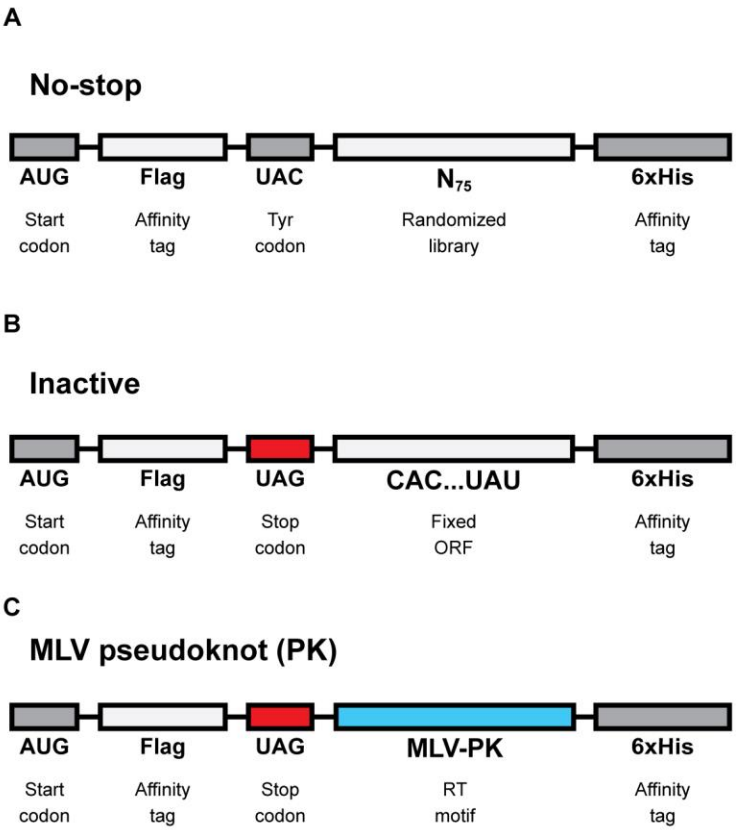
Interrogation of eukaryotic stop codon readthrough signals by *in vitro* RNA selection

AUTHORS: Andrew V. Anzalone, Sakellarios Zairis, Annie J. Lin, Raul Rabadan & Virginia W. Cornish

Supplementary Materials Contents



Sample	Sequence
1	PQDGNSVAWVVGSHACDWSREHVGC
2	SNFKAIPLCLTLRTRVPDLERDRVI
3	HTAEVTQWGESRCRGALRTFGRCKL
4	RFISRADQASMRSIRICGALLGSKQ
5	SRCVFGIHRICGQGCFRQLILVTRV
6	ATWQMIVSWRSISCRAIYGRLQIRG
7	SLTGSNSERISRGYRTKAHSGGRIL
8	DSSAHFPNICSGFVLVRILGVAGTC
9	TYVLFGKRGPMGAQKNWRVQGWSES
10	GNVRWFPGDLRRHVTLRVMSQSNKC
11	PRIRCYARKVQIRELAPRGRHLGCS
12	GRIIYAYVADGEIRSVFTTPGDEVS

Sample	Notes
1	No mutations
2	No mutations
3	No mutations
4	*TAG(Stop)→TGG(Trp)
5	*TAG(Stop)→TAT(Tyr)
6	*ΔA35→No stops, FLAG in frame
7	*TAG(Stop)→TGG(Trp)
8	No mutations
9	*TAG(Stop)→TGG(Trp)
10	*TAG(Stop)→TGG(Trp)

MLV-PK : Inactive	MLV-PK	Inactive	Relative MLV Enrichment
1 : 1	5 / 6	1 / 6	5
1 : 10	9 / 12	3 / 12	30
1 : 100	3 / 11	8 / 11	37.5
1 : 1,000	0 / 13	13 / 13	--

Figure S1. Mock *in vitro* selection experiments. A single mRNA-display selection cycle was performed using the control constructs shown. Selection products were cloned, then individual colonies were sampled and analyzed by Sanger sequencing. (A) A control library containing 75 randomized nucleotides, lacking a deliberately installed stop codon. Assuming equal nucleotide probability weightings, each library member has a 30% probability of containing zero stop codons. All twelve post-selection library members that were sequenced were found to encode open reading frames devoid of stop codons (amino acid sequences shown). The p-value of this result is 5×10^{-7}. (B) A single sequence control construct with one deliberately installed stop codon. Post-selection products showed enrichment for stop codon mutations or mutations that placed the UAG out of frame. (C) A single sequence control construct with one deliberately installed stop codon and the readthrough signal from the Murine Leukemia Virus positioned appropriately downstream. When translating this mRNA display template in a single reaction alongside the inactive construct shown in (B), post-selection enrichment for the readthrough signal was observed as shown in the adjacent table.

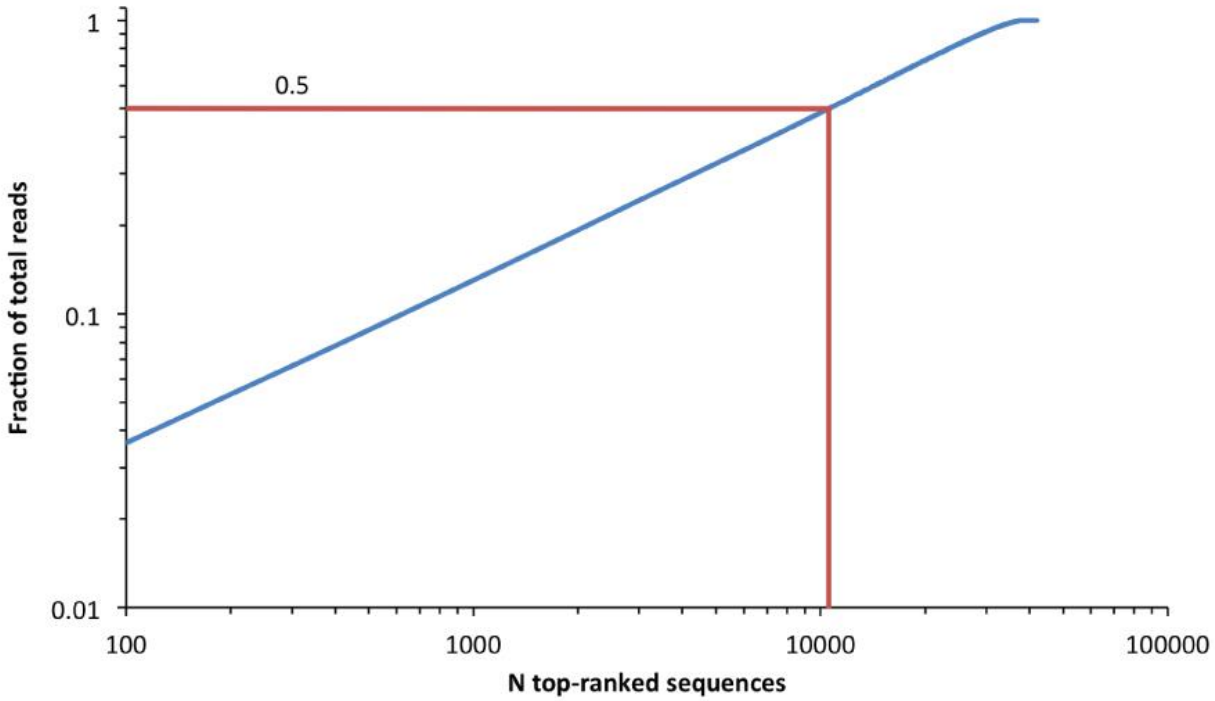
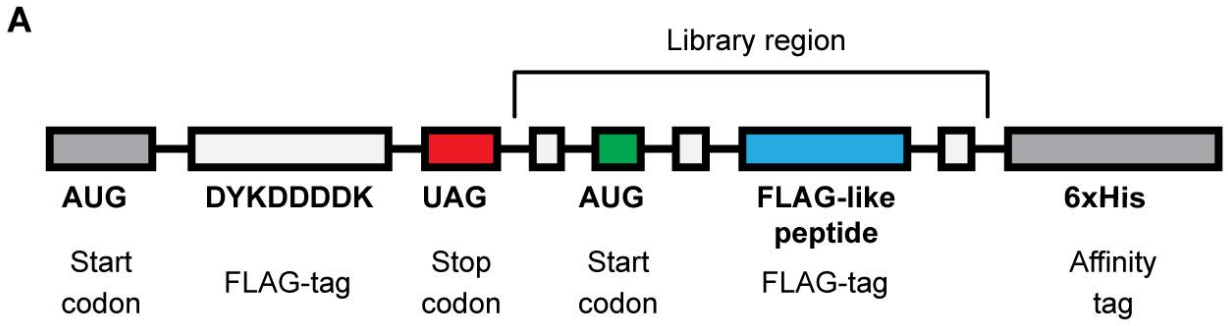


Figure S2. Distribution of post-selection sequencing reads. Plotted is the cumulative fraction of total reads contributed by the top N-ranked sequences ordered by abundance. The top 100 unique sequences account for roughly 3.5% of total reads. Just over 10,000 sequences account for 50% of total sequencing reads.



B

LMGNQMLLHVDYKDSDRSRKSDSAI
QEMSRYAIDYKDHDLSGSPGNGDVI
TGQTMSDLAMQISDDYKLGDTGHDG
MERTEVRTDYKDGDRTGIRVSSGVS
LSEMKGLSGSDLGSLVTDYKDADVG
GLHMSSDYKDLDTPIPNGRALNLGV
DYLTMKDYKDADLSLPRGDTIVAGL
EAATYRAHMDYKNDDVHPDLVVERG
LFRFCAGEYQIGMADERDYKDADVK
WVYIDQQR*SVMQLLFDYKDLDKTV

Figure S3. Enrichment of FLAG-like peptide encoding sequences. (A) Sequences that encode AUG start codons and downstream FLAG-like peptide epitopes were enriched by *in vitro* selection. Presumably, translation initiation was possible downstream of the UAG stop codons, therefore circumventing the requirement for stop codon readthrough. These sequences demonstrated no readthrough activity in cellular assays. (B) Examples of sequences enriched by this mechanism. Methionine start codons are shown in green, FLAG-like peptide epitopes are shown in blue, and the asterisk denotes a stop codon. Sequences containing putative FLAG epitopes were removed from the readthrough dataset and compiled into a 'decoy' sequence set.

