Yao Figure S1

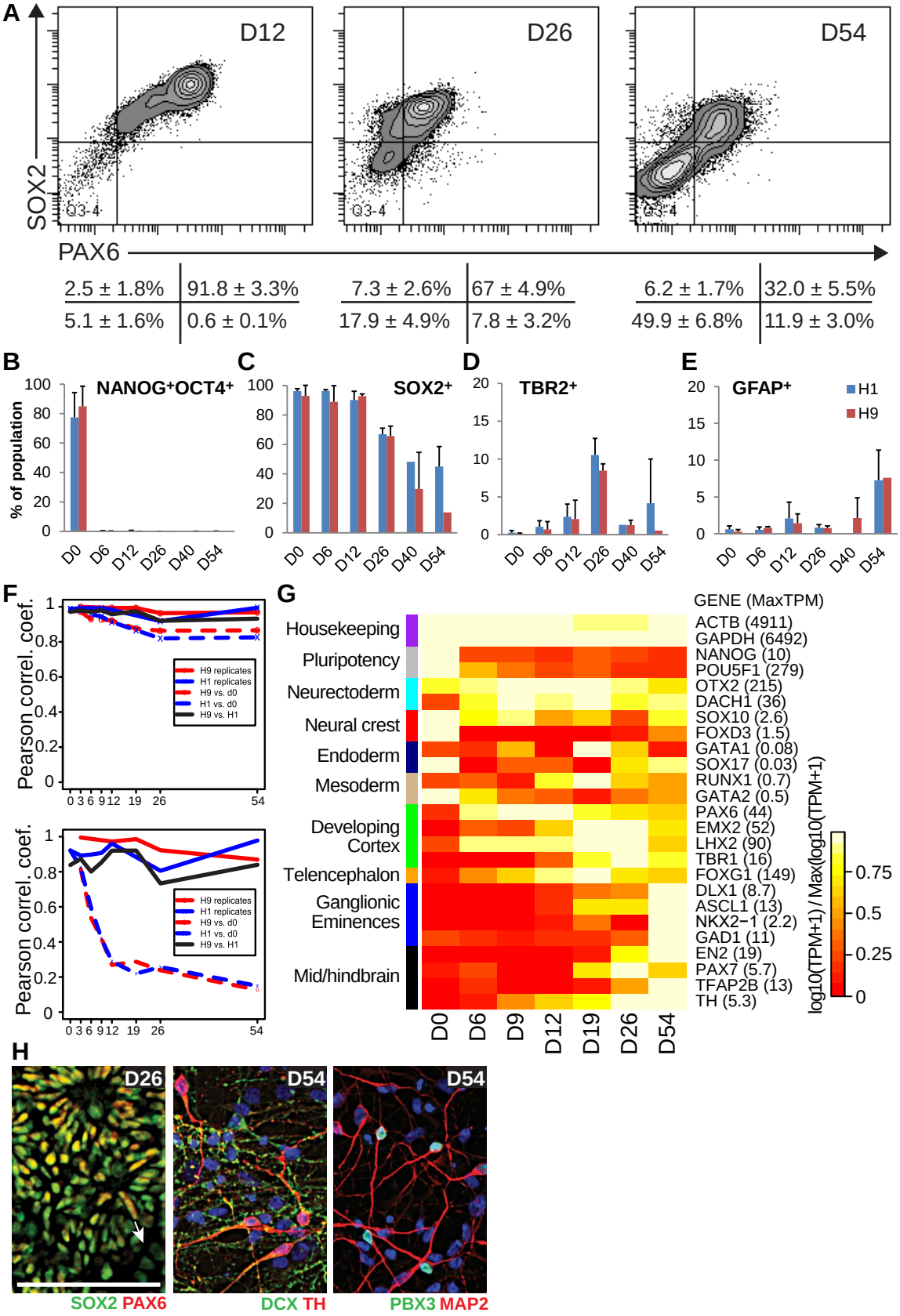

Figure S1: Reproducible generation of cortical and non-cortical cells. Related to Figure 1. (A) SOX2 and PAX6 flow staining shows $SOX2^{+}PAX6^{-}$ cells increase from D12 (2.5 ± 1.8%) to D26 (7.3 ± 2.6%). (B-E) Quantitation of flow cytometry analysis showing percentage of cells positive for the indicated antibody stain. Data is mean ± SD where three different differentiations were profiled. H1 D40 and H9 D54 represent one and two biological samples, respectively. (F) Plots showing Pearson's correlation between indicated samples for all genes (top) and differentially expressed genes (bottom) profiled from all cells of a well (~1×10^{6} cells) from indicated stages by RNA-Seq. All samples were generated from at least two differentiations. (G) Gene expression as in Figure 1E but normalized the maximum expression level of each gene. (H) Representative images of immunostaining indicated non-cortical cells that are labeled by SOX2 but not PAX6 at D26 (arrow). Expression of TH or PBX3 with DCX or MAP2 at D54. Scale bar 100 μm.

Yao Figure S2

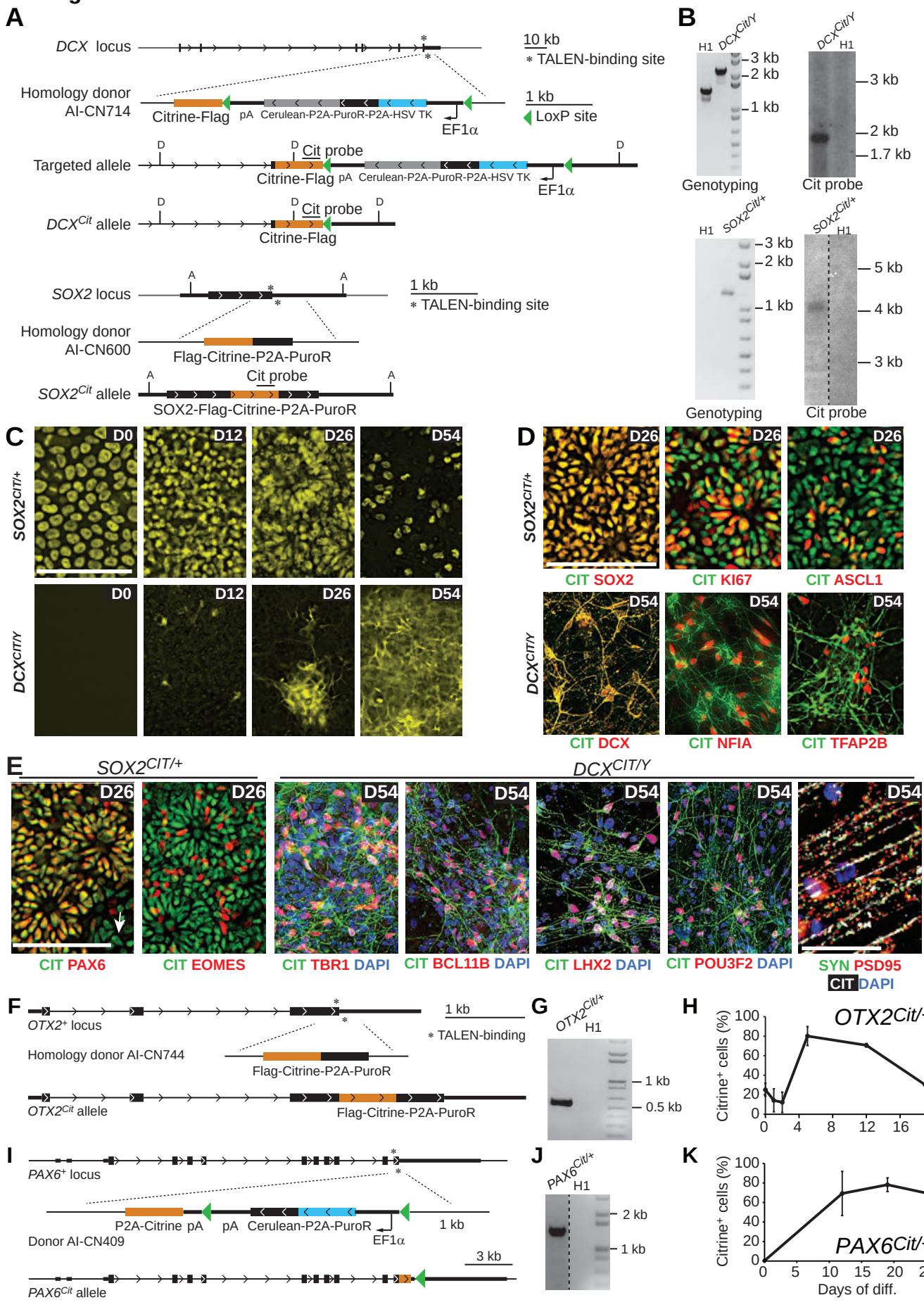

Figure S2: DCX, SOX2, OTX2 and PAX6 reporter lines. Related to Figure 1. (A) The targeting strategy to generate *DCX*$^{Cit/Y}$ and *SOX2*$^{Cit/+}$ reporter cell lines. (B) Long-range genotyping PCR and Southern blot of the resulting lines. The dashed line denotes a region of the blot that was deleted to place the control lane adjacent to the stem cell clone. Restriction enzymes (A, AflII; D, DraIII) are indicated. (C) Endogenous citrine expression from *DCX*$^{Cit/Y}$ and *SOX2*$^{Cit/+}$ reporter cell lines at D0, D12, D26 and D54. (D) Additional immunostaining on *DCX*$^{Cit/Y}$ and *SOX2*$^{Cit/+}$ reporter cell lines. All scale bars: 100 μm. (E) Representative images of immunostaining of D26 SOX2$^{Cit/+}$ and D54 DCX$^{Cit/+}$ reporter lines. Arrow marks cells that express SOX2 but not PAX6. Scale bar: 100 μm in B; 100 μm in H except for SYN/PSD/CIT/DAPI micrograph where scale bar is 25 μm. Transgenic *OTX2*$^{Cit/+}$ and *PAX6*$^{Cit/+}$ H1 human embryonic stem cell reporter lines. (F) The targeting strategy to generate the *OTX2*$^{Cit/+}$ reporter line with (G) 5' junction genotyping PCR and (H) percent citrine positive cells +/- SD from *OTX2*$^{Cit/+}$ lines measured by flow cytometry (n = 3). (I) the targeting strategy to generate the *PAX6*$^{Cit/+}$ reporter line with (J) 5' junction genotyping PCR. The dashed line denotes a portion of the gel removed for visual clarity. (K) Percent citrine positive cells +/- SD from the *PAX6*$^{Cit/+}$ line measured by flow cytometry (n = 3).

