D12\_P\_S100A11 relative to the major cluster D12\_P in the SmartSeq2 dataset. \*\*p < 0.01; ns, not significant (Fisher's exact test). See also Figure S5.

detected, with 33 spanning more than one cell type (Figures 7C and S6F–S6I; Table S2). 24 clones contained POU3F2+, but not LHX2+, cell types; 8 clones contained LHX2+, but not POU3F2+, cell types; and only 5 clones contained both POU3F2+ and LHX2+ cell types, which is fewer than would be observed if these cell types assort into clones independently (p < 0.01 by Fisher's exact test; Figure 7C; Table S2). The five clones containing both POU3F2+ and LHX2+ cell types contained more cells

than the others (median = 19 cells versus median = 2.5 cells, p = 0.016 by two-sided Kolmogorov-Smirnov test), suggesting they arose from more primitive neural progenitors. D54\_N\_DLX1 interneurons were detected in clones with LHX2+ or POU3F2+ cells, and “other” cell types were found in clones composed mainly of like cell types. These fate-mapping data directly validate the computational lineage tree, indicating that the majority of neural progenitors at day 27 are committed to



(legend on next page)

distinct region-specific fates, yet some primitive progenitors remain uncommitted.

To test if differences in cell fate of day 26 progenitors were due to cell-autonomous differences in lineage potential and to bypass possible biases from viral fate mapping, we generated single-cell clones by seeding day 26 progenitors at clonal density to differentiate for 4 weeks on mouse astrocytes (yielding  $2.3 \pm 1.4$  spatially resolved colonies per well; **Figures 7D** and **S6J**). Colonies varied in size (1–1,000 cells), but those with cortical lineage potential (containing LHX2<sup>+</sup> cells) typically contained more cells than colonies lacking LHX2<sup>+</sup> cells (**Figure S6K**). Cell-type-specific antibodies were used for immunostaining to assess cell-type composition within colonies (850 colonies analyzed over  $n = 5$  independent experiments). We marked human cells with human nuclear antigen (HNA), neurons with TUBB3 (TuJ1) and cortical lineage potential cells with LHX2. These markers were multiplexed with cell-type markers identified from transcriptomics. One such cell-type marker is BCL11B, which marks D54\_N\_NEUROD6 and D54\_N\_PPP1R17 cell types (**Figures 2E** and **4D**). Colonies containing LHX2<sup>+</sup> cells much more likely included neurons expressing BCL11B<sup>+</sup> neurons than those lacking LHX2<sup>+</sup> cells (66%  $\pm$  14% versus 9%  $\pm$  14%,  $p < 0.001$ , unpaired t test; **Figures 7E** and **7F**), similar to NFIA<sup>+</sup>TuJ1<sup>+</sup> cortical neurons (**Figure S6L**). In contrast, colonies containing LHX2<sup>+</sup> cells were less likely to contain TFAP2B<sup>+</sup> neurons (**POU3F2**<sup>+</sup> clusters) than those lacking LHX2<sup>+</sup> cells (3%  $\pm$  6% versus 24%  $\pm$  9%,  $p < 0.01$ , unpaired t test; **Figures 7G** and **7H**). These results directly link prominent cell type markers to the LHX2 or the POU3F2 lineage tree branches. Other cell-type markers that we observed in our sequencing dataset (POU3F2, FOXP2, CALB2 [calretinin], and CRABP1) demonstrated no significant colony associations with LHX2 (**Figures S6M–S6P**). Together, these clonal outgrowth results independently corroborate the distinct cortical and mid/hindbrain branches of the human neural single-cell lineage tree. Moreover, they suggest that region-specific branches are caused by differences in cell-autonomous commitment among day 26 progenitors rather than stochastic and/or non-cell-autonomous phenomena.

## DISCUSSION

We present a comprehensive characterization of human brain cell types generated from hESCs and define 41 cell types consisting of 14 progenitor, 5 transitional, 19 neuronal, and 3 non-

neural cell types. Our data demonstrate that multiple cortical and mid/hindbrain neuronal and progenitor cell types are produced in our 2D in vitro differentiations, and these regional identities can be toggled by modulating canonical Wnt/β-catenin signaling. In addition, we chart the molecular overlap between cultured and primary cortical cell types, using both atlas and single-cell analysis from human primary cortex. A Bayesian algorithm was used to generate a lineage tree and chart the trajectory of our regionally diverse cell types, while highlighting regulators of the lineage tree. Finally, we confirmed these lineage relationships through clonal analyses of fate and lineage potential to directly demonstrate that hESC-derived neural progenitors have cell-autonomously committed identities. This study charts human brain region-specific development, which is essential for accurately modeling and understanding the logic and uniqueness of human brain.

