Stem Cell Reports, Volume *12*

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, Mari Olsen, Max Hauessler, Sol Katzman, Sofie R. Salama, and David Haussler

Supplemental Information

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

Andrew R. Field[1,2], Frank M.J. Jacobs[3,7], Ian T. Fiddes[2,3], Alex P. R. Phillips[3], Andrea M. Reyes-Ortiz[3], Erin LaMontagne[3], Lila Whitehead[1], Vincent Meng[1], Jimi L. Rosenkrantz[4], Mari Olsen[4], Max Hauessler[2,3], Sol Katzman[2], Sofie R. Salama[2,3,4,5,6], David Haussler[2,3,4,5]

[1]Molecular, Cell, and Developmental Biology, University of California Santa Cruz, Santa Cruz, California, United States of America
[2]Genomics Institute, University of California Santa Cruz, Santa Cruz, California, United States of America
[3]Biomolecular Engineering, University of California Santa Cruz, Santa Cruz, California, United States of America
[4]Howard Hughes Medical Institute, University of California, Santa Cruz, Santa Cruz, California, United States of America
[5]Senior and Corresponding Authors
[6]Lead contact
[7]Present address: University of Amsterdam, Swammerdam Institute for Life Sciences, Amsterdam, The Netherlands

Supplemental Experimental Procedures

iPSC generation

Primate primary fibroblasts were grown as adherent cultures in MEM Alpha (ThermoFisher) supplemented with 10% Gibco FBS (ThermoFisher) and 1% Pen-Strep (ThermoFisher). Integration-free chimpanzee induced pluripotent stem cells were produced at Applied StemCell from S008919 primary fibroblasts (Yerkes Primates, Coriell) by episomal reprogramming using the Y4 plasmid combination described in Okita et al., 2011. Integration-free Sumatran orangutan induced pluripotent stem cells were generated using the CytoTune 2.0 Sendai Reprogramming kit (ThermoFisher) from 11045-4593 primary fibroblasts obtained from the Frozen Zoo® (http://institute.sandiegozoo.org/resources/frozen-zoo%C2%AE). Both chimpanzee and orangutan iPSCs were initially established on mouse embryonic fibroblasts with KSR-15 (KO DMEM/F-12 + 20% KOSR, 1% NEAA, 1% GlutaMAX, 1% Pen-Strep, and 0.1mM 2-mercaptoethanol supplemented with 15 ng/mL bFGF) media and were transferred to feeder-free conditions on Matrigel (Corning) with mTeSR-1 (Stem Cell Tech) for chimpanzee or vitronectin (ThermoFisher) with Essential-8 Flex (ThermoFisher) for orangutan. Pluripotency was confirmed by immunofluorescence staining of pluripotency markers, RT-PCR, teratoma assay, and karyotype (Fig. S1).

Teratoma Assay

Mice were anesthetized by intraperitoneal injection with 100mg/kg ketamine (MWI Veterinary Supply). 2 subcutaneous injections of 1 to 5 million cells suspended in 30% Matrigel (Corning) were made in the dorsolateral or ventral lateral areas of NOD-SCID mice (NOD.CB17-Prkdc^{scid}/NCrCrl, Charles River) similar to Prokhorova et al., 2009.

Mice were observed for up to 12 weeks for the appearance of tumors in the injected areas. The animals were euthanized by cervical dislocation and teratomas were harvested, fixed in 4% paraformaldehyde, saturated in 30% sucrose in PBS, embedded in Tissue Freezing Medium™ (Triangle Biomedical Sciences), and frozen for cryostat sectioning. Sections of the tumors were stained with hematoxylin (Mayer's Hematoxylin Solution, Sigma) & eosin (Eosin Y solution, Sigma) and analyzed for the generation of all three germ layers.

Karyotyping

Chimpanzee and orangutan iPSC lines were confirmed to have a stable wild type 48/XX karyotype through at least passage 32 or 36, respectively (Fig. S1I&J). Karyotyping services were performed by Cell Line Genetics or the Coriell Institute for Medical Research.

Cortical organoid generation

The Eiraku et al., 2008 protocol was optimized for use with human ESCs, rhesus ESCs, chimpanzee iPSCs, and orangutan iPSCs. Human H9 and rhesus LyonESC1 embryonic stem cells were cultured on mouse embryonic fibroblasts with KSR-8 media (KO DMEM/F-12 + 20% KOSR, 1% NEAA, 1% GlutaMAX, 1% Pen-strep, and 0.1mM 2-mercaptoethanol supplemented with 8ng/mL bFGF). Embryonic stem cells were manually lifted from MEF feeders and allowed to self-form into embryoid bodies on low attachment plates (Corning) in KSR media. Chimpanzee and orangutan induced pluripotent stem cells were grown in feeder-free conditions on Matrigel (Corning) with

mTeSR-1 (Stem Cell Tech) or vitronectin (ThermoFisher) with Essential-8 Flex media (ThermoFisher), respectively, and 10,000 cells per EB were aggregated using AggreWell-800 plates (Stem Cell Technologies) in Aggrewell media (Stem Cell Technologies) supplemented with 10 uM Y-27632 rock inhibitor (ATCC) and transferred to low attachment plates (Corning) on day 2. Both methods supplemented the respective media with 500ng/mL DKK1 (Peprotech), 500 ng/mL NOGGIN (R & D Systemes), 10 uM SB431542 (Sigma), and 1 uM Cyclopamine V. californicum (VWR) for the first 18 days of differentiation. The media was changed to Neurobasal (Invitrogen) supplemented with N2 (Gibco) and 1 uM Cyclopamine on day 18. At this time, chimpanzee and orangutan neurospheres were also supplemented with 10ng/mL bFGF and 10ng/mL EGF to improve survivability in Neurobasal media. After day 26, all cultures were grown in Neurobasal/N2 media without any added factors. Total RNA was extracted at weekly time points for each species. This timeline was adjusted accordingly in rhesus, harvesting on days 6, 11, 17, 22, and 28, to account for differences in gestational timing (described below).

