HHS Public Access

Author manuscript

Cell Rep. Author manuscript; available in PMC 2024 July 19.

Published in final edited form as:

Cell Rep. 2024 April 23; 43(4): 114074. doi:10.1016/j.celrep.2024.114074.

UPF1 regulates mRNA stability by sensing poorly translated coding sequences

Damir Musaev[1], **Mario Abdelmessih**[1,2], **Charles E. Vejnar**[1], **Valeria Yartseva**[1,3], **Linnea A. Weiss**[1], **Ethan C. Strayer**[1], **Carter M. Takacs**[1,4], **Antonio J. Giraldez**[1,5,6,7,*]

[1]Department of Genetics, Yale University School of Medicine, New Haven, CT 06510, USA

[2]AstraZeneca, Waltham, MA 02451, USA

[3]Kenai Therapeutics, San Diego, CA, USA

[4]University of New Haven, West Haven, CT 06516, USA

[5]Yale Stem Cell Center, Yale University School of Medicine, New Haven, CT 06510, USA

[6]Yale Cancer Center, Yale University School of Medicine, New Haven, CT 06510, USA

[7]Lead contact

SUMMARY

Post-transcriptional mRNA regulation shapes gene expression, yet how *cis*-elements and mRNA translation interface to regulate mRNA stability is poorly understood. We find that the strength of translation initiation, upstream open reading frame (uORF) content, codon optimality, AU-rich elements, microRNA binding sites, and open reading frame (ORF) length function combinatorially to regulate mRNA stability. Machine-learning analysis identifies ORF length as the most important conserved feature regulating mRNA decay. We find that Upf1 binds poorly translated and untranslated ORFs, which are associated with a higher decay rate, including mRNAs with uORFs and those with exposed ORFs after stop codons. Our study emphasizes Upf1's converging role in surveilling mRNAs with exposed ORFs that are poorly translated, such as mRNAs with long ORFs, ORF-like 3′ UTRs, and mRNAs containing uORFs. We propose that Upf1 regulation of poorly/untranslated ORFs provides a unifying mechanism of surveillance in regulating mRNA stability and homeostasis in an exon-junction complex (EJC)-independent nonsense-mediated decay (NMD) pathway that we term ORF-mediated decay (OMD).

*Correspondence: antonio.giraldez@yale.edu.

AUTHOR CONTRIBUTIONS

D.M., M.A., and A.J.G. designed the project. D.M. performed the experiments and data analysis. M.A. provided R script for *cis*-element decay correlation and the machine-learning pipeline that D.M. adapted for further analysis. C.E.V. advised D.M. on computational methods and wrote a Python script to analyze 3′ UTR MPRA that D.M. adapted for further data analysis. V.Y. and C.M.T. provided 3′ UTR MPRA decay datasets. E.C.S. provided a script for k-mer analysis that D.M. adapted for further data analysis. D.M., L.A.W., and A.J.G. wrote the paper.

DECLARATION OF INTERESTS

D.M. and A.J.G. have a pending patent on mRNA regulatory *cis*-elements. A.J.G. is also a founder of RESA Therapeutics Inc., a company specializing in RNA technology for therapeutics.

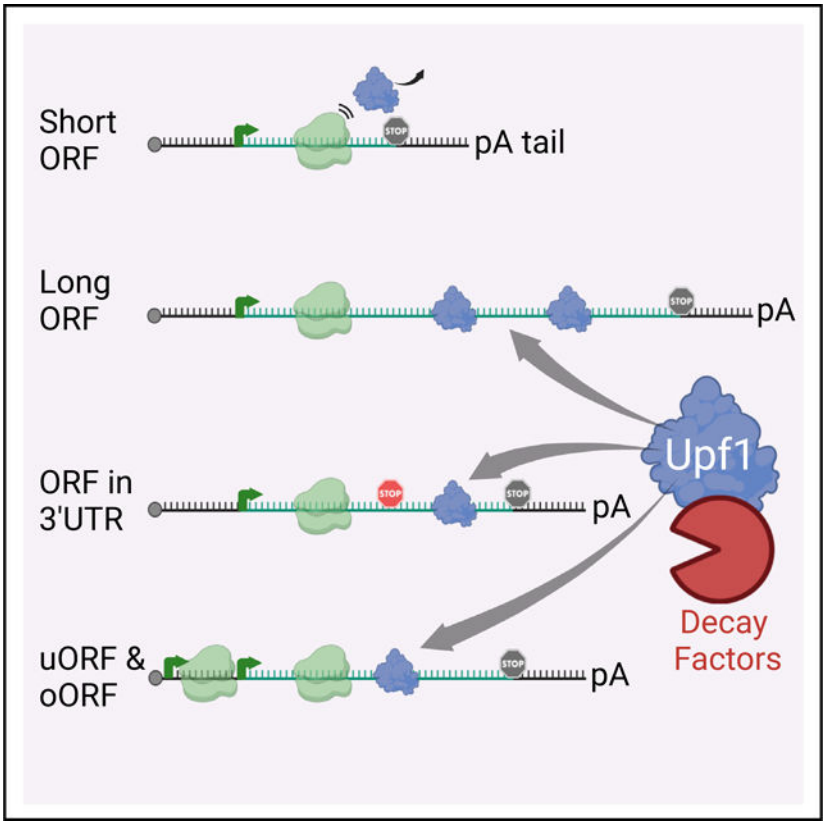
SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.celrep.2024.114074.