Our inferred lineage tree branches mimic established regional differences in neurogenesis. Although the inferred lineage tree was supported by our fate-mapping experiments, some cell types could not be mapped with high confidence into the lineage tree. The cell types are lineally related as shown by clonal analysis, but the lineage algorithm requires the identification of intermediate cell types to allow the high-confidence prediction of linkage. Furthermore, the algorithm requires the asymmetric distribution of transcription factor expression. If cell sampling failed to recover a short-lived or rare intermediate, or if there were too few asymmetrically expressed transcription factors, our algorithm could not generate a high-confidence linkage.

Despite the forebrain-mid/hindbrain branch point at day 26, Wnt/β-catenin signaling instructs the mid/hindbrain lineage branch prior to day 12 (**Figures 6D–6F** and **S5F**). This is consistent with previous hESC experiments (Maroof et al., 2013; Maury et al., 2015) and mouse genetics (Schüller and Rowitch, 2007), leading to Wnt target gene expression in D12\_P\_S100A11 cells (**Figure 6F**), but not yet mid/hindbrain markers (**Figure 2E**). The mid/hindbrain state could be specified in these cells prior to expression of markers; alternatively, a lack of observed S100A11 cells at day 19 could have made this lineage branch impossible to detect. Regardless, use of WNT to bias the lineage tree will help produce more uniform populations of targeted brain region cell types from hESCs.

Our lineage tree shows how early-born human neural progenitors give rise to early neurons with different regional identities. Neural pan-progenitor genes like VIM, PAX6, and SOX2 are expressed in progenitors from day 12, but the vRG marker

**Figure 7. Clonal Analysis Confirms Distinct POU3F2 and LHX2 Branches of the Human Brain Lineage Tree**

- (A) Schematic of viral barcoding experiment, indicating time of infection and collection.  
(B) Representative fluorescence micrographs captured from live differentiating clones, exhibiting non-neuronal (clone A) and neuronal morphologies (clone B). Images are from montages of entire wells. Scale bar represents 250 μm (clone A) or 500 μm (clone B).  
(C) All multicellular clones are shown as individual rows and each day (D)54 cell type as columns. All single-cell clones are combined in the bottom row ("singlets"). The size of each dot indicates the number of cells per cell type from a clone. Dots are colored as cell types in **Figure 2E**.  
(D) Clonal analysis of cell-autonomous lineage potential was performed by re-plate D26 progenitors at clonal density on feeder mouse astrocytes and then analyzing outgrown colonies for cell composition at D54 by immunostaining.  
(E and G) Example colonies are stained with antibodies for HNA plus TuJ1 (blue), LHX2 (green), and either BCL11B (E, red) or TFAP2B (G, red). Colonies were grouped into categories of LHX2-containing and non-LHX2-containing. Scale bar, 50 μm.  
(F) Colonies that contain LHX2<sup>+</sup> cells are more likely to contain BCL11B<sup>+</sup> cells as compared to colonies lacking LHX2<sup>+</sup> cells.  
(H) Colonies that contain LHX2<sup>+</sup> cells are less likely to contain TFAP2B<sup>+</sup> cells as compared to colonies lacking LHX2<sup>+</sup> cells.  
In (F) and (H), five independent differentiations were analyzed, and 25–38 colonies per immunostaining cocktail per experiment were inspected for the presence of cell types. \*\*\* $p < 0.001$ , \*\* $p < 0.01$  by unpaired t test. See also **Figures S6** and **S7** and **Table S2**.

