Stem Cell Reports, Volume *13*

Supplemental Information

Human Cortical Organoids Expose a Differential Function of GSK3 on Cortical Neurogenesis

Alejandro López-Tobón, Carlo Emanuele Villa, Cristina Cheroni, Sebastiano Trattaro, Nicolò Caporale, Paola Conforti, Raffaele Iennaco, Maria Lachgar, Marco Tullio Rigoli, Berta Marcó de la Cruz, Pietro Lo Riso, Erika Tenderini, Flavia Troglio, Marco De Simone, Isabel Liste-Noya, Giuseppe Macino, Massimiliano Pagani, Elena Cattaneo, and Giuseppe Testa

Supplemental information

1. Supplemental figures S1 to S5 legends
2. Supplemental experimental procedures
3. References related to experimental procedures
4. Supplemental tables. List of differentially expressed genes. List of GSEA categories (Excel files).

1. Supplemental figure legends

Figure S1. Expression profiles for gene signatures defining specific developmental stages. Related to Figure 3. Stripcharts for expression levels (in log2cpm) from bulk RNA sequencing at day 18 (A), day 50 (B) and day 100 (C). Unexposed samples are visualized in turquoise, CHIR-exposed samples in orange. Blue and red boxes highlight respectively genes down- or up-regulated by CHIR treatment, considering the following thresholds from RNASeq bulk differential expression analysis: *: nominal p-value < 0.05; **: nominal p-value < 0.01; ***: FDR < 0.05 and absolute $\log_2$FC > 1.

Figure S2. Expression profiles for genes related to dorsalising/ventralizing fate. Stripcharts for expression levels (in log2cpm) from bulk RNA sequencing at day 18 (A), day 50 (B) and day 100 (C). Unexposed samples are visualized in turquoise, CHIR-exposed samples in orange. Blue and red boxes highlight respectively genes down- or up-regulated by CHIR treatment, considering the following thresholds from RNASeq bulk differential expression analysis: *: nominal p-value < 0.05; **: nominal p-value < 0.01.

Figure S3. Breakdown of stage-wise differential expression analysis. Related to Figure 3. A. Top-100 genes according to absolute loading value for the first component from Principal Component Analysis. **B.** Results of gene ontology enrichment analysis performed for Biological Process GO domain on the 100 genes selected as detailed in A; p-values for the top-10 categories are shown. **C**. Gene expression profiles (as z-scores) for the genes identified in bulk RNASeq as modulated by GSK3 treatment (FDR < 0.05, absolute $\log_2$FC > 1) at day 18 (A), day 50 (B) and day 100 (C). **D**. Results of gene ontology enrichment analysis performed on the DEGs identified at day 50, divided for up-regulated and down-regulated genes. Bar plots depict the p-values for the top-10 Cellular Component GO terms.

Figure S4. Overlap of Louvain clusters to cell identities within human fetal data single cell transcriptomes. Related to Figure 4. A. For each sub-panel, cells (represented as dots) are colored according to the expression levels of paired combinations of representative cell type markers (DCX, STMN2 = mature neurons), (ENO2, ST18 = early neural progenitors), (FAM107A, HOPX = outer radial glia), (SMOC1, S100B = intermediate progenitors) (CDK1, CDC20, MKI67 = proliferating progenitors) (CXCL14 = choroid/Astro). **B.** Heatmap of the overlaps between marker genes characterizing the internal organoids clusters and external gene signatures of the relevant single cell clusters from human fetal brains (Nowakowski et al., 2017). Colors are based on the log2Enrichment computed for each pair of comparisons. * indicates an overlap with p-value < 0.05.

Figure S5. UMAP plots for proliferation/maturity distribution and pseudotime decomposition. Related to Figure 6. A. Visualization of expression levels of proliferation/progenitor markers (in red: CDK1, MKI67) coupled with maturity markers (in green: DCX, STMN2), (in yellow: overlapping expression). UMAPS are visualized separately for each timepoint and condition. **B.** First and second components inferred from condition-wise diffusion components and visualized as color gradient on UMAP. The two extremes of the color scale (origin-blue and dark red-outcome) represent the distance in the relative diffusion component. CTL D100 is the only condition with an end point in oRG area.

2. Supplemental experimental procedures

hPSC culture. hPSC lines were cultured under feeder-free conditions on Matrigel (BD Biosciences) coated dishes (diluted 1:40 matrigel:DMEM/F12) and grown in Tesr™ E8™ medium (Stem Cell Technologies). Cells were passaged upon treatment with ReLeSR (Stem Cell Technologies). All differentiation procedures were performed on iPSC lines with at least 15 passages after reprogramming. Pluripotent lines came from different individuals representing either human iPSC or ESC lines previously described (Adamo et al., 2015; González et al., 2014). All the cultures were regularly tested and maintained mycoplasma free.

Table 1. Distribution of hPSC lines used in this study.