Adjustment of rhesus macaque time points

When optimizing the ESC cortical neurosphere differentiation protocol for macaque ES cells, we noticed that macaque ESC colonies and neurospheres grow faster and form neural rosettes earlier than human. Indeed, macaque ESCs are reported to have a faster doubling time (~0.8x that of hESCs; Amit et al., 2000; Fluckiger et al., 2006). Macaque also has a shorter gestation period and a predicted faster progression of neurodevelopmental events (Workman et al., 2013). Therefore, in order to most reliably

compare the gene expression events that occur during human and macaque ESC cortical neurosphere differentiation we needed to adjust for the intrinsic difference in neurodevelopmental speed.

A time point adjustment was implemented based on semi-quantitative RT-PCR analysis comparing expression levels of *TBR1* and *CTIP2* over the course of ESC cortical neurosphere differentiation (Fig. S2). The best linear fit of human and macaque expression level dynamics is obtained when the relative timing for macaque cortical neurosphere differentiation is adjusted by a factor of 0.8, whereby human day 21 is most comparable to macaque day 17 and human day 35 is most comparable to macaque day 28 (Fig. S2A; right panel). Human, chimpanzee, and orangutan time points are indicated in weeks where the equivalent macaque time points are the time point most comparable to the corresponding human week: W1*, W2*, W3*, W4*, and W5* (days 6, 11, 17, 22, 28, respectively). No adjustment was found necessary for the other species in this study.

Immunofluorescence Staining

Cortical organoids were fixed for 15 minutes in 4% paraformaldehyde and saturated in 30% sucrose for 24 hours prior to being embedded in Tissue Freezing Medium™ (Triangle Biomedical Sciences) and frozen for cryostat sectioning. Sections of 16-18uM were adhered to glass slides and fixed a second time in 4% paraformaldehyde for 10 minutes. Cells in 2-dimensional culture were grown on acid etched coverslips and fixed for 10 minutes with 4% paraformaldehyde prior to staining. Samples were incubated at

4°C in blocking solution (3% BSA and 0.1% Triton X-100 in PBS) for 4 hours. Primary antibody incubation was performed overnight at 4°C in blocking solution. Secondary antibody incubation was for 1-4 hours at room temperature in blocking solution. Samples were mounted with SlowFade® Gold antifade reagent (Invitrogen).

Primate Genome Alignment and Annotation

A progressive Cactus (Paten et al., 2011) whole genome alignment was generated between the human hg19 assembly, chimpanzee panTro4 assembly, orangutan ponAbe2 assembly, and rhesus macaque rheMac8 assembly. This alignment was used as input to the Comparative Annotation Toolkit (https://github.com/ComparativeGenomicsToolkit/Comparative-Annotation-Toolkit, citation pending) along with the FANTOM5 (Hon et al., 2017) lv3 annotation set. This process projects the annotations from human to the other primates in the alignment. Subsequent filtering and post-processing produces a high quality comparative annotation set. RNA-seq obtained from SRA (https://www.ncbi.nlm.nih.gov/sra) were used to help guide the annotation process. The resulting annotations were used for the initial differential expression analysis below. Filtered FANTOM transcript models available on GEO (GSE106245): hg19.fantom.lv3.filtered.gtf.gz (human), panTro4.fantom.lv3.transmap.filtered.gtf.gz (chimpanzee), ponAbe2.fantom.lv3.transmap.filtered.gtf.gz (orangutan), rheMac8.fantom.lv3.transmap.filtered.gtf.gz (rhesus).

RNA-Seq library preparation

Total RNA was collected from our organoid cultures by TRIzol (ThermoFisher) extraction and depleted of ribosomal RNA by Ribo-zero (Epicentre). Bulk strand-specific total-transcriptome RNA sequencing libraries were prepared using dUTP marking during second strand synthesis either with the TruSeq Stranded Total RNA Library Prep Gold kit (Illumina) or homebrew components (Parkhomchuk et al. 2009).

RNA-Sequencing Analysis

Paired-end Illumina reads were trimmed from the 3' end of read1 and read2 to 100x100bp for human and rhesus libraries and 80x80bp for chimpanzee and orangutan based on sequence quality. Bowtie2 v2.2.1 (Langmead et al., 2012) was used with the "--very-sensitive" parameter to filter reads against the repeatMasker library (Smit et al., 2015) for each respective species which were removed from further analysis. STAR v2.5.1b (Dobin et al., 2012) was used to map RNA-seq reads to hg19 (human, Genome Reference Consortium GRCh37, 2009), panTro4 (chimpanzee, CGSC Build 2.1.4, 2011), ponAbe2 (orangutan, WUSTL Pongo_albelii-2.0.2, 2007), and rheMac8 (rhesus macaque, Baylor College of Medicine HGSC Mmul_8.0.1, 2015) respective to the origin species. STAR was run with the default parameters with the following exceptions: --outFilterMismatchNmax 999, --outFilterMismatchNoverLmax 0.04, --alignIntronMin 20, --alignIntronMax 1000000, and --alignMatesGapMax 1000000. STAR alignments were converted to genomic position coverage with the bedtools command genomeCoverageBed -split (Quinlan et al., 2010). Coverage for each gene in a gene model for its species was derived by summing the position coverage over all the exonic positions of the gene as defined by the annotation sets (GEO, GSE106245:

hg19.fantom.lv3.filtered.gtf.gz [human], panTro4.fantom.lv3.transmap.filtered.gtf.gz [chimpanzee], ponAbe2.fantom.lv3.transmap.filtered.gtf.gz [orangutan], rheMac8.fantom.lv3.transmap.filtered.gtf.gz [rhesus]) for initial analysis. DESeq2 v1.14.1 (Love et al., 2014) was used to provide basemean expression values and differential expression analysis across the time course in each species. Total gene coverage for a gene was converted to read counts by dividing the coverage by N+N (100+100 for human and rhesus and 80+80 for chimpanzee and orangutan) since each paired-end NxN mapped read induces a total coverage of N+N across its genomic positions. Results are available in Table S1 and at GEO, GSE106245: GSE106245_deseq_baseMean_allF2.txt.gz. A visualization of this data is available at the UCSC Genome Browser as a Track Hub in the Public Hubs section, with Hub Name: Primate x4 NeuroDiff and Human CRISPRa.

