

*Supplemental Material for:*

**The Interplay Between Translational Efficiency, Poly(A) Tails, MicroRNAs, and Neuronal  
Activation**

Timothy J. Eisen<sup>1,2</sup>, Jingyi Jessica Li<sup>3</sup>, and David P. Bartel<sup>1,2\*</sup>

<sup>1</sup>Howard Hughes Medical Institute, Cambridge, MA, 02142, USA

<sup>2</sup>Whitehead Institute for Biomedical Research, Cambridge, MA, 02142, USA

<sup>3</sup>Department of Statistics, Department of Biomathematics, and Department of Human Genetics,  
University of California, Los Angeles, CA 90095, USA

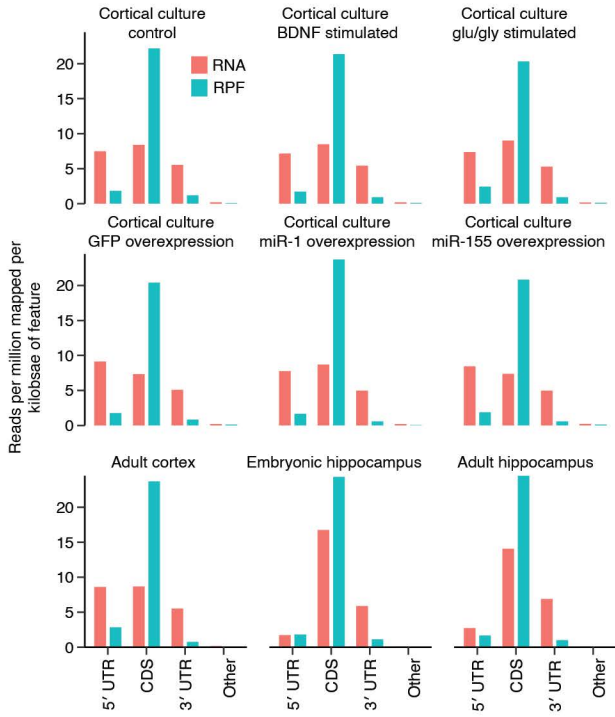
\*Correspondence to: [dbartel@wi.mit.edu](mailto:dbartel@wi.mit.edu).

Running title: Poly(A) Tails, MicroRNAs, and Neuronal Activation

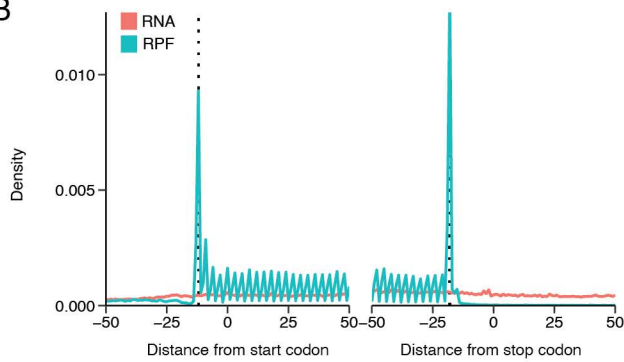
Keywords: poly(A) tails, microRNAs, neurons, translational efficiency, PAL-seq, ribosome  
profiling

