Supplemental Information

Structurally conserved primate lncRNAs are transiently expressed during human cortical differentiation and influence cell type specific genes

Andrew R. Field, Frank M.J. Jacobs, Ian T. Fiddes, Alex P. R. Phillips, Andrea M. Reyes-Ortiz, Erin LaMontagne, Lila Whitehead, Vincent Meng, Jimi L. Rosenkrantz, Max Hauessler, Sol Katzman, Sofie R. Salama, David Haussler

Figure S1

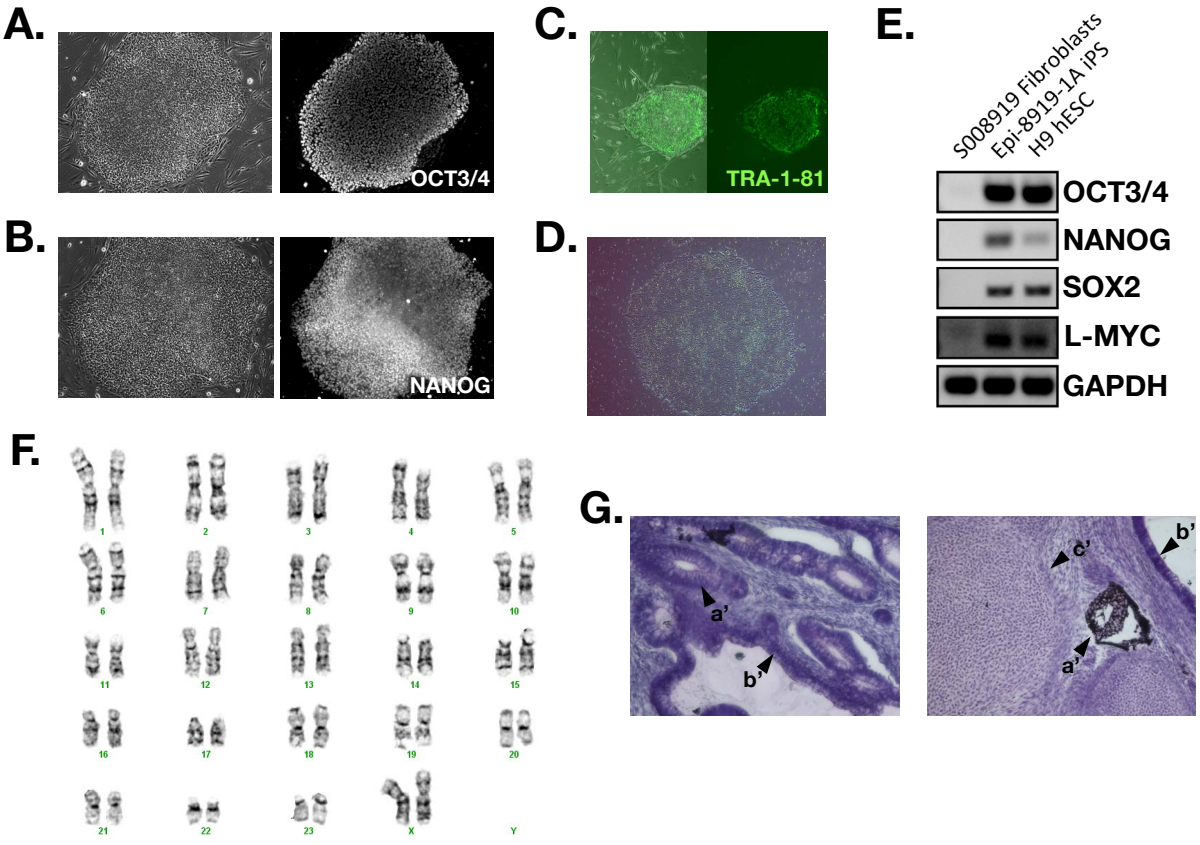

Figure S1, related to Figure 1. Chimpanzee iPSC verification. Immunofluorescence staining of Epi-8919-1A chimpanzee iPSC colonies displays *OCT3/4* (A) and *NANOG* (B) expression. (C) Live staining of an early iPSC colony for Tra-1-81. (D) Brightfield image of a chimpanzee iPSC colony on Matrigel feeder-free conditions. (E) RT-PCR products visualized on agarose gel comparing expression of *OCT3/4*, *NANOG*, *SOX2*, *L-MYC*, and *GAPDH* in S008919 starting chimpanzee fibroblasts, reprogrammed Epi-8919-1A chimpanzee iPSCs, and human H9 ESCs. (F) Wildtype 48, XX karyotype was confirmed in chimpanzee iPSCs at passage 32 after reprogramming. (G) Haemotoxylin and Eosin staining of teratomas derived from Epi-8919-1A iPSCs shows the generation of all three germ layers: (a') ectoderm (neural rosettes and pigmented cells), (b') endoderm (gut), and (c') mesoderm (cartilage).

