

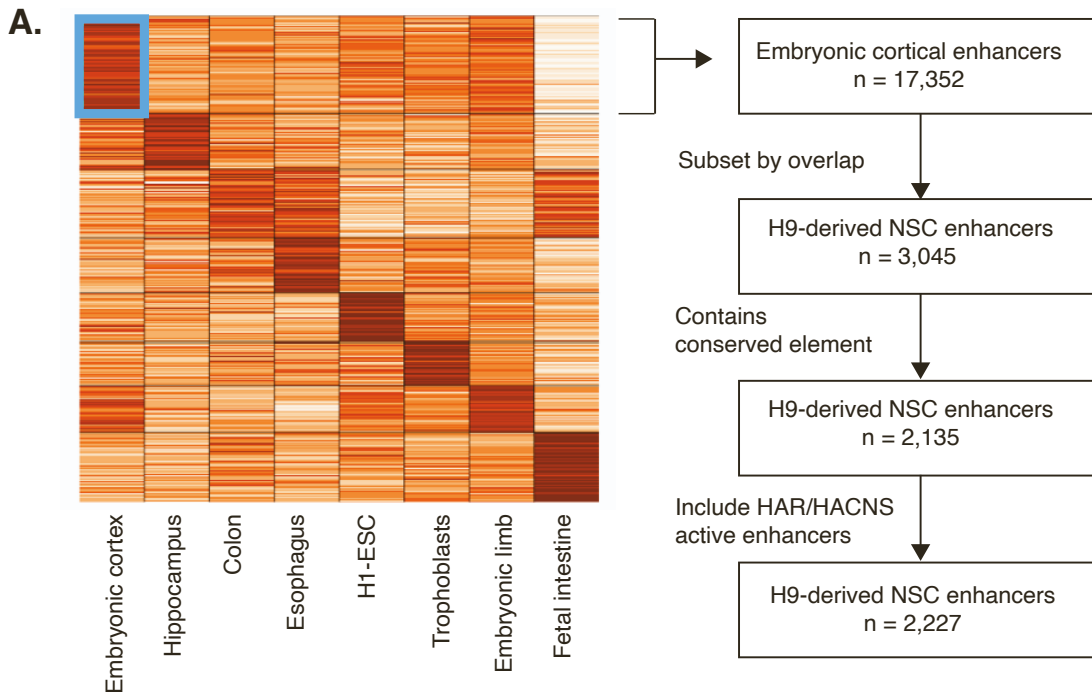
Cell Reports, Volume 43

Supplemental information

**Massively parallel disruption of enhancers active
in human neural stem cells**

Evan Geller, Mark A. Noble, Matheo Morales, Jake Gockley, Deena Emera, Severin Uebbing, Justin L. Cotney, and James P. Noonan

Supplementary Figures



B. GREAT, enriched GO Biological Function

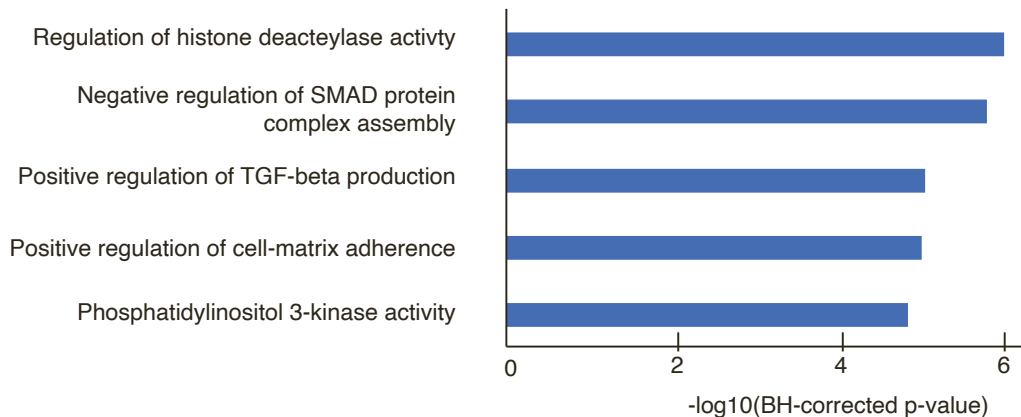


Figure S1. Identification of cortical-enriched enhancers active in hNSC (A) (Left) K-means clustering with embryonic, fetal, and other adult cell and tissue types to identify cortical-associated H3K27ac intervals. (Right) Filtering scheme for the identification of active H3K27ac intervals overlapping active H3K27ac intervals in hNSC. **(B)** Gene ontology enrichments identified by GREAT for cortical-enhancers active in hNSC. P-values are derived from the binomial test reported by GREAT.

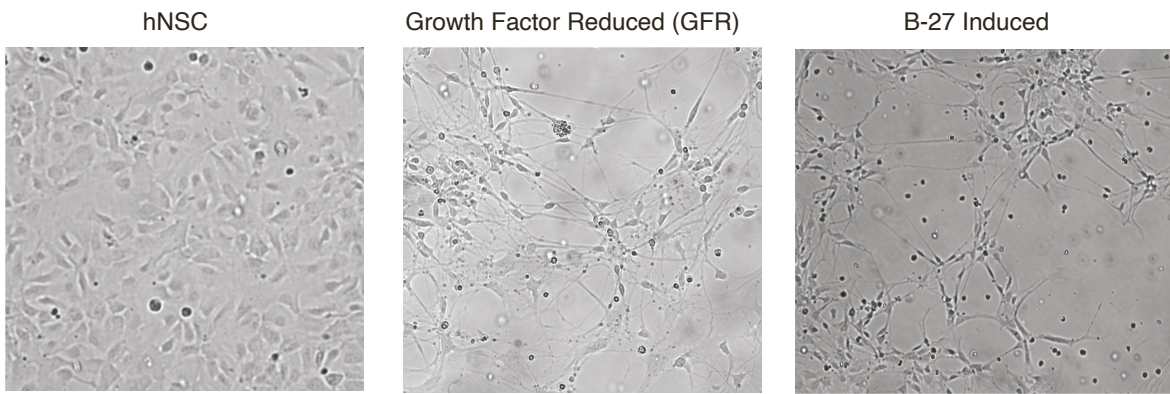
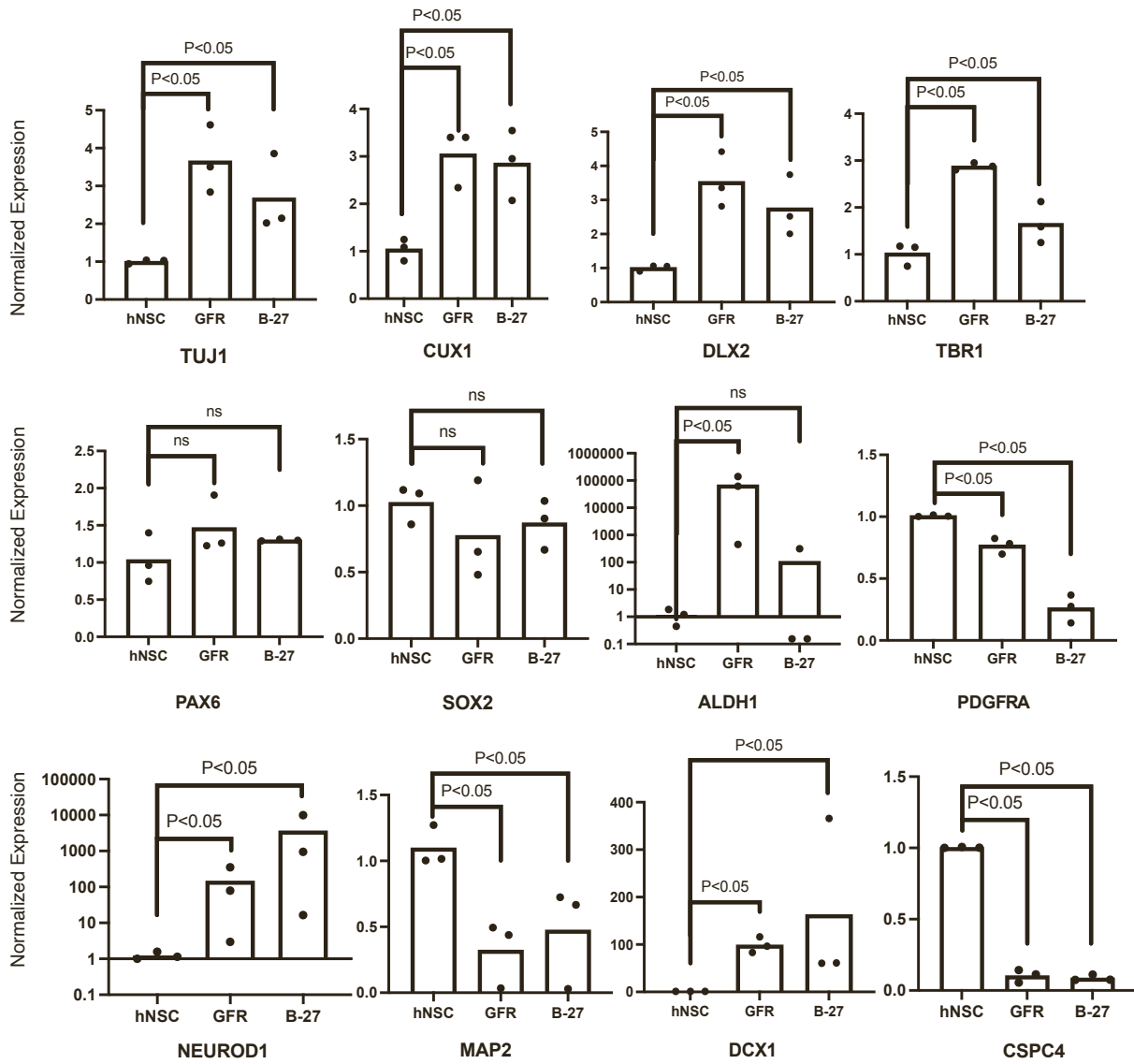
A.**B.**

