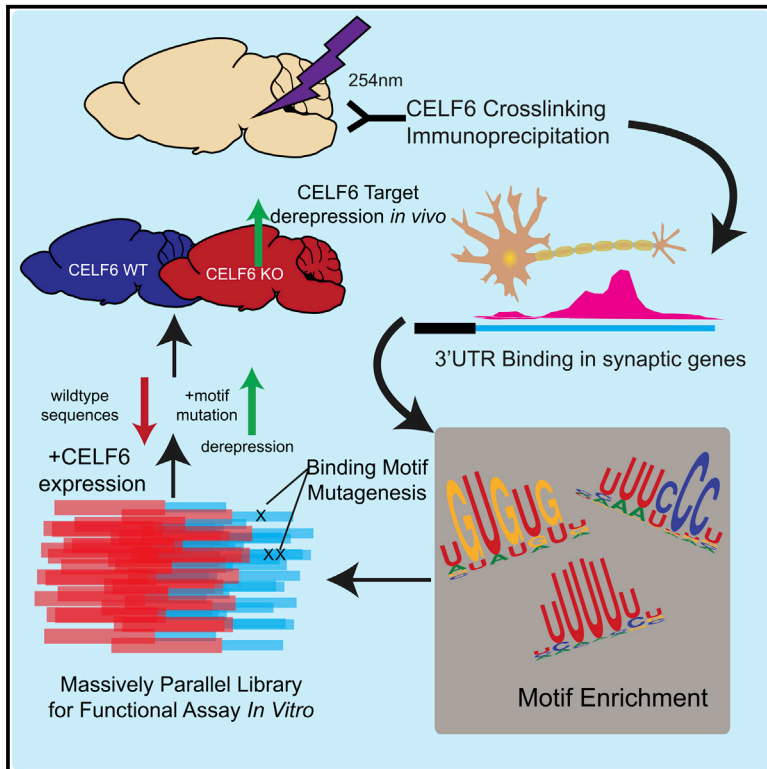


# CLIP and Massively Parallel Functional Analysis of CELF6 Reveal a Role in Destabilizing Synaptic Gene mRNAs through Interaction with 3' UTR Elements

## Graphical Abstract



## Authors

Michael A. Rieger, Dana M. King, Haley Crosby, Yating Liu, Barak A. Cohen, Joseph D. Dougherty

## Correspondence

jdougherty@wustl.edu

## In Brief

Rieger et al. assay the function of the RNA-binding protein CELF6 by defining its targets in the brain. They show that CELF6 largely binds 3' UTRs of synaptic mRNAs. Using a massively parallel reporter assay, they further show that CELF6 and other CELFs are associated with lower mRNA abundance and that targets are derepressed in *Celf6*-knockout mice *in vivo*.

## Highlights

- CELF6 primarily associates with 3' UTRs of synaptic genes in the mouse brain
- CELF6:sequence interaction is assayed using a massively parallel reporter assay
- CELF3–6 all result in lower mRNA levels with few changes to translation efficiency
- CELF6 targets are derepressed *in vivo* in *Celf6*-knockout mice



## Article

# CLIP and Massively Parallel Functional Analysis of CELF6 Reveal a Role in Destabilizing Synaptic Gene mRNAs through Interaction with 3' UTR Elements

Michael A. Rieger,<sup>1,2,3</sup> Dana M. King,<sup>1</sup> Haley Crosby,<sup>1,2</sup> Yating Liu,<sup>1,2</sup> Barak A. Cohen,<sup>1</sup> and Joseph D. Dougherty<sup>1,2,4,\*</sup><sup>1</sup>Department of Genetics, Washington University School of Medicine, St. Louis, MO 63110, USA<sup>2</sup>Department of Psychiatry, Washington University School of Medicine, St. Louis, MO 63110, USA<sup>3</sup>Present address: Salk Institute for Biological Studies, La Jolla, CA, USA<sup>4</sup>Lead Contact\*Correspondence: [jdougherty@wustl.edu](mailto:jdougherty@wustl.edu)<https://doi.org/10.1016/j.celrep.2020.108531>**SUMMARY**

CELF6 is a CELF-RNA-binding protein, and thus part of a protein family with roles in human disease; however, its mRNA targets in the brain are largely unknown. Using cross-linking immunoprecipitation and sequencing (CLIP-seq), we define its CNS targets, which are enriched for 3' UTRs in synaptic protein-coding genes. Using a massively parallel reporter assay framework, we test the consequence of CELF6 expression on target sequences, with and without mutating putative binding motifs. Where CELF6 exerts an effect on sequences, it is largely to decrease RNA abundance, which is reversed by mutating UGU-rich motifs. This is also the case for CELF3–5, with a protein-dependent effect on magnitude. Finally, we demonstrate that targets are derepressed in CELF6-mutant mice, and at least two key CNS proteins, FOS and FGF13, show altered protein expression levels and localization. Our work finds, in addition to previously identified roles in splicing, that CELF6 is associated with repression of its CNS targets via the 3' UTR *in vivo*.

**INTRODUCTION**

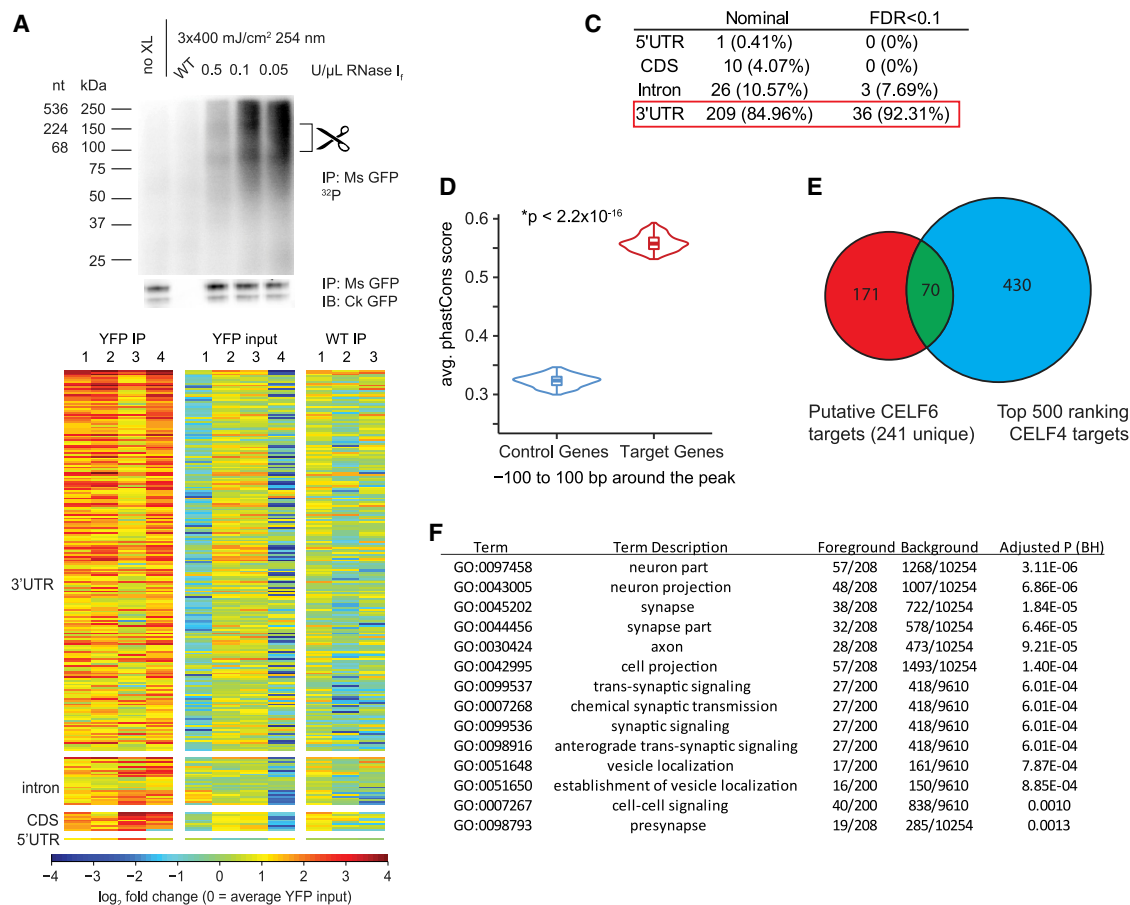
RNA-binding proteins (RBPs) regulate the life cycle of messenger RNAs (mRNA), from transcription, splicing, and nuclear export to localization, maintenance, translation, and degradation (Vindry et al., 2014; Xiong et al., 2015). The CUGBP and ELAV-like factor (CELF) family of RBP has six members (CELF1–6), which can be divided into two groups based on sequence homology: CELF1–2, which are expressed ubiquitously, and CELF3–6, which show enrichment in the CNS (Dasgupta and Ladd, 2012). The most studied, CELF1, was first characterized in relation to myotonic muscular dystrophy pathogenesis (Lee and Cooper, 2009). It binds CUG- and (U)GU-rich motifs (Timchenko et al., 1996; Wang et al., 2015) and has been shown to promote exon skipping (Wang et al., 2015) as well as mRNA degradation via recruitment of deadenylases (Moraes et al., 2006). CELF3–6, however, have not been as well characterized.

We identified CELF6 to be both enriched in serotonin-producing neurons and disrupted in an individual with autism (Dougherty et al., 2013). CELF6 showed expression in the hypothalamus and in several monoaminergic cell populations commonly targeted in psychiatry (Maloney et al., 2016). Likely as a consequence of its role in those cells, it is also associated with conditioned behaviors: *Celf6*-knockout (KO) mouse pups show altered ultrasonic vocalization, while adults exhibit deficits in fear conditioning and loss of conditioned place preference to cocaine (Dougherty et al., 2013; Maloney et al., 2019). Although capable

of regulating splicing *in vitro* (Ladd et al., 2004), CELF6 has recently been shown to interact with the 3' untranslated region (UTR) as well, allowing it to stabilize p21 mRNA in colorectal cancer cells (Liu et al., 2019). Its targets and function in the CNS, however, are not well described. Here, we sought to explore the molecular role of CELF6 in the brain to better understand how its disruption may alter behavior.

To define the function of CELF6 *in vivo*, we performed cross-linking immunoprecipitation and sequencing (CLIP-seq) in the brain and found CELF6 to be overwhelmingly associated with 3' UTRs of target mRNAs, many coding for proteins involved in synaptic transmission. UTRs showed enrichment for UGU-containing motifs, consistent with prior *in vitro* findings (Ray et al., 2013), as well as other motifs. We cloned UTR elements under CLIP-seq peaks into the 3' UTR of a reporter construct and measured reporter library expression and abundance on translating ribosomes while modifying CELF6 expression and sequence integrity. We found CELF6-associated UTR sequences were repressive in general, CELF6 could function to enhance this reduction to mRNA levels, and this effect could be abrogated by motif mutation. We also observed this effect across CELF3–6. In the *Celf6*-KO mouse brain, we find CELF6 targets to be generally derepressed, validating our findings *in vivo*. Taken together, we show that CELF6 is largely associated with reductions to transcript abundance of its neuronal targets and thus may have an important role regulating cellular functions in the brain.





**Figure 1. CELF6 Primarily Associates with 3' UTRs of Target mRNAs In Vivo**

(A) CELF6 CLIP, lanes: no crosslink (no XL), wild-type (WT) control, and three concentrations of RNase I. Top: RNA <sup>32</sup>P autoradiogram. Bottom: anti-GFP immunoblot. Scissors mark region (80–150 kDa) above CELF6-HA/YFP (78 kDa) isolated for sequencing. Immunoblot detects two bands, corresponding to sizes of known isoforms of CELF6.

(B) Log<sub>2</sub> counts-per-million (CPM) RNA in four CELF6-HA/YFP<sup>+</sup> replicates and three WT replicates differentially abundant regions. Heatmaps show enrichment in HA/YFP<sup>+</sup> immunoprecipitate (IP) samples relative to input and WT controls. CPM is normalized to mean YFP input by row.

(C) Summary of differential enrichment analysis of genes showing nominally significant ( $p < 0.05$ ), and Benjamini-Hochberg (FDR  $< 0.1$ ) enrichment in HA/YFP<sup>+</sup> IP samples relative to both HA/YFP<sup>+</sup> input and WT IP controls. No genes were found in intergenic regions.

(D) Average phastCons score for  $\pm 100$  bp around CLIP peaks in targets compared with control genes not exhibiting enrichment in CLIP ( $p < 2.2 \times 10^{-16}$ , two-tailed t test).

(E) Overlap of CELF6 and CELF4 targets identified in Wagnon et al. (2012).

(F) DAVID gene ontology analysis of CELF6 CLIP targets (241 unique genes).

## RESULTS

### CELF6 Associates with Conserved Elements in 3' UTRs of Target mRNAs In Vivo

We first sought to use CLIP to identify CELF6 targets *in vivo*. Unfortunately, available antibodies against CELF6 protein do not function well for immunoprecipitation against endogenous protein (data not shown). We previously developed a BAC transgenic mouse expressing epitope-tagged CELF6-YFP/HA (78 kDa), which exhibits endogenous patterns of anatomical and subcellular (cytoplasmic and nuclear) expression, with only slight overexpression compared with native CELF6 (Maloney et al., 2016). Because we have previously validated antibodies

for GFP (and related YFP) pull-down (Doyle et al., 2008), we designed our CLIP studies using that line.

To establish the *in vivo* RNA-binding properties of CELF6, we first performed CLIP with anti-EGFP antibodies on CELF6-YFP/hemagglutinin (HA) mice followed by radiolabeling of nucleic acid (Figure 1A). Controls included an immunoprecipitate (IP) from uncrosslinked samples and IP from wild-type (WT) tissue. As expected, there was a lack of detectable RNA in IP from WT tissue and uncrosslinked YFP<sup>+</sup> tissue. To capture the targets of CELF6 *in vivo*, we purified a region of 60–200 nucleotides in size (80–150 kDa) from lysates of CELF6-YFP/HA<sup>+</sup> brains. Similar to the Vidaki et al. (2017) study of Mena, we found that the stringent IP conditions of standard CLIP protocols were incompatible

with our anti-GFP IP (not shown). Thus, to enable statistical background estimation, we also collected samples from size-selected input RNA (controlling for differences in RNA abundance), as well as pull-down from WT littermate brains (to control for non-specific binding). We defined high-confidence targets as those showing statistical enrichment in CLIP samples compared with both controls and we defined lower-confidence targets as showing enrichment against only one control.

We prepared sequencing libraries from those samples using an adaptation of the eCLIP workflow (Van Nostrand et al., 2017) (Methods S1). All samples were sequenced to a similar depth, and, on average, 94% of reads mapped to genic regions (Table S1). Next, to define specific sites of CELF6 binding, we called peaks genome-wide using Piranha (Uren et al., 2012) and performed differential enrichment analysis using edgeR comparing CLIP to controls for read counts under peaks (Robinson et al., 2010). In the event that genome-wide background count estimation in Piranha would limit discovery, we also summed reads mapping to smaller subgenic features: 5' UTR, coding sequence (CDS), intron, or 3' UTR and performed differential enrichment analysis restricted to those regions. In total, we identified significant enrichment across 241 genes combined across analysis methods (Table S2).

CELF proteins have been shown to function both in splicing and post-transcriptional regulation. To understand how CELF6 may function in the brain, we examined the distribution of CLIP reads across genes. We hypothesized that splicing-related functions would correspond to increased density in internal coding exons and introns, whereas post-transcriptional regulation would be reflected as increased density in UTRs. Figure 1B shows a heatmap of CLIP targets across all samples relative to controls. Most enriched regions are in 3' UTRs (Figure 1C; >80%). Additionally, UTRs showing CELF6 binding were more conserved than expected by chance (Figure 1D). We also compared our CLIP hits to the top-ranking targets of the related protein CELF4 (Wagnon et al., 2012). Figure 1E shows this overlap—with 70/241 CELF6 target genes in common with CELF4 target genes (also noted in Table S2). The Wagnon et al. (2012) study was performed on different brain regions, cell types, and at different ages, and thus, this overlap is notable. CELF6 targets were also disproportionately involved in synaptic transmission (Figure 1F). Taken together, these data suggest CELF6 may regulate neuronal function by binding UTRs of synaptic gene transcripts, altering their stability or translation.

### CELF6-Bound 3' UTRs Are Enriched for U-, UGU-, and CU-Containing Motifs

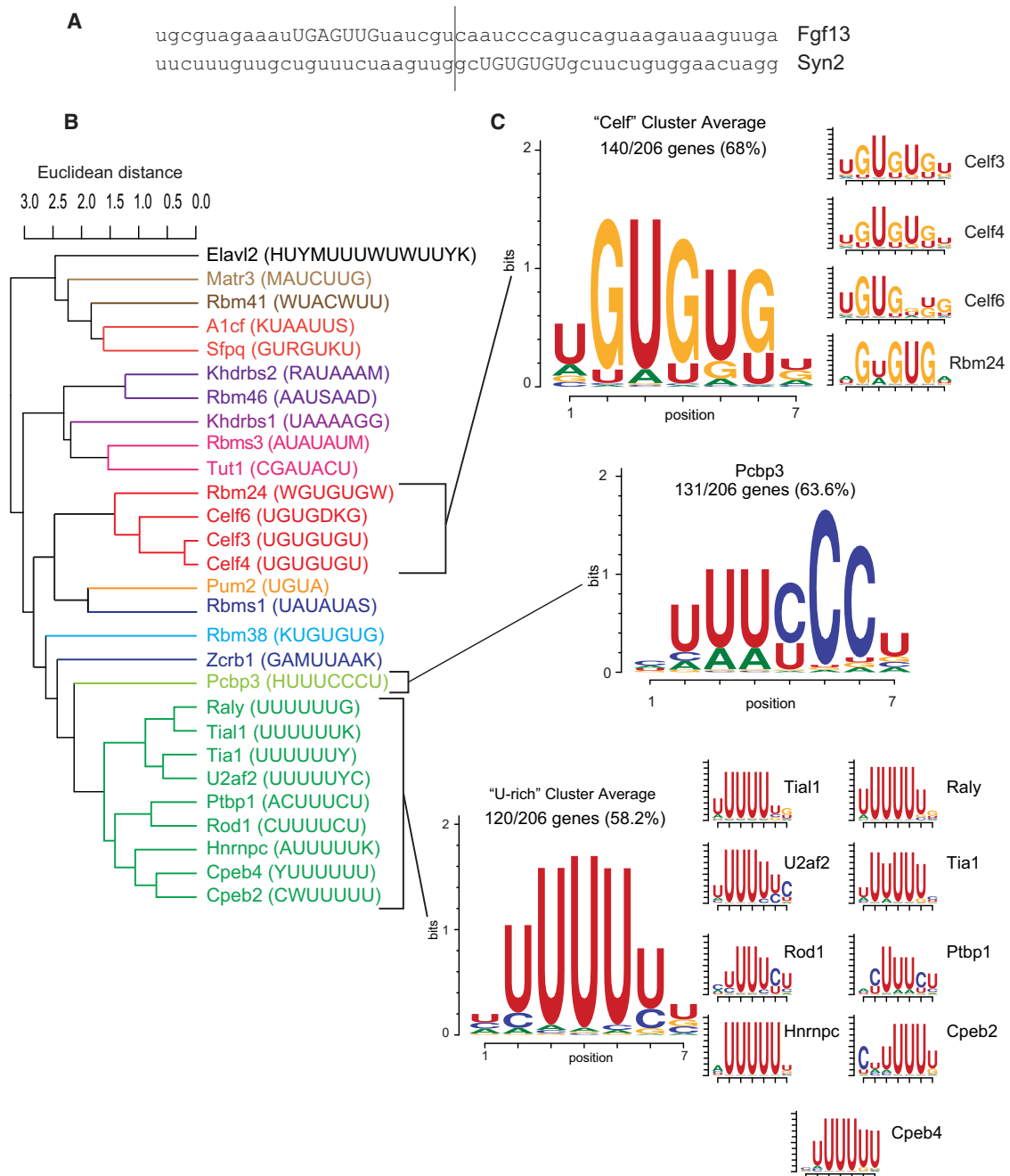
Previous study *in vitro* has identified CELFs as preferring UGU-rich sequences (RNACompete; Ray et al., 2013) and manual inspection under CELF6 clip peaks found matches to those motifs (Figure 2A). To scan CLIP targets systematically and potentially identify other preferred sequences in the brain, we used MEME-SUITE to search for motif enrichment. We analyzed 50 nucleotide segments centered under 544 peaks. Those included 25 3' UTR peaks (25 unique genes) identified using Piranha genome wide. However, we performed a second pass of peak calling in those UTRs identified via summation across the UTR (as described in the preceding section). Doing so enabled us to

discover an additional 519 peaks in 191 UTRs, including new peaks in 10 genes identified in the initial genome-wide pass. We suspect the additional discovery is due to incorrect background estimation when using read counts across an entire gene or genome wide versus more local background-count estimation. UTRs varied between 1 and 9 peaks with a median of 2, with 75% of the UTRs containing 1–3 peaks. We compared those sequences to 544 control sequences sampled randomly from UTRs of brain-expressed genes showing no CLIP enrichment.

We found 29 enriched motifs at a false discovery rate (FDR) < 0.05, including the RNACompete binding motifs for CELFs. We also found U-rich motifs of the TIA/CPEB families to be enriched, among others. Because RBPs can exhibit degenerate preferences (Auweter et al., 2006; Harvey et al., 2018), for a more holistic summary, we clustered enriched motifs by the Euclidean distance between their position probability matrices (PPMs) in the CISBP-RNA database. As expected, enriched motifs clustered partially by RBP family (Figure 2B), and motifs within a cluster were highly similar. After clustering, we scanned our sequences again to determine which showed matches to the cluster-average PPMs, as “meta-motifs.” We found that 519/544 sequences showed significant matches to at least one cluster (Tables S3 and S4).

A few of these clusters were highly recurrent (Figure 2C; Tables S3 and S4); 140/206 unique genes' 3' UTRs (68%) showed at least one match to the “CELF”-cluster ([U/A]GUGU[G/U][UGA]), and 131/206 UTRs (63.6%) showed at least one match to the PCBP3 motif, which forms its own cluster with a central UUU[C/U]CC sequence. PCBP3 binds both double- and single-stranded nucleic acids, known primarily as a transcription factor (Choi et al., 2009), though a related protein (PCBP4) has also been shown to regulate mRNA stability (Scoumanne et al., 2011). 120/206 (58.2%) genes showed at least one match to the “U-rich” cluster, whose members all possess a central stretch of 4–5 Us. Proteins in this cluster include TIA1, involved in stress granule localization (Gilks et al., 2004); CPEBs, involved in polyadenylation (Vilalba et al., 2011); and HNRNPC, involved in both mRNA stability and localization (Nakielny and Dreyfuss, 1996; Shetty, 2005). Moreover, 61/206 genes (29.6%) had  $\geq 1$  match to both the CELF cluster and the U-rich cluster, and 65/206 genes (31.5%) had  $\geq 1$  match to both the CELF and PCBP3 clusters. Thus, CELF6 binding appears to be associated with combinations of known motifs for both CELFs and other RBPs, and not just UGU-rich sequences.

We also tested whether CELF6 peaks had location biases within UTRs that might suggest function, such as an enrichment near a polyadenylation signal (PAS). We found a small increase in the number of predicted PASs per transcript compared with length-matched UTRs from control genes ( $p = 0.00023$ ; Figure S1C), as well as decreased distance between CLIP peaks and a PAS ( $p = 5.3E-6$ ; Figure S1D). In the case of UTRs with only one predicted PAS, this was true only for the PAS [ATATAA], but not [AATAAA] ( $p = 0.00154$ ; Figure S1E). Although these findings suggest that CELF6's proximity may influence choice of PAS, an examination of alternative polyadenylation *in vivo* in *Celf6*-KO mice did not show CELF6 to alter the probability of alternative 3' UTR usage (not shown, available at GEO: GSE160293).



**Figure 2. CELF6-Associated 3' UTRs Are Enriched for U-, UG-, and CU-Containing Motifs**

(A) Example 50-nt regions under CLIP peaks.

(B) Clustering of enriched motifs identified with MEME-SUITE. Color coding shows cluster membership.

(C) Motif logos for the clusters found under the CLIP peak. Logos represent the average of position probability matrices (PPMs) for each motif cluster. Side panels show individual PPMs making up each cluster's membership.

### Massively Parallel Reporter Assay Defines the Function of CELF6-Bound Motifs

Motif analysis indicates that sequence specificity mediates interactions between CELF6 and RNA *in vivo*, so we next sought to understand the consequence of that association, such as on

translation or transcript stability. We used a massively parallel reporter assay framework for post-transcriptional regulatory element sequencing (PTRE-seq) to evaluate a large number of target UTR sequences at once (Cottrell et al., 2018). For PTRE-seq, we sub-cloned 410 independent, CLIP-defined

UTR elements, each 120 bp long and centered under CLIP peaks, into the 3' UTR of a tdTomato expression plasmid (Figure 3A). That library included 254 high-confidence targets showing enrichment against both CLIP controls (input and WT/non-specific pull-down). We also included cloneable elements from 156 additional elements under peaks that only met the input comparison criterion (lower stringency, highlighted in green dots going forward). As these performed similarly, we included both sets in subsequent figures. For reproducibility, each UTR element was included in the library six times, with unique 9-bp barcodes. As a comparison group, we also included 410 elements with all significant motif matches mutated. Sequences and their mutations are shown in Table S5.

