Lalli *et al.* Supplemental Material

High-throughput single-cell functional elucidation of neurodevelopmental disease-associated genes reveals convergent mechanisms altering neuronal differentiation

Matthew A. Lalli[1,2], Denis Avey[1,2], Joseph D. Dougherty[1,3], Jeffrey Milbrandt[1,*], Robi D. Mitra[1,2,*]

1. Department of Genetics, Washington University in St. Louis School of Medicine, St. Louis MO 63110, USA
2. Edison Family Center for Genome Sciences and Systems Biology, Washington University in St. Louis School of Medicine, St. Louis MO 63110, USA
3. Department of Psychiatry, Washington University in St. Louis School of Medicine, St. Louis MO 63110, USA

*Corresponding Authors

Table of Contents

Supplemental Methods

LUHMES cell culture

Proliferating LUHMES cells were maintained in DMEM/F12 media supplemented with 1% N2 and 40 ng/mL basic fibroblast growth factor (bFGF) and 1% penicillin-streptomycin. Cells were grown in T25 flasks coated with poly-ornithine and fibronectin. For differentiation, bFGF was withdrawn from the media and tetracycline was added (1 µg /mL) to repress the v-Myc transgene. For time-course RNA-sequencing experiments, neurotrophic factors (cAMP and GDNF) were added to the differentiation media. These factors were withheld from repression experiments to increase the sensitivity to detect perturbations (Scholz et al. 2011). dCas9-KRAB LUHMES were maintained in blasticidin-containing media (10 µg/mL) to prevent transgene silencing.

Human iPSC-derived neural progenitor cell (XCL4) RNA-seq Experiments

Tet-On dCas9-KRAB XCL4 were plated at a density of 200,000 cells per well in a 12-well plate on Matrigel and infected in triplicate with individual guide RNAs targeting ASD genes. Next day, media containing doxycycline (2µg/mL) and puromycin (1 µg/mL) was added to induce dCas9-KRAB and select for guide-containing cells. Cells were passaged 4 days after selection. mRNA was collected 4 days after replating (8 days of repression).

Bulk RNA Barcoding and sequencing (BRB-seq) was performed with minor modifications (Alpern et al. 2019). We performed first-strand reverse transcription (RT) reactions in 10 µL total volume using 250 ng of RNA from each sample, 1 µL of 10 µM barcoded oligo-dT-VN primers modified to mimic the 10x Genomics v2 chemistry. RNA was mixed with barcoded oligo-dT primers, water, and dNTPs and hybridized by incubation at 65 ˚C for 5 minutes and immediately transferred onto ice. 1 uL of template switch oligo (TSO), 0.5 µL of Maxima H Minus Reverse Transcriptase, and 5X RT buffer were added and

samples are incubated at 45 °C for 90 min for reverse transcription. After RT, six samples were pooled together and purified using a nucleospin column (Machery-Nagel) and eluted with 23 µL of elution buffer.

We designed barcoded oligoDT-VN oligos to mimic 10x Genomics v2 chemistry (partial seq1, 16 bp cell barcode extracted randomly from the 10x Genomics safelist, and a 10 bp UMI (5N + 5V). Barcoded primer sequences are listed in Supplemental Table S6.

Barcoded, pooled first-strand reactions (23 µL) were mixed with 1 µL partial seq1 primer (10 µM), 1 µL SMART primer (10 µM), and 25 µL 2X KAPA HiFi HotStart ReadyMix (Roche). 10 cycles of PCR with a long extension time (98˚ 20 seconds, 60˚ 30 seconds, 72˚ 6 minutes) were performed. cDNA was purified with 0.6X AMPure XP (Beckman Coulter) magnetic beads. DNA was eluted with 20 µL water and concentration was measured using the Tapestation D5000 ScreenTape (Agilent). 600 pg of product were tagmented with barcoded N7 primers and P5-index-seq1 primers using the Nextera XT kit (Illumina).

BRB-seq libraries were sequenced on a HiSeq2500 paired-end with 28 x 91 reads.

gRNA cloning

gRNAs direct dead Cas9 (dCas9) to a window 25-75 nucleotides downstream of the gene's transcription start site. gRNAs were screened for sequence features predicting high activity and no off-target effects (Sanson et al. 2018). Lentiviral gRNA expression vectors were created by annealing two complementary oligonucleotides encoding gRNAs at 100 µM (IDT DNA) with sticky-ends and ligating annealed products into BsmB1 digested CROP-seq-opti vector using Golden Gate assembly.

Each Golden Gate assembly reaction contained 6.5 µL water, 1 µL 1:10 diluted annealed oligos, 1 µl T4 ligase, 1 µl T4 ligase buffer, and 0.5 µL BsmB1. Reactions were incubated at 16 ˚C for 10 minutes (ligation) and 55 ˚C for 10 minutes (restriction) for 4 cycles. 1 µL of Golden Gate mixture was transformed into 30 µL Stellar Competent cells and plated onto ampicillin-containing agar plates. Individual colonies were miniprepped after colony PCR and all constructs were verified by Sanger sequencing (Genewiz).

gRNA Detection

gRNA-enrichment PCR primers are provided in Supplemental Table S7. Post cDNA amplification products are used for PCR (1 µL per 25 µL PCR reaction). Three separate gRNA-enrichment PCRs are performed per 10x Genomics Single Cell Chip lane and each is sequenced to around a million reads on a MiSeq. Look-up tables of gRNA and cell barcodes are generated using custom Python scripts with a detection threshold of 20 UMIs per gRNA-cell barcode combination. Look-up tables for each experiment are provided in Supplemental Code. Cell barcodes from filtered high-quality cells were matched against this table and only cells with a single gRNA were retained for analysis.

Bioinformatic Analyses

All sequencing reads were mapped to hg38 using STAR (v2.5.1b). The output filtered gene expression matrices were imported into R (v 3.5.1) for further analysis.

- *Single-cell quality control*

Quality control was performed in Seurat by computing the number of transcripts per cell and the percentage of mitochondrial gene expression. Cells with more than 500 but fewer than 7500 detected genes, and less than 8% mitochondrial gene expression were retained.

- *gRNA Effect Size Estimation*

We estimated gRNA knock-down efficiency using the mimosca.run_model() command in the MIMOSCA toolkit (Dixit et al. 2016). We ran this analysis on both replicates pooled or separately.

- *Pseudotime Enrichment and Depletion Analysis*

Cells were grouped by each target gene and the proportion of cells in each pseudotime state was calculated. We computed the expected and observed number of cells in each pseudotime state compared to either all cells or cells with non-targeting gRNAs. Because most gRNAs did not alter pseudotime state membership, both approaches provided similar results. We performed χ-squared tests on these counts and used a lenient threshold of $p < 0.05$ to determine whether any target gene knock-down induced an alteration in pseudotime. For these genes, we then looked for a proportion change of > |0.2| to indicate enrichment or depletion. For overall pseudotime, we used the average pseudotime of all cells with a given gene knock-down and compared means with a *t*-test.