Figure S2. Human neural stem cell (hNSC) neuronal differentiation. (A) Representative brightfield images of hNSC (*left*), growth factor reduced (*middle*), and B27 induced (*right*) cell culture after 21 days post differentiation. (B) Normalized gene expression readout by qPCR for neuronal cell type markers. Bar plots represent housekeeping gene (EEF2) normalized expression values relative to hNSC. Individual data points represent the average of three biological replicates. Statistical tests were performed using a two-way Student's T-test.

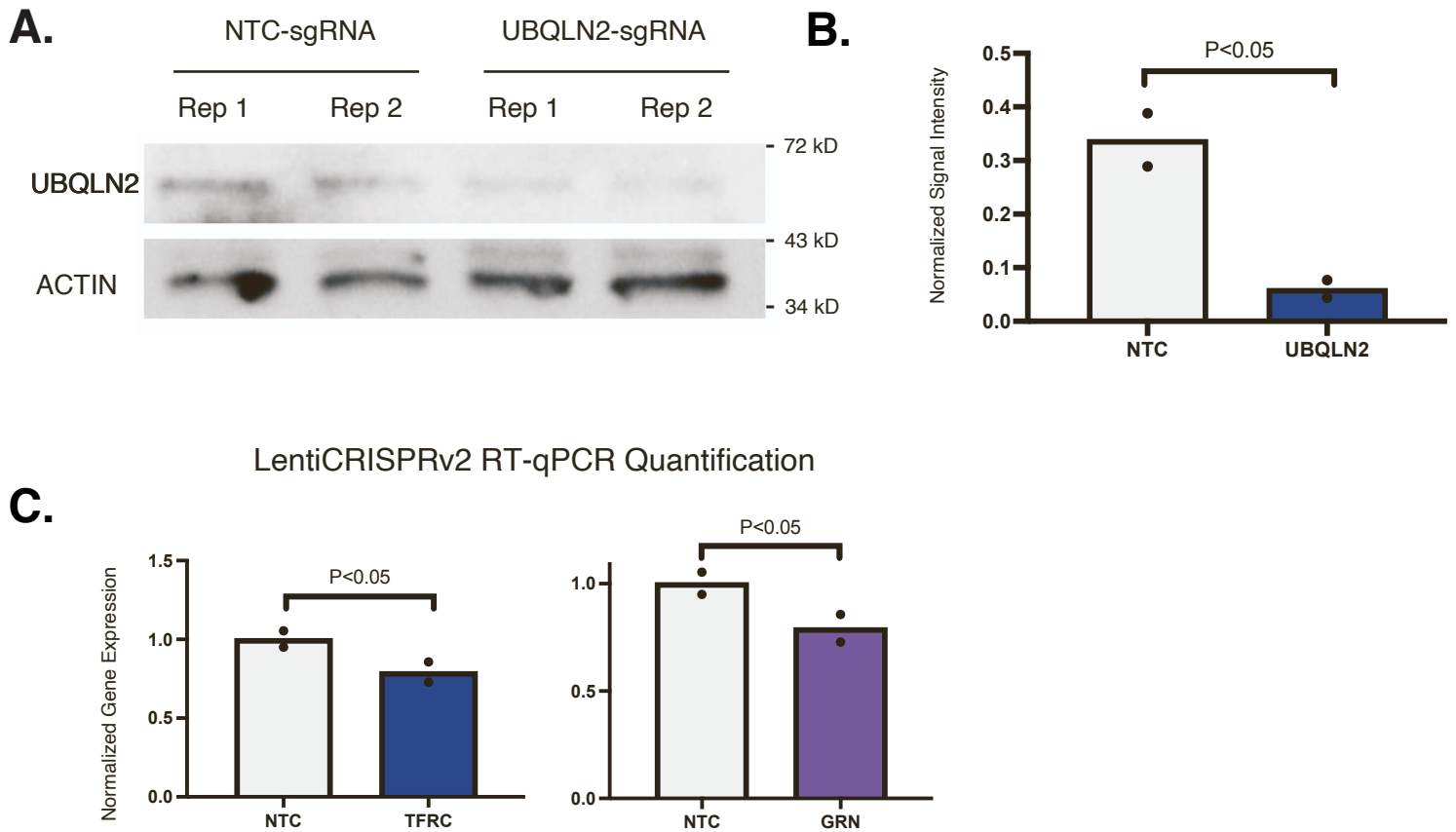
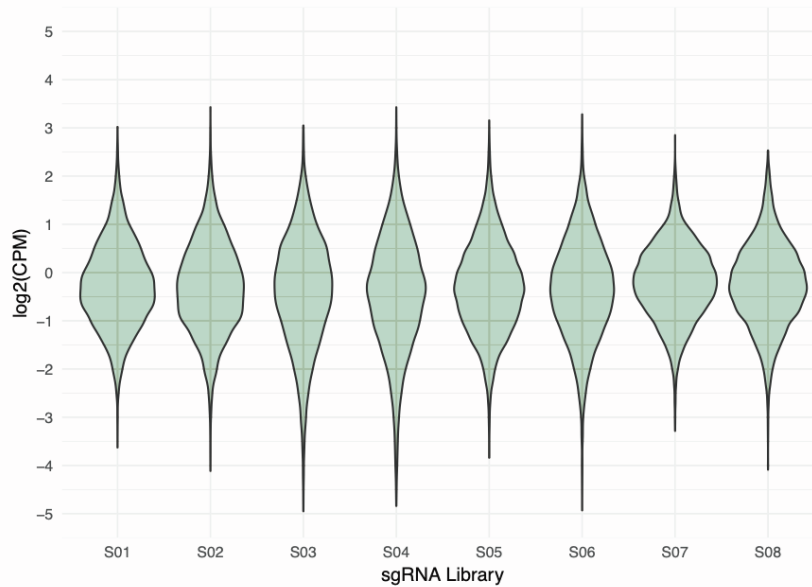


