Cell Reports, Volume *26*

Supplemental Information

Cell-Type-Specific Profiling of Alternative Translation Identifies Regulated Protein Isoform Variation in the Mouse Brain

Darshan Sapkota, Allison M. Lake, Wei Yang, Chengran Yang, Hendrik Wesseling, Amanda Guise, Ceren Uncu, Jasbir S. Dalal, Andrew W. Kraft, Jin-Moo Lee, Mark S. Sands, Judith A. Steen, and Joseph D. Dougherty

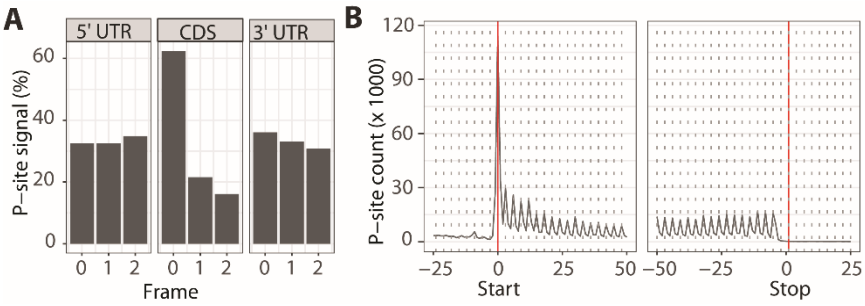

Figure S1. Neuron/glia culture ribosome footprints indicate bona fide translation. *Related to Figure 1.* **A)** More than 60% of P-sites detected in the footprints mapping to the CDS shows the translation of the correct reading frame. **B)** P-sites detected in the footprints mapping to the CDS exhibit 3-nucleotide periodicity of ribosomes. Data are from a 'no HHT, No KCl' sample, and the same pattern was observed with other samples.