Figure S1

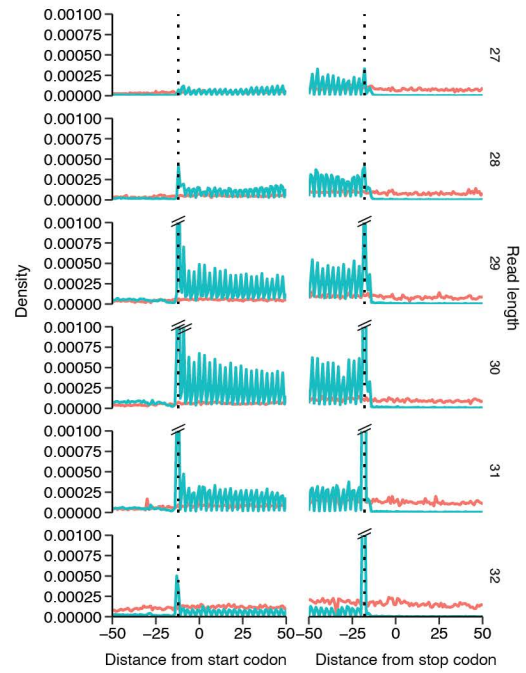
A



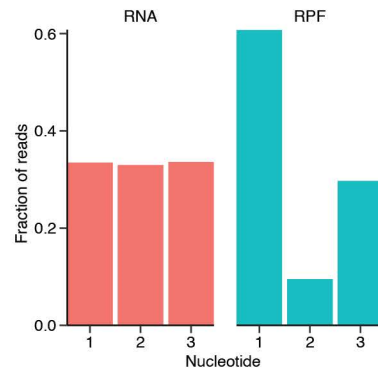
B



C

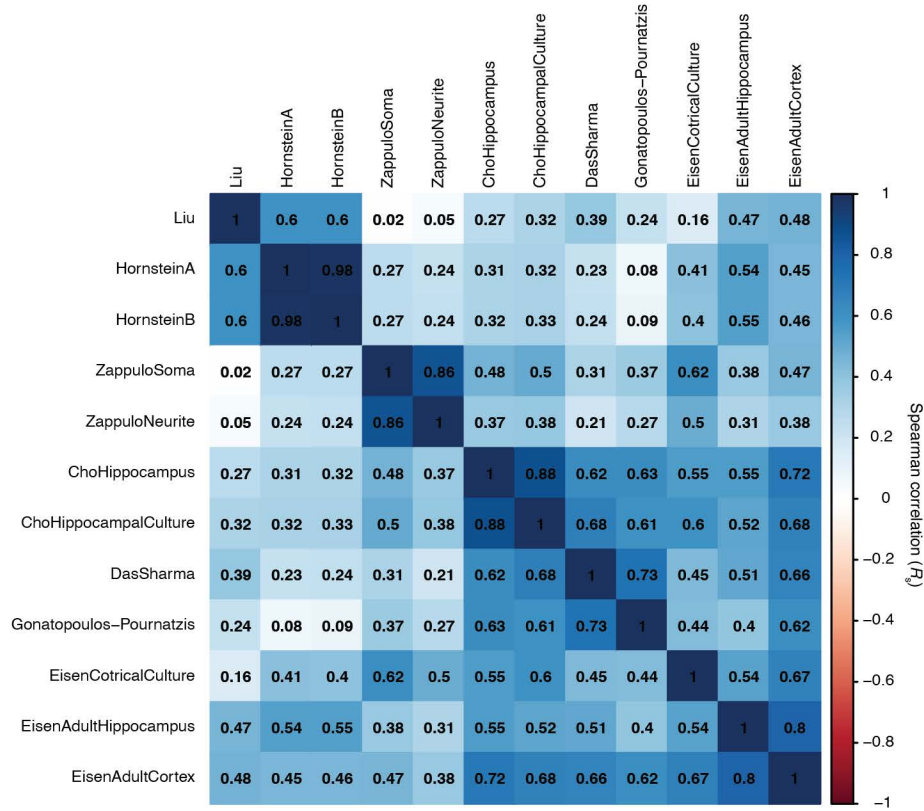


D



**Supplemental Figure S1.** Assessment of ribosome profiling data. (A) Distribution of reads across CDSs and UTRs for the nine ribosome profiling and matched RNA-seq datasets in the main-text figures of this study. For each dataset, the fraction of reads per million mapped reads per kilobase of either CDS or UTR is displayed for the RNA-seq data (red) and the ribosome profiling data (green). Reads that mapped to introns or to regions >10 kb upstream or downstream of the annotated transcription start and end, respectively, contributed to the total, but not to the CDS or UTR categories. Analysis was performed using RSeQC software (Wang et al. 2012). (B) Periodicity in ribosome profiling data. Shown is a meta-transcript from an unstimulated primary cortical culture, generated by plotting the distances between the 5'-most nucleotide of each gene-mapping read and either the first nucleotide of the start codon (left plot) or the last nucleotide of the stop codon (right plot). As expected, 3-nt periodicity was observed in the ribosome profiling but not the RNA-seq data. Dashed lines at -12 and -18 indicate the positions of the 5' end of the RPF inferred for a ribosome with the start codon in its A site or stop codon in its P site, respectively. (C) Periodicity of RPFs of different lengths. Plotted are the data in (B), subsetted by read length, as indicated at the right of each plot. Diagonal black lines indicate that the value exceeded the y axis. Different read lengths resulted primarily from 3' heterogeneity, as they did not alter the positions of the major peaks corresponding to ribosomes with a start codon in their A sites or stop codon in their P sites. (D) Fraction of mRNA-seq and RPF tags mapping to each of the three codon nucleotides in panel (B).

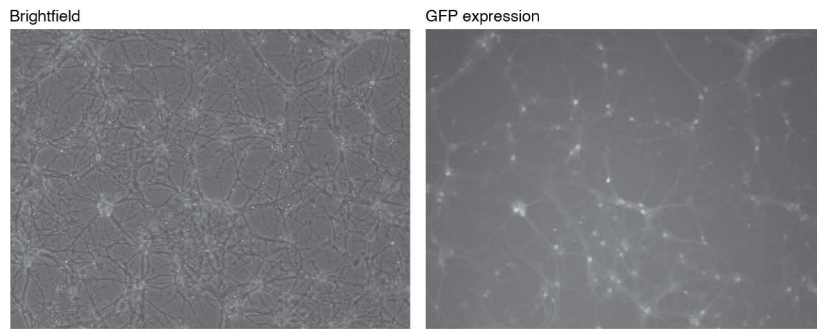
Figure S2



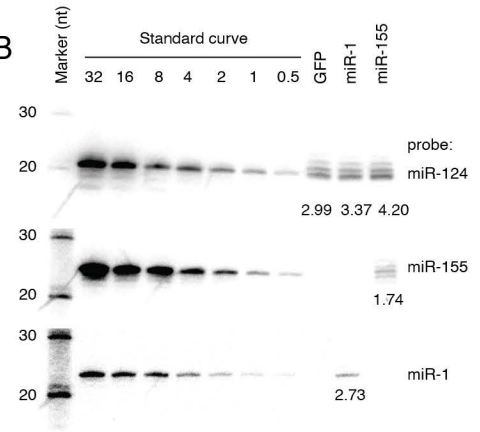
**Supplemental Figure S2.** Agreement between TE measurements from different studies profiling murine brain or neuronal samples. Spearman correlations for pairwise comparisons of TE values ( $\log_2$ ) are plotted, with a color indicating their strength (key). Data were from primary hippocampal culture and tissue (Cho et al. 2015), adult neural stem cells (Liu et al. 2018), mouse cortex (Das Sharma et al. 2019), two biological replicates from brain (Hornstein et al. 2016), soma and neurites from neurons induced from mESCs (Zappulo et al. 2017), and mESC-derived neurons (Gonatopoulos-Pournatzis et al. 2020). Also included were the cortical-culture, adult-hippocampus, and adult-cortex datasets from this study. The datasets were ordered by hierarchical clustering using the hclust command, and results were plotted using the corrplot package in R (Wei and Simko 2017). The analysis considered only mRNAs that passed the expression cutoff in all datasets ( $n = 3,577$ ).