Figure S3. Validation of CRISPR-mediated knockdown in hNSC. (A) Western blot showing UBQLN2 expression following transduction with non-targeting control sgRNA (*left*) and UBQLN2-targeting sgRNA (*right*). (B) Knockdown of UBQLN2 determined by ACTIN-normalized signal intensity for non-targeting control sgRNA and UBQLN2-targeting sgRNA. Individual data points represent the average of two biological replicates. Statistical tests were performed using a two-way Student's T-test.

A. Individual sgRNA representation for each plasmid sub-library



B.

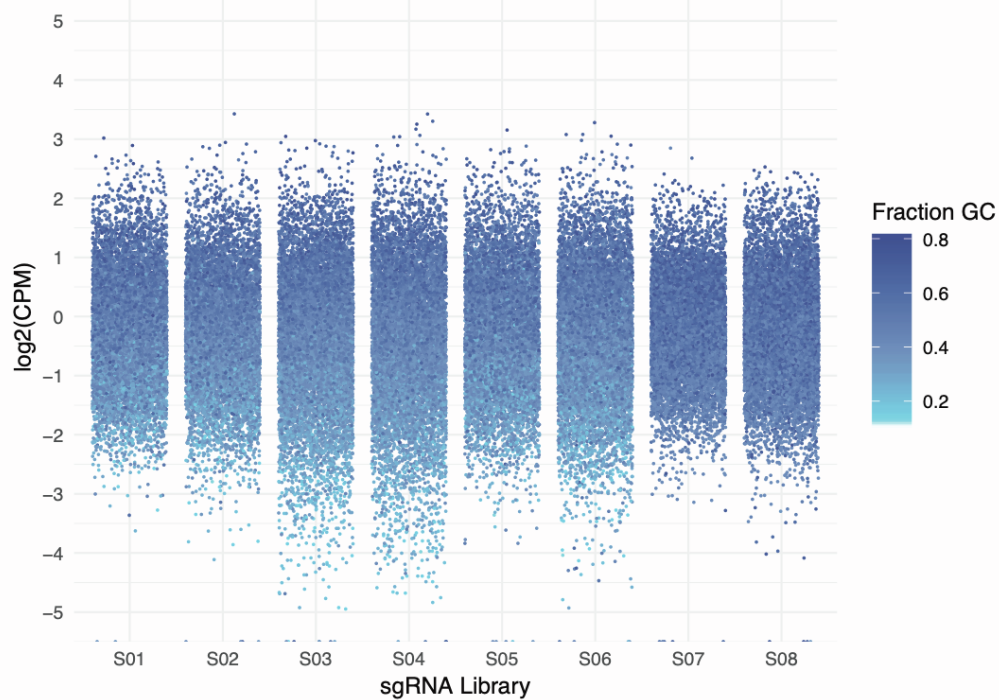


Figure S4. Quality control of sgRNA-Cas9-GFP plasmid libraries. Plasmid libraries S01-S06 are enhancer targeting and S07-S08 are gene targeting. (A) Violin plots of individual sgRNA representation relative to mean representation. $\log_2(\text{counts per million})$ are shown for each sub-library included in the study. (B) Jitter plots showing the percentage GC-content of individual sgRNAs for each sub-library.

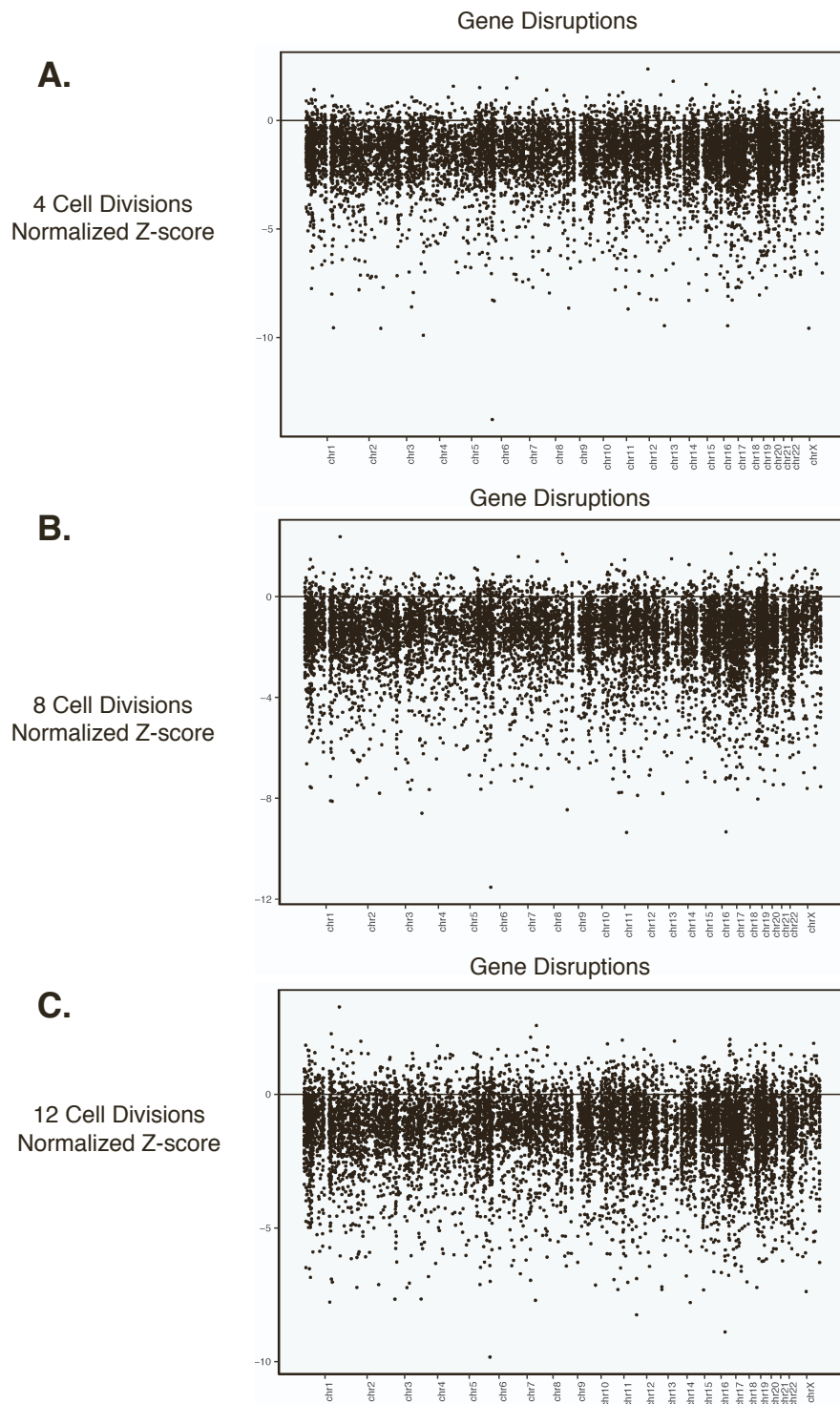


Figure S5. Genome-wide plots of gene disruption biological effects on cell proliferation. The biological effects on cell proliferation are normalized Z-scores relative to mean and standard deviation of the genomic background controls. Plots are shown for all genomic loci at 4 cell divisions (**A**), 8 cell divisions (**B**), and 12 cell divisions (**C**).