ANXA1 and oRG markers *FAM107A*, *HOPX*, and *TNC* (Pollen et al., 2015; Thomsen et al., 2016) are only expressed at day 54, indicating that these cells likely reflect radial glia prior to oSZ formation. This early staging is consistent with our results mapping the cultured cells to the earliest stages of the human developmental atlas. We also detected early-formed human neurons. The earliest (day 12) *DCX*<sup>+</sup> neurons express *TBR1* and weak levels of *EOMES* and *RELN*. Perhaps these day 12 *DCX*<sup>+</sup> neurons correspond to predecessor cells that express *TBR1* and are formed during neural tube closing (Bystron et al., 2006). Cells with highest *RELN* levels (presumptive Cajal-Retzius cells; Hevner et al., 2003) were detected later at days 26 and 40, and these frequently expressed *EOMES*. Although *EOMES* is a canonical marker of proliferative IPCs in vivo, most in vitro *EOMES*<sup>+</sup> cells were not proliferative and instead had a strong neuronal signature. These cell types resemble preplate cells that co-express *EOMES* and *TBR1* (Figure 3E; Bulfone et al., 1999). Indeed, 56% of genes differentially expressed in neurons from the *LHX2*<sup>+</sup> branch are detected in mouse preplate or preplate-derived cells. We believe that future human embryonic brain single-cell atlases are required for comparison and to guide the hESC modeling field.

The cortex develops through the sequential creation of cellular layers that exhibit unique molecular and morphological properties. We did not detect some classical (especially upper) layer markers (e.g., *SATB2* and *CUX2*). This could be due to the limited time in culture (54 days), and progenitors may still be producing early-born neurons. Also, classical layer markers may have poor predictive utility at mid-gestation stages; they are only weakly expressed in fetal macaque with dramatic regional and temporal variation (Bakken et al., 2016). Furthermore, neurons from primary mid-gestational tissues appear relatively homogeneous (Figure 5B; Camp et al., 2015; Pollen et al., 2014). These considerations spotlight single-cell RNA-seq as a high-value methodology for cultured cell characterization and suggest cortical progenitors will likely need to be cultured longer to yield upper-layer cortical neurons.

In summary, our observations suggest our culture system models the earliest steps of human brain development, including regional patterning, which is vital, because primary samples at these stages are exceedingly rare. This study represents an advance for the field in terms of breadth and depth of cell characterization and provides benchmark datasets to understand the origins and diseases of the human brain.

## EXPERIMENTAL PROCEDURES

### Genome Engineering and hESC Culture

H1 or H9 hESCs (WiCell) were maintained with mTeSR1 media (STEMCELL Technologies) on Matrigel (BD) or hES media (DMEM/F12 with 20% Knockout Serum Replacement [KSR]; Life Technologies) on CF-1 MEFs (GlobalStem). The TALEN genes targeting *SOX2*, *DCX*, and *OTX2* were made using the fast ligation-based automatable solid-phase high-throughput (FLASH) method (Reyon et al., 2012), and *PAX6* using the REAL method (Sander et al., 2011). Engineered lines were generated as previously described (Martinez et al., 2015). See *Supplemental Experimental Procedures* for more details.

### hESC Neural Differentiation

hESCs were seeded for a 12-day CI phase in NIM media with SMAD inhibitors (Chambers et al., 2009) and cyclopamine (Stemgent), reseeded for the progenitor expansion (PE) phase at days 12 and 19 in neural stem cell culture

media (NSCM) with EGF (Thermo Fisher) and bFGF (Ciccolini and Svendsen, 1998; Tropepe et al., 1999), and finally re-seeded at day 26 for ND with the neurogenic and neurotrophic factors BDNF (R&D Systems), GDNF (R&D Systems), NT3 (R&D Systems), and cAMP (Sigma) (Hu et al., 2010). Differentiations were validated by immunostaining at days 12, 26, and 54 and by flow cytometry at day 26. For more details, see *Supplemental Experimental Procedures*.

### Antibody Staining

Cells were fixed in 4% PFA for 15 min and blocked and permeabilized in 10% normal goat serum with 0.1% Triton X-100. Primary and secondary antibodies are listed in Table S3. Brain pieces were fixed overnight in 4% PFA/PBS at 4°C, cryoprotected in 30% sucrose in PBS for 24–48 hr, and then embedded in OCT. Cryosections were cut on the coronal plane at 20 μm thickness and stained as described above (see Table S3 for antibodies).