We assayed the effect of each element on both transcript abundance and final translation levels, as assessed by ribosome occupancy of those reporters. We used transient cotransfection of the library with an EGFP/RPL10A construct that tags ribosomes with GFP enabling translating ribosome affinity purification (TRAP). This method allows assessment of RNA abundance on ribosomes by anti-GFP pull-down and has previously been shown to be sensitive to UTR elements that regulate translation (Heiman et al., 2008). To assess transcript abundance, we also collected total RNA from the same cells. We opted for an *in vitro* culture system with largely undetectable levels of CELF6 expression: SH-SY5Y neuroblastoma cells (Figure 3B), so that we could control CELF6 expression exogenously, confirmed by RT-PCR (Figure S2).

We assessed expression of PIRE-seq library sequences by counting barcodes in sequencing libraries prepared from transient transfection replicates. Pearson correlation in barcode counts between replicates was >0.9, indicating good reproducibility. To account for any differences in starting abundance within the library itself, log<sub>2</sub> counts per million (CPM) in RNA barcodes were further normalized to log<sub>2</sub> CPM from sequencing the starting DNA plasmid pool (normalized count data hereafter termed “log<sub>2</sub> expression”). After removing barcodes that were absent or poorly measured across all samples, the final analyzed library contained 379 UTR element pairs (reference and mutant, 229 high-confidence targets and 150 lower-stringency targets); all of which were represented by three to six barcodes per element, across 153 total genes (104 high-confidence genes and 49 lower-stringency genes).

To determine the relationship between CELF6 expression and element sequence, we analyzed log<sub>2</sub> TRAP levels using a 2 × 2 factorial design linear mixed-effects model, fitting fixed effects of the element sequence (“reference” or “mutant”) and the CELF6 expression condition (control [CTL] or CELF6 expression) and the interaction of the condition X sequence, with a random intercept term for replicate sample, treating an element’s barcodes within a sample as a repeated measure. Individual models were fitted for each UTR element pair. A summary of effects across all library elements and estimates of R<sup>2</sup> are shown in Table S6. At nominal p < 0.05, 293/379 (77.3%) of elements showed some significant effect (main effects of either sequence or CELF6 expression or a sequence-X CELF6-expression interaction, 283/379 [74.7%] at FDR < 0.1).

### CELF6-Associated Motif Sequences Decrease RNA Abundance and Corresponding Ribosome Occupancy

We first examined the role of the element sequence itself on ribosome occupancy by comparing the UTR elements with and without motif mutations in the TRAP data in the empty-vector condition. Among the 77.3% of elements showing any significant effect, 88% (258/293 elements) showed a main effect of sequence, regardless of CELF6 expression. Looking at the distribution of log<sub>2</sub>-fold changes in expression between reference and mutant sequences under CTL transfection conditions, 220/258 (85.3%) of them were repressive when compared with their mutated counterparts (median log<sub>2</sub>-fold change: −0.47; 1.39-fold decrease). CELF6-bound elements decreased ribosome occupancy, even in the absence of CELF6 expression.

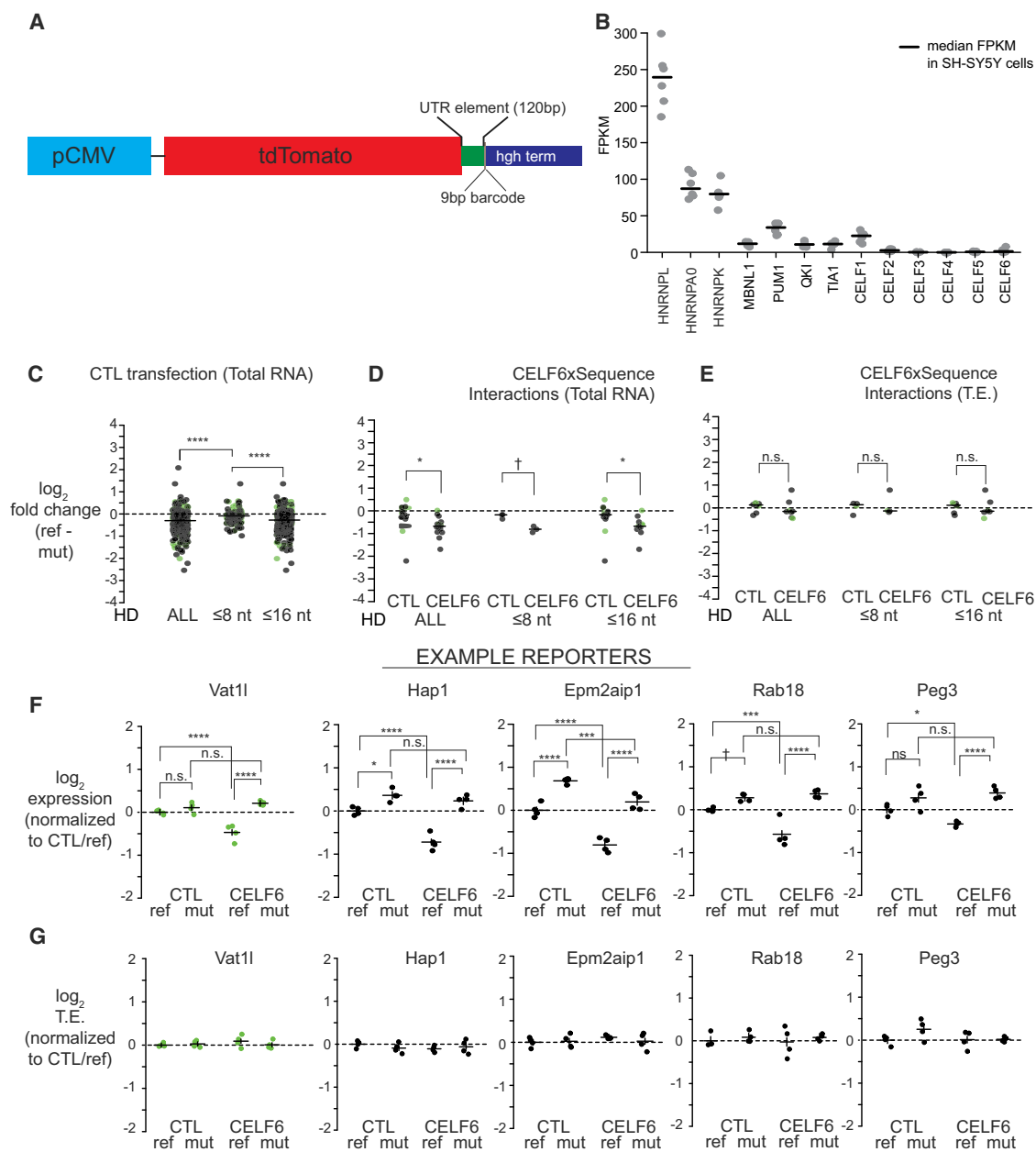
There are two primary mechanisms by which a sequence can alter ribosome occupancy—altering translation efficiency (TE), usually reflecting the loading of mRNAs onto ribosomes, or altering mRNA stability (Garneau et al., 2007; Hinnebusch et al., 2016; Ivshina et al., 2014; Piqué et al., 2008). To assess TE, log<sub>2</sub> expression for each barcode in the TRAP samples was normalized to input RNA to account for their overall abundance (log<sub>2</sub> TE). We then fit our model independently for both input RNA expression and TE. A summary of those effects is shown in Tables S7 and S8.

To assess relative influence of either mechanism, we first looked at input RNA expression; 285/379 elements (75.2%) showed any significant effect (282/379, 74.4% at FDR < 0.1). Similar to what was found in the analysis of TRAP RNA levels, 89.8% of them (256/285) showed a main effect of sequence mutation, regardless of CELF6 expression. Similar to ribosome occupancy, 86.3% had fold changes less than 0, indicating that reference sequences are repressive when compared with their mutated counterparts (Figure 3C) (median log<sub>2</sub>-fold change: −0.47; 1.39-fold decrease). We did not observe a similar trend in TE. Only 34/379 sequences (8.97%) showed any nominally significant effect on TE, and 0 showed any significant effect at FDR < 0.1; 15/34 showed a significant effect of sequence; however, six showed negative log<sub>2</sub>-fold changes, and nine showed positive fold changes. This indicates that alteration of transcript abundance is the primary mechanism leading to decreased ribosomal occupancy, rather than changes in TE.

Significant changes to expression could be incurred because of large numbers of mutated bases. Thus, we looked at changes between reference and mutant for elements with smaller amounts of mutation (Hamming distance [HD] of ≤8, the approximate size of a motif, or ≤16 nucleotides; maximum HD = 25 for any element). Expression of reference sequences was generally lower than mutant sequences for these subsets, with 72.9% at <0 and 84.8% at <0 for HD ≤8 and HD ≤16, respectively (Figure 3C). Thus, even a modest amount of mutation is capable of elevating the expression levels of most elements.

### CELF6 Decreases RNA Abundance in a Sequence-Dependent Manner

We next assessed interactions between CELF6 expression and sequence, focusing on RNA expression; 16 elements (15 unique genes) showed a nominally significant interaction (p < 0.05)



**Figure 3. CELF6-CLIP-Enriched Motif Sequences Represent a Set of Repressive Elements**

(A) PTRE-seq reporter (pmrPTRE\_AAV).

(B) Fragments per kilobase of transcript per million reads (FPKM) levels are shown for several RNA-binding proteins in SH-SY5Y cells ( $n = 8$ ). Lines represent median FPKM.

(C) Reporter RNA  $\log_2$ -fold changes in expression (reference versus mutant) in CTL transfections, for all element pairs, as well as Hamming distance (HD)  $\leq 8$  or  $\leq 16$ .

(D) Reporter RNA  $\log_2$ -fold changes between reference and mutant in CTL and CELF6 expression conditions, from elements showing significant condition X sequence interactions.

(E) Reporter RNA  $\log_2$ -fold changes in translation efficiency (TE) between reference and mutant in CTL and CELF6 conditions, for elements in (D). Data points in (C)–(E) are  $\log_2$ -fold changes (reference versus mutant) per condition, averaged by replicate over barcodes and then over replicates. Lines represent medians. Comparisons between conditions or sequence mutation were assessed by Mann-Whitney U tests.

(F)  $\log_2$  expression across five example reporter library elements in CTL, CELF6, reference, and mutant conditions.

(G)  $\log_2$  TE across example reporters in (F). Data points in (F) and (G) are averaged across barcodes for each biological replicate.

Lines in (F) and (G) represent average  $\pm$  95% confidence intervals, normalized to CTL/reference condition. Post hoc pairwise comparisons in (F) and (G) computed using the multcomp package in R. n.s.,  $p > 0.1$ ; \* $p < 0.05$ , \*\* $p < 0.001$ , \*\*\* $p < 0.0001$ , \*\*\*\* $p < 1E-5$ . Green points represent library elements selected from less-stringent CELF6 target criteria (enrichment over input alone).

Scale bar in (A) represents 0.1 mm as indicated.

between CELF6 expression and sequence. Figure 3D plots the average  $\log_2$ -fold change between reference and mutant sequence for both the CTL and CELF6 conditions across each subsetted level of sequence mutation. The fold changes between reference and mutant sequences were more negative (median: 1.4-fold) under CELF6 expression. That is driven by repression of the reference sequence, rather than an elevation of the mutant sequence by CELF6 (median  $\log_2$ -fold change CELF6-CTL reference =  $-0.37$ , CELF6-CTL mutant  $-0.05$ ;  $p = 9E-4$ , Mann-Whitney U test). This indicates the CELF6 affects those elements by further decreasing mRNA levels, and that effect is sequence dependent.

As mentioned earlier, few effects were observed on TE by comparison. Only 8/34 elements with any effect on TE showed interactions with CELF6 expression at  $p < 0.05$ , with 0 showing interactions at FDR  $< 0.1$  (Figure 3E). Additionally, we did not detect a consistent direction of CELF6 effect on TE. Looking at changes to reference sequence TE when CELF6 is overexpressed, only two elements showed a trend to difference compared with CTL, one showing a decrease (Lin7c,  $\log_2$ -fold change:  $-0.45$ ;  $p = 0.05$ ) and one showing an increase (Enc1,  $\log_2$ -fold change:  $+0.71$ ;  $p = 0.052$ ). Other elements had different sources of interaction, e.g., a difference between CELF6 and CTL only in mutated elements (e.g., Prkacb,  $\log_2$ -fold change in mutated sequence only:  $+0.29$ ;  $p = 0.002$ ). By contrast, looking at input RNA expression, 14/16 elements with nominally significant interactions showed  $\log_2$ -fold changes in the reference between CELF6 and CTL, changes that were negative. Examples of element expression are shown in Figures 3F, and Figure 3G shows TE. These findings indicate that, for elements showing interactions between CELF6 and element sequence, there is a repression of the reference sequence via a decrease in RNA abundance with CELF6 expression, and that effect is abolished after mutation of the conserved motifs. Therefore, we conclude that, overall, when CELF6 exerts a measurable effect in our system, that effect is to decrease RNA abundance. Where impact on TE occurs, it is not generalizable in the direction of the effect.

To further validate these findings and confirm an effect on final protein levels, we overexpressed five individual reporters, with and without mutation, along with EGFP-tagged CELF6 or EGFP alone, then measured reporter expression by tdTomato fluorescence (Figure 4). For 4/5 cases, repression upon CELF6 expression was the same direction and magnitude to that observed by PTRE-seq. For 3/4, the repression observed by CELF6 was significantly reversed by motif mutation. This shows that this finding ultimately affects target protein levels.

### CELF3–5 Also Decrease RNA Abundance

CELF3, CELF4, and CELF6 binding preferences determined by RNAcompete are highly similar (Figure 2), and as a group, CELF3–6 are more similar in amino acid identity than CELF1 or CELF2 are (Dasgupta and Ladd, 2012). Therefore, we hypothesized that the repression we observed with CELF6 would also be true of CELF3–5. We transiently transfected our PTRE-seq library along with His/Xpress-tagged human CELF3, CELF4, or CELF5 used previously to study these proteins (Figure S2) (Ladd et al., 2001, 2004). We then performed the analysis of mRNA abundance described in the preceding section, refitting

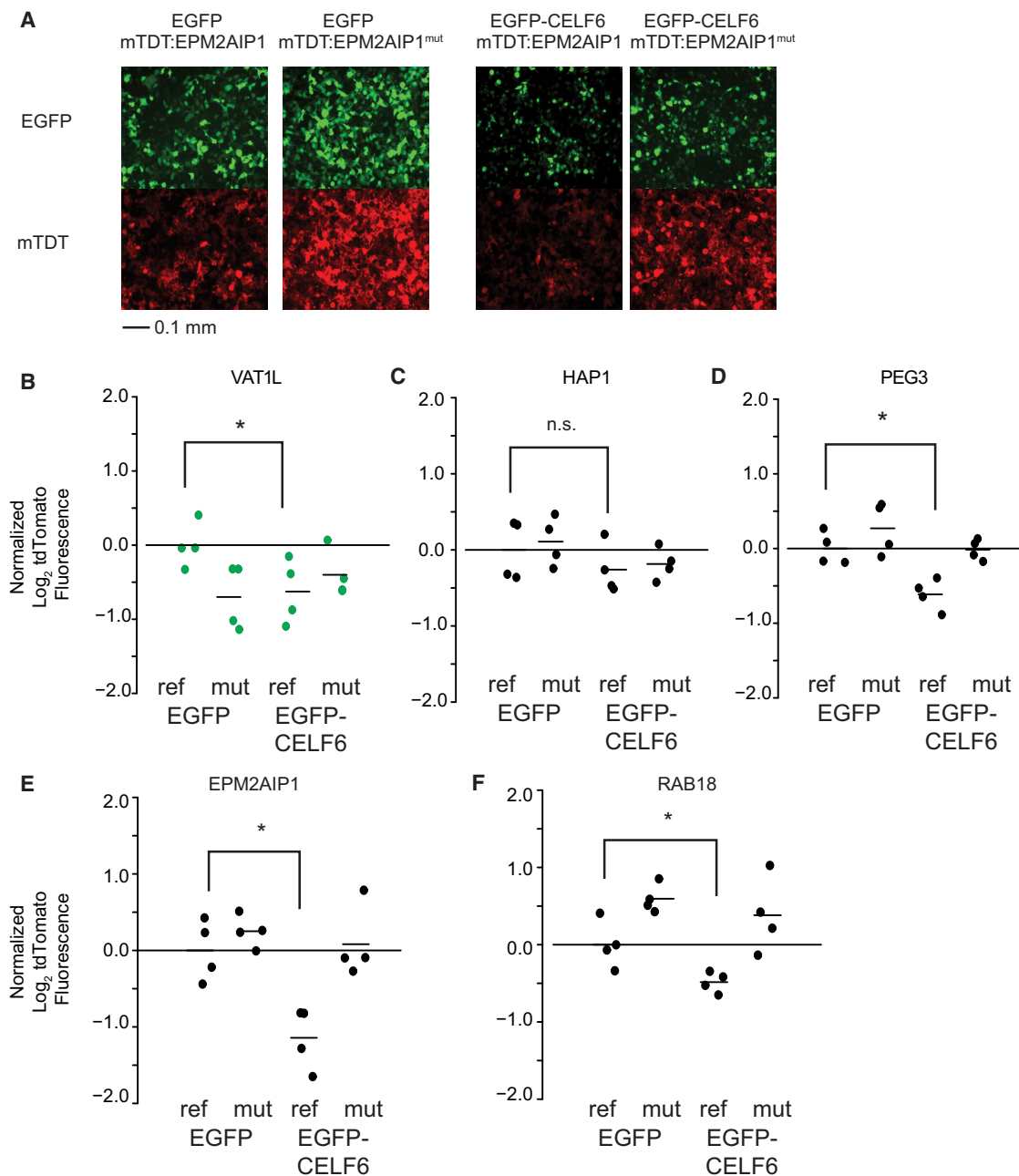
the linear mixed models to include data from these new expression conditions (Figure 5).

There were 63/379 ( $p < 0.05$ ; 42/379 with FDR  $< 0.1$ ) elements representing 63/153 genes showing a nominally significant sequence by CELF-expression interactions in the total reporter RNA. Looking at the  $\log_2$ -fold change in the reference versus the mutant sequence, CELF3 and CELF4 showed the largest differences (Figure 5A). The median  $\log_2$ -fold change in the reference versus the mutant sequence in the CELF3 expression condition was  $-1.22$ , with comparable change under CELF4 expression ( $-1.06$ ) (CTL median  $\log_2$ -fold change:  $-0.36$ ; CTL versus CELF3  $p = 9.5E-13$ ; CTL versus CELF4  $p = 1.2E-11$ ; and CELF3 versus CELF4  $p = 0.31$ ; Mann-Whitney U tests). CELF6 and CELF5 also showed comparable fold changes, and those were intermediate between the CTL condition and CELF3 or CELF4 conditions (CELF5:  $-0.58$ , CELF6:  $-0.57$ ; CELF5 versus CTL,  $p = 0.0025$ ; CELF6 versus CTL,  $p = 0.0016$ ; CELF5 versus CELF6,  $p = 0.57$ ). These rankings were consistent at both low ( $HD \leq 8$ ) and higher ( $HD \leq 16$ ) levels of sequence mutation (Figures 5B and 5C). Thus, expression of any of these CELFs was able to reduce abundance of reference reporters, an effect that could be abolished by mutation. Furthermore, CELF3 and CELF4 were associated with the strongest effects, whereas CELF5 and CELF6 showed more moderate effects. Thus, within these CELF proteins, CELFs 5/6 and CELFs 3/4 appear to form distinct subgroups with respect to their effects on mRNA abundance.

When looking at TE, there were fewer significant effects. Nine elements showed significant interactions of condition by element sequence ( $p < 0.05$ , 0/379 at FDR  $< 0.1$ ). However, when plotted in terms of their fold change between reference and mutant sequence, we were again unable to generalize a direction of effect on TE across elements. Median  $\log_2$ -fold changes in TE were near 0, with no significant differences across conditions (Figures 5D–5F). Thus, CELF expression decreased overall ribosome occupancy of reference elements, but largely by disrupting reporter transcript abundance. Total reporter expression for key examples are shown in Figure 5G with their respective TE shown in Figure 5H. Among significant CELF by sequence interactions, 6/63 showed significant reductions between a CELF condition and CTL in the mutant sequence (Slc39a6, Olfm1, Slc25a1, Pcdh17, Zbtb5, and Socs5). However, even though these mutant sequences appeared to respond to the CELF condition, their respective reference sequences were still more negative (Table S7). Thus, we find, in general, the effect of CELF by sequence interaction to be a CELF-associated decrease in reporter levels for elements with intact motifs.

Because expression of various CELFs may overlap in the brain, we next tested combinatorial effects of co-expressing CELF proteins. We transfected CELF6 construct with one of CELF3–5 in equimolar proportions. Among elements showing significant sequence-by-condition interactions, expression of CELF3 and CELF6 together resulted in repression similar to CELF3 by itself (median CELF3/CELF6  $\log_2$ -fold change reference versus mutant:  $-1.23$ , CELF3 alone:  $-1.22$ ;  $p = 0.37$ ). This was also true of CELF4 (median CELF4/CELF6  $\log_2$ -fold change:  $-1.13$ , CELF4 alone:  $-1.06$ ;  $p = 0.34$ ). Thus, the effect of CELF4 and CELF3 appears to be dominant, or at least





#### Figure 4. Individual mtdTomato UTR Element Constructs Replicate Interactions of CELF6 Expression and Sequence

Individual UTR elements were cloned as in Figure 3A and transfected along with either EGFP or EGFP-CELF6

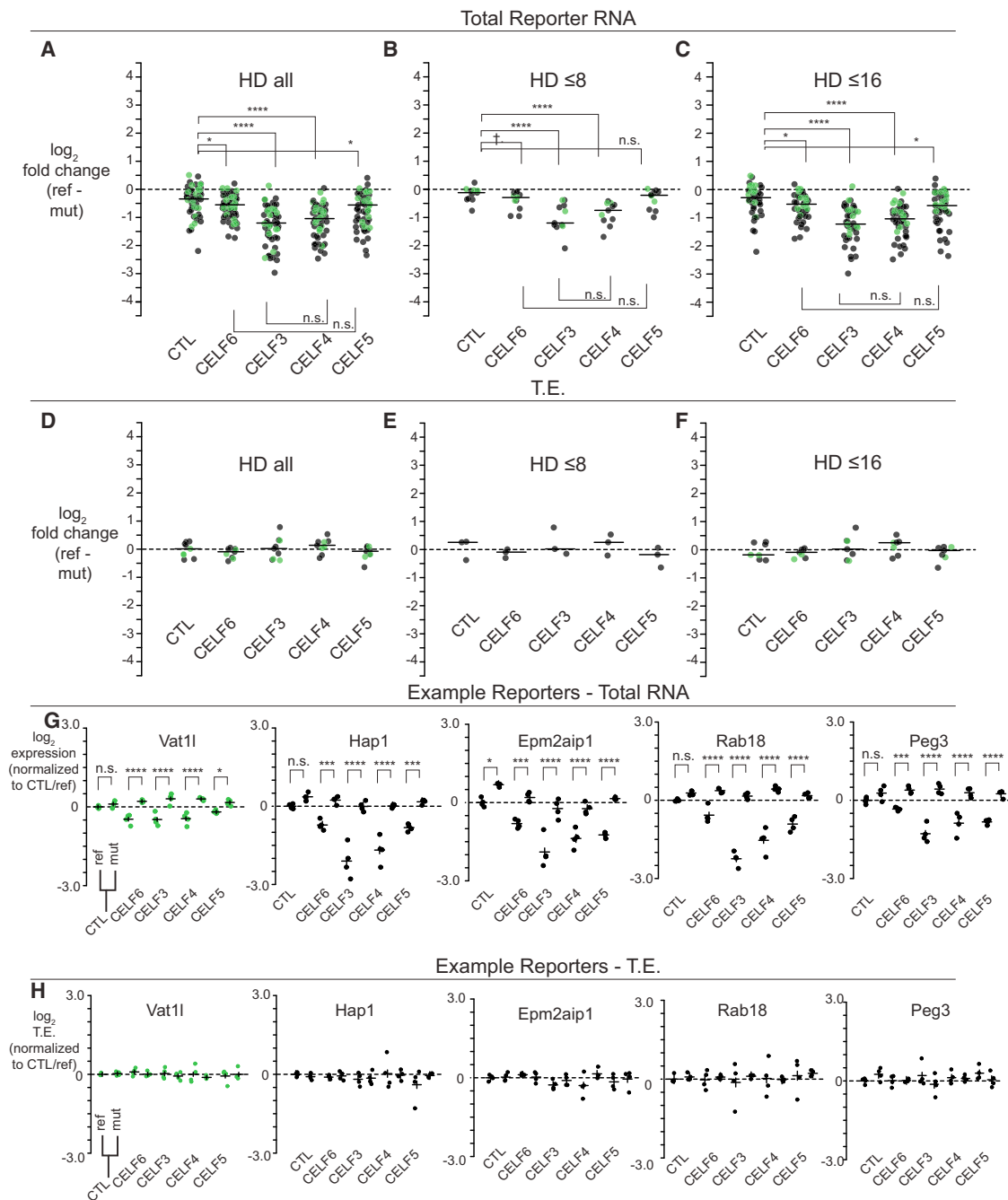
(A) Example of live-cell epifluorescent images from a transfection experiment showing EGFP or EGFP-CELF6 expression (top panels) and either tdTomato:EPM2AIP1 or tdTomato:EPM2AIP1(mutant) (bottom panels).