LncRNA annotation analysis, structure conservation, and expression estimates

Cufflinks v2.0.2 suite (Trapnell et al., 2010; Trapnell et al., 2012) was used to assemble transcript predictions of potentially unannotated lncRNAs in each species and the Cuffmerge tool was used to combine these annotations with FANTOM5 transcripts. The resulting Cufflinks-assembled and merged transcript sets were then projected through the cactus alignment (Stanke et al., 2008) to each of the other three genomes. Guided by the Cufflinks annotation set in each genome, these projections from the other genomes were assigned a putative gene locus (GEO, GSE106245: hg19.cufflinks.filtered.gtf.gz [human], panTro4.cufflinks.filtered.gtf.gz [chimpanzee], ponAbe2.cufflinks.filtered.gtf.gz [orangutan], rheMac8.cufflinks.filtered.gtf.gz [rhesus]).

Transcripts were assigned a lncRNA ID number based on overlap with block loci defined by co-expression of transcripts within 10kb at at least one time time point, expression from the same strand, uninterrupted by antisense transcription, and no overlap with known protein-coding genes ("lncRNA ID" column in Table S1 and in GEO, GSE106245).

RSEM v1.3.0 (Li and Dewey, 2011) was used to provide TPM expression values for these newly generated transcripts (Table S1, GSE106245: GSE106245_hg19.RSEM.gene_expression.tsv.gz [human], GSE106245_panTro4.RSEM.gene_expression.tsv.gz [chimpanzee], GSE106245_ponAbe2.RSEM.gene_expression.tsv.gz [orangutan], GSE106245_rheMac8.RSEM.gene_expression.tsv.gz [rhesus]). Expressed lncRNAs were assessed using the homGeneMapping tool from the AUGUSTUS toolkit (Konig et al., 2016). homGeneMapping makes use of cactus alignments to project annotation features in all pairwise directions, providing an accounting of features found in other genomes. homGeneMapping was provided both the Cufflinks transcript assemblies as well as expression estimates derived from the combination of the week 0 to week 5 RNA-seq experiments in all four species. The results of this pipeline were combined with the above transcript projections to ascertain a set of lncRNA loci that appear to have human specific expression, human-chimp specific expression, great-ape specific expression, and expressed in all primates (Table S2 & S3). For this analysis, a locus was considered expressed in the current reference genome if one or more transcripts had RNA-seq support for every single one of its intron junctions, and considered

expressed in another genome if the transcripts that mapped from that genome to the current reference had RNA-seq support for any of its intron junctions. All single-exon transcripts were filtered out to reduce noise. To eliminate the possibility of the specificity results being skewed by assembly gaps or alignment error, loci which appeared to have sub-tree specific expression were checked against the cactus alignment to ensure that there was a matching locus in each other genome. If a genome appeared to be missing sequence, then this locus was flagged as having incomplete information.

3' Single Cell RNA-sequencing

Human H9 embryonic stem cells were grown on vitronectin with E8-Flex media (ThermoFisher). 10,000 cells per EB were aggregated using AggreWell-800 plates (Stem Cell Technologies) in Aggrewell media (Stem Cell Technologies) supplemented with 10 uM Y-27632 rock inhibitor (ATCC) and transferred to low attachment plates (Corning) on day 2. Aggrewell media was supplemented with 500ng/mL DKK1 (Peprotech), 500 ng/mL NOGGIN (R & D Systemes), 10 uM SB431542 (Sigma), and 1 uM Cyclopamine V. californicum (VWR) for the first 18 days of differentiation. The media was changed to Neurobasal (Invitrogen) supplemented with N2 (Gibco), 1 uM Cyclopamine, 10ng/mL bFGF, and 10ng/mL EGF on day 18. After day 26, neurospheres were grown in Neurobasal/N2 media without any added factors. Single cell suspensions for 10X Genomics Chromium scRNA-seq were prepared with TrypLE (ThermoFisher) and handled according to the 10X protocol RevA (version 1 chemistry) for undifferentiated hESCs and week 5 cortical organoids and RevB (version 2 chemistry) for weeks 1 and 2 cortical organoids. Cell count, quality, and viability was

assessed using Trypan Blue (ThermoFisher) on a TC20 automated cell counter (BioRad). Single cell suspensions were made aiming for 1500-3000 cells captured per library. The data was analyzed by CellRanger v1.2 (10X Genomics) using a custom annotation set based on FANTOM5 lv3 (Hon et al., 2017) (GEO, GSE106245: hg19.fantom.lv3.filtered.gtf.gz) and visualized using the Loupe Cell Browser v1.0.0 (10X Genomics). The gene by cell data matrices are available in supplemental data files in matrix market exchange format (http://math.nist.gov/MatrixMarket/formats.html) (GEO, GSE106245: sch9froz_wk01_af111_filtered_gene_bc_matrix.tgz [week 1 gene by cell matrix], sch9froz_wk01_af111_possorted.bam [week 1 position sorted BAM file], sch9froz_wk01_af112_filtered_gene_bc_matrix.tgz [week 1 gene by cell matrix], sch9froz_wk01_af112_possorted.bam [week 1 position sorted BAM file], sch9wild_wk00_af104_filtered_gene_bc_matrix.tgz [week 0 gene by cell matrix], sch9wild_wk00_af104_possorted.bam [week 0 position sorted BAM file], sch9wild_wk02_af106_filtered_gene_bc_matrix.tgz [week 2 gene by cell matrix], sch9wild_wk02_af106_possorted.bam [week 2 position sorted BAM file], sch9wild_wk02_af107_filtered_gene_bc_matrix.tgz [week 2 gene by cell matrix], sch9wild_wk02_af107_possorted.bam [week 2 position sorted BAM file], sch9wild_wk05_af102_filtered_gene_bc_matrix.tgz [week 5 gene by cell matrix], sch9wild_wk05_af102_possorted.bam [week 5 position sorted BAM file], sch9wild_wk05_af103_filtered_gene_bc_matrix.tgz [week 5 gene by cell matrix], sch9wild_wk05_af103_possorted.bam [week 5 position sorted BAM file]).

Cell clusters were identified and manually curated by the expression of canonical cell markers using a combination of graphical and K-means clustering from the CellRanger v1.2 pipeline (10X Genomics) as a guide (Table S4).