- *BRB-seq Analysis*

Because we performed 3' RNA capture with oligos mimicking 10x Genomics chemistry, reads are processed using CellRanger (v 2.1.0). Gene counts for each sample are output by CellRanger which demultiplexes all BRB-seq samples and counts number of unique molecular identifiers (UMIs) for each transcript. Gene counts are then input into edgeR for standard bulk RNA-seq analysis (Robinson et al. 2010).

Supplemental Figures

Figure S1

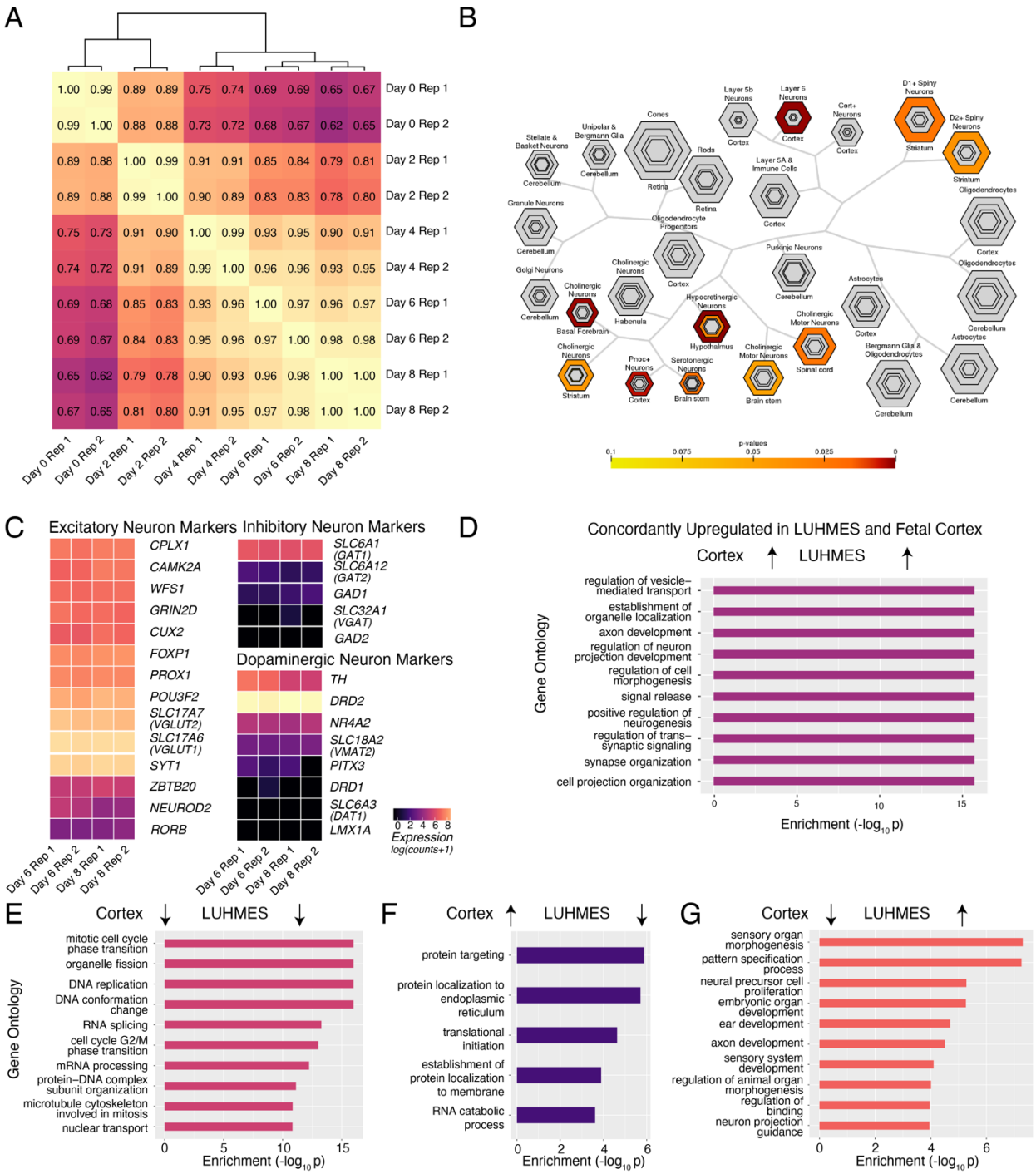

Figure S1: Time-course RNA-seq highlights the utility of LUHMES as a neurodevelopmental model system. A) Pearson correlation of RNA-sequencing counts between all pairs of RNA-seq samples shows high reproducibility of differentiation.

Samples cluster in progressive temporal order by day of differentiation. Hierarchical clustering indicates Day 4-8 samples are more related to each other than to early neuronal progenitor cells (Day 0 and Day 2 samples). n = 2 replicates for each timepoint. **B)** Cell-type Specific Enrichment Analysis (CSEA) of Day 8 differentiated LUHMES shows transcriptional similarities to a variety of neurological disease relevant neuronal cell types. **C)** Differentiated LUHMES (day 6 and day 8) express excitatory neuronal markers, dopaminergic markers, but not inhibitory markers. **D-G)** Related to Fig 1C. Quadrant specific Gene Ontology analysis highlights concordant and discordant pathways across developing fetal cortex and LUHMES differentiation.