(B–F) tdTomato  $\text{log}_2$  fluorescence, with lines representing means normalized to EGFP/reference sequence.

Significance was assessed by two-way ANOVA, with post hoc pairwise comparisons using the multcomp package in R. n.s.,  $p > 0.1$ ;  $^{\dagger}p < 0.1$ ,  $^*p < 0.05$ ,  $^{**}p < 0.001$ . Green points represent library elements selected from less-stringent CELF6 target criteria (enrichment over HA/YFP input alone). Scale bar in (A) represents 0.1 mm as indicated.

maximal, in these co-transfections, not because of differences in the level of each CELF's expression (Figure S2). When CELF5 and CELF6 were expressed together, the median  $\text{log}_2$ -fold change of reference versus mutant sequence was more negative

in the CELF5/CELF6 condition compared with CELF6 alone by approximately 20% (median  $\text{log}_2$ -fold change in CELF5/CELF6:  $-0.72$ , CELF6 alone:  $-0.57$ ;  $p = 0.037$ ), although that comparison did not reach significance compared with CELF5



**Figure 5. CELF3–5 Show Redundancy in Ability to Enhance Repression of CELF6-CLIP-Enriched UTR Elements**

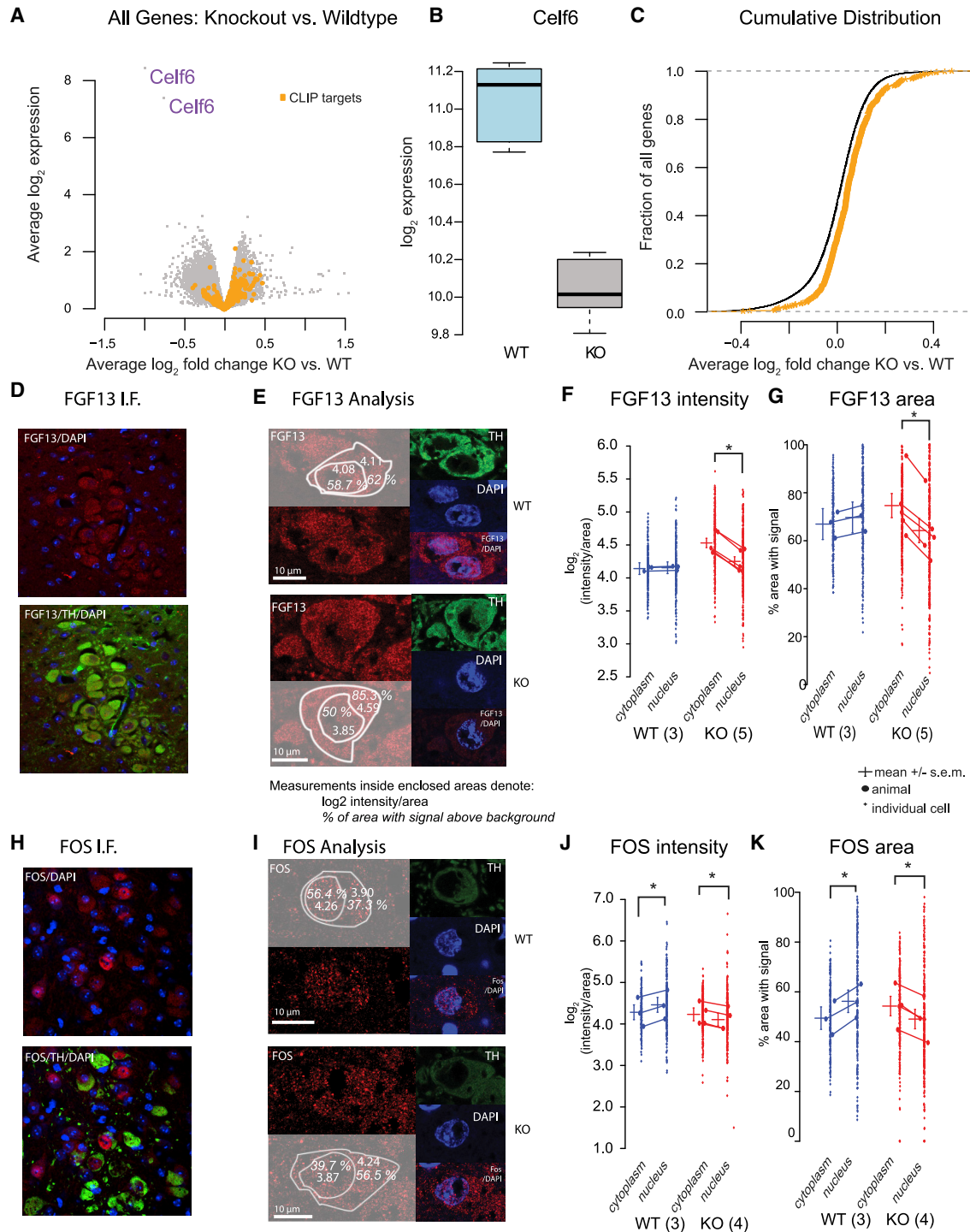
(A–C) Log<sub>2</sub>-fold change across CELF conditions considering all reporter elements, or reporter elements with HD ≤ 8 or ≤ 16, for any element with significant condition X sequence interactions.

(D–F) Log<sub>2</sub>-fold change in TE for conditions, as in (A)–(C). Data points in (A)–(F) were averaged for redundant barcodes, then, across replicates, with lines representing medians. Statistical comparisons in (A)–(F) were assessed by Mann-Whitney U tests.

(G) Log<sub>2</sub> expression across five example reporter library elements across CELF, reference, and mutant conditions.

(H) Log<sub>2</sub> TE across example reporters in (G). Data points in (G) and (H) are averaged across barcodes, with each dot representing a replicate.

As in Figure 3, horizontal lines in (G) and (H) represent average expression or TE ± 95% confidence intervals, normalized to the CTL/reference. Post hoc pairwise comparisons in (G) and (H) computed with multcomp in R. n.s., p > 0.1; †p < 0.1, \*p < 0.05, \*\*p < 0.001, \*\*\*p < 0.0001, \*\*\*\*p < 1E–5. Green points represent library elements selected from less-stringent CELF6 target criteria (enrichment over HA/YFP input alone).



**Figure 6. CELF6 Targets Identified by CLIP Show Increased Expression in *Celf6*-Knockout Brains**

(A) Volcano plot showing gene-expression changes in *Celf6* KO versus WT mouse brains.

(B) *Celf6* mRNA is decreased in KO brains as expected ( $p < 1 \times 10^{-6}$ ).

(C) CLIP targets (orange) are derepressed in KO brains compared with other genes ( $p < 1.1 \times 10^{-9}$ , Welch's t test).

(D) Example of anti-FGF13 staining in LC.

(E) IF Analysis, left: FGF13 staining (red) with overlay showing quantification. Example images were selected to be near the mean of the data. Right: tyrosine hydroxylase (TH) and DAPI staining.

(legend continued on next page)

alone (CELF5 alone:  $-0.58$ ;  $p = 0.19$ ). Nevertheless, that may indicate that CELF5 and CELF6 can act additively. Overall, our reporter assays show that CELF proteins suppress mRNA abundance of targets containing specific sequence motifs.

### CELF6 Regulates Transcript Abundance in the Brain

We next determined whether CELF6 affects target transcript abundance *in vivo* when CELF6 is absent. We examined targets by microarray analysis, comparing eight WT and eight *Celf6*-KO brains. Microarray probes measuring *Celf6* are reduced, confirming loss of transcript, in the KO mouse (Figure 6B). Effects of CELF6 loss on total brain RNA abundance are modest: whereas several hundred genes show a nominally significant change in RNA abundance (Figure 6A), the median  $\log_2$ -fold change is 0.24, and only the *Celf6* probes survive multiple testing corrections. However, examining the distribution of fold changes in CLIP targets compared with randomly selected probes, showed increased abundance in KO brains ( $p < 1.11E-09$ , Welch's *t* test) (Figure 6C). This is consistent with a role for CELF6 in decreasing target stability *in vivo*. Given that CELF6 expression is restricted regionally (Maloney et al., 2016) and brain CELFs may have overlapping activity, modest changes are unsurprising. This experiment also combines direct effects on CELF6 targets and indirect effects on gene expression because of development in the KO background. However, despite those limitations, we see nominally significant ( $p < 0.05$ ) regulation of several high-confidence CELF6 CLIP targets, including important neuronal signaling genes, such as *Reln* and *Fgf13*, as well as genes meeting lower-stringency criteria, such as *Fos* and *Mecp2* (Figure S3); 20/21 measurement CELF6 targets change in the expected direction, showing increased mRNA levels in the KO, a result highly unlikely to be due to chance ( $p < 0.0005$ ,  $\chi^2$  test). This is driven by the 3' UTR CLIP targets, ( $p < 2.2E-08$ ; Figure 6C) because we did not detect changes in target genes identified via 5' UTR or intronic binding. A few targets identified via binding in the CDS show a similar magnitude of increase in the KO; however, that did not reach statistical significance. Thus, overall, our analysis indicates CELF6 deletion is generally associated with an increase in the RNA abundance of its targets, *in vivo*, as it does in reporter assays *in vitro*.

A systems analysis of the consequences of CELF6 loss is also intriguing (Figure S3C). Gene Ontology of the 200 most-upregulated genes in the KO brains ( $p < 0.05$ ), regardless of CLIP status, reveals significant enrichment for terms for transcription regulators (transcription factor complex,  $p < 0.8E-3$  with Benjamini-Hochberg [B-H] correction) and other components of the nucleus, as well as synaptic genes (growth cone,  $p < 3.47E-4$ , and dendritic spine,  $4.02E-3$ ). Downregulated genes showed no such consistent enrichment, with only a single category

showing downregulation (flagellum,  $p < 9.62E-3$ ). Thus, loss of CELF6 appears to dysregulate both nuclear transcription and synaptic genes *in vivo*.

### CELF6 Regulates FOS and FGF13 Protein Abundance and Localization in the Locus Coeruleus

Finally, we tested the effect of *Celf6* KO on target protein levels and localization *in vivo*. CELF6 is expressed sparsely in the brain, and immunoblots would be uninformative. We, therefore, began with CLIP targets in which antibodies had been previously validated for immunofluorescence in the brain. We focused on fibroblast growth factor 13 (FGF13), a non-canonical FGF previously shown to bind intracellularly to voltage-gated sodium channels and regulate their subcellular spread (Pablo et al., 2016). Additionally, we looked at FOS (a lower-stringency target), an immediate early gene stimulated by neuronal activity and involved in downstream synaptic changes required for phenomena such as long-term potentiation (Abraham et al., 1991). We assayed expression in the locus coeruleus (LC), a stress-responsive neuromodulatory nucleus, identifiable by location and tyrosine hydroxylase (TH) staining, with robust CELF6 expression in WT mice (Maloney et al., 2016). To enable detection of FOS, animals were first subjected to restraint stress to activate the LC. We confirmed the antibodies produced a detectable signal in the LC under those conditions (Figures 6D and 6H). Because localization is important to the function of both proteins, the FGF13 or FOS signal was quantified for both fluorescence intensity and subcellular compartment (cytoplasmic or nuclear), as determined by DAPI or TH staining, respectively (Figures 6E–G and 6I–6K).

In the case of FGF13, we observed a significant increase (~20%) in the cytoplasmic KO mean expression (Figure 6E;  $\log_2$  (intensity/area) cytoplasm = 4.53, nucleus = 4.25;  $p < 0.001$ ;  $\log_2$ -fold change, cytoplasm – nucleus = 0.28), whereas we did not observe a significant change between WT cellular compartments ( $\log_2$ -fold change =  $-0.013$ ). Additionally, we observed a significant increase (~10%) in KO cytoplasmic FGF13 spread (Figure 6F) quantified as the percentage of cytoplasmic area containing a signal above background (KO cytoplasmic mean = 74.6%, KO nuclear mean = 64.2%;  $p < 0.0001$ ). In the WT, we did not detect a significant difference, although we observed an opposing trend with ~3% change in the area with a signal (WT cytoplasm = 66.9%, WT nucleus = 64.2%;  $p = 0.066$ ).

In the case of FOS, we observed a small increase in signal (~10%) in the cytoplasm compared with the nucleus (Figure 6H; KO cytoplasmic  $\log_2$  intensity/area = 4.23, KO nucleus = 4.1;  $p < 0.0001$ ;  $\log_2$ -fold change cytoplasm – nucleus = 0.13) and the opposing trend (~12% change) in the WT (WT cytoplasm = 4.28, WT nucleus = 4.46;  $p < 0.0001$ ,  $\log_2$ -fold change =  $-0.18$ ). The percentage of area containing a FOS signal was

(F) FGF13 fluorescence intensity ( $\log_2$  pixel intensity/area), with small dots representing cells and large dots averaged by animal (WT:  $n = 3, 92, 45$ , and 71 cells; KO:  $n = 5, 69, 19, 41, 99$ , and 69 cells). Lines represent grand means  $\pm$  95% confidence intervals. (G–K) FGF13 area (percentage of region with an antibody signal above background) for the same cells as in (E). (G)–(I), as in (H)–(K), but for FOS immunostaining (WT:  $n = 3, 68, 70$ , and 62 cells; KO:  $n = 4, 94, 53, 44$ , and 98). Intensity and area with signal were analyzed using linear mixed models fitting fixed effects of subcellular localization (cytoplasm, nucleus), genotype (WT, KO), and their interaction, using individual cells per animal as a repeated-measure/random effect. Significant effects were determined by ANOVA with Tukey post hoc multiple comparisons between means computed in R using the *lsmeans* package. Significant comparisons in (D)–(K) are for  $p < 0.001$ . Scale bars in (E) and (I) represent 10  $\mu$ m as indicated.

enriched in the nucleus of WT animals (Figure 6I; ~7% higher in the nucleus;  $p < 0.0001$ ; cytoplasm = 49.4%, nucleus = 56.2%) but enriched in the cytoplasm of KO animals (~5% higher in the cytoplasm;  $p < 0.0001$ ; cytoplasm = 54.3%, nucleus = 49.1%) The effects that are small in magnitude, however, are both consistent with the model in which CELF6 represses mRNA levels and protein expression in WT animals, which are de-repressed in the cytoplasm in KO animals.

## DISCUSSION

In this study, we identify the *in vivo* targets of CELF6 in the brain. Our methods required us to use a tagged protein, but that protein exhibits the endogenous expression pattern and is only modestly overexpressed (Maloney et al., 2016). We found that CELF6 primarily binds to 3' UTRs of mRNAs, with many targets shared by CELF4 (Wagnon et al., 2012). UTR elements under CELF6 CLIP peaks are enriched for several motifs previously identified (Ray et al., 2013) and are similar to CELF1 preferences (Wang et al., 2015). Using PTRE-seq (Cottrell et al., 2018), we evaluated the effect of CELF6 across many binding elements. Although our *in vitro* assays are not direct binding assays, we show that CELF6 and other CELFs are associated with downregulation of target 3' UTRs, and this can be mitigated by motif mutation. Future work on this protein will involve understanding its binding affinity and effect on mRNA half-life. Although CELF6 has been shown to regulate alternative splicing (Ladd et al., 2004), we find very few significant targets outside of 3' UTR regions, suggesting that CELF6 in the brain may be more involved with post-transcriptional regulation. UTR binding is consistent with recent work showing CELF6 is able to regulate the stability of p21 (Liu et al., 2019) and FBP (Yang et al., 2020) in cancer cells. We found very few changes to splicing in *Celf6* KO mice (not shown; available at GEO: GSE160293) and few effects to TE, suggesting that CELF6 acts on its brain targets primarily to enhance mRNA degradation. We show these same targets are regulated *in vivo* and that loss of CELF6 can alter both protein intensity and subcellular distribution.

Our work here raises a number of biological questions. Although most of the PTRE-seq library was repressive, many elements were not further sensitive to CELF6 expression. This suggests that either CELF6 has additional functions on these transcripts not assayed (e.g., RNA localization) or that its functional effect on these sequences depends on cellular context. Although CELF2-6 RNA levels were largely undetectable in SH-SY5Y cells, the complement of other RBPs certainly differs from that of neurons. We did detect levels of MBNL1 in SH-SY5Y cells, an RBP that has been shown to antagonize CELF1 (Wang et al., 2015) leading to mRNA stabilization. It has also been shown that CELF1, CELF2, and CELF6 can all antagonize MBNL1's splicing functions (Ohsawa et al., 2015). If MBNL1 and CELF6 can act antagonistically, then the resistance to CELF6 of some constructs in culture may be due to MBNL1 expression. Antagonistic activity of RBPs on mRNA translation and stability has also been observed for CELF2 and HUR (Sureban et al., 2007) and ELAVL1 and ZFP36 (Mukherjee et al., 2014). Such activity between RBPs may be mediated by binding-site proximity, and when RBP levels are altered, access to sites

changes, and so does regulation (see Plass et al., 2017, study of AU-rich elements). In our analysis of CELF6 CLIP targets, we found enriched binding motifs across several RBP families. Future work will explore how interactions between sequence elements affects target CELF sensitivity.

It may be that CELF6-association downregulation of targets is due to enhancement of mRNA decay. Indeed, CELF1 has been shown able to recruit poly(A) ribonuclease (PARN) to RNA targets to facilitate mRNA decay, and Moraes et al (2006) found that CELF1 could associate with PARN *in vitro*. We tested co-immunoprecipitations of CELF6 with PARN but did not find evidence of an association *in vitro* (data not shown). However, it may associate with other proteins that mediate decay. The *Xenopus* homolog of the CELF proteins, EDEN-BP, has also been shown to regulate deadenylation, and oligomerization of the protein is required for that activity (Cosson et al., 2006). Future work will also explore whether CELF interactions (homo- or heterodimers) exist and can affect regulation. Finally, structure-function analysis mapping the domains involved in any such interactions will be of great interest in future studies.

CELF6 target enrichment for synaptic mRNAs and altered localization of FOS and FGF13 make it interesting to speculate that there is a role for CELF6 in regulating neuronal mRNA localization. Neurons are known to carefully regulate localization near activated spines to allow local translation to influence the development and strength of synaptic connections. However, because CELF6 expression is limited to sparse populations of neuronal cells poorly characterized with respect to local mRNA regulation, foundational work will be required before that can be fully assessed.

In sum, we have presented identification of CELF6 CNS binding targets *in vivo* and show that CELF6 can regulate mRNA abundance, both *in vitro* and *in vivo*. Some of these targets may mediate the behavioral phenotypes of the CELF6 KO mouse, which include communicative, exploratory, and reward system deficits (Dougherty et al., 2013; Maloney et al., 2019). Robust changes in response to more modest changes in RNA abundance indicate that mRNA translation must be carefully tuned for normal brain function and that subtle disruption of mRNA levels can substantially change the organism. Our data present an opportunity for further investigation into which targets, in which cell types, can regulate behavior.

## STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead Contact
  - Materials Availability
  - Data and Code Availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Animal Models
  - Cell Culture
- METHOD DETAILS
  - CLIP

- CLIP-Seq Sequencing Library Preparation
- Total RNA-Seq of SH-SY5Y cells
- Total RNA-Seq and CLIP-Seq Raw Data Processing
- Peak Calling and Read Counting
- CLIP Motif Enrichment Analysis
- PTRE-Seq Reporter Library Preparation
- PTRE-Seq Reporter Library Transfection
- PTRE-Seq Sequencing Library Preparation
- RT-PCR and qRT-PCR
- PTRE-Seq Reporter Validation
- Agilent Gene Microarray
- FGF13 and FOS Immunofluorescence
- **QUANTIFICATION AND STATISTICAL ANALYSIS**
  - Defining CELF6 CLIP Targets by Differential Expression Analysis
  - PTRE-Seq Barcode Counting and Normalization
  - PTRE-Seq Linear Mixed Model Analysis
  - PTRE-Seq Validation Analysis
  - Agilent Expression Microarray Analysis
  - Immunofluorescence Quantification

#### SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at <https://doi.org/10.1016/j.celrep.2020.108531>.

#### ACKNOWLEDGMENTS

This work was supported by the NIH (5R00NS067239, 5R21MH099798, 5R01HG008687, and R01MH116999) and the Simons Foundation. We thank Eric Wang, Eric Van Nostrand, Jean Schaffer, and Joshua Langmade for advice and training on CLIP-seq; Kyle Cottrell and Sergej Djuranovic for help with PTRE-seq assays; and Hemangi Chaudhari, Kristina Sakers, and Bernard Mulvey for consultation. We also thank Allison Lake, Susan Maloney, Affua Akuffo, Jasbir Dalal, and Nilambari Pisat, who were involved in generation of mice, data collection, and advice; the Thomas Cooper laboratory for CELF expression constructs; and Gregory Pitt for the FGF13. Finally, we thank Tomás Lagunas, Jr., Anthony Fischer, Stephen Plassmeyer, Sarah Koester, and Kayla Nygaard for reviewing this manuscript.

#### AUTHOR CONTRIBUTIONS

M.A.R., J.D.D., D.M.K, and B.A.C. designed the experiments. M.A.R. performed most of the experiments. M.A.R. and J.D.D. wrote the manuscript. H.C. performed *in vivo* immunofluorescence experiments and wrote the associated methods. Y.L. performed additional bioinformatic analysis. M.A.R., J.D.D., D.M.K., and B.A.C. revised manuscript.

#### DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: October 9, 2018  
Revised: August 14, 2020  
Accepted: November 25, 2020  
Published: December 22, 2020

#### REFERENCES

Abraham, W.C., Dragunow, M., and Tate, W.P. (1991). The role of immediate early genes in the stabilization of long-term potentiation. *Mol. Neurobiol.* *5*, 297–314.

Ascano, M., Hafner, M., Cekan, P., Gerstberger, S., and Tuschl, T. (2012). Identification of RNA-protein interaction networks using PAR-CLIP. *Wiley Interdiscip. Rev. RNA* *3*, 159–177.

Auweter, S.D., Oberstrass, F.C., and Allain, F.H.-T. (2006). Sequence-specific binding of single-stranded RNA: is there a code for recognition? *Nucleic Acids Res.* *34*, 4943–4959.

Bailey, T.L. (2011). DREME: motif discovery in transcription factor ChIP-seq data. *Bioinformatics* *27*, 1653–1659.

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* *67*, 1–47.

Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* *30*, 2114–2120.

Choi, H.S., Hwang, C.K., Song, K.Y., Law, P.-Y., Wei, L.-N., and Loh, H.H. (2009). Poly(C)-binding proteins as transcriptional regulators of gene expression. *Biochem. Biophys. Res. Commun.* *380*, 431–436.

Corcoran, D.L., Georgiev, S., Mukherjee, N., Gottwein, E., Skalsky, R.L., Keene, J.D., and Ohler, U. (2011). PARalyzer: definition of RNA binding sites from PAR-CLIP short-read sequence data. *Genome Biol.* *12*, R79.

Cosson, B., Gautier-Courteille, C., Maniey, D., Ait-Ahmed, O., Lesimple, M., Osborne, H.B., and Paillard, L. (2006). Oligomerization of EDEN-BP is required for specific mRNA deadenylation and binding. *Biol. Cell* *98*, 653–665.

Cottrell, K.A., Chaudhari, H.G., Cohen, B.A., and Djuranovic, S. (2018). PTRE-seq reveals mechanism and interactions of RNA binding proteins and miRNAs. *Nat. Commun.* *9*, 301.

Dasgupta, T., and Ladd, A.N. (2012). The importance of CELF control: molecular and biological roles of the CUG-BP, Elav-like family of RNA-binding proteins. *Wiley Interdiscip. Rev. RNA* *3*, 104–121.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* *29*, 15–21.

Dougherty, J.D., Maloney, S.E., Wozniak, D.F., Rieger, M.A., Sonnenblick, L., Coppola, G., Mahieu, N.G., Zhang, J., Cai, J., Patti, G.J., et al. (2013). The disruption of *Celf6*, a gene identified by translational profiling of serotonergic neurons, results in autism-related behaviors. *J. Neurosci.* *33*, 2732–2753.