Figure S2

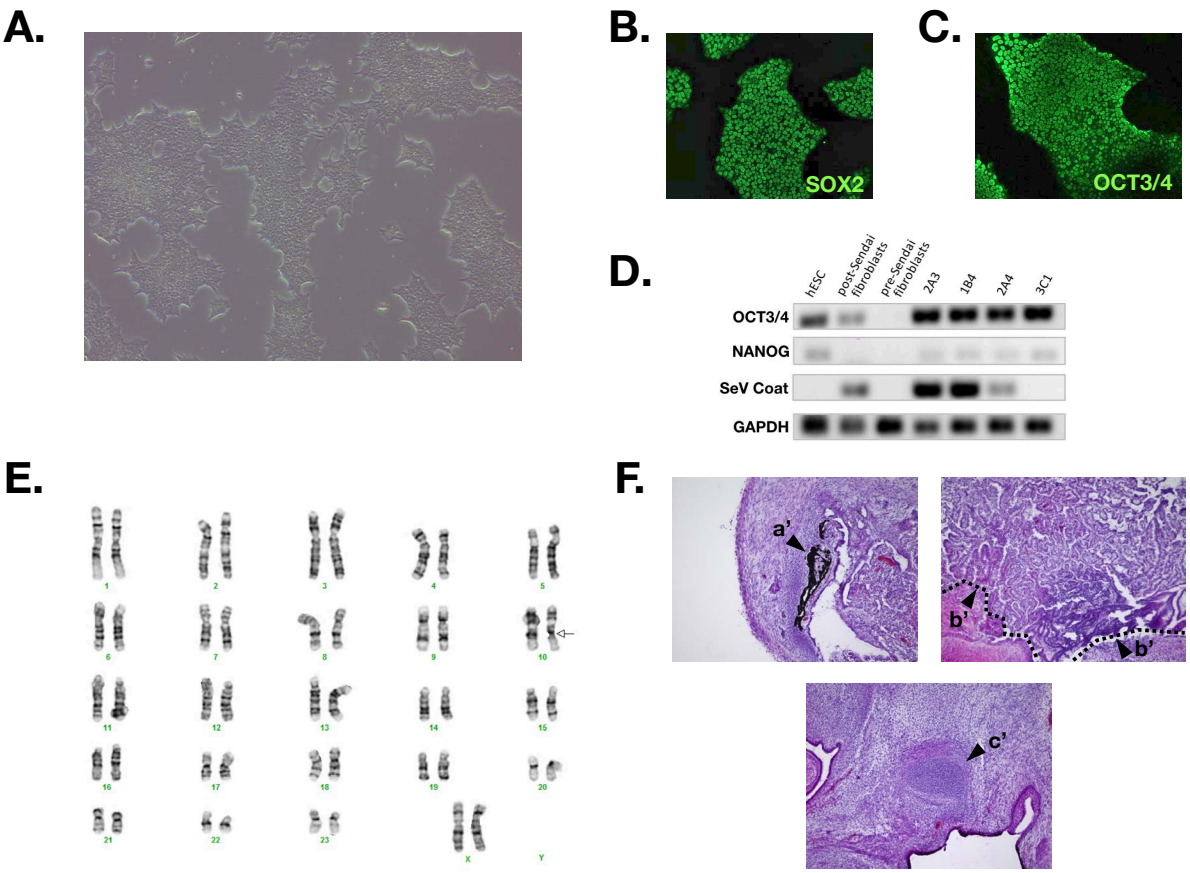

Figure S2, related to Figure 1. Orangutan iPSC verification. (A) Brightfield image of Jos-3C1 Sumatran orangutan iPSC colonies on vitronectin. Immunofluorescence staining displays *SOX2* (B), and *OCT3/4* (C) expression. (D) RT-PCR products visualized on agarose gel comparing expression of *OCT3/4*, *NANOG*, Sendai Virus coat protein, and *GAPDH* in human H9 ESCs, orangutan fibroblasts post- and pre-transduction with Sendai Virus, and 4 clones of orangutan iPSCs shows Sendai Virus has cleared in clone 3C1. (E) Wildtype 48, XX karyotype was confirmed in orangutan iPSCs at passage 36 after reprogramming. An inversion in chromosome 10 is naturally occurring in the wild Sumatran orangutan population (Locke et al. 2011) and was present in the original fibroblasts. (F) Haemotoxylin and Eosin staining of teratomas derived from Jos-3C1 iPSCs shows the generation of all three germ layers: (a') ectoderm (pigmented cells), (b') endoderm (gut), and (c') mesoderm (cartilage).