Experiment/Line	809-1-5 (iPSC1)	KOLF2CL (iPSC2)	MIFF1-3 (iPSC3)	3391B (iPSC4)	HUES8 (ESC)
ICC		X	X	X	X
Rosette formation assay		X	X	X	

Growth curve			X	X	X
Bulk RNAseq	X	X	X	X	X
Single cell RNAseq	X	X	X	X	X

Neural induction and lumen quantification. hPSC cells were plated at a density of 0.7×10^5 cells cm^2 on Matrigel (BD Biosciences) coated dishes in grow medium supplemented with 10 µM ROCK inhibitor (Y-2763221, Cell Guidance System). Cell cultures were expanded for two days until they were 70% confluent. The starting differentiation medium includes DMEM/F12 (Life Technologies) with N2 and B27 without retinoic acid (Life Technologies), supplemented with 500 nM LDN193189 (Sigma) and 10 µM SB431542 (Tocris). Cells were fixed at day 20 in 4% (wt/vol) paraformaldehyde (PFA) for 15 minutes at room temperature (RT) and washed 3 times with phosphate-buffered saline (PBS). Cells were then permeabilized with PBS containing 0.5% Triton X-100 (Sigma) and blocked with 10% (vol/vol) normal goat serum (NGS; Vector) for 1 hour at RT. Next, cells were incubated overnight at 4°C. The following primary antibodies and dilutions were used: Anti-PALS1, (Proteintech 117710-1-AP). Object identification module of Cell Profiler software (v.2.1.1) was used to automatically quantify lumens (number and size of Pals1 positive areas) of neural rosettes.

Cortical organoid protocol. hPSCs were plated onto cell cycle-arrested mouse embryonic feeders (MEFs)(Millipore) for one passage, colonies grown for at least 48h and then enzymatically detached by incubation with 0.7 mg/ml dispase (Invitrogen: 17105-041) for approx. 30 min. Suspended colonies were subsequently transferred into ultra-low-attachment 100 mm plastic plates (Corning) in FGF2-free knockout serum medium. For the first 24 h (day 0), the medium was supplemented with the ROCK inhibitor Y-27632 (EMD Chemicals). For neural induction, dorsomorphin (Merck, 5 µM) and SB-431542 (Tocris, 10 µM) were added to the medium until day 5. From day 6 onward, organoids were moved to neural medium (NM) containing Neurobasal (Invitrogen 10888), B-27 serum substitute without vitamin A (Invitrogen 12587), GlutaMax 1:100 (Fisher 35050071), 100 U/ml penicillin and streptomycin (Invitrogen) and 50 mM b-Mercaptoethanol (Gibco 31350010). The NM was supplemented with 20 ng/ml FGF2 (Thermo) and 20 ng/ml EGF (Tocris) for 19 days with daily medium change in the first 10 days, and every other day for the subsequent 9 days. On day 12, floating organoids were moved to orbital shaker (VWR Standard Orbital Shaker, Model 1000) and kept on constant shaking at 50 rpm to promote nutrient and oxygen exchange. To induce neurogenesis, FGF2 and EGF were replaced with 20 ng/ml BDNF (Peprotech) and 20 ng/ml NT3 (Peprotech) starting at day 25, while from day 43 onwards only NM without growth factors was used for medium changes every other day.

Primary and secondary antibodies used and quantifications. Primary antibodies were prepared in PBS + 5% normal donkey serum (Jackson Immuno Resarch) overnight at 4°C. The following primary antibodies and dilutions were used: anti-PAX6, 1:200 (Biolegend PRB-278B); anti-KI67, 1:200 (Abcam ab15580); anti-NESTIN, 1:500 (Millipore MAB5326); anti–DCX, 1:1000 (BD Biosciences 611706); anti-TBR1, 1:200 (Abcam ab31940), anti-HOPX, 1:50 (Sigma-Aldrich HPA030180), Anti CTIP2 1:200 (Abcam ab18465). After primary incubation, sections were washed three times with PBS and the incubated with appropriate secondary antibodies: anti-rabbit Alexa 488-conjugated (Bethyl A090-516D2) anti-mouse, Alexa 594-conjugated (Bethyl A110-305B4) (Molecular Probes, Invitrogen) diluted 1:500 in blocking solution and incubated for 2 h at RT. Before mounting, sections were incubated with Hoechst 33258 (5 µg/mL; Molecular Probes, Invitrogen) or DAPI 1:5000 (Merck), as indicated on each caption. Quantification of nuclear markers was done by using the automatic cluster counter ITCN plugin from FIJI (v.1.49 NIH-USA) for at least 3 organoids from 3 independent lines. Images were RGB converted and cluster counts were done over pre-defined concentric grids from VLS. The relative number of positive cells was calculated as a percentage of total DAPI+ cells. Images were acquired with a Leica DMI 6000B microscope (10x, 20x and 40x objectives) and analyzed with LAS-AF imaging software and then processed using Image J (v1.49 NIH, USA) to adjust contrast for optimal RGB rendering. Semi-quantitative measurements were made in imageJ, the background noise was removed with a sliding paraboloid filter of dimension 500 px, then a gaussian filter was applied. A positive cell was considered the local maxima with a minimum threshold of signal to noise. To evaluate the density of cells we consider organoid slice stained with DAPI, from the images of the whole organoid we remove, with a sliding paraboloid filter, the background and then apply a common threshold on all the conditions. The area of the resulting mask is considered as area occupied by nuclei and used as normalization.