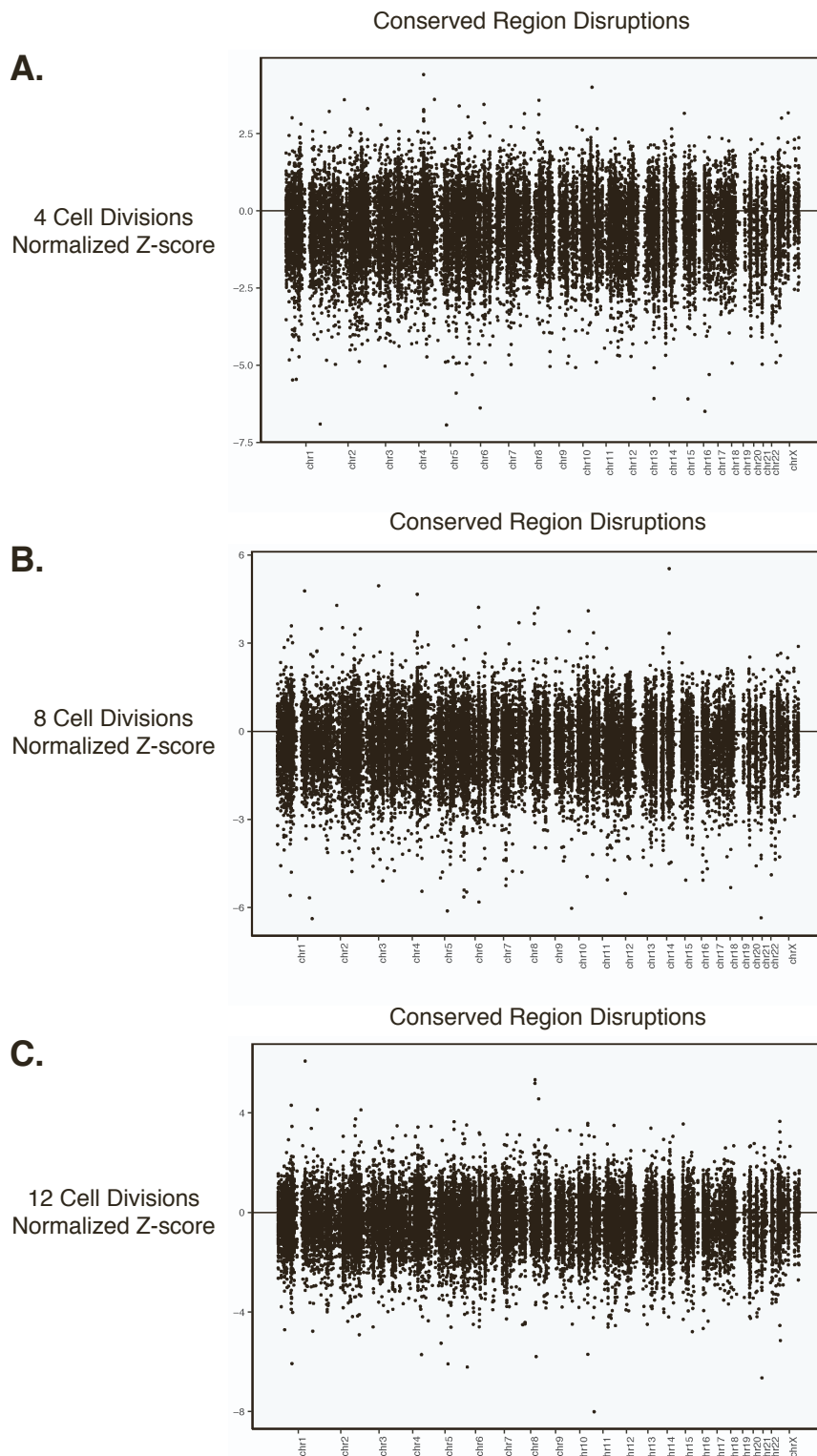


Figure S6. Genome-wide plots of conserved region disruption biological effects on cell proliferation. The biological effects on cell proliferation are normalized Z-scores relative to mean and standard deviation of the genomic background controls. Plots are shown for all genomic loci at 4 cell divisions (**A**), 8 cell divisions (**B**), and 12 cell divisions (**C**).