CRISPRa assay

The CRISPR-activation (CRISPRa) assay was modified from Konermann et al., 2014. HEK293FT cells were cultured with DMEM+GlutaMAX (ThermoFisher) supplemented with 10% FBS without antibiotic. Each well of a 6-well plate was seeded with 500k cells and co-transfected at 60-70% confluence the next day using Xfect reagent (Takara) with dCas9-VP64_Blast (Feng Zhang, addgene #61425), MS2-p65-HSF1_Hygro (Feng Zhang, addgene #61426), and a combination of 5 custom guide RNAs per target in the custom plasmid 783 (gift from S. Carpenter, UCSC) for a total of 7.5 g DNA in a ratio of 1:1:2, respectively. Guides were designed so there were no potential off-target sites with 1 mismatched base and those with 2 mismatches were not within 1kb of a known transcription start site. Transfected cells were selected at 24 hours by incubation with 2 g/mL puromycin until harvest. RNA was harvested at 48 hours after transfection using TRIzol reagent (ThermoFisher) and RNA was extracted using Direct-zol columns (ZYMO). Quantitect SYBR® Green RT-PCR (Qiagen) was used with 50ng of total RNA per reaction, 4 replicates per condition. Relative expression was calculated by ddCt normalized to HEK293FT transfection with non-targeting scrambled control guides. RNA-seq libraries were prepared in biological triplicates with the NEXTflex Rapid Directional qRNA-Seq Library Prep Kit (PerkinElmer). Differential expression analysis was performed as described above with reads trimmed to 90x90bp and using the

human Cufflinks generated transcripts (hg19.cufflinks.filtered.gtf.gz) for DESeq quantification (Table S1). Results are available at GEO, GSE120702. A visualization of this data is available at the UCSC Genome Browser as a Track Hub in the Public Hubs section, with Hub Name: Primate x4 NeuroDiff and Human CRISPRa.

Table S5. Antibodies, guides and primers. Details regarding antibodies, the targeting guide sequences used in plasmid 783 targeting each TrEx lncRNA (or random SCRAM sequence) and primers used in this study.

Immunofluorescence Primary Antibodies				
Target	**Animal**	**Source**	**Catalog Number**	**Dilution Used**
CTIP2/BCL11B	rat	Abcam	ab18465	1:500
Nanog	rabbit	Abcam	ab21624	1:1000
OCT3/4	rabbit	Abcam	ab19857	1:500
OCT3/4	rabbit	Santa Cruz Biotech	sc-9081	1:50
PAX6	mouse	DSHB	na	1:100
PAX6	mouse	Santa Cruz Biotech	sc-53108	1:50
Sendai Virus Particle	rabbit	MBL	PD029	1:1000
SOX2	mouse	Abcam	ab79351	1:100
SOX2	rabbit	Abcam	ab97959	1:1000
SOX2	mouse	Santa Cruz Biotech	sc-398254	1:50
SSEA4	mouse	Abcam	ab16287	1:100
TBR1	rabbit	Abcam	ab31940	1:200
TBR2/EO	rabbit	Abcam	ab31940	1:200

MES				
Tra-1-60	mouse	Abcam	ab16288	1:500
Tra-1-81	mouse	Abcam	ab16289	1:200
Tra-1-81 DyLight 488 live stain	pre-conjugated	Stemgent	cat#09-0069	1:100
Immunofluorescence Secondary Antibodies				
Fluorophore	**Raised in**	**Raised against**	**Company**	**Catalog Number**
Cy5	Donkey	Mouse	Jackson ImmunoResearch Laboratories, Inc	715-175-150
Cy3	Donkey	Rabbit	Jackson ImmunoResearch Laboratories, Inc	711-165-152
AlexaFluor 488	Donkey	Mouse	ThermoFisher	A-21202
AlexaFluor 594	Donkey	Rat	ThermoFisher	A-21209
AlexaFluor 647	Donkey	Rabbit	Abcam	ab150075
AlexaFluor 488	Donkey	Rat	ThermoFisher	A-21208
AlexaFluor 555	Donkey	Rabbit	Abcam	ab150074
AlexaFluor 647	Donkey	Mouse	ThermoFisher	A-31571
Targeting guide sequences				

SCRAM 2	AAGATGAAAG GAAAGGCGTT			
SCRAM F7	GTCCATACGC ATAATCACCG			
SCRAM F8	ACTTACCTCC GGACCCCCAT			
SCRAM F9	CCTACACGAC GAACGCAGGT			
TrEx108 sgR1	AATCTTGATTC CTCTCCTAT			
TrEx108 sgR2	CTGCAAACAT ATAAGTTAGA			
TrEx108 sgR3	TTTGTGTGCA CTACAGAGGA			
TrEx108 sgR4	ACAGACACCA ATGCTTTAAA			
TrEx108 sgR5	TACTTAATCCT TCGTGACTA			
TrEx4039 sgR1	CTTACCCTGA ACAATTAGAG			
TrEx4039 sgR2	GAGATCCTCT CTAATTGTTC			
TrEx4039 sgR3	AATTGATCACT GCTAACTCC			
TrEx4039 sgR4	CTCTCTAATTG TTCAGGGTA			
TrEx4039 sgR5	GAATGCACTT ATCATCAGCA			
TrEx5008 sgR1	GTGGCTACAT TTCTTTGCAC			

TrEx5008 sgR2	GTTAAATATCC CTGTGCTTG			
TrEx5008 sgR3	TAACAACTAAA AGAGCTGGT			
TrEx5008 sgR4	CATTGCAATG ACAAGTGAAT			
TrEx5008 sgR5	CATGTCTACTA CAAATCTTA			
TrEx8168 sgR1	CGGGAGTTGA GGGTGCCGAG			
TrEx8168 sgR2	AGGGCCGGG ATGCTGGTGC C			
TrEx8168 sgR3	AGTCAGGTCC ACGGGAGAGC			
TrEx8168 sgR4	AAGACCATGC TGAAGGATAA			
TrEx8168 sgR5	GCCTTTGGTT TCCATGCAGC			
PCR Primers				
Gene ID	**Use**	**Fwd Sequence**	**Rev Sequence**	
EBNA-1 Plasmid	PCR	ATCAGGGCCAAGACAT AGAGATG	GCCAATGCAACTT GGACGTT	
Episomal KLF4	PCR	CCACCTCGCCTTACAC ATGAAGA	TAGCGTAAAAGG AGCAACATAG	
Episomal L-MYC	PCR	GGCTGAGAAGAGGATG GCTAC	TTTGTTTGACAGG AGCGACAAT	
Episomal LIN28	PCR	AGCCATATGGTAGCCT CATGTCCGC	TAGCGTAAAAGG AGCAACATAG	
Episomal	PCR	CATTCAAACTGAGGTA	TAGCGTAAAAGG	