Bulk transcriptome analysis. *Differential gene expression.* Gene expression quantification at the gene level was performed by Salmon (version 0.8.2) (Patro et al., 2017), using hg38 RefSeq annotation. To estimate differential expression, the matrix of gene counts was analyzed by edgeR (version 3.20.9) (Robinson et al., 2009). For each time point, genes with an expression level of at least 2 cpm (count per million) in at least 3 samples were selected for the analysis. Small genes, ribosomal genes and fusion genes were excluded. After TMM normalization, differential expression analysis comparing treated to untreated samples was performed using a likelihood ratio test on the coefficients of a negative binomial model. Significantly modulated genes were selected setting an absolute value of log2 fold change (Log2FC) higher than 1 and a false discovery rate (FDR) lower than 5%. Log2 cpm values, were

used for heatmap representation of gene expression profiles (visualized as z-scores). Heatmaps were produced with pheatmap R package (version 1.0.10, Raivo Kolde (2018). pheatmap: Pretty Heatmaps.). Analyses were performed in R version 3.4.4. Functional annotation of biological functions was performed by Gene ontology analysis and Gene set enrichment analysis (GSEA) using as set source H1 collection from the Molecular Signature Database (Liberzon et al., 2015). *Gene Ontology Enrichment Analysis.* Gene ontology enrichment analysis for the Cellular Component domain of the ontology was performed on the 898 DEGs identified at day 50, split in up-regulated and down-regulated genes. The pool of tested genes (as selected for differential expression analysis) was used as background. The analysis was performed by topGO (version 2.30.1) (Adrian Alexa and Jorg Rahnenfuhrer 2016). topGO: Enrichment Analysis for Gene Ontology), relying on Fisher test and Weight01 method to take into account ontology hierarchy; minimum node size was set at 15. After imposing an enrichment cut-off of 2, a 0.01 p-value cut off was applied to select significantly enriched GO terms. Barplot in supplementary figure 4 shows the top-10 categories ranked for p-value.

Gene set enrichment analysis. GSEA was applied to each developmental stage with GSAA software, version 1.2 (Xiong et al., 2014). Raw reads for the same genes tested for differential expression were analyzed by GSAASeqSP (permutation type 'gene set').

Single cell library preparation and sequencing. Briefly, a small volume (6 - 8 µl) of single-cell suspension at a density of 1000 cells/µl was mixed with RT-PCR master mix and immediately loaded together with Single-Cell 3′ gel beads and partitioning oil into a single-cell 3′ Chip. The gel beads were coated with unique primers bearing 10× cell barcodes, unique molecular identifiers (UMI) and poly(dT) sequences. The chip was then loaded onto a Chromium instrument (10× Genomics) for single-cell GEM generation and barcoding. RNA transcripts from single cells were reverse-transcribed within droplets to generate barcoded full-length cDNA using Clontech SMART technology. After emulsion disruption, cDNA molecules from one sample were pooled and preamplified. Finally, amplified cDNAs were fragmented, and adapter and sample indices were incorporated into finished libraries which were compatible with Illumina sequencing. The final libraries were quantified by Qubit system (Thermo) and calibrated with an in-house control sequencing library. The size profiles of the pre-amplified cDNA and sequencing libraries were examined by Agilent Bioanalyzer 2100 using a High Sensitivity DNA chip (Agilent). Two indexed libraries were equimolarly pooled and sequenced on Illumina NOVAseq 6000 platform using the v2 Kit (Illumina, San Diego, CA) with a customized paired-end, dual indexing format according to the recommendation by 10× Genomics. Using proper cluster

density, a coverage around 250 M reads per sample (2000–5000 cells) were obtained corresponding to at least 50,000 reads/cell.

Single cell transcriptome analysis. Before downstream analyses, data deriving from the 11 samples was integrated by Seurat v3.0-alpha analytical framework (Stuart et al., 2018). After normalization, anchors for data integration were identified considering 3000 anchor points (genes) and 40 dimensions. For data reduction, UMAP was applied with 50 nearest neighbors (nn); cluster initial positions were set considering PAGA node position (Scanpy v1.3.1) (Wolf et al., 2018). On the integrated dataset, clusters were identified by applying Louvain with Multilevel Refinement from Seurat with resolution parameter at 0.7. This resulted in the identification of 15 clusters. For cluster annotation, we applied the FindMarker Seurat function, using MAST as test and filtering for up-regulated genes with adjusted P value < 0.05. The obtained lists were compared in an overlap analysis with gene lists derived from two published single cell datasets: from the WGCNA analysis of single cell clusters of human fetal brains (paper by Nowakowski et al, DOI: 10.1126/science.aap8809), data were downloaded from the web portal at https://cells.ucsc.edu/?ds=cortex-dev and upregulated genes (vg_diff.float>1.5) were selected for each relevant cluster, from the differential expression analysis of single nuclei clusters (Table S6 of the paper by Amiri et al, DOI: 10.1126/science.aat6720), upregulated genes (fold change > 0 and adjusted P value < 0.05) were selected for each relevant cluster. P values and enrichment coefficients were computed relative to a universe of 3000 genes used for all the other single cell analysis. Cluster-specific expression levels of biologically-relevant genes identified among the top dysregulated were visualized by violin plots, stratified for stage and treatment. To test for unbalance in the number of highly expressing cells for representative genes per cluster or cluster set, a threshold of expression at the 90th percentile was fixed, and the number of cells above or below the threshold in treated and untreated cells was compared by Fisher test. The identification of the threshold and sub-sequent analysis was performed on the sub-sampled dataset, separately for Day 50 and Day 100.
Cell cycle analysis were performed using Scanpy function score_genes_cell_cycle, relying on the genes from (Kowalczyk et al., 2015). Diffusion map algorithm for dimensionality reduction was performed with Scanpy with 50 nn. Pseudotime analysis for lineage branching reconstruction was applied using wishbone algorithm (Setty et al., 2016). The analysis was performed on the complete dataset, as well as separately for each of the four biological conditions in order to infer stage or treatment-selective trajectories; the origin was identified with the same method applied on complete dataset. Trajectories were reproduced defining using at least 3 different markers. Partition-based

graph abstraction (PAGA) algorithm was applied on the complete dataset, as well as separately for each of the four biological conditions and plotted with layout Reingold Tilford. The position of the nodes identified on the complete dataset was exploited in the graph for each biological condition.