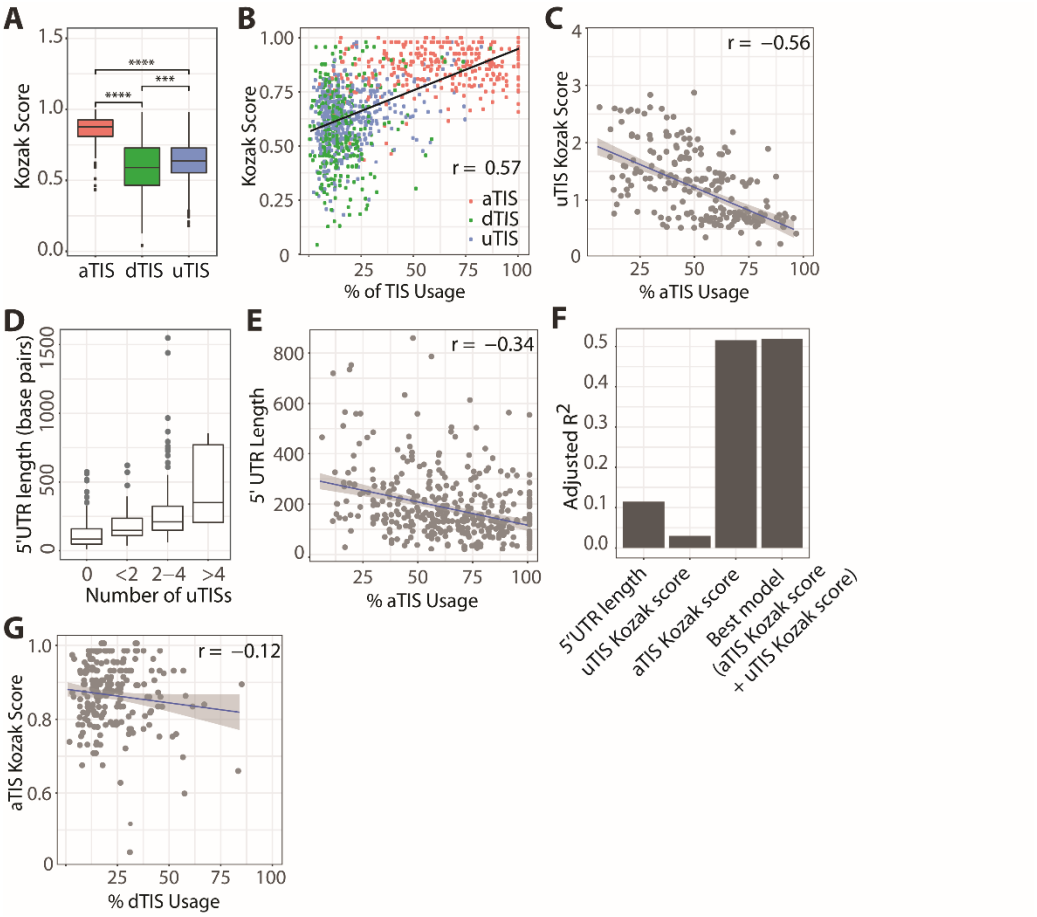

Figure S2. Identified TISs reveal the complexity of initiation from neural transcripts. *Related to Figure 1.* **A)** Kozak strength varies significantly between TISs. Overall comparison of Kozak strengths was done using post hoc Tukey's Honestly Significance Difference test and showed a highly significant difference (p = 0, not shown in the box). Pair-wise comparison was done using *t* tests (***: $p \leq 0.001$; ****: $p \leq 0.0001$). **B)** Comparison of TIS usage and Kozak strength shows that TISs with higher Kozak scores, especially aTISs, are more favorably used for initiation **(C)** uTISs with higher Kozak scores inhibit initiation from aTISs. The Y axis represents the sum of the kozak scores of all uTIS in a given transcripts. **(D-E)** Longer 5'UTRs contain more uTISs (D) and inhibit initiation from aTISs more strongly than shorter 5'UTR (E). **(F)** Regression-based model prediction shows the Kozak scores of aTIS and uTIS as the strongest determinants of initiation from aTISs (R-squared 0.519). **(G)** aTISs with higher Kozak scores moderately reduce initiation from dTISs. r = Pearson correlation coefficient. TIS usage was calculated as the P-site counts at a TIS over total P-site counts across all TISs on a given transcript. Kozak Strength was calculated by comparing the TIS context sequences with a position weight matrix derived from an a polysome profiling experiment determining ribosome loading on synthetic sequences as described in Sample et. al., 2019 and shown in arbitrary unit in the Y axis, with 1 indicating the best strength.