OCT3/4		AGGG	AGCAACATAG	
Episomal SOX2	PCR	TTCACATGTCCCAGCACTACCAGA	TTTGTTTGACAGGAGCGACAAT	
GAPDH	RT-PCR	CCAGGTGGTCTCCTCT	CCCTGTTGCTGTAGCC	
L-MYC	RT-PCR	GCGAACCCAAGACCCAGGCCT	CAGGGGGTCTGCTCGCACCGT	
NANOG	RT-PCR	TTTCAGAGACAGAAATACCTC	TCACACCATTGCTATTCTTCG	
OCT3/4	RT-PCR	CTTGCTGCAGAAGTGGGTGGAGGAA	CTGCAGTGTGGGTTTCGGGCA	
SeV coat protein	RT-PCR	GGATCACTAGGTGATATCGAG	ACCAGACAAGAGTTTAAGAGA	
TrEx108	RT-PCR	ATTCTGTGGAGGGAGGGACT	TGCAGCATTTGCTTACCTTG	
TrEx4039	RT-PCR	CAGGGAAAGCCTGCAATTTA	TAATGCTTGCCGACTCATCA	
TrEx5008	RT-PCR	GTGACACAGACAGGCGACAG	GTGCTTCCAGTTGTTGCAGA	
TrEx8168	RT-PCR	CAGGCCAGACAGAGGAGATT	GCGGTAAGGTGGACTAGCAA	

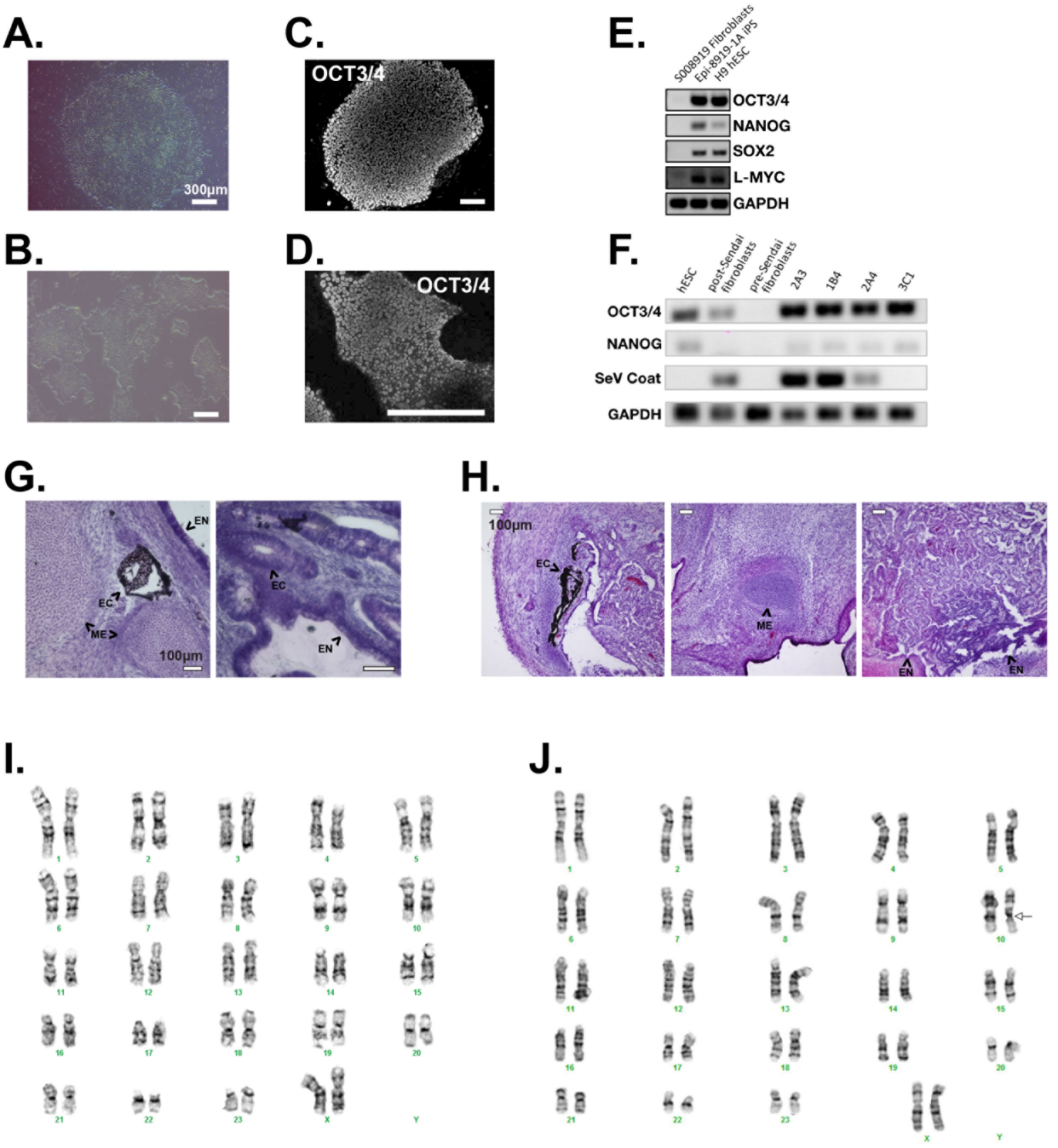

Figure S1, related to Figure 1. Chimpanzee and Orangutan iPSC verification. Brightfield images of chimpanzee Epi-8919-1A (A) and orangutan 3C1 (B) iPSC colonies show ESC morphology on feeder-free conditions. Immunofluorescence staining of Epi-8919-1A (C) and 3C1 (D) iPSCs display nuclear *OCT3/4* expression across the entire colony. Scale bars indicate 300µm (A-D). (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) 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. Haemotoxylin and Eosin staining of teratomas derived from chimpanzee (G) and orangutan (H) iPSC lines shows the generation of all three germ layers: (EC) ectoderm (neural rosettes and pigmented cells), (EN) endoderm (gut), and (ME) mesoderm (cartilage). Scale bars indicate 100μm (G-H). Wildtype 48, XX karyotype was confirmed in chimpanzee (I) and orangutan (J) iPSCs at passage 32 and 36, respectively, after reprogramming. An inversion in chromosome 10 observed in our orangutan line is naturally occurring in the wild Sumatran orangutan population (Locke et al. 2011) and was present in the original fibroblasts.