3. References related to experimental procedures

Adamo, A., Atashpaz, S., Germain, P.-L., Zanella, M., D'Agostino, G., Albertin, V., Chenoweth, J., Micale, L., Fusco, C., Unger, C., et al. (2015). 7q11.23 dosage-dependent dysregulation in human pluripotent stem cells affects transcriptional programs in disease-relevant lineages TL - 47. Nat. Genet. *47 VN-r*, 132–141.

González, F., Zhu, Z., Shi, Z.D., Lelli, K., Verma, N., Li, Q. V., and Huangfu, D. (2014). An iCRISPR platform for rapid, multiplexable, and inducible genome editing in human pluripotent stem cells. Cell Stem Cell.

Kowalczyk, M.S., Tirosh, I., Heckl, D., Rao, T.N., Dixit, A., Haas, B.J., Schneider, R.K., Wagers, A.J., Ebert, B.L., and Regev, A. (2015). Single-cell RNA-seq reveals changes in cell cycle and differentiation programs upon aging of hematopoietic stem cells. Genome Res.

Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J.P., and Tamayo, P. (2015). The Molecular Signatures Database Hallmark Gene Set Collection. Cell Syst.

Nowakowski, T.J., Bhaduri, A., Pollen, A.A., Alvarado, B., Mostajo-Radji, M.A., Di Lullo, E., Haeussler, M., Sandoval-Espinosa, C., Liu, S.J., Velmeshev, D., et al. (2017). Spatiotemporal gene expression trajectories reveal developmental hierarchies of the human cortex. Science (80-.).

Patro, R., Duggal, G., Love, M.I., Irizarry, R.A., and Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. Nat. Methods.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2009). edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics.

Setty, M., Tadmor, M.D., Reich-Zeliger, S., Angel, O., Salame, T.M., Kathail, P., Choi, K., Bendall, S., Friedman, N., and Pe'Er, D. (2016). Wishbone identifies bifurcating developmental trajectories from single-cell data. Nat. Biotechnol.

Wolf, F.A., Angerer, P., and Theis, F.J. (2018). SCANPY: Large-scale single-cell gene expression data analysis. Genome Biol.

4. Supplemental tables.

Table 2. List of differentially expressed genes.

Table 3. List of GSEA categories (Excel files).

Figure S1.