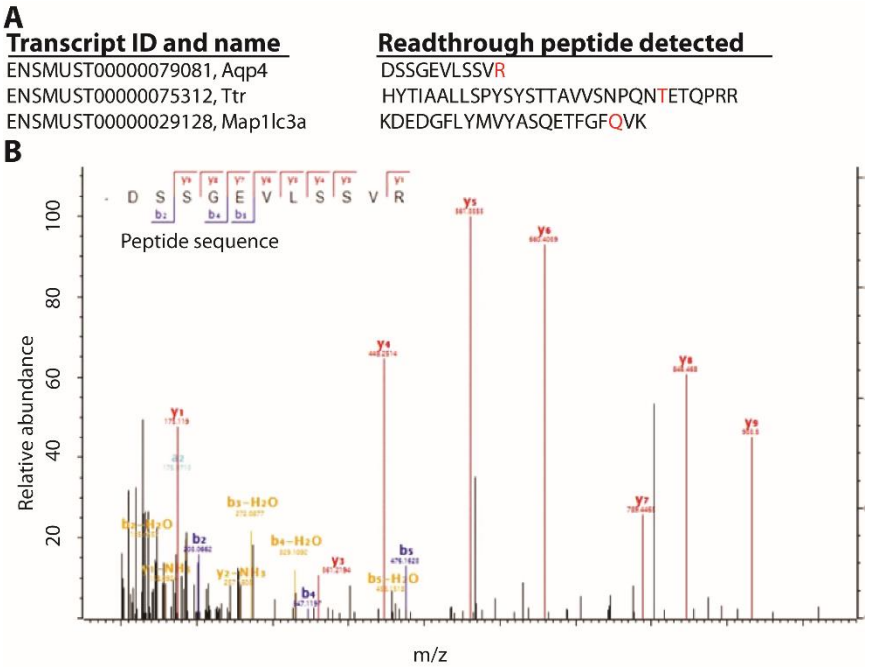

Figure S3. Detection of readthrough peptides in mouse brain mass spectrometry data. *Related to Figures 4 and 6.* **(A)** Peptides corresponding to three readthrough regions are shown. Stop codon-recoded amino acids are in red. **(B)** Tandem mass spectra and mass-to-charge ratios (m/z) of b and y product ions corresponding to the Aqp4 readthrough peptide is shown. Mouse brain mass spectrometry data reported by Sharma *et al* 2015 were examined setting the posterior error probability cut off value at 0.01.

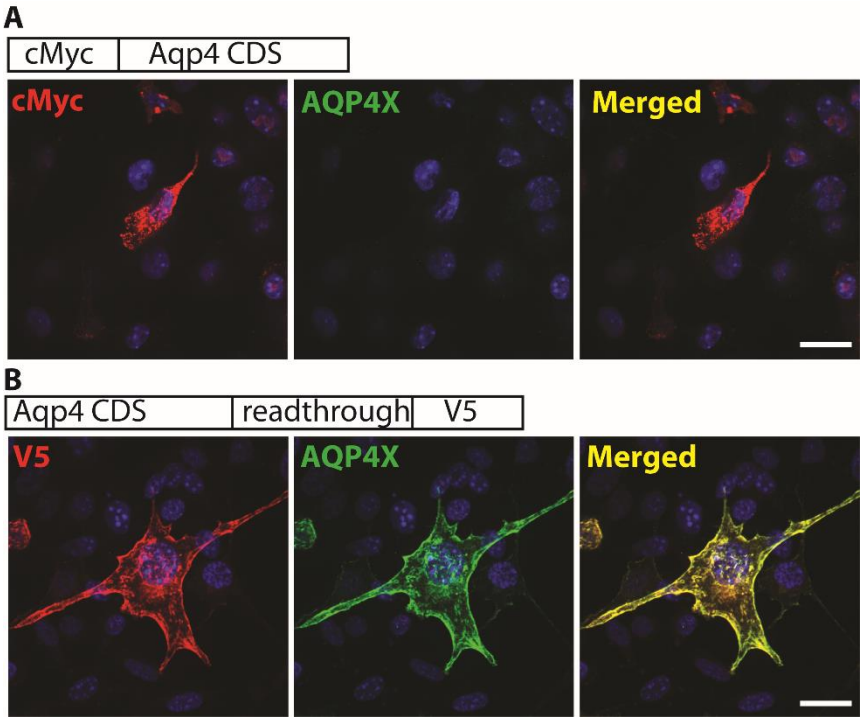

Figure S4. Validation of anti-Aqp4X antibody. *Related to Figures 5 and S5.* DBT glioblastoma cells were transfected with constructs expressing normal Aqp4 with an N-terminal cMyc tag **(A)** or Aqp4X with a C-terminal V5 tag **(B)**. Immunostaining of the transfected cells shows that a cMyc-positive cell does not express Aqp4X, whereas a V5-positive cell does. Scale bars = 20 μM.