Figure S2

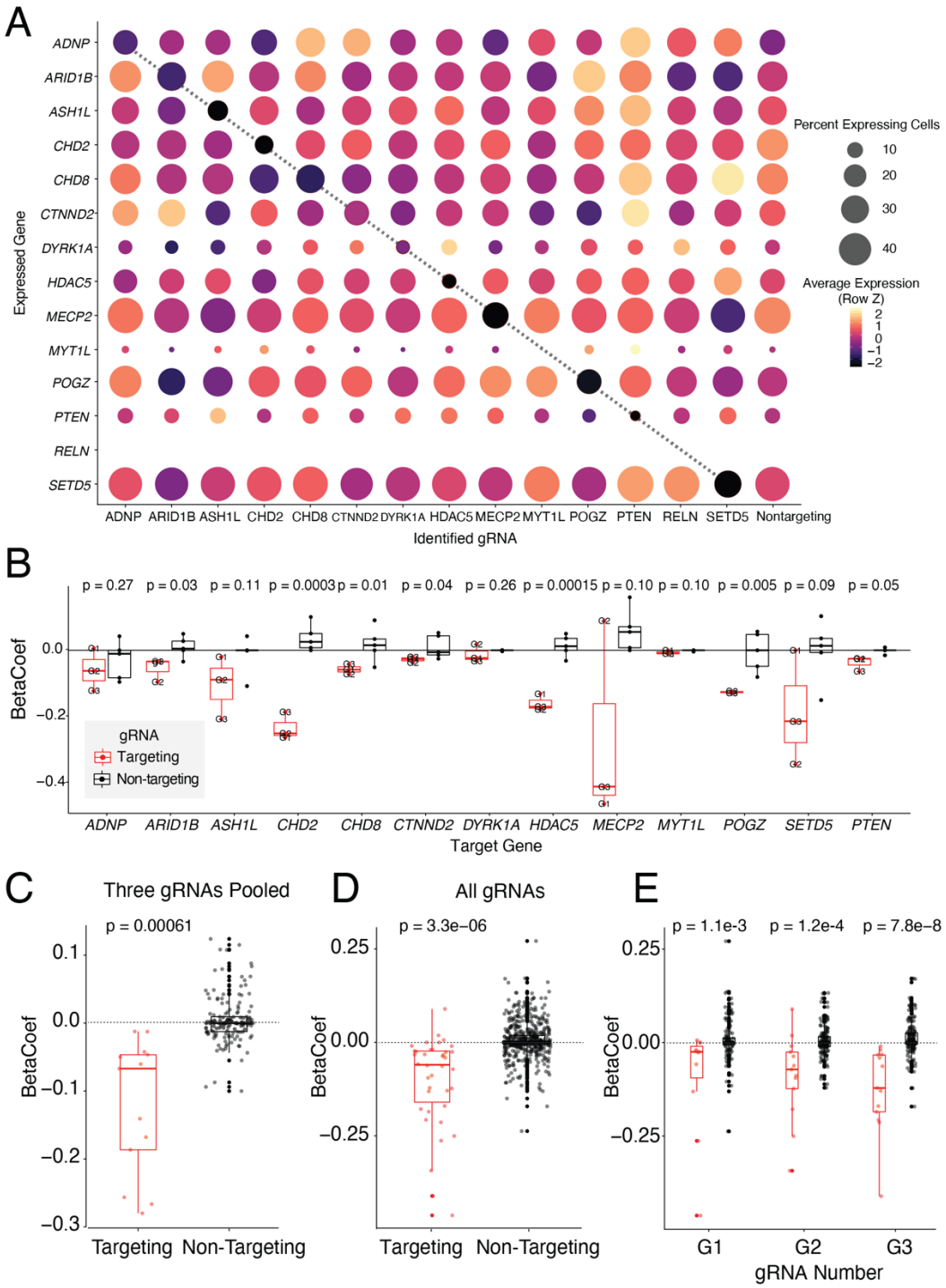

Figure S2: Estimating gRNA knock-down efficiency across targeted genes. A) Seurat DotPlot representation of the expression of ASD candidate genes (rows) across cells grouped by targeted gene (columns). Expression values are Z-normalized within a row across cells grouped by targeted gene. Cooler colors indicating lower expression. The

dark diagonal (dotted line) reflects efficient repression in cells with a guide targeting a given gene. *RELN* was not appreciably detected in scRNA-seq data. **B)** MIMOSCA beta coefficients for 3 on-target and 5 non-targeting guides are plotted for each target gene. Beta < 0, marked with dotted line, represents repression. For all genes, most guides have some degree of repression and this was significant ($p < 0.05$, one-tailed *t*-test) for 7 genes by this analysis. **C)** Collectively, targeting gRNAs have significant, on-target activity (negative beta coefficients) when all 3 gRNAs are merged or **D)** when gRNAs are analyzed separately. **E)** All 3 designed gRNAs for each gene have significant on-target repression in aggregate and Beta coefficients were not significantly different by guide number.

Figure S3

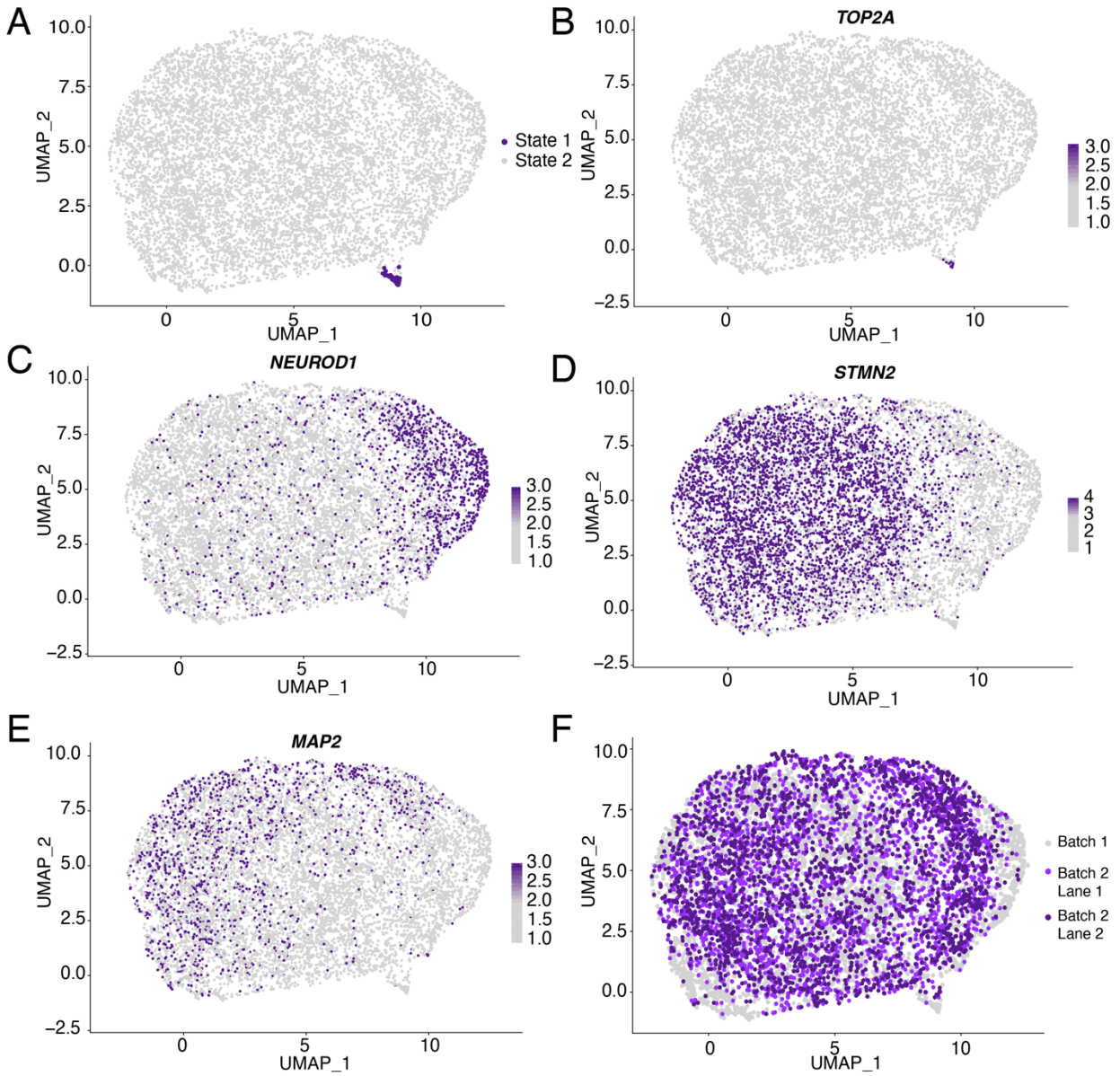

Figure S3: Global structure in transcriptional data. A) UMAP clustering of transcriptomes reveals minor global variation and two cell states. B) Cell State 1 (45 cells) is enriched for expression of the DNA-replication associated gene *TOP2A*, a marker of proliferation; hence virtually all cells are post-mitotic. **C-E)** Neuronal marker genes show variable patterns of gene expression across UMAP space. **F)** No batch effects were apparent on global clustering.