Figure S3

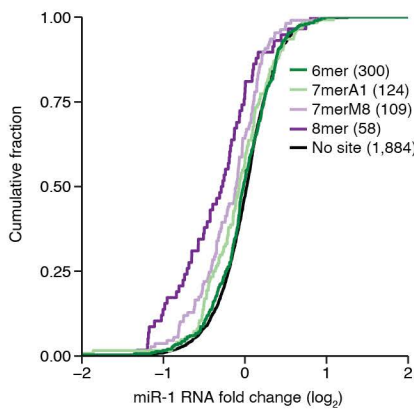
A



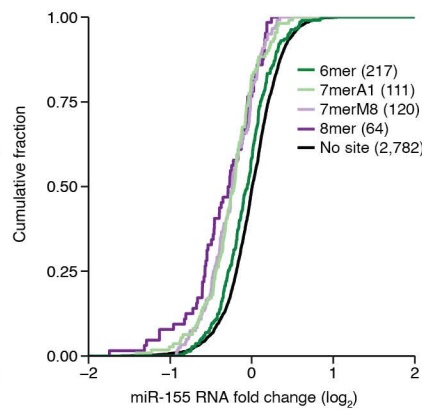
B



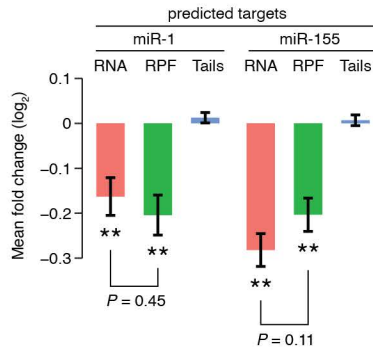
C



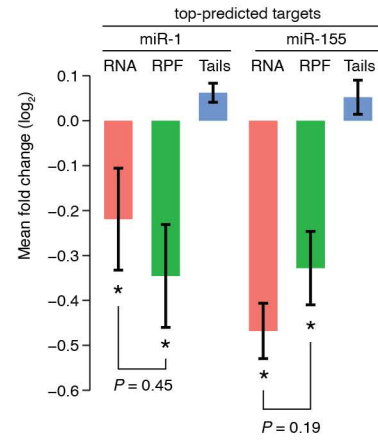
D



E



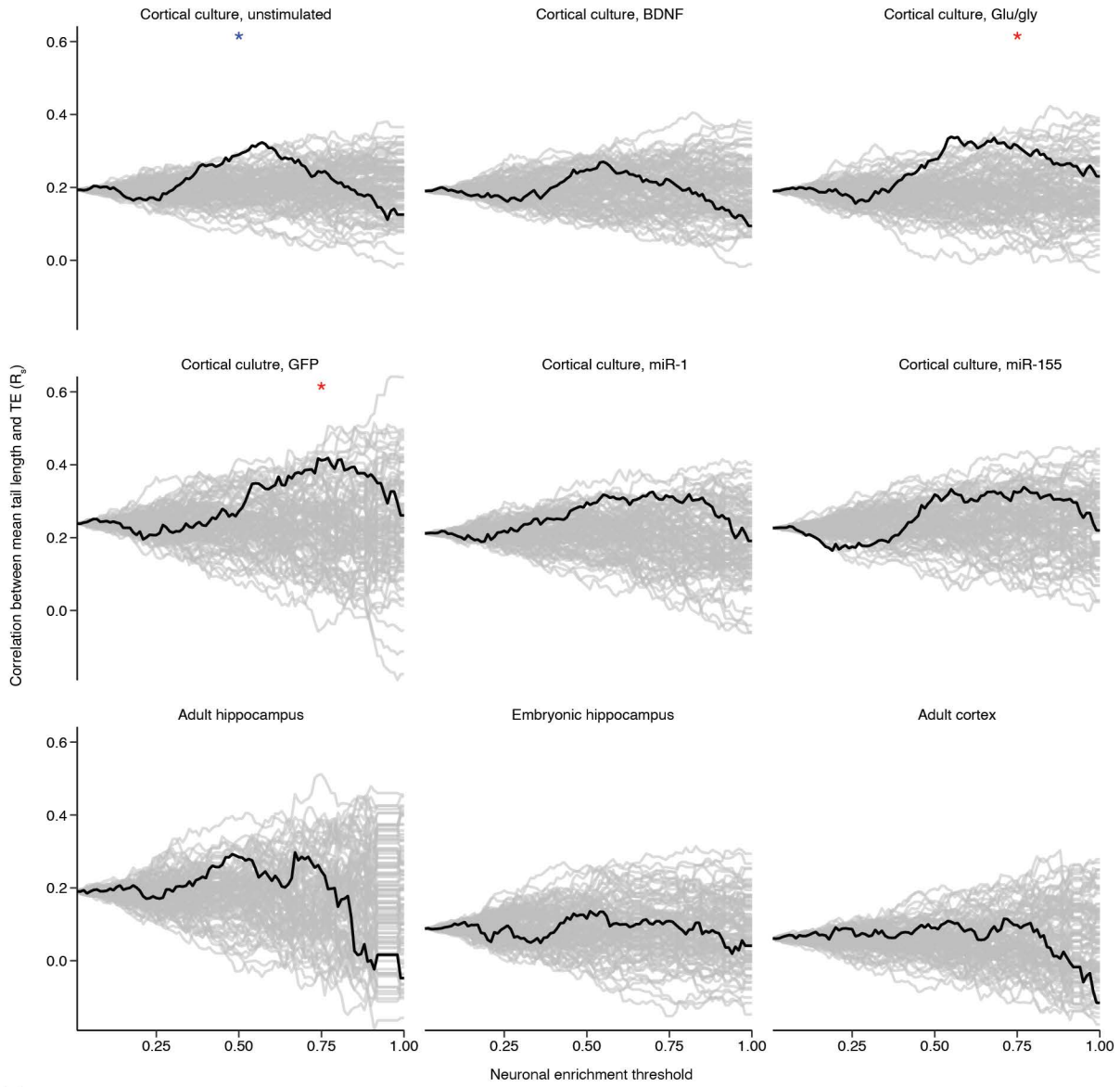
F



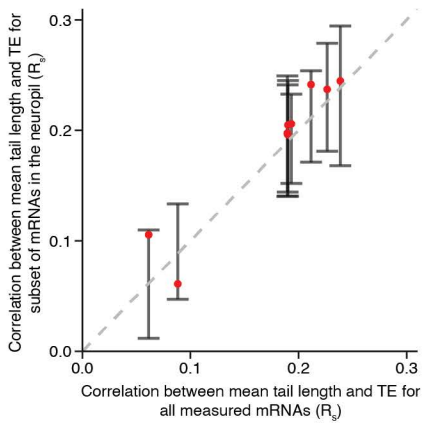
**Supplemental Figure S3.** The effects of microRNAs on their predicted targets in neurons. (A) Transduction of most neurons. Shown is a representative field of the neuronal cultures transduced with miRNA-expressing virus, imaged using brightfield (left) and fluorescence (488 nm, right) microscopy. (B) Ectopic expression of miRNAs in neurons. An RNA blot of a gel loaded with the indicated concentrations (fmol/ $\mu$ g) of synthetic miRNA standards and 1  $\mu$ g of total RNA from cultured neurons expressing either GFP only, miR-1, or miR-155 was probed for an endogenous miRNA (miR-124), then successively stripped and re-probed for miR-1 and miR-155. Quantification (in fmol/ $\mu$ g total RNA), based on a standard curve generated from the synthetic standards, is shown below each band observed from cultured neurons (including all bands for quantification of miR-124 and miR-155). (C) Repression-predicted targets of miR-1 after ectopically expressing this miRNA in neurons. Predicted miR-1 targets were classified on the basis of whether they had at least one seed-matched site of the indicated type (Bartel 2009) in their 3' UTR, and cumulative distributions of mRNA fold changes observed upon expression of miR-1 are plotted. For comparison, the distribution of fold changes observed for mRNAs that lacked a seed-matched site throughout their entire sequence is also plotted. Numbers of mRNAs in each class are shown in parentheses. The expected hierarchy of site-type efficacy was observed (Bartel 2009). (D) Repression predicted targets of miR-155 after ectopically expressing this miRNA in neurons. Otherwise, this panel is analogous to (C). (E) No detectable effect of miRNAs on TE or tail length of predicted targets enriched in the synaptic neuropil of the hippocampus. Analysis was as in Figure 3B, except it considered only mRNAs orthologous to those identified in the synaptic neuropil of rat hippocampus (Cajigas et al. 2012). Values for predicted targets (n = 109 and 78 for miR-1 and miR-155, respectively) were compared to those of mRNAs with no site to the transduced miRNA (n = 245 and 355 for no-site cohorts for miR-1 and miR-155, respectively). (F) No detectable effect of miRNAs on TE or tail length of top-predicted targets enriched in the synaptic neuropil of the hippocampus. This analysis was as in (E), except it considered the values for only the top-predicted targets (n = 15 and 7 for miR-1 and miR-155, respectively).