Doyle, J.P., Dougherty, J.D., Heiman, M., Schmidt, E.F., Stevens, T.R., Ma, G., Bupp, S., Shrestha, P., Shah, R.D., Doughty, M.L., et al. (2008). Application of a translational profiling approach for the comparative analysis of CNS cell types. *Cell* *135*, 749–762.

Fox, J., and Weisberg, S. (2019). *An R Companion to Applied Regression* (Sage).

Garneau, N.L., Wilusz, J., and Wilusz, C.J. (2007). The highways and byways of mRNA decay. *Nat. Rev. Mol. Cell Biol.* *8*, 113–126.

Gilks, N., Kedersha, N., Ayodele, M., Shen, L., Stoecklin, G., Dember, L.M., and Anderson, P. (2004). Stress granule assembly is mediated by prion-like aggregation of TIA-1. *Mol. Biol. Cell* *15*, 5383–5398.

Grant, C.E., Bailey, T.L., and Noble, W.S. (2011). FIMO: scanning for occurrences of a given motif. *Bioinformatics* *27*, 1017–1018.

Harvey, R.F., Smith, T.S., Mulrone, T., Queiroz, R.M.L., Pizzinga, M., Dezi, V., Villeneuve, E., Ramakrishna, M., Lilley, K.S., and Willis, A.E. (2018). Trans-acting translational regulatory RNA binding proteins. *Wiley Interdiscip. Rev. RNA* *9*, e1465.

Heiman, M., Schaefer, A., Gong, S., Peterson, J.D., Day, M., Ramsey, K.E., Suárez-Fariñas, M., Schwarz, C., Stephan, D.A., Surmeier, D.J., et al. (2008). A translational profiling approach for the molecular characterization of CNS cell types. *Cell* *135*, 738–748.

Hinnebusch, A.G., Ivanov, I.P., and Sonenberg, N. (2016). Translational control by 5'-untranslated regions of eukaryotic mRNAs. *Science* *352*, 1413–1416.

Hothorn, T., Bretz, F., and Westfall, P. (2008). Simultaneous inference in general parametric models. *Biom. J.* *50*, 346–363.

Huang, W., Sherman, B.T., and Lempicki, R.A. (2009a). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* *4*, 44–57.

- Huang, W., Sherman, B.T., and Lempicki, R.A. (2009b). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* **37**, 1–13.
- Ivshina, M., Lasko, P., and Richter, J.D. (2014). Cytoplasmic polyadenylation element binding proteins in development, health, and disease. *Annu. Rev. Cell Dev. Biol.* **30**, 393–415.
- Kishore, S., Jaskiewicz, L., Burger, L., Hausser, J., Khorshid, M., and Zavolan, M. (2011). A quantitative analysis of CLIP methods for identifying binding sites of RNA-binding proteins. *Nat. Methods* **8**, 559–564.
- Ladd, A.N., Charlet, N., and Cooper, T.A. (2001). The CELF family of RNA binding proteins is implicated in cell-specific and developmentally regulated alternative splicing. *Mol. Cell Biol.* **21**, 1285–1296.
- Ladd, A.N., Nguyen, N.H., Malhotra, K., and Cooper, T.A. (2004). CELF6, a member of the CELF family of RNA-binding proteins, regulates muscle-specific splicing enhancer-dependent alternative splicing. *J. Biol. Chem.* **279**, 17756–17764.
- Lebedeva, S., Jens, M., Theil, K., Schwanhäusser, B., Selbach, M., Landthaler, M., and Rajewsky, N. (2011). Transcriptome-wide analysis of regulatory interactions of the RNA-binding protein HuR. *Mol. Cell* **43**, 340–352.
- Lee, J.E., and Cooper, T.A. (2009). Pathogenic mechanisms of myotonic dystrophy. *Biochem. Soc. Trans.* **37**, 1281–1286.
- Lenth, R.V. (2016). Least-squares means: the R package lmeans. *J. Stat. Softw.* **69**, 1–33.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079.
- Liao, Y., Smyth, G.K., and Shi, W. (2013). The subread aligner: fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Res.* **41**, e108.
- Liu, G., Zhang, Q., Xia, L., Shi, M., Cai, J., Zhang, H., Li, J., Lin, G., Xie, W., Zhang, Y., and Xu, N. (2019). RNA-binding protein CELF6 is cell cycle regulated and controls cancer cell proliferation by stabilizing p21. *Cell Death Dis.* **10**, 688.
- Maere, S., Heymans, K., and Kuiper, M. (2005). BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics* **21**, 3448–3449.
- Maloney, S.E., Khangura, E., and Dougherty, J.D. (2016). The RNA-binding protein Celf6 is highly expressed in diencephalic nuclei and neuromodulatory cell populations of the mouse brain. *Brain Struct. Funct.* **221**, 1809–1831.
- Maloney, S.E., Rieger, M.A., Al-Hasani, R., Bruchas, M.R., Wozniak, D.F., and Dougherty, J.D. (2019). Loss of CELF6 RNA binding protein impairs cocaine conditioned place preference and contextual fear conditioning. *Genes Brain Behav.* **18**, e12593.
- McLeay, R.C., and Bailey, T.L. (2010). Motif enrichment analysis: a unified framework and an evaluation on ChIP data. *BMC Bioinformatics* **11**, 165.
- Memczak, S., Jens, M., Elefsinioti, A., Torti, F., Krueger, J., Rybak, A., Maier, L., Mackowiak, S.D., Gregersen, L.H., Munschauer, M., et al. (2013). Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature* **495**, 333–338.
- Moore, M.J., Zhang, C., Gantman, E.C., Mele, A., Darnell, J.C., and Darnell, R.B. (2014). Mapping argonaute and conventional RNA-binding protein interactions with RNA at single-nucleotide resolution using HITS-CLIP and CIMS analysis. *Nat. Protoc.* **9**, 263–293.
- Moraes, K.C.M., Wilusz, C.J., and Wilusz, J. (2006). CUG-BP binds to RNA substrates and recruits PARN deadenylase. *RNA* **12**, 1084–1091.
- Mukherjee, N., Corcoran, D.L., Nusbaum, J.D., Reid, D.W., Georgiev, S., Hafner, M., Ascano, M., Jr., Tuschl, T., Ohler, U., and Keene, J.D. (2011). Integrative regulatory mapping indicates that the RNA-binding protein HuR couples pre-mRNA processing and mRNA stability. *Mol. Cell* **43**, 327–339.
- Mukherjee, N., Jacobs, N.C., Hafner, M., Kennington, E.A., Nusbaum, J.D., Tuschl, T., Blakeshear, P.J., and Ohler, U. (2014). Global target mRNA specification and regulation by the RNA-binding protein ZFP36. *Genome Biol.* **15**, R12.
- Nakagawa, S., and Schielzeth, H. (2013). A general and simple method for obtaining R<sup>2</sup> from generalized linear mixed-effects models. *Methods Ecol. Evol.* **4**, 133–142.
- Nakielnny, S., and Dreyfuss, G. (1996). The hnRNP C proteins contain a nuclear retention sequence that can override nuclear export signals. *J. Cell Biol.* **134**, 1365–1373.
- Van Nostrand, E.L., Gelboin-Burkhardt, C., Wang, R., Pratt, G.A., Blue, S.M., and Yeo, G.W. (2017). CRISPR/Cas9-mediated integration enables TAG-eCLIP of endogenously tagged RNA binding proteins. *Methods* **118–119**, 50–59.
- Ohsawa, N., Koebis, M., Mitsuhashi, H., Nishino, I., and Ishiura, S. (2015). ABLIM1 splicing is abnormal in skeletal muscle of patients with DM1 and regulated by MBNL, CELF and PTBP1. *Genes Cells* **20**, 121–134.
- Ouwenga, R.L., and Dougherty, J. (2015). Fmrp targets or not: long, highly brain-expressed genes tend to be implicated in autism and brain disorders. *Mol. Autism* **6**, 16.
- Pablo, J.L., Wang, C., Presby, M.M., and Pitt, G.S. (2016). Polarized localization of voltage-gated Na<sup>+</sup> channels is regulated by concerted FGF13 and FGF14 action. *Proc. Natl. Acad. Sci. USA* **113**, E2665–E2674.
- Piqué, M., López, J.M., Foissac, S., Guigó, R., and Méndez, R. (2008). A combinatorial code for CPE-mediated translational control. *Cell* **132**, 434–448.
- Plass, M., Rasmussen, S.H., and Krogh, A. (2017). Highly accessible AU-rich regions in 3′ untranslated regions are hotspots for binding of regulatory factors. *PLoS Comput. Biol.* **13**, e1005460.
- Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842.
- R Core Team (2014). R: A Language and Environment for Statistical Computing (R Foundation for Statistical Computing).
- Ray, D., Kazan, H., Cook, K.B., Weirauch, M.T., Najafabadi, H.S., Li, X., Gueroussov, S., Albu, M., Zheng, H., Yang, A., et al. (2013). A compendium of RNA-binding motifs for decoding gene regulation. *Nature* **499**, 172–177.
- Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47.
- Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140.
- Scoumanne, A., Cho, S.J., Zhang, J., and Chen, X. (2011). The cyclin-dependent kinase inhibitor p21 is regulated by RNA-binding protein PCBP4 via mRNA stability. *Nucleic Acids Res.* **39**, 213–224.
- Shetty, S. (2005). Regulation of urokinase receptor mRNA stability by hnRNP C in lung epithelial cells. *Mol. Cell. Biochem.* **272**, 107–118.
- Sugimoto, Y., König, J., Hussain, S., Zupan, B., Curk, T., Frye, M., and Ule, J. (2012). Analysis of CLIP and iCLIP methods for nucleotide-resolution studies of protein-RNA interactions. *Genome Biol.* **13**, R67.
- Sureban, S.M., Murmu, N., Rodriguez, P., May, R., Maheshwari, R., Dieckgrafe, B.K., Houchen, C.W., and Anant, S. (2007). Functional antagonism between RNA binding proteins HuR and CUGBP2 determines the fate of COX-2 mRNA translation. *Gastroenterology* **132**, 1055–1065.
- Timchenko, L.T., Timchenko, N.A., Caskey, C.T., and Roberts, R. (1996). Novel proteins with binding specificity for DNA CTG repeats and RNA CUG repeats: implications for myotonic dystrophy. *Hum. Mol. Genet.* **5**, 115–121.
- Ule, J., Jensen, K.B., Ruggiu, M., Mele, A., Ule, A., and Darnell, R.B. (2003). CLIP identifies Nova-regulated RNA networks in the brain. *Science* **302**, 1212–1215.
- Uren, P.J., Bahrami-Samani, E., Burns, S.C., Qiao, M., Karginov, F.V., Hodges, E., Hannon, G.J., Sanford, J.R., Penalva, L.O.F., and Smith, A.D. (2012). Site identification in high-throughput RNA-protein interaction data. *Bioinformatics* **28**, 3013–3020.
- Vidaki, M., Drees, F., Saxena, T., Lanslots, E., Taliaferro, M.J., Tatarakis, A., Burge, C.B., Wang, E.T., and Gertler, F.B. (2017). A requirement for mena,

an actin regulator, in local mRNA translation in developing neurons. *Neuron* 95, 608–622.e5.

Villalba, A., Coll, O., and Gebauer, F. (2011). Cytoplasmic polyadenylation and translational control. *Curr. Opin. Genet. Dev.* 21, 452–457.

Vindry, C., Vo Ngoc, L., Krays, V., and Gueydan, C. (2014). RNA-binding protein-mediated post-transcriptional controls of gene expression: integration of molecular mechanisms at the 3' end of mRNAs? *Biochem. Pharmacol.* 89, 431–440.

Wagnon, J.L., Briese, M., Sun, W., Mahaffey, C.L., Curk, T., Rot, G., Ule, J., and Frankel, W.N. (2012). CELF4 regulates translation and local abundance of a vast set of mRNAs, including genes associated with regulation of synaptic function. *PLoS Genet.* 8, e1003067.

Wang, E.T., Ward, A.J., Cherone, J., Wang, T.T., Giudice, J., Treacy, D., Freese, P., Lambert, N.J., Saxena, T., Cooper, T.A., et al. (2015). Antagonistic

regulation of mRNA expression and splicing by CELF and MBNL proteins. *Genome Res.* 25, 858–871.

Xiong, H.Y., Alipanahi, B., Lee, L.J., Bretschneider, H., Merico, D., Yuen, R.K.C., Hua, Y., Guerussov, S., Najafabadi, H.S., Hughes, T.R., et al. (2015). RNA splicing. The human splicing code reveals new insights into the genetic determinants of disease. *Science* 347, 1254806.

Yang, X., Zhao, L., Pei, J., Wang, Z., Zhang, J., and Wang, B. (2020). CELF6 modulates triple-negative breast cancer progression by regulating the stability of FBP1 mRNA. *Breast Cancer Res. Treat.* 183, 71–82.

Zhang, C., and Darnell, R.B. (2011). Mapping in vivo protein-RNA interactions at single-nucleotide resolution from HITS-CLIP data. *Nat. Biotechnol.* 29, 607–614.



## STAR★METHODS

### KEY RESOURCES TABLE

| REAGENT or RESOURCE                                     | SOURCE                                 | IDENTIFIER  |
|---|--|---|
| <b>Antibodies</b>                                       |  |   |
| Mouse anti-EGFP(1)                                      | MSKCC                                  | Clone 19F7; RRID: AB_2716736  |
| Mouse anti-EGFP(2)                                      | MSKCC                                  | Clone 19C8; RRID: AB_2716737  |
| Chicken anti-GFP  | AVES                                   | GFP-1020; RRID: AB_10000240   |
| Mouse anti-TH   | MilliporeSigma                         | MAB318; RRID: AB_2201528  |
| Rabbit anti-FGF13                                       | Geoffrey Pitt lab                      | N/A   |
| Goat anti-cFOS  | Santa Cruz                             | sc-52-g; RRID: AB_2629503   |
| Alexa Fluor 488 Donkey anti-Mouse                       | ThermoFisher                           | A21202; RRID: AB_141607   |
| Alexa Fluor 546 Donkey anti-Rabbit                      | ThermoFisher                           | A10040; RRID: AB_2534016  |
| Alexa Fluor 647 Donkey anti-Goat                        | ThermoFisher                           | A21447; RRID: AB_2535864  |
| Horseradish Peroxidase Goat anti-chicken IgY            | AVES                                   | H-1004; RRID: AB_2313517  |
| <b>Bacterial Strains</b>                                |  |   |
| NEB® 5-alpha Competent <i>E. coli</i> (High Efficiency) | New England Biolabs                    | C2987   |
| <b>Deposited Data</b>                                   |  |   |
| CLIP-Seq raw data                                       | This paper                             | GEO: GSE118623  |
| PTRE-Seq raw data                                       | This paper                             | GEO: GSE118623  |
| <b>Experimental Models: Cell Lines</b>                  |  |   |
| Human: SH-SY5Y (female, age 4 years)                    | ATCC                                   | CRL-2266  |
| <b>Experimental Models: Organisms/Strains</b>           |  |   |
| Mouse: CELF6-HA/YFP                                     | Dougherty Lab                          | JD2078  |
| Mouse: Celf6 <sup>-/-</sup> ("KO")                      | Dougherty Lab                          | Brj56   |
| <b>Oligonucleotides</b>                                 |  |   |
| See <a href="#">Table S9</a>                            | Integrated DNA Technologies            | N/A   |
| <b>Recombinant DNA</b>                                  |  |   |
| His/Xpress-CELF3 plasmid                                | Thomas Cooper lab                      | N/A   |
| His/Xpress-CELF4 plasmid                                | Thomas Cooper lab                      | N/A   |
| His/Xpress-CELF5 plasmid                                | Thomas Cooper lab                      | N/A   |
| His/Xpress-CELF6 plasmid                                | Thomas Cooper lab                      | N/A   |
| EGFP-CELF6  | This paper                             | N/A   |
| pEGFP-C1  | Clontech                               | N/A   |
| pcDNA3.1 His-A  | Thermo Fisher                          | V38520  |
| EGFP-RPL10a   | <a href="#">Heiman et al., 2008</a>    | N/A   |
| pQC membrane TdTomato IX                                | Addgene                                | 37351   |
| pmrPTRE-AAV   | This paper                             | N/A   |
| pmrPTRE-AAV with CELF6 CLIP target UTR library          | This paper                             | N/A   |
| <b>Software and Algorithms</b>                          |  |   |
| Trimmomatic   | <a href="#">Bolger et al., 2014</a>    | <a href="http://www.usadellab.org/cms/?page=trimmomatic">http://www.usadellab.org/cms/?page=trimmomatic</a>   |
| STAR  | <a href="#">Dobin et al., 2013</a>     | <a href="https://github.com/alexdobin/STAR">https://github.com/alexdobin/STAR</a>   |
| FGBio   | Fulcrum Genomics                       | <a href="https://github.com/fulcrumgenomics/fgbio">https://github.com/fulcrumgenomics/fgbio</a>   |
| Picard Tools  | Broad Institute                        | <a href="https://broadinstitute.github.io/picard">https://broadinstitute.github.io/picard</a>   |
| Samtools  | <a href="#">Li et al., 2009</a>        | <a href="http://samtools.sourceforge.net/">http://samtools.sourceforge.net/</a>   |
| Bedtools  | <a href="#">Quinlan and Hall, 2010</a> | <a href="https://github.com/arq5x/bedtools2/blob/master/docs/content/overview.rst">https://github.com/arq5x/bedtools2/blob/master/docs/content/overview.rst</a> |
| Piranha   | <a href="#">Uren et al., 2012</a>      | <a href="http://smithlabresearch.org/software/piranha/">http://smithlabresearch.org/software/piranha/</a>   |

(Continued on next page)

**Continued**

| REAGENT or RESOURCE  | SOURCE                  | IDENTIFIER  |
|--|-------------------------|---|
| Subread (featureCounts for read counting)  | Liao et al., 2013       | <a href="http://subread.sourceforge.net/">http://subread.sourceforge.net/</a>   |
| R  | R Core Team, 2014       | <a href="https://www.r-project.org/">https://www.r-project.org/</a>   |
| edgeR (R package)  | Robinson et al., 2010   | <a href="https://bioconductor.org/packages/release/bioc/html/edgeR.html">https://bioconductor.org/packages/release/bioc/html/edgeR.html</a> |
| Discriminative Regular Expression Motif Elicitation (DREME)                                  | Bailey, 2011            | <a href="http://meme-suite.org/tools/dreme">http://meme-suite.org/tools/dreme</a>   |
| Analysis of Motif Enrichment (AME)   | McLeay and Bailey, 2010 | <a href="http://meme-suite.org/tools/ame">http://meme-suite.org/tools/ame</a>   |
| Find Individual Motif Occurrences (FIMO)   | Grant et al., 2011      | <a href="http://meme-suite.org/tools/fimo">http://meme-suite.org/tools/fimo</a>   |
| CISBP-RNA Database: Catalog of Inferred Sequence Binding Preferences of RNA binding proteins | Ray et al., 2013        | <a href="http://cisbp-rna.ccb.utoronto.ca/index.php">http://cisbp-rna.ccb.utoronto.ca/index.php</a>   |
| lme4 (R package)   | Bates et al., 2015      | <a href="https://cran.r-project.org/web/packages/lme4/index.html">https://cran.r-project.org/web/packages/lme4/index.html</a>               |
| multcomp (R package)   | Hothorn et al., 2008    | <a href="https://cran.r-project.org/web/packages/multcomp/index.html">https://cran.r-project.org/web/packages/multcomp/index.html</a>       |
| car (R package)  | Fox and Weisberg, 2019  | <a href="https://cran.r-project.org/web/packages/car/index.html">https://cran.r-project.org/web/packages/car/index.html</a>                 |
| limma (R package)  | Ritchie et al., 2015    | <a href="https://bioconductor.org/packages/release/bioc/html/limma.html">https://bioconductor.org/packages/release/bioc/html/limma.html</a> |
| ImageJ   | NIH                     | <a href="https://imagej.nih.gov/">https://imagej.nih.gov/</a>   |
| lsmmeans (R package)   | Lenth, 2016             | <a href="https://cran.r-project.org/web/packages/lsmmeans/">https://cran.r-project.org/web/packages/lsmmeans/</a>                           |

**RESOURCE AVAILABILITY**

**Lead Contact**

Requests for information, resources, and reagents should be directed to Joseph D. Dougherty ([jdougherty@genetics.wustl.edu](mailto:jdougherty@genetics.wustl.edu)).

**Materials Availability**

Materials are available as found above in the [Key Resources Table](#). Plasmids generated in this study (pmrPTRE-AAV, pmrPTRE-AAV with CELF6 target UTR library), please see [Lead Contact](#) above. Raw sequencing data, as well as processing and analysis code, can be found under [Data and Code Availability](#).

**Data and Code Availability**

All software packages can be found in [Key Resources Table](#). Raw and processed sequencing data from CLIP-Seq can be accessed through the National Center for Biotechnology Information's Gene Expression Omnibus under accession number GSE118623. UCSC Genome Browser session showing CLIP-Seq data associated with NCBI GEO Accession GSE118623 can be viewed following this link:

<http://genome.ucsc.edu/cgi-bin/hgTracks?>

[hgS\\_doOtherUser=submit&hgS\\_otherUserName=mrieger&hgS\\_otherUserSessionName=celf6\\_clip\\_mm10](http://genome.ucsc.edu/cgi-bin/hgTracks?hgS_doOtherUser=submit&hgS_otherUserName=mrieger&hgS_otherUserSessionName=celf6_clip_mm10)

Code used to call peaks, perform MEME-SUITE motif enrichment clustering, and analyze the PTRE-Seq library data can be found at: <https://github.com/clevermizo/CELF6CLIPMS>

**EXPERIMENTAL MODEL AND SUBJECT DETAILS**

**Animal Models**

All protocols involving animals were approved by the Animal Studies Committee of Washington University in St. Louis. Cages were maintained by our facility on a 12 hr: 12 hr light:dark schedule with food and water supplied ad libitum.

**Genotyping**

Genotyping of all mice was performed using a standard protocol. Animals were genotyped from toe clip tissue lysed by incubation at 50°C in Tail Lysis Buffer for 1 h to overnight (0.5 M Tris-HCl pH 8.8, 0.25 M EDTA, 0.5% Tween-20) containing 4 μL/mL 600 U/mL Proteinase K enzyme (EZ BioResearch), followed by heat denaturation at 99°C for 10 min. 1 μL Crude lysis buffer was used as template for PCR with 500 nM forward and reverse primers, as specified, using 1X Quickload Taq Mastermix (New England Biolabs) with

the following cycling conditions: 94°C 1 min, (94°C 30 s, 60°C 30 s, 68°C 30 s) x 30 cycles, 68°C 5 min, 10°C hold. All referenced PCR primer sequences listed below are found in [Table S9](#).

#### Mice for CLIP

Mice used in CLIP experiments derived from 4 litters of CELF6-HA/YFP x C57BL6/J crosses. Each sample used for sequencing was generated by pooling tissue from 3-4 CELF6-HA/YFP+ animals or WT animals from each litter. Pooling tissue was required to generate sufficient material for successful CLIP library generation. Genotyping was performed using HA-F/YFP-R pair (presence or absence of HA-YFP cassette) and Actb-F/Actb-R pair (internal PCR control). Genotyping was performed on post-natal day 7 to select animals for pools, and then performed again on tissue collected after sample processing on post-natal day 9 to confirm genotypes. Sex of animals was noted but not controlled in statistical analysis of CLIP-Seq data as all pools contained animals of both sexes.

(1) Litter 1: 8 total animals.

CELF6-HA/YFP Pool 1: 4 animals: 1 male and 3 females.

CTL Pool 1: 4 animals: 3 males and 1 female.

(2) Litter 2: 8 total animals.

CELF6-HA/YFP Pool 2: 4 animals: 2 males and 2 females.

CTL Pool from this litter was excluded because upon confirmatory regenotyping it was determined that 1 CTL animal in the pool was actually CELF6-HA/YFP+.

(3) Litter 3: 6 total animals.

CELF6-HA/YFP Pool 3: 3 animals: 2 females and 1 male.

CTL Pool 2: 3 animals: 2 males and 1 female.

(4) Litter 4: 8 total animals.

CELF6-HA/YFP Pool 4: 4 animals: 2 females, 2 males.

CTL Pool 3: 4 animals: 3 males, 1 female.

#### Mice for Agilent Expression microarray

8 Celf6<sup>+/+</sup> (WT) (3 males and 5 females) and 8 Celf6<sup>-/-</sup> (KO) animals (4 males and 4 females) derived from 13 litters of Celf6<sup>+/+</sup> X Celf6<sup>+/+</sup> crosses were used to generate tissue for the RNA microarray experiment used to assay expression of CELF6 CLIP targets. Animals were genotyped as above using Celf6<sup>genoF</sup>/Celf6<sup>genoR</sup> primer pair and age when tissue was harvested ranged between 3.5-9 months.