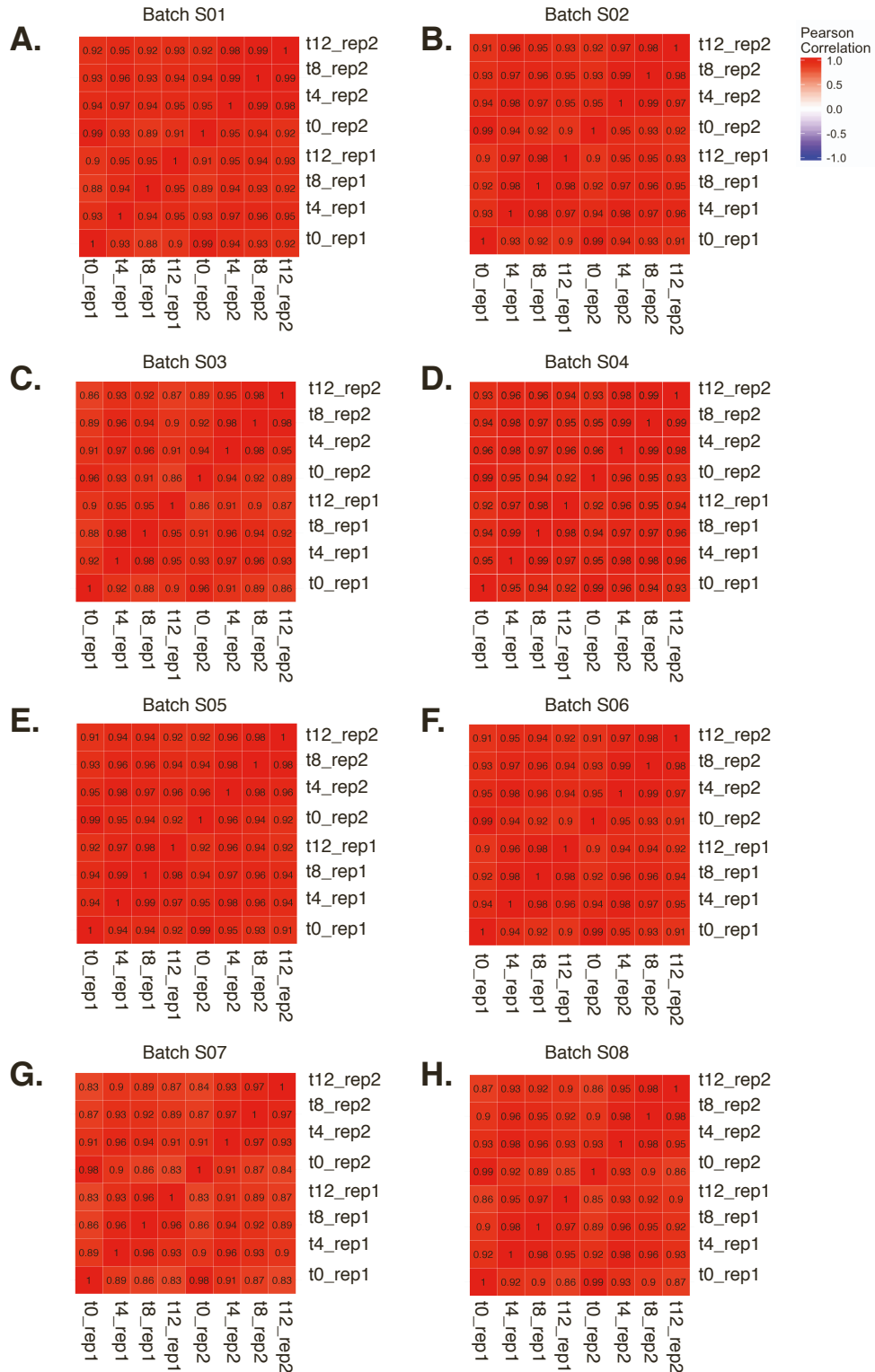


Figure S7. Lentivirus-integrated sgRNA read count correlation plots. (A-H) For each sub-library, the Pearson correlation coefficient is reported for t0, t4, t8, and 12 for rep1 and rep2 read counts sequenced and mapped by high-throughput sequencing.

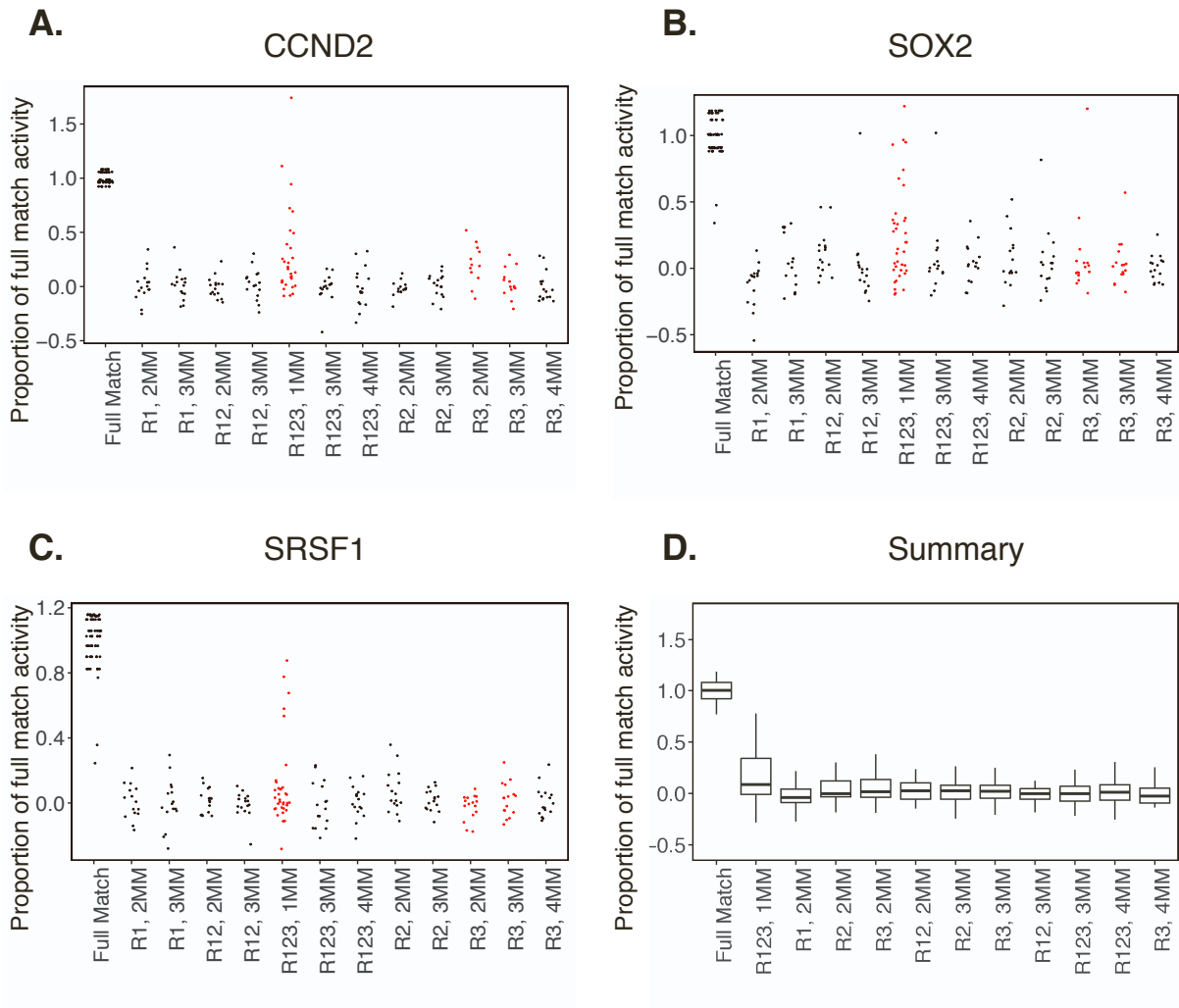


Figure S8. Proportion of on-target activity quantified by specificity controls. The sgRNA scoring scheme was used to test potential off-target effects on cell proliferation. Mismatches were introduced to the on-target sgRNA sequence for CCND2 (A), SOX2 (B), and SRSF1 (C). R1, R2, and R3 are regions of the 20 nt sgRNA sequences where mismatches were introduced. For each condition, 1MM (1 mismatch) 2MM (2 mismatches), 3MM (3 mismatches), or 4MM (4 mismatches) were introduced within the region to generate specificity control sgRNAs ($n > 15$ sgRNAs per condition). Jitter plots for each region and mismatch pattern are shown (A-C). Jitter plots compatible with the scoring criteria to be included in the study (black) and those not meeting scoring criteria (red) are shown. (D) Summary of on-target activity across CCND2, SOX2, and SRSF1 when mismatches are introduced.