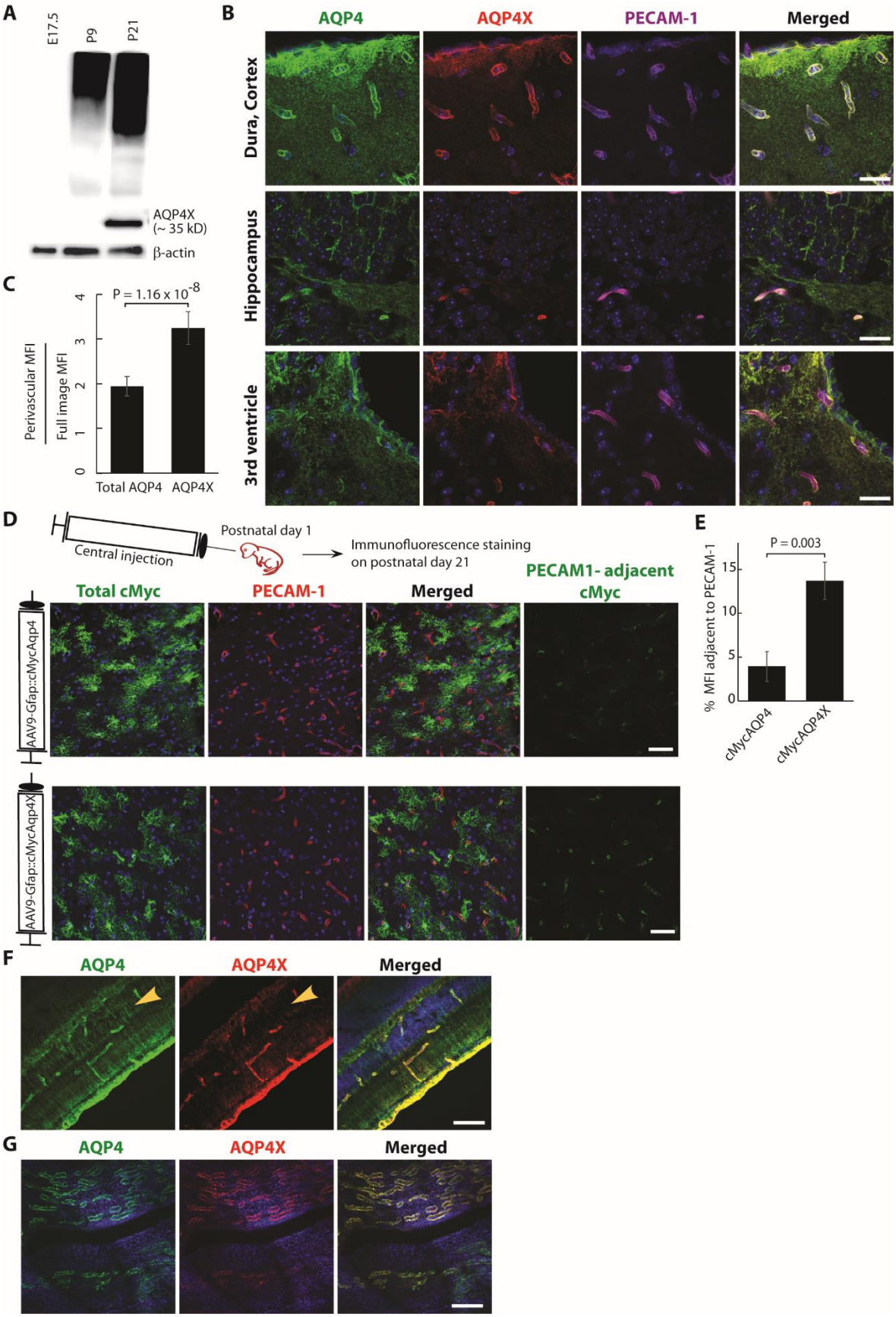
A
E17.5
P9
P21
AQP4X
(~ 35 kD)
β-actin
B
AQP4
AQP4X
PECAM-1
Merged
Dura, Cortex
Hippocampus
3rd ventricle
C
P = 1.16 x 10-8
Perivascular MFI
Full image MFI
0
1
2
3
4
Total AQP4
AQP4X
D
Central injection
Postnatal day 1
Immunofluorescence staining
on postnatal day 21
Total cMyc
PECAM-1
Merged
PECAM1- adjacent
cMyc
AAV9-Gfap::cMycAqp4
AAV9-Gfap::cMycAqp4X
E
P = 0.003
% MFI adjacent to PECAM-1
0
5
10
15
cMycAQP4
cMycAQP4X
F
AQP4
AQP4X
Merged
G
AQP4
AQP4X
Merged