Figure S4

A



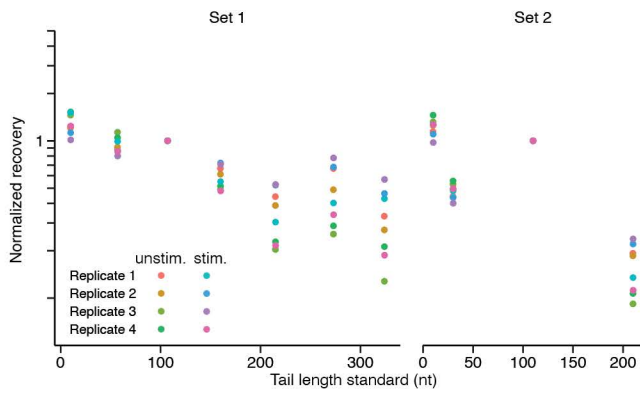
B



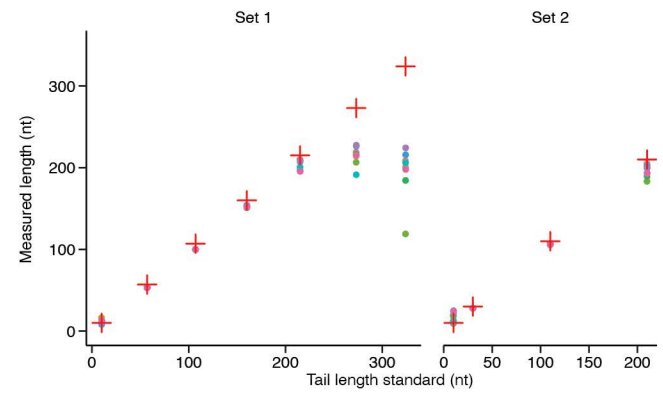
**Supplemental Figure S4.** Investigation of the extent to which neuronally enriched mRNAs exhibit stronger correlations between tail length and TE. (A) Little evidence of stronger correlations between tail length and TE when focusing on neuronally enriched mRNAs. For each of the nine samples for which tail length and TE were measured (including only the DIV14 cortical culture neurons), mRNAs were ranked by their ratio of neuronal to non-neuronal expression (Zhang et al. 2014), and these rankings were binned into 100 quantiles. Spearman correlations between tail length and TE were then plotted, using each of these quantiles as a threshold to define the set of mRNAs that were evaluated (black line), with larger threshold values indicating increasing stringency of neuronal-biased expression. For example, a value at a threshold of 0.90 indicated the Spearman correlation when considering only the 10% most neuronally enriched mRNAs. To assess the results expected by chance, correlations for 100 cohorts of randomly ordered mRNAs were also plotted for each sample (grey lines). Blue asterisks denote significance using an empirical  $p < 0.05$  based on the 50<sup>th</sup> quantile and red asterisks denote significance based on the 75<sup>th</sup> quantile. Significance at the 90<sup>th</sup> quantile was also tested but there are no samples with significance at this quantile. (B) Little improvement of correlations between tail length and TE when considering neuropil-enriched mRNAs. Plotted are the Spearman correlations (red dots) between tail length and TE for each of the 9 datasets in (A), considering either all measured mRNAs (x axis) or the subset homologous to mRNAs enriched in the synaptic neuropil of a rat (y axis) (Cajigas et al. 2012). Error bars indicate the 5<sup>th</sup> and 95<sup>th</sup> quantiles of correlations observed for 100 random samples of each dataset, each sample containing the same number of mRNAs as in the neuropil-enriched data.