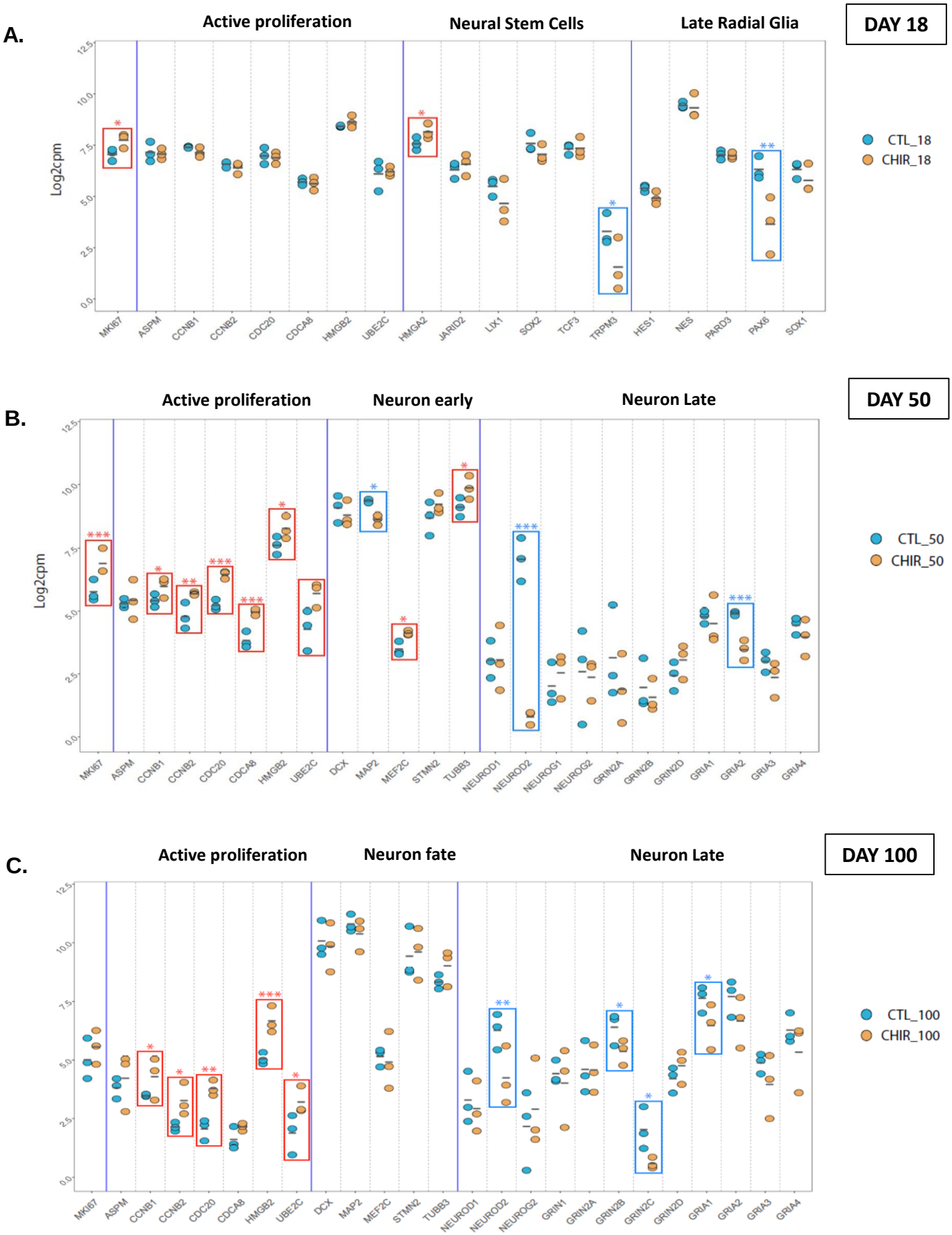

A.
Active proliferation
Neural Stem Cells
Late Radial Glia
DAY 18
Log2cpm
MKI67
ASPM
CCNB1
CCNB2
CDC20
CDCA8
HMGB2
UBE2C
HMGA2
JARID2
LIX1
SOX2
TCF3
TRPM3
HES1
NES
PARD3
PAX6
SOX1
CTL_18
CHIR_18
B.
Active proliferation
Neuron early
Neuron Late
DAY 50
Log2cpm
DCX
MAP2
MEF2C
STMN2
TUBB3
NEUROD1
NEUROD2
NEUROG1
NEUROG2
GRIN2A
GRIN2B
GRIN2D
GRIA1
GRIA2
GRIA3
GRIA4
CTL_50
CHIR_50
C.
Active proliferation
Neuron fate
Neuron Late
DAY 100
GRIN1
GRIN2C
CTL_100
CHIR_100

Figure S2.

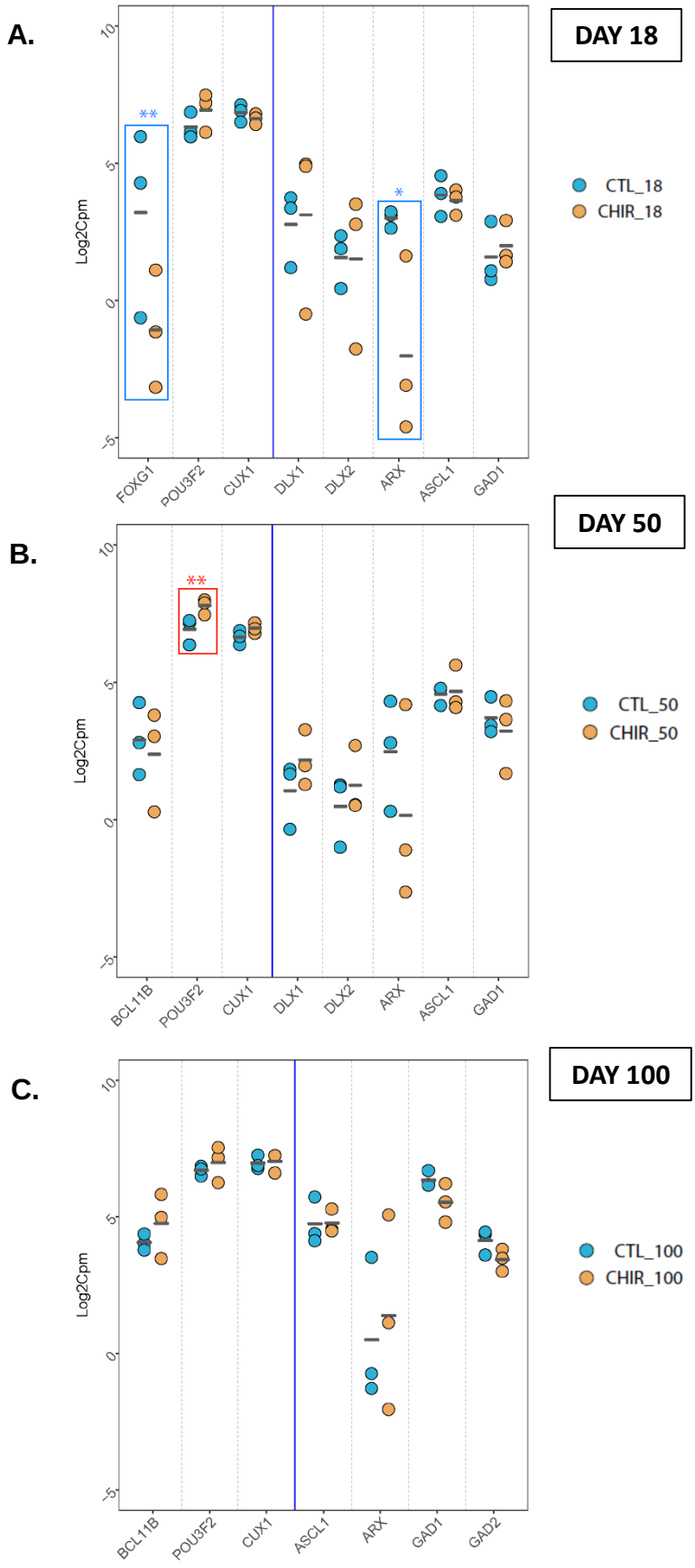

A.
DAY 18
Log2Cpm
10
5
0
-5
**
*
FOXG1
POU3F2
CUX1
DLX1
DLX2
ARX
ASCL1
GAD1
CTL_18
CHIR_18
B.
DAY 50
Log2Cpm
**
BCL11B
POU3F2
CUX1
DLX1
DLX2
ARX
ASCL1
GAD1
CTL_50
CHIR_50
C.
DAY 100
Log2Cpm
BCL11B
POU3F2
CUX1
ASCL1
ARX
GAD1
GAD2
CTL_100
CHIR_100

Figure S3.