In brief

Musaev et al. explore conserved regulatory mechanisms of mRNA. They report that mRNAs with long ORFs undergo accelerated decay and reduced translation. Upf1 is a key factor in recognizing mRNAs with long ORFs, uORFs, and poorly translated mRNAs, and its phosphorylation is critical for these binding preferences.

Graphical Abstract



INTRODUCTION

Gene expression is vital for determining cell fate and homeostasis, and it is strongly influenced by post-transcriptional mRNA regulation. Regulatory *cis*-elements within mRNAs are recognized by *trans*-factors, such as microRNAs (miRNAs) and RNA-binding proteins (RBPs), to control their half-lives and translation.[1,2] mRNA regulation occurs for both stable systems, such as terminally differentiated cells, and dynamic systems undergoing cellular or developmental transitions, such as developing embryos. One example of a major cellular reprogramming event that occurs during embryonic development is the maternal to zygotic transition (MZT) that involves large-scale remodeling of the transcriptome.[3,4] Thousands of maternally provided mRNAs are post-transcriptionally targeted for clearance as zygotic gene expression initiates. The stability of these mRNAs can be influenced by *cis*-elements throughout their sequence. For example, translation initiation repressive elements, such as upstream open reading frames (uORFs) in the 5′ untranslated region (UTR) or

overlapping ORFs (oORFs) can lead to mRNA decay.[5–7] Another layer of control over mRNA stability is provided by codon optimality in the open reading frames (ORFs) in stable and developing systems, from yeast to vertebrates.[8–10] In the 3′ UTRs, members of a conserved family of miRNAs (miR-430, miR-427, and miR-309 in zebrafish,[11] *Xenopus*,[12] and *Drosophila*,[13] respectively) are key players in targeting maternal mRNAs for decay. Additionally, in zebrafish 3′ UTR, massive parallel reporter assays (MPRA) have identified regulatory motifs,[14–16] but how these different motifs work together to shape mRNA stability is poorly understood. While many *cis*-elements in 5′ UTRs, ORFs, and 3′ UTRs have been associated with mRNA stability, a comprehensive understanding of their relative contributions and mechanisms of action across eukaryotes remains elusive.[4,17]

The regulation of mRNA decay is controlled by RBPs and their interplay with the translation and decay machineries is a sequence-specific process. A fundamental sequence characteristic that differs among 5′ UTRs, ORFs, and 3′ UTRs is their GC richness. In vertebrates such as *Homo sapiens*, *Mus musculus*, *Rattus norvegicus*, and *Gallus gallus*, 5′ UTRs are the most GC rich (~60%), and GC content progressively decreases around 2-fold from 5′ UTR to ORF to 3′ UTR.[18] This conserved sequence bias might be the result of selective pressure by various RBPs reminiscent of conserved mechanisms of gene regulation. In fact, increasing GC content in the 3′ UTR can increase mRNA decay, but the underlying mechanism is still poorly understood.[15,19]

In this study, we investigated *cis*-elements that affect mRNA stability in zebrafish and HEK293 cells, using both non-steady (maternal mRNAs in zebrafish) and steady-state (HEK293 cells) systems to compare our findings across distantly related vertebrates. Using a random-forest machine-learning approach, we identified several mRNA *cis*-elements, including the Kozak sequence, uORFs, miRNAs, AU-rich elements, codon optimality, and ORF length, as important predictors of mRNA stability. We focused on understanding the mechanism of ORF length-directed decay, which led us to identify Upf1 as a common protein factor regulating three convergent decay mechanisms driven by mRNAs with uORFs, low translation, and long ORFs in a pathway that we termed ORF-mediated decay (OMD).