### Calcium Imaging

Cells were loaded with 4 μM FURA2-AM (Thermo Fisher) in ND at room temperature for 30 min. Ca<sup>2+</sup> activity was recorded using a 40× objective for 5-min intervals. Images were captured with 300-ms exposures at both 340 nm and 380 nm. Nikon NIS-Elements software was used to analyze events with measurements greater than 0.006 relative fluorescent units (RFUs) above baseline.

### Fetal Brain Tissue Processing

Human fetal tissue was received from the Laboratory of Developmental Biology who collected, staged, and dissected these tissues utilizing the University Washington IRB protocol 41557. Human fetal tissue was donated with written informed consent and in compliance with the Health and Human Services Code of Federal Regulations (CFR) 45 Part 46.101-124 (subparts B, D, E). Specimen age in days post conception was estimated by fetal foot length. Cortical tissue was identified by morphology, physically disrupted, and then enzymatically dissociated to single cells. Cells were then washed, filtered, counted, and fixed for storage at –80°C until processing by FRISCR. See *Supplemental Experimental Procedures* for more details.

### Single-Cell Transcriptomics

Single cells were sorted on a FACS Aria (BD) into 96-well collection plates and stored at –80°C. We prepared libraries as previously reported (Hashimshony et al., 2012), with a few modifications (see *Supplemental Experimental Procedures*). FRISCR was carried out as previously described (Thomsen et al., 2016), and SmartSeq2 sequencing libraries were prepared as previously reported (Picelli et al., 2013). Libraries were then quantified and sequenced on a HiSeq (Illumina). Details on mapping RNA-seq data, clustering to identify cell types, and comparing data between CelSeq and SmartSeq2, atlases, and FRISCR are provided in *Supplemental Experimental Procedures*. Metadata for each profiled single cell can be found in Data S2.

### Lineage Inference

We used a Bayesian method (S.R., unpublished data). Briefly, we analyzed all possible triplets of cell types present at adjacent time points, and then based on only transcription factor expression data, we used a Bayesian formulation to identify the highest-probability triplets with linked topologies. Triplets that showed strong evidence of topological linkages were assembled iteratively into a tree that was rooted at day 12. When we observed conflicting triplets, we selected the topology with higher probability. See *Supplemental Experimental Procedures* for more details.

### Viral Clonal Analysis

A barcoded retroviral library was constructed by inserting a random 10-bp barcode into the 3' end of *tdTomato* and then packaging it into retrovirus particles (pseudotyped with VSV-G). At day 27, 4 to 6 × 10<sup>3</sup> infectious units (IFUs) were inoculated per well of a 24-well plate (1.5 × 10<sup>6</sup> cells), to yield 20–50 *tdTomato*<sup>+</sup> colonies per well at day 54. Daily fluorescent images were taken for some clones to monitor expansion. Single cells were processed by SmartSeq2 as described above. Barcodes and transcriptomes were identified, and random forest classification was used to determine the cell type identity. See *Supplemental Experimental Procedures* for more details.

### Progenitor Potential Assay by Clonal Outgrowth

Single cells from day 26 were plated onto primary mouse astrocytes at clonal density (ten cells per well of a 96-well plate). Single-cell colonies were differentiated for 4 weeks in ND media and then fixed and analyzed by immunostaining (Table S3). See [Supplemental Experimental Procedures](#) for details.

### ACCESSION NUMBERS

The accession number for the raw and normalized population and single-cell RNA-seq data from hESC-derived cells is GEO: GSE86894. The accession number for primary fetal human single-cell RNA-seq data is dbGaP: phs001205.v1.p1.

### SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, seven figures, five tables, and two data sets and can be found with this article online at <http://dx.doi.org/10.1016/j.stem.2016.09.011>.