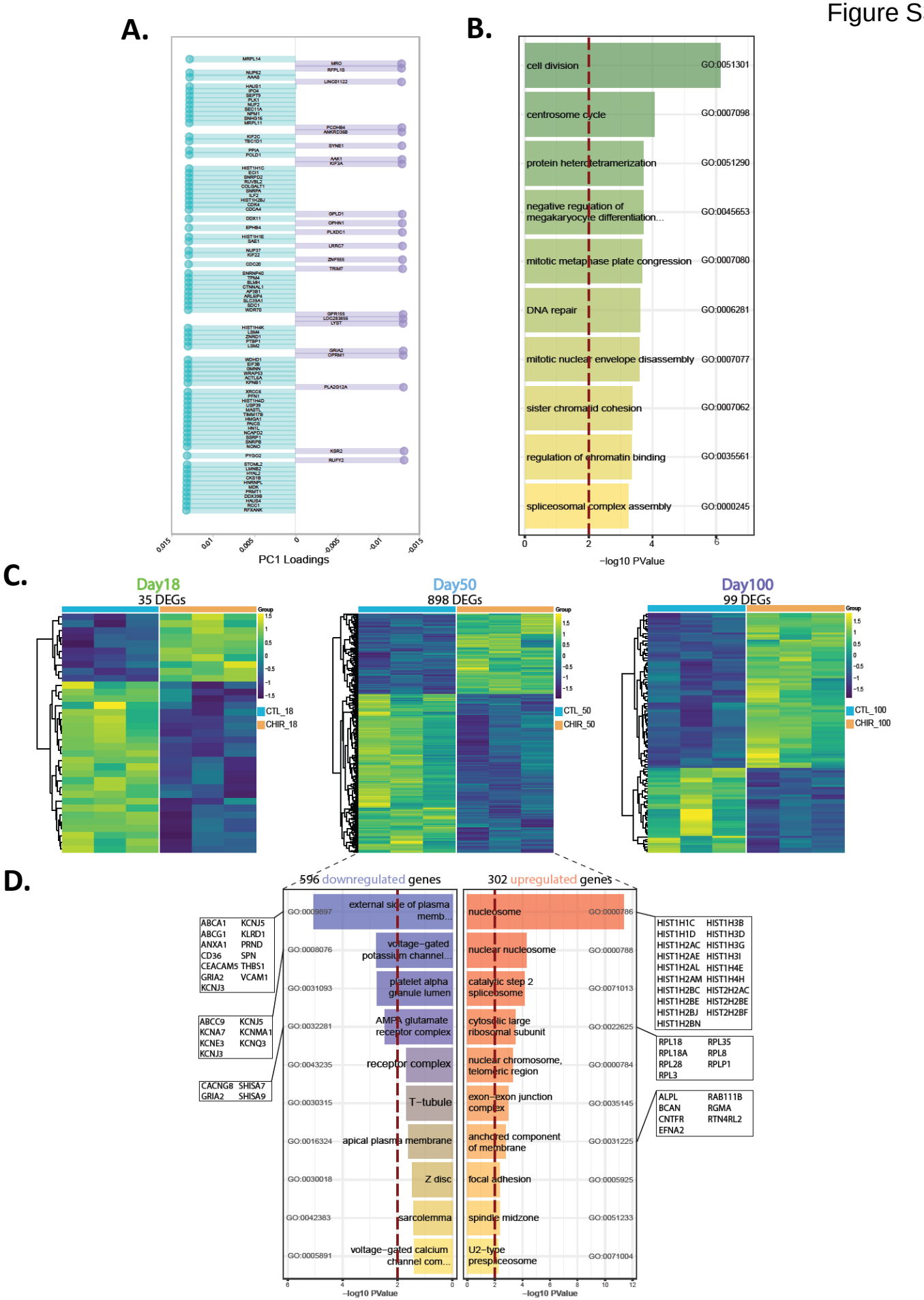
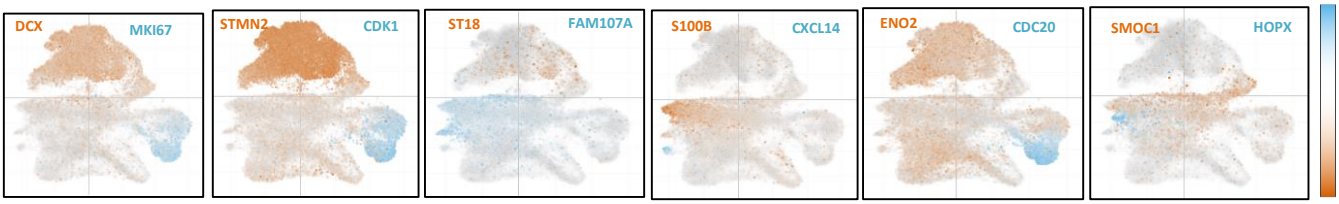