Figure S3

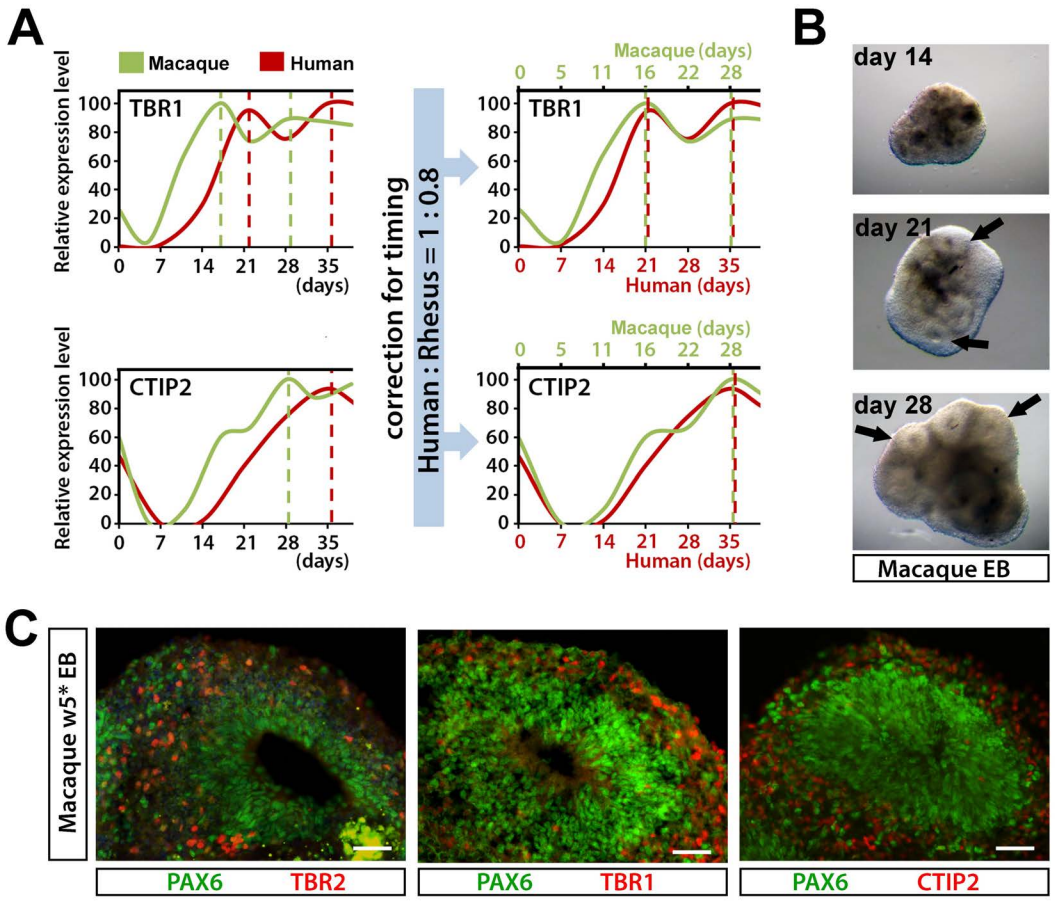

Figure S3, related to Figure 1 and 2. Determination of comparable time points between human and rhesus cortical organoids. A) Relative expression levels of TBR1 (upper graphs) and CTIP2 (lower graphs) in organoids isolated at multiple time points throughout human (red) and macaque (green) ESC cortical organoid differentiation, as determined by semi-quantitative RT-PCR. On the left, values are plotted using the actual time-scale, dotted lines indicate peak expression for human (red) and macaque (green) organoids. On the right, the same relative expression values are plotted using different timescales for human (lower X-axis) and macaque (upper axis) values. Macaque times are adjusted by a factor of 0.8. B) Light microscopy images of macaque ESC-derived cortical organoids at various time points. Black arrows indicate neural rosettes. C) IF staining with cortical neuron markers at day 28 (W5*): PAX6 and TBR2 (left),TBR1 (middle) and CTIP2 (right). Scale bars=50 um. EB, embryoid body (ESC cortical organoids).