#### Cell Culture

SH-SY5Y neuroblastoma cells (ATCC CRL-2266) were maintained at 5% CO<sub>2</sub>, 37°C, 95% relative humidity in 1:1 Dulbecco Modified Eagle Medium/Nutrient Mixture F-12 (DMEM/F12 GIBCO) supplemented with 10% fetal bovine serum (FBS Sigma). Under maintenance conditions, cells were also incubated with 1% Penicillin-streptomycin (Thermo), but antibiotics were not used during transient transfections. Cell passage was performed with 0.25% Trypsin-EDTA (Thermo).

## METHOD DETAILS

### CLIP

Our CLIP procedure is modeled after the procedure of [Wang et al. \(2015\)](#). Post-natal day 9 mice were euthanized by rapid decapitation, and brains were dissected. Cortices and cerebella were removed, retaining basal forebrain, striatum, diencephalon, colliculi, and hindbrain regions, which are the brain regions with highest CELF6 expression (9). Dissected tissue was flash frozen in liquid nitrogen and then powdered with a mortar and pestle cooled with liquid nitrogen and kept on dry ice in 10 cm Petri dishes until use. Crosslinking was performed using 3 rounds of 400 mJ/cm<sup>2</sup> dosage of 254 nm ultraviolet radiation, with Petri dishes on dry ice, in a Stratalinker UV crosslinker. After each round of crosslinking, powder in the dishes was redistributed to allow for even crosslinking. After crosslinking, powders were kept on wet ice and incubated with 1 mL lysis buffer (50 mM Tris-HCl pH 7.4, 100 mM NaCl, 1X cOmplete EDTA-free protease inhibitor (Sigma), 0.04 U/μL recombinant RNasin (Promega), 10 mM activated sodium orthovanadate, 10 mM NaF). Recombinant RNasin does not inhibit RNase I which was used for RNase digestion in CLIP and was added to prevent other environmental RNase activity. To obtain both cytoplasmic and nuclear fractions in the lysate, lysis buffer was supplemented with NP40 (Sigma CA630) detergent (final concentration 1%) and subjected to mechanical homogenization in a teflon homogenizer 10 times, and lysates were allowed to incubate on ice for 5 min. For RNase digestion, RNase I<sub>f</sub> (New England Biolabs) was diluted to final concentration 0.5, 0.1, or 0.05 U/mL per lysate for radiolabeling. For control radiolabeled samples (no crosslink and WT tissue immunoprecipitates), the highest (0.5 U/mL) concentration of RNase was used. For samples used for sequencing, 0.05 U/mL final concentration was used. RNase-containing lysates were incubated in a thermomixer set to 1200 RPM at 37°C for 3 min and then clarified at 20,000 xg for 20 min. 2% input lysate was saved for input samples for sequencing. Per immunoprecipitation, 120 μL of Dynabeads M280 streptavidin (Thermo) were incubated with 17 μL 1 mg/mL biotinylated Protein L (Thermo), and 36 μg each of mouse anti-EGFP clones 19F7 and 19C8 antibodies (MSKCC) for 1 h. Beads were prepared in batch for all immunoprecipitations and then washed five times with 0.5% IgG-free bovine serum albumin (Jackson ImmunoResearch) in 1X PBS, followed by three washes in lysis buffer. Clarified lysates were incubated with coated, washed beads for 2 h at 4°C with

end-over-end rotation and then washed in 1 mL of wash buffer (50 mM Tris-HCl pH 7.4, 350 mM NaCl, 1% NP-40, 0.04U/ $\mu$ L RNasin) four times, for 5 min with end-over-end rotation at 4°C. For radiolabeling experiments, 60% of washed bead volume was reserved for immunoblotting and added to 20  $\mu$ L of 1X Bolt-LDS non-reducing sample buffer (Thermo), and 40% proceeded to radioactive labeling. Beads for radioactive labeling were subsequently washed 3x200  $\mu$ L in PNK wash buffer (20 mM Tris-HCl pH 7.4, 10 mM MgCl<sub>2</sub>, 0.2% Tween-20, 2.5 U/ $\mu$ L RNasin) and then incubated with 10  $\mu$ L PNK reaction mixture (1X PNK reaction buffer (New England Biolabs), 4  $\mu$ Ci  $\gamma$ -<sup>32</sup>P-ATP (Perkin Elmer), 10 U T4 PNK (New England Biolabs)) for 5 min, at 37°C. After labeling, samples were washed in 3x200  $\mu$ L PNK wash buffer to remove unincorporated label, and then added to 10  $\mu$ L of 1X Bolt LDS non-reducing sample buffer. All samples in sample buffer were heated for 10 min at 70°C and then separated on a 4%–12% gradient NuPAGE Bis/Tris gels (Thermo) and then transferred to PVDF membranes with 10% methanol for 6 h at constant 150 mA. Samples for immunoblot were blocked for 1 h in block solution (5% nonfat dried milk in 0.5% Tween-20/1X TBS), and then overnight with 1:1000 chicken anti-GFP antibodies (AVES) with rocking at 4°C. Blots were washed 3 times for 5 min in 0.5% Tween20/1X TBS and then incubated with 1:5000 anti-chicken HRP secondary antibodies (AVES) for 1 h at room temperature and treated with Biorad Clarity enhanced chemiluminescence reagents for 5 min and chemiluminescent data acquired with a Thermo MyECL instrument. Radioactive signal was acquired using an Amersham Typhoon Imaging System and a BAS Storage Phosphor screen (GE Healthcare Life Sciences).

### CLIP-Seq Sequencing Library Preparation

For CLIP-Seq, EGFP immunoprecipitated WT and HA-YFP+ tissue and 2% input samples were purified from PVDF membranes as follows. Membrane slices were cut with a clean razor according to the diagram in [Figure 1A](#), from unlabeled samples as has been performed in eCLIP (14). PVDF membrane slices were incubated in 1.7 mL microcentrifuge tubes with 200  $\mu$ L Proteinase K buffer (100 mM Tris-HCl pH 7.4, 50 mM NaCl, 10 mM EDTA, 1% Triton X-100) containing 40  $\mu$ L of 800 U/mL Proteinase K (NEB) and incubated in a horizontal shaker at 250 RPM, 37°C for 1 h. Horizontal shaking reduces the need to cut the membrane into small pieces per sample which is seen in many protocols, and 1% Triton X-100 in the Proteinase K buffer facilitates increased yield from the membrane by preventing binding of Proteinase K to the membrane. 200  $\mu$ L of fresh 7M Urea/Proteinase K buffer was then added to slices, and tubes were incubated an additional 20 min with horizontal shaking at 250 RPM, 37°C. RNA was purified by addition of 400  $\mu$ L of acid phenol/chloroform/isoamyl alcohol and shaken vigorously for 15 s and allowed to incubate 5 min on the bench. RNA samples were centrifuged at 20,000 xg for 10 min. Aqueous layers were purified using a Zymo RNA Clean and Concentrator-5 column. Output from CLIP RNA samples was estimated for total concentration using an Agilent Bioanalyzer, and approximately 0.5 ng of RNA was used to prepare next generation sequencing libraries. The full protocol for sequencing library preparation is given in [Methods S1](#). Although this protocol is based on eCLIP, [Methods S1](#) is generalized for any RNA-Seq preparation.

### Total RNA-Seq of SH-SY5Y cells

For preliminary RNA-Seq of SH-SY5Y cells, eight replicate sub-confluent (approximately 80%) 10 cm dishes (TPP) were harvested by dissociation from the plate by pipetting and centrifugation at 500 x g, 5 min at room temperature. Pellets were lysed according to TRAP protocol lysis conditions ([Methods S2](#)), and RNA was purified using Trizol LS, followed by treatment for 15 min at 37°C with DNase I (NEB) and cleanup using Zymo RNA Clean and Concentrator-5 columns (Zymo Research). RNA samples were assessed by Agilent TapeStation with RIN<sub>v</sub> values between 8-10. RNA was fragmented and prepared into libraries for next generation sequencing using [Methods S1](#) as for CLIP-Seq samples.

### Total RNA-Seq and CLIP-Seq Raw Data Processing

As all total CLIP-Seq and Total RNA-Seq samples were prepared using a unified library preparation procedure, raw data processing used the same set of methods and tools (see [Key Resources Table](#)) for strand-specific quantification of sequencing reads containing unique molecular identifiers (See [Methods S1](#)) to collapse amplification duplicates. Briefly, CLIP libraries were sequenced in a 2 x 40 paired-end mode on an Illumina Next-Seq. Unique Molecular Identifier (UMI) sequences were extracted from Illumina Read 2, and reads were trimmed for quality using Trimmomatic. Using STAR, remaining reads were aligned to ribosomal RNA, and unalignable reads corresponding to non-rRNA were aligned to the mm10 mouse reference genome and assembled into BAM formatted alignment files. BAM files were annotated with UMI information using the FGBio Java package. PCR duplicates assessed by their UMIs were removed from the BAM files using Picard.

### Peak Calling and Read Counting

To call peaks genome-wide using Piranha, the genome was windowed into 100 bp contiguous windows with 50% overlap using Bedtools, and strand-specific read counts were determined for each window. To ensure that called peaks would have the same boundaries for each replicate sample, peaks were called on a merged BAM file across all YFP-HA+ immunoprecipitated CLIP samples. Background counts for Piranha were estimated on a gene-by-gene basis to control for differences in overall level of expression. Piranha p values for significant peaks were adjusted for multiple testing using Benjamini-Hochberg, and all peaks called with False Discovery Rate < 0.1 were kept for further analysis and stored as a Gene Transfer Format (GTF) file. In practice we found that Piranha peaks varied widely, with some peaks called with widths on the order of several kilobases despite having a clear local maximum. To more narrowly count reads near peak maxima in a consistent way across all peaks, Piranha peak boundaries were truncated to a

width of 100 base pairs around the peak maximum. As an alternate approach for discovery, we also wanted to consider total read counts across subgenic annotated regions (5' UTR, 3'UTR, coding sequence (CDS) and introns), for which annotations were retrieved from the UCSC Genome Browser, in the event that CELF6 bound promiscuously without much sequence specificity though perhaps restricted to particular region of a gene. In order to ensure correct mapping of reads to splice sites, the table of intron annotations was allowed to overlap the surrounding exons by 10 bases. In the end, this procedure produced 5 GTF files:

- Genome-wide peaks calls (100bp around called maxima from merged BAM file)
- 3'UTR
- CDS
- 5'UTR
- Introns

Each of these GTFs was used as a template for strand-specific feature counting in individual samples using the featureCounts program in the Subread package. For sums across regions (e.g., 3'UTRs) that showed enrichment in CLIP (see [Quantification and Statistical Analysis](#) below), we repeated the peak calling procedure in significantly enriched regions to determine whether any peaks missed in the genome-wide calling could be identified locally in enriched regions. We found most significantly enriched regions, such as 3'UTRs, had clearly definable peaks which were not determined in genome-wide peak calling. This may be due to misestimation of correct background in genome-wide calling. In genome-wide calling, background was estimated on a per-gene basis, but it may be that peaks are still missed because a transcript model is needed for most accurate calling. Determining enriched regions and then calling peaks on a small, localized region, seemed to ameliorate this issue.

SH-SY5Y Total RNA-seq samples were counted based on a GTF of all UCSC-annotated gene exons to derive total gene counts.

### CLIP Motif Enrichment Analysis

50 bp regions under the maxima of peaks in CLIP target 3'UTRs were used in MEME Suite and compared to 50 bp regions sampled randomly from the 3'UTRs of non-targets – genes which exhibited 0 or negative fold enrichment in YFP-HA+ CLIP samples compared to input and WT controls. The AME tool was used to search motif enrichment against the CISBP-RNA database ([Ray et al., 2013](#)), using the maximum match score with the ranksum test.

The position probability matrices (PPMs) for CISBP-RNA motifs showing significant enrichment in CLIP 3'UTR peaks were hierarchically clustered ([Figures 2B and 2C](#)) by using vectorized PPMs for each motif and the *hclust()* function in R (using the “complete” method). To determine whether or not there were significant matches to a cluster, rather than an individual CISBP-RNA motif PPM, average PPMs were computed by averaging the PPMs across all cluster members. These averaged PPMs were then used with the MEME-Suite FIMO tool to determine whether individual peaks had significant matches to each of the clusters ([Tables S3 and S4](#)).

### PTRE-Seq Reporter Library Preparation

All oligos used for library preparation are provided in [Key Resources Table](#). We generated the pmrPTRE-AAV backbone from an existing mtdTomato construct by PCR amplification and subcloning of the following elements: CMV promoter and a T7 promoter, PCR amplified and subcloned into the MluI restriction site of pQC membrane TdTomato IX. CMV and T7 promoters were amplified from pcDNA3.1 using pCMV\_T7-F/ pCMV\_T7-R and Phusion polymerase (NEB). Then, in order to add a NheI-KpnI restriction enzyme cassette into the 3'UTR, the entire pmrPTRE-AAV plasmid was amplified (pmrPTRE-AAV\_Full\_F/R) and recircularized using Infusion HD (Clontech). The correct backbone sequence of pmrPTRE-AAV was confirmed by Sanger sequencing.

Originally, 462 sequences of 120 bp were considered for cloning into the library across significant genes. These included 295 peaks in high confidence targets (meeting enrichment against both input and non-specific IP controls, see [Quantification and Statistical Analysis](#)), and 167 peaks in lower confidence targets showing enrichment only over input. Mutations to motifs found by AME were made as follows. The location of significant matches to motifs were determined using FIMO. Next, for each matched motif, the PPM representing this motif in CISBP-RNA was used to determine the choice of mutation at each base. Bases showing PPM probability of 0.8 or greater were mutated to the base showing the minimum value of PPM at that position. If all other three bases showed equal probability, a base was randomly selected from the three. The procedure was repeated at each position for each motif match, before moving to the next matching motif. Where motifs overlapped, lower ranking motifs (based on FIMO score) did not override mutations already made based on higher ranking motifs. After completion, this generated a set of 473 mutant elements which ranged between 1 - 25 mutated nucleotides. Subsequently, sequences were scanned for poly A signals and restriction enzyme sites which would interfere with cloning (NheI, KpnI), which were removed. A final set of 410 sequences (254 high confidence, 156 meeting only input control criterion) and their paired mutant sequences were synthesized by Agilent Technologies with 120bp of UTR sequence in addition six unique 9 bp barcodes per sequence and priming sites for amplification and cloning (final length 210 bp). These sequences and their mutated forms are given in [Table S5](#).

Obtained synthesized sequences were amplified with 4 cycles of PCR using Phusion polymerase with primers GFP-F and GFP-R. We selected these priming sites as these are standard primers in our laboratory used for genotyping that result in robust amplification. The library was PAGE purified and concentration of recovered library was estimated by Agilent TapeStation. The library was

digested with NheI and KpnI enzymes and ligated into pmrPTRE-AAV with T4 Ligase (Enzymatics). In order to ensure high likelihood of obtaining all library elements, we prepared our plasmid pool from approximately 40,000 colonies.

### PTRE-Seq Reporter Library Transfection

His/Xpress-tagged CELF3,4,5,&6 were obtained from the laboratory of Thomas Cooper. For the four-plasmid experiments, 2500 ng containing equimolar amounts of two His/Xpress-CELF constructs, an EGFP-RPL10a construct, and the CELF6 PTRE-Seq library were prepared with Lipofectamine 2000 in Opti-MEM I (GIBCO). For three-plasmid experiments, remaining mass was substituted with empty pcDNA3.1-His. SH-SY5Y cells were trypsinized and incubated in 10 cm dishes with Lipofectamine/DNA complexes overnight in DMEM/F12 supplemented with 10% FBS. The following day media was replaced with fresh DMEM/F12 supplemented with 10% FBS. Cells were pelleted for Translating Ribosome Affinity Purification (TRAP) and total RNA extraction 40 h post-transfection. TRAP and total RNA extraction were performed according to the [Methods S2](#). RNA quality was assessed by Agilent TapeStation, and all samples had RIN<sub>e</sub> values > 8. The procedure for TRAP in [Methods S2](#) is based on ([Heiman et al., 2008](#)) with additional modifications that have been optimized in our laboratory. Five replicates per condition were generated in batches balanced for all conditions. In each case, replicates were transfected from newly thawed aliquots of cells passaged once before transfection to control for cell passage. Read counts from one batch were found to cluster separately from all others after sequencing, and data from this batch were excluded. The final data were analyzed from four replicates per condition.

### PTRE-Seq Sequencing Library Preparation

PTRE-Seq sequencing libraries were prepared by cDNA synthesis using pmrPTRE-AAV antisense oligo for library specific priming, and Superscript III Reverse Transcriptase (Thermo) according to the protocol shown in [Methods S3](#). After cDNA synthesis, cDNA libraries were enriched with PCR using Phusion polymerase (Thermo), and pmrPTRE-AAV antisense and sense oligos using 18 cycles. In parallel, plasmid pool DNA was also amplified for sequencing the original plasmid pool. Purified PCR products were digested with NheI and KpnI enzymes and ligated to 4 equimolar staggered adapters to provide sequence diversity for sequencing on the NextSeq. Ligated products were amplified with Illumina primers as in CLIP-Seq library preparation ([Methods S1](#)) and subjected to 2x40 paired-end next generation sequencing on an Illumina NextSeq.

### RT-PCR and qRT-PCR

To confirm exogenous expression of CELF constructs used in PTRE-Seq experiments, 20 ng of total RNA was converted into cDNA using qScript cDNA synthesis kit (Quantabio, kit employs a mix of both oligo-dT and random hexamer priming) and diluted 4-fold. For RT-PCR, 4  $\mu$ L of diluted cDNA was used in PCR reactions with 500 nM forward and reverse primers, ([Figure S2](#); [Table S9](#)), using 1X Quickload Taq Mastermix (New England Biolabs) with the following cycling conditions: 94°C 1 min, (94°C 30 s, 60°C 30 s, 68°C 30 s) x 25 cycles, 68°C 5 min, 10°C hold, and then separated by 2% agarose and stained with ethidium bromide. For qRT-PCR, 4  $\mu$ L of diluted cDNA was combined with 500 nM His/Xpress-pcDNA-F and His/Xpress-pcDNA-R primers or HsActb-F/HsActb-R, and 1X PowerUP SYBR Green Master Mix (Thermo). Each sample/primer combination was run in 4 technical replicates on a Viiia7 Real Time PCR System (Thermo) using the following cycling program: 50°C 2 min, 95°C 2 min (95°C 1 s, 60°C 30 s) x 40 cycles, followed by dissociation step: 95°C 15 s 1.6°/s ramp, 60°C 1 min 1.6°/s ramp, 95°C 15 s 0.15°/s ramp. Samples were run alongside no template and no reverse transcription controls to ensure reactions were free of non-target contamination, and dissociation curves were inspected to ensure the absence of non-target amplicons.  $C_T$  values for each sample were averaged across technical replicates and transformed by first computing the  $\Delta C_T = C_T^{\text{His/Xpress}} - C_T^{\text{HsActb}}$ , and then computing relative log<sub>2</sub> expression (“ $\Delta\Delta C_T$ ”):  $= -(\Delta C_T^{\text{sample}} - \Delta C_T^{\text{reference}})$ , where the reference was taken to be the average  $\Delta C_T$  across the CELF6 expression condition.

### PTRE-Seq Reporter Validation

For validation of individual library element reporters ([Figure 4](#)), CELF6 CDS was subcloned from His/Xpress-CELF6 into pEGFP-C1 using EcoRI and BamHI restriction sites. Individual Vat1l, Hap1, Peg3, Epm2aip1, and Rab18 reference and mutant sequences were synthesized using Integrated DNA Technologies (IDT) gBlocks Gene Fragments and cloned into pmrPTRE-AAV with NheI and KpnI as above.

50 ng equimolar library element reporters with either pEGFP-C1 (EGFP alone) or EGFP-CELF6 was transiently transfected into SH-SY5Y cells in 96-well plates (TPP). 40 h post-transfection, media was removed and replaced with warm PBS (1.8 mM KH<sub>2</sub>PO<sub>4</sub>, 10 mM Na<sub>2</sub>HPO<sub>4</sub>, 2.7 mM KCl, 137 mM NaCl, pH 7.4). tdTomato and GFP fluorescence were determined by BioTek Instruments Cytation 5 Cell Imaging Multi-Mode Reader with internal temperature maintained at 37°C. 96-well plates were prepared in four replicate batches, where each 96-well plate contained one replicate well for all 20 conditions (five reference, five mutant reporters, in both EGFP-CELF6 and EGFP-only transfections). Log<sub>2</sub>-transformed fluorescence intensity measurements were z-score normalized on each plate to account for batch-to-batch differences in transfection efficiency and fluorescence intensity. Data in [Figure 4](#) are shown further normalized to the average value for each reporter in the EGFP-only, reference sequence condition. Example epifluorescent images were obtained using a Leica DMI3000 B microscope with 20X magnification. Monochromatic images were acquired with QCapture software (QImaging), using gain = 1, offset = 1, exposure = 205 ms for both red and green fluorescent filter sets. 16-bit grayscale images were converted to RGB color and minimally brightness-adjusted for presentation using Adobe Photoshop CS2.

### Agilent Gene Microarray

Brains from eight WT and eight Celf6 mutant mice (see) were extracted, frozen in liquid nitrogen, and crushed into a fine powder, from which RNA was extracted using QIAGEN RNEasy columns on a Qiacube robot, following manufacturer's protocol. RNA was DNase I treated, and RNA quantity and integrity were confirmed using Agilent Bioanalyzer. cDNA was prepared and chemically labeled with Kreatech ULS RNA labeling kit (Kreatech Diagnostics) and Cy5-labeled cDNAs were hybridized to Agilent Mouse v2 4x44K microarrays (G4846A-026655). Hybridization of the labeled cDNAs was done in Agilent 2x gene expression hybridization buffer, Agilent 10x blocking reagent and Kreatech Kreamblock onto Agilent 4x44K V2 microarrays at 65°C for 20 min. Slides were scanned on an Agilent C-class Microarray scanner. Gridding and analysis of images was performed using Agilent Feature Extraction V11.5.1.1.

### FGF13 and FOS Immunofluorescence

3 Celf6<sup>+/+</sup> (WT) (2 males and 1 female) and 3 Celf6<sup>-/-</sup> (KO) animals (2 males and 1 female) derived from 2 litters of Celf6<sup>+/+</sup> × Celf6<sup>-/-</sup> crosses were used to generate tissue for immunofluorescence. Mice were perfused with PBS followed by 4% paraformaldehyde (PFA) in PBS. Brains were dissected then post-fixed overnight in 4% PFA, cryoprotected in increasing concentrations of sucrose over 48 h, and frozen in blocks of OCT (Tissue-Tek), sectioned on a cryostat to 40 microns, and stored at 4°C in PBS/0.25% Sodium Azide until use. Sections were blocked with 5% serum for 1 h at room temperature, immersion-stained with primary antibodies overnight at room temp, washed 3x with PBS, then incubated with appropriate Alexa Fluor-labeled secondary antibodies for 90 min, counterstained with DAPI, and sections mounted on slides. Primary antibodies included rabbit anti-FGF13 (a gift from Geoffrey Pitt) and mouse anti-TH. Primary antibodies for FOS stain included goat anti-cFos, and mouse anti-TH (see Key Resources Table). Immunofluorescent imaging was conducted by an experimenter blind to genotype, focusing on the locus coeruleus (as identified by TH staining) using a Zeiss LSM 510 laser-scanning confocal microscope and accompanying software.

## QUANTIFICATION AND STATISTICAL ANALYSIS

### Defining CELF6 CLIP Targets by Differential Expression Analysis

Currently there is no standard statistical approach for identification of targets from CLIP data. Methods typically include clustering aligned sequences within individual CLIP RNA samples or replicate averages, with varying probabilistic modeling approaches for assessing signal-to-noise in read density, but rarely take into account variance across replicates or differential abundance compared to control samples (Ascano et al., 2012; Corcoran et al., 2011; Kishore et al., 2011; Lebedeva et al., 2011; Memczak et al., 2013; Moore et al., 2014; Mukherjee et al., 2011; Sugimoto et al., 2012; Ule et al., 2003; Wang et al., 2015; Zhang and Darnell, 2011). Furthermore, CLIP studies frequently fail to account for differences in the starting abundance of possible target mRNAs, thus reported targets are frequently biased toward highly expressed genes (Ouwenga and Dougherty, 2015). Here, we adopted a strategy of counting reads (as described above) and using standard differential expression analysis tools (edgeR) to make statistical inference on counted features in target immunoprecipitated samples compared to controls, with the hypothesis being that true targets will be enriched in target HA-YFP CLIP samples over both WT controls (representing non-specific immunoprecipitation pull down) and input samples (accounting for differences in starting abundance of possible target mRNAs). WT controls are a key recommendation in discussion of CLIP methods using engineered tags (Van Nostrand et al., 2017).