A.
PC1 Loadings
B.
cell division GO:0051301
centrosome cycle GO:0007098
protein heterotetramerization GO:0051290
negative regulation of megakaryocyte differentiation... GO:0045653
mitotic metaphase plate congression GO:0007080
DNA repair GO:0006281
mitotic nuclear envelope disassembly GO:0007077
sister chromatid cohesion GO:0007062
regulation of chromatin binding GO:0035561
spliceosomal complex assembly GO:0000245
−log10 PValue
C.
Day18
35 DEGs
CTL_18
CHIR_18
Day50
898 DEGs
CTL_50
CHIR_50
Day100
99 DEGs
CTL_100
CHIR_100
D.
596 downregulated genes
302 upregulated genes
GO:0009897 external side of plasma memb...
GO:0008076 voltage−gated potassium channel...
GO:0031093 platelet alpha granule lumen
GO:0032281 AMPA glutamate receptor complex
GO:0043235 receptor complex
GO:0030315 T−tubule
GO:0016324 apical plasma membrane
GO:0030018 Z disc
GO:0042383 sarcolemma
GO:0005891 voltage−gated calcium channel com...
ABCA1 KCNJ5
ABCG1 KLRD1
ANXA1 PRND
CD36 SPN
CEACAM5 THBS1
GRIA2 VCAM1
KCNJ3
ABCC9 KCNJ5
KCNA7 KCNMA1
KCNE3 KCNQ3
KCNJ3
CACNG8 SHISA7
GRIA2 SHISA9
−log10 PValue
GO:0000786 nucleosome
GO:0000788 nuclear nucleosome
GO:0071013 catalytic step 2 spliceosome
GO:0022625 cytosolic large ribosomal subunit
GO:0000784 nuclear chromosome, telomeric region
GO:0035145 exon−exon junction complex
GO:0031225 anchored component of membrane
GO:0005925 focal adhesion
GO:0051233 spindle midzone
GO:0071004 U2−type prespliceosome
HIST1H1C HIST1H3B
HIST1H1D HIST1H3D
HIST1H2AC HIST1H3G
HIST1H2AE HIST1H3I
HIST1H2AL HIST1H4E
HIST1H2AM HIST1H4H
HIST1H2BC HIST2H2AC
HIST1H2BE HIST2H2BE
HIST1H2BJ HIST2H2BF
HIST1H2BN
RPL18 RPL35
RPL18A RPL8
RPL28 RPLP1
RPL3
ALPL RAB11B
BCAN RGMA
CNTFR RTN4RL2
EFNA2
−log10 PValue

Figure S4.