Yao Figure S3

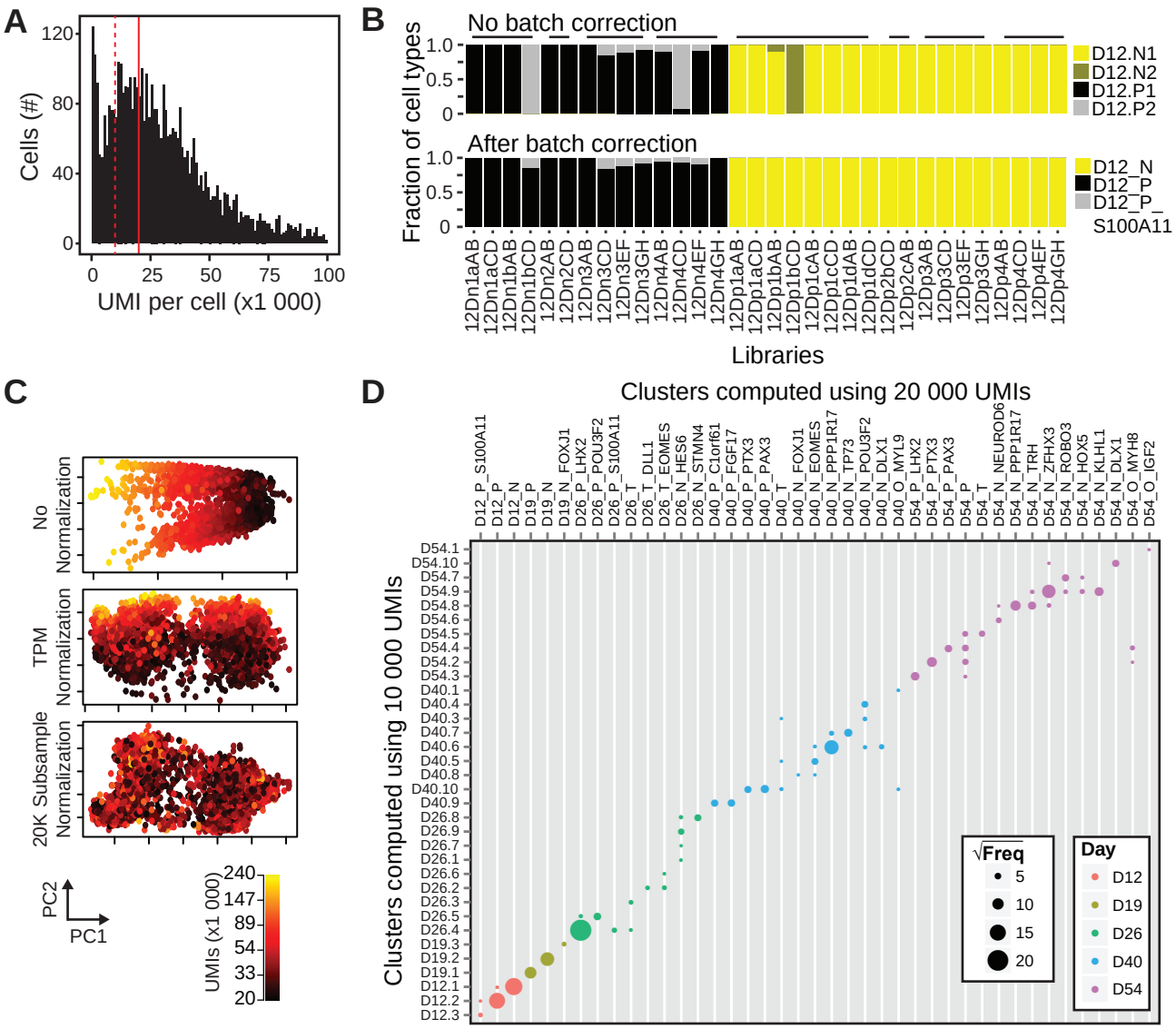

Figure S3. Batch correction and subsampling of CelSeq data. Related to Figure 2. (A) Histogram showing number of transcripts (UMI) per cell profiled by CelSeq (4368 cells). The red line indicates the threshold for inclusion in the study (>20,000 UMIs/cell) while the dashed red line indicated 10,000 UMIs/cell. (B) Barplot showing the number of cell types per D12 library before and after batch correction. Each library consists of 22 cells and bars across top indicate libraries from a single plate of sorted cells. (C) PCA analysis of all cells, dark color indicates low read, number while ligher indicates high read number. Only subsampling removes the clustering by read number. (D) Cell cluster correspondence between cells subsampled to 10,000 UMI and cells subsampled to 20,000 UMIs. Most 20,000 UMI subsampled clusters mapped to one or two 10,000 UMI clusters.

Yao Figure S4

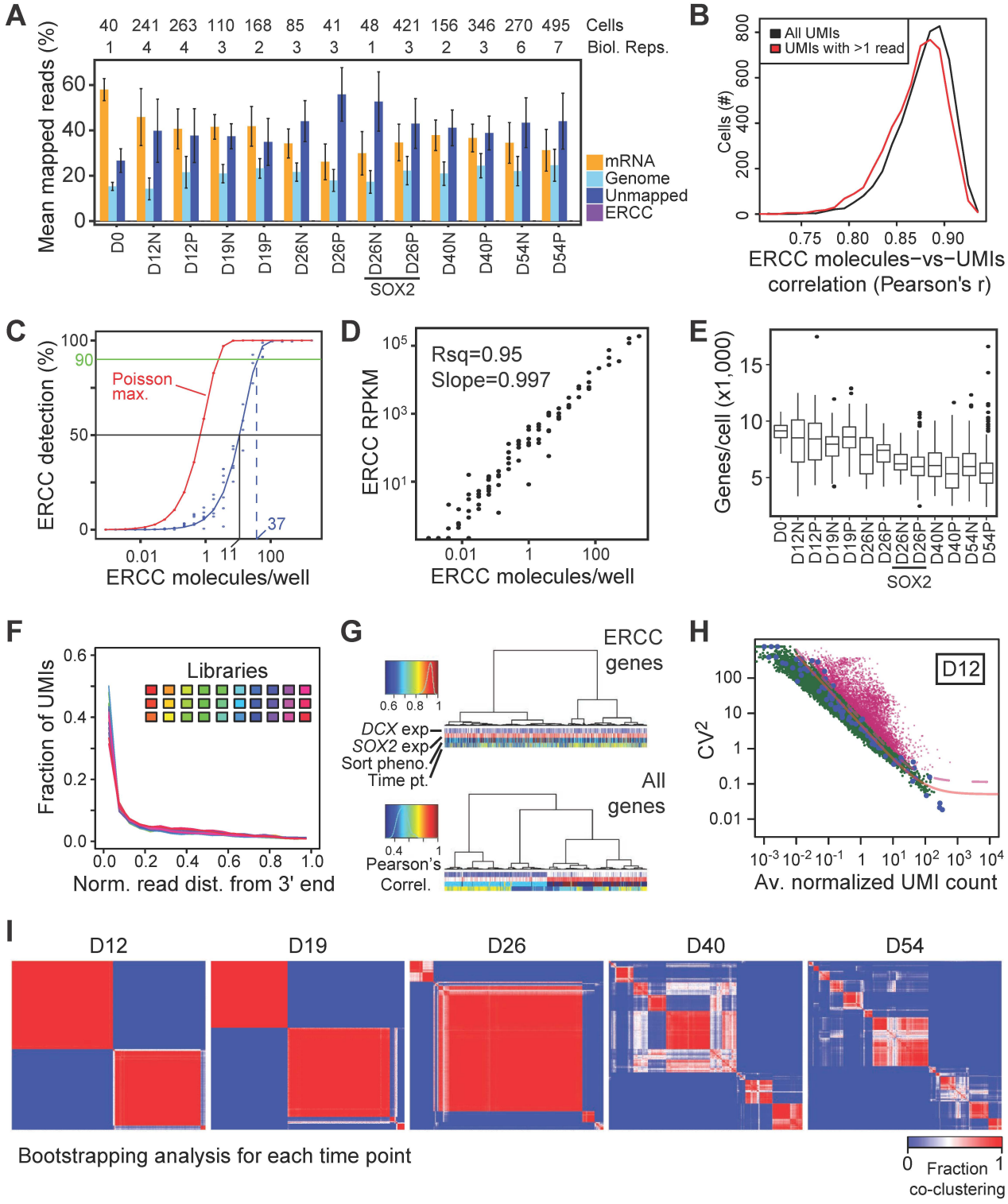

Figure S4. CelSeq single-cell RNA-Seq metrics. Related to Figure 2. (A) Read statistics for 2,684 included CelSeq samples, bar plot shows mean ± SD. Numbers of cells and biological replicates included in analysis are shown above the plot. (B) Pearson's correlation between known ERCC spike-in number and UMI detection across all single cells based on all ERCC UMIs (black), or UMIs represented by >1 read. (C) Sensitivity of CelSeq measured by ERCCs. Poisson estimate for complete detection (red), and actual ERCC detection (blue). ERCC data is the average of all wells for each spiked-in RNA species. A transcript at 11 and 37 copies per cell will be detected 50% and 90% of the time, respectively. (D) ERCC RPKM scatter plot shows linear amplification. (E) Boxplot showing the number of genes detected ≥ 1 UMI per cell for indicated timepoint and populations. (F) 3'gene distribution of all detected genes by library for all libraries in the study (each colored differently) show no batch bias in read distribution. (G) Cell-cell Pearson's correlation coefficient is shown for (top) ERCCs only and (bottom) all genes. (H) Representative variance versus expression plot for all D12 single cells. Blue dots are ERCCs, and genes higher than the variance threshold (dashed line) are shown as pink dots. (I) Bootstrapping co-clustering results for each time point as shown in Figure 2D.