Figure S4

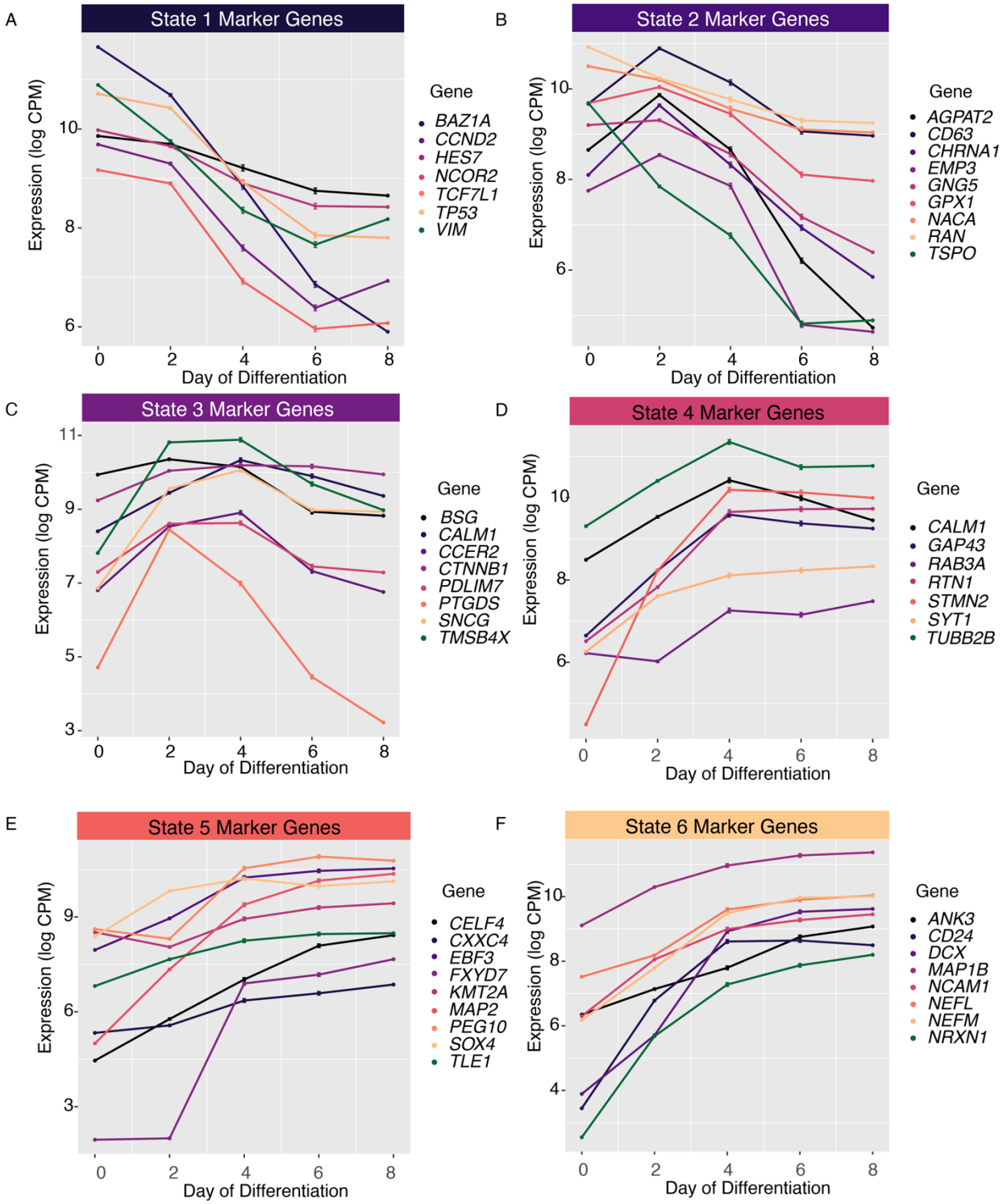

Figure S4: Pseudotime state marker gene expression patterns in neuron differentiation time-course dataset show pseudotime states correspond to day of differentiation. (Related to Figure 3). Expression of positive marker genes from each of

the six pseudotime states in bulk RNA-seq differentiation dataset. **A)** State 1 marker genes tend to have high expression in differentiation day 0-2 cells, and include cell cycle genes (*CCND2),* proliferation genes (*TP53),* and the neural stem cell marker vimentin *(VIM)***. B)** State 2 marker genes have highest expression in day 2-4 cells**. C)** State 3 marker also peak in differentiation days 2-4. **D-F)** States4-6 express marker genes that have high expression in differentiation days 6-8. **F**) State 6 Marker genes include the mature neuronal markers *DCX, MAP1B, NCAM1,* and *NEFL.* Together, these data support the interpretation of pseudotime as an axis of progressive differentiation.