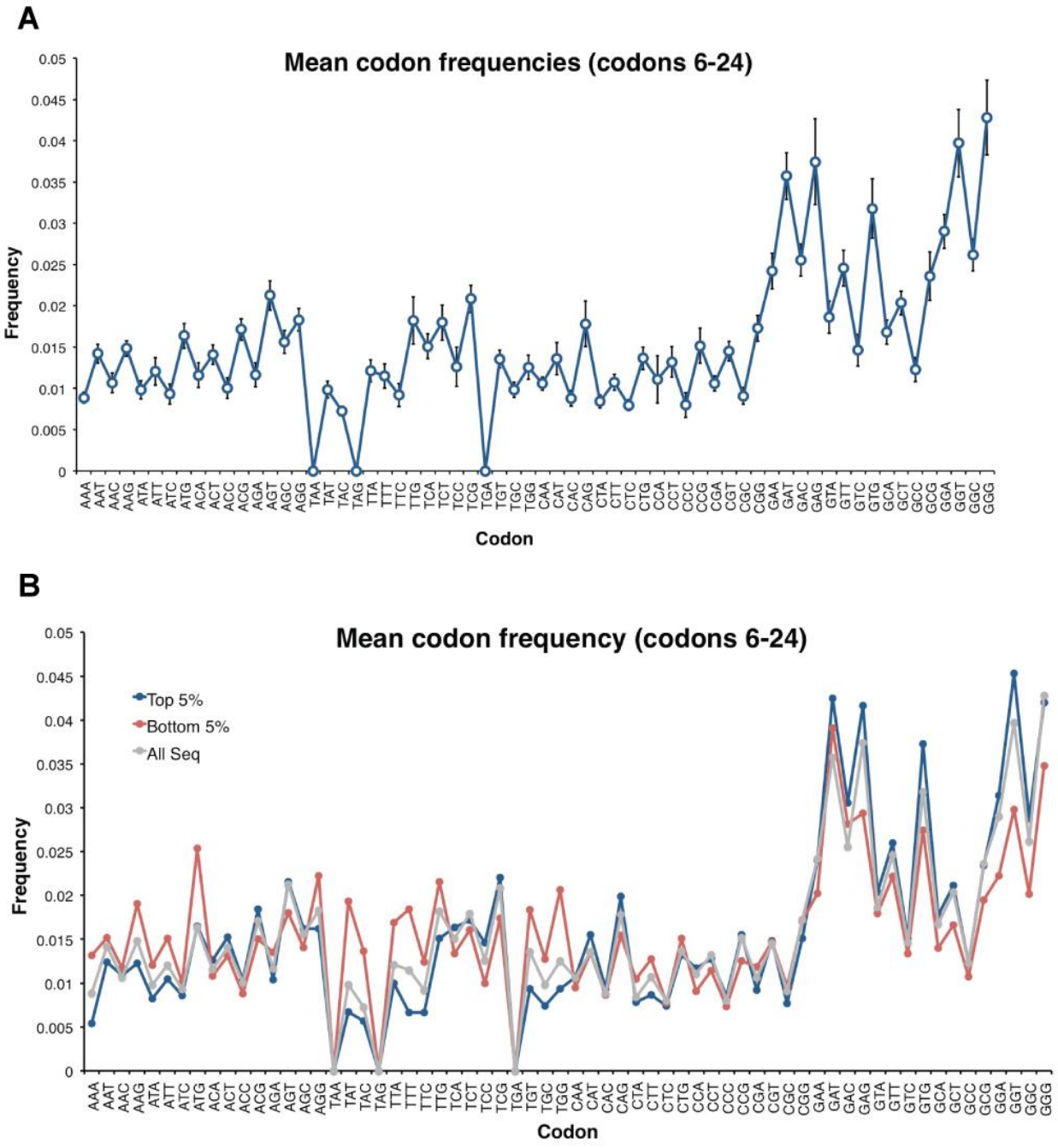


Figure S4. Codon frequency distributions in post-selection enriched sequences. Downstream library codon frequencies were assessed to determine potential codon usage bias, and for normalization of relative enrichments. (A) Mean triplet sequence frequency for codon positions 6 through 24 (nucleotides 16-72) within the library. Error bars represent the standard deviation. (B) Same plot as in (A), with the addition of codon frequency distributions for the sets of sequences ranked in the top 5% and bottom 5% by abundance.

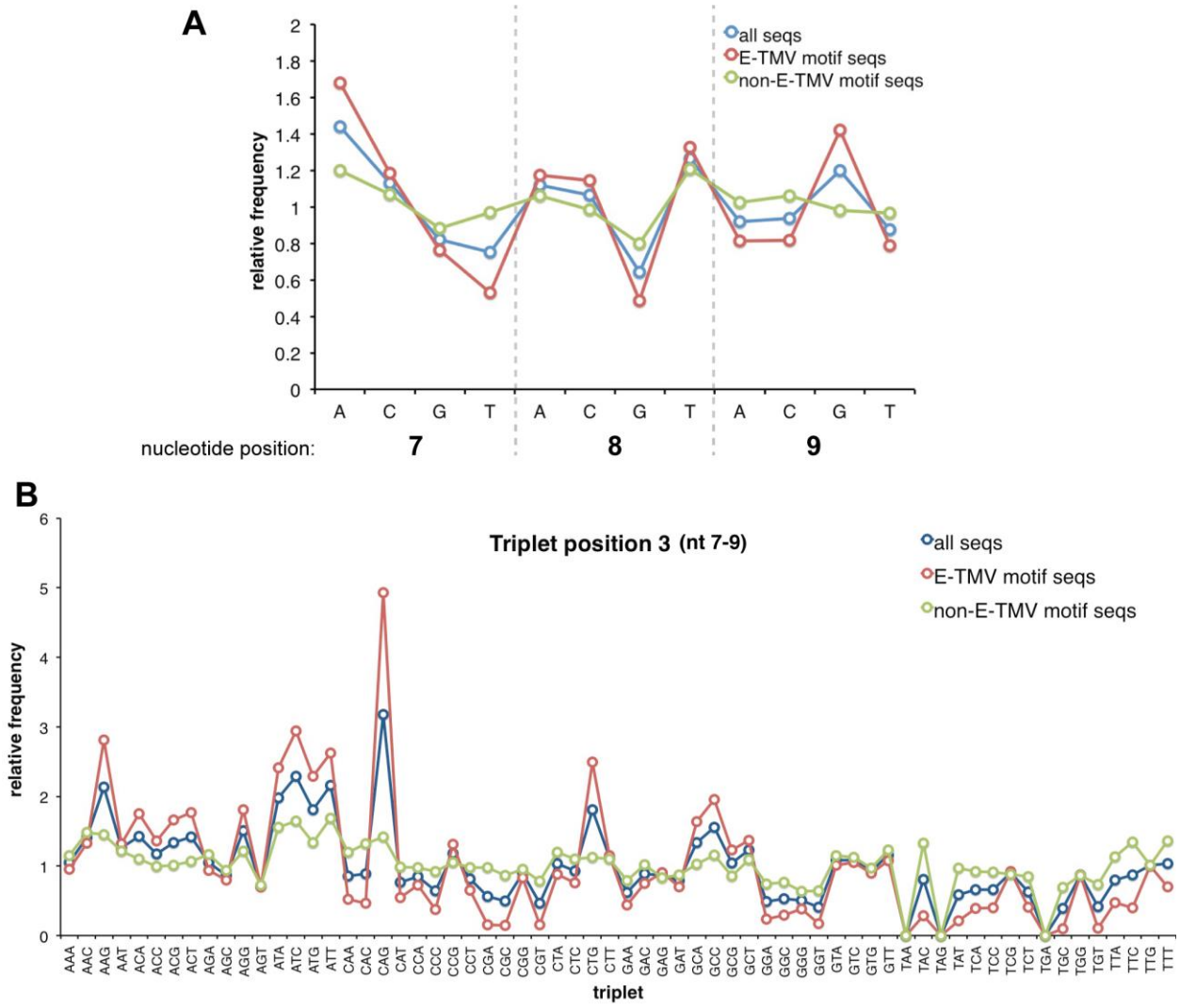


Figure S5. Sequence enrichment normalized to library prevalence. (A) Nucleotide frequencies in positions 7 through 9 were normalized to downstream nucleotide frequencies in the corresponding reading frame (i.e. position 7 is normalized to the mean of positions 16, 19, 22, ... , 70). Frequency distributions are plotted for all readthrough sequences, those that fall into the E-TMV motif set, and those that do not fall in the E-TMV motif set. (B) Triplet frequencies for triplet position 3 (nucleotides 7-9) were normalized to downstream codon frequencies (see Supp. Fig. 4). Frequency distributions are plotted for the full set of readthrough sequences, the E-TMV motif set, and the non-E-TMV motif set.

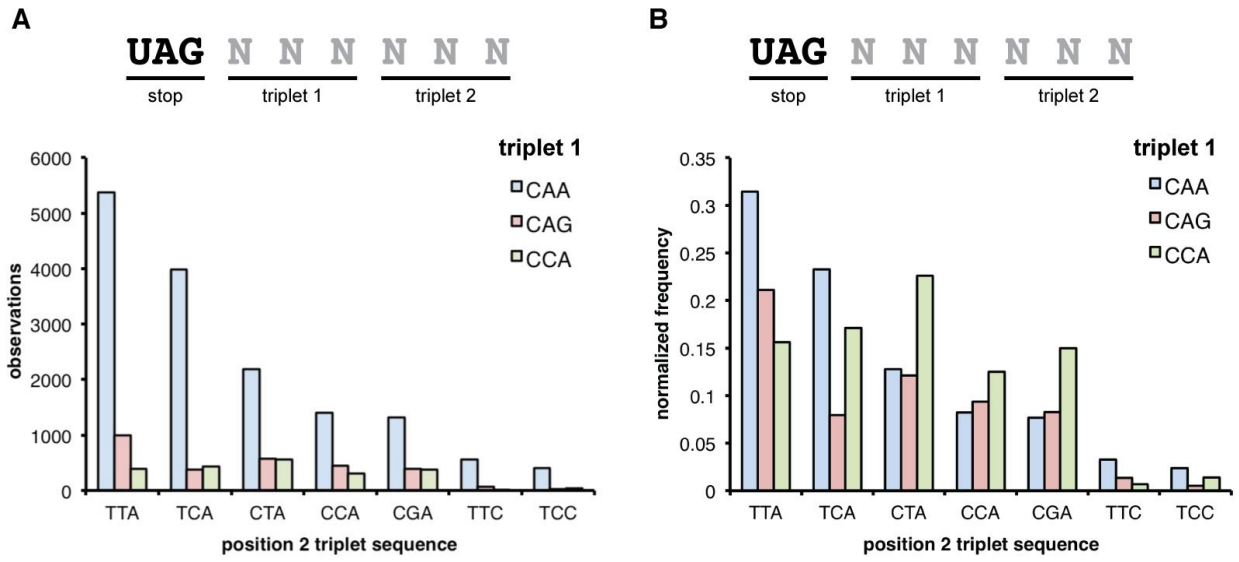


Figure S6. E-TMV motif sequence pairings. The E-TMV motif comprises a set of hexanucleotide sequences that begin with the triplets CAA, CAG, and CAA. These are followed by a subset of triplets including TTA, TCA, CTA, CCA, and CGA. When CAA occupies the first triplet position, TTC and TCC are also frequently observed. (A) Unique sequence observations containing the E-TMV motif triplet pairings are plotted as indicated. (B) E-TMV motif triplet pairings normalized to the total number of CAA, CAG, or CAA containing sequences.