Figure S4

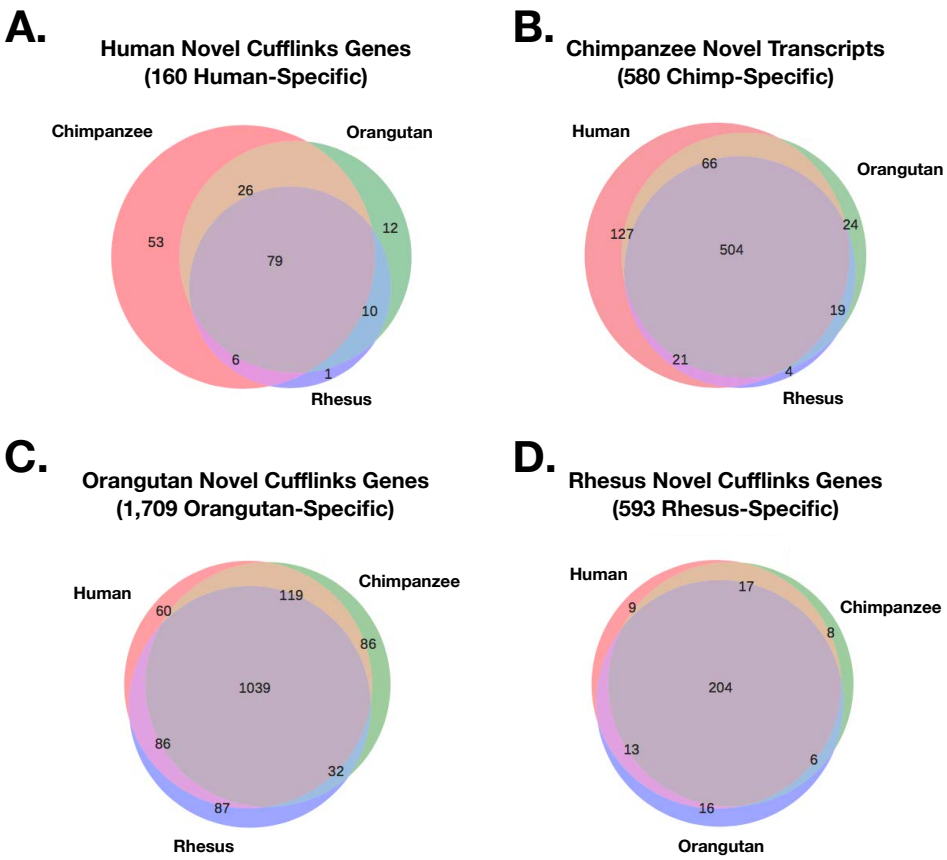

Figure S4, related to Figure 3. Conservation of novel Cufflinks gene loci detected in each species. Gene loci with no overlap to FANTOM5 lv3 (Hon, et al. 2017) (or its transMap equivalent, see Methods) were identified by Cufflinks. Unique loci were aligned pairwise using transMap to each other genome and evaluated for intron boundary retention as with FANTOM identified lncRNAs. Venn diagrams depicting the conservation of novel loci from human (A), chimpanzee (B), orangutan (C), and rhesus (D) to each of the other species. Fewer novel human gene loci were detected presumably due to being the origin species of the FANTOM transcripts. The orangutan numbers are expected to be inflated due to the relatively poor genome assembly and the resulting poor alignment to other genomes.