Yao Figure S5

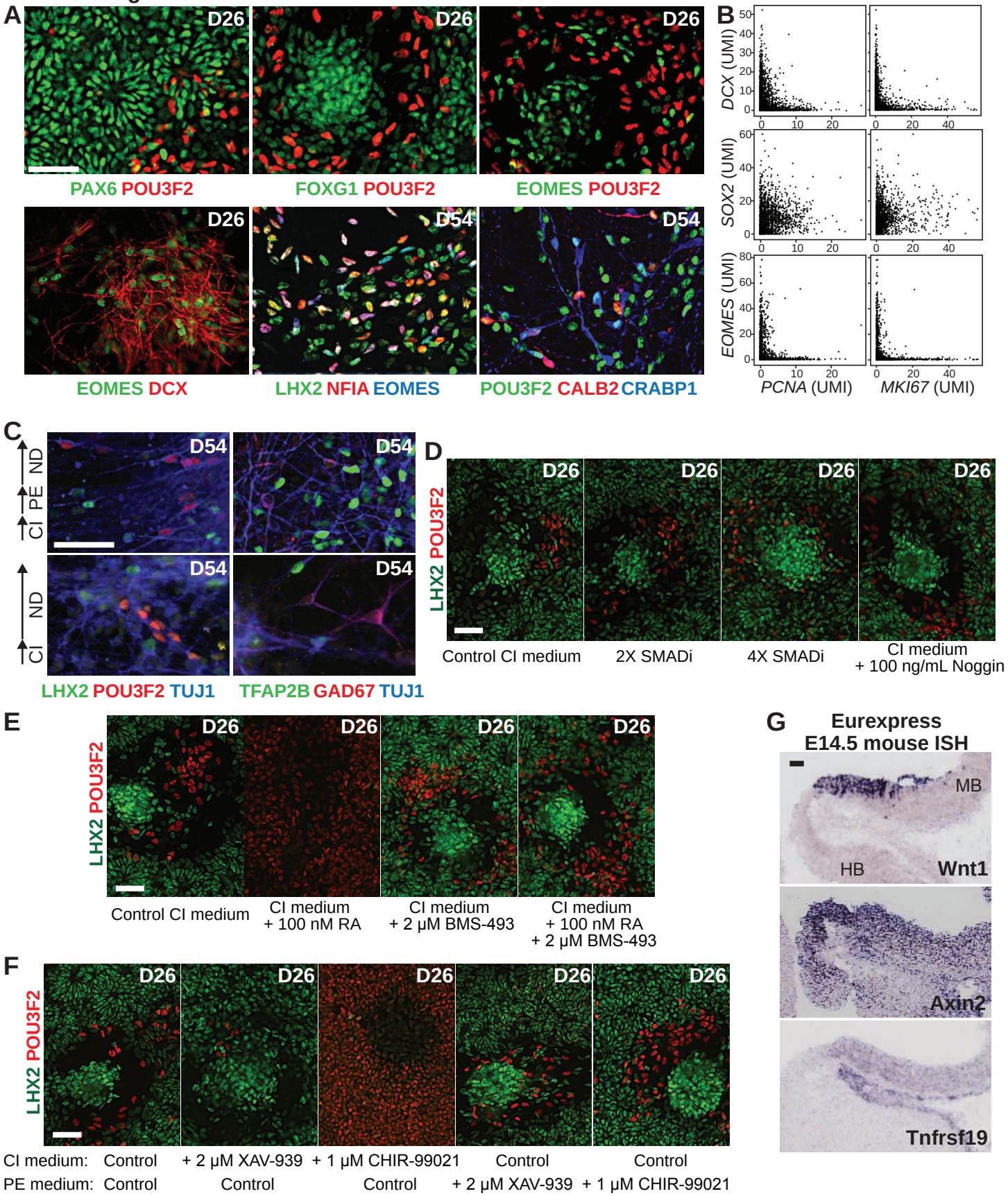

Figure S5: Immunostaining of cell-type markers and testing differentiation conditions. Related to Figures 3 and 6. (A) Immunocharacterization of cell types using markers identified from our single cell RNA-Seq analysis. *Top:* POU3F2 is rarely co-expressed with forebrain markers. *Bottom:* EOMES is expressed in DCX^+ cells at D26 and co-expressed with some NFIA and LHX2 expressing cells at D54. POU3F2 is also co-expressed with some CALB2 cells and some CRABP1 cells. (B) Cell cycle gene (*MKI67* and *PCNA*) co-expression with progenitor (*SOX2*) and neuronal (*DCX*) marker genes and *EOMES*. Each dot represents one cell and the values are the number of detected transcripts (UMIs) per cell. (C) The chief neuron types classes are generated regardless of the growth factor-containing PE phase. Parallel differentiations were performed with CI, PE, and ND phases (control differentiation), or by omitting PE phase and analyzing by immunostaining for forebrain ($LHX2^+$), mid/hindbrain ($POU3F2^+$ and $TFAP2B^+$), and inhibitory ($GAD67^+$) at D54. (D) D26 $POU3F2^+$ progenitors are unaffected by increased blockade of BMP and/or TGF-beta signaling pathways. Parallel differentiations were performed with control CI medium (5 μM SB431542 and 50 nM LDN193189), 2x SMAD inhibition (10 μM SB431542 and 100 nM LDN193189), 4x SMAD inhibition (20 μM SB431542 and 200 nM LDN193189), or 100 ng/mL recombinant Noggin protein added. (E) D26 $POU3F2^+$ progenitors are unaffected by blocking retinoic acid signaling (2 μM BMS-493) although they can be generated by exogenous retinoic acid (100 nM RA) which is blockable by BMS-493. (F) Canonical Wnt/beta-catenin pathway is active during CI phase but not during PE phase to induce mid/hindbrain identity. 2 μM XAV-939 inhibits (and 1 μM CHIR-99021 promotes) $POU3F2^+$ progenitor formation only when added during CI phase and not during PE phase. (G) Canonical Wnt/beta-catenin pathway target genes *Axin2* and *Tnfrsf19* are expressed in a Wnt1-responsive pattern near the mid/hindbrain boundary in mouse. Images of E14.5 mouse mid/hindbrain region are downloaded from the Eurexpress ISH database (Diez-Roux et al., 2011). Scale bar for all images is 50 μm..