Figure S5. Readthrough confers the perivascular localization ability to AQP4 in the central nervous system. *Related to Figure 5.* **A)** Western blot on brain lysates detects the expected 35 kD band in the adult brain but not in the developing brain. Note that high molecular weight smear accompanied the expected band, as is commonly seen for other AQP4 antibodies89. **B)** Immunostaining of the adult mouse brain shows diffused expression of AQP4 and PECAM 1-overlapping perivascular expression of AQP4X in different regions. **C)** Quantification of MFIs from B shows the signal from anti-AQP4X is significantly more perivascular than the total (AQP4 + AQP4X) signal from anti-AQP4. **D)** AAV9-mediated astrocytic expression of cMyc-tagged AQP4X- (an extra stop codon added to prevent readthrough) or AQP4X+ (stop mutated to allow constitutive readthrough) in the brain followed by immunostaining shows that cMycAQP4 is less contiguous with the perivascular PECAM-1 than cMycAQP4X. Regions of interest were drawn around PECAM-1 to distinguish perivascular cMyc. **E)** Quantification of cMyc MFIs from D shows that perivascular cMyc accounts for 4% and 14% of the total cMyc in AAV9-Gfap::cMycAqp4X- injected brain and AAV9-Gfap::cMycAqp4X+ injected brain, respectively. **F, G)** Immunofluorescence staining of the retina (F) and kidney (G) from adult mice shows the expression of AQP4X in these tissues. Note that AQP4X largely perivascular in the retina but completely overlaps with AQP4 in the kidney. The nuclear stain DAPI is removed from the single channel images for the retina to highlight the presence of AQP4 outside the perivascular region (arrow). For C and E: 3 mice; 6 sections/mouse. Unpaired t-test; error bars, standard errors of mean. MFI, mean fluorescence intensity; AAV9, Adeno-associated virus 9. Scale bars = 20 μM in B, 50 μM in D, G and G.