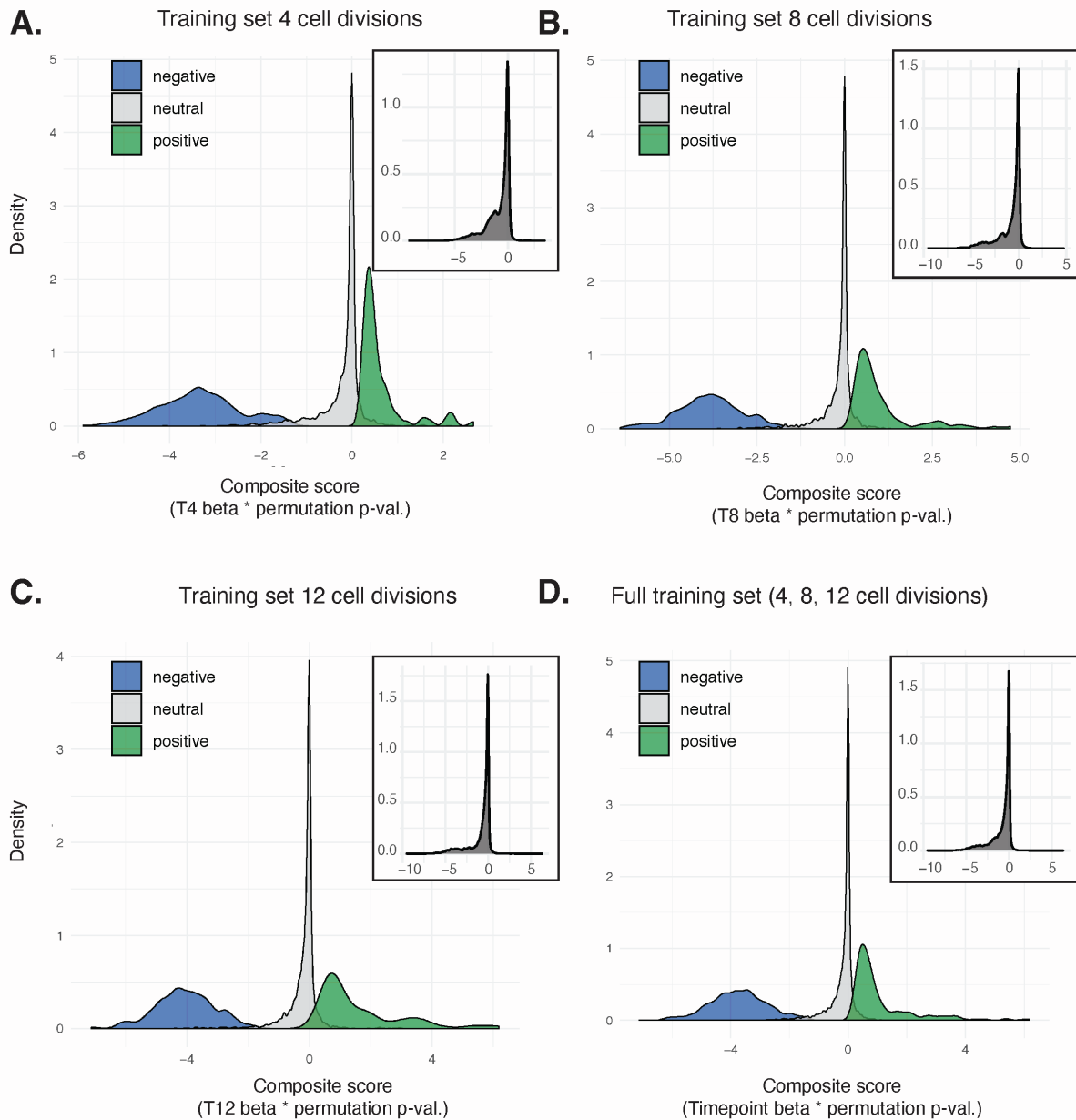


Figure S9. Training set for linear-discriminant analysis and the identification of proliferation phenotypes. Density plots for negative (blue), neutral (light gray), and positive (green) proliferation phenotypes are shown for 4 cell divisions (A), 8 cell divisions (B), and 12 cell divisions (C). (D) The full training set across all timepoints included in the study. Each panel includes the full distribution of data collected during the experiment (dark gray, inset). The training set was defined as follows: the ‘decreasing’ population includes 500 genes decreasing proliferation across a panel of cancer cell lines (Wang, et al. 2015), the ‘neutral’ population is all regions within the genomic background, and the ‘increasing’ population is comprised of the top 1% of proliferation increasing regions identified at each time point (4, 8, and 12 cell divisions)

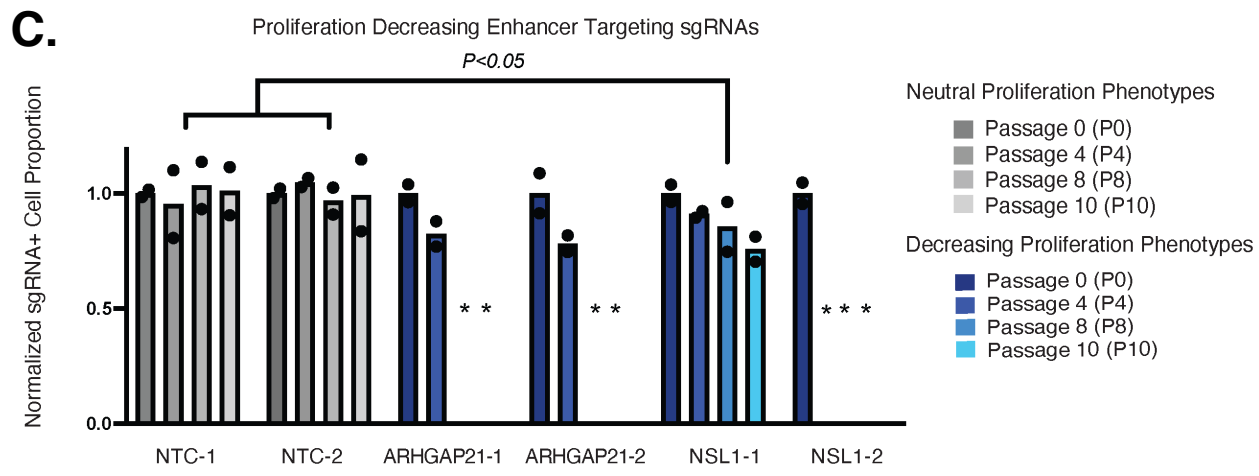
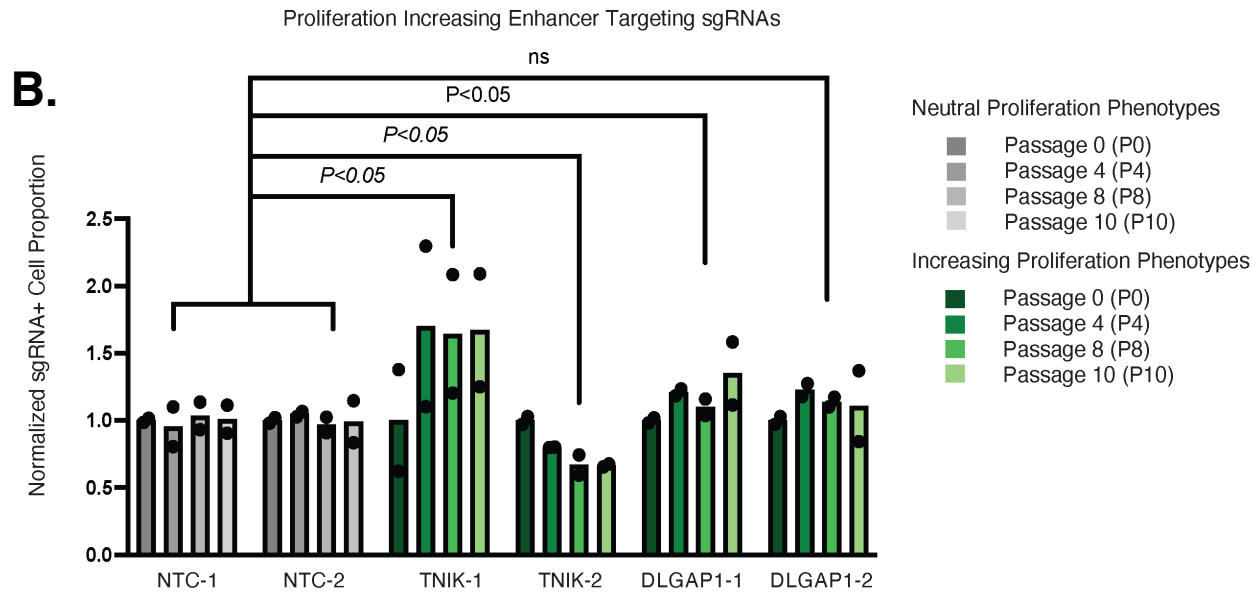
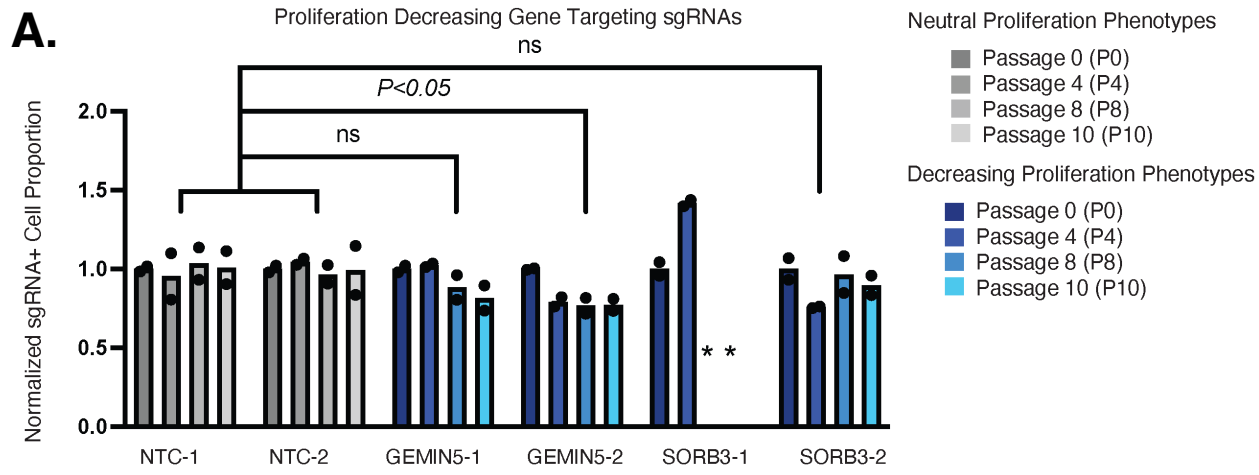