Yao Figure S6

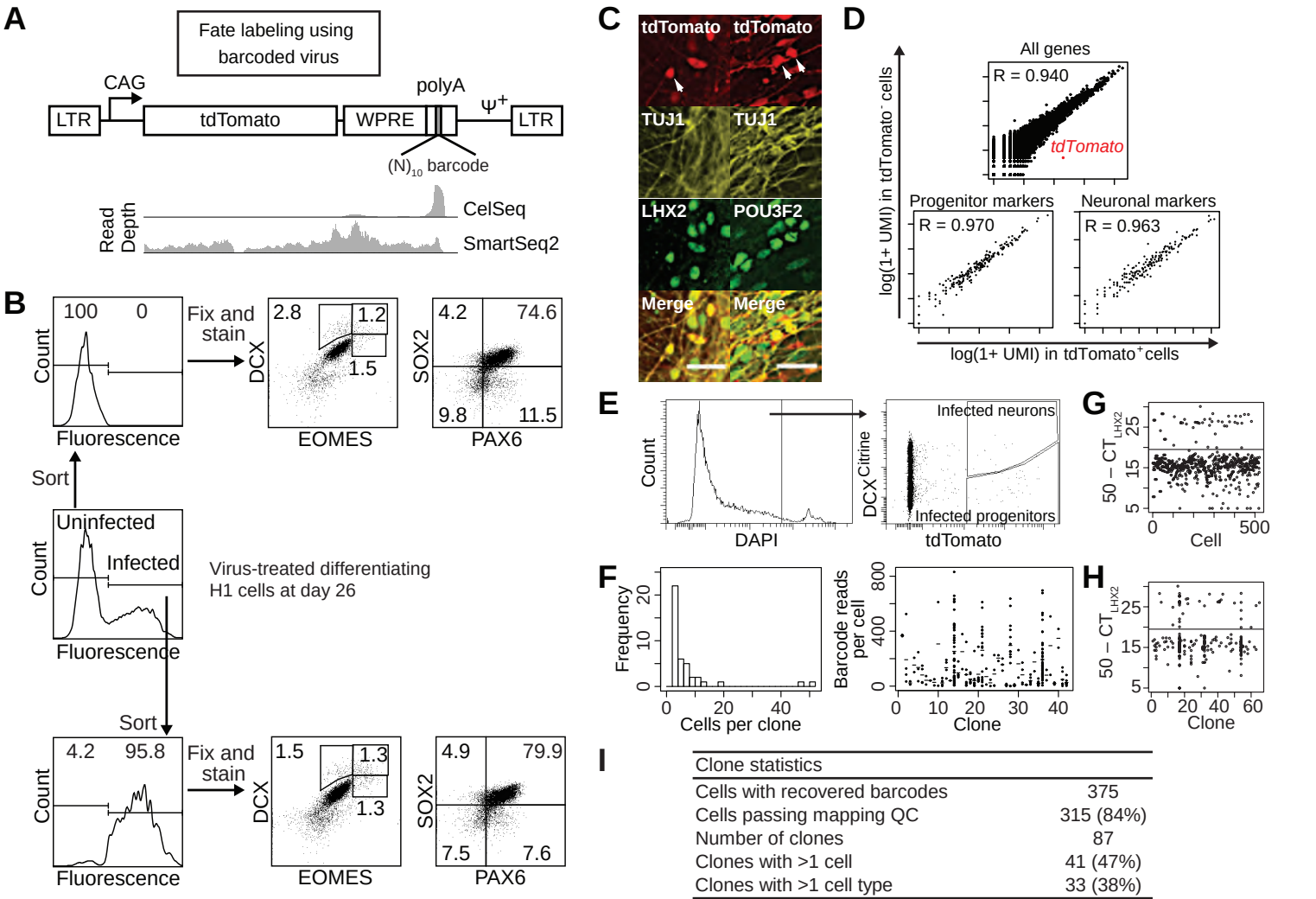

Figure S6. Techniques for clonal analysis of lineage fate and potential from hESC-derived neural progenitors. Related to figure 7. Data are shown from barcoded virus fate labeling (A-I) and clonal potential outgrowths (J-P). (A) Schematic of tdTomato-expressing barcoded virus for fate labeling studies. Also shown are sequence read coverages of the viral transcript using both CelSeq and SmartSeq2 in a representative single cell. Abbreviations are: LTR, long terminal repeat; CAG, chicken alpha globin; WPRE, woodchuck hepatitis virus posttranscriptional regulatory element; Ψ^+, retroviral packaging element. (B) Sorting/staining strategy for assessing viral tropism. Differentiating H1 cells infected at D24 were sorted by fluorescence level at D26, reanalyzed to confirm proper sorting, then fixed and stained with antibodies against DCX, EOMES, SOX2, and PAX6. Population percentages of the indicated gates are shown. (C) Fluorescence micrographs of clones with tdTomato co-expressed with LHX2 or POU3F2. Scale bar: 25 μm. (D) Correlation of gene expression (measured by RNA-Seq) between bulk population samples of D54 tdTomato$^+$ cells (infected at D27) and tdTomato$^-$ cells (uninfected); libraries were generated by CelSeq. Correlations are shown between infected and uninfected cells using all genes (*top*), progenitor gene markers (*bottom left*), and neuronal gene markers (*bottom right*). Differential expression of the *tdTomato* viral transcript is highlighted in red. (E) Gating strategy for sorting single tdTomato$^+$ D54 cells for sequencing. tdTomato is multiplexed with the *DCX^{Cit}* reporter to select for infected neurons or progenitors. (F) Barcoded clone read statistics. *Left* shows a histogram of number of cells detected per clone, and *right* shows the number of barcoded *tdTomato* reads per cell within clones. (G) Quantitative RT-PCR of *LHX2* on single cells and (H) clones conducted on amplified single cell cDNA libraries. (I) Summary statistics for barcoded clones. (J) Representative well of clonally outgrown day 54 colonies from day 26 progenitors seeded on astrocytes at 10 cells per well in a 96 well plate, stained for HNA + TuJ1 to locate human cells and neurons. Most wells have spatially separated colonies as shown. Scale 1 mm (full) or 200 μm (insets). (K) Histogram of approximate colony sizes by cell number. (L-P) Representative LHX2-containing and non-LHX2-containing colonies stained for HNA + TuJ1 (blue), LHX2 (green). Shown in red are: NFIA (L), FOXP2 (M), POU3F2 (N), CALB2 (Calretinin, O) or CRABP1 (P). Scale 50 μm. In L, arrows mark NFIA+TuJ1+ human neurons, but NFIA also marks LHX2+TuJ1- human progenitors as well as HNA- mouse astrocytes. The stripcharts compare the subtype marker incidences in LHX2-containing colonies to their incidences in non-LHX2-containing colonies by unpaired t-test in five independent differentiations. * $P < .05$, n.s., not significant ($P > .05$).

Yao Figure S7

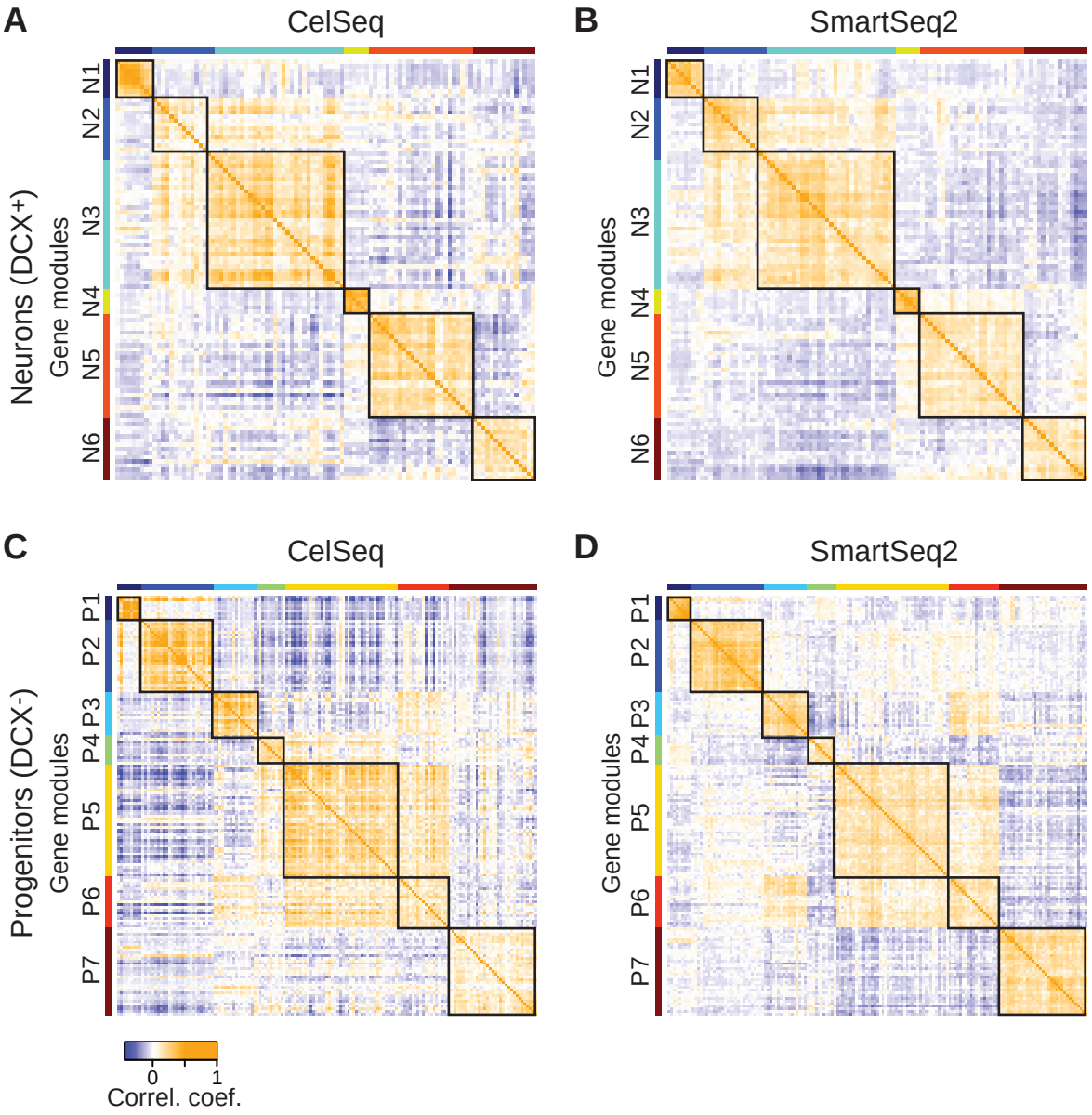

Figure S7. Mapping SmartSeq2 data to CelSeq cell types. Related to Figure 7. Heatmaps showing Pearson correlation of marker genes that are conserved between CelSeq (A,C) and SmartSeq2 (B,D) in DCX^+ (A,B) or DCX^- (C,D) cells from D54 cultures. Module representation of genes is shown with color bar and modules were derived using WGCNA from high variance genes from the CelSeq and SmartSeq2 datasets. Thus, N5 and N6, and P5 and P6 aren't distinct in the heatmap.

SUPPLEMENTAL TABLES LEGENDS

Table S1: List of module-associated genes. Related to Figures 2, 4 and S7. Excel file contains multiple tabs that show genes names and module identities that are associated with the analysis.The first tab includes gene excluded from clustering analysis. Abbreviations are: CC/M= gene ontology (GO) enriched for synthesis phase of cell cycle, M=GO enriched mitosis, RP= GO enriched for ribosomal-associated genes, MTRNR=MTRNR family genes, LINE= genes with highly repetitive LINE elements, TUBB=module marked by *TUBB* gene, Steroid=GO enrichment in steroid biology.

Table S2: List of all clonally related cells identified by viral barcoding. Related to Figure 7 and S6. Table shows the clone number, barcode sequence, cell name, number of barcode reads, that best mapping cell type, and the probability of that mapping.

Table S3: List antibodies used in the study. Related to Experimental Procedures.

Primary antibodies	Dilution	Species	Isotype	Vendor	Catalogue #
BCL11B	1:500	Rat	IgG	Abcam	ab18465
Calretinin	1:1000	Mouse	IgG1	EMD Millipore	MAB1568
Calretinin	1:1000	Rabbit	IgG	EMD Millipore	AB5054
CRABP1	1:1000	Mouse	IgG2b	Abcam	ab2816
CRYAB	1:500	Mouse	IgG1	Abcam	ab13496
DCX	1:500	Rabbit	IgG	Abcam	ab18723
EOMES	1:500	Chicken	IgY	EMD Millipore	ab15894
EOMES	1:500	Rabbit	IgG	Abcam	ab23345
FOXG1	1:2000	Rabbit	IgG	Abcam	ab18259
FOXP2	1:4000	Rabbit	IgG	Abcam	AB16046
GAD67, cl. 1G10.2	1:250	Mouse	IgG2a	EMD Millipore	MAB5406
GFAP	1:500	Rabbit	IgG	Abcam	ab7260
GFP	1:2000	Chicken	IgY	Abcam	ab13970
HNA	1:1000	Mouse	IgG1	EMD Millipore	MAB1281
HOPX	1:100	Rabbit	IgG	Santa Cruz Biotechnology, Inc.	sc-30216
Ki67	1:1000	Mouse	$IgG1_k$	BD Pharmingen	550609
Ki67	1:1000	Rabbit	IgG	Abcam	ab15580
LHX2 cl.6G2	1:500	Mouse	IgG1	Abcam	ab130256
LHX2 cl.6G2	1:1000	Mouse	IgG1	Thermo Fisher	MA5-15834
MAP2	1:1000	Mouse	IgG1	Sigma-Aldrich	M2320
MAP2	1:1000	Rabbit	IgG	Novus Biologicals	NBP1-40606
MASH1	1:500	Mouse	IgG1	BD Pharmingen	556604
Neu N	1:250	Chicken	IgY	Aves Labs, Inc.	NUN
Neu N	1:100	Mouse	IgG1	EMD Millipore	MAB377
NF1A	1:1000	Rabbit	serum	Active Motif	39397
OTX2	1:1000	Rabbit	IgG	EMD Millipore	AB9566
PAX6, cl. P3U1	1:50	Mouse	IgG1	DSHB	PAX6
PBX3	1:1000	Rabbit	IgG	Abcam	ab183849
POU3F2	1:3000	Rabbit	IgG	Cell Signalling	12137
POU3F2, cl.8C4.2	1:500	Mouse	$IgG1_k$	EMD Millipore	MABD51