### AUTHOR CONTRIBUTIONS

Genome engineering was done by A.J., A.R.K., R.A.M., R.C.M., A.M.N., N.K.N., L.J.T., and M.W.S. and planned by J.S.G. and A.K. Stem cell banking was done by A.M.N., N.K.N., and R.A.M. Y.W. and S.Y. adapted the differentiation protocol. R.A.M., H.M., M.A.F., B.W.G., and Y.W. conducted all differentiations. Immunostaining was done by Y.W., H.M., M.F., and A.M.N. Calcium imaging was done by B.W.G. and R.A.M. and planned by Y.W. Live single-cell RNA-seq was piloted by B.P.L. and C.Y. and done by A.R.K. and E.R.T. Bioinformatic analysis of RNA-seq data was done by Z.Y. and V.M. C.L.T. analyzed developing mouse atlas data. The lineage algorithm was developed by L.F. and S.R. and applied by V.M., with help from S.M. Primary human tissue was obtained with help from I.A.G. and processed by J.K.M., R.D.H., and S.I.S. FRISCR analysis was conducted by E.R.T. Viral barcoding experiments were conducted by S.K. and non-viral clonal analysis by J.K.M. All sorting was done by N.V.S. and S.B. Experiments were conceived by J.S.G., B.P.L., Y.W., J.K.M., S.K., V.M., J.K.M., Z.Y., and S.R. The paper was written by J.K.M., B.P.L., Y.W., V.M., S.K., J.S.G., and S.R. S.R., C.L.T., and J.W.P. provided program leadership.

### ACKNOWLEDGMENTS

We wish to thank the Allen Institute founders, P.G. Allen and J. Allen, for their vision, encouragement, and support. We wish to thank A. Bernard and E. Lein for assistance with human specimen procurement, J. Miller for assistance in analysis of BrainSpan Atlas of the Developing Human Brain, and Luis Puelles for consultation on developmental neuroanatomy. Human primary samples were received from the Birth Defects Research Laboratory (BDRL). BDRL is supported by NIH award 5R24HD000836 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development. S.R. was supported in part by NIH directors pioneer award 5DP1MH099906-03 and National Science Foundation grant PHY-0952766.