Figure S5

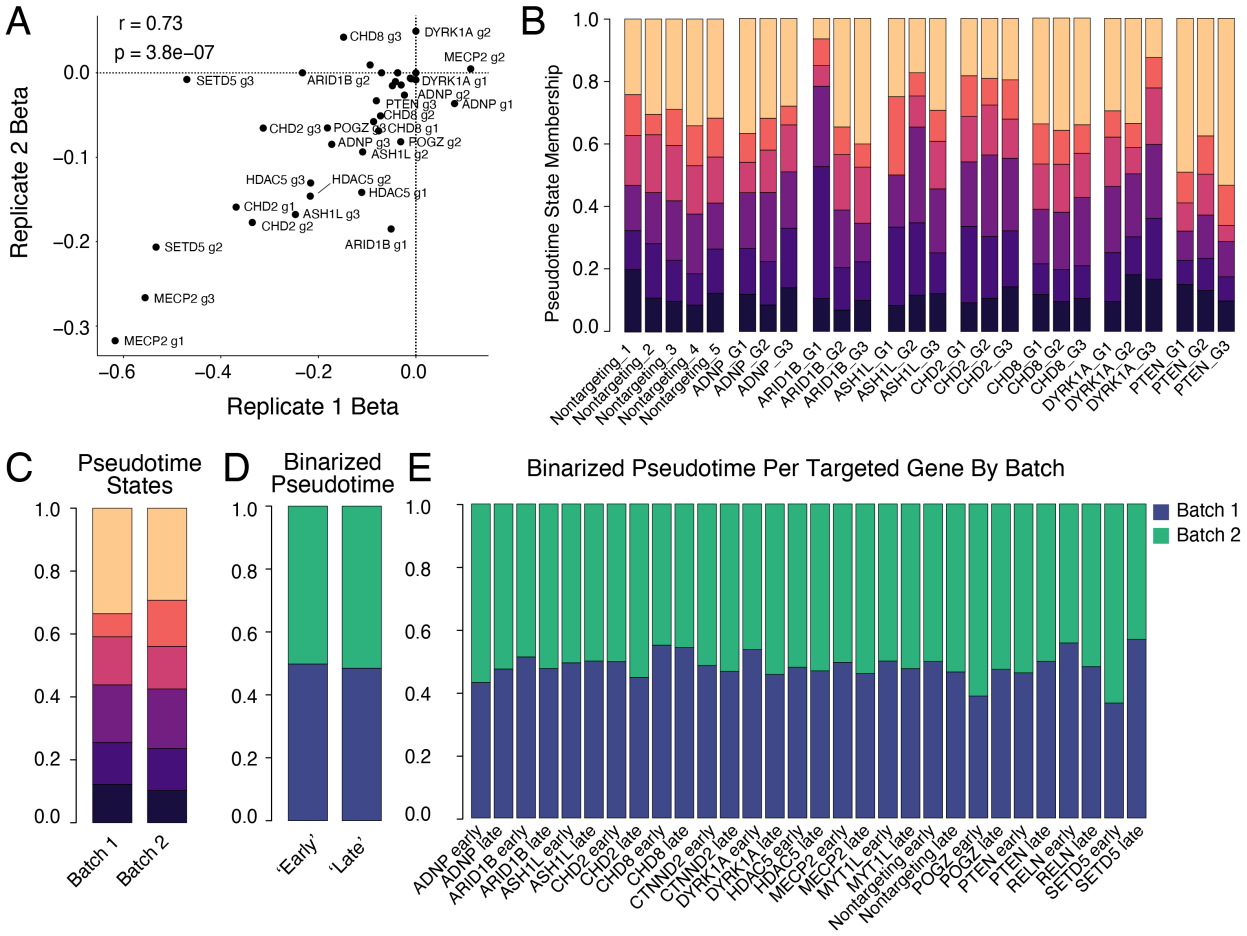

Figure S5: Variation across guides and batches are low. A) MIMOSCA Beta estimates for individual gRNA knock-down efficiencies are highly reproducible across replicate experiments. **B)** Pseudotime state memberships are highly concordant across most guides targeting the same gene. **C)** Each replicate had roughly the same proportion of cells in each pseudotime state. **D)** Binarized pseudotime ('Early' vs 'Late') proportions were equivalent across replicates. **E)** Binarized pseudotime proportions were equivalent across almost all target genes and stages *e.g.* altered pseudotime status was reproducible and not driven by batch effects across replicate experiments.

Figure S6

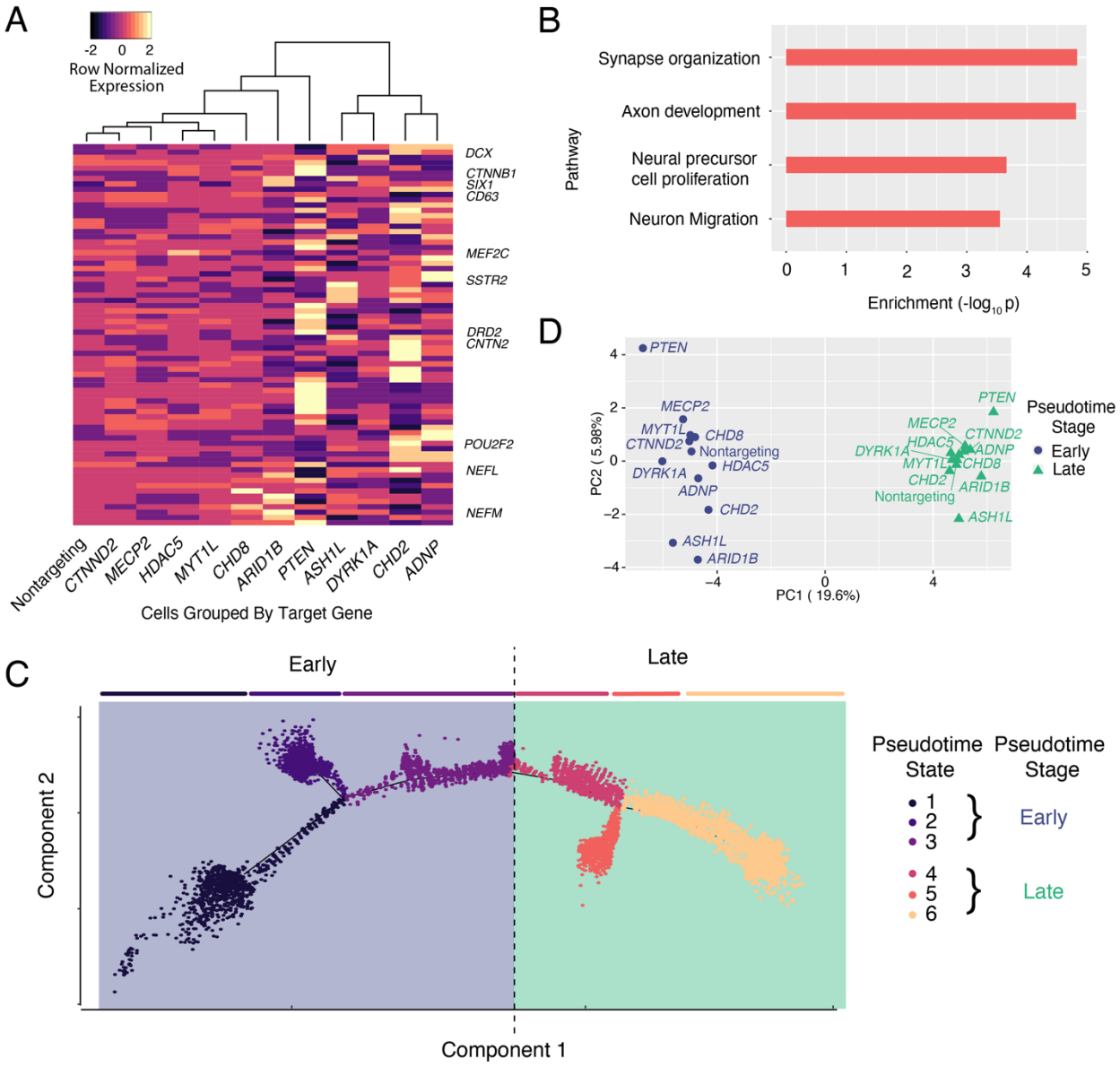

Figure S6: Single-cell differential expression analysis reveals variation in neuron differentiation genes. A) Hierarchical clustering of single-cell expression of recurrently dysregulated genes. Each column represents the average expression profile of all cells having any of the 3 guides targeting the listed gene. Each row is a gene that was found to be differentially expressed in at least 3 knock-down conditions. These genes include the neuronal maturation markers *DCX, NEFL,* and *NEFM.* The 'Delayed Differentiation' module uncovered by pseudotime analysis clusters together and shows a decrease of differentiation markers. *PTEN* shows the opposite expression pattern. **B)** Biological process enrichment of recurrently dysregulated genes highlights variation in neuronal maturation processes, and neural precursor cell proliferation. **C)** Diagram of pseudotime

stratification. Pseudotime states 1-3 were assigned as an ‘Early’ pseudotime stage. Pseudotime states 4-6 were assigned as ‘Late’ stage. Expression profiles for cells in each knockdown condition were re-computed for ‘Early’ and ‘Late’ stages. **D)** Principal component analysis of transcriptional profiles after stratifying by pseudotime status shows that PC1 corresponds to pseudotime and explains 19.6% of the variance in the data. PC2 may correspond to early-stage effects of specific ASD-gene repression, and clusters the ‘Delayed Differentiation’ module genes together, with *PTEN* on the opposite spectrum.