PSD-95, cl. 108E10	1:300	Mouse	IgG	Synaptic Systems	124011
Reelin, cl. 142	1:1000	Mouse	IgG1	EMD Millipore	MAB5366
ROBO3	1:200	Goat	IgG	R&D Systems	AF3076
SATB2, cl. SATBA4B10	1:100	Mouse	IgG1	Abcam	Ab51502
SOX2	1:200	Rabbit	IgG	EMD Millipore	ab5603
Synaptophysin	1:1000	Rabbit	IgG	Abcam	ab68851
TBR1	1:1000	Rabbit	IgG	Abcam	ab31940
TFAP2B	1:250	Rabbit	IgG	Cell Signalling	2509
TH	1:1000	Rabbit	IgG	Pel-Freez	P40101-150
TH	1:3000	Mouse	IgG1	ImmunoStar	22941
TUJ1	1:1000	Mouse	IgG2a	BioLegend	801201
TUJ1	1:500	Rabbit	IgG	Covance	MRB-435P

Secondary antibodies	**Fluorophore**	**Isotype target**	**Vendor**	**Cat #**
Alexa Fluor Donkey anti-Goat	488	IgG	Thermo Fisher	A11055
Alexa Fluor Goat anti-Chicken	488	IgG	Thermo Fisher	A11039
Alexa Fluor Goat anti-Chicken	594	IgG	Thermo Fisher	A11042
Alexa Fluor Goat anti-Chicken	647	IgG	Thermo Fisher	A21449
Alexa Fluor Goat anti-Mouse	350	IgG	Thermo Fisher	A11045
Goat anti-Mouse	BV421	IgG	Becton Dickinson	563846
Goat anti-Mouse	BV480	IgG	Becton Dickinson	564877
Alexa Fluor Goat anti-Mouse	488	IgG	Thermo Fisher	A11001
Alexa Fluor Goat anti-Mouse	488	IgG1	Thermo Fisher	A21121
Alexa Fluor Goat anti-Mouse	488	IgG2a	Thermo Fisher	A21131
Alexa Fluor Goat anti-Mouse	488	IgG2b	Thermo Fisher	A21141
Alexa Fluor Goat anti-Mouse	555	IgG	Thermo Fisher	A-21424
Alexa Fluor Goat anti-Mouse	555	IgG1	Thermo Fisher	A21127
Alexa Fluor Goat anti-Mouse	555	IgG2a	Thermo Fisher	A21137
Alexa Fluor Goat anti-Mouse	555	IgM	Thermo Fisher	A21426
Alexa Fluor Goat anti-Mouse	594	IgG	Thermo Fisher	A11032
Alexa Fluor Goat anti-Mouse	594	IgG1	Thermo Fisher	A21125
Alexa Fluor Goat anti-Mouse	647	IgG	Thermo Fisher	A21236
Alexa Fluor Goat anti-Mouse	647	IgG1	Thermo Fisher	A21240
Alexa Fluor Goat anti-Mouse	647	IgG2a	Thermo Fisher	A21241
Alexa Fluor Goat anti-Mouse	647	IgG2b	Thermo Fisher	A21242
Alexa Fluor Goat anti-Rabbit	350	IgG	Thermo Fisher	A11046
Goat anti-Rabbit	BV480	IgG	BD Horizon	564879
Alexa Fluor Goat anti-Rabbit	488	IgG	Thermo Fisher	A11034
Alexa Fluor Goat anti-Rabbit	555	IgG	Thermo Fisher	A21429
Alexa Fluor Goat anti-Rabbit	594	IgG	Thermo Fisher	A11037
Alexa Fluor Goat anti-Rabbit	647	IgG	Thermo Fisher	A21245
Alexa Fluor Goat anti-Rat	350	IgG	Thermo Fisher	A21093

Goat anti-Rat	BV421	IgG	BioLegend	405414
Goat anti-Rat	BV480	IgG	Becton Dickinson	564878
Alexa Fluor Goat anti-Rat	488	IgG	Jackson ImmunoResearch	112-545-167
Alexa Fluor Goat anti-Rat	488	IgG	Thermo Fisher	A11006
Alexa Fluor Goat anti-Rat	594	IgG	Thermo Fisher	A11007
Alexa Fluor Goat anti-Rat	647	IgG	Jackson ImmunoResearch	112-605-062
Alexa Fluor Goat anti-Rat	647	IgG	Thermo Fisher	A21247

Table S4: List of primers used in the study. Related to Experimental Procedures. Excel file contains multiple tabs that list primers used in the study.

Table S5: LHX2$^+$ branch gene expression in Allen Developing Mouse Brain Atlas. Related to Figure 3 Excel sheet with a list of genes scored for their presence (1) or absence (0) in the regions indicated at E13.5 and E15.5. The genes selected were differentially up-regulated in the *LHX2*$^+$ neurons relative to the *POU3F2*$^+$ neurons at D40 and D54, and were also represented in the data set of the Allen Developing Mouse Brain Atlas.

SUPPLEMENTAL DATA

Data S1: Heatmaps reported in Figures 2D, 5B, D-E. Multiple tabs in an Excel spreadsheet showing data from heatmaps normalized and arranged as presented in the main figures.

Data S2: Single cell mapping and metadata document. Related to Figures 2-5, 7, S3-4, Table S2. Excel sheet with three tabs associating single cell names (cell_ID) to replicate (group, or batch and diff, or batch and sample), day of differentiation (day), cell sorting phenotype information (phenotype), cluster name (if mapped to clusters), and general sequencing and mapping statistics data. The three tabs have CelSeq, SmartSeq2, and primary Fetal single cell data mapping and metadata. Virally barcoded cells are found in the SmartSeq2 file in rows that say "yes" under the "Viral BC?" column. CelSeq abbreviations are total reads (Reads), reads with a cell-specific barcode (BC), reads mapping to: transcriptome and genome (TG), transcriptome (T), introns (I), ERCC spike-in controls (ERCC). (p) denotes percentage, (R) denotes reads, (UMI) denotes unique molecular identifier. "Genes" is the number of genes detected per cell. ERCC_R is the R value, and ERCC_slope is the slope of ERCC amplification of each cell. For Fetal and SmartSeq2 samples the metrics include mapping to message RNA (mRNA), mitochondrial RNA (mtRNA), ribosomal RNA (rRNA), non-coding RNAs (NC), genome, ERCC, and tdTomato (tdT). "Genes" is the number of genes detected per cell. Note, *tdTomato* will only be detected from virally infected cells. Raw and normalized single cell RNA-Seq data from hESC-derived cells can be obtained at NCBI GEO: GSE86894. Raw and normalized primary fetal human single cell RNA-Seq data can be found at dbGaP: phs001205.v1.p1.

SUPPLEMENTAL EXPERIMENTAL PROCEDURES

hESC culture and genomic targeting. H1 hESCs (WA01; WiCell) were maintained with mTeSR1 media (Stem Cell Technologies) on Matrigel (BD). *PAX6*-targeted cells were maintained with hES media (DMEM/F12 with 20% KSR; Thermo Fisher) on CF-1 MEFs (GlobalStem). ESCs at p38-41 were treated with 1 μM thiazovivin (StemRD) one day prior to electroporation. On the day of electroporation, single cells were obtained by Accutase treatment (Thermo Fisher) and electroporated with the following conditions and reagents: Neon electroporator, resuspension buffer R, 100 μL electroporation tip (Thermo Fisher); 1050 V, 30 ms pulse width, 2 pulses; 1.5×10^6 cells; 1-3 μg of each TALEN plasmid and 3-12 μg of the HDR donor plasmid. The cells were then treated with 2 μM thiazovivin for 24 h following electroporation for recovery. After recovery, cells were drug-selected with 1 μg/mL puromycin (Thermo Fisher) for three days on Matrigel or DR4 MEFs (GlobalStem). For targeting *PAX6* and *DCX*, drug selection cassettes were removed by transfecting Cre recombinase mRNA (*in vitro* transcribed with mMessage mMachine, Thermo Fisher) with Stemfect RNA transfection reagent (Stemgent). *DCX*-targeted cells were further treated with 2 μM ganciclovir (Sigma) for three days. Individually picked $SOX2^{Cit/+}$ and $OTX2^{Cit/+}$ colonies and $DCX^{Cit/Y}$ and $PAX6^{Cit/+}$ single-cell clones were validated by PCR and had normal karyotypes (Cell Line Genetics). Southern blotting (Lofstrand) confirmed single-copy insertion of citrine in $SOX2^{Cit/+}$ and $DCX^{Cit/Y}$.

hESC neural differentiation.
hESCs were dissociated with Accutase and plated on Matrigel-coated 24-well plates at $2.5x10^5$ cells/cm^2 in DMEM/F12 (#11330-032), 1× N2, 1× B27 without vitamin A, 2 mM Glutamax, 100 μM non-essential amino acids, 0.5 mg/mL BSA, 1X Pen-Strep, and 100 μM 2-mercaptoethanol (referred to as basal medium; all from Thermo Fisher) with 20 ng/mL FGF2 (Thermo Fisher) and 2 μM thiazovivin. Cortical induction was initiated by changing to basal medium with 5 μM SB431542 (StemRD), 50 nM LDN193189 (Reagents Direct) and 1 μM cyclopamine (Stemgent) (referred to as NIM). NIM was changed daily for 11 days. On day 12, cells were dissociated and seeded on Matrigel-coated 24-well plates at 5×10^5/cm^2 in basal medium with 20 ng/mL FGF2 and 2 μM thiazovivin. Progenitor expansion was initiated on D13 by changing to serum-free human neural stem cell culture medium (NSCM, #A10509-01 from Thermo Fisher) containing 20 ng/mL FGF2 and 20 ng/mL EGF. NSCM was changed daily for 6 days. Cultures were passaged once more on D19 with Accutase and replated at $5x10^5$ cells/cm^2. On D26, cells were dissociated with Accutase and seeded on 24-well plates sequentially coated with poly-D-lysine (Millipore) and laminin (Thermo Fisher) at 1×10^5 cells/cm^2 in basal medium supplemented with 20 ng/mL FGF2 and 2μM thiazovivin. On D27, medium was changed to a 1:1 mixture of DMEM/F12 and Neurobasal medium (#21103-049) supplemented with 100 μM cAMP (Sigma), 10 ng/mL BDNF (R&D Systems), 10 ng/mL GDNF (R&D Systems) and 10 ng/mL NT-3 (R&D Systems) (referred to as ND). Cells were maintained in ND medium for 4 weeks until day 54 with half medium change every other day.