Samples counted using Subread were imported into R using edgeR. To be included in differential analysis, counted regions were required to possess a minimum of 10 read counts in at least 3 samples. Differential testing was then performed in edgeR against read counts deriving from WT samples or YFP-HA+ input samples. We defined CLIP targets as those having positive fold change enrichment in YFP-HA+ CLIP samples compared to both WT CLIP and input samples, with edgeR p values < 0.05 from the edgeR two sample negative binomial exact test. To test less stringent cut-offs, we also examined some targets meeting p < 0.05 for just the YFP-HA versus Input as noted in figure legends. Target peaks and gene information are found in Table S2. Analysis of enrichment for gene ontology terms was performed in DAVID (Huang et al., 2009a, 2009b) using genes of known expression in the brain as the background list rather than the entire mouse genome.

### PTRE-Seq Barcode Counting and Normalization

Barcode counts from sequencing read FASTQ files for each element were determined using a Python script. Read counts were imported into R using edgeR and converted to counts-per-million (CPM) to normalize for differences in library size. Elements showing no counts in the DNA plasmid pool sequencing were removed. Expression was then computed as:

$$\log_2 \text{expression} = \log_2 \frac{CPM_{RNA}}{CPM_{DNA}}$$

Translation efficiency was computed as:

$$\log_2 TE = \log_2 \text{expression}_{TRAP} - \log_2 \text{expression}_{TotalRNA}$$

We required that analyzed barcodes have at least 10 counts in at least four samples to be included in analysis. Additionally, we required that all elements in the library have a minimum of three out of the original six barcodes present, and present for both reference and mutant sequences.

### PTRE-Seq Linear Mixed Model Analysis

The final set of elements after filtering on expression and numbers of barcodes was 379 across 153 unique gene UTRs, comprising 229 high confidence (meeting both input and wt/non-specific IP controls) and 150 lower confidence (meeting input control only) CLIP targets. Individual linear mixed models were performed using lme4 fitting the following:

$$\log_2(\text{expression} + \text{TE}) \sim \text{condition} * \text{sequence} + (1|\text{sample})$$

where barcodes were used as repeated-measures for each sample, per element. Fixed effect terms of condition referred to either: (a) CTL or CELF6 expression for analyses in Figure 3, or (b) CTL, CELF6, CELF3, CELF4, CELF5, CELF3/6, CELF4/6, CELF5/6 for analyses related to Figure 5. The fixed effect term of sequence was either (a) reference or (b) mutated sequence. Omnibus Analysis of Deviance tests for significant effects of fixed effect terms were computed using type II ANOVA R with the car package with Satterthwaite degrees of freedom. Estimates of  $R^2$  were determined according to the procedure of Nakagawa and Schielzeth (2013) which provides a simple method for obtaining these estimates from non-linear and empirical model fits in a “percentage of variance explained”-interpretative sense. Omnibus p values for fixed effects are also reported in Tables S6, S7, and S8 for models alongside Benjamini-Hochberg adjusted False Discovery Rates for TRAP, input expression, and translation efficiency (TE) respectively.

Figures 3 and 5 show analysis of these across all individual models and for subsets of elements with smaller numbers of mutations introduced ( $\leq 8$  or  $\leq 16$ ) to discern whether radically mutating elements has exerted a strong effect. Because these subsets are nested and because this bird’s-eye view analysis is pooling independently fitted models, we have assessed significance between them using non-parametric Mann-Whitney U tests for differences between medians. Post hoc multiple comparisons significance and confidence intervals reported for individual elements in Figures 3 and 4, in order to identify the sources of interaction, were computed using the multcomp R package.

### PTRE-Seq Validation Analysis

Validation of reporters in Figure 4 was assessed using the car package in R to compute Type II two-way ANOVA with main effects of sequence and CELF6 exogenous expression and their interaction, with post hoc pairwise multiple comparisons determined using multcomp.

### Agilent Expression Microarray Analysis

Features were extracted with Agilent Feature Extraction software and processed with the limma package in R to background correct (using normexp) and quantile normalize between arrays. Genotype of knockout samples was confirmed via decrease of Celf6 mRNA expression level, and differential expression analysis was conducted using limma with empirical bayes (eBayes) analysis.  $\log_2$  fold change of CLIP targets was compared to an equivalent number of randomly sampled probes by Welch’s t test. Gene Ontologies Analysis was conducted in Cytoscape using the BiNGO Module (Maere et al., 2005), with Benjamini-Hochberg correction and a display cutoff of  $p < 0.01$ . Cellular component results are displayed.

### Immunofluorescence Quantification

Positive FOS and FGF13 immunofluorescent signal was quantified using the NIH Image analysis software, ImageJ. Cell bodies and nuclei were demarcated via freehand outlining by experimenter blind to genotype based on TH and DAPI stain of cells. Intensity of signal was measured in each cell, as well as within the nucleus of each cell, and normalized by total pixel area. Fraction of area containing positive signal in each cell and within each nuclei were also measured using a standardized threshold for positive signal. Signal intensity/area as well as fraction of area containing signal for each protein was analyzed using a linear mixed effects model in R with lme4, with individual quantified regions of interest as a repeated-measure for each animal, with fixed effects of region (cytoplasm versus nucleus) and genotype (WT versus KO). Significance was determined using Type II ANOVA with Satterthwaite degrees of freedom using the car package in R, with post hoc multiple pairwise comparisons determined using the multcomp package.



**Cell Reports, Volume 33**

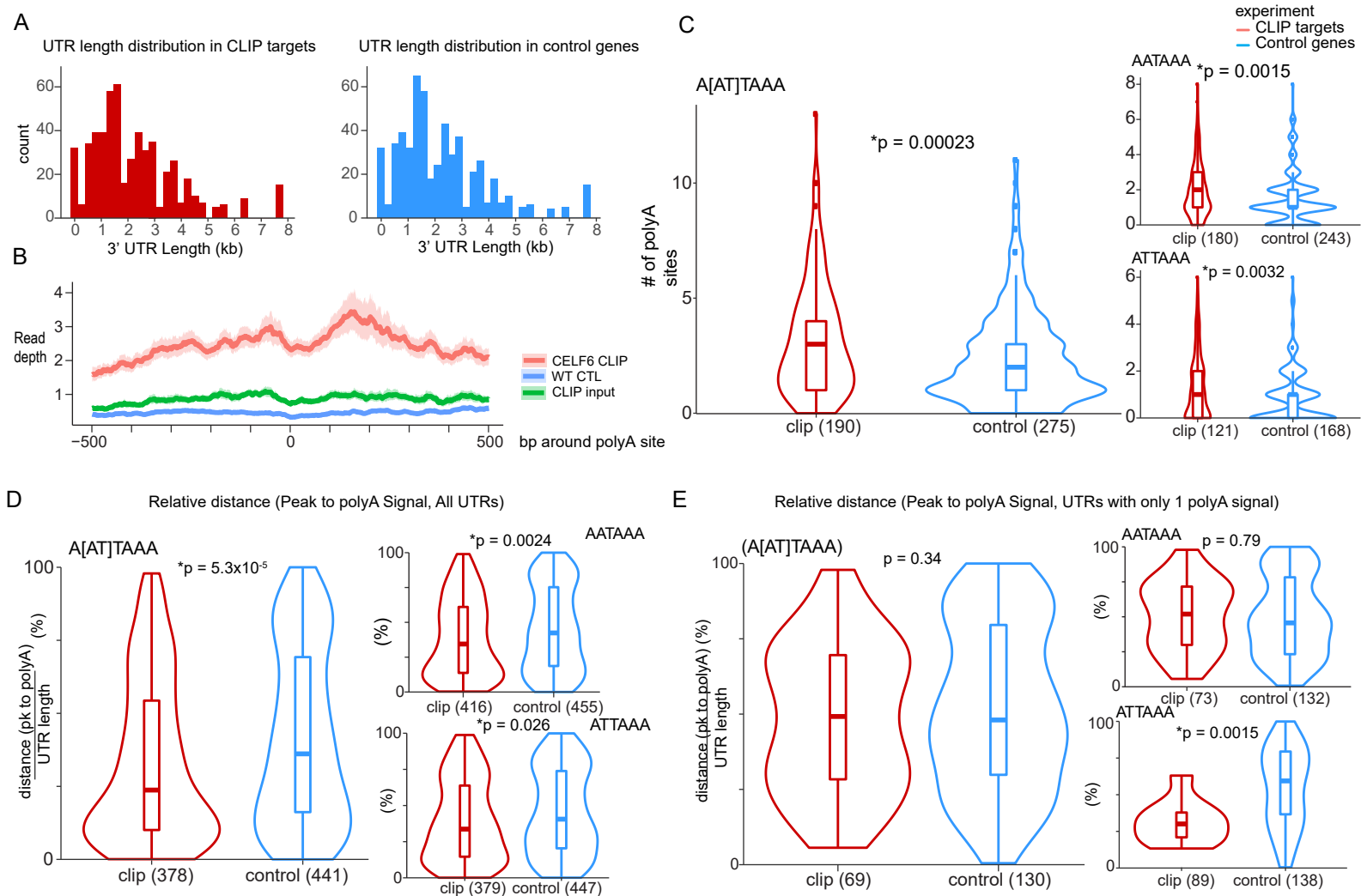
**Supplemental Information**

**CLIP and Massively Parallel Functional Analysis of  
CELF6 Reveal a Role in Destabilizing Synaptic Gene  
mRNAs through Interaction with 3' UTR Elements**

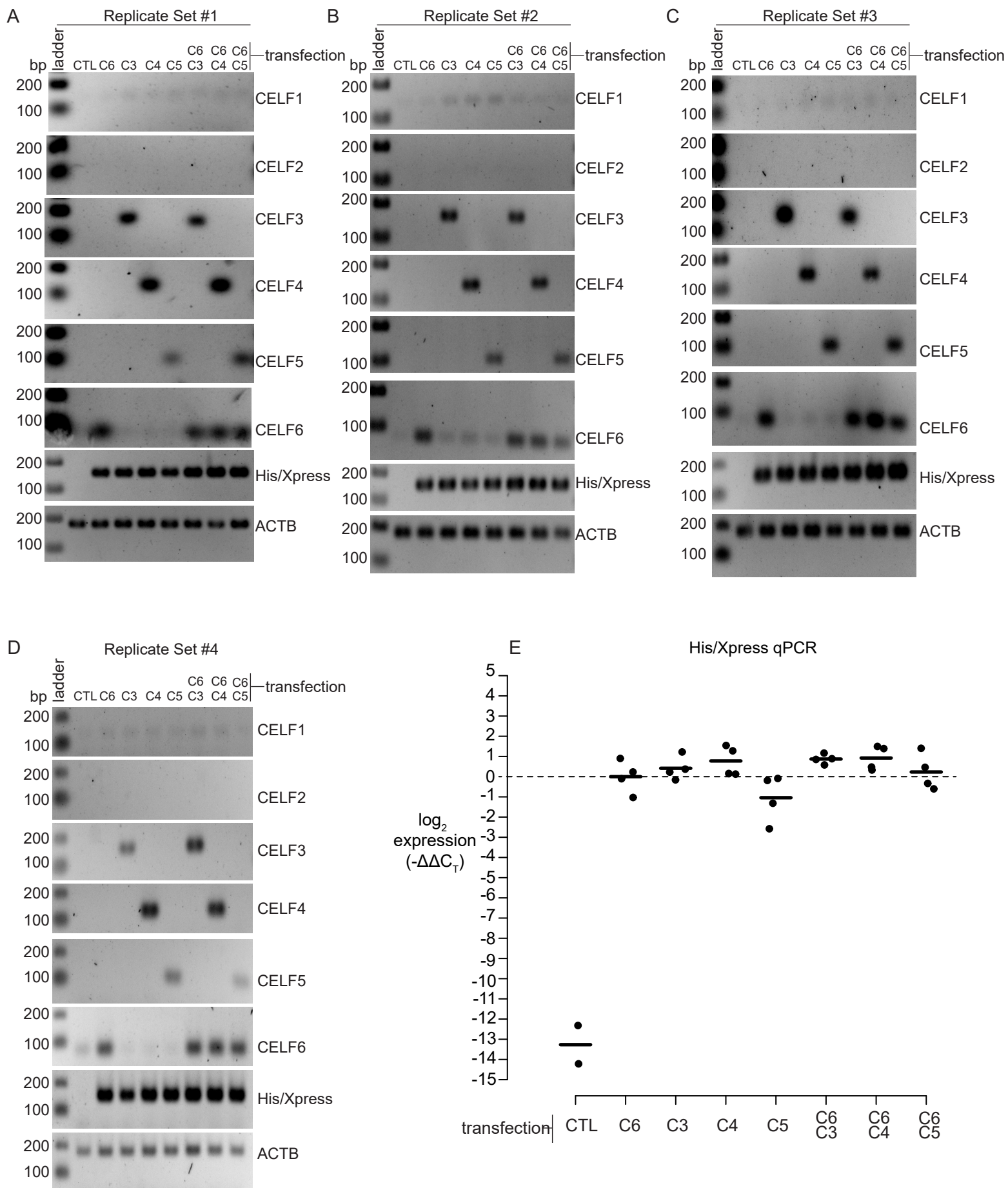
**Michael A. Rieger, Dana M. King, Haley Crosby, Yating Liu, Barak A. Cohen, and Joseph D. Dougherty**

| Sample      | Genotype | Fraction | total pairs (millions) | % duplication | unique aligners (millions) | 3UTR  | CDS   | 5UTR  | introns |
|-------------|----------|----------|------------------------|---------------|----------------------------|-------|-------|-------|---------|
| yfp-1-ip    | yfp      | ip       | 9.9                    | 6.59          | 5.9                        | 19.2  | 25.53 | 4.5   | 50.77   |
| wt-1-ip     | wt       | ip       | 15.6                   | 4.64          | 9.7                        | 15.53 | 27.98 | 4.91  | 51.57   |
| yfp-1-input | yfp      | input    | 4.6                    | 6.15          | 2.7                        | 12.4  | 20.47 | 12.58 | 54.55   |
| wt-1-input  | wt       | input    | 4.9                    | 6.96          | 2.7                        | 13.5  | 19.15 | 11.38 | 55.97   |
| yfp-2-ip    | yfp      | ip       | 6.9                    | 4.48          | 4.1                        | 14.7  | 27    | 10.2  | 48.1    |
| wt-2-ip     | wt       | ip       | 3.2                    | 5.14          | 1.6                        | 10.86 | 27.8  | 10.28 | 51.06   |
| yfp-2-input | yfp      | input    | 29.3                   | 19.48         | 21.2                       | 11.61 | 16.12 | 4.2   | 68.06   |
| wt-2-input  | wt       | input    | 5.1                    | 3.49          | 3.9                        | 11.54 | 16.44 | 4.49  | 67.53   |
| yfp-3-ip    | yfp      | ip       | 5.8                    | 35.53         | 3.2                        | 15.12 | 26.36 | 7.16  | 51.36   |
| yfp-3-input | yfp      | input    | 10.7                   | 9.11          | 7.8                        | 15.11 | 18.69 | 4.14  | 62.06   |
| yfp-4-ip    | yfp      | ip       | 6.7                    | 26.52         | 3.3                        | 21.48 | 26.19 | 9.01  | 43.31   |
| wt-3-ip     | wt       | ip       | 14.4                   | 49.98         | 3.7                        | 15.15 | 31.49 | 7.21  | 46.15   |
| yfp-4-input | yfp      | input    | 2.2                    | 56.86         | 0.8                        | 13.43 | 17.16 | 8.36  | 61.06   |
| wt-3-input  | wt       | input    | 9.8                    | 20.42         | 6.7                        | 16.46 | 21.24 | 4.15  | 58.14   |

**Table S1: Sequencing results for CLIP-Seq samples, Related to Figure 1.** Table showing CLIP-Seq samples 1-4, CLIP-Seq input samples 1-4, and WT CTL IP samples 1-3 with: total read pairs (millions) surviving quality trimming, % duplication as estimated by unique molecular identifiers, uniquely aligning reads (millions), % of uniquely aligning reads aligning to 3'UTR, CDS, 5'UTR, or intronic subgenic regions.

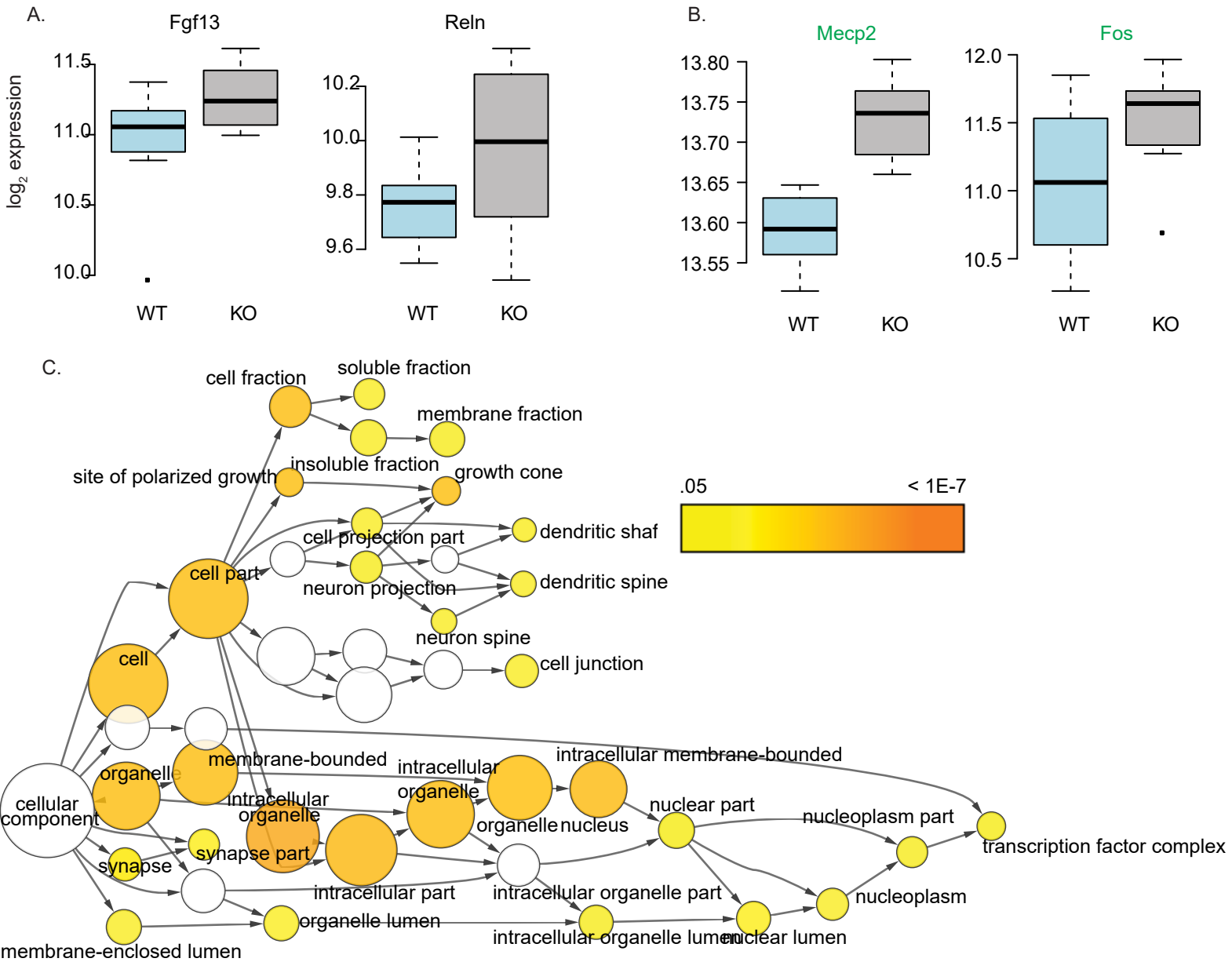


**Figure S1: Proximity of CELF6 CLIP peaks to poly-adenylation sites, Related to Figure 1.** (A) Length distribution of 3' UTRs of putative CELF6 CLIP-targets (N=544) (left, red) and length distribution of 3' UTRs from control genes showing no enrichment or depletion in CELF6 CLIP compared to immunoprecipitates from WT littermate controls and CELF6-YFP input samples (randomly selected length-matched, N=544). (B) Read depth around poly-adenylation (AATAAAA or ATTAAA) signals in CELF6 CLIP or control UTRs. (C) Number of polyA sites per UTR in CELF6 CLIP-targets or length-matched controls for either AATAAAA/AT-TAAA (left) or AATAAAA and ATTAAA separately (right). (D) Distance from CLIP peak to closest polyA site expressed as % of length of the UTR. Controls employed randomly selected coordinates in lieu of peaks. (E) Same as (D) for UTRs with only a single polyA site. Significance in C-E tested determined by t-test.



**Figure S2: Expression of His/Xpress-tagged CELF constructs in PTRE-Seq replicates, Related to Figure 3 and Figure 5.**

**Figure S2: Expression of His/Xpress-tagged CELF constructs in PTRE-Seq replicates, Related to Figure 3 and Figure 5.** CELF6 CLIP UTR element PTRE-Seq library was transiently expressed in neuroblastoma SH-SY5Y along with EGFP-RPL10a and the following constructs: pcDNA3.1 (CTL), (C6), His/Xpress-CELF3 (C3), His/Xpress-CELF4 (C4), His/Xpress-CELF5 (C5). Constructs were transfected singly or in combination with His/Xpress-CELF6, and exogenous expression was confirmed by RT-PCR with 25 cycles, using primers for CELF1-6, His-Xpress tag, or ACTB as a loading control, separated by 2% agarose, and stained with ethidium bromide. Results shown from (A) Replicate set #1, (B) Replicate set #2, (C) Replicate set #3, (D) Replicate set #4. (E) Quantitative real-time PCR (40 cycles) showing log<sub>2</sub> expression level of constructs (relative to ACTB), using the His-Xpress tag primer set across conditions, and normalized to the average of the C6 condition. 2 out of 4 CTL samples showed amplification with His/Xpress tag primers in excess of 35 cycles, the remaining 2 samples did not show any amplification. Points show individual sample values and lines show means.



**Figure S3: Gene Ontology of Up-regulated Genes in CELF6 KO brain, Related to Figure 6.** (A) & (B) expression of example high confidence (*Fgf13*, *Reln*) or lower stringency (*Mecp2*, *Fos*) CLIP targets in WT and KO mouse brain (see Figure 6). (C) Gene Ontology network of 200 most upregulated transcripts ( $p < 0.05$ ) in CELF6 KO brain tissue highlights alterations of synaptic and transcriptional genes. Shading denoted Benjamini-Hochberg adjusted p-values for enriched GO terms. Network constructed in Cytoscape using BiNGO.

| <b>Table S9: Oligonucleotides, Related to Key Resource Table.</b>         |  |  |
|---|--|--|
| <b>Mouse genotyping</b>   |  |  |
| Sequence (5'-3')  | Oligo Name   | Amplicon Size  |
| TTAAGCGTAGTCTGGGACGTCGTA<br>TGGGT   | HA-F   | ] 480 bp   |
| CTACGTCCAGGAGCGCACCATCTT<br>CTT   | YFP-R  |  |
| AGAGGGAAATCGTGCGTGAC<br>CAATAGTGATGACCTGGCCGT                             | Actb-F<br>Actb-R   | ] 100 bp   |
| CCCTGCCACCTAGCTCTTCAGGTT<br>ATGGCTGAGCTCTTTCTTGAGAAG<br>TAC               | Celf6geno-F<br>Celf6geno-R   | ] 415 bp (WT), 188 bp<br>(KO)  |
| <b>Modified eCLIP-Seq Next Generation Sequencing Library Prep</b>         |  |  |
| Sequence (5'-3')  | Oligo Name   | Ordering Specs   |
| /5Phos/rArGrArUrCrGrGrArArGrArGr<br>CrGrUrCrGrUrGrUrArG/3SpC3/            | A01m adapter   | RNase-Free HPLC<br>purification, Storage 200<br>µM in H <sub>2</sub> O                           |
| ACACGACGCTCTTCCGA   | AR17 primer  | Storage 200 µM in Tris<br>pH 7.8   |
| /5Phos/NNNNNNNNNAGATCGGAA<br>GAGCACACGTCTG/3SpC3/                         | Rand103tr3<br>adapter  | PAGE purification,<br>machine mixing for [N] <sub>10</sub> ,<br>Storage 200 µM in Tris<br>pH 7.8 |
| AATGATACGGCGACCACCGAGATC<br>TACTCTTTCCCTACACGACGCT<br>CTTCCGATC*T         | Illumina Universal<br>F  | PAGE purification,<br>Storage 100 µM in Tris<br>pH 7.8   |
| CAAGCAGAAGACGGCATAACGAGAT<br>xxxxxxxxGTGACTGGAGTTCAGAC<br>GTGTGCTCTTCCG*A | Illumina R with<br>Index ([x] <sub>9</sub> )<br>sequence for<br>sample<br>multiplexing | PAGE purification,<br>Storage 100 µM in Tris<br>pH 7.8   |
| <b>PTRE-Seq Library Generation &amp; Sequencing Library Preparation</b>   |  |  |
| Sequence (5'-3')  | Oligo Name   | Ordering Specs   |
| TGACACGCGTGTGACATTGATTAT<br>TGACTAGTTA                                    | pCMV_T7-F  | Storage 100 µM in Tris<br>pH 7.8   |
| TGACGGATCCTCCCTATAGTGAGT<br>CGTATTAATTT                                   | pCMV_T7-R  | "  |
| TAAGCTAGCCTGGTACCGGCATCC<br>CTGTGACCCCTC                                  | pmrPTRE_AAV_F<br>ull_F   | "  |
| GGTACCAGGCTAGCTTACTTGAC<br>AGCTCGTCCATGCCGTAC                             | pmrPTRE_AAV_F<br>ull_R   | "  |
| CCTACGGCGTGCAAGTGCCTCAGC<br>CGGCGAGCTGCACGCTGCGTCCT<br>C                  | GFP-F<br>GFP-R   | "<br>"   |
| GGCACTGGAGTGGCAACT  | pmrPTRE<br>antisense   | "  |
| GCATGGACGAGCTGTACAAG<br>ACACTCTTTCCCTACACGACGCTC<br>TTCCGATCTTCATGTA*C    | pmrPTRE sense<br>KpnI_overhang_1   | "<br>PAGE purification,<br>Storage 100 µM in Tris<br>pH 7.8                                      |

|  |                       |   |
|--|-----------------------|---|
| ACACTCTTTCCCTACACGACGCTC<br>TTCCGATCTCAGGTGTA*C        | KpnI_overhang_2       | “ |
| ACACTCTTTCCCTACACGACGCTC<br>TTCCGATCTGTTCTGTGTA*C      | KpnI_overhang_3       | “ |
| ACACTCTTTCCCTACACGACGCTC<br>TTCCGATCTAGCAGCTGTA*C      | KpnI_overhang_4       | “ |
| /5phos/A*TGAAGATCGGAAGAGCGT<br>CGTGTAGGGAAAGAGTGT-3    | KpnI_complement<br>_1 | “ |
| /5phos/A*CCTGAGATCGGAAGAGC<br>GTCGTGTAGGGAAAGAGTGT-3   | KpnI_complement<br>_2 | “ |
| /5phos/A*GGAACAGATCGGAAGAG<br>CGTCGTGTAGGGAAAGAGTGT-3  | KpnI_complement<br>_3 | “ |
| /5phos/A*GCTGCTAGATCGGAAGA<br>GCGTCGTGTAGGGAAAGAGTGT-3 | KpnI_complement<br>_4 | “ |
| /5phos/C*TAGAGATCGGAAGAGCA<br>CACGTCTG                 | NheI_overhang         | “ |
| CAGACGTGTGCTCTTCCGATC*T                                | NheI_complement       | “ |

**Table S9: Oligonucleotides, Related to Key Resource Table.** Oligonucleotides for genotyping, reporter library generation, and sequencing library preparation.