Figure S7

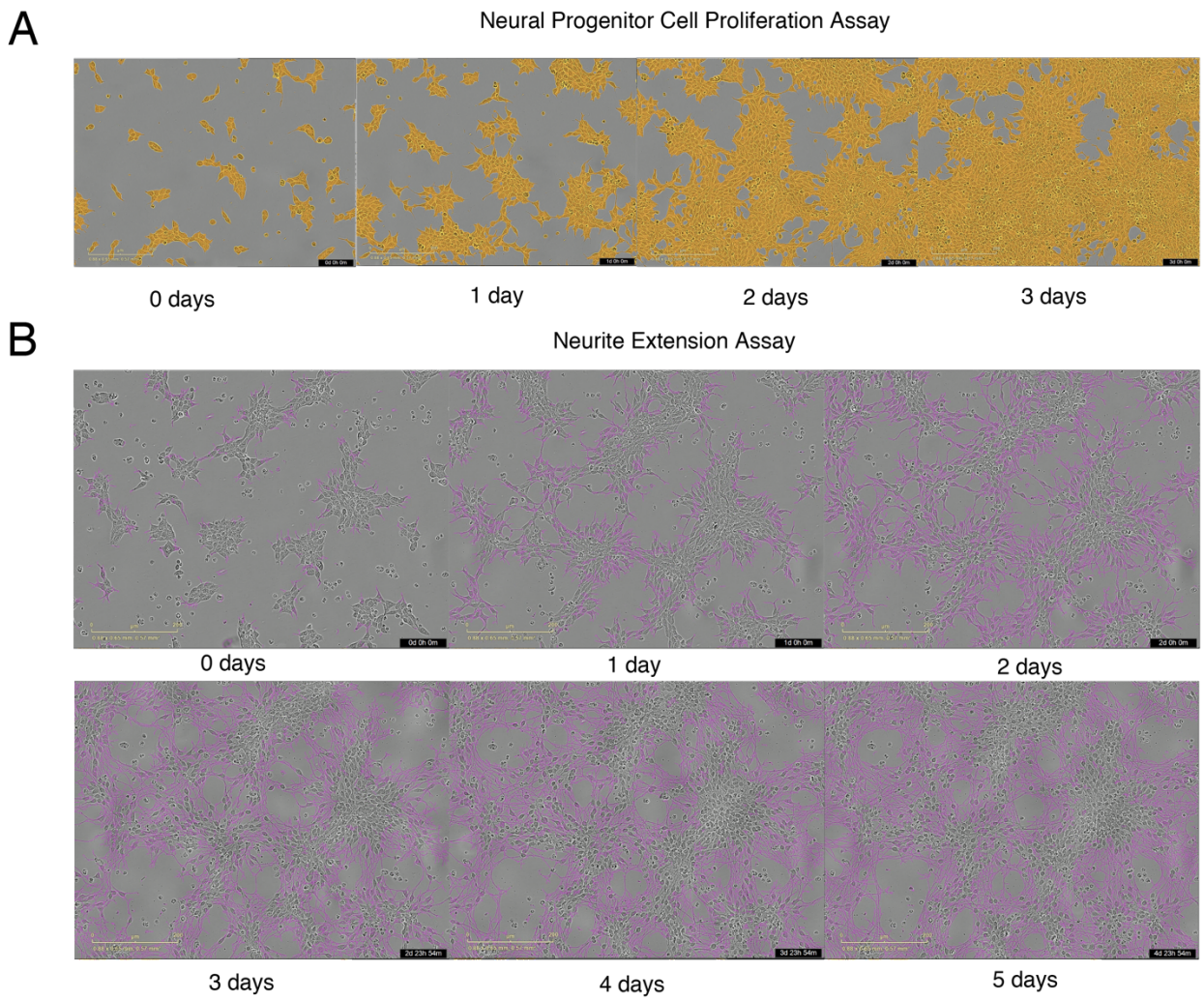

Figure S7: Time-lapse imaging of neural progenitor cell proliferation and neurite extension. A) Neural progenitor cell proliferation is measured by creating a cell mask (orange) and computing the area of confluence at each time point. **B)** Neurite extension is measured with the NeuroTrack assay in the IncuCyte software. Neurite masks are shown (purple). Neurite extension lengths are normalized by cell cluster area. For both assays, images were acquired every 4 hours. Proliferation assay was performed for 3 days. Neurite extension was monitored for 5 days.