RESULTS

Exploring mRNA stability features

We set out to identify potential *cis*-elements that regulate mRNA stability across 5′ UTRs, ORFs, and 3′ UTRs. We hypothesized that the use of two complementary systems, the developing zebrafish embryo and human HEK293 cells, would allow us to identify regulatory elements across non-steady and steady-state systems that are conserved in vertebrates. In zebrafish embryos, we conducted a time-course sequencing analysis of poly(A)-selected mRNAs during the first 6 h post fertilization (hpf). We analyzed the degradation of 3,041 maternally deposited mRNAs that exhibit transcriptional quiescence,[20,21] thereby mitigating the confounding influence of newly synthesized mRNA. In HEK293 cells, we calculated the half-lives of 4,206 mRNAs using TimeLapse-seq, which relies on pulse metabolic labeling of mRNAs with 4-thiouridine.[22,23] We noted a spectrum of mRNA stability, prompting us to investigate whether distinct *cis*-elements could be linked

to these diverse decay behaviors. We therefore categorized the mRNAs according to their onset of decay during zebrafish development (Figures 1A and 1B) or quartiles of half-life in HEK293 cells using supervised cluster analysis and searched for *cis*-elements that were significantly enriched in each group. Out of 86,213 features, which included uORFs, 5′ UTR GC content, codons, amino acids, 3′ UTR GC content, and 6- to 8-nt k-mers in 3′ UTRs (Figures 1A; Table S1), we identified a total of 560 and 3,754 *cis*-elements that were enriched for specific patterns of decay in zebrafish and HEK293, respectively (see STAR Methods). To determine whether these *cis*-elements significantly correlated with either mRNA stability or decay, we compared the level of decay between those mRNAs enriched (>80th %) or depleted (<20th %) for each individual feature within each species. This analysis identified 201 *cis*-elements in zebrafish and 160 in HEK293 cells that correlate with mRNA stability or decay (Figures 1C and 1D; Table S2 and S3). Additionally, in zebrafish we identified the developmental time window during which each *cis*-element was significantly associated with differential mRNA stability (Figure 1C). Taken together, our analysis identified *cis*-regulatory elements that are significantly associated with differential mRNA decay in two divergent vertebrate species.

Exploring *cis*-element correlation with mRNA stability

Our analysis identified both conserved and non-conserved features regulating mRNA stability in zebrafish embryos and HEK293 cells (Figures 1C and 1D). Among the conserved elements, a high Kozak score in the 5′ UTR was significantly associated with mRNA stability (zebrafish $p = 7.2e{-}10$, HEK293 $p = 1.1e{-}15$, unpaired two-sample Wilcoxon test), while the number of uORFs, high uORF-Kozak score, and long uORFs correlated with mRNA decay (Figures 1C; Tables S2 and S3). Notably, in zebrafish, this effect was most significant at the 2- to 6-hpf developmental interval, coinciding with the start of zygotic genome activation. We found that enrichment for certain codons, such as AAG (Lys), correlated with mRNA stabilization, while enrichment for codons such as TCT (Ser) and CAG (Gln) correlated with mRNA decay. Our findings regarding codon identity in zebrafish are consistent with previous studies from several laboratories, including ours,[9,10,17] and extend these findings by identifying the developmental time points when these correlations are most significant. In addition, enrichment in proline repeats and ORF length correlated with mRNA decay in both systems. These results highlight the conserved roles of numerous *cis*-elements within 5′ UTRs and ORFs across zebrafish and HEK293 cells. Several AU-rich motifs resembling the AU-rich element UAUUUAU[25] significantly correlated with mRNA decay in both HEK293 (AUUUAU, $p = 1.12e{-}39$), and zebrafish (UAUUUAU, $p = 1.23e{-}10$) (Figures 1C and 1D). On the other hand, some features correlated with opposite functions in mRNA stability between the two species. For example, enrichment in valine repeats or codons such as ATC (Ile) correlated with mRNA stability in HEK293 cells but mRNA decay in zebrafish. Other features appeared to be specific for a given system (Figure 1C). The element that most significantly correlated with decay in the 3′ UTR in zebrafish was the presence of the miR-430 motif (Figure 1C), which clears many maternally provided mRNAs after zygotic genome activation[11] (which occurs at 2 hpf). Together, this analysis highlights both unique and conserved regulatory features that mediate mRNA stability in zebrafish embryonic development and steady-state HEK293 cells.

Next, we asked whether these *cis*-elements function combinatorially to regulate mRNA stability. We compared the stability of gene groups enriched in one or multiple *cis*-elements, including codon optimality, ORF length, uORFs, the miR-430 seed, and AU-rich elements, all of which strongly correlate with stability. Notably, we observed that mRNAs enriched in optimal codons (top 50%) with short ORFs (top 50%), and lacking uORFs, miR-430 (zebrafish), or an AU-rich element (HEK293), were the most stable. Conversely, we found that mRNAs enriched in non-optimal codons (top 50%) with long ORFs (top 50%) and containing uORFs, miR-430 seeds (zebrafish), or AU-rich elements (HEK293) were the least stable mRNAs (Figures 1E and 1F, $p = 4.5e{-}25$ and 2.9e–22, respectively; unpaired two-sample Wilcoxon test). A stepwise introduction of these mRNA decay elements correlated with decreased stability of mRNAs, and the corollary was observed for stabilizing elements (Figures 1E and 1F). Taken together, our findings demonstrate the combinatorial effect of *cis*-elements on mRNA decay and stability and highlight the conserved function of these elements across species.