## Methods S1. Generalized Library Prep with rRNA Depletion & tagging with unique molecular identifiers (UMI), Related to Figure 1.

This protocol is based on the method found in:

Van Nostrand, Eric L., et al. "Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP)." *Nature methods* 13.6 (2016): 508-514.

and adapted for both CLIP-Seq & total RNA-Seq. The starting material is purified RNA.

This protocol is written presuming a strip tube format. Make sure to have a Permagen Labware strip tube magnet, strip tube mini-centrifuge, and 10  $\mu$ L 8-channel multichannel pipettor, and 300  $\mu$ L 8-channel multichannel pipettor.

### rRNA Probe Annealing

Materials: NEBNext rRNA depletion kit (E6310)

RNA should be 10ng-1 $\mu$ g in a 12 $\mu$ L volume

*The NEBNext rRNA depletion kit functions by hybridizing a proprietary mix of DNA probes complementary to rRNA. RNase H then degrades RNA:DNA hybrids, leaving unhybridized RNA intact. Afterwards, DNA probes are degraded by DNase I treatment.*

1. Per reaction assemble: 1  $\mu$ L NEBNext rRNA depletion solution 2  $\mu$ L Probe hybridization buffer 12  $\mu$ L RNA Sample
2. Vortex and spin down.
3. Thermal cycling program (with lid at 105°C) is as follows:

|         |            |
|---------|------------|
| 95°C    | 2 min      |
| 95-22°C | 0.1°C/sec* |
| 22°C    | 5 min hold |

\* On our Biorad Thermal Cyclers, we have to set the following for second stage:  
731 total cycles, 1 sec  
95C start, -0.1deg/cycle
4. Spin and place on ice.

### RNase H Digestion

1. Make a master mix. Per reaction add:

|                                   |
|-----------------------------------|
| 2 $\mu$ L NEBNext RNase H         |
| 2 $\mu$ L RNase H Reaction Buffer |
| 1 $\mu$ L H <sub>2</sub> O        |
2. Add 5  $\mu$ L to 15  $\mu$ L rRNA annealing reaction (20  $\mu$ L final volume)
3. Vortex, spin, and heat to 37°C for 30 minutes with lid at 40°C

### DNase I Digestion

1. Make a master mix. Per reaction add:

|                                   |
|-----------------------------------|
| 5 $\mu$ L DNase I Reaction Buffer |
| 2.5 $\mu$ L DNase I (RNase-free)  |
| 22.5 $\mu$ L H <sub>2</sub> O     |
2. Add 30  $\mu$ L to 20  $\mu$ L reaction (50  $\mu$ L final volume)
3. Vortex, spin, and heat to 37°C for 30 minutes with lid at 40°C

## Clean up with MyONE Silane Beads

### Materials:

MyONE Silane Beads (Thermo Scientific 37002D)

Buffer RLT (Qiagen, any RNA kit or Product # 79216)

5M NaCl, 100% EtOH, 75% EtOH, 80% EtOH (later steps in this protocol)

Strip tube magnet (Permagen Labware 0.2 mL PCR Strip Magnetic Separator)

1.7 mL tube magnet

*MyONE Silane and Agencourt RNAClean are not the same technology. MyONE Silane can purify small fragments given the right proportion of EtOH, Agencourt RNAClean bottoms out at 100 nt.*

1. Separate 20  $\mu$ L of MyONE Silane beads per sample on magnet and remove storage buffer. (For 10 samples, separate 200  $\mu$ L, etc. Use 1.7 mL tube for batch preparation.)
2. Wash beads in batch with 900  $\mu$ L of Qiagen Buffer RLT.
3. Resuspend beads in 150  $\mu$ L/sample of Qiagen Buffer RLT (3 starting sample volumes) and 5  $\mu$ L per sample 5M NaCl. (For large numbers of samples, you may need to use a 15 mL conical. For 10 samples of 50  $\mu$ L each, resuspend beads in 1500  $\mu$ L Buffer RLT + 50  $\mu$ L 5M NaCl.)
4. Split rRNA depletion reaction to two sets of strip tubes (50  $\mu$ L  $\rightarrow$  25 & 25). This is to ensure strips can accommodate total volume.
5. To each sample add. 77.5  $\mu$ L Beads, RLT, NaCl with multichannel. Mix by pipetting up and down 10 times.
6. Add 154  $\mu$ L 100% EtOH (1.5 mix volumes) to each strip tube with multichannel.
7. Mix by pipetting up and down and rotate samples at room temp for 15 minutes.
8. Separate on magnet for 30 seconds and remove supernatant.
9. Wash beads with 0.2 mL 75% EtOH. Pipette to fully resuspend and move to new strip. At this step, combine strips that were split in step 4.
10. Separate on magnet for 30 seconds. Remove wash with multichannel.
11. Wash 2 more times with 75% EtOH. Add wash buffer and let sit for 30 seconds on magnet and remove with multichannel.
12. Dry 5 minutes on magnet. Remove excess EtOH with vacuum or by pipette which may collect at bottom.
13. Resuspend in 10  $\mu$ L of H<sub>2</sub>O and let sit for 5 minutes off magnet. Then to clean up put back on magnet, separate, and move eluates to new strip tubes.

*Optional: Resuspend in >10  $\mu$ L and assess a small amount by Agilent TapeStation or Agilent Bioanalyzer to confirm loss of small (18S) and large (28S) rRNA peaks.*

*Note: Contamination with MyONE Silane beads does not appear to inhibit any downstream steps, so don't worry about a small amount of magnetic beads coming along. At the very end of library preparation, however, you do want to ensure libraries are bead-free before pooling for sequencer.*

## Fragmentation

*If doing CLIP or another prep where RNA samples are already fragmented, use mix components in step (1) and skip to dephosphorylation reaction.*

1. Per reaction assemble:
  - 1.2  $\mu$ L 10X Antarctic Phosphatase Buffer (NEB M0289)
  - 9.3  $\mu$ L rRNA-depleted sample RNA
2. Thermal cycler 94°C (lid at 105°C) for 5-15 minutes
3. Move to ice.

*We did 15 minutes starting with 1 µg into rRNA depletion and had peak fragments between 80-100 nt by TapeStation. You may want to do an experiment with trial RNA and monitor peak size of fragments with TapeStation or Bioanalyzer (or even a RNA polyacrylamide gel). Smaller amounts probably need less fragmentation time.*

## **Dephosphorylation Reaction**

*RNase I digestion in CLIP, and heat based fragmentation for total RNA-Seq both leave 3' phosphates. These must be removed before adapter ligation. Triton-X 100 is added to 1% based upon personal communication with NEB Tech Support that this improves dephosphorylase activity of T4 PNK to >90%. We have not determined empirically whether multiple dephosphorylation enzymes are really necessary.*

1. Make a master mix. Per reaction add 1.5 µL of:
  - 0.5 µL rRNasin (Promega N2511)
  - 1 µL Antarctic Phosphatase (NEB M0289)
2. Add 1.5 µL to each sample.
3. Vortex samples briefly and spin.
4. Heat to 37°C for 30 minutes with lid at 40°C.
5. Make a master mix. Per reaction add:
  - 2.5 µL T4 PNK Buffer
  - 2.5 µL 10% Triton-X 100
  - 0.5 µL T4 PNK (NEB M0201)
  - 7.5 µL H<sub>2</sub>O
6. Add 13 µL master mix to each sample (final volume 25 µL).
7. Vortex briefly and spin down.
8. Heat to 37°C for 30 minutes with lid at 40°C

## **Clean up with MyONE Silane Beads (Abbreviated, see above for full protocol)**

1. 20 µL of MyONE Silane beads per sample.
2. Wash beads in batch with 900 µL RLT.
3. Resuspend beads in 75 µL/sample of Qiagen Buffer RLT (3 volumes, last step is 25 µL) and 2.5 µL/sample 5M NaCl.
4. Add 77.5 µL NaCl/RLT/beads to samples & 154 µL 100%EtOH
5. Mix and rotate samples at room temp for 15 minutes.
6. Separate on magnet for 30 seconds and remove supernatant.
7. Wash with 0.2 mL 75% EtOH and move to new strip.
8. Wash 2 more times with 0.2 mL 75% EtOH.
9. Dry 5 minutes on magnet.
10. Remove residual EtOH.
11. Resuspend in 9~9.5 µL of H<sub>2</sub>O and let sit off magnet for 5 minutes.
12. Separate on magnet and move 8.5 µL eluate to strip tubes containing A01m adapter. (See next section).

*Optional: Resuspend in >9 µL and assess a small amount by Agilent TapeStation to confirm size shift as a result of fragmentation. Elution volume is slightly larger than 8.5 µL to ensure you can move volume safely to next set of tubes.*

## A01m Ligation

A01m Adapter:

/5Phos/rArGrArUrCrGrGrArArGrArGrCrGrUrCrGrUrGrUrArG/3SpC3/

(IDT: Purify at 250 nmol RNA Oligo scale using RNase-Free HPLC Purification, store at 200  $\mu$ M in H<sub>2</sub>O, aliquoted, at -80° C. SpC3 is a molecule which is a carbon chain that can be added to oligos. It blocks any ligation at its own end, since there is no 3' OH group, thus adapter chains are not possible.)

T4 RNA Ligase High Concentration (NEB M0437)

1. Strip tube should contain 9  $\mu$ L total: 8.5  $\mu$ L dephosphorylated RNA fragments & 0.5  $\mu$ L 40  $\mu$ M A01m.
2. Heat to 65°C for 2 minutes with lid at 105°C.
3. Place on ice for 1 minute.
4. Make a master mix. Per reaction add:
  - 1.5  $\mu$ L DMSO (100%)
  - 2.0  $\mu$ L RNA Ligase Buffer (10x)
  - 2.0  $\mu$ L ATP (10 mM)
  - 0.5  $\mu$ L Promega rRNasin (40 U / mL)
5. Add 6  $\mu$ L mix to each sample.
6. Vortex briefly and spin down.
7. Per reaction add 4  $\mu$ L PEG8K (50%). (Cut pipette tip for easier pipetting.)
8. Vortex briefly and spin down.
9. Add 1  $\mu$ L T4 RNA Ligase High Conc (30 U/  $\mu$ L).
10. Vortex briefly and spin down.
11. Tape down horizontally into a container and place on a shaker at 250 rpm for 2 hours at room temp.

## Clean up with MyONE Silane Beads (Abbreviated, see above for full protocol)

*NaCl is not added as there is NaCl in the RNA Ligase buffer. According to Eric Van Nostrand, the EtOH percentage is changed to favor larger fragments and not unligated adapter. My own experiments with MyONE Silane are somewhat inconclusive as to whether that matters.*

1. Separate 10  $\mu$ L of MyONE Silane beads per sample.
2. Wash beads in batch with 900  $\mu$ L of Qiagen Buffer RLT.
3. Resuspend beads in 60  $\mu$ L/sample of Qiagen Buffer RLT.
4. Add 60  $\mu$ L beads/RLT to each sample and mix.
5. Add 52.5  $\mu$ L EtOH (0.75 mix volumes) and mix.
6. Rotate samples at room temp for 15 minutes.
7. Separate on magnet for 30 seconds and remove supernatant.
8. Wash with 0.2 mL 75% EtOH. Resuspend and move to new tube.
9. Separate on magnet for 30 seconds. Wash 2 more times with 75% EtOH with resuspending (30 seconds on magnet).
10. Dry 5 minutes on magnet.
11. Remove residual EtOH.
12. Resuspend in 7.5~8  $\mu$ L of H<sub>2</sub>O and let sit for 5 minutes. Separate on magnet and move 7  $\mu$ L to a new set of strip tubes containing 1.5  $\mu$ L of 20  $\mu$ M AR17 primer.

## Reverse Transcription

AR17 primer: ACACGACGCTCTTCCGA

Order as standard primer. Store in H<sub>2</sub>O at 200  $\mu$ M at -20°C.

Working dilution is 20  $\mu$ M.

Thermo Superscript RT III First Strand Synthesis system (Thermo 18080051)

1. Strip tubes should contain 1.5  $\mu$ L AR17 (20  $\mu$ M) and 7  $\mu$ L A01m-ligated RNA
2. Heat to 65°C for 2 minutes with lid at 105°C
3. Place on ice for 1 minute.
4. Make a master mix. Per reaction add (total 11.5  $\mu$ L mix):
  - 2.0  $\mu$ L SSRTIII 10x Buffer
  - 2.0  $\mu$ L dNTPs (10 mM)
  - 4.0  $\mu$ L MgCl<sub>2</sub> (25 mM)
  - 2.0  $\mu$ L DTT (100 mM)
  - 0.6  $\mu$ L RnaseOUT
  - 0.9  $\mu$ L SSRTIII Enzyme
5. *(Old kits fail!)*
6. Add 11.5  $\mu$ L master mix to each 8.5  $\mu$ L sample (f.v. 20  $\mu$ L).
7. Vortex briefly and spin down.
8. Heat to 50°C for 45 minutes.  
*ExoSAP-It degrades primers and dNTPs, and thus only true RNA:cDNA hybrids remain intact.*
9. Per reaction add 3.5  $\mu$ L ExoSAP-It (Thermo 78200.200.UL).
10. Vortex briefly and spin down.
11. Heat to 37°C for 15 minutes.
12. Per reaction add 1  $\mu$ L EDTA (0.5M). Vortex briefly and spin down.
13. Per reaction add 3  $\mu$ L 1M NaOH. Vortex briefly and spin down.
14. Heat to 70°C for 12 minutes in thermal cycler to degrade RNA.
15. Per reaction add 3  $\mu$ L 1M HCl to neutralize pH. (Final volume is 30.5  $\mu$ L.)

### **Clean up with MyONE Silane Beads (Abbreviated, see above for full protocol)**

*Changes to EtOH added and 80% in wash are based on eCLIP protocol but not clear to me why.*

1. Separate 10  $\mu$ L of MyONE Silane beads per sample.
2. Wash beads in batch with 900  $\mu$ L of Qiagen Buffer RLT.
3. Resuspend beads in 91.5  $\mu$ L/sample of Qiagen Buffer RLT (3 starting sample volumes).
4. Add 91.5  $\mu$ L Beads in Buffer RLT to each sample and mix.
5. 111  $\mu$ L EtOH (0.91 mix volumes) and mix.
6. Rotate samples at room temp for 15 minutes.
7. Separate on magnet for 30 seconds and remove supernatant.
8. Wash with 0.2 mL 80% EtOH. Resuspend and move to new tube.
9. Separate on magnet for 30 seconds. Wash 2 more times with 80% EtOH with resuspending (30 seconds on magnet).
10. Dry 5 minutes on magnet.
11. Remove residual EtOH.
12. Resuspend in 9.5~10  $\mu$ L of H<sub>2</sub>O and let sit for 5 minutes. Separate on magnet and move 9  $\mu$ L to a new set of strip tubes containing 0.5  $\mu$ L of 80  $\mu$ M Rand103tr3 adapter.

## Rand103tr3 Ligation

*NEB has two protocols using T4 RNA Ligase High Concentration. The ligation protocol for a ssRNA oligo to an RNA molecule has a 2 hour incubation, but the protocol for ligating to DNA says to proceed overnight. T4 RNA Ligase may be less efficient with ssDNA than it is with RNA but in any case, we have used overnight ligation for this step and have not tested as to whether that is necessary or whether shorter amount of times are equivalent.*

Rand103tr3 Adapter:

/5Phos/NNNNNNNNNAGATCGGAAGAGCACACGTCTG/3SpC3/

Purify at 100 nmol scale using PAGE Purification, and when asked for random Ns, we used the "Machine Mixing" option (there are two options, hand mixing and machine mixing). Hand mixing might be better to ensure equimolar probabilities of random incorporation.

Store in H<sub>2</sub>O at 200 μM in -20°C. Working dilution is 80 μM.

1. Strip tubes should contain 0.5 μL Rand103tr3 (80 μM) and 9 μL cDNA
2. Heat to 65°C for 2 minutes with lid at 105°C.
3. Place on ice for 1 minute
4. Per reaction add (total 5.5 μL mix):
  - 1.5 μL DMSO (100%)
  - 2.0 μL RNA Ligase Buffer (10x)
  - 2.0 μL ATP (10 mM)
5. Add 5.5 μL mix to each sample.
6. Vortex briefly and spin down.
7. Per reaction add 4 μL PEG8K (50%). (Cut pipette tip for easier pipetting.)
8. Vortex briefly and spin down.
9. Add 1 μL T4 RNA Ligase High Conc (30 U/ μL).
10. Vortex briefly and spin down.
11. Tape down horizontally into a container and place on a shaker at 250 rpm overnight.

*Note: Because MyONE Silane allows for purification of small things, some cDNA generated from free A01m adapter can make it through to this step, thus A01m:Rand103tr3 dimers are possible. The final PCR product of this adapter is 139 nt. This can be removed by size selection after PCR, and you can run a negative control (no starting RNA) through the protocol to verify the adapter. If you start with a high concentration of RNA in the protocol, it seems that very little of this gets made, but if the amount of RNA is limiting, it becomes more prevalent.*

## Clean up with MyONE Silane Beads

*Unclear to me why the eCLIP protocol switches back to washes in 75% EtOH but we assume that lower percentage is higher stringency here.*

1. Separate 10 μL of MyONE Silane beads per sample.
2. Wash beads in batch with 900 μL of Qiagen Buffer RLT.
3. Resuspend beads in 60 μL/sample of Qiagen Buffer RLT (3 starting sample volumes).
4. Add 60 μL Beads in Buffer RLT to each sample and mix.
5. Add 60 μL EtOH (0.75 mix volumes) and mix.
6. Rotate samples at room temp for 15 minutes.
7. Separate on magnet for 30 seconds and remove supernatant.
8. Wash with 0.2 mL 75% EtOH. Resuspend and move to new tube.

9. Separate on magnet for 30 seconds. Wash 2 more times with 75% EtOH with resuspending (30 seconds on magnet).
10. Dry 5 minutes on magnet.
11. Remove residual EtOH.
12. Resuspend in 10  $\mu$ L of H<sub>2</sub>O and let sit for 5 minutes.

### Trial Library PCR

NEBNext Q5 Ultra II Q5 Master Mix (NEB M0544)

Universal Primer

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATC\*T

(The \* is a phosphorothioate bond, which helps preserve the primer and prevent degradation. Make up at 100  $\mu$ M and use 10  $\mu$ M in reaction.)

Index Primers

Can be anything containing the Illumina Read2 priming site with an index that is also compatible with the adapter ligated template. Here is an SIC index primer (index in lower case)

CAAGCAGAAGACGGCATAACGAGATtccgtattaGTGACTGGAGTTCAGACGTGTGCTCTTCCGA

1. Make a master mix. Per sample:
  - 10  $\mu$ L Q5 Ultra Master Mix
  - 1  $\mu$ L 10  $\mu$ M NEBNext Universal primer
  - 7  $\mu$ L H<sub>2</sub>O
2. In individual tubes, combine:
  - 1  $\mu$ L 10  $\mu$ M Index primer
3. 1  $\mu$ L adapter ligated cDNA
4. Add 18  $\mu$ L master mix to index primer/cDNA.

4.0 Run using Q5 Ultra PCR program

|       |           |
|-------|-----------|
| 98°C  | 30 sec    |
| (98°C | 10 sec    |
| 55°C  | 15 sec    |
| 65°C  | 60 sec) - |

*Repeat for N cycles and pull samples at desired cycle numbers and keep on ice.*

|      |       |
|------|-------|
| 65°C | 5 min |
| 12°C | Hold  |

Run 2.5% agarose gel/1X TBE. Determine cycle showing robust amplification of library and not high MW overamplification bands.

We use PCR followed by gel rather than qPCR to determine cycle amplification because it is easier to appreciate the size of your library as well. We test a few cycles in the range of 10-20. In practice, for 1  $\mu$ g of input RNA for total RNA-Seq, we found 10 cycles to be sufficient.

## Preparative Library PCR

Cycle number to use should reflect proportional amount of cDNA input. Suppose we 1  $\mu\text{L}$  in a test reaction in the previous step. If we use 8  $\mu\text{L}$  here, that is 8 fold more starting material, which is 3 base-2 logarithm units. If 15 cycles is determined in the previous step, then use  $15-3 = 12$  cycles in the preparative PCR.

1. Make a master mix. Per sample:
  - 25  $\mu\text{L}$  Q5 Ultra Master Mix
  - 2.5  $\mu\text{L}$  10  $\mu\text{M}$  NEBNext Universal primer
  - 12  $\mu\text{L}$  H<sub>2</sub>O
2. In individual tubes, combine:
  - 2.5  $\mu\text{L}$  10  $\mu\text{M}$  Index primer
3. 8  $\mu\text{L}$  adapter ligated cDNA
4. Add 39.5  $\mu\text{L}$  master mix to each sample.
5. Run using Q5 Ultra PCR program
  - 98°C            30 sec
  - (98°C            10 sec
  - 55°C            15 sec
  - 65°C            60 sec ) x desired cycles
  - 65°C            5 min
  - 12°C            Hold

## SPRI Purification

*Purification is a size selection step using altered polyethylene glycol concentration with Beckman Coulter AMPure XP beads (Beckman Coulter Product # A63881). This selects for things between 200 - 400 bp (We have empirically optimized this using a DNA ladder.) You add the beads directly from a well mixed container without washing them first. It works using the PEG in the bead storage buffer. This is preferable to gel purification because gel purification results in heavy loss of yield compared to magnetic bead based purification. It is also preferable to electroelution after gel purification because electroelution does not scale well to large numbers of samples.*

*Size selection is important because, especially for small amounts of starting material, the A01m:Rand103tr3 adapter dimer is a prevalent species, and will soak up a lot of reads.*

1. Bring volume to 100  $\mu\text{L}$  with 50  $\mu\text{L}$  H<sub>2</sub>O.
2. Add 80  $\mu\text{L}$  of AMPure XP. Mix 10 times by pipetting and incubate 5 minutes.
3. Separate on magnet.
4. Move supernatant to new tube.
5. Add 40  $\mu\text{L}$  of AMPure XP. Mix 10 times by pipetting and incubate 5 minutes.
6. Separate on magnet 2-3 minutes.
7. Wash 2x30s 80% EtOH. It is really important with SPRI beads not to disturb them on the magnet. Just add and let sit for thirty seconds. Remove with multichannel.
8. Dry 5 min.
9. Remove any residual ethanol which collects.
10. Elute in 10  $\mu\text{L}$  of 10 mM Tris pH 7.8.
11. Assay by TapeStation or Bioanalyzer.