Figure S10. Proliferation-phenotype validation assay for protein-coding genes and promoter-interacting conserved regions within enhancers. Proliferation phenotypes for individual sgRNAs were determined by transduction in hNSC and tracking the proportion of GFP-positive cells over time, relative to the non-targeting controls. **(A)** Proliferation-decreasing protein-coding targets, **(B)** proliferation-increasing enhancer targets, and **(C)** proliferation-decreasing enhancer targets were included. In **B** and **C**, enhancer targets are labeled based on their putative gene target. Experimental conditions that resulted in cell culture collapse due to low viability are indicated by an asterisk (*). Individual data points represent the average of two biological replicates. Statistical tests were performed using a two-way Student's T-test. P values were not calculated for targets that exhibited cell culture collapse.

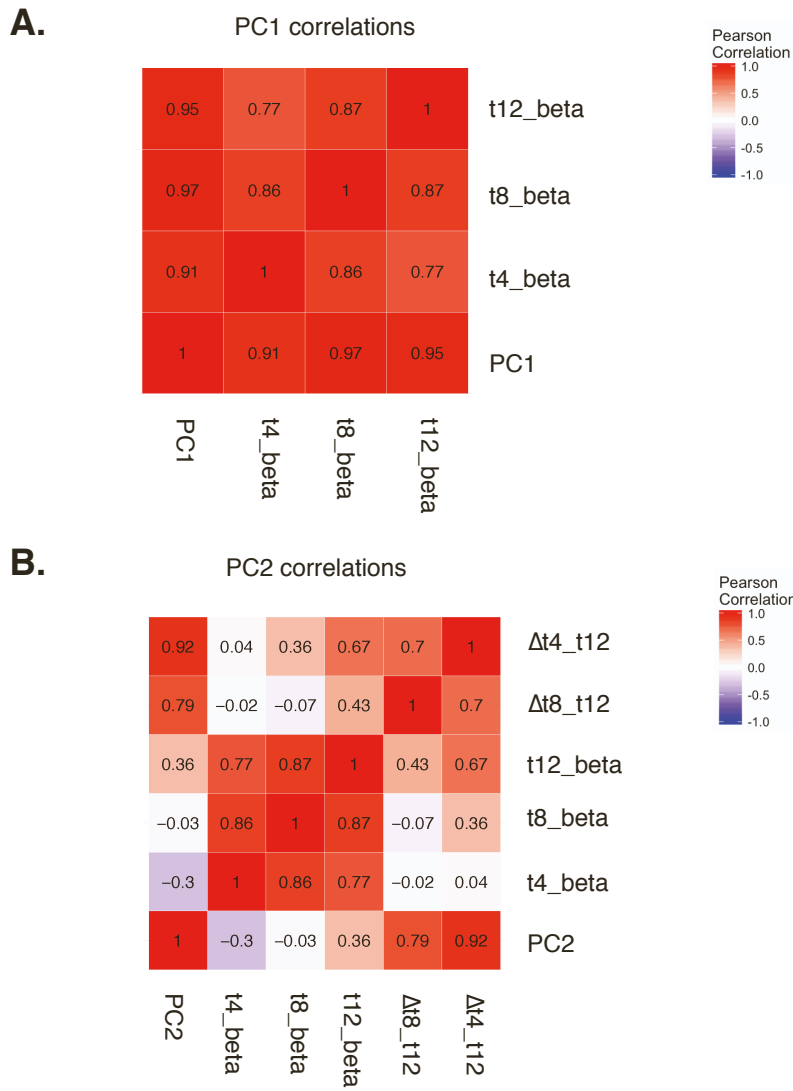




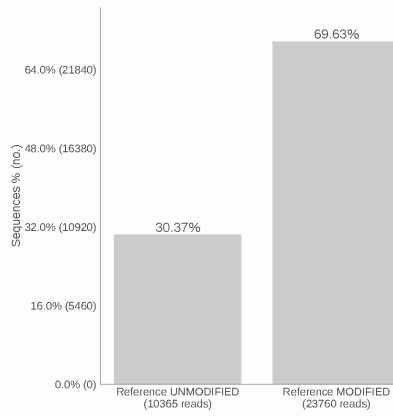
Figure S11. PC1 and PC2 correlation matrices following principal component analysis. (A) PC1 Pearson correlations with Beta scores across all targeted regions (genes, enhancers, genomic background controls) and time points. **(B)** PC2 correlations with joint beta scores and change in beta scores calculated from 8 to 12 cells divisions and from 4 to 12 cell divisions.

A.