Quality of differentiations was routinely assessed by immunostaining at D12 (PAX6 and DCX), at D26 (LHX2, SOX2, EOMES, POU3F2, and TBR1), and at D54 (MAP2 costained with TBR1, CTIP2, SATB2) (Table S3). In addition flow cytometry at D26 (EOMES and SOX2 and PAX6) was performed. Typically EOMES at day 26 proved the most valuable quality control metric (~10% of cells by both flow cytometry and immunostaining) and predicted failure at D54. Specifically when EOMES was low (<1% of cells) differentiations failed and were typically dominated by $POU3F2^+$ cell types and/or non-neural "other" cell types. These failed differentiations were eliminated from further analysis, typically ~ 20% of experiments (5 of 19 experiments in 2016).

Fetal brain tissue processing.
We identified cortical pieces by morphology and partitioned them into one half for fixation, sectioning, and immunostaining, and the other half for single cell harvesting. For dissociation, we minced the tissue into small pieces (approx. 0.25 - 0.5 mL total volume) with #5 forceps (Fine Science Tools) in Ca^{2+}- and Mg^{2+}-free HBSS (Thermo Fisher). Minced pieces were then treated with 2 mL trypsin solution for 20 min at 37 °C (Ca^{2+}- and Mg^{2+}-free HBSS, 10 mM HEPES, 2 mM $MgCl_2$, 0.25 mg/mL bovine pancreatic trypsin (EMD Millipore), 10 μg/mL DNase I (Roche), 100 nM TTX (Tocris), 20 μM DNQX (Tocris), and 50 μM DL-AP5 (Tocris), pH 7.6). We quenched digestion with 6 mL of ice-cold Quenching Buffer (440 mL Leibovitz L-15 medium, 50 mL water, 5 mL 1M HEPES pH 7.3–7.4, 5 mL 100× Pen-Strep, 20 mg/mL bovine serum albumin (Sigma), 100 μg/mL trypsin inhibitor (Sigma), 10 μg/mL DNase I, 100 nM TTX, 20 μM DNQX, and 50 μM DL-AP5). We then pelleted the samples (220×g, 4 min, 4°C) and resuspended with 1 mL of quenching buffer and triturated on ice with a P1000 pipette set to 1 mL, using 25 gentle

cycles up and down without forming bubbles. We then diluted the cell suspension to 30 mL in Staining Medium (440 mL Leibovitz L-15 medium, 50 mL water, 5 mL 1M HEPES pH 7.3–7.4, 5 mL 100× Pen-Strep, 20 mL 77.7 mM EDTA pH 8.0 [prepared from Na_2H_2EDTA], 1 g bovine serum albumin, 100 nM TTX, 20 μM DNQX, and 50 μM DL-AP5), filtered through a 45 micron cell filter, pelleted (220 × g, 10 min, 4°C), resuspended in 5 mL staining medium, and counted on a hemocytometer (typically ~20-40 M live cells isolated per cortical piece at ~50% viability). Cells were then fixed (4% PFA in PBS, 15 min on ice), rinsed twice (PBS, 0.2% w/v molecular biology grade BSA (Gemini Bio-Products), 0.25% v/v RNasin Plus (Promega)), and kept frozen in this buffer at -80°C at 10 M cells / mL until processed for FRISCR as in previously described (Thomsen et al., 2016).

Single cell sorting. To generate single cell suspensions, hESC-derived cultures were dissociated from plates using Accutase (ThermoFisher) at 37°C. Light trituration using a P1000 pipette was done every 5 min until nearly all clumps had been dissociated (up to 1 h). Cell suspension was washed and filtered through a 40 μm cell strainer. Cells were washed in PBS with 1% FBS and stained with 0.5-1 μg/mL DAPI. Single-cell suspensions were loaded onto a FACSAria II SORP (Becton Dickinson) and sorted directly into PCR strip tubes or plates held in chilled aluminum blocks. Doublets and dead cells were excluded based on forward scatter, side scatter and DAPI fluorescence. Sorting was done using the 130 μm nozzle with the sort mode set to single cell. Accuracy of single-cell sorts was confirmed by sorting DAPI-stained fixed cells onto a dry well of a 96-well plate and analyzing by fluorescence microscopy.

Single cell transcriptomics. We prepared libraries using the CelSeq protocol as previously reported (Hashimshony et al., 2012) with a few modifications. Single cells were sorted with a FACSAria (BD) into 96-well plates containing 1.2 μL 2× CellsDirect Buffer (Thermo Fisher) with 0.1 μL of External RNA Controls Consortium (ERCC) control RNAs diluted to 1×10^{-6} molecules (Thermo Fisher). After sorting, plates were then frozen and stored at -80°C. For library preparation, plates were thawed on ice. mRNA was reverse transcribed using 1.25 pmol or 0.15625 pmol of oligoT primer carrying a cell-specific 8 NT barcode and a 5 NT unique molecular identifier (UMI) (Islam et al., 2014) (see Table S4). Barcode design ensured at least three nucleotide differences from any other barcode. Samples were incubated in a PCR machine (Tetrad, BioRad) at 70 °C with a 70 °C heated lid for 3 min, spun, and heated again for two more minutes. Samples were reverse transcribed using Superscript III (Thermo Fisher) for two hours at 50 °C with a 52 °C lid and subsequently treated with 1 μL of ExoSAP-IT (Affymetrix). Samples were cooled on ice for second strand synthesis, where Second Strand Synthesis Buffer, dNTPs, DNA Polymerase, and RNAse H (NEB) were added to the samples for a 10 μL total volume and incubated at 16 °C for 2 h. Single cells were pooled by 24 wells per library, with each library containing a water-only well and an ERCC-only well. Single cell pools or population RNA libraries were purified with an equal volume of RNA Clean Beads (Beckman Coulter), linearly amplified at 37 °C for 15 h using the HiScribe T7 High Yield RNA Synthesis kit (NEB), and treated with DNAse I (Thermo Fisher). RNA was fragmented using the NEBNext RNA Fragmentation Module (NEB), purified using an equal volume of RNA Clean Beads, and visualized (RNA Pico Kit, Bioanalyzer 2100, Agilent). The RNA fragments were repaired by treating with Antarctic Phosphatase and Polynucleotide Kinase (NEB) and purified with an equal volume of RNA Clean Beads. cDNA libraries were made using the NEBNext Small Library Prep Kit according to the manufacturer's protocol, except Superscript III was used for the RT step. Index primers were used in PCR amplification. Libraries were purified using an equal volume of RNA Clean Beads and were quantified on the Bioanalyzer using the DNA High Sensitivity Kit (Agilent). Approximately 160-200 nmol of a pool of libraries were size selected to exclude species <180 bp on a 2% Dye-Free cassette on the Pippin Prep (Sage) and Speed Vac concentrated to approximately 14 μL. Libraries were then quantified by qRT-PCR using p5 (5'-AATGATACGGCGACCACCGAGA-3') and p7 (5'-CAAGCAGAAGACGGCATACGAGAT-3') primers and visualized (DNA High Sensitivity Kit, Bioanalyzer 2100). Library pools were then sequenced on an Illumina HiSeq using a custom read1 primer (5'-TCTACACGTTCAGAGTTCTACAGTCCGACGATC-3') and the Illumina primer HP10. Standard Illumina primers HP12 and HP11 were used for the index read and the transcript read, respectively. PE50 kits (Illumina) were used for sequencing with read lengths of 25 nt, 6 nt, and 47 nt for read1, index, and read2, respectively. A list of primers are in Table S4.

FRISCR was carried out as recently described (Thomsen et al., 2016), and SmartSeq2 sequencing libraries were prepared as previously reported (Picelli et al., 2013). After reverse transcription and template switching, we amplified cDNA with KAPA HotStart HIFI 2× ReadyMix (Kapa Biosystems) for 22 cycles for RNA from single primary cortical cells. We purified PCR products using Ampure XP beads (Beckman Coulter). We quantified cDNA using a High Sensitivity DNA Chip (Agilent) on a Bioanalyzer 2100 or with the Quant-iT PicoGreen dsDNA Assay Kit (Thermo Fisher) on an Enspire plate reader (PerkinElmer). We used 1 ng of cDNA to generate RNA-Seq libraries using the

Nextera XT library prep system (Illumina). We carried out sequencing of human cortical cells the on Illumina HiSeq using 50 base paired-end reads.