Received: February 23, 2016

Revised: July 18, 2016

Accepted: September 29, 2016

Published: October 27, 2016

### REFERENCES

- Bakken, T.E., Miller, J.A., Ding, S.L., Sunkin, S.M., Smith, K.A., Ng, L., Szafer, A., Dalley, R.A., Royall, J.J., Lemon, T., et al. (2016). A comprehensive transcriptional map of primate brain development. *Nature* 535, 367–375.
- Bertacchi, M., Pandolfini, L., Murenu, E., Viegi, A., Capsoni, S., Cellerino, A., Messina, A., Casarosa, S., and Cremisi, F. (2013). The positional identity of mouse ES cell-generated neurons is affected by BMP signaling. *Cell. Mol. Life Sci.* 70, 1095–1111.
- Bulfone, A., Martinez, S., Marigo, V., Campanella, M., Basile, A., Quaderi, N., Gattuso, C., Rubenstein, J.L.R., and Ballabio, A. (1999). Expression pattern of the *Tbr2* (Eomesodermin) gene during mouse and chick brain development. *Mech. Dev.* 84, 133–138.
- Bystron, I., Rakic, P., Molnár, Z., and Blakemore, C. (2006). The first neurons of the human cerebral cortex. *Nat. Neurosci.* 9, 880–886.
- Camp, J.G., Badsha, F., Florio, M., Kanton, S., Gerber, T., Wilsch-Bräuner, M., Lewitus, E., Sykes, A., Hevers, W., Lancaster, M., et al. (2015). Human cerebral organoids recapitulate gene expression programs of fetal neocortex development. *Proc. Natl. Acad. Sci. USA* 112, 15672–15677.
- Chambers, S.M., Fasano, C.A., Papapetrou, E.P., Tomishima, M., Sadelain, M., and Studer, L. (2009). Highly efficient neural conversion of human ES and iPS cells by dual inhibition of SMAD signaling. *Nat. Biotechnol.* 27, 275–280.
- Ciccolini, F., and Svendsen, C.N. (1998). Fibroblast growth factor 2 (FGF-2) promotes acquisition of epidermal growth factor (EGF) responsiveness in mouse striatal precursor cells: identification of neural precursors responding to both EGF and FGF-2. *J. Neurosci.* 18, 7869–7880.
- Colasante, G., Lignani, G., Rubio, A., Medrihan, L., Yekhlef, L., Sessa, A., Massimino, L., Giannelli, S.G., Sacchetti, S., Caiazzo, M., et al. (2015). Rapid Conversion of Fibroblasts into Functional Forebrain GABAergic Interneurons by Direct Genetic Reprogramming. *Cell Stem Cell* 17, 719–734.
- Edri, R., Yaffe, Y., Ziller, M.J., Mutukula, N., Volkman, R., David, E., Jacob-Hirsch, J., Malcov, H., Levy, C., Rechavi, G., et al. (2015). Analysing human neural stem cell ontogeny by consecutive isolation of Notch active neural progenitors. *Nat. Commun.* 6, 6500.
- Espuny-Camacho, I., Michelsen, K.A., Gall, D., Linaro, D., Hasche, A., Bonnefont, J., Bali, C., Orduz, D., Bilheu, A., Herpoel, A., et al. (2013). Pyramidal neurons derived from human pluripotent stem cells integrate efficiently into mouse brain circuits *in vivo*. *Neuron* 77, 440–456.
- Florio, M., Albert, M., Taverna, E., Namba, T., Brandl, H., Lewitus, E., Haffner, C., Sykes, A., Wong, F.K., Peters, J., et al. (2015). Human-specific gene ARHGAP11B promotes basal progenitor amplification and neocortex expansion. *Science* 347, 1465–1470.
- Gaspard, N., Bouschet, T., Hourez, R., Dimidschstein, J., Naeije, G., van den Ameele, J., Espuny-Camacho, I., Herpoel, A., Passante, L., Schiffmann, S.N., et al. (2008). An intrinsic mechanism of corticogenesis from embryonic stem cells. *Nature* 455, 351–357.
- Ha, A., Perez-Iratxeta, C., Liu, H., Mears, A.J., and Wallace, V.A. (2012). Identification of Wnt/β-catenin modulated genes in the developing retina. *Mol. Vis.* 18, 645–656.
- Hanashima, C., Li, S.C., Shen, L., Lai, E., and Fishell, G. (2004). Foxg1 suppresses early cortical cell fate. *Science* 303, 56–59.
- Hashimshony, T., Wagner, F., Sher, N., and Yanai, I. (2012). CEL-Seq: single-cell RNA-Seq by multiplexed linear amplification. *Cell Rep.* 2, 666–673.
- Hevner, R.F., Neogi, T., Englund, C., Daza, R.A., and Fink, A. (2003). Cajal-Retzius cells in the mouse: transcription factors, neurotransmitters, and birthdays suggest a pallial origin. *Brain Res. Dev. Brain Res.* 141, 39–53.
- Hirata, T., Suda, Y., Nakao, K., Narimatsu, M., Hirano, T., and Hibi, M. (2004). Zinc finger gene fez-like functions in the formation of subplate neurons and thalamocortical axons. *Dev. Dyn.* 230, 546–556.
- Hook, V., Brennan, K.J., Kim, Y., Toneff, T., Funkelstein, L., Lee, K.C., Ziegler, M., and Gage, F.H. (2014). Human iPSC neurons display activity-dependent neurotransmitter secretion: aberrant catecholamine levels in schizophrenia neurons. *Stem Cell Reports* 3, 531–538.
- Hu, B.Y., Weick, J.P., Yu, J., Ma, L.X., Zhang, X.Q., Thomson, J.A., and Zhang, S.C. (2010). Neural differentiation of human induced pluripotent stem cells follows developmental principles but with variable potency. *Proc. Natl. Acad. Sci. USA* 107, 4335–4340.
- Klein, A.M., Mazutis, L., Akartuna, I., Tallapragada, N., Veres, A., Li, V., Peshkin, L., Weitz, D.A., and Kirschner, M.W. (2015). Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell* 161, 1187–1201.