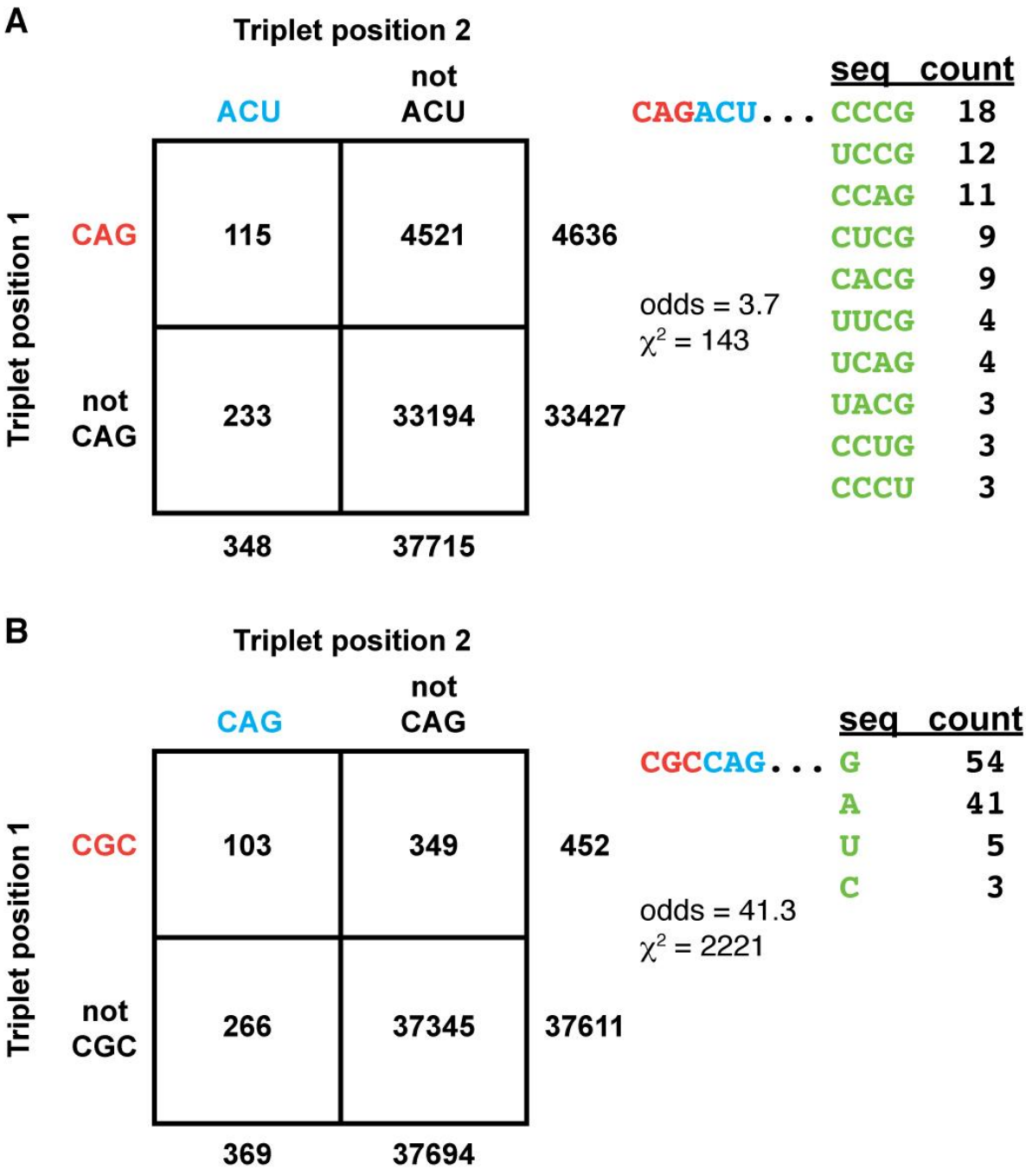


Figure S7. Extended primary sequence motifs. (A) The CAG-ACU pairing was discovered by Fisher's exact test. The contingency table for this association is shown with the odds ratio (OR) and chi-squared (χ^2) statistic. The table to the right shows the ranked abundances of tetranucleotide sequences following CAGACU. (B) The CGC-CAG pairing. The contingency table for this association is shown with the odds ratio (OR) and chi-squared (χ^2) statistic. The table to the right shows the ranked abundances of nucleotides following CGCCAG, forming a consensus sequence of CGCCAGR

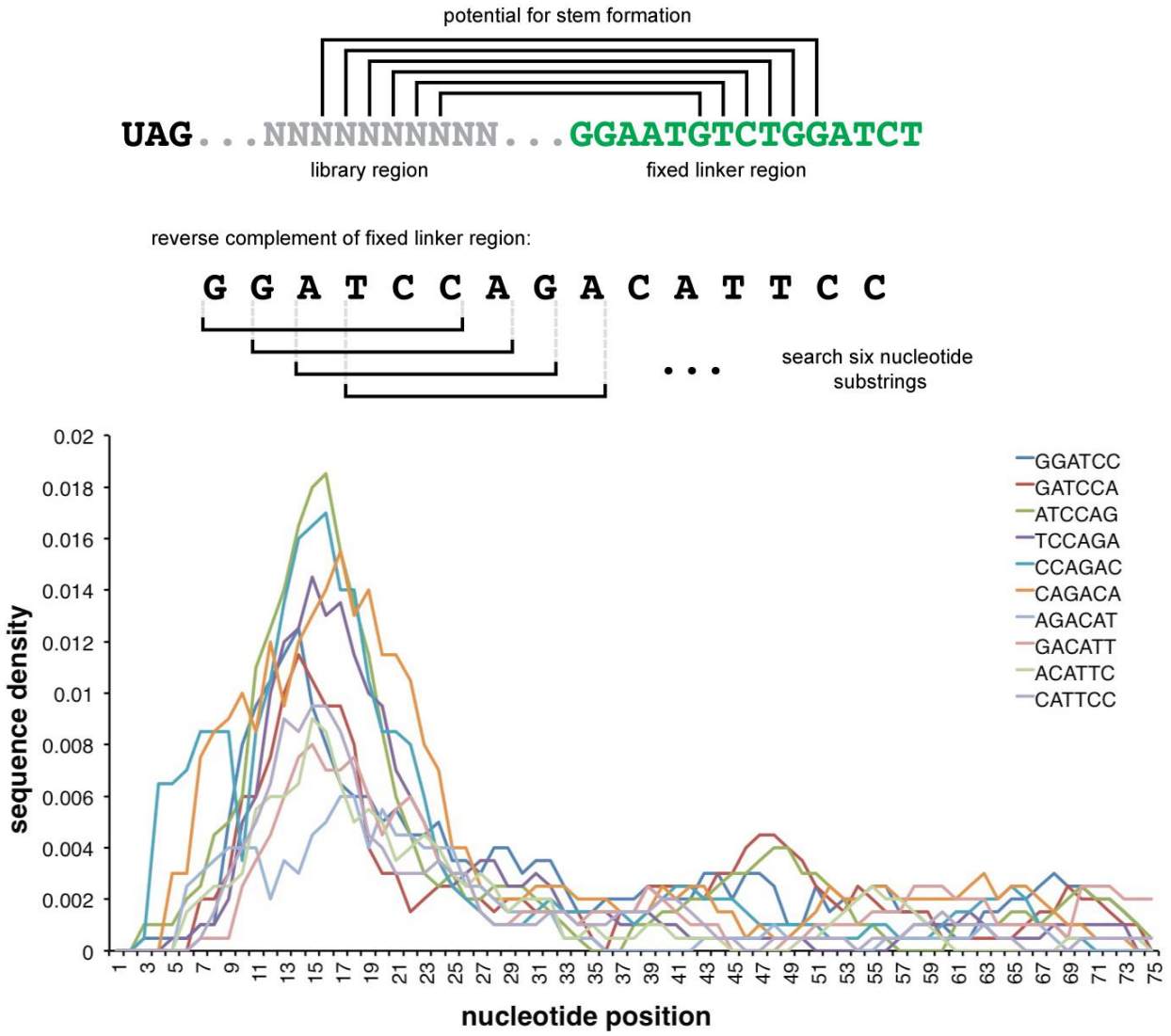


Figure S8. Downstream primary sequence enrichment. Library members that contain sequence strings that are complementary to a downstream fixed linker sequence have the capacity to form RNA stem structures. This was evaluated by mapping the density of 6-nucleotide windows of complementary hexanucleotide strings over the length of the 75-nucleotide library region. Density of individual hexanucleotide strings is shown. See **Fig. 3a** for averaged substring density.

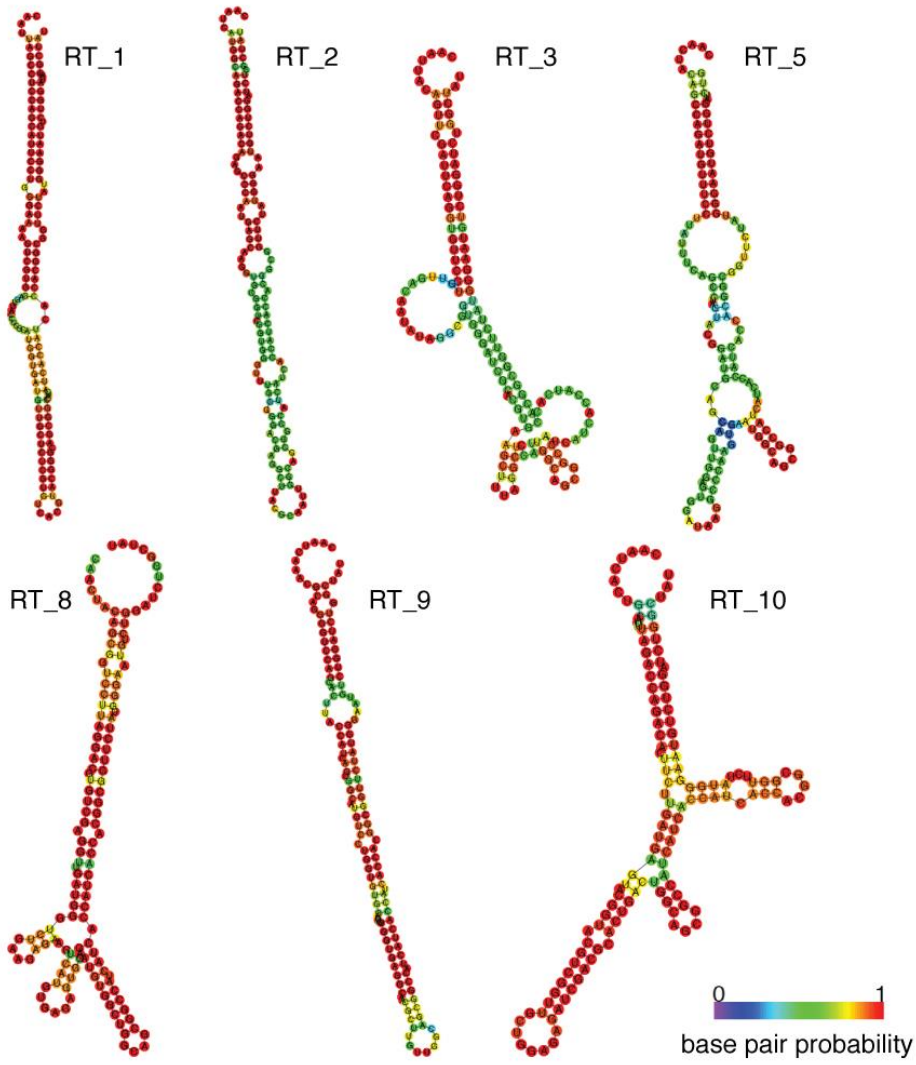


Figure S9. Predicted secondary structures for top ranked library members. Extensive secondary structure predictions are shown for 7 out of the top 10 top ranked library members (rank indicated by the numbering RT_N). Base coloring represents base pairing probability, as indicated by the legend.