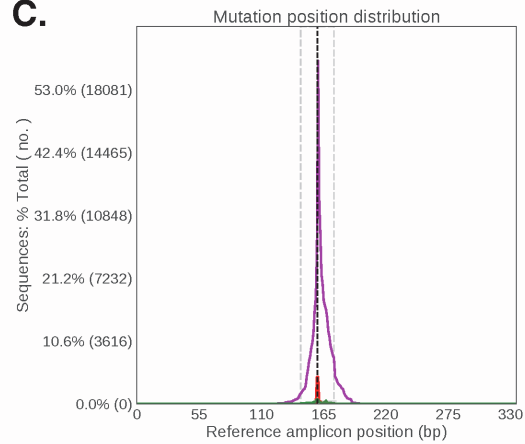
Putative NPAS3 Enhancer
Human-Accelerated Region HACNS96

Human	T	C	C	T	T	G	G	G	T	G	T	T	A	A	G	C	A	G	C	C		
Chimpanzee	T	T	C	T	T	G	G	G	T	G	T	T	A	A	A	C	A	G	C	C		
Mouse	T	T	C	T	T	G	G	G	T	G	T	T	A	A	A	C	A	G	C	C		
TFBS	TBX2  TBX20 																					
Modified Alleles	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	C	C	
	T	C	C	T	T	G	G	-	-	-	-	-	-	-	-	-	C	A	G	C	C	
	T	C	C	T	T	G	G	-	-	-	-	-	-	-	-	G	C	A	G	C	C	
	T	C	C	T	T	G	G	-	-	-	-	-	-	-	-	-	-	-	-	G	C	C
	T	C	C	T	T	G	G	-	-	-	-	-	-	-	-	-	-	-	-	-	C	C
	-	-	-	-	-	-	-	-	-	-	-	-	A	A	G	C	A	G	C	C	-	-
	T	C	C	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	T	-	-	-	-	-	-	-	-	-	T	A	A	G	C	A	G	C	C	-	-	-

B.



C.



D.

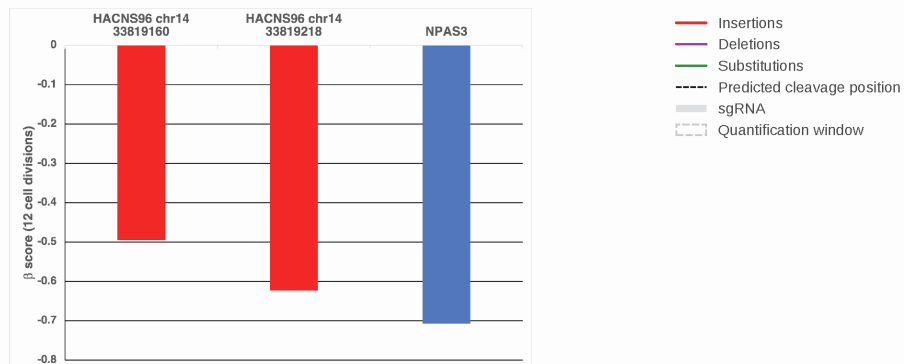


Figure S12. Example of a proliferation-altering disruption that affects human-specific substitutions within a HAR (*HACNS96*). (A) (Top) Alignment of *HACNS96* with orthologous sequences in chimpanzee, and mouse, adapted from the 46-way UCSC Multiz alignment (GRCh37/hg19). A human-specific mutation (red) overlaps with predicted TBX2/TBX20 transcription factor binding site (JASPAR 2018). (Bottom) Modified alleles resulting from sgRNA-Cas9-induced disruption. (B) Quantification of allele modification following sgRNA-Cas9 disruption. (C) Mutational spectrum around the predicted cleavage site for sgRNA-Cas9 disruptions. (D) The effects of two independent disruptions in *HACNS96* (shown in red) and disruption of *NPAS3* (shown in blue). The starting coordinate for each sgRNA target in *HACNS96* is shown (in GRCh37/hg19).

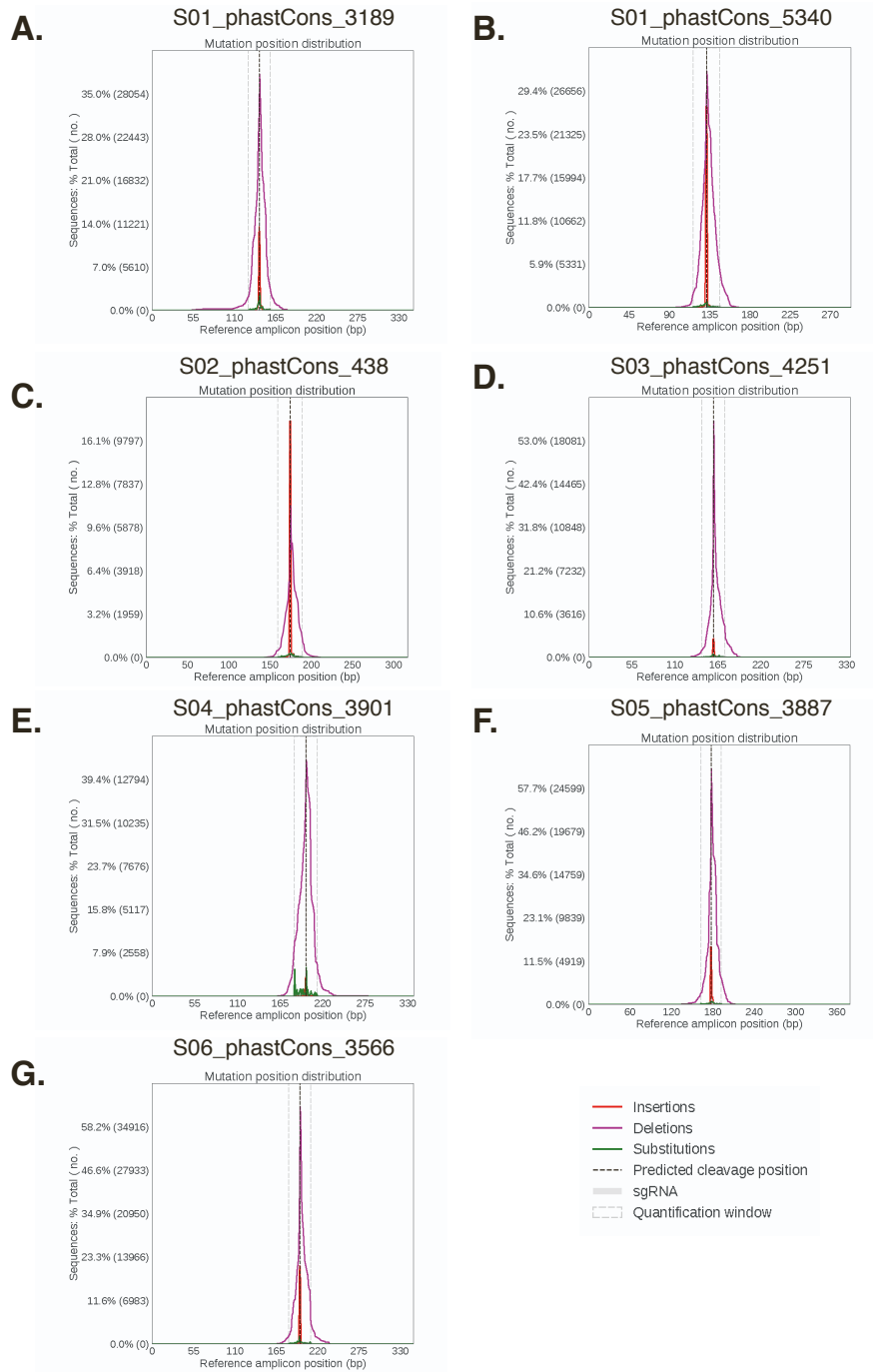


Figure S13. Mutational spectra for disruptions within individual conserved regions. (A-G) Mutation types are quantified at the predicted cleavage site for each individual sgRNA targeting the conserved region (see Table S10). Conserved region identifiers and genomic coordinates (GRCh37/hg19) are reported in Table S3.