Mapping sequencing data. For CelSeq, reads were de-multiplexed by CelSeq index, allowing for one sequence mismatch. The indexes were designed with a minimum Hamming distance of 3, allowing one mismatch to be error-corrected. The transcript reads for each cell were aligned to the RefSeq transcriptome (downloaded March 2013) using Tophat with default parameters (Trapnell et al., 2009). Unaligned reads were then aligned to the genome using Bowtie (Langmead et al., 2009), followed by alignment to the ERCC spike-in controls. At this point, any remaining unaligned reads were then mapped to the genome again using GSNAP (Wu and Nacu, 2010), allowing for soft clipping. After alignment, all reads mapping to exons of the same gene were collapsed by their Unique Molecular Identifier (UMI) and were used to generate the overall cell-by-cell gene expression matrix. Because CelSeq is strongly 3′ biased and the full 3′ UTR isn't always represented in the RefSeq transcriptome, if a read mapped within 1kb of the 3′ end of a gene and was in the same orientation as the gene, we assessed it as a transcript-mapping read for that gene. Non-exon read mapping percentages were calculated, but these reads were not used in the gene expression analysis. For normalization, cells with fewer than 20,000 total cellular UMIs were discarded, and data for all remaining cells was randomly subsampled to 20,000 total cellular UMIs. We also calculated clusters based on a 10,000 UMI threshold cutoff, which resulted in the inclusion of more cells, and minor changes to some of the clusters relative to the 20,000 UMI cutoff, but not the appearance of new cell types that we had not seen previously with the 20,000 UMI cutoff. ERCC UMIs were subsampled to the same ratio (20,000/total cellular UMIs) for each cell to generate the final data set. Mapping SmartSeq2 sequence reads was done as previously described (Thomsen et al., 2016).

Batch effect correction. Using CelSeq we have noticed substantial batch effects similar to other recent papers using the molecularly similar MARS-Seq protocols (Paul et al., 2015). Presumably these batch effects arise both from cell-to-cell crosstalk as well as from other library-specific noise. In order to estimate and correct for these batch effects we first identify genes that are differentially expressed among batches using DESeq2 (adjusted p-value < .005 and fold-change >2), because these genes account for the most prominent batch-related effects, and form WGCNA gene modules that drive formation of batch specific clusters. Then we compute the per-library median expression values and per-experiment overall median expression values (each experiment contains 2-8 libraries) for these genes. we correct each library by subtracting the difference between per-experiment overall and per-library median gene expression values. Correction of library bias from these top library specific genes is sufficient to eliminate library specific gene modules that bias clustering.

Initial clustering. Cells were clustered to determine cell types based on the CelSeq single-cell gene expression data. Cells with >20,000 transcripts are clustered iteratively based on WGCNA modules (Zhang and Horvath, 2005) of genes with variance greater than technical noise determined by DESeq2 (Brennecke et al., 2013). Due to strong temporal gene expression variation across cells from different time points, we perform clustering independently for cells at each age group. WGCNA gene modules that were present or absent in fewer than four cells, or had low statistical power to discriminate clusters were removed. We calculated a "de.score" to determine the power of gene modules to discriminate clusters. For this, all genes from the evaluated module were used to divide cells into two clusters. Based on those two cell clusters, differentially expressed genes were identified (adjusted Pvalue ≤ 0.01 and > 2 fold change), and a de.score was calculate as the sum of $-\log_{10}$ (adjusted Pvalues) for all differential expressed genes. Modules that corresponded to cell cycle states, ribosomal proteins, mitochondrial genes, and cell stress condition were removed from downstream analysis (Table S1). The cells are then clustered based on eigengenes from the remaining gene modules using hierarchical clustering with Ward's method. The number of clusters is selected dynamically such that the de.score between every pair of clusters is at least 80.

Bootstrapping to determine final clusters and to assess cluster robustness. We randomly removed 20% of the cells present in the CelSeq dataset and recomputed clustering 100 times. The number of times that cells co-cluster after each iteration is directly used to compute a pairwise co-clustering frequency. These co-clustering frequencies were input into a standard hierarchical clustering algorithm with Ward's method, with the de.score of at least 80 terminal criterion for tree cutting (same as for initial clustering above) to establish final clusters. Furthermore these pairwise co-clustering frequencies were also used to assess cluster robustness as measured by both the amount of intra-cluster co-clustering (the "tightness" of each cluster), and also the amount of inter-cluster co-clustering (the "relatedness" of each pair of clusters), which are both easily visualized in the constellation diagram (Figure 2F).

***Cross-platform comparison*.** To compare hESC-derived cell types to developmental mouse/human brain atlases, we normalized datasets by first subtracting the minimum expression value per gene, then divided by the max expression value per gene or the median maximum value of all genes if that maximum value fell below this threshold. This was done to to ensure that the transformed data sets are on a similar scale between 0 and 1. Due to differences in platforms, resolution, and temporal stages between hESC-derived cell types and atlases, comparisons using all genes were misleading. Thus, we conducted our comparisons using conserved co-expressed genes modules identified by WGCNA (Langfelder and Horvath, 2007) that distinguished both the hESC-derived cell types and brain regions. To determine the consensus gene modules, we computed the similarity matrices for the hESC-derived cell type markers using WGCNA TOMsimilarity function for both datasets, and used the per-element minimum of the two similarity matrices as the consensus. The genes were then hierarchically clustered based on the consensus similarity matrix. To ensure the coherence of each gene module, we calculated the correlation between module members (genes) and the module eigengene, then removed the module members with Pearson correlation <0.2. The statistical significance of the conserved gene module is calculated using modulePreservation function (Langfelder et al., 2011). We then established the Spearman correlation between hESC-derived cells and the atlas regions using the genes from the consensus gene modules (Table S1).

***Mapping of viral barcode cells*.** The virally infected and barcoded cells were profiled by SmartSeq2. All major gene modules were shared between the CelSeq and SmartSeq2 datasets with some subtle differences. To map the expression profiles of these virally marked and SmartSeq2-profiled cells to CelSeq clusters, SmartSeq2 gene expression was first normalized as described above in "Cross-platform comparison". Then the genes that distinguish CelSeq clusters and were also detected in the SmartSeq2 dataset were used to train a random forest classifier (using the "randomForest" R package (Breiman, 2001)) based on the CelSeq datasets, which is then used to predict the cluster membership of the SmartSeq2 amplified virally barcoded cells. To improve classification accurarcy of small clusters, the classifier used 50 randomly selected cells per cluster as a training set (with replacement sampling in case if the cluster had fewer than 50 total cells). The sampling and classification precedure was repeated for 100 times to assess the robustness of the prediction results. For each cell, we report the top prediction with the corresponding confidence level (Table S2).

Comparison with primary fetal cortical cells. Primary fetal cortical cells with >100,000 mRNA reads, > 1000 ERCC reads and >30% mRNA mapping percentage were selected for clustering. The selected cells were clustered using iterative WGCNA approach discussed above but with no batch effect correction. Only hESC-derived cells from $LHX2^{+}$ clusters from Day 26 and 54 were compared to primary fetal cortical cells. We first identified conserved gene modules between hESC-derived and primary cells, which corresponded to the progenitors, intermediate progenitors and neurons respectively. Then we searched for non-conserved genes that are correlated conserved gene modules only among fetal or hESC-derived cells. To do this, we computed the eigengene for each conserved modules for both fetal and hESC-derived cells. To select genes that are specific to the hESC-derived cells, we chose genes that have correlation > 0.6 with an eigengenes in hESC-derived cells, and correlation < 0.2 with the corresponding eigengene in fetal cells. Similarly, to select genes that are specific to the fetal cells, we chose genes that have correlation > 0.6 with an eigengenes in fetal cells, and correlation < 0.3 with the corresponding eigengenes in hESC-derived cells.

Lineage inference. We used a Bayesian inference method to be described elsewhere (S. Ramanathan, personal communication) to assess the probability that any given triplet of cell types represents a lineage relationship. In general, a gene whose distribution of expression is significantly lower in one cell type as compared to two others provides evidence about that cell type not being an intermediate (or progenitor) state, as formalized in this equation:

$$p\left(g_i^{A,B,C} \middle| T = \mathcal{A}, \beta_i = 1, \{C\}\right) = \frac{1}{2}\begin{Bmatrix} p\left(g_i^{A,B,C} \middle| \mu_B^i < \mu_A^i, \mu_B^i < \mu_C^i, \{C\}\right) \\ +p\left(g_i^{A,B,C} \middle| \mu_C^i < \mu_A^i, \mu_C^i < \mu_B^i, \{C\}\right) \end{Bmatrix}$$

Where $p\left(g_i^{A,B,C} \middle| T = \mathcal{A}, \beta_i = 1, \{C\}\right)$ represents the probability distribution of expression of gene i in cell types A, B, and C, given a topology where A is the intermediate state and gene i is a key gene (signified by $\beta_i = 1$) under a clustering $\{C\}$. μ_A^i and σ_A^i represent the mean and standard deviation of the expression of the gene in cell type A, with analogous expressions for cell types B and C.

For a given set of three cell types, four lineage topologies are possible: each of the three types could be the intermediate state, or there could be no significant evidence for any of them to be the intermediate state. The probability of a given topology, given the expression data, can be calculated as follows:

$$p(T|\{g_i^{A,B,C}\}) \propto \prod_i \left(1 + \frac{3}{2}\mathcal{O}_i[1 - p(\mu_T^i \text{is min}|g_i^{A,B,C})]\right),$$

where, $\mathcal{O}_i = p(\beta_i = 1|g_i^{A,B,C})/p(\beta_i = 0|g_i^{A,B,C})$ is the odds of gene i being a transition gene and thus having a unique minimum. The term $p(\mu_T^i \text{is min}|g_i^{A,B,C})$ is the probability that the mean μ_T^i of the distribution of the expression levels of gene i in the root cell type T is less than the mean in the other two cell types. Using this framework, we analyzed all possible triplets of cell types present at adjacent time points using only transcription factor expression data, and identified the highest-probability triplets with linked topologies. Triplets that showed strong evidence of topological linkages were assembled manually into a tree which was rooted at D12 For each successive time point, we selected only those triplets containing any types from that time point, the previous time point, and the following time point, and linked these triplets to build the tree; and any given triplet suggests either a branch split or a progression of cell types, and the overall tree was assembled manually. In cases where there were conflicting triplets, we selected the topology with the higher overall probability. Once the putative tree was built, we identified the predicted asymmetrically regulated transcription factors for each triplet, using both the Bayesian inference algorithm as well as differential gene expression among all pairs of types using the DESeq package in R. These asymmetrically regulated transcription factors contributing to various triplets are listed in Figure 6C. The grouping of cell types in Figure 6B indicates that there were no triplets that separated those types into a meaningful topology.