Figure S5

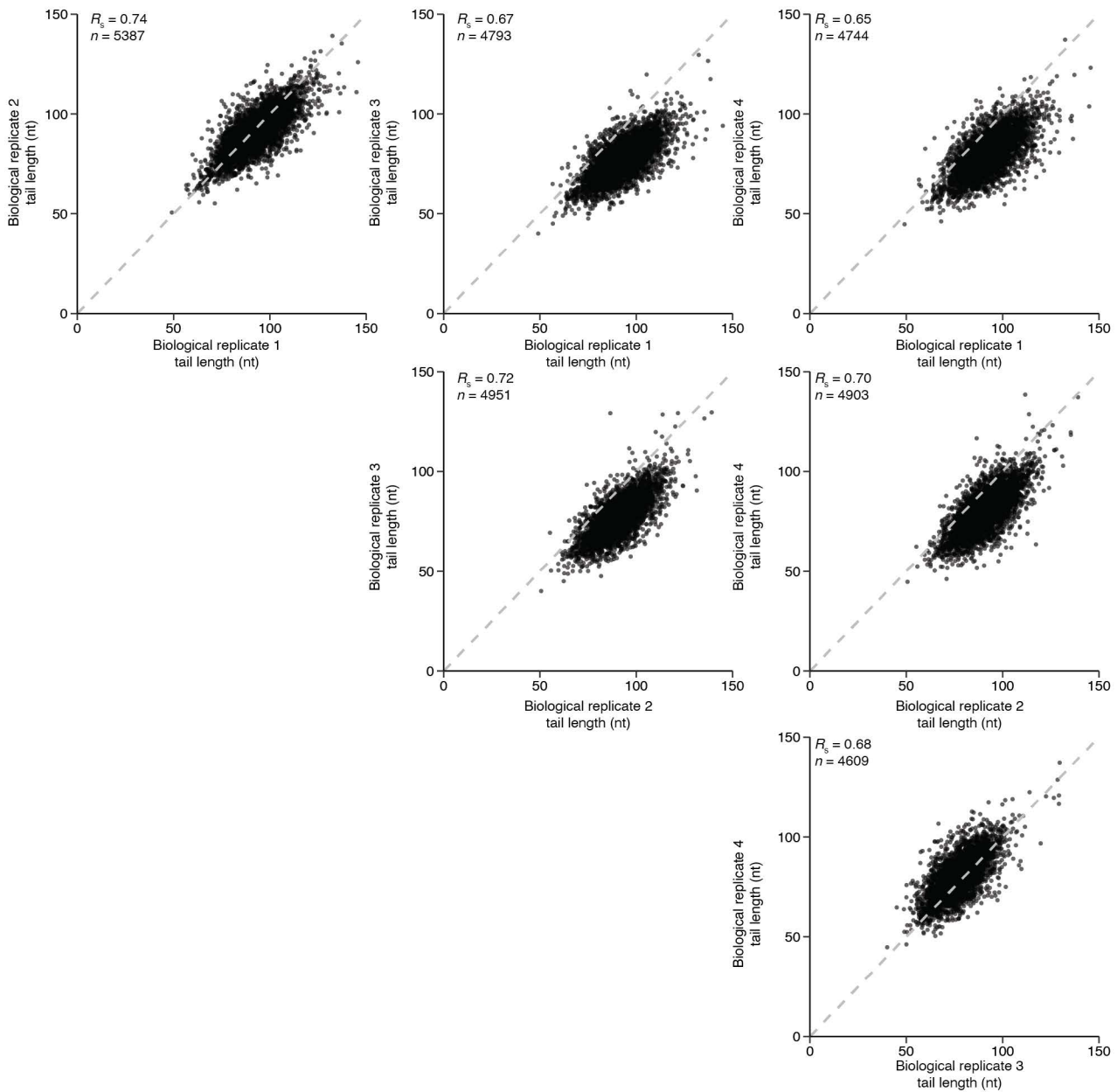
A



B

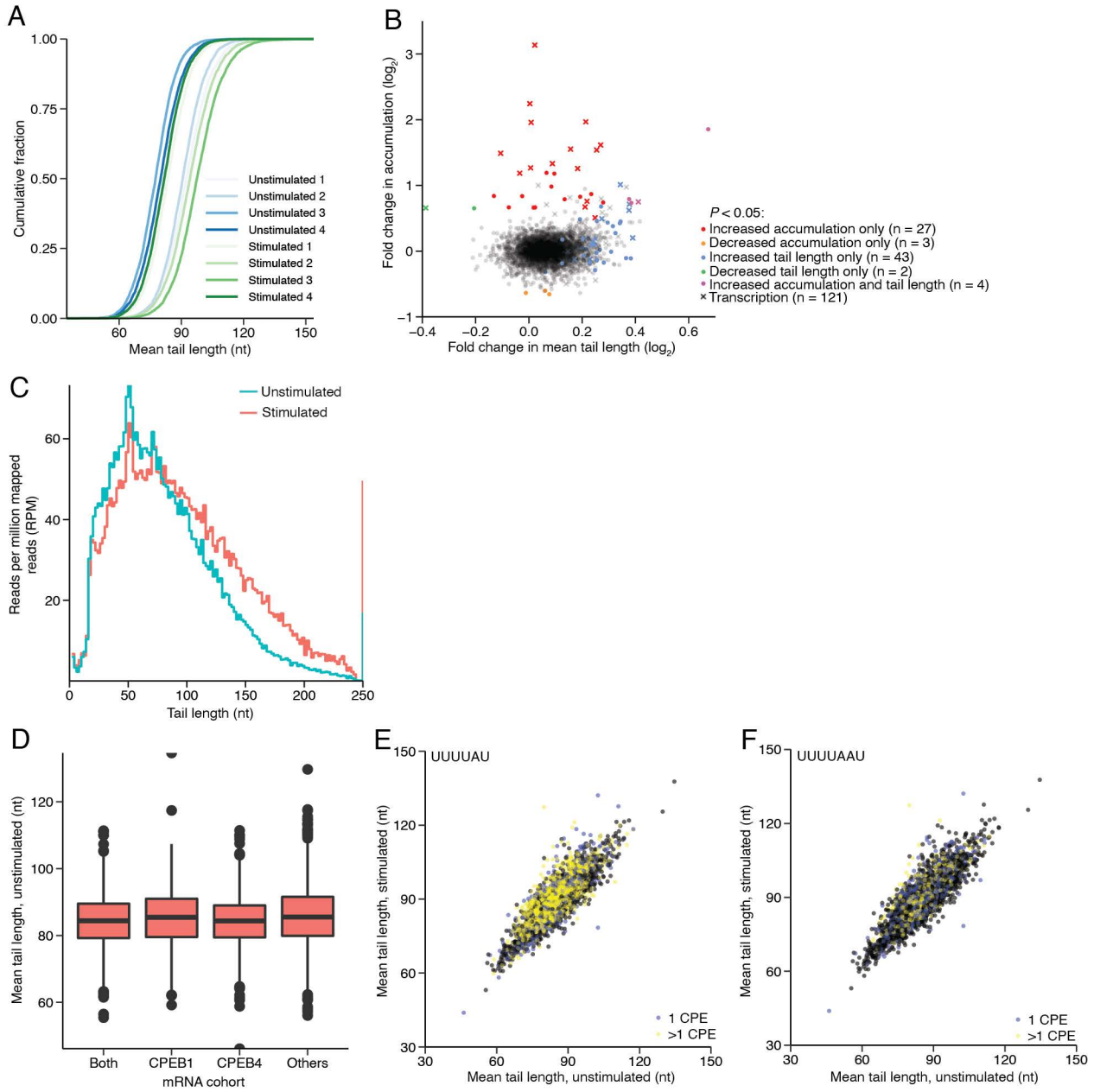


C



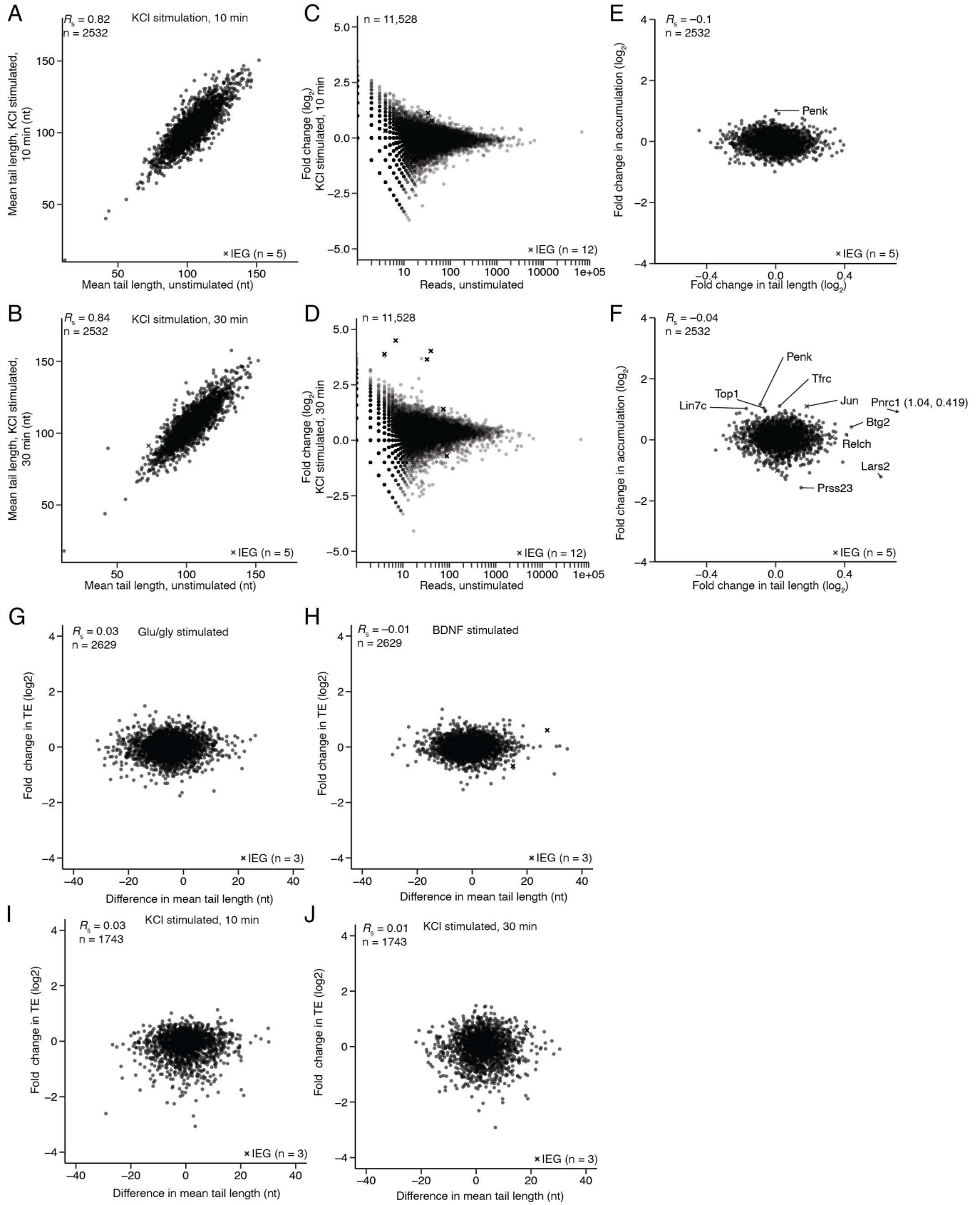
**Supplemental Figure S5.** Accuracy and reproducibility of tail-length measurements. (A) Recovery of synthetic RNA standards with tails of known lengths. Two sets of standards (Table S1) were added to total RNA prior to preparation of each PAL-seq v2 library. The tag counts corresponding to each standard in each of the eight PAL-seq datasets used in this study are plotted, after normalizing first to the ratios expected based on quantification of the original spike mixture and then to either the 107-nt standard (set 1) or the 110-nt standard (set 2). The 10-nt tail-length standards used quantification from read 2 of the PAL-seq run instead of read 1 because of very poor recovery of these barcodes in read 1. This poor recovery was attributed to the 10-nt barcodes containing a stretch of three As after a stretch of three non-As at their 3' end, and Klenow polymerase incorporating Ts across from these six nucleotides, thereby obscuring barcode calling from read 1. (B) Tail-length measurements of the synthetic standards. For each standard in each dataset, the mean tail length measured using PAL-seq is plotted as a function of the tail length measured from PAGE, with red crosses denoting expected measurements. Standards with tail lengths > 250 nt were not expected to be measured accurately because their tail length exceeded the length of the sequencing read used to measure the tail. For two samples (unstimulated replicates 1 and 3), the 10-nt standard of set 1 had insufficient reads to be included in this analysis. (C) Pairwise comparisons of tail-length measurements for biological replicates in which the mice had not been exposed to light (n = 4). Plotted are mean tail lengths for mRNAs exceeding the 50-tag cutoff in both samples. Dashed line is  $y = x$ .