Figure S5

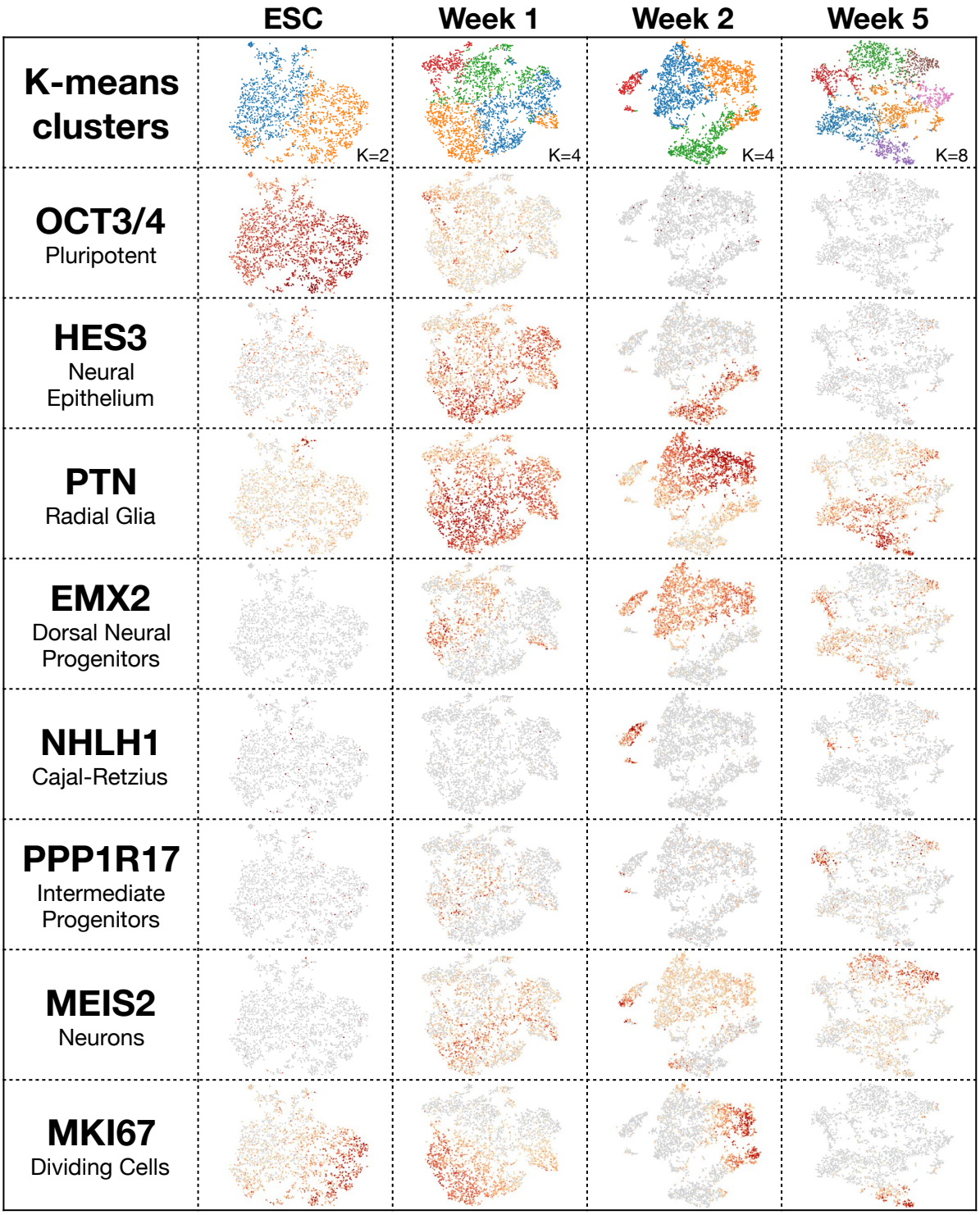

Figure S5, related to Figure 4. Human cortical organoid single cell RNA-sequencing time course. t-SNE plots from single cell RNA-seq displaying the expression of *OCT3/4* (pluripotent cells), *HES3* (NE), *PTN* (RG), *EMX2* (dorsal neural progenitors), *NHLH1* (CR), *PPP1R17* (intermediate progenitors), *MEIS2* (neurons), and

MKI67 (dividing cells) and showing the increasing cell heterogeneity as time progresses in weeks 0, 1, 2, and 5 of neural differentiation in from human ESCs.