Barcoded viral plasmid library construction. The CAG-tdTomato-WPRE-polyA cassette from the Ai9 plasmid (a gift from Hongkui Zeng, Addgene plasmid 22799) was cloned into the backbone of pMXs-SOX2 (Addgene plasmid 13367) via InFusion cloning (Clontech) to create pMXs-Ai9 (the CAG promoter was PCR-amplified with SK463 and SK475, tdTomato was amplified with SK471 and SK474, the backbone was amplified with SK476 and SK477). A 10-bp sequence of the polyA signal was replaced with a 10-bp degenerate barcode library (Integrated DNA Technologies), generating pMXs-Ai9-BC by cloning a DNA fragment generated by annealing and extending two primers, one of which contained the library (SK597 and SK599). To preserve library diversity, the entire ligation reaction was transformed into chemically competent bacteria; additionally, the bacterial transformation reactions were not plated to isolate single bacterial colonies for liquid cultures. Instead, 1% of each transformation reaction was plated to estimate maximal library size, and the remaining 99% was placed immediately into selective liquid culture for large-scale plasmid prep. For all primer sequences, see Table S4.

Viral packaging. The barcoded plasmid library pMXs-Ai9-BC was packaged into retrovirus using the packaging cell line Plat-A (Cell Biolabs) and pseudotyped with the coat protein VSV-G. In each batch of four to ten 10 cm plates of Plat-A cells, each 10 cm dish of Plat-A cells plated at 9 x 10^4 cells/cm^2 was transfected with pMXs-Ai9-BC (6 µg) and pMD2.G (0.24 µg; Addgene plasmid 12259) using Lipofectamine LTX (24 µL) with Plus reagent (6 µL; Thermo Fisher). Fifteen-20 h following transfection, the transfection media was replaced with standard cell culture media (10% serum in DMEM). At 48 hours after transfection, cell culture supernatant was collected and viral particles were concentrated using PEG-It or Retro-Concentin (System Biosciences). Viral pellets were resuspended at 1% of initial supernatant volume in cold PBS and stored at -80 °C in aliquots.

Viral titering. Concentrations of viral preparations were measured by serial dilution infections in 293T plated at confluence in 24-well plates. Serially diluted samples of viral preps were used to inoculate the 293T cultures with 10 µg/mL polybrene (Millipore). The culture plates were immediately centrifuged for 1 h at room temperature at 800 x g, then returned to 37°C. Two days later, two to three representative microscope fields were imaged per serial dilution, and numbers of fluorescent cells were counted per field. Final viral titer was estimated based on the cell counts, surface area per microscopic field, cell culture vessel surface area, and serial dilution of inoculum.

Pre-screening virally barcoded cells. To reduce the number of un-related single cell that would be sequenced, we determined clonally related cells first. Briefly, 1 uL of amplified cDNA libraries were re-amplified using the following primers: SK0631, SK0632 (Table S4). These primers span a region of the viral tdTomato 3' UTR containing the degenerate 10 bp barcode. Reactions were amplified using Phusion Flash High-Fidelity PCR Master Mix (Thermo Fisher) according to the manufacturer, products were purified with Agencourt AMPure XP beads (Beckman Coulter), and 15 ng sequenced with the primer SK0530 (Table S4) at Genewiz (South Plainfield, NJ). To ensure both branches were represented in our data set we assessed expression of *LHX2* in the amplified cDNA libraries (Figure S6G-H).

GAPDH was detected with primers: SK0645, SK0646, and *LHX2* with SK0649, SK0650 (Table S4). Primers pairs were designed to span introns. qRT-PCR was performed using qScript One-Step SYBR Green RT-qPCR (Quant Biosciences) according to the manufacturer. Reactions underwent the following thermal cycling conditions: 94°C for 2 minutes for initial denaturation; 50 cycles of 94°C x 30 sec, 55°C x 30 sec, 72°C x 30 sec for amplification using a Lightcycler 480 (Roche Life Science). Single-cell cDNA libraries were chosen to continue with Nextera library preparation for RNA sequencing. All cells from all multicellular clones were sequenced (with or without expression of *LHX2*) except for clones 83 and 46 that had 288, and 73 cells recovered. Only 50 cells were sequence cell were sequenced from these two clones, including all the $LHX2^+$ cells.

Progenitor potential assay by clonal outgrowth. Mouse astrocytes were prepared from P0-P2 pup SVZs as a feeder layer for colony formation. Dissected and minced SVZs from five pups were then treated with 0.5 mL trypsin solution for 6 min at 37°C, quenched with 2 mL ice-cold quenching buffer, pelleted (220×g, 4 min, 4°C), resuspended with 1 mL quenching buffer, triturated with a P1000 pipet using 15-20 gentle cycles up and down, then plated all in a 15-cm dish with DMEM high glucose (#11995-065) plus 10% FBS and Pen-Strep. Cells were passaged twice 1:4 after 7-10 days each, then postmitotic astrocytes were plated in 96-well plates at 10^4 cells per well and maintained with medium changes every two-three weeks until human cell seeding. Temporary treatment with clodrosomes was used in some cases to eliminate microglia when contamination was severe (50 ug/mL for 3 days).

For colony formation, differentiating hESC progenitors at D26 were seeded at clonal density (10 cells per well) on astrocytes which typically led to approximately two colonies per well. Colonies were divided into two differentiation conditions, which were found post hoc to yield similar cell types within colonies and so the data for both differentiation conditions were pooled. In the first condition the colonies were differentiated immediately for four weeks in ND medium. In the second condition the colonies were were initially grown for two weeks in a modified NSCM medium prior to four weeks of ND for differentiation. The modified NSCM consisted of: 69 mL DMEM-F12 (ThermoFisher), 26 mL Neurobasal-A (ThermoFisher), 2 mL B-27 minus vitamin A (ThermoFisher), 1 mL N-2 [#17502-048], 0.5 mL GlutaMAX (ThermoFisher), 1 mL Pen-Strep (ThermoFisher), 100 µL 50 mM 2-mercaptoethanol (Sigma), 100 µL 10 mM Y-27632 dihydrochloride (Tocris), 20 ng/mL EGF (ThermoFisher), 20 ng/mL bFGF (ThermoFisher), and 20 ng/mL IGF-1 (R&D Systems).

After differentiation all the colonies were fixed (4% PFA, 15 min, 4°C), rinsed with PBS, then processed for immunostaining. Twenty-five to thirty-eight colonies were analyzed per experiment, per differentiation condition, per staining cocktail. Immunostaining (Table S3) of these colonies was performed using five-channel imaging with the dyes BV421, BV480, Alexa 488, Alexa 555, and Alexa 647. Three channels were occupied by HNA + TuJ1 (BV421 channel), POU3F2 (BV480 channel), and LHX2 (488 channel), and then 555 and 647 channels were used for other markers to interrogate co-occurrence with LHX2 and/or POU3F2. These markers were CRABP1, CALB2 (Calretinin), FOXP2, GAD67, BCL11B, TFAP2B, and NFIA. HNA + TuJ1 was used to find colonies and detect presence of neurons (>95% of colonies). Colonies were then scored as containing POU3F2 or LHX2 or any of the other markers co-stained (a binary yes or no for each marker). Fine analysis of co-stained cell types was tabulated for each clone (e.g., $FOXP2^+LHX2^+TuJ1^+$ cells) but not used for analysis. For BCL11B , only bright $BCL11B^+$ cells were considered to be true positive cells because many cells had dim expression of BCL11B including many $TuJ1^-$ cells.

SUPPLEMENTAL REFERENCES

Breiman, L. (2001). Random Forests. Machine Learning *45*, 5-32.

Brennecke, P., Anders, S., Kim, J.K., Kolodziejczyk, A.A., Zhang, X., Proserpio, V., Baying, B., Benes, V., Teichmann, S.A., Marioni, J.C., *et al.* (2013). Accounting for technical noise in single-cell RNA-seq experiments. Nature methods *10*, 1093-1095.

Diez-Roux, G., Banfi, S., Sultan, M., Geffers, L., Anand, S., Rozado, D., Magen, A., Canidio, E., Pagani, M., Peluso, I., *et al.* (2011). A high-resolution anatomical atlas of the transcriptome in the mouse embryo. PLoS biology *9*, e1000582.

Hashimshony, T., Wagner, F., Sher, N., and Yanai, I. (2012). CEL-Seq: single-cell RNA-Seq by multiplexed linear amplification. Cell reports *2*, 666-673.

Islam, S., Zeisel, A., Joost, S., La Manno, G., Zajac, P., Kasper, M., Lonnerberg, P., and Linnarsson, S. (2014). Quantitative single-cell RNA-seq with unique molecular identifiers. Nature methods *11*, 163-166.

Langfelder, P., and Horvath, S. (2007). Eigengene networks for studying the relationships between co-expression modules. BMC systems biology *1*, 54.

Langfelder, P., Luo, R., Oldham, M.C., and Horvath, S. (2011). Is my network module preserved and reproducible? PLoS computational biology *7*, e1001057.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol *10*, R25.

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., Bjorklund, A.K., Faridani, O.R., Sagasser, S., Winberg, G., and Sandberg, R. (2013). Smart-seq2 for sensitive full-length transcriptome profiling in single cells. Nature methods *10*, 1096-1098.

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., *et al.* (2016). Fixed single-cell transcriptomic characterization of human radial glial diversity. Nature methods *13*, 87-93.

Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. Bioinformatics *25*, 1105-1111.

Wu, T.D., and Nacu, S. (2010). Fast and SNP-tolerant detection of complex variants and splicing in short reads. Bioinformatics *26*, 873-881.

Zhang, B., and Horvath, S. (2005). A general framework for weighted gene co-expression network analysis. Stat Appl Genet Mol Biol *4*, Article17.