Figure S6



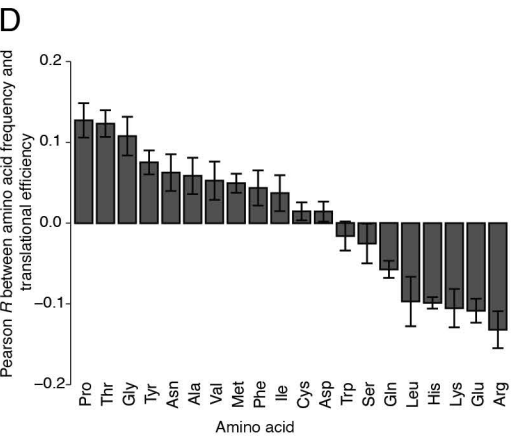
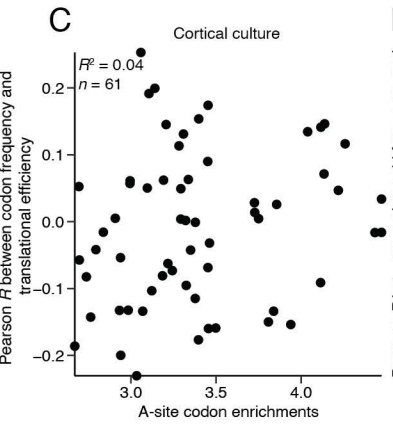
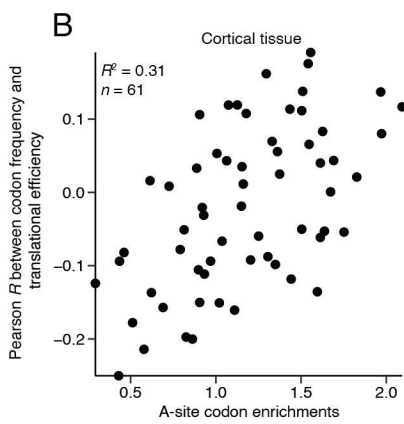
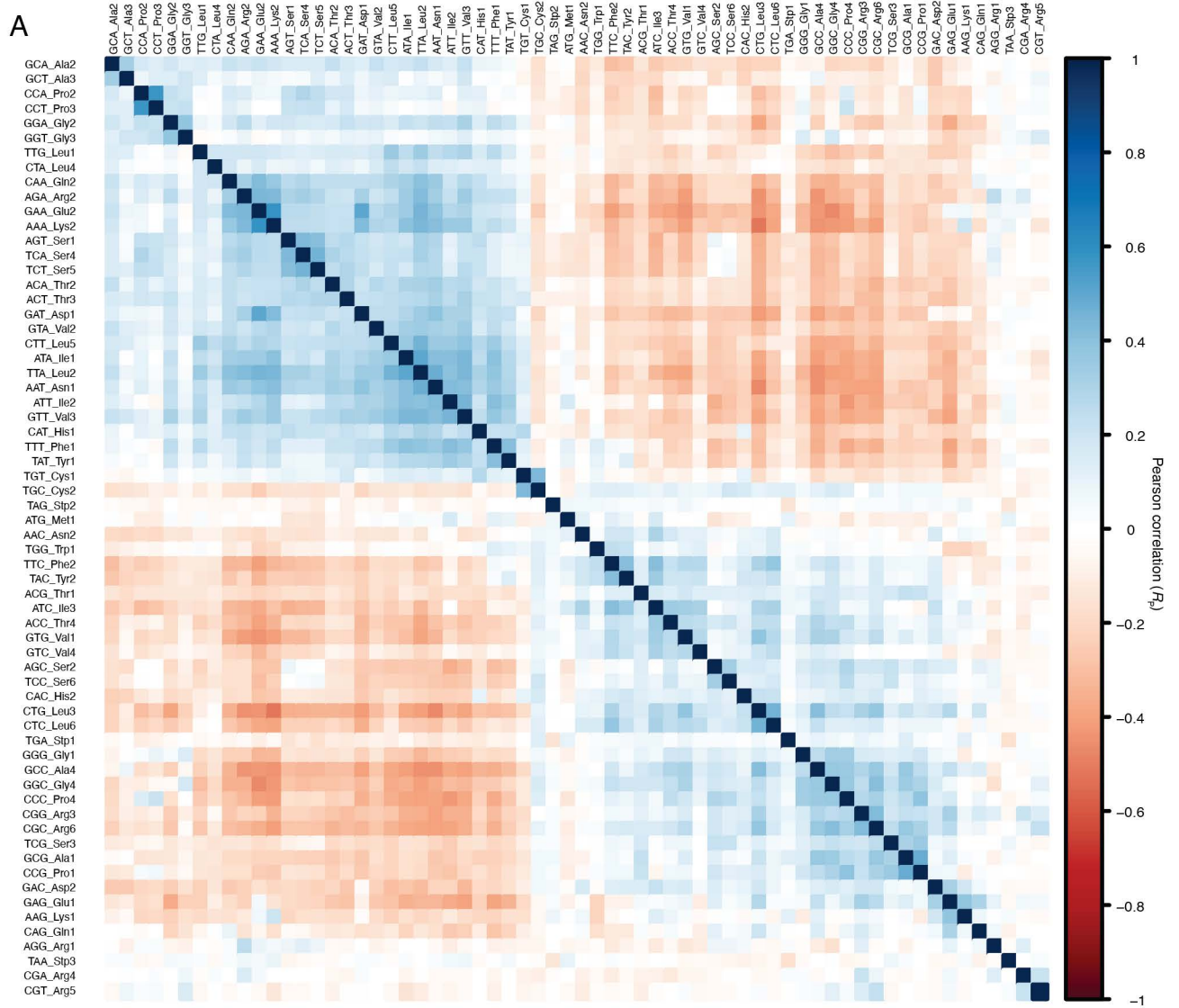
**Supplemental Figure S6.** Additional analyses of tail-length changes observed after stimulation of the visual cortex. (A) The distributions of mean tail lengths of mRNAs from visual-cortex samples. Although variability was observed between samples, no consistent difference was observed after stimulation. Each sample considered mRNAs from the same 4,321 genes. (B) The inferred influence of altered transcription on mRNA accumulation. Analysis was as in 5D, except crosses denote mRNAs with significant changes in transcription, as measured by single-nuclei RNA-seq (Hrvatin et al. 2018). mRNAs marked as undergoing transcriptional changes are those that were reported to change significantly in any cell type of the visual cortex 1 h after light stimulation. (C) Tail-length distributions of the subset of mRNAs that had increased mean tail lengths following stimulation but lacked significant changes in abundance. Shown are histograms combining results for the 43 mRNAs classified as having “increased tail length only” in Figure 5D. Data are plotted in 2-nt bins, with the y axis indicating reads per million reads mapped (RPM). (D) No detectable relationship between mRNA tail length and CPEB binding. Shown are distributions of mean tail lengths for mRNAs reported to be bound by either CPEB1 (n = 350), CPEB4 (n = 738), both CPEB1 and CPEB4 (n = 738), or neither protein (others, n = 2,486), as indicated by RNA immunoprecipitation followed by sequencing (RIP-seq) from the mouse striatum (Parras et al. 2018) (line, median; box, 25<sup>th</sup> and 75<sup>th</sup> percentiles; whiskers, 1.5x interquartile range; data beyond whiskers plotted as points). Shown are results from analysis of mean tail lengths from the four unstimulated datasets. (E) No indication that mRNAs containing CPE motifs were differentially polyadenylated in response to light stimulation. Mean tail lengths for the stimulated cohort are compared to those for the unstimulated cohort (n = 4 mice per cohort). Points for mRNAs with one CPE motif, defined as UUUUUAU in their 3' UTR (Charlesworth et al. 2013; Reyes and Ross 2016) are in blue. Points for mRNAs with more than one CPE motif are in yellow. In cases where an mRNA had multiple 3' UTR isoforms, the maximum number of CPE motifs for any isoform was used. (F) The same analysis as in (E), except defining the CPE motif as UUUUAAU.