Figure S6

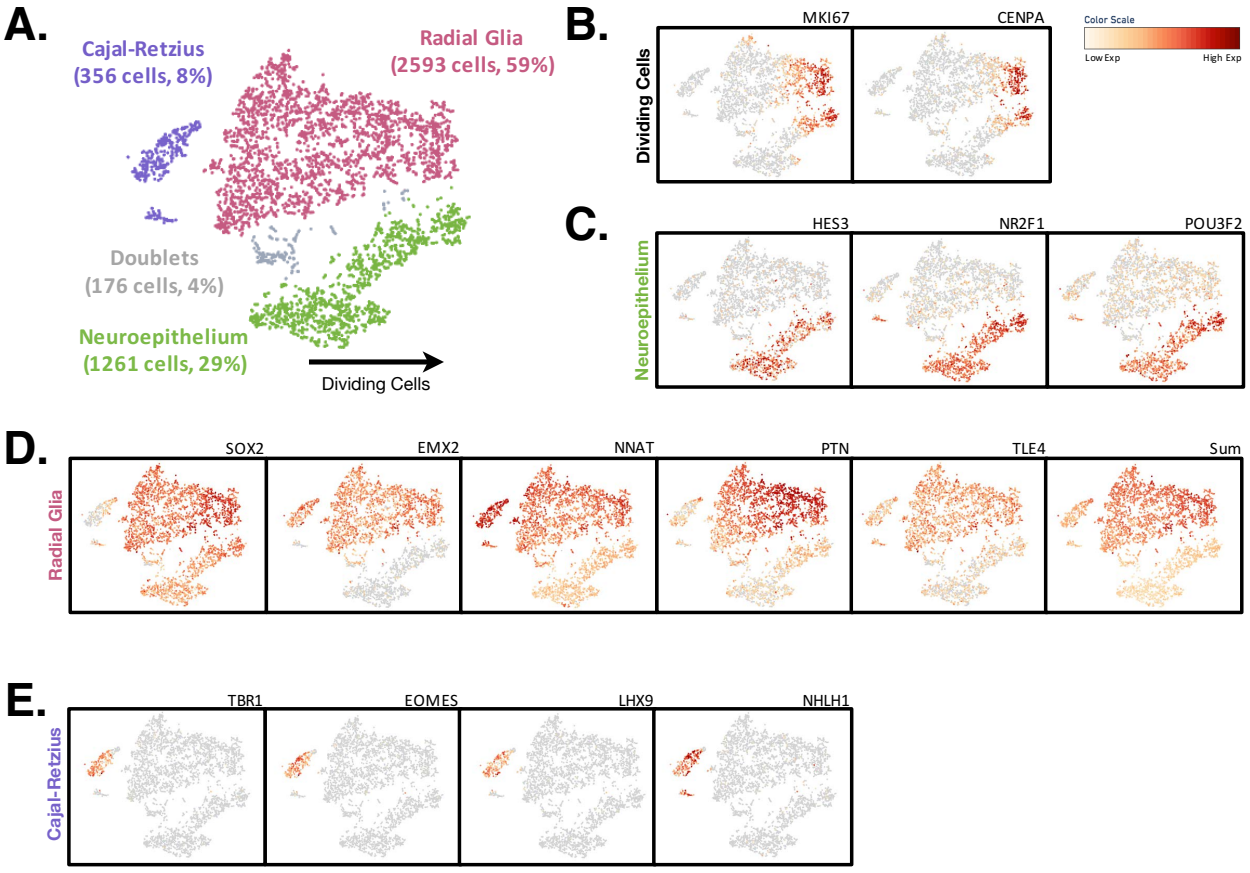

Figure S6, related to Figure 4. Cell type detection in week 2 cortical organoid single cell RNA-seq. (A) Shown is a tSNE plot of cell types detected in week 2 single cell RNA-sequencing libraries. Putative cell types were manually curated by a combination of K-means clustering, Louvain graphical clustering, and canonical cell type marker expression. Both the radial glia and neuroepithelium clusters showed expression of *MKI67* and *CENPA* concentrated toward one end of the tSNE plot indicating dividing cells as would be expected for these cell populations (B). 1261 cells (29%) most closely fit undifferentiated neuroepithelial cells with strong expression of HES3 and NR2F1 (C). Though we also see POU3F2, which is commonly associated with midbrain development, since we do not see midbrain markers later in differentiation, it is possible that this cell state occurs prior to brain region specification. The largest fraction of cells at this time point, 2593 cells (59%), exhibited the radial glia markers *SOX2*, *EMX2*, *NNAT*, *PTN*, and *TLE4* (D). 356 cells (8%) were found to have transcriptional profiles consistent with Cajal-Retzius cells predominantly expressing *TBR1*, *EOMES*, *LHX9*, and *NHLH1* (E).

Table S1, related to Figure 3. Gene model expression and intron boundary conservation analysis. Human, chimpanzee, orangutan, and rhesus Cufflinks generated transcripts were analyzed for their conservation in other species' genomes. Each transcript model was tested for mapping to other genomes ("mappable_genomes"), whether they had Cufflinks generated transcript models that spanned at least one common intron junction in those genomes ("annotated_genomes"), and whether they met a minimal expression threshold of 0.1TPM ("expressed_genomes").

Table S2, related to Figure 3. Novel Cufflinks predicted transcripts expression and intron boundary conservation. Cufflinks transcripts with no overlap with ENSEMBL and FANTOM5 were analyzed for transcript structure and expression conservation in other species. The separate tabs represent unique transcripts expressed in each respective species. Each transcript model was tested for mapping to other genomes ("mappable_genomes"), whether they had Cufflinks generated transcript models that spanned at least one common intron junction in those genomes ("annotated_genomes"), and whether they met a minimal expression threshold of 0.1TPM ("expressed_genomes").