## Methods S2: Translating Ribosome Affinity Purification (TRAP), Related to Figure 3 and Figure 5.

This protocol based on previous TRAP protocols cited in this manuscript (below) and is intended as an easy guide to use at the bench. Some modifications have been made to streamline the procedure based on several years of optimization in the Dougherty Lab.

Dougherty, Joseph D., Susan E. Maloney, David F. Wozniak, Michael A. Rieger, Lisa Sonnenblick, Giovanni Coppola, Nathaniel G. Mahieu et al. "The disruption of Celf6, a gene identified by translational profiling of serotonergic neurons, results in autism-related behaviors." *Journal of Neuroscience* 33, no. 7 (2013): 2732-2753.

Heiman, Myriam, Anne Schaefer, Shiaoqing Gong, Jayms D. Peterson, Michelle Day, Keri E. Ramsey, Mayte Suárez-Fariñas et al. "A translational profiling approach for the molecular characterization of CNS cell types." *Cell* 135, no. 4 (2008): 738-748.

Doyle, Joseph P., Joseph D. Dougherty, Myriam Heiman, Eric F. Schmidt, Tanya R. Stevens, Guojun Ma, Sujata Bupp et al. "Application of a translational profiling approach for the comparative analysis of CNS cell types." *Cell* 135, no. 4 (2008): 749-762.

### Bead Prep

Streptavidin MyONE T1 beads (Thermo 65601) have 2x the binding capacity of Streptavidin M-280 beads (Thermo 11205D). The quantities below are intended to coat beads with 2-fold molar excess of immunological components (anti-EGFP antibodies & Pierce biotinylated protein L (Thermo 29997)), based upon the Dyanbeads User Guide: [https://assets.thermofisher.com/TFS-Assets/LSG/manuals/MAN0015761\\_DynabeadsMyOneStreptavidin\\_T1\\_UG.pdf](https://assets.thermofisher.com/TFS-Assets/LSG/manuals/MAN0015761_DynabeadsMyOneStreptavidin_T1_UG.pdf)

Add **60  $\mu\text{L}$  x N IPs** of Streptavidin MyOne T1 **or 120  $\mu\text{L}$  x N IPs** of Streptavidin M-280 beads to a tube. Put the tube on the magnet stand and give it a minute to separate. Remove storage buffer (0.1%BSA/1XPBS)

1. Prepare a binding mixture containing:

17  $\mu\text{L}$  of 1 ( $\mu\text{g}$ )/( $\mu\text{L}$ ) Protein L (17  $\mu\text{g}$ ) **x N IPs**

20  $\mu\text{L}$  of 1.78\* ( $\mu\text{g}$ )/( $\mu\text{L}$ ) anti-EGFP clone 19F7 (36  $\mu\text{g}$ ) **x N IPs**

36  $\mu\text{L}$  of 1\* ( $\mu\text{g}$ )/( $\mu\text{L}$ ) anti-EGFP clone 19C8 (36  $\mu\text{g}$ ) **x N IPs**

127  $\mu\text{L}$  of 1X PBS (f.v. 200  $\mu\text{L}$ ) **x N IPs**

- 1.0 actual concentration of antibody will vary lot to lot.
2. Resuspend beads in the antibody/Protein L mixture by pipetting.
3. Incubate beads and antibody/Protein L mixture at room temperature for at least 1 hour with end-over-end rotation (or up to overnight at 4C with end-over-end rotation).
4. Put beads on magnet stand and give it a minute to separate.
5. Discard supernatant.
6. Resuspend beads in 1 mL 1XPBS/0.1% BSA. Give it a minute in suspension and a minute on the stand to wash.
7. Repeat step 7 4 times (total of 5 washes).
8. Resuspend beads in 1 mL of Wash Buffer. Give it a minute in suspension and a minute on the stand to wash.
9. Repeat step 9 2 times (total of 3 washes).
10. After last wash, resuspend in 1.05xN IPsx100 $\mu\text{L}$  Wash Buffer (5% more than the number of IPs.)

*This step allows you to distribute 100  $\mu\text{L}$  equally to all your IP tubes from the batch of beads. 5% extra volume ensures that you can do this equally. You will find if you resuspend your beads in 500  $\mu\text{L}$  of lysis*

*buffer and try to put 100  $\mu$ L in each of 5 tubes, you will be unable to do so as the detergent in lysis buffer makes this difficult. You can also plan for N+1 IPs, but this works just as well.*

11. Distribute 100  $\mu$ L to **N** tubes equal to the number of actual IPs.
12. Keep on ice until you are ready to use them. Remove the Wash buffer you used to aliquot the beads before use.

## Homogenization and Lysis

*We use the (1mL) proportions for <half a brain or a near confluent 10cm dish of cells, and the 2 mL proportions for a whole brain. See referenced papers for recommendations for volume homogenization buffer per mg tissue, and this may require in-house optimization.*

1. Dissect tissue.
2. Keep glass mortar(s) on ice and fill each with (1 mL | 2 mL) of lysis buffer.
3. Move tissue into glass homogenizer.
4. Take drill fitted with teflon pestle and homogenize up and down 6 times at medium/high power with power drill.
5. After homogenization, pour off homogenate into a 1.7 mL tube (if homogenizing in 1 mL) or some larger size of tube that will accommodate your volume (2 mL tubes are nice).
6. Spin 2000 xg, 10 min, 4°C.
7. Remove (800  $\mu$ L | 1.6 mL) of homogenate from Step 6 and move to a new tube and add:  
(0.1 mL | 0.2 mL) 10% NP40 (final concentration 1%)  
(0.1 mL | 0.2 mL) 300 mM DHPC (final concentration 30 mM)
8. Invert to mix and incubate on ice for 10 minutes.
9. Spin at 20,000xg, 15 minutes, 4C.
10. Measure total lysate. Take 0.1 volumes as Input sample and bring total volume to 250  $\mu$ L with Wash Buffer. Add 750  $\mu$ L Trizol LS and store at -80C until you are ready to extract RNA.
11. Take the remaining 0.9 volumes and resuspend your beads in it.
12. IP for 2 hours.
13. After IP put the samples on the stand and let sit for a minute. Remove the supernatant and discard (you can also take this supernatant to compare to input if you like, in which case repeat step 10 above).
14. Resuspend in 1 mL of High Salt Wash Buffer. Let sit for a minute in suspension on ice, then a minute on the stand.
15. Repeat step 14 3 more times (total of 4 washes).
16. Resuspend beads in 250 uL Wash Buffer. Add 750  $\mu$ L of Trizol LS and store at -80C until you are ready to extract RNA.

## RNA Extraction

1. Bring samples to room temperature if they have been stored at -80C and incubate at room temperature for 5 minutes. If you haven't done so already, take out the Glycoblue (stored at -20) and bring to room temperature.
2. Add 0.2 mL chloroform.
3. Shake vigorously by hand for 15 seconds.
4. Incubate for 7 minutes on bench.  
*At this point, you can use Phase Lock Gel Heavy tubes to help you separate phases later. Pellet Phase Lock Gel 12,000 xg, 30 seconds. Add your sample to the pelleted Phase Lock Gel. Mix well but do not vortex.*
5. Centrifuge at 12,000xg, 15minutes, 4C.
6. Vortex the Glycoblue well and spin briefly.

7. Add 2  $\mu\text{L}$  of Glycoblue to new tubes equal to the number of samples you have.
8. Remove the aqueous phase to a new tube (the upper layer) from step (5) to the tubes with glycoblue and mix well.
9. Add 0.7 volumes of 100% isopropanol.
10. Incubate on the bench for 10 minutes.
11. Centrifuge at 4C, max speed ( $\geq 12,000\times g$ ), for 15 minutes.
12. Pour off the supernatant. Do not pipette off the supernatant.
13. Add 1 mL of 80% EtOH. (Dislodge the pellet and invert several times to wash.)
14. Repeat the centrifugation in step 11.
15. Pour off the supernatant. Be careful not to dislodge pellet.
16. Leave tubes open on your tube rack while you prepare the DNase treatment mix (3-5 minutes. Do not overdry!). (This paper and protocol uses NEB Rnase-free DNase I (M0303) but we have also used this protocol with Qiagen DNase and this works well. Likely most DNase kits will be appropriate).

DNase Treatment Mix:

2  $\mu\text{L}$  10X NEB DNase I buffer x **n+1 samples**

87  $\mu\text{L}$  H<sub>2</sub>O x **n+1 samples**

3  $\mu\text{L}$  Qiagen DNase I x **n+1 samples**

17. Resuspend the pellet in 20  $\mu\text{L}$  of DNase Treatment Mix.
18. Incubate at 37 C for 15 minutes.
19. Clean up RNA with Zymo RNA Clean & Concentrator 5 (> 17 nt), Qiagen RNeasy kit (>200 nt), or MyONE Silane Dynabeads (see Methods S1)\*.

## Reagent Recipes

### 0.1% BSA/1XPBS

Add IgG-Free Bovine Serum Albumen (100 mg per 10 mL) to 1X PBS and allow to rock gently for >10 minutes to go into solution (1% BSA). Dilute 1% BSA/1XPBS 10-fold.

### 10% NP40

Carefully make up 10% v/v NP40 (IGEPAL CA-630, Sigma) by pipetting 100% NP40 into H<sub>2</sub>O and let rock for >10 minutes to fully dilute.

### 300 mM DHPC

Resuspend 100 mg 07:0 PC (DHPC, Avanti polar lipids) in 692  $\mu\text{L}$  of H<sub>2</sub>O. To avoid foaming, add H<sub>2</sub>O and let it rock for a few minutes and then transfer to Eppendorf tubes.

### 100 mg/mL cycloheximide

Dissolve cycloheximide in 100% methanol at 100 mg/mL.

7X Roche cOmplete EDTA-free protease inhibitor

Crush tablet in a small volume (~300-500  $\mu\text{L}$ ) of water. Pipette up and down to start to dissolve. Bring volume to 1.5 mL and mix until fully dissolved.

*Add DTT, RNasin, Superase-In, protease inhibitor, and cycloheximide just prior to use.*

## Homogenization Buffer

10 mM HEPES pH 7.4  
150 mM KCl  
5 mM MgCl<sub>2</sub>  
0.5 mM dithiothreitol  
1 μL/mL rRNasin  
1 μL/mL SUPERase-in  
1 X cOmplete EDTA-free protease inhibitor  
100 μg/mL cycloheximide

### **Wash Buffer**

10 mM HEPES pH 7.4  
150 mM KCl  
5 mM MgCl<sub>2</sub>  
1 % NP-40  
0.5 mM dithiothreitol  
1 μL/mL rRNasin  
1 μL/mL SUPERase-in  
1 X cOmplete EDTA-free protease inhibitor  
100 μg/mL cycloheximide

### **High Salt Wash Buffer**

10 mM HEPES pH 7.4  
350 mM KCl  
5 mM MgCl<sub>2</sub>  
1 % NP-40  
0.5 mM dithiothreitol  
1 μL/mL rRNasin  
1 μL/mL SUPERase-in  
1 X cOmplete EDTA-free protease inhibitor  
100 μg/mL cycloheximide

## Methods S3: Preparing PTRE-Seq samples for Next Generation Sequencing, Related to Figure 3 and Figure 5.

The input into this procedure for generating sequencing libraries from PTRE-Seq samples is either extracted and purified library RNA or library plasmid DNA. To prepare for sequencing from plasmid DNA, skip cDNA synthesis and proceed to “**ds cDNA amplification and plasmid DNA amplification**”. If processing samples in parallel, the samples can be synchronized from the amplification step forward as indicated. Procedure is based on mRNA synthesized off of the pmrPTRE\_AAV backbone (see Key Resource Table). Plasmid map available upon request.

For multiplexing, this protocol assumes use of a Permagen Labware strip tube magnet separator ( see Methods S1).

### cDNA Synthesis

pmrPTRE Antisense oligo: GGCCTGGAGTGGCAACT

Superscript III Reverse Transcriptase: Thermo 18080093

(Note the only apparent differences between buying the enzyme alone or in the First Strand Synthesis kit is that the  $MgCl_2$  is included in the 5X First Strand Buffer in product 18080093, and the enzyme kit alone does not include dNTPs).

1. Add 2.5  $\mu$ L of 12  $\mu$ M pmrPTRE Antisense oligo to each tube.
2. Add 8.0  $\mu$ L of RNA/water to each tube containing 10 ng of RNA. ( This protocol, cycle numbers etc. in downstream steps, was developed with starting 10 ng of RNA, however could be scaled up or potentially even down, though some cycle numbers may have to be optimized to avoid over-amplification. No rRNA depletion is performed as the library is amplified using oligonucleotide primers specific to the pmrPTRE\_AAV plasmid.)
3. Heat to 65°C for 2 minutes with lid at 105°C. Place immediately on ice for 1 minute.
4. Make mastermix, per reaction:
  - 4  $\mu$ L 5X First Strand Buffer
  - 2  $\mu$ L 10 mM dNTPs
  - 2  $\mu$ L 100 mM dithiothreitol
  - 0.6  $\mu$ L Promega rRNasin (N2515)
  - 0.9  $\mu$ L SSRT III RT enzyme
5. Add 9.5  $\mu$ L of mastermix to each sample.
6. Incubate all at 50°C for 45 minutes.
7. Add 3.5  $\mu$ L of ExoSAP-It (Thermo 78200.200.UL).
8. Vortex & spin
9. Heat to 37°C for 15 minutes. (f.v. 23.5)
10. Add 1  $\mu$ L 0.5 M EDTA per reaction. Vortex & spin. (f.v. 24.5)
11. Add 3  $\mu$ L 1 M NaOH. Vortex and Spin. (f.v. 27.5)
12. Heat to 70C for 12 minutes.
13. Add 3  $\mu$ L 1M HCl. Vortex and Spin. (f.v. 30.5)
14. Clean up with MyONE Silane Dynabeads (see Methods S1 for full procedural description). Briefly:
  - a. Wash beads with 900  $\mu$ L RLT
  - b. Resuspend in 91.5  $\mu$ L/sample buffer RLT
  - c. Add 91.5  $\mu$ L to each sample and mix (f.v. 122)
  - d. Add 111  $\mu$ L 100% EtOH. Incubate 15 minutes. (f.v. 233)
  - e. Wash MyONE silane style 3x0.2mL 80% EtOH. First wash, fully resuspend. Washes 2&3 just wash on magnet.

- f. Dry 5 minutes.
- g. Remove any excess EtOH.
- h. Elute in 12  $\mu\text{L}$  50 mM Tris pH 7.8 and incubate 5 minutes before clarifying.

### ds cDNA amplification and plasmid DNA amplification

NEB Phusion High-Fidelity Polymerase (NEB M0530)

PTRE sense oligo: GCATGGACGAGCTGTACAAG

PTRE antisense oligo: as in cDNA synthesis

*If using Plasmid DNA, dilute to approximately 5 ng/ $\mu\text{L}$ .*

1. Seed 20  $\mu\text{L}$  PCR reaction with 8  $\mu\text{L}$  of eluted cDNA sample from the previous step, or 8  $\mu\text{L}$  of 5 ng/ $\mu\text{L}$  DNA samples. Per reaction:
  - 10.0  $\mu\text{L}$  2X Phusion HF Master Mix (NEB M0530)
  - 1.0  $\mu\text{L}$  10  $\mu\text{M}$  PTRE UTR antisense 1 (x25= 25)
  - 1.0  $\mu\text{L}$  10  $\mu\text{M}$  PTRE UTR sense 1 (x25 = 25)
  - 8.0  $\mu\text{L}$  cDNA
2. Add 12  $\mu\text{L}$  of master mix to each 8  $\mu\text{L}$  sample.
3. Run following program:
  - 98°C 30s
  - 98°C 10s, 60°C 10s, 72°C 15s | 12 cycles (plasmid DNA) or 18 cycles (cDNA)
  - 72°C 10 min
  - 4°C  $\infty$

Product size: 245 bp

Product sketch:

```

mtdTomato          STOP NheI  Inserted 3'UTR element
5' -GCATGGACGAGCTGTACAAGTAAgctagc{X}120 bp insert
3' -CGTACCTGCTCGACATGTTTCATTcgatcg{X}complement

                BARCODE
5' -GAATTCcGATATCaCTCGAG{N}9bp barcode
3' -CTTAAGgCTATAGtGAGCTC{N}complement

KpnI              HGH terminator
5' -GGTACCGGCATCCCTGTGACCCCTCCCCAGTGCCTCTCCTGGCCCTGGAAGTTGCCACTC
3' -CCATGGCCGTAGGGACACTGGGGAGGGGTACGGAGAGGACCGGGACCTTCAACGGTGAG

5' -CAGTGCC
3' -GTCACGG
  
```

Note the location of the KpnI & NheI sites. This procedure adapts the Illumina P1 priming site to the KpnI restriction site, such that the 9 bp element barcode will be read as the beginning of Read 1 of sequencing.

## AMPure 80/40 selection

*Purification/size selection steps use Beckman Coulter AMPure XP beads (Beckman Coulter A63881) to size select to a range ~200-400 bp thus purifying the product from the previous step with high yield. We recommend this against gel extraction for simplicity and yield, however 2% agarose gels were used in development to verify correct size of amplicon (See Methods S1).*

1. Bring AMPure XP beads to room temperature ~30 min.
2. Bring volume to 100  $\mu$ L with H<sub>2</sub>O (add 80  $\mu$ L).
3. Vortex beads and pipette 80  $\mu$ L and add to PCR products. Mix by pipetting up and down 10x.
4. Incubate on bench 5 minutes.
5. Put on strip tube magnet for 3 minutes.
6. Re-vortex beads and pipette 40  $\mu$ L to a clean strip tube.
7. Harvest supernatant (~160  $\mu$ L recoverable) and add to new tube with 40  $\mu$ L beads in it. Mix by pipetting up and down 10x.
8. Incubate on bench 5 minutes.
9. Put on strip tube magnet 3 minutes.
10. Wash beads on magnet with 200  $\mu$ L 80% EtOH for 30 seconds. Do not disturb pellet. Add wash, let sit, then remove.
11. Repeat for total of two washes.
12. Dry beads for 5 minutes and remove any excess EtOH with vacuum or by pipette.
13. Resuspend beads in 10.5  $\mu$ L Buffer EB (so that it will be possible to recover 10  $\mu$ L) and incubate on bench 5 minutes.
14. Add to magnet for 1 minute and harvest 10  $\mu$ L.

## NheI/KpnI Digest

This procedure digests the double stranded products generated above to prepare for adapter ligation. Uses NEB enzymes NheI-HF (R3131) and KpnI-HF (R3142).

1. Remove 10  $\mu$ L of purified PCR product and add to:
  - 2  $\mu$ L 10X CutSmart Buffer
  - 1  $\mu$ L NheI HF
  - 1  $\mu$ L KpnI HF
  - 6  $\mu$ L H<sub>2</sub>O
2. Mix well, spin, and incubate in thermal cycler for 1 hour at 37°C with lid set to 40°C.
3. During protocol optimization, we verified digest by running products out on an Agilent TapeStation instrument. Digested product 160 bp.
4. AMPure 100/50 selection
5. This selection has altered volumes of AMPure XP to select against undigested PCR product. Purity of products was confirmed by Agilent TapeStation.
6. Bring volume to 100  $\mu$ L with H<sub>2</sub>O (80  $\mu$ L).
7. Vortex beads and pipette 100  $\mu$ L and add to digested products. Mix by pipetting up and down 10x.
8. Incubate on bench 5 minutes.
9. Put on strip tube magnet for 3 minutes.
10. Re-vortex beads and pipette 50  $\mu$ L to a clean strip tube.
11. Harvest supernatant and add to new tube with 50  $\mu$ L beads in it. Mix by pipetting up and down 10x.
12. Incubate on bench 5 minutes.
13. Put on strip tube magnet 3 minutes.

14. For QC, save the first bead pellet. This is the first round of selection. Mostly selects out high MW species, but may select out a bit of desired product too.
15. Wash beads on magnet with 200  $\mu$ L 80% EtOH for 30 seconds. Do not disturb pellet. Add wash, let sit, then remove.
16. Repeat for total of two washes.
17. Dry beads for 5 minutes and remove any excess EtOH with vacuum or by pipette.
18. Resuspend beads in 10.5  $\mu$ L Buffer EB (so that it will be possible to recover 10  $\mu$ L) and incubate on bench 5 minutes.
19. Add to magnet for 1 minute and harvest 10  $\mu$ L.

### Staggered Adapter Preparation

In order to generate sufficient sequence diversity for next generation sequencing, the KpnI/P1 adapter containing a priming site for the Illumina P1, is ligated as a mix of 4 adapters which are staggered in length (see Table S9 and below):

|  |                   |
|--|-------------------|
| ACACTCTTTCCCTACACGACGCTCTTCCGATCTTCATGTA*C             | KpnI_overhang_1   |
| ACACTCTTTCCCTACACGACGCTCTTCCGATCTCAGGTGTA*<br>C        | KpnI_overhang_2   |
| ACACTCTTTCCCTACACGACGCTCTTCCGATCTGTTCTGTA*<br>C        | KpnI_overhang_3   |
| ACACTCTTTCCCTACACGACGCTCTTCCGATCTAGCAGCTGT<br>A*C      | KpnI_overhang_4   |
| /5phos/A*TGAAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTG<br>T-3    | KpnI_complement_1 |
| /5phos/A*CCTGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGT<br>GT-3   | KpnI_complement_2 |
| /5phos/A*GGAACAGATCGGAAGAGCGTCGTGTAGGGAAAGA<br>GTGT-3  | KpnI_complement_3 |
| /5phos/A*GCTGCTAGATCGGAAGAGCGTCGTGTAGGGAAAGA<br>GTGT-3 | KpnI_complement_4 |
| /5phos/C*TAGAGATCGGAAGAGCACACGTCTG                     | NheI_overhang     |
| CAGACGTGTGCTCTTCCGATC*T                                | NheI_complement   |

1. Prepare synthesized KpnI adapters and NheI/P2 adapters as 100  $\mu$ M in 1X TE or Tris pH 8 buffer.
2. Make 1:1 equimolar mixes (f.c. 10  $\mu$ M each adapter) of "overhang" and "complement" (final volume is arbitrary).
3. Heat adapter mixes to 95C for 3 minutes.
4. Move to bench and let cool to room temperature to anneal double stranded adapters.
5. Move to ice.
6. Make a 1:1:1:1 equimolar mix of KpnI double stranded adapters (final concentration 1  $\mu$ M each adapter).
7. Make a 1  $\mu$ M dilution of NheI adapter.



Adapters are now ready for ligation. Barcode sequences will be the first nucleotides read in Read 1 of illumina sequencing.

## Adapter Ligation

T4 DNA Ligase (L6030-LC)

1. Combine 10  $\mu$ L purified/digested PCR product in the follow reaction  
6  $\mu$ L H<sub>2</sub>O  
2  $\mu$ L Enzymatics T4 DNA Ligase buffer 10X  
0.5  $\mu$ L 1  $\mu$ M NheI/P2 adapter  
0.5  $\mu$ L 1  $\mu$ M mix of KpnI/P1 adapters 1-4  
1  $\mu$ L of Enzymatics T4 DNA Ligase
2. Incubate at 16 C for 1 hour
3. Purify products with Ampure 80/40 procedure (above) and elute in 10.5  $\mu$ L EB, and harvest 10  $\mu$ L.

After adapter ligation, products will be ~220 bp. Ligation can be confirmed by Agilent TapeStation.

## Preparative PCR for Sequencing

As in Methods S1, this procedure is performed using NEB Q5 Ultra II PCR master mix (M0544). Trial PCR was run followed by agarose gel electrophoresis to optimize cycle number.

Illumina Univ. F: AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATC\*T

Index Reverse Primer:

CAAGCAGAAGACGGCATAACGAGATxxxxxxxxGTGACTGGAGTTCAGACGTGTGCTCTTCCG\*A

“xxxxxxxx” represents a 9bp unique sample index for demultiplexing.

1. Master mix:  
25  $\mu$ L NEB Q5 Ultra II pcr master mix  
2.5  $\mu$ L 10  $\mu$ M universal primer  
12  $\mu$ L H<sub>2</sub>O
2. Aliquot 39.5 to each adapter-ligated DNA sample.
3. Add 2.5  $\mu$ L of index primer.
4. Run with Q5 program for 9 cycles:  
98°C 30 s  
(98°C 10 s | 55°C 15 s | 65°C 1 min) X 9  
65°C 5 min  
12°C hold
5. Purify with 80/40 AMPure XP procedure as above.
6. Assay purity and molar concentration of NGS library PCR products by Agilent TapeStation.