Figure S8

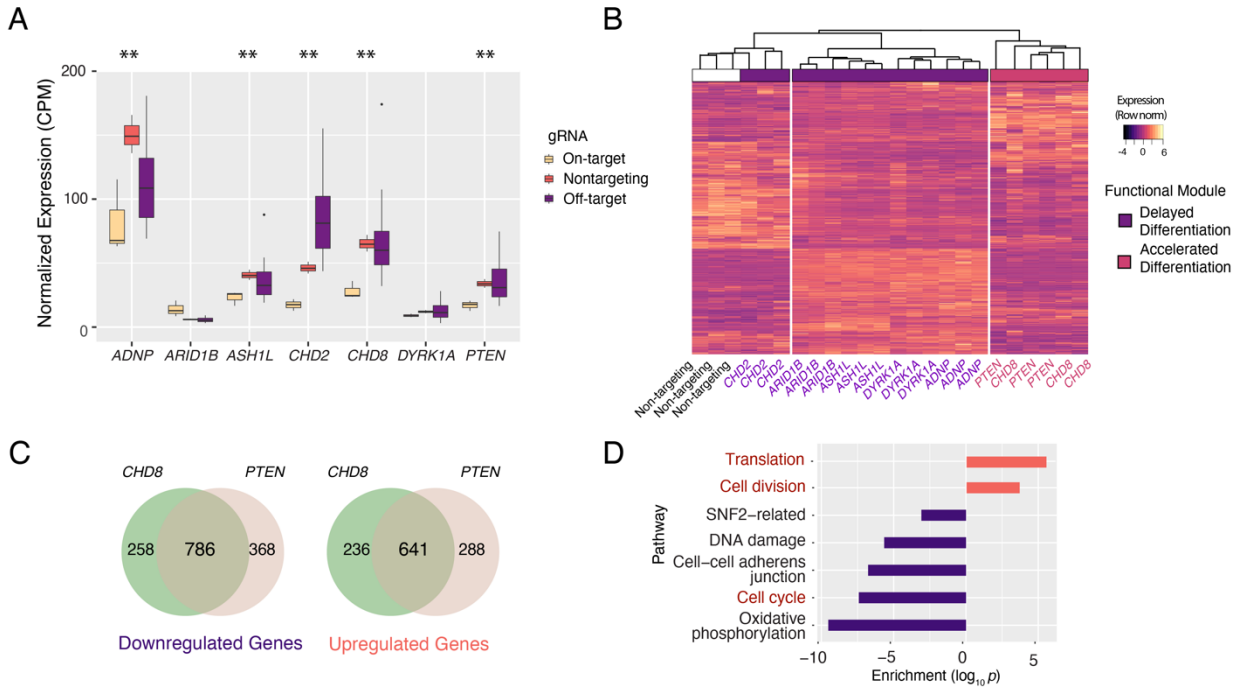

Figure S8: ASD-gene repression in iPSC-derived neural progenitor cells confirms module gene membership and transcriptional convergence. A) Gene expression level in RNA-seq data confirms efficient repression for 5/7 targeted genes. Low detection of *ARID1B* and *DYRK1A* precludes interpretation. (** = p < 1e-5). **B)** Unsupervised, hierarchical clustering of RNA-seq replicates by highly variable genes clusters functional module genes together. **C**) *CHD8* and *PTEN* repression elicits shared transcriptional consequences on down- and up-regulated genes. **D)** These genes are enriched for increased cell division, translation, and decreased cell-cycle genes. Downregulated genes are plotted with negative enrichment scores.

Figure S9

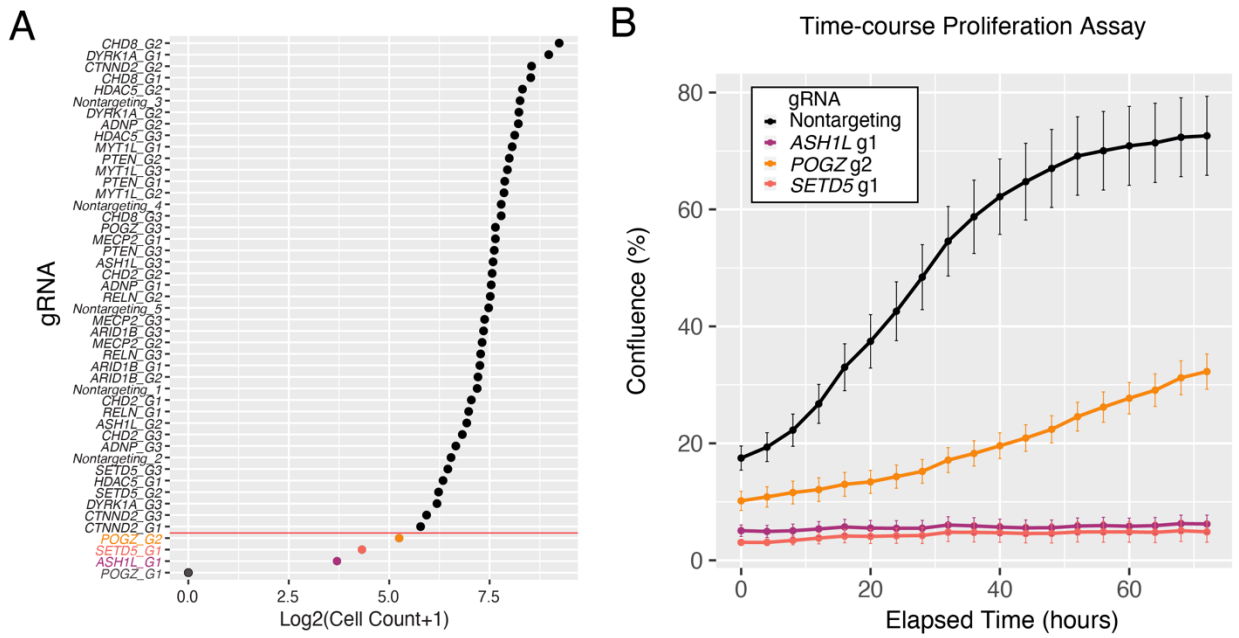

Figure S9: gRNA abundance in final single-cell dataset. A) Four gRNAs were significantly depleted from the pooled single-cell experiment (χ-squared test, $p < 0.01$). **B)** Proliferation assay in LUHMES infected with individual depleted gRNAs reveals severe reductions in neural progenitor cell proliferation.

Figure S10

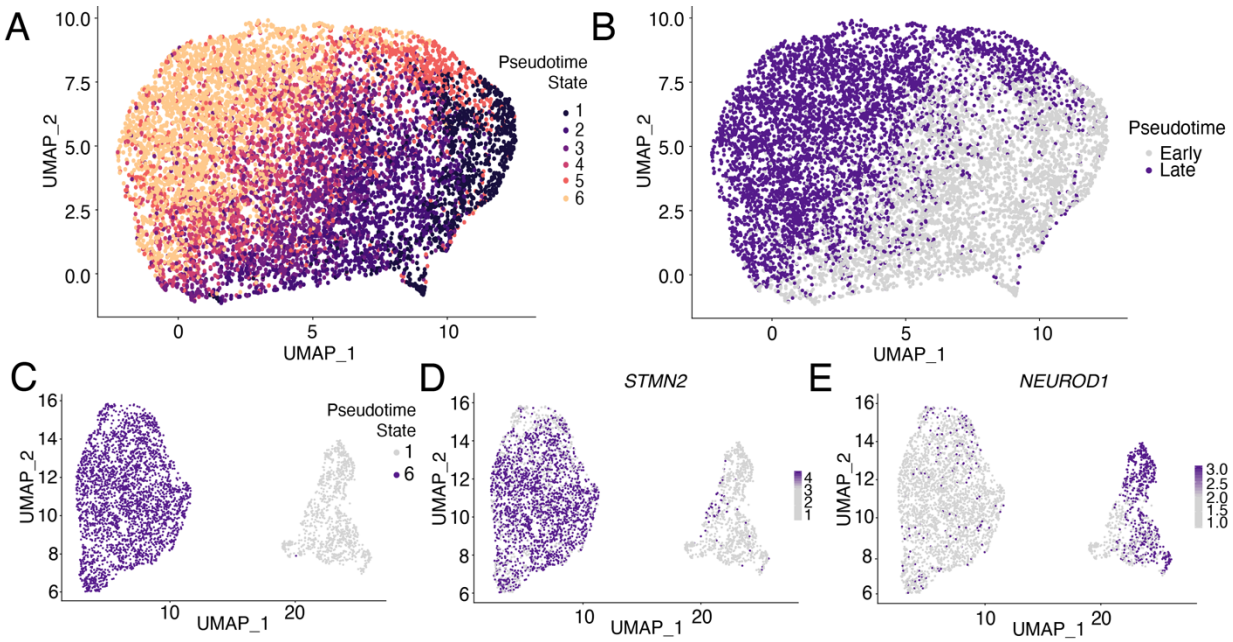

Figure S10: Pseudotime is apparent in global clustering and cells in extreme pseudotime states cluster separately. A) Pseudotime states labels (see Figure 3A) were transferred into UMAP space and show a continuous diagonal trajectory from lower right to upper left. **B)** Early and Late Pseudotime stages labels bisect the UMAP plot. **C)** Clustering only cells in the extreme pseudotime states 1 and 6 reveals two discrete neuronal populations. **D)** The left cluster (state 6) expresses later neuronal marker *STMN2*. **E)** The right cluster (state 1) expresses early neuronal transcription factor *NEUROD1.*

Supplemental Tables

Supplemental Tables S1-S3 are provided as separate XLS documents due to their length.