Table S3, related to Figure 3. TrEx conservation analysis. Human TrEx lncRNAs were defined by max expression at weeks 1 through 4 and expression levels below 50% of maximum at weeks 0 and 5 as determined by RSEM TPM values. LncRNAs that were conserved in other species by maximum expression greater than 0.1 TPM and shared at least one intron junction with a human transcript model were tested for their conservation of a TrEx pattern in other genomes. TrEx patterns were considered conserved ("1" for "true" and "0" for "false") if max expression for the species was at weeks 1 through 4 and below 50% maximal expression level at week 0 as determined by RSEM TPM values. Blank cells indicate that the human locus was not mappable in the respective genome.

Table S4, related to Figure 4. Week 2 organoid manually curated single cell RNA-seq clusters and top distinguishing genes. The first tab lists the cell barcodes associated with each manually curated cell type. The second tab lists the top 20 distinguishing genes of each manually curated cell type using "globally distinguishing genes" tool in the 10X Loupe Cell Browser v1.0.0 (10X Genomics).

Table S5, related to Figure 4, 5, and 6. Cell Ranger top correlated and anti-correlated genes in cells expressing TrEx lncRNAs. Each tab represents one of the 8 TrEx lncRNAs, respectively. Cells expressing each of the lncRNAs were grouped into a curated cluster and compared to cells with no expression of the lncRNA ("no exp") using the "locally distinguishing genes" tool in the 10X Loupe Cell Browser v1.0.0 (10X Genomics). The top 20 and bottom 20 genes correlated and anti-correlated with cells expressing each TrEx lncRNA are displayed.

Table S6, related to Figure 4, 5, and 6. Pairwise Pearson correlations of lncRNAs to other genes in week 2 organoids. Each tab represents one of the 8 target TrEx

lncRNAs, respectively. Pearson correlations were calculated for all genes compared pairwise to TrEx lncRNA expression in week 2 organoids. Gene correlations with a p-value lower than 1E-09 are displayed.

Table S7, related to methods. Materials list. Primer sequences, sgRNA guide sequences, primary and secondary antibodies used in this study are provided on separate tabs, respectively.