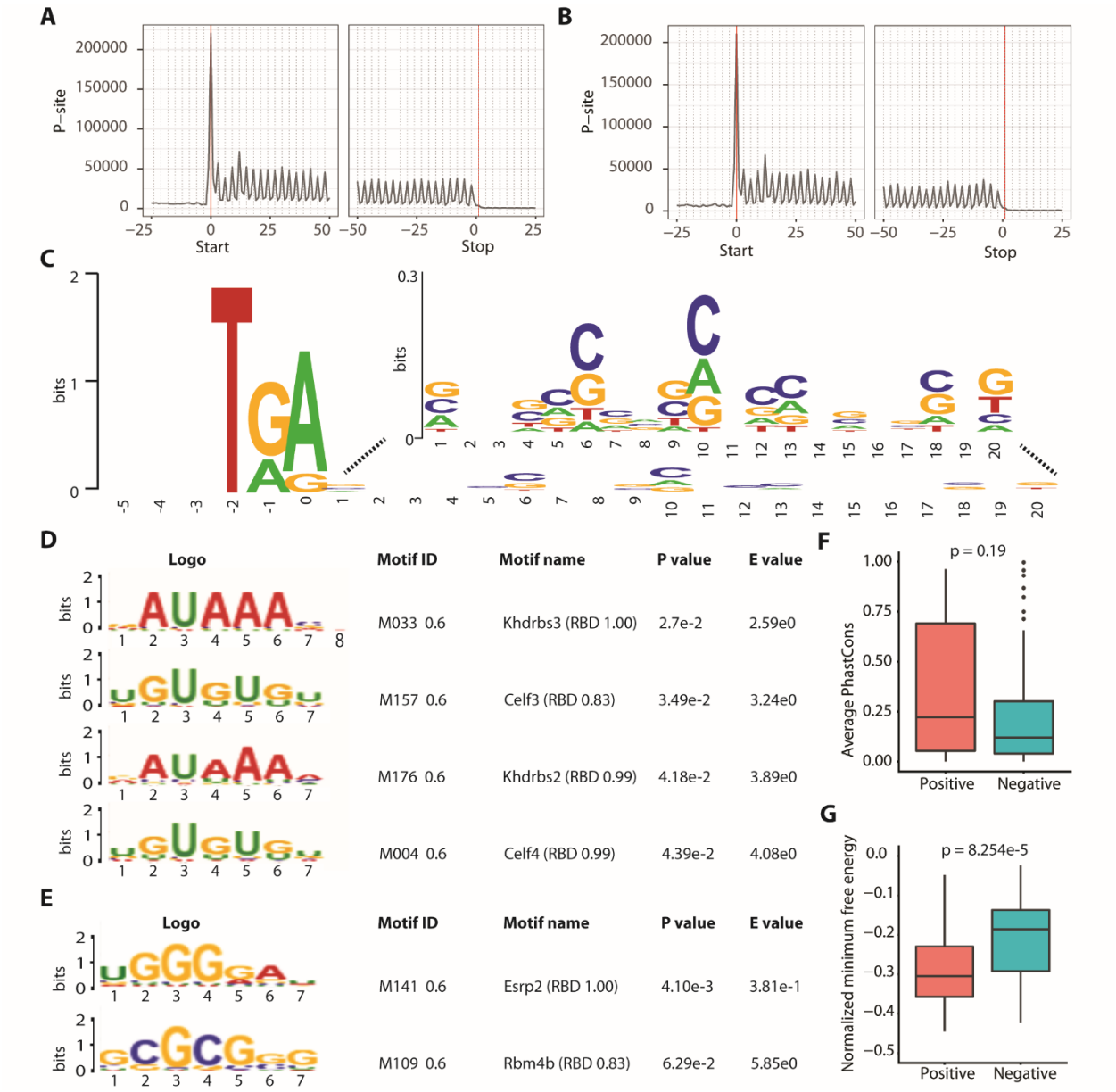

Figure S6. Readthrough candidates show 3-nucleotide periodicity in the CDS and specific sequence features in the readthrough region. *Related to Figure 6.* **A, B)** P sites detected globally in *Snap25::Rpl10a-Egfp* TRAP-RF and *Aldh1l1::Rpl10a-Egfp* TRAP-RF are plotted in A and B, respectively. **C)** Position weight matrix reveals TGA as the commonest stop codon in transcripts undergoing readthrough. Inset shows the magnification of the 20-nucleotide readthrough window. **D, E)** Analysis of motif enrichment for readthrough candidates using shuffled sequences as controls (D) and readthrough-negative candidates as controls (E) shows the overrepresented motifs for specific RNA-binding proteins. **F)** Conservation analysis of sequence between the first and second stop codons suggests that readthrough positive regions tend to be conserved, but not significantly, as compared to readthrough negative regions. Fisher's exact test was performed on the odds ratios of positive and negative groups. PhastCons scores greater than 0.5 are considered to represent conservation arbitrarily. **G)** Wilcoxon test to compare minimum free energies reveals a more stable secondary structure (less free energy) between the first and second stop codons of readthrough candidates. 48 in vivo readthrough candidates were used

in all analyses (2 were removed due to transcript ID problem). 46 readthrough-negative transcripts with matched lengths between the first and second stop codons were used in E, F and G. In C, D and E, Y axis represents conservation in bits.

Table S2. Peptide products of alternative translation initiation sites detected in the mouse brain mass spectrometry data from Sharma et al, 2015. *Related to Figure 2 and Table 1.* "+" and "-" indicate downstream and upstream relative to the canonical initiation site, respectively.