- Konopka, G., Friedrich, T., Davis-Turak, J., Winden, K., Oldham, M.C., Gao, F., Chen, L., Wang, G.Z., Luo, R., Preuss, T.M., and Geschwind, D.H. (2012). Human-specific transcriptional networks in the brain. *Neuron* **75**, 601–617.
- Lancaster, M.A., Renner, M., Martin, C.A., Wenzel, D., Bicknell, L.S., Hurles, M.E., Homfray, T., Penninger, J.M., Jackson, A.P., and Knoblich, J.A. (2013). Cerebral organoids model human brain development and microcephaly. *Nature* **501**, 373–379.
- Langfelder, P., and Horvath, S. (2007). Eigengene networks for studying the relationships between co-expression modules. *BMC Syst. Biol.* **1**, 54.
- Langfelder, P., Luo, R., Oldham, M.C., and Horvath, S. (2011). Is my network module preserved and reproducible? *PLoS Comput. Biol.* **7**, e1001057.
- Lui, J.H., Nowakowski, T.J., Pollen, A.A., Javaherian, A., Kriegstein, A.R., and Oldham, M.C. (2014). Radial glia require PDGFD-PDGFR $\beta$  signalling in human but not mouse neocortex. *Nature* **515**, 264–268.
- Macosko, E.Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, A.R., Kamitaki, N., Martersteck, E.M., et al. (2015). Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* **161**, 1202–1214.
- Mariani, J., Simonini, M.V., Palejev, D., Tomasini, L., Coppola, G., Szekely, A.M., Horvath, T.L., and Vaccarino, F.M. (2012). Modeling human cortical development in vitro using induced pluripotent stem cells. *Proc. Natl. Acad. Sci. USA* **109**, 12770–12775.
- Maroof, A.M., Keros, S., Tyson, J.A., Ying, S.W., Ganat, Y.M., Merkle, F.T., Liu, B., Goulburn, A., Stanley, E.G., Elefantiy, A.G., et al. (2013). Directed differentiation and functional maturation of cortical interneurons from human embryonic stem cells. *Cell Stem Cell* **12**, 559–572.
- Martinez, R.A., Stein, J.L., Krostag, A.R., Nelson, A.M., Marken, J.S., Menon, V., May, R.C., Yao, Z., Kaykas, A., Geschwind, D.H., and Grimley, J.S. (2015). Genome engineering of isogenic human ES cells to model autism disorders. *Nucleic Acids Res.* **43**, e65.
- Maury, Y., Côme, J., Piskorowski, R.A., Salah-Mohellibi, N., Chevaleyre, V., Peschanski, M., Martinat, C., and Nedelec, S. (2015). Combinatorial analysis of developmental cues efficiently converts human pluripotent stem cells into multiple neuronal subtypes. *Nat. Biotechnol.* **33**, 89–96.
- Miller, J.C., Tan, S., Qiao, G., Barlow, K.A., Wang, J., Xia, D.F., Meng, X., Paschon, D.E., Leung, E., Hinkley, S.J., et al. (2011). A TALE nuclease architecture for efficient genome editing. *Nat. Biotechnol.* **29**, 143–148.
- 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.
- Paul, F., Arkin, Y., Giladi, A., Jaitin, D.A., Kenigsberg, E., Keren-Shaul, H., Winter, D., Lara-Astiaso, D., Gury, M., Weiner, A., et al. (2015). Transcriptional heterogeneity and lineage commitment in myeloid progenitors. *Cell* **163**, 1663–1677.
- Picelli, S., Björklund, A.K., Faridani, O.R., Sagasser, S., Winberg, G., and Sandberg, R. (2013). Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nat. Methods* **10**, 1096–1098.
- Pollen, A.A., Nowakowski, T.J., Shuga, J., Wang, X., Leyrat, A.A., Lui, J.H., Li, N., Szpankowski, L., Fowler, B., Chen, P., et al. (2014). Low-coverage single-cell mRNA sequencing reveals cellular heterogeneity and activated signaling pathways in developing cerebral cortex. *Nat. Biotechnol.* **32**, 1053–1058.
- 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.
- Porter, F.D., Drago, J., Xu, Y., Cheema, S.S., Wassif, C., Huang, S.P., Lee, E., Grinberg, A., Massalas, J.S., Bodine, D., et al. (1997). Lhx2, a LIM homeobox gene, is required for eye, forebrain, and definitive erythrocyte development. *Development* **124**, 2935–2944.
- Qian, X., Nguyen, H.N., Song, M.M., Hadiono, C., Ogden, S.C., Hammack, C., Yao, B., Hamersky, G.R., Jacob, F., Zhong, C., et al. (2016). Brain-region-specific organoids using mini-bioreactors for modeling ZIKV exposure. *Cell* **165**, 1238–1254.