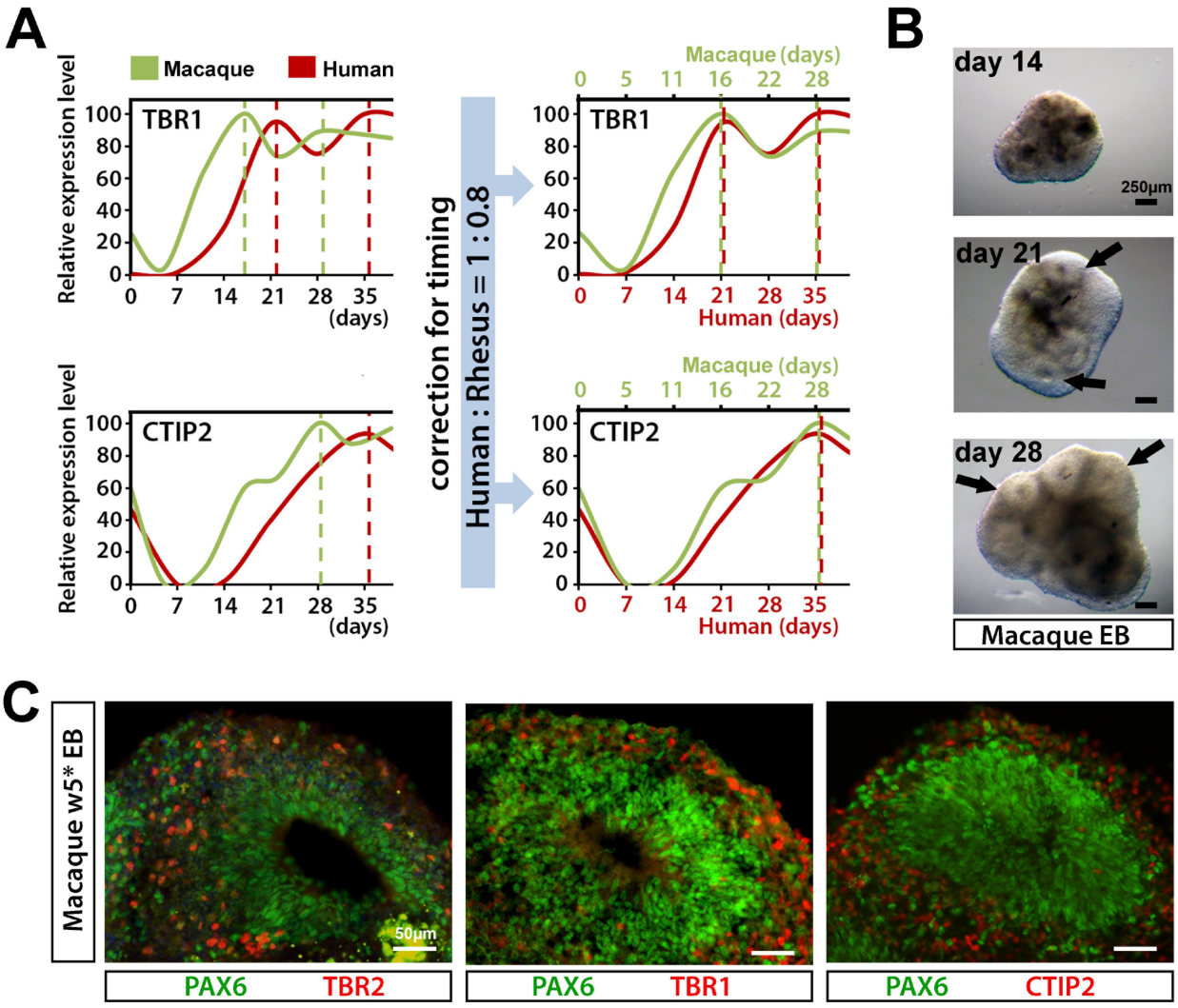

Figure S2, related to Figures 1 and 2. Determination of comparable time points between human and rhesus neurospheres. (A) Relative expression levels of TBR1 (upper graphs) and CTIP2 (lower graphs) in neurospheres isolated at multiple time points throughout human (red) and macaque (green) ESC cortical neurosphere 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) neurospheres. 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 neurospheres at various time points. Black arrows indicate neural rosettes. Scale bars indicate 250μm. (C) IF staining with cortical neuron markers at day 28 (W5*): PAX6 and TBR2 (left),TBR1 (middle) and CTIP2 (right). Scale bars indicate 50μm. EB, embryoid body (ESC cortical neurosphere).