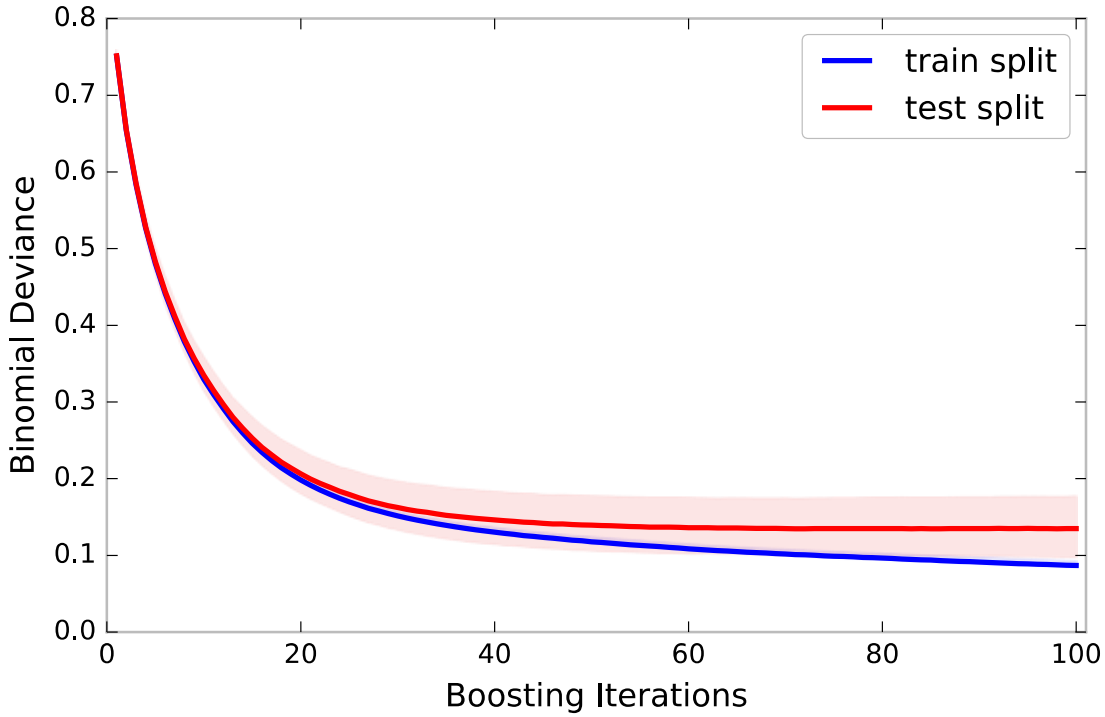


Figure S10. Classification performance on held out data. No evidence of overfitting, captured by test set loss, is observed over the 100 rounds of boosting under 10-fold cross validation.

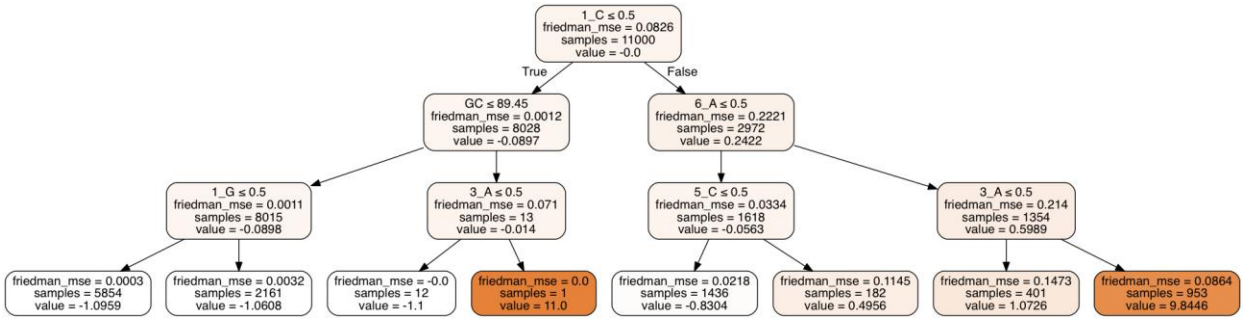


Figure S11. Decision trees as weak rules. The first weak rule incorporated into the boosted model discriminates, at the top level, on having cytosine at the 1st position. Depending on the status, having GC content greater than 89.45% or having guanine at the 6th position is the next discriminating feature, and so forth. An additive model of 100 such decision trees (weak rules) comprise the overall classifier.

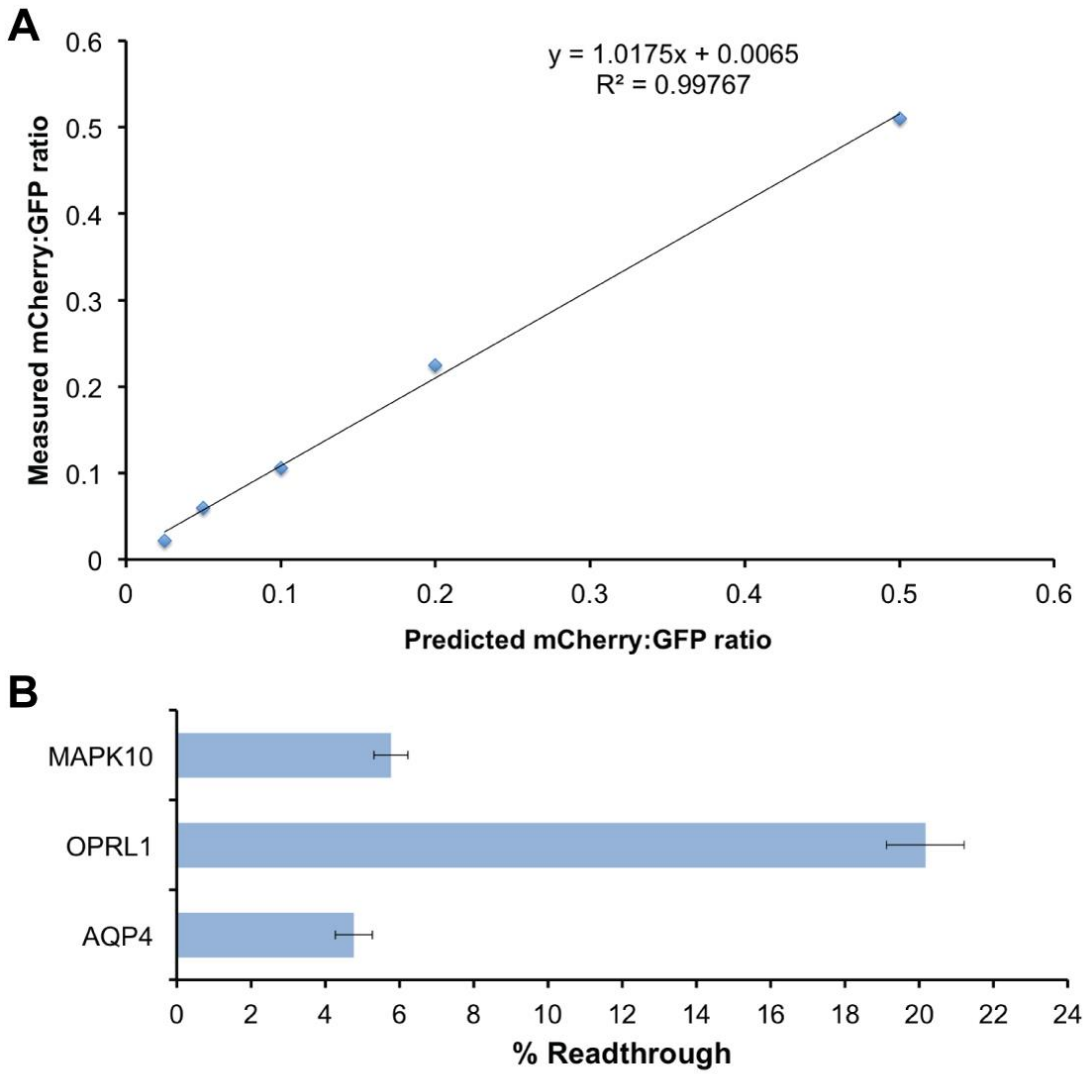


Figure S12. Validation of CMV-dualFP vector for readthrough assays in HEK293T cells. (A) A full-length EGFP-mCherry fusion construct was co-transfected with a construct containing a stop codon separating the EGFP and mCherry open reading frames (without a readthrough element). These two plasmids were transfected at different ratios, EGFP and mCherry fluorescence recorded, and the fluorescence ratio normalized to the single EGFP-mCherry plasmid transfection (set to 1). (B) Measurement of readthrough activity for three previously reported human 3'-UTRs in the dual-FP reporter.

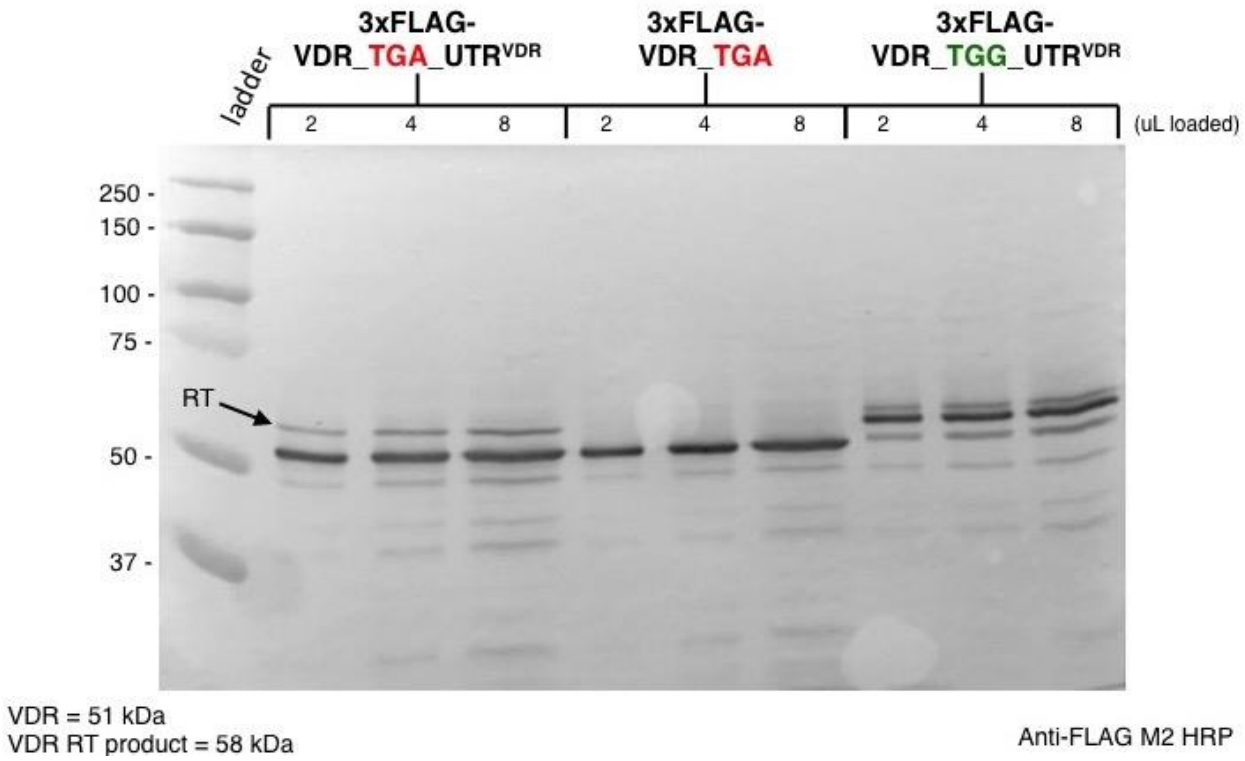


Figure S13. Full western blot of VDR readthrough in HEK293T cells. 293T cells were transfected with VDR expression plasmids as described in the Materials and Methods section. Lane 1: ladder. Lanes 2-4, 3xFLAG-tag VDR encoding its cognate UGA stop codon and the first 201 nucleotides of its 3'-UTR, with 2, 4, or 8 uL of lysate loaded. Lanes 5-7, 3xFLAG-tag VDR encoding its cognate UGA stop codon and no additional VDR 3'-UTR sequence, with 2, 4, or 8 uL of lysate loaded. Lanes 8-10, 3xFLAG-tag VDR encoding a UGG tryptophan codon in place of the cognate UGA stop codon and the first 201 nucleotides of its 3'-UTR, with 2, 4, or 8 uL of lysate loaded.