Figure S7



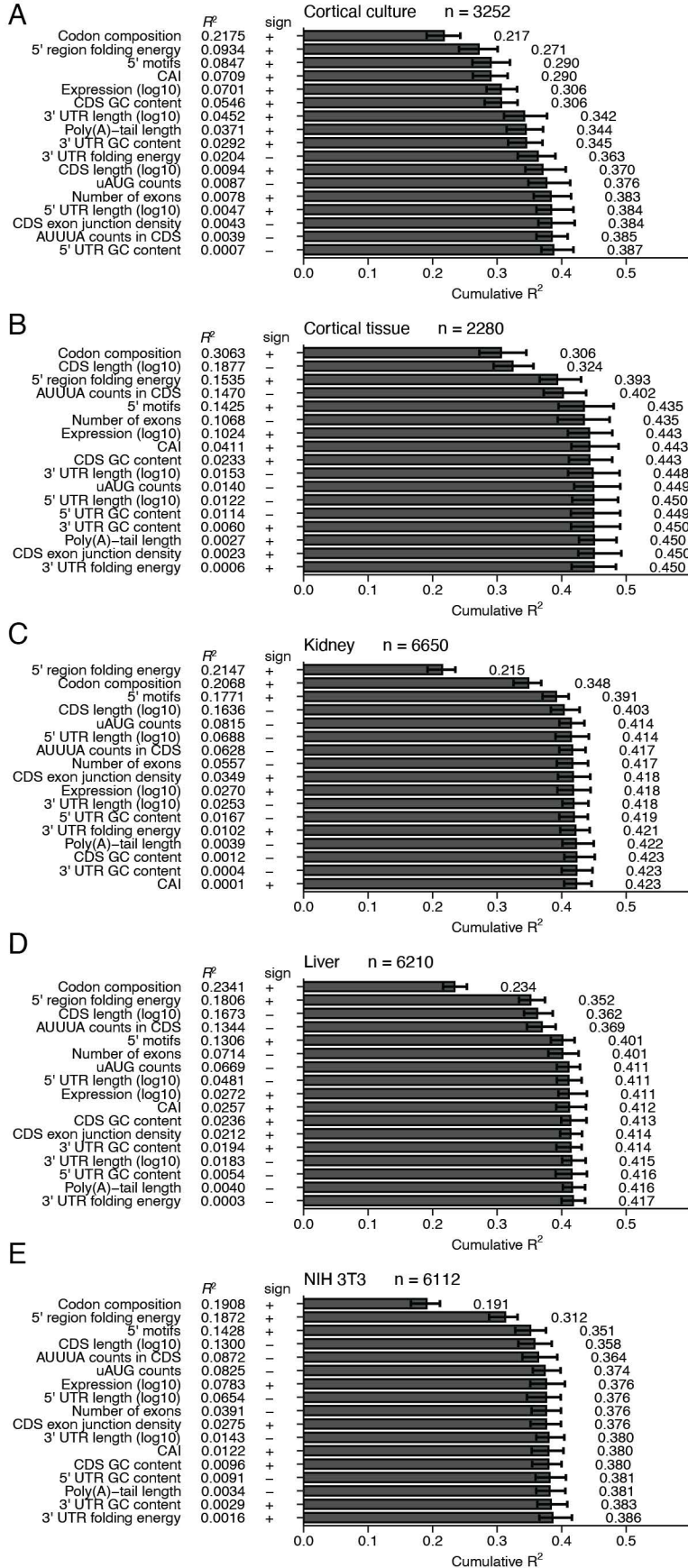
**Supplemental Figure S7.** (A) The effect of KCl stimulation for 10 min on mean tail lengths. For mRNAs of each gene that passed the 50-tag expression cutoff, mean tail length in the stimulated sample was compared to that in the untreated control. Values for IEGs are indicated with an “x”. (B) The effect of KCl stimulation for 30 min on mean tail lengths; otherwise, as in (A). (C) The effect of KCl stimulation for 10 min on gene expression, as measured by PAL-seq tags. For mRNAs of each gene, the number of tags in the stimulated samples was compared to that in the untreated control. (D) The effect of KCl stimulation for 30 min on gene expression; otherwise as in (C). (E) Comparison of change in expression with change in tail length observed after KCl stimulation for 10 min. Gene labels were added for all points with a  $\log_2$  fold change in expression of  $> 1$  or  $< -1.5$ , or  $\log_2$  fold change in tail length of  $> 0.4$ . (F) Comparison of change in expression with change in tail length observed during KCl stimulation for 30 min, otherwise, as in (E). (G–J) Comparison of changes in TE and differences in mean tail length observed during glu/gly (G), BDNF (H), 10 min of KCl (I), or 30 min of KCl (J) stimulation. Values for IEGs are indicated with an “x”.

Figure S8



**Supplemental Figure S8.** Codon composition and a relationship between amino-acid composition and TE observed in cortical cultures. (A) Codon composition across mRNAs expressed in DIV14 primary cortical culture neurons. For each of the 61 coding codons, its frequency in mRNAs that passed cutoffs for expression ( $n = 5,128$ ) was determined. The Pearson correlation between these frequencies was calculated for every pair of codons, and the resulting data were clustered using Euclidean distance. Positive values indicate that the two codons were more likely to appear in the same mRNA, whereas negative values indicate that the two codons were more likely to be mutually exclusive. (B) The relationship between inferred translocation rates and association between codons and TE for cortical tissue. 30-nt RPFs were used to infer the codon present in the ribosomal A site, excluding any read that mapped to the first 50 nt of the ORF or to non-coding regions. Any read with a 5' position that mapped out of the reading frame was also discarded. The RPF counts for each A-site codon were then normalized to the corresponding RNA-seq reads. The resulting log<sub>2</sub> enrichment values are plotted on the x axis, comparing these values to the Pearson correlation between the codon and TE (y axis, as in Figure 7). (C) As in (B), except plotting the A-site enrichments and TE–codon correlations for the DIV14 unstimulated primary cortical culture. (D) The relationship between amino-acid composition and TE observed in cortical cultures. The TE and amino-acid frequency were compared for all measured transcripts in each of the six primary-culture samples. The resulting Pearson R value is plotted as a bar plot, with error bars denoting the standard deviation observed between different samples. Correlations are ranked by the mean R value. Otherwise, this panel is as in Figure 7.

Figure S9



**Supplemental Figure S9.** Linear models quantifying correlates of translation using all features considered in this study. (A) Model for the primary cortical culture sample used in Figure 8 built without discarding any features. Features are arranged by decreasing order of the  $R^2$  between that feature and TE, rather than by AIC ordering. Otherwise, this panel is as in Figure 8. (B), (C), (D), and (E) Models for adult cortex, kidney (Castelo-Szekely et al. 2017), liver (Janich et al. 2015), and NIH-3T3 cells (Subtelny et al. 2014), respectively, generated and displayed as in (A).