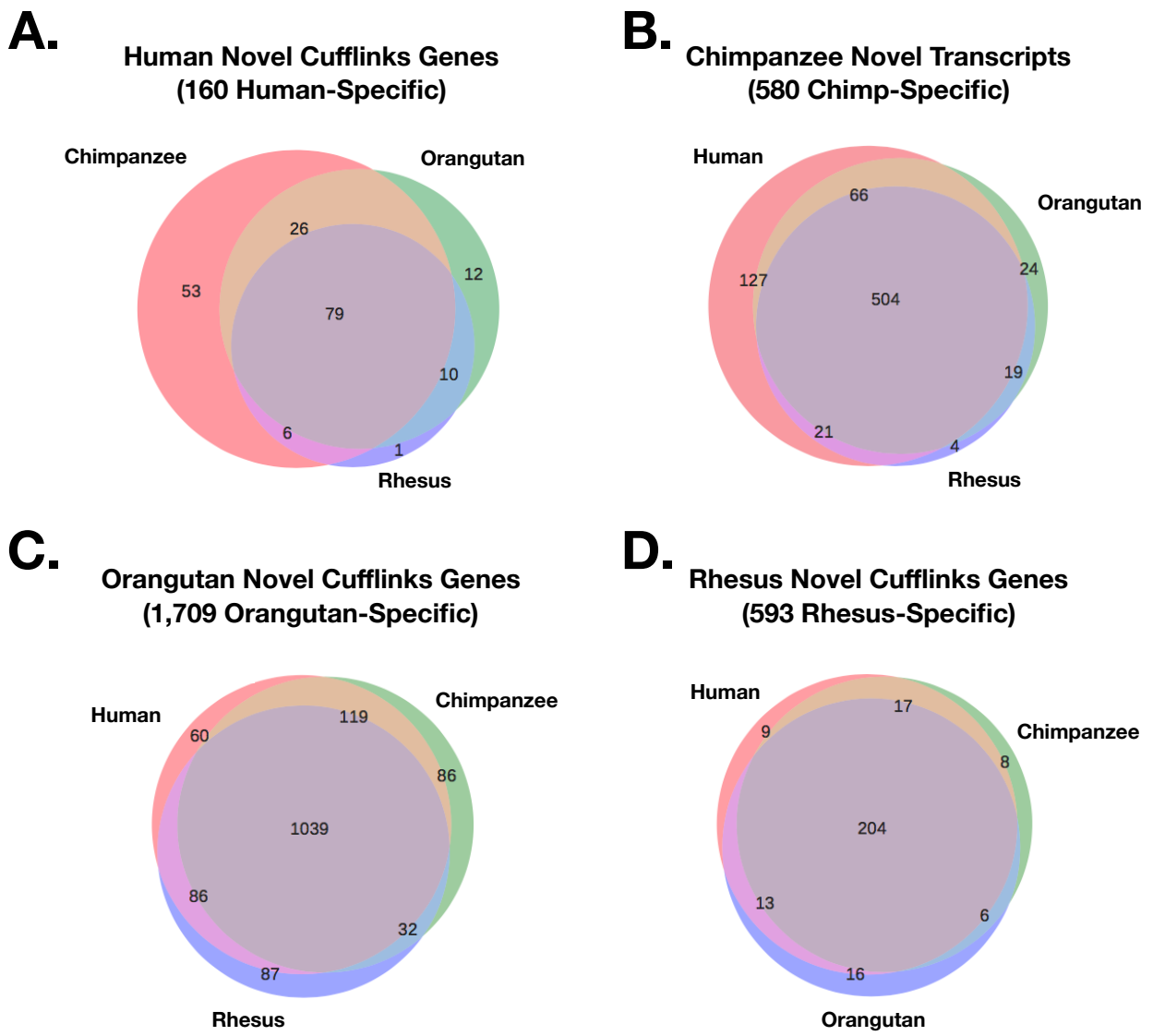

Figure S3, related to Figure 3. Conservation of novel Cufflinks gene loci detected in each species. Gene loci with no overlap to FANTOM5 lv3 (Hon, et al. 2017) (or its transMap equivalent, see Methods) were identified by Cufflinks. Unique loci were aligned pairwise using transMap to each other genome 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.

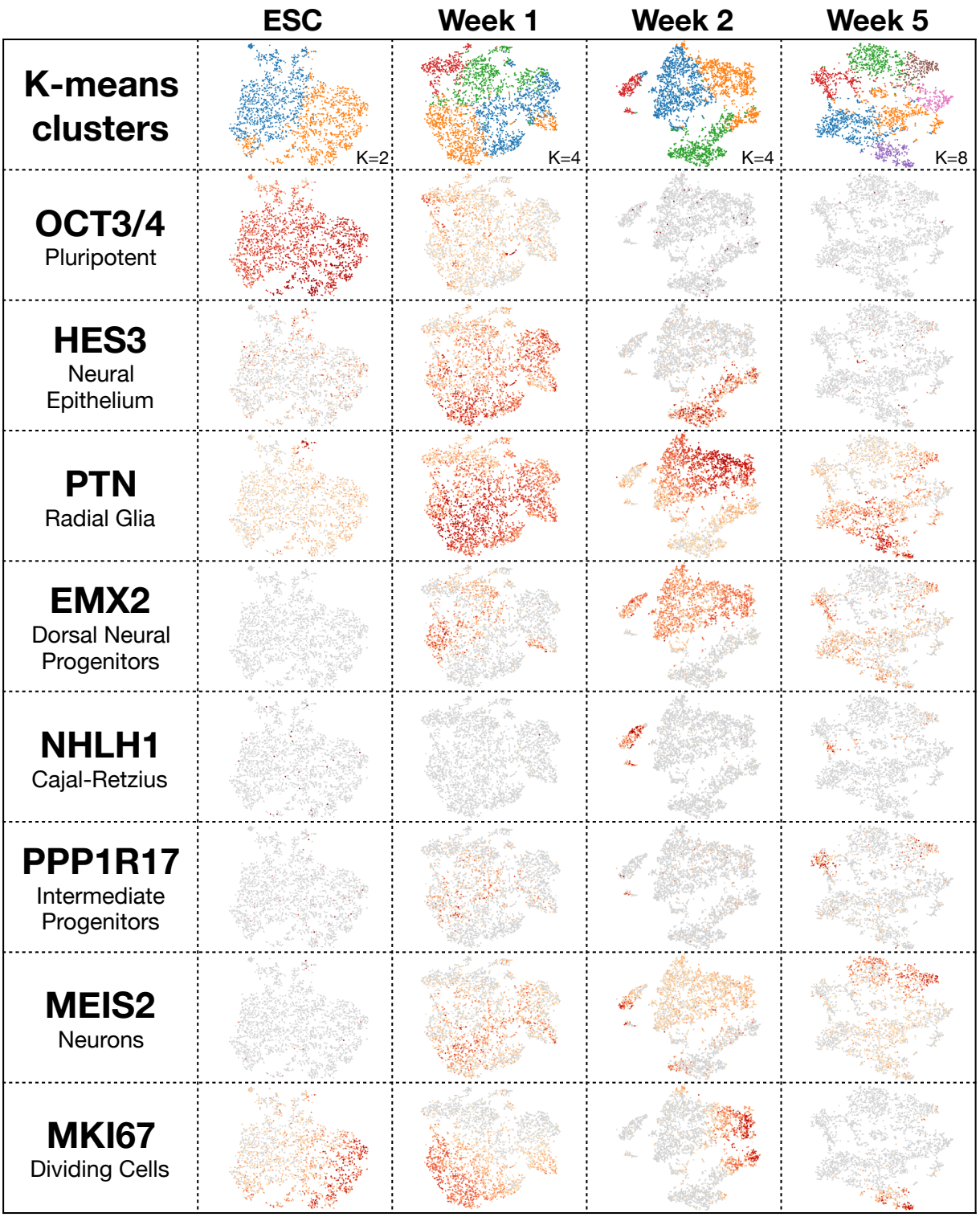

Figure S4, related to Figure 4. Human neurosphere single cell RNA-sequencing time course. t-SNE plots from single cell RNA-seq displaying the expression of *OCT3/4* (pluripotent cells), *HES3* (NE), *PTN* (RG), *EMX2* (dorsal neural progenitors), *NHLH1* (CR), *PPP1R17* (intermediate progenitors), *MEIS2* (neurons), and *MKI67* (dividing cells)

show the increasing cell heterogeneity as time progresses in weeks 0, 1, 2, and 5 of neural differentiation in from human ESCs.