- Reilly, S.K., Yin, J., Ayoub, A.E., Emera, D., Leng, J., Cotney, J., Sarro, R., Rakic, P., and Noonan, J.P. (2015). Evolutionary genomics. Evolutionary changes in promoter and enhancer activity during human corticogenesis. *Science* **347**, 1155–1159.
- Reyon, D., Tsai, S.Q., Khayter, C., Foden, J.A., Sander, J.D., and Joung, J.K. (2012). FLASH assembly of TALENs for high-throughput genome editing. *Nat. Biotechnol.* **30**, 460–465.
- Ricciardi, S., Ungaro, F., Hambrock, M., Rademacher, N., Stefanelli, G., Brambilla, D., Sessa, A., Magagnotti, C., Bach, A., Giarda, E., et al. (2012). CDKL5 ensures excitatory synapse stability by reinforcing NGL-1-PSD95 interaction in the postsynaptic compartment and is impaired in patient iPSC-derived neurons. *Nat. Cell Biol.* **14**, 911–923.
- Sander, J.D., Cade, L., Khayter, C., Reyon, D., Peterson, R.T., Joung, J.K., and Yeh, J.R. (2011). Targeted gene disruption in somatic zebrafish cells using engineered TALENs. *Nat. Biotechnol.* **29**, 697–698.
- Schüller, U., and Rowitch, D.H. (2007). Beta-catenin function is required for cerebellar morphogenesis. *Brain Res.* **1140**, 161–169.
- Shi, Y., Kirwan, P., Smith, J., Robinson, H.P., and Livesey, F.J. (2012). Human cerebral cortex development from pluripotent stem cells to functional excitatory synapses. *Nat. Neurosci.* **15**, 477–486.
- Shin, J., Berg, D.A., Zhu, Y., Shin, J.Y., Song, J., Bonaguidi, M.A., Enikolopov, G., Nauen, D.W., Christian, K.M., Ming, G.L., and Song, H. (2015). Single-cell RNA-seq with Waterfall reveals molecular cascades underlying adult neurogenesis. *Cell Stem Cell* **17**, 360–372.
- 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.
- Takahashi, K., and Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663–676.
- Tasic, B., Menon, V., Nguyen, T.N., Kim, T.K., Jarsky, T., Yao, Z., Levi, B., Gray, L.T., Sorensen, S.A., Dolbeare, T., et al. (2016). Adult mouse cortical cell taxonomy revealed by single cell transcriptomics. *Nat. Neurosci.* **19**, 335–346.
- Thompson, C.L., Ng, L., Menon, V., Martinez, S., Lee, C.K., Glattfelder, K., Sunkin, S.M., Henry, A., Lau, C., Dang, C., et al. (2014). A high-resolution spatiotemporal atlas of gene expression of the developing mouse brain. *Neuron* **83**, 309–323.
- 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.
- Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., Lennon, N.J., Livak, K.J., Mikkelsen, T.S., and Rinn, J.L. (2014). The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* **32**, 381–386.
- Treutlein, B., Brownfield, D.G., Wu, A.R., Neff, N.F., Mantalas, G.L., Espinoza, F.H., Desai, T.J., Krasnow, M.A., and Quake, S.R. (2014). Reconstructing lineage hierarchies of the distal lung epithelium using single-cell RNA-seq. *Nature* **509**, 371–375.
- Tropepe, V., Sibilia, M., Ciruna, B.G., Rossant, J., Wagner, E.F., and van der Kooy, D. (1999). Distinct neural stem cells proliferate in response to EGF and FGF in the developing mouse telencephalon. *Dev. Biol.* **208**, 166–188.
- van de Leemput, J., Boles, N.C., Kiehl, T.R., Corneo, B., Lederman, P., Menon, V., Lee, C., Martinez, R.A., Levi, B.P., Thompson, C.L., et al. (2014). CORTECON: a temporal transcriptome analysis of in vitro human cerebral cortex development from human embryonic stem cells. *Neuron* **83**, 51–68.
- Vierbuchen, T., Ostermeier, A., Pang, Z.P., Kokubu, Y., Südhof, T.C., and Wernig, M. (2010). Direct conversion of fibroblasts to functional neurons by defined factors. *Nature* **463**, 1035–1041.
- Zeisel, A., Muñoz-Manchado, A.B., Codeluppi, S., Lönnberg, P., La Manno, G., Juréus, A., Marques, S., Munguba, H., He, L., Betsholtz, C., et al. (2015). Brain structure. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. *Science* **347**, 1138–1142.
- Zhang, B., and Horvath, S. (2005). A general framework for weighted gene co-expression network analysis. *Stat. Appl. Genet. Mol. Biol.* **4**, Article17.