A


DCX
MKI67
STMN2
CDK1
ST18
FAM107A
S100B
CXCL14
ENO2
CDC20
SMOC1
HOPX

B

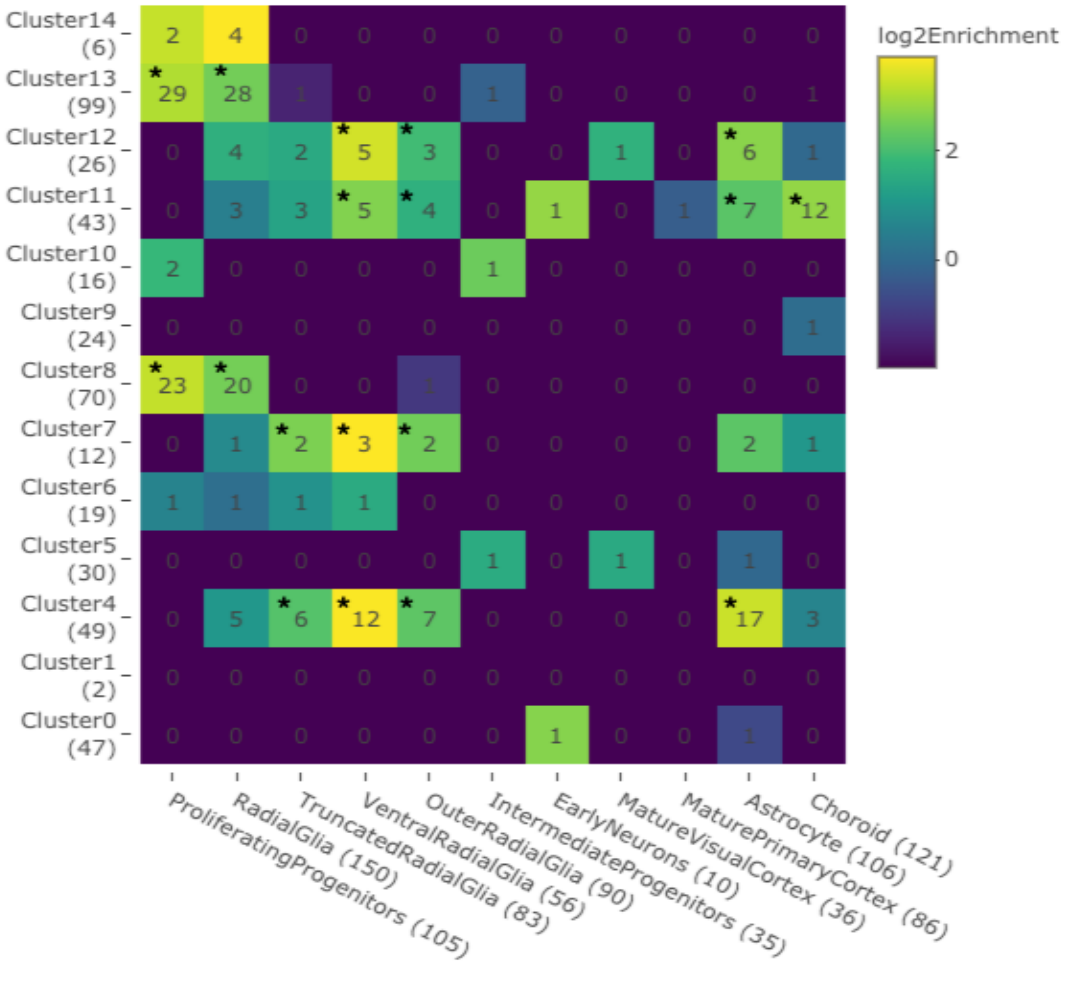
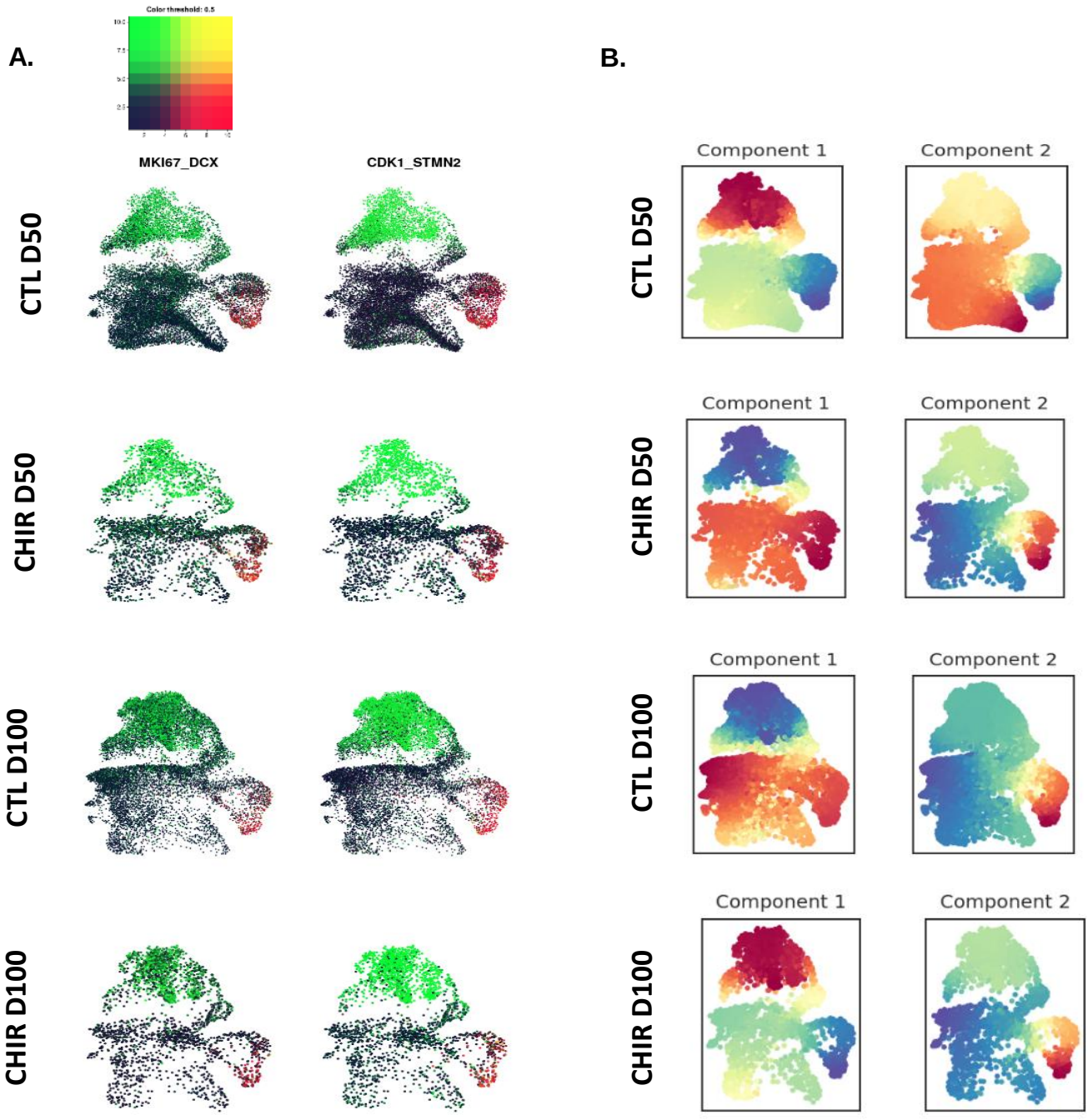

log2Enrichment
2
0
Cluster14 (6)
Cluster13 (99)
Cluster12 (26)
Cluster11 (43)
Cluster10 (16)
Cluster9 (24)
Cluster8 (70)
Cluster7 (12)
Cluster6 (19)
Cluster5 (30)
Cluster4 (49)
Cluster1 (2)
Cluster0 (47)
ProliferatingProgenitors (105)
RadialGlia (150)
TruncatedRadialGlia (83)
VentralRadialGlia (56)
OuterRadialGlia (90)
IntermediateProgenitors (35)
EarlyNeurons (10)
MatureVisualCortex (36)
MaturePrimaryCortex (86)
Astrocyte (106)
Choroid (121)

Figure S5.


A.
Color threshold: 0.5
MKI67_DCX
CDK1_STMN2
CTL D50
CHIR D50
CTL D100
CHIR D100
B.
Component 1
Component 2
CTL D50
CHIR D50
CTL D100
CHIR D100