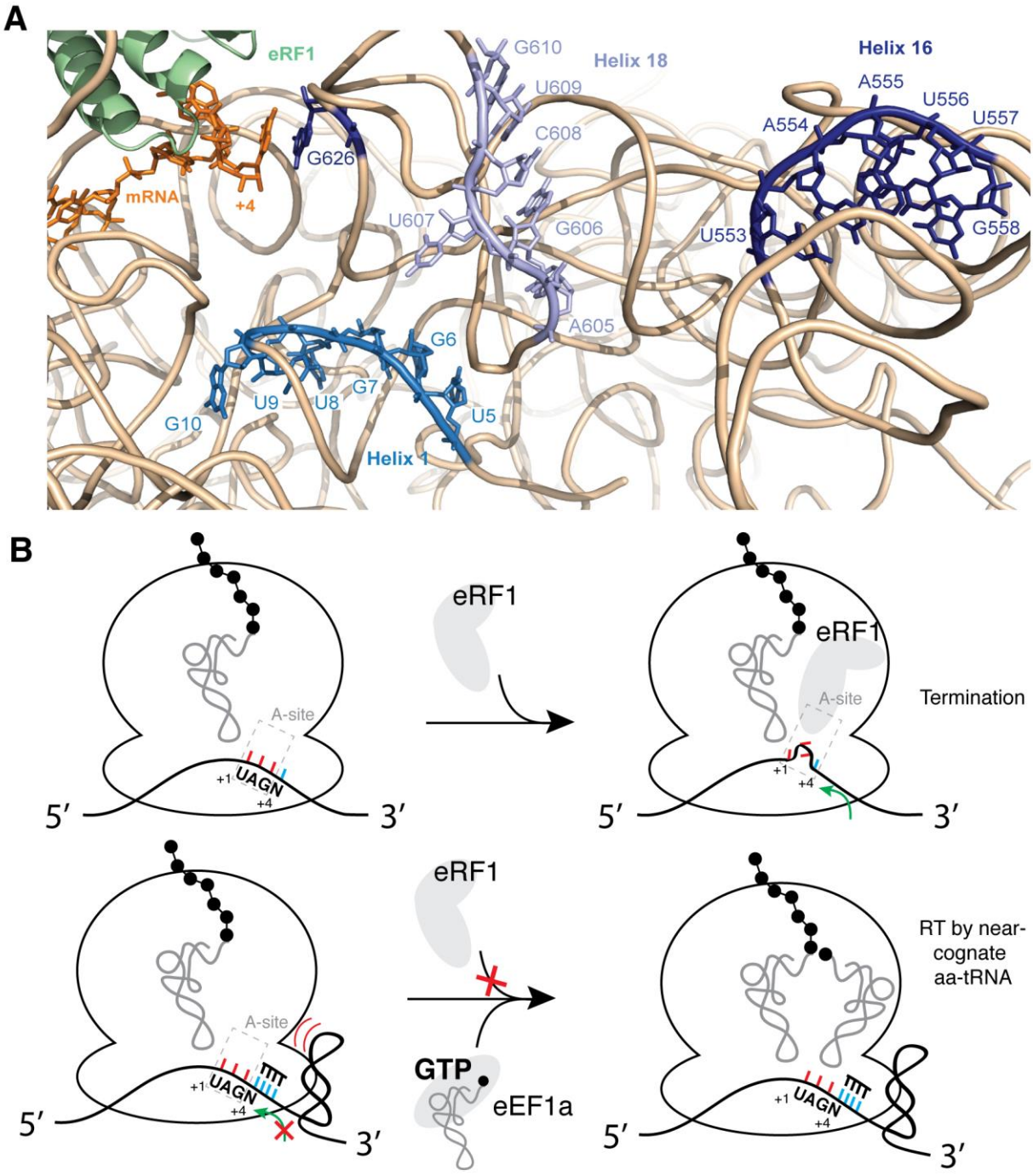


Figure S14. Complementary sequences in 18S rRNA and mechanistic interpretations. (A) Cryo-EM structure of eRF1 in active recognition of a stop codon, generated from pdb 3jag. eRF1 (green) is shown binding to the stop codon within the mRNA (orange). Two stretches of sequence in helix 1 (marine) and helix 18 (steel blue), located in proximity to the mRNA tunnel, are complementary to primary sequence readthrough motifs CAAYYA and CAGACU. Helix 16 also contains a stretch of nucleotides (dark blue) complementary to the TMV motif, but is distant from the decoding center. (B) Mechanistic interpretation of readthrough by primary sequence and secondary structure motifs. Normal stop codon recognition by eRF1 induces a contraction of

the mRNA within the A-site. This motion can be inhibited by mRNA sequences that interact, perhaps by direct base pairing, with the ribosomal small subunit. Furthermore, RNA structural elements situated at the entryway of the mRNA tunnel could inhibit the requisite mRNA motions. This inhibition of translation termination provides time for decoding stop codons with near-cognate aminoacyl-tRNAs.

Table S1. DNA sequences of *in vitro* selection constructs.

Construct	Sequence
No-stop Lib	**TCTAATACGACTCACTATA**GGG*ACAATTACTATTTACAATTACA*ATGGACTACAAGGACGAC GACGACAAGTACNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNGGCAGCGGCCATCATCACCATCACCACG GCGGTTCTATGGGAATGTCTGGATCTGGCTAT
Inactive	**TCTAATACGACTCACTATA**GGG*ACAATTACTATTTACAATTACA*ATGGACTACAAGGACGAC GACGACAAGACCCTAGATGACTAGCACCGAGAACCCGGCGCCACGCAATGGAACGTCCTT AACTCCGGCAGGCAATTAAAGGGAACGTATGGCAGCGGCCATCATCACCATCACCACGGC GGTTCTATGGGAATGTCTGGATCTGGCTAT
MLV	**TCTAATACGACTCACTATA**GGG*ACAATTACTATTTACAATTACA*ATGGACTACAAGGACGAC GACGACAAGACCCTAGATGACTAGGGAGGTCAGGGTCAGGAGCCCCCCCCTGAACCCAG GATAACCCTCAAAGTCGGGGGGCAACCCGTCGGCAGCGGCCATCATCACCATCACCACGG CGGTTCTATGGGAATGTCTGGATCTGGCTAT
RT-Lib	**TCTAATACGACTCACTATA**GGG*ACAATTACTATTTACAATTACA*ATGGACTACAAGGACGAC GACGACAAGTAGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNGGCAGCGGCCATCATCACCATCACCACG GCGGTTCTATGGGAATGTCTGGATCTGGCTAT

Bold sequence: **T7 promoter**; italic sequence: *TMV enhancer*; underlined sequence: FLAG-tag; bold underline: **stop codon**; double underline: His_6-tag.

Table S2. Oligonucleotide used in this study.

Construct	Sequence
RT-Lib ultramer	GACTACAAGGACGACGACGACAAGTAGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNGGCAGCGGCCATCATCACCATC
Control Lib ultramer	GACTACAAGGACGACGACGACAAGTACNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNGGCAGCGGCCATCATCACCATC
CT-for	TCTAATACGACTCACTATAGGGACAATTACTATTTACAATTACAATGGACTACAAGGACGACGACGACAAGTAC
RT-for	TCTAATACGACTCACTATAGGGACAATTACTATTTACAATTACAATGGACTACAAGGACGACGACGACAAGTAG
RT-rev	ATAGCCAGATCCAGACATTCCCATAGAACCGCCGTGGTGATGGTGATGATGGCCGCTGCC
AVA95	TTTTTTTTTTTTATAGCCAGATCC
Phospho-A_{27}CC-Puro	pAAAAAAAAAAAAAAAAAAAAAAAAAAACC-puromycin
AVA314	TTCCCTACACGACGCTCTTCCGATCTTACAAGGACGACGACGACAAGTAG
AVA315	TTCCCTACACGACGCTCTTCCGATCT**N**TACAAGGACGACGACGACAAGTAG
AVA316	TTCCCTACACGACGCTCTTCCGATCT**NN**TACAAGGACGACGACGACAAGTAG
AVA319	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATC
AVA320	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCATGGTGATGATGGCCGCTGCC
AVA321	CAAGCAGAAGACGGCATACGAGATCGTGATGTGACTGGAGTTCAGACGTGTGCTC

Table S3. Plasmids used in this study.