SN	Transcript ID	Gene name	Alternative initiation site	Frame
1.	ENSMUST00000020657	Ube2b	-24	In frame
2.	ENSMUST00000067664	Ywhae	-66	In frame
3.	ENSMUST00000021933	Ctsl	-27	In frame
4.	ENSMUST00000030134	Rad23b	-277	Out of frame
5.	ENSMUST00000021674	Fos	-46	Out of frame
6.	ENSMUST00000100802	Nufip2	-45	In frame
7.	ENSMUST00000067664	Ywhae	-49	Out of frame
8.	ENSMUST00000110082	Calm1	-100	Out of frame
9.	ENSMUST00000103008	Sdcbp	-54	In frame
10.	ENSMUST00000027377	Igfbp5	-405	In frame
11.	ENSMUST00000111230	Tagln2	7	Out of frame
12.	ENSMUST00000031131	Uchl1	15	In frame
13.	ENSMUST00000106255	Cap1	9	In frame
14.	ENSMUST00000031565	Fscn1	12	In frame
15.	ENSMUST00000112172	Tmsb4x	7	Out of frame
16.	ENSMUST00000112229	Gpm6b	12	In frame
17.	ENSMUST00000110082	Calm1	12	In frame
18.	ENSMUST00000108857	Atox1	6	In frame

Table S3. Transcripts showing at least 1% readthrough identified in vitro. *Related to Figure 4.*

SN	Transcript symbol	Readthrough rate
1.	*2010107E04Rik*	0.033
2.	*Aqp4*	0.382
3.	*Brk1*	0.048
4.	*Ctxn1*	0.327
5.	*Dlx1*	0.185
6.	*Dynll2*	0.106
7.	*Gpm6a*	0.056
8.	*Mdh1*	0.154
9.	*Mlc1*	0.055
10.	*Mt1*	0.024
11.	*Mt2*	0.041
12.	*Ncan*	0.014
13.	*Pea15a*	0.015
14.	*Pfn2*	0.015
15.	*Plat*	0.102
16.	*Ptms*	0.027
17.	*Ssna1*	0.322
18.	*Tmem261*	0.230

Table S4: Transcripts with at least 1% readthrough in vivo. *Related to Figure 6.* Readthrough rates are an average of two Snap25::Rpl10a-Egfp and two Aldh1l1::Rpl10a-Egfp samples for neuron- and astrocyte-enriched transcripts, respectively, and an average of all four for cell type-nonspecific transcripts.

Neuronal		Astrocytic		Non cell type-specific	
Transcript	**Readthrough rate**	**Transcript**	**Readthrough rate**	**Transcript**	**Readthrough rate**
Atp1a3	0.017	*Aldoc*	0.065	*2010107E04Rik*	0.051
Celf5	0.24	*Aqp4*	0.189	*Arl3*	0.026
Clstn3	0.041	*Atp1a2*	0.025	*Brk1*	0.023
Cox6a1	0.023	*Clu*	0.014	*Camk2n1*	0.035
Ctxn1	0.18	*Csrp1*	0.014	*Fam195b*	0.08
Dynll2	0.02	*Cst3*	0.019	*Lmtk3*	0.319
Eef1a2	0.024	*Fam19a5*	0.15	*Ssna1*	0.294
Fam163b	0.182	*Manbal*	0.078	*Triap1*	0.17
Fxyd7	0.015	*Fopnl*	0.445	*Tmem88b*	0.33
Hint1	0.064	*Map4*	0.317	*Ubl3*	0.095
Lynx1	0.035	*Mbp*	0.012		
Map1lc3a	0.025	*Mlc1*	0.139		
Mdh1	0.071	*Mt1*	0.098		
Pld3	0.014	*Mt2*	0.037		
Prr24	0.209	*Ndrg2*	0.024		
Ptms	0.023	*Pea15a*	0.023		
Serp2	0.179	*Sparc*	0.009		
Thy1	0.014	*Smim7*	0.119		
Tmem240	0.041	*Ttr*	0.033		
Tmem261	0.209				
Yaf2	0.243				