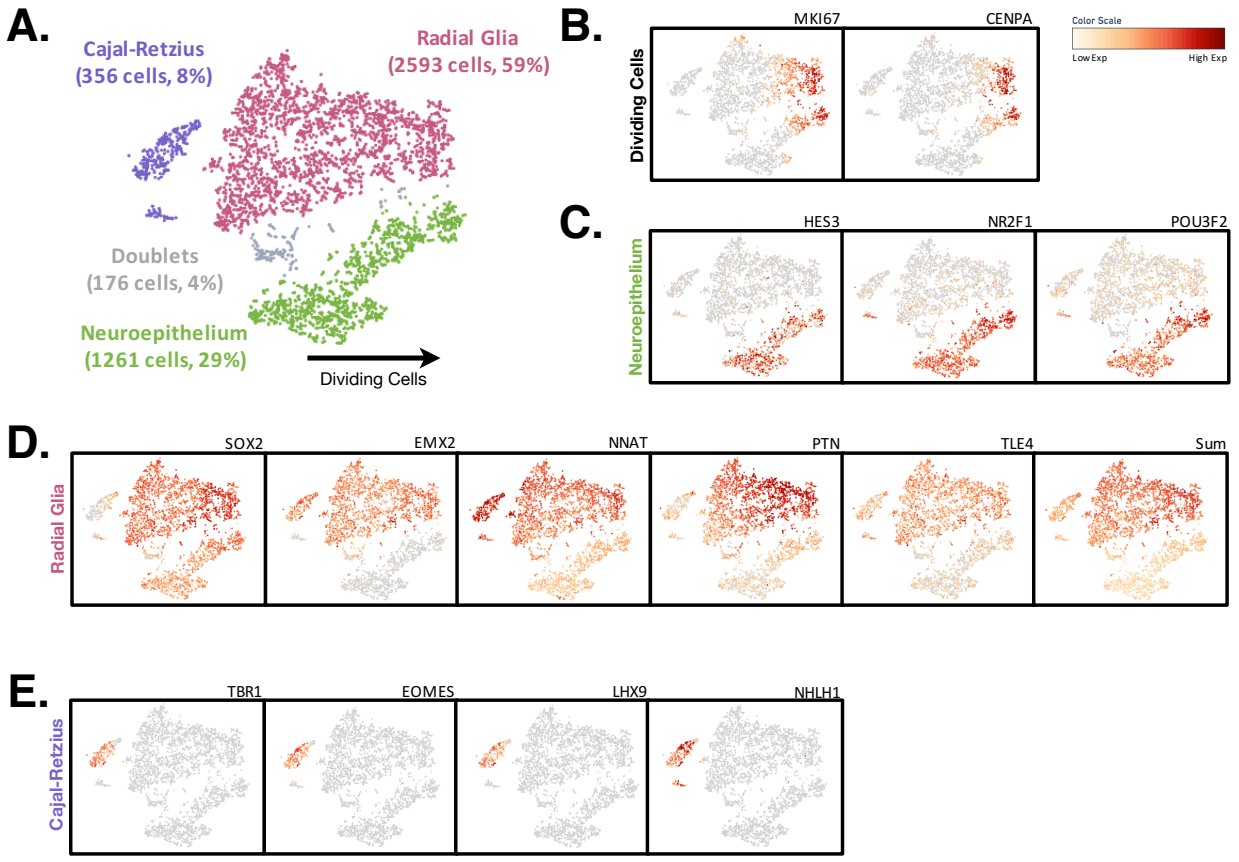

Figure S5, related to Figure 4. Cell type detection in week 2 neurosphere single cell RNA-seq. (A) Shown is a tSNE plot of cell types detected in week 2 single cell RNA-sequencing libraries. Putative cell types were manually curated by combination of K-means clustering, Louvain graphical clustering, 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).

Supplemental Data

Table S1, related to Figures 2, 3, 4, and 5. RNA-sequencing gene expression results of organoid differentiation time course and CRISPRa experiments. The first tab shows the gene expression values calculated by DESeq of cortical neuron differentiation across all species using the FANTOM5 lv3 annotations (Hon, et al. 2017). The tabs labeled "NeuroDiff_RSEM_human", "NeuroDiff_RSEM_chimpanzee", "NeuroDiff_RSEM_orangutan", and "NeuroDiff_RSEM_rhesus" display the gene expression values calculated by RSEM using Cufflinks generated transcript annotations for the human, chimpanzee, orangutan, and rhesus macaque time course respectively. The last tab labeled "CRISPRa_all" shows the DESeq expression values of each CRISPRa experiment compared against the scrambled non-targeting guide controls using the Cufflinks generated transcript annotations.

Table S2, related to Figure 3. Gene model expression and intron boundary conservation analysis for human and chimpanzee. 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"). This table lists the results in the human and chimpanzee genomes.

Table S3, related to Figure 3. Gene model expression and intron boundary conservation analysis for orangutan and rhesus. 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"). This table lists the results in the orangutan and rhesus macaque genomes.

Table S4, related to Figures 3 and 4. Novel Cufflinks predicted transcripts expression and intron boundary conservation, TrEx conservation analysis and week2 organoid manually curated single cell RNA-Seq clusters and top distinguishing genes. Tabs 1-4-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”). Tab 5-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.
Tab 6-lists the cell barcodes associated with each manually curated cell type cluster from human week 2 organoid scRNA-Seq. Tab 7 lists the top 20 distinguishing genes of each manually curated cell type cluster using "globally distinguishing genes" tool in the 10X Loupe Cell Browser v1.0.0 (10X Genomics).