Plasmid	Construct details	Comments
pCR-NegRT	T7-TAG-InactiveRT	Kan^R, Amp^R
pCR-MLV-PK	T7-TAG-MLV-PK	Kan^R, Amp^R
p425-dualFP[§]	p425-pTHD3-yEGFP-mCherry	Leu2, 2-micron, Amp^R
p425-UAG-RT1	p425-pTHD3-yEGFP-UAG-RT1-mCherry	Leu2, 2-micron, Amp^R
p425-UAG-RT2	p425-pTHD3-yEGFP-UAG-RT2-mCherry	Leu2, 2-micron, Amp^R
p425-UAG-RT3	p425-pTHD3-yEGFP-UAG-RT3-mCherry	Leu2, 2-micron, Amp^R
p425-UAG-RT4	p425-pTHD3-yEGFP-UAG-RT4-mCherry	Leu2, 2-micron, Amp^R
p425-UAG-RT5	p425-pTHD3-yEGFP-UAG-RT5-mCherry	Leu2, 2-micron, Amp^R
p425-UAG-RT6	p425-pTHD3-yEGFP-UAG-RT6-mCherry	Leu2, 2-micron, Amp^R
p425-UAG-RT7	p425-pTHD3-yEGFP-UAG-RT7-mCherry	Leu2, 2-micron, Amp^R
p425-UAG-RT8	p425-pTHD3-yEGFP-UAG-RT8-mCherry	Leu2, 2-micron, Amp^R
p425-UAG-RT9	p425-pTHD3-yEGFP-UAG-RT9-mCherry	Leu2, 2-micron, Amp^R
p425-UAG-RT10	p425-pTHD3-yEGFP-UAG-RT10-mCherry	Leu2, 2-micron, Amp^R
p425-UAG-RT11	p425-pTHD3-yEGFP-UAG-RT11-mCherry	Leu2, 2-micron, Amp^R
p425-UAG-RT12	p425-pTHD3-yEGFP-UAG-RT12-mCherry	Leu2, 2-micron, Amp^R
p425-UAG-RT13	p425-pTHD3-yEGFP-UAG-RT13-mCherry	Leu2, 2-micron, Amp^R
p425-UAG-RT14	p425-pTHD3-yEGFP-UAG-RT14-mCherry	Leu2, 2-micron, Amp^R
p425-UAG-RT15	p425-pTHD3-yEGFP-UAG-RT15-mCherry	Leu2, 2-micron, Amp^R
p425-UAG-RT16	p425-pTHD3-yEGFP-UAG-RT16-mCherry	Leu2, 2-micron, Amp^R
p425-UAG-RT17	p425-pTHD3-yEGFP-UAG-RT17-mCherry	Leu2, 2-micron, Amp^R
p425-UAG-RT18	p425-pTHD3-yEGFP-UAG-RT18-mCherry	Leu2, 2-micron, Amp^R
p425-UAG-RT19	p425-pTHD3-yEGFP-UAG-RT19-mCherry	Leu2, 2-micron, Amp^R
p425-UAG-RT20	p425-pTHD3-yEGFP-UAG-RT20-mCherry	Leu2, 2-micron, Amp^R
p425-UAG-RT21	p425-pTHD3-yEGFP-UAG-RT21-mCherry	Leu2, 2-micron, Amp^R
p425-UAG-RT22	p425-pTHD3-yEGFP-UAG-RT22-mCherry	Leu2, 2-micron, Amp^R
p425-UAG-RT23	p425-pTHD3-yEGFP-UAG-RT23-mCherry	Leu2, 2-micron, Amp^R
p425-UAG-RT24	p425-pTHD3-yEGFP-UAG-RT24-mCherry	Leu2, 2-micron, Amp^R
p425-UAG-RT25	p425-pTHD3-yEGFP-UAG-RT25-mCherry	Leu2, 2-micron, Amp^R
p425-UAA-RT1	p425-pTHD3-yEGFP-UAA-RT1-mCherry	Leu2, 2-micron, Amp^R
p425-UAA-RT2	p425-pTHD3-yEGFP-UAA-RT2-mCherry	Leu2, 2-micron, Amp^R
p425-UAA-RT8	p425-pTHD3-yEGFP-UAA-RT8-mCherry	Leu2, 2-micron, Amp^R
p425-UAA-RT9	p425-pTHD3-yEGFP-UAA-RT9-mCherry	Leu2, 2-micron, Amp^R
p425-UAA-RT10	p425-pTHD3-yEGFP-UAA-RT10-mCherry	Leu2, 2-micron, Amp^R
p425-UAA-RT11	p425-pTHD3-yEGFP-UAA-RT11-mCherry	Leu2, 2-micron, Amp^R

p425-UGA-RT1	p425-pTHD3-yEGFP-UGA-RT1-mCherry	Leu2, 2-micron, Amp^R
p425-UGA-RT2	p425-pTHD3-yEGFP-UGA-RT2-mCherry	Leu2, 2-micron, Amp^R
p425-UGA-RT8	p425-pTHD3-yEGFP-UGA-RT8-mCherry	Leu2, 2-micron, Amp^R
p425-UGA-RT9	p425-pTHD3-yEGFP-UGA-RT9-mCherry	Leu2, 2-micron, Amp^R
p425-UGA-RT10	p425-pTHD3-yEGFP-UGA-RT10-mCherry	Leu2, 2-micron, Amp^R
p425-UGA-RT11	p425-pTHD3-yEGFP-UGA-RT11-mCherry	Leu2, 2-micron, Amp^R
pCMV-dualFP‡	pCMV-EGFP-mCherry	via Addgene 20972; Amp^R
pCMV-RT1	pCMV-EGFP-RT1-mCherry	Amp^R
pCMV-RT3	pCMV-EGFP-RT3-mCherry	Amp^R
pCMV-RT9	pCMV-EGFP-RT9-mCherry	Amp^R
pCMV-RT18	pCMV-EGFP-RT18-mCherry	Amp^R
pCMV-AQP4	pCMV-EGFP-AQP4-mCherry	Amp^R
pCMV-MAPK10	pCMV-EGFP-MAPK10-mCherry	Amp^R
pCMV-OPRL1	pCMV-EGFP-OPRL1-mCherry	Amp^R
pCMV-PCSK9	pCMV-EGFP-PCSK9-mCherry	Amp^R
pCMV-IL18	pCMV-EGFP-IL18-mCherry	Amp^R
pCMV-PTDSS2	pCMV-EGFP-PTDSS2-mCherry	Amp^R
pCMV-VGLL2	pCMV-EGFP-GLL2-mCherry	Amp^R
pCMV-ERVV2	pCMV-EGFP-ERVV2-mCherry	Amp^R
pCMV-SYT12	pCMV-EGFP-SYT12-mCherry	Amp^R
pCMV-RXFP2	pCMV-EGFP-RXFP2-mCherry	Amp^R
pCMV-CCNE1	pCMV-EGFP-CCNE1-mCherry	Amp^R
pCMV-PVRL3	pCMV-EGFP-PVRL3-mCherry	Amp^R
pCMV-VDR	pCMV-EGFP-VDR -mCherry	Amp^R
pCMV-trACP2	pCMV-EGFP-trACP2-mCherry	Amp^R
pCMV-ACP2	pCMV-EGFP-ACP2-mCherry	Amp^R
pCMV-EDN1	pCMV-EGFP-EDN1-mCherry	Amp^R
pCMV-CDKN2B	pCMV-EGFP-CDKN2B-mCherry	Amp^R
pCMV-KCNF1	pCMV-EGFP-KCNF1-mCherry	Amp^R
pCMV-MYNN	pCMV-EGFP-MYNN-mCherry	Amp^R
pCMV-NKX2-4	pCMV-EGFP-NKX2-4-mCherry	Amp^R
pCMV-KIF3C	pCMV-EGFP-KIF3C-mCherry	Amp^R
pCMV-PRR20A	pCMV-EGFP-PRR20A-mCherry	Amp^R
pCMV-FGD5	pCMV-EGFP-FGD5-mCherry	Amp^R
pCMV-WHSC1	pCMV-EGFP-WHSC1-mCherry	Amp^R

pCMV-SFTA2	pCMV-EGFP-SFTA2-mCherry	Amp^R
pCMV-LEPROTL1	pCMV-EGFP-LEPROTL1-mCherry	Amp^R
pVDR-TGA	pCMV-3xFLAG-VDR-TGA-Δutr	via Addgene 11680; Kan^R
pVDR-TGA-utr	pCMV-3xFLAG-VDR-TGA-utr(1-201)	Kan^R
pVDR-TGG-utr	pCMV-3xFLAG-VDR-TGG-utr(1-201)	Kan^R

§ p425-dualFP plasmid was constructed previously as described in ref. Anzalone, et al. *Nat. Methods* **2016**, *13*, 453; contains NheI and AatII restriction sites for cloning new RT inserts.
‡ pCMV-dualFP plasmids contain NheI and BamHI restriction sites for cloning new RT inserts.

Table S4. Sequences of synthetic RT constructs.

Construct	Sequence
GFP-mCherry fusion	GGCAGCGGCGACTACAAGGACGACGACGACAAG**TAT**GACACTGCTCGAAGAAAAGATACG AAGGAGGCGGCCTGGCCGGAGCACTACCGCATCGACGCGTATTCGAATACTGGCAGCGGC CATCATCACCATCACCACGGCGGTTCTATGGGAATGTCTGGA
RT-1	GGCAGCGGCGACTACAAGGACGACGACGACAAG**TAG**GACACTGCTCGAAGAAAAGATAC GAAGGAGGCGGCCTGGCCGGAGCACTACCGCATCGACGCGTATTCGAATACTGGCAGCGG CCATCATCACCATCACCACGGCGGTTCTATGGGAATGTCTGGA
RT-2	GGCAGCGGCGACTACAAGGACGACGACGACAAG**TAG**CAATTAGCTCGAAGAAAAGATAC GAAGGAGGCGGCCTGGCCGGAGCACTACCGCATCGACGCGTATTCGAATACTGGCAGCGG CCATCATCACCATCACCACGGCGGTTCTATGGGAATGTCTGGA
RT-3	GGCAGCGGCGACTACAAGGACGACGACGACAAG**TAG**CAATTACAGCGAAGAAAAGATAC GAAGGAGGCGGCCTGGCCGGAGCACTACCGCATCGACGCGTATTCGAATACTGGCAGCGG CCATCATCACCATCACCACGGCGGTTCTATGGGAATGTCTGGA
RT-4	*C*GGCAGCGGCGACTACAAGGACGACGACGACAAG**TAG**CAATTAGGACGAAGAAAAGATA CGAAGGAGGCGGCCTGGCCGGAGCACTACCGCATCGACGCGTATTCGAATACTGGCAGCG GCCATCATCACCATCACCACGGCGGTTCTATGGGAATGTCTGGA
RT-5	GGCAGCGGCGACTACAAGGACGACGACGACAAG**TAG**CAGTTAGCTCGAAGAAAAGATAC GAAGGAGGCGGCCTGGCCGGAGCACTACCGCATCGACGCGTATTCGAATACTGGCAGCGG CCATCATCACCATCACCACGGCGGTTCTATGGGAATGTCTGGA
RT-6	GGCAGCGGCGACTACAAGGACGACGACGACAAG**TAG**CAACGAGCTCGAAGAAAAGATAC GAAGGAGGCGGCCTGGCCGGAGCACTACCGCATCGACGCGTATTCGAATACTGGCAGCGG CCATCATCACCATCACCACGGCGGTTCTATGGGAATGTCTGGA
RT-7	GGCAGCGGCGACTACAAGGACGACGACGACAAG**TAG**CCACTAGCTCGAAGAAAAGATAC GAAGGAGGCGGCCTGGCCGGAGCACTACCGCATCGACGCGTATTCGAATACTGGCAGCGG CCATCATCACCATCACCACGGCGGTTCTATGGGAATGTCTGGA*GACGTC*
RT-8	GGCAGCGGCGACTACAAGGACGACGACGACAAG**TAG**CGCCAGGCTCGAAGAAAAGATAC GAAGGAGGCGGCCTGGCCGGAGCACTACCGCATCGACGCGTATTCGAATACTGGCAGCGG CCATCATCACCATCACCACGGCGGTTCTATGGGAATGTCTGGA
RT-9	GGCAGCGGCGACTACAAGGACGACGACGACAAG**TAG**CTAGGCGCTCGAAGAAAAGATAC GAAGGAGGCGGCCTGGCCGGAGCACTACCGCATCGACGCGTATTCGAATACTGGCAGCGG CCATCATCACCATCACCACGGCGGTTCTATGGGAATGTCTGGA
RT-10	*GCTAGC*GGCAGCGGCGACTACAAGGACGACGACGACAAG**TAG**CTATCCGCTCGAAGAAAA GATACGAAGGAGGCGGCCTGGCCGGAGCACTACCGCATCGACGCGTATTCGAATACTGGC AGCGGCCATCATCACCATCACCACGGCGGTTCTATGGGAATGTCTGGA
RT-11	GGCAGCGGCGACTACAAGGACGACGACGACAAG**TAG**CAGACTCCCGGAAGAAAAGATAC GAAGGAGGCGGCCTGGCCGGAGCACTACCGCATCGACGCGTATTCGAATACTGGCAGCGG CCATCATCACCATCACCACGGCGGTTCTATGGGAATGTCTGGA
RT-12	GGCAGCGGCGACTACAAGGACGACGACGACAAG**TAG**CAAACTCCCGGAAGAAAAGATAC GAAGGAGGCGGCCTGGCCGGAGCACTACCGCATCGACGCGTATTCGAATACTGGCAGCGG CCATCATCACCATCACCACGGCGGTTCTATGGGAATGTCTGGA
RT-13	GGCAGCGGCGACTACAAGGACGACGACGACAAG**TAG**CAGCATCCCGGAAGAAAAGATAC GAAGGAGGCGGCCTGGCCGGAGCACTACCGCATCGACGCGTATTCGAATACTGGCAGCGG CCATCATCACCATCACCACGGCGGTTCTATGGGAATGTCTGGA
RT-14	GGCAGCGGCGACTACAAGGACGACGACGACAAG**TAG**CAGACTGCTCGAAGAAAAGATAC GAAGGAGGCGGCCTGGCCGGAGCACTACCGCATCGACGCGTATTCGAATACTGGCAGCGG CCATCATCACCATCACCACGGCGGTTCTATGGGAATGTCTGGA
RT-15	GGCAGCGGCGACTACAAGGACGACGACGACAAG**TAG**CAATTAGCTCCCAGACAAGATACG