Table S4

GO:0044770, Name:cell cycle phase transition	GO:0010721, Name:negative regulation of cell development
APP	ADGRG1
CALM1	APP
CENPF	B2M
CARM1	CARM1
CUL5	CBFA2T2
CDK2AP2	CDK5
CLASP2	CNTN2
DCTN1	CTNNB1
DYNC1H1	DRAXIN
EZH2	EIF4E
EIF4E	GAL
FBXL15	ITGB1BP1
KHDRBS1	ITM2C
MECP2	LSM1
MNAT1	
NES	
PHLDA1	
PPP2R2A	
TAOK1	
TUBA1A	
TUBA4A	
ZNF385A	

Supplemental Table S4: Cell-cycle phase transition genes and negative regulation of cell development genes were differentially expressed after repression of 'Delayed Differentiation' genes in early pseudotime stage cells.

Table S5

Primer Name	Primer Sequence
ADNP qPCR fw	CATGGGAGGATGTAGGACTGT
ADNP qPCR rv	ATGGACATTGCGGAAATGACT
CTNND2 qPCR fw	AGGTCCCCGTCCATTGATAG
CTNND2 qPCR rv	ACTGGTGCTGCAACATCTGAA
PTEN qPCR fw	TTTGAAGACCATAACCCACCAC
PTEN qPCR rv	ATTACACCAGTTCGTCCCTTTC
DYRK1A qPCR fw	AAGAAGCGAAGACACCAACAG
DYRK1A qPCR rv	TTTCGTAACGATCCATCCACTTT
CHD8 qPCR fw	CTGCACAGTCACCTCGAGAA
CHD8 qPCR rv	TGGTTCTTGCACTGGTTCAG
HDAC5 qPCR fw	GTACCCAGTCCTCCCCTGC
HDAC5 qPCR rv	GCACATGCACTGGTGCTTTA

ACTB qPCR fw	CATGTACGTTGCTATCCAGGC
ACTB-qPCR rv	CTCCTTAATGTCACGCACGAT

Table S5: Primers used for qPCR.

Table S6

Primer Name	Primer Sequence
pSeq1-BC1-UMI-dt-VN	CTACACGACGCTCTTCCGATCTCTGATAGCATGGTCATNNNNNVVVVVTTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
pSeq1-BC2-UMI-dtVN	CTACACGACGCTCTTCCGATCTCACAGTAGTTAGGGTGNNNNNVVVVVTTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
pSeq1-BC3-UMI-dtVN	CTACACGACGCTCTTCCGATCTGTAACTGCATGGTCTANNNNNVVVVVTTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
pSeq1-BC4-UMI-dtVN	CTACACGACGCTCTTCCGATCTACTGAACCAGTGGGATNNNNNVVVVVTTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
pSeq1-BC5-UMI-dtVN	CTACACGACGCTCTTCCGATCTAACACGTTCAGTTCGANNNNNVVVVVTTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
pSeq1-BC6-UMI-dtVN	CTACACGACGCTCTTCCGATCTTATCAGGGTTTAGCTGNNNNNVVVVVTTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
pSeq1-BC7-UMI-dtVN	CTACACGACGCTCTTCCGATCTACTTTCATCGTAGGAGNNNNNVVVVVTTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
pSeq1-BC8-UMI-dtVN	CTACACGACGCTCTTCCGATCTAACGTTGGTAGCGTCCNNNNNVVVVVTTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
pSeq1-BC9-UMI-dtVN	CTACACGACGCTCTTCCGATCTCTCATTACAGACGCCTNNNNNVVVVVTTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
pSeq1-BC10-UMI-dtVN	CTACACGACGCTCTTCCGATCTTGACTAGCAGGGTTAGNNNNNVVVVVTTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
pSeq1-BC11-UMI-dtVN	CTACACGACGCTCTTCCGATCTGCGGGTTAGTAATCCCNNNNNVVVVVTTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
pSeq1-BC12-UMI-dtVN	CTACACGACGCTCTTCCGATCTTCTCATAGTTGTGGAGNNNNNVVVVVTTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
SMART-TSO	AAGCAGTGGTATCAACGCAGAGTrGrGrG
Partial seq1	CTACACGACGCTCTTCCGATCT
SMART	AAGCAGTGGTATCAACGCAGAGT
P5-index-seq1	AATGATACGGCGACCACCGAGATCTACACAGGACA ACACTCTTTCCCTACACGACGCTCTTCCGATCT

Table S6: BRB-seq oligo names and sequences. r denotes RNA bases.

Table S7

P5-index-Seq1-fw	AATGATACGGCGACCACCGAGATCTACACAGGACAACACTCTTTCCCTACACGACGCTCTTCCGATCT
P7-Index-Seq2-10x-sgRNA	CAAGCAGAAGACGGCATACGAGATCGGGCAACGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGTGGAAAGGACGAAACA*C*C*G

Table S7: Primers for gRNA specific amplification. * denotes phosphorothioate modification to reduce mispriming due to proof-reading polymerase

Supplemental References

Alpern D, Gardeux V, Russeil J, Mangeat B, Meireles-Filho ACA, Breysse R, Hacker D, Deplancke B. 2019. BRB-seq: ultra-affordable high-throughput transcriptomics enabled by bulk RNA barcoding and sequencing. *Genome Biol* **20**: 71.

Dixit A, Parnas O, Li B, Chen J, Fulco CP, Jerby-Arnon L, Marjanovic ND, Dionne D, Burks T, Raychowdhury R, et al. 2016. Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens. *Cell* **167**: 1853-1866.e17.

Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**: 139–140.

Sanson KR, Hanna RE, Hegde M, Donovan KF, Strand C, Sullender ME, Vaimberg EW, Goodale A, Root DE, Piccioni F, et al. 2018. Optimized libraries for CRISPR-Cas9 genetic screens with multiple modalities. *Nat Commun* **9**: 5416.

Scholz D, Pöltl D, Genewsky A, Weng M, Waldmann T, Schildknecht S, Leist M. 2011. Rapid, complete and large-scale generation of post-mitotic neurons from the human LUHMES cell line. *J Neurochem* **119**: 957–971.