	AAGGAGGCGGCCTGGCCGGAGCACTACCGCATCGACGCGTATTCGAATACTGGCAGCGGC CATCATCACCATCACCACGGCGGTTCTATGGGAATGTCTGGATCTGGC
RT-16	GGCAGCGGCGACTACAAGGACGACGACGACAAG**TAG**CAATTAGCTGACATTTCCGATACG AAGGAGGCGGCCTGGCCGGAGCACTACCGCATCGACGCGTATTCGAATACTGGCAGCGGC CATCATCACCATCACCACGGCGGTTCTATGGGAATGTCTGGATCTGGC
RT-17	GGCAGCGGCGACTACAAGGACGACGACGACAAG**TAG**CAATTAGCTCATCCAGAAGATACG AAGGAGGCGGCCTGGCCGGAGCACTACCGCATCGACGCGTATTCGAATACTGGCAGCGGC CATCATCACCATCACCACGGCGGTTCTATGGGAATGTCTGGATCTGGC
RT-18	GGCAGCGGCGACTACAAGGACGACGACGACAAG**TAG**CAATTAGCTCATCCAGACGATACG AAGGAGGCGGCCTGGCCGGAGCACTACCGCATCGACGCGTATTCGAATACTGGCAGCGGC CATCATCACCATCACCACGGCGGTTCTATGGGAATGTCTGGATCTGGC
RT-19	GGCAGCGGCGACTACAAGGACGACGACGACAAG**TAG**CAATTAGCTCATCCAGACATTACG AAGGAGGCGGCCTGGCCGGAGCACTACCGCATCGACGCGTATTCGAATACTGGCAGCGGC CATCATCACCATCACCACGGCGGTTCTATGGGAATGTCTGGATCTGGC
RT-20	GGCAGCGGCGACTACAAGGACGACGACGACAAG**TAG**CAATTAGCTCATCCAGACATTCCGG AAGGAGGCGGCCTGGCCGGAGCACTACCGCATCGACGCGTATTCGAATACTGGCAGCGGC CATCATCACCATCACCACGGCGGTTCTATGGGAATGTCTGGATCTGGC
RT-21	GGCAGCGGCGACTACAAGGACGACGACGACAAG**TAG**CAATTAGCTCATCCAGACATTCCCC ATAGAGGCGGCCTGGCCGGAGCACTACCGCATCGACGCGTATTCGAATACTGGCAGCGGCC ATCATCACCATCACCACGGCGGTTCTATGGGAATGTCTGGATCTGGC
RT-22	GGCAGCGGCGACTACAAGGACGACGACGACAAG**TAG**CAATTAGCTCATCCAGACATTCCCC ATAGAGGCGGCCTGGCCGGAGCACTACCGCATCGACGCGTATTCGAATACTGGCAGCGGCC ATCATCACCATCACCACGGCGGTTCTATGGGAACTGGC
RT-23	GGCAGCGGCGACTACAAGGACGACGACGACAAG**TAG**GACACTGCTCATCCAGACATTCCCC ATAGAGGCGGCCTGGCCGGAGCACTACCGCATCGACGCGTATTCGAATACTGGCAGCGGCC ATCATCACCATCACCACGGCGGTTCTATGGGAATGTCTGGATCTGGC
RT-24	GGCAGCGGCGACTACAAGGACGACGACGACAAG**TAG**CGCCAGGCTCATCCAGACATTCCCC ATAGAGGCGGCCTGGCCGGAGCACTACCGCATCGACGCGTATTCGAATACTGGCAGCGGCC ATCATCACCATCACCACGGCGGTTCTATGGGAATGTCTGGATCTGGC
RT-25	GGCAGCGGCGACTACAAGGACGACGACGACAAG**TAG**CTAGGCGCTCATCCAGACATTCCCC ATAGAGGCGGCCTGGCCGGAGCACTACCGCATCGACGCGTATTCGAATACTGGCAGCGGCC ATCATCACCATCACCACGGCGGTTCTATGGGAATGTCTGGATCTGGC

For yeast or mammalian vector constructs, sequences were cloned between EGFP and mCherry open reading frames; bold: stop codon; single underline: primary sequence RT motif; double underline: stem forming sequence. For constructs containing UAA or UGA stop codons, only the stop codon portion of the construct was altered.

Table S5. Sequences of human 3'-UTR constructs.

Construct	Sequence
VDR	ctcggcatggacgagctgtacaaagctagcggcagcggcGGCAATGAGATCTCC**TGA**CTAGGACAGCCTGTGGC GGTGCCTGGGTGGGGCTGCTCCTCCAGGGCCACGTGCCAGGCCCGGGGCTGGCGGCTACT CAGCAGCCCTCCTCACCCCGTCTGGGGTTCAGCCCCTCCTCTGCCACCTCCCCTATCCcacgg cggttctatgggaatgtctggggatccaccggtcgccaccatggtgagcaagg
ACP2	ctcggcatggacgagctgtacaaagctagcggcagcggcgactacaaggacgacgacgacaagCACGCC**TGA**CAACCACTCA GCCCCCTTCCCTCCACCTCC**A**AGGGGAGGTGGGCTGGGCCCTCGCTCCTGACTGTTGCTGC TCCCggcagcggccatcatcaccatcaccacggcggttctatgggaatgtctggggatccaccggtcgccaccatggtgagcaagg
trACP2	ctcggcatggacgagctgtacaaagctagcggcagcggcgactacaaggacgacgacgacaagCACGCC**TGA**CAACCACTCA GCCCCCTTCCCTCCACCTCCcacggcggttctatgggaatgtctggggatccaccggtcgccaccatggtgagcaagg
PVRL3	ctcggcatggacgagctgtacaaagctagcggcagcggcgactacaaggacgacgacgacaagAGGGAGTGGTATGTT**TAG**C AACCACTGAATGTGACTTTACTATGTACAATGTTCATTCACACTAGTAGATCATTTTCAGA TTGTTCATACTTTTTCTTGAGGAAGAATAAGCTTTTTCAAGTTGATGGCAGCGGCCATCAT CACCATCACcacggcggttctatgggaatgtctggggatccaccggtcgccaccatggtgagcaagg
CCNE1	ctcggcatggacgagctgtacaaagctagcggcagcggcGGGCCGGAAATGGCG**TGA**CCACCCCATCCTTCTCC ACCAAAGACAGTTGCGCGCCTGCTCCACGTTCTCTTCTGTCTGTTGCAGCGGAGGCGTGCG TTTGCTTTTACAGATATCTGTATGGAAGAGTGTTTCTTCCACAACAGAAGTATTTCTcacggcg gttctatgggaatgtctggggatccaccggtcgccaccatggtgagcaagg
RXFP2	ctcggcatggacgagctgtacaaagctagcggcagcggcgactacaaggacgacgacgacaagATGAAACCAGTTTCC**TAG**C AATCATTTTGGATCACTGGACTTTCAGTGGACTACCTAAAACAGGGGACAGCTTTTGGAAG AGGCAGCGGCCATCATCACCATCACcacggcggttctatgggaatgtctggggatccaccggtcgccaccatggtgagcaa gg
SYT12	ctcggcatggacgagctgtacaaagctagcggcagcggcgactacaaggacgacgacgacaagGCTGTCCGGCGAAAC**TAG**C AACCAGGGCGGGCCAGTTGGGCAATGGAGCTGCTGGAGCCCGGTACCCACTCAGCTCTGT CTGATGCCCTCTCCATAGCCCAGCTGGAGCCGTGTACACTGGGGTCCCCTGGCAGAGTCCT CATGACCCATCCTggcagcggccatcatcaccatcaccacggcggttctatgggaatgtctggggatccaccggtcgccaccatggtga gcaagg
ERVV2	ctcggcatggacgagctgtacaaagctagcggcagcggcGAGGAATTTTCTCTC**TGA**GACAGAGCAAGAGAGG GAGACCCTGATGACTTCTTCGCCCCATGTCAGCAGGAAGTAGTTACAGAAGACCCACGAC GTCCTTACAACCAGAGCTTTTCAGGGTCTCCATCTCTTGAGGAGGGAAATATTAGGGTAcac ggcggttctatgggaatgtctggggatccaccggtcgccaccatggtgagcaagg
VGLL2	ctcggcatggacgagctgtacaaagctagcggcagcggcGCATCCCTCCTGAGC**TGA**TCTGCTGACCCAGGGTT TCCCCTTCCCCTTCCCTTCTGGCCAGCCTTGGAGGCTCAGCATCTGTGCCTCTCACTCCGTG GATTAGGACATGGGGGAAGGCAGAGACTTCAGACTTCTCAGTGTGTTGGGAAAACCcacggc ggttctatgggaatgtctggggatccaccggtcgccaccatggtgagcaagg
PTDSS2	ctcggcatggacgagctgtacaaagctagcggcagcggcGCACCAACTCCAAAC**TGA**CCTGGGCCGTGGCTGCC TCGTGAGCCTCCCAGAGCCCAGGCCTCCGTGGCCTCCTCCTGTGTGAGTCCCACCAGGAGC CACGTGCCCGGCCTTGCCCTCAAGGTTTTTTGCTTTTCTCCTGTGCACCTGGCGAGGcacggcg gttctatgggaatgtctggggatccaccggtcgccaccatggtgagcaagg
IL18	ctcggcatggacgagctgtacaaagctagcggcagcggcGTTCAAAACGAAGAC**TAG**CTATTAAAATTTCATGC CGGGCGCAGTGGCTCACGCCTGTTATCCCAGCCCTTTGGGAGGCTGAGGCGGGCAGATCA CCAGAGGTCAGGTGTTCAAGACCAGCCTGACCAACATGGTGAAACCTCATCTCTACTAcacg gcggttctatgggaatgtctggggatccaccggtcgccaccatggtgagcaagg
PCSK9	ctcggcatggacgagctgtacaaagctagcggcagcggcTCCCAGGAGCTCCAG**TGA**CAGCCCCATCCCAGGAT GGGTGTCTGGGGAGGGTCAAGGGCTGGGGCTGAGCTTAAAAATGGTTCCGACTTGTCCCT CTCTCAGCCCTCCATGGCCTGGCACGAGGGGATGGGGATGCTTCCGCCTTTCCGGGGCcacg gcggttctatgggaatgtctggggatccaccggtcgccaccatggtgagcaagg
EDN1	ctcggcatggacgagctgtacaaagctagcggcagcggcgactacaaggacgacgacgacaagCGAGCACATTGG**TGA**CAGA CCTTCGGGGCCTGTCTGAAGCCAggcagcggccatcatcaccatcaccacggcggttctatgggaatgtctggggatccaccg

	gtcgccaccatggtgagcaagg
CDKN2B	ctcggcatggacgagctgtacaaagctagcggcagcggcgactacaaggacgacgacgacaagACAGCCACGGGGGACTGACGCCAGGTTCCCCAGCCGCCCACAACGACTTTATTTTCTTACCCAATTTCCCACCCCCACCCACCTAATTCGATGAAGGCTGCCAACGGGGAGCGGCGGAAAGCCTGTAAGCCTGCAAGCCTGTCTGAGACTCACggcagcggccatcatcaccatcaccacggcggttctatgggaatgtctggggatccaccggtcgccaccatggtgagcaagg
KCNF1	tgctgctggtattacccatggtatggatgaattgtacaaagctagcggcagcggcCTCCAGAGTTGCAAGTGACAGGAGGGCCCCTCAGGCAGAGATGGACCAGGCGGTGGACAGATGGGTAGATGTGGCAGGCATGTCATCGACAGCACAGAAGGGCTGTCCTGTGTCCCCCCAACCCTCCCCTGGACAGACTCTGAAGGCCCTCCCcacggcggttctatgggaatgtctggatctggcgacgtcatggtttcaaaaggtgaagaagataatatg
MYNN	tgctgctggtattacccatggtatggatgaattgtacaaagctagcggcagcggcTTACAACAATTATACTGACTTTGTAAGGAATATGGAATTGCTAAGATATCATTGGTAGCAAACATCTCTGGTAAGGTGCATATATTCATATTAAATTCCCATTCATTTGAGTTGcacggcggttctatgggaatgtctggatctggcgacgtcatggtttcaaaaggtgaagaagataatatg
NKX2-4	tgctgctggtattacccatggtatggatgaattgtacaaagctagcggcagcggcTATGGCAGGACGTGGTGACAGCGAGGGCGCCCCGGGGCTAGGTCCTGGTGCACCCGAAGGGTCTGCAAGAAACTGCTAGAACGGATGGGGGAGGCATGCGAAAACCGACTCGCGCGTGTAAACTGCGATTCAAAACGAACCGGTCTCAGAGAGcacggcggttctatgggaatgtctggatctggcgacgtcatggtttcaaaaggtgaagaagataatatg
KIF3C	tgctgctggtattacccatggtatggatgaattgtacaaagctagcggcagcggcGTGGCGGACCATGAGTGACAACCATCACGTCAGGCTGCCCATCCAATAGACTCCTGGGATGGGGCAGCCAACCCTGGCTCATCTCATCTGCCGCTTGGTGCGTGTGCGTGTGCGTGCATGTGCGTGTGCGTGTGTGCAGGGGTGAGAATCTGGcacggcggttctatgggaatgtctggatctggcgacgtcatggtttcaaaaggtgaagaagataatatg
PRR20A	tgctgctggtattacccatggtatggatgaattgtacaaagctagcggcagcggcACCGACATTGCGTATTGACCACTATCTGCCACCCACGTTGTTCCCAGCCTCCCTTCCTTCCACCTGGACGTTCCCCCCAGCCCCACTTCTGCTCCACTCCTCCCCCGACTGGACCTGAAGCCTGAGCTTCCCCTGAACTTGGAGTACGCAACTcacggcggttctatgggaatgtctggatctggcgacgtcatggtttcaaaaggtgaagaagataatatg
FGD5	tgctgctggtattacccatggtatggatgaattgtacaaagctagcggcagcggcGATGCGAGTGTGTTATAGCAGTTATCAAAGCATGTGGACTTGTAACAAATTCTTAGGTCAATATGTGAATGCTTTTAGAAGCcacggcggttctatgggaatgtctggatctggcgacgtcatggtttcaaaaggtgaagaagataatatg
WHSC1	tgctgctggtattacccatggtatggatgaattgtacaaagctagcggcagcggcGTCACAGAGGGCAAATAGCGCCAGGCGGCCGCTTGGCCGGATCCAGGGGCGGTGCAGGGCGGCCGGCCCTGCCTGCGGGAGAGGGCGAGCATGAACTGGCCCGGAGGACCCAGCTCGAGCCGCCAGGACACAGACGTACAGGCCTCCTCGGGAcacggcggttctatgggaatgtctggatctggcgacgtcatggtttcaaaaggtgaagaagataatatg
SFTA2	tgctgctggtattacccatggtatggatgaattgtacaaagctagcggcagcggcGTTGTCTGCAACACATGACAGCCATTGAAGCCTGTGTCCTTCTTGGCCCGGGCTTTTGGGCCGGGGATGCAGGAGGCAGGCCCCGACCCTGTCTTTCAGCAGGCCCCCACCCTCCTGAGCGGCAATAAAcacggcggttctatgggaatgtctggatctggcgacgtcatggtttcaaaaggtgaagaagataatatg
LEPROTL1	tgctgctggtattacccatggtatggatgaattgtacaaagctagcggcagcggcCCCTCTCAGGCACAATGACAACTACTGCTCAGTGCCAGACACTGCACCATGcacggcggttctatgggaatgtctggatctggcgacgtcatggtttcaaaaggtgaagaagataatatg
AQP4	ctcggcatggacgagctgtacaaagctagcggcagcggcGAGGTATTGTCTTCAGTATGACTAGAAGATCGCACTGAAAGCAGACAAGACTCCTTAGAACTGTCCTCAGATTTCCTTCCACCCATTAAGGAAACAGATTTGTTAcacggcggttctatgggaatgtctggggatccaccggtcgccaccatggtgagcaagg
MAPK10	ctcggcatggacgagctgtacaaagctagcggcagcggcCTGGGTTGTTGCAGGTGACTAGCCGCCTGCCTGCGAAACCCAGCGTTCTTCAGGAGAcacggcggttctatgggaatgtctggggatccaccggtcgccaccatggtgagcaagg
OPRL1	ctcggcatggacgagctgtacaaagctagcggcagcggcGAGACGGTACCGCGGCCCGCATGACTAGGCGTGGACCTGCCCATGGTGCCTGTCAGCCCGCAGAGCCCATCTACGCCCAACACAGAGCTCACACAGGTCACTGCTCTCcacggcggttctatgggaatgtctggggatccaccggtcgccaccatggtgagcaagg

Lower case: linker and flanking homology; upper case: derived from gene; bold underlined: stop codon

Table S6. Dual-FP reporter sequences.

Construct	Sequence
Yeast dual-FP	ATGTCTAAAGGTGAAGAATTATTCACTGGTGTTGTCCCAATTTTGGTTGAATTAGATGGTG ATGTTAATGGTCACAAATTTTCTGTCTCCGGTGAAGGTGAAGGTGATGCTACTTACGGTAA ATTGACCTTAAAATTTATTTGTACTACTGGTAAATTGCCAGTTCCATGGCCAACCTTAGTC ACTACTTTCGGTTATGGTGTTCAATGTTTTGCTAGATACCCAGATCATATGAAACAACATG ACTTTTTCAAGTCTGCCATGCCAGAAGGTTATGTTCAAGAAAGAACTATTTTTTTCAAAGA TGACGGTAACTACAAGACCAGAGCTGAAGTCAAGTTTGAAGGTGATACCTTAGTTAATAG AATCGAATTAAAAGGTATTGATTTTAAAGAAGATGGTAACATTTTAGGTCACAAATTGGA ATACAACTATAACTCTCACAATGTTTACATCATGGCTGACAAACAAAAGAATGGTATCAA AGTTAACTTCAAAATTAGACACAACATTGAAGATGGTTCTGTTCAATTAGCTGACCATTAT CAACAAAATACTCCAATTGGTGATGGTCCAGTCTTGTTACCAGACAACCATTACTTATCCA CTCAATCTGCCTTATCCAAAGATCCAAACGAAAAGAGAGACCACATGGTCTTGTTAGAAT TTGTTACTGCTGCTGGTATTACCCATGGTATGGATGAATTGTACAAA*GCTAGC******GACGTC* ATGGTTTCAAAAGGTGAAGAAGATAATATGGCTATTATTAAAGAATTTATGAGATTTAAA GTTCATATGGAAGGTTCAGTTAATGGTCATGAATTTGAAATTGAAGGTGAAGGTGAAGGT AGACCATATGAAGGTACTCAAACTGCTAAATTGAAAGTTACTAAAGGTGGTCCATTACCA TTTGCTTGGGATATTTTGTCACCACAATTTATGTATGGTTCAAAAGCTTATGTTAAACATCC AGCTGATATTCCAGATTATTTAAAATTGTCATTTCCAGAAGGTTTTAAATGGGAAAGAGTT ATGAATTTTGAAGATGGTGGTGTTGTTACTGTTACTCAAGATTCATCATTACAAGATGGTG AATTTATTTATAAAGTTAAATTGAGAGGTACTAATTTTCCATCAGATGGTCCAGTTATGCA AAAAAAAACTATGGGTTGGGAAGCTTCATCAGAAAGAATGTATCCAGAAGATGGTGCTTT AAAAGGTGAAATTAAACAAAGATTGAAATTAAAAGATGGTGGTCATTATGATGCTGAAGT TAAAACTACTTATAAAGCTAAAAAACCAGTTCAATTACCAGGTGCTTATAATGTTAATATT AAATTGGATATTACTTCACATAATGAAGATTATACTATTGTTGAACAATATGAAAGAGCTG AAGGTAGACATTCAACTGGTGGTATGGATGAATTATAT**TAA**
Mammalian dual-FP	ATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGAC GGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTAC GGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACC CTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGC AGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTT CAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGG TGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACA AGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACG GCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCG ACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACT ACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCC TGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTACA AA*GCTAGC******GGATCC*ATGGTGAGCAAGGGCGAGGAGGATAACATGGCCATCATCAAGG AGTTCATGCGCTTCAAGGTGCACATGGAGGGCTCCGTGAACGGCCACGAGTTCGAGATCG AGGGCGAGGGCGAGGGCCGCCCCTACGAGGGCACCCAGACCGCCAAGCTGAAGGTGACC AAGGGTGGCCCCCTGCCCTTCGCCTGGGACATCCTGTCCCCTCAGTTCATGTACGGCTCCA AGGCCTACGTGAAGCACCCCGCCGACATCCCCGACTACTTGAAGCTGTCCTTCCCCGAGG GCTTCAAGTGGGAGCGCGTGATGAACTTCGAGGACGGCGGCGTGGTGACCGTGACCCAGG ACTCCTCCCTGCAGGACGGCGAGTTCATCTACAAGGTGAAGCTGCGCGGCACCAACTTCC CCTCCGACGGCCCCGTAATGCAGAAGAAGACCATGGGCTGGGAGGCCTCCTCCGAGCGGA TGTACCCCGAGGACGGCGCCCTGAAGGGCGAGATCAAGCAGAGGCTGAAGCTGAAGGAC GGCGGCCACTACGACGCTGAGGTCAAGACCACCTACAAGGCCAAGAAGCCCGTGCAGCTG CCCGGCGCCTACAACGTCAACATCAAGTTGGACATCACCTCCCACAACGAGGACTACACC ATCGTGGAACAGTACGAACGCGCCGAGGGCCGCCACTCCACCGGCGGCATGGACGAGCTG TACAAG**TAA**

italic: restriction sites; ****: RT insert; bold: stop codon, GFP encoded upstream of ****, mCherry encoded downstream.