Machine learning highlights the significance of *cis*-elements

Modeling can allow us to predict the relative contributions of *cis*-elements to mRNA regulation. We therefore constructed a random-forest machine-learning model to predict mRNA decay in order to determine which *cis*-elements contribute most significantly to mRNA stability. The model was trained on datasets comprising the 201 zebrafish and 160 HEK293 cell *cis*-elements that are significantly associated with mRNA stability or decay (Figure 1A). To validate the efficacy of our model, we performed 5-fold cross-validation on randomly selected training data and found that predicted and experimental mRNA decay values correlated with R values of 0.44 and 0.56 for zebrafish and HEK293 cells, respectively (Figures 2A and 2B, Pearson correlation coefficient), values that are consistent with other systems.[26] While the machine-learning model allows us to compare the relative contributions of *cis*-elements to mRNA fate, it does not specify their correlation with mRNA decay or stability. We therefore compared the average feature importance generated by machine-learning analysis with the difference in stabilities of genes enriched in and depleted of a given *cis*-element, thus allowing us to identify the features that contribute most significantly to mRNA stability or decay (Figures 2C and 2D). Notably, in HEK293 cells, serine and histidine frequencies were the most effective predictors of mRNA decay, while in zebrafish it was the 3′ UTR motif GCACTT, the binding site for miR-430, and was specific to zebrafish (Figures 2C and 2D).

Interestingly, many predictive *cis*-elements were conserved. We identified 52 predictive *cis*-elements that were shared between zebrafish and HEK293 cells (Figure 2E). Furthermore, their corresponding feature importance values exhibited a significant correlation ($R = 0.77$, $p = 2.7e{-}11$). Elements controlling mRNA translation initiation, such as uORF-Kozak and Kozak score of the main ORF, play a critical role in mRNA decay and stability, respectively, thus highlighting the conserved role of translation in mRNA decay. Strikingly, ORF length emerged as the most important conserved feature for predicting mRNA stability across both systems, yet the underlying mechanism is unknown.[27–30] Altogether, our machine-learning analysis identified the regulatory elements that are most relevant in the regulation of mRNA stability in steady and non-steady systems.

ORF length is associated with mRNA decay

Using our model, we identified ORF length as a conserved and principal predictor of mRNA decay. To validate this prediction, we first analyzed the effect of ORF length on mRNA decay by comparing the decay rates of mRNAs with long (>80th %) versus short (<20th %) ORFs using the zebrafish strictly maternal mRNA dataset. In agreement with our predictive model, we found that mRNAs with longer ORFs exhibited significantly higher levels of decay than those with shorter ORFs (Figure 3A, Kolmogorov-Smirnov test (KS) $p = 9.5e–19$), even when stable short ribosomal proteins were excluded (Figure S1D, KS $p = 5e–18$).

The individual contribution of ORF length to mRNA decay can be confounded by other factors present on endogenous transcripts, such as codon optimality or the length of the 3′ UTRs (Figure S2C).[9,17] For example, we found that, in zebrafish, increasing ORF length correlates with 3′ UTR length (Figure S1C, Pearson 0.14; $p = 8e–12$; and Figure S1E, $p = 8.7e–6$), and, therefore, 3′ UTR length is also correlated with mRNA decay (Figures S1B, $p = 2e–4$). Regardless, ORF length is a driving factor of mRNA stability (Table S4, partial correlation of ORF length given 3′ UTR length, −0.184, $p = 1.1e–20$; 3′ UTR length given ORF length, −0.074, $p = 0.0002$). Based on these observations, we next set out to directly test the effect of ORF length on mRNA decay within a reporter system. We designed a short ORF reporter of 300 bp in length, while the long reporter consisted of this short sequence repeated six times in tandem, resulting in an ORF length of 1,800 bp. Both reporters had common 5′ and 3′ UTRs where the 3′ UTR lacked known decay motifs, including miR-430 target sequences, AU-rich elements, and C-rich decay motifs.[16] Injection of these reporters into one-cell-stage zebrafish embryos followed by northern blot analysis revealed that mRNA reporters with longer ORFs decayed faster than those with shorter ones. This effect was observed regardless of whether the ORF was enriched in optimal (codon adaptation Index [CAI] = 0.91) or non-optimal codons (CAI = 0.6), with a stronger decay observed for non-optimal ORFs (Figures 3B and 3C). Taken together, our machine-learning model and experimental data both indicate that increasing ORF length reduces mRNA stability.

Translation efficiency is ORF-length dependent and affects stability

We wondered whether the increased rate of mRNA decay for long ORFs could be due to the interplay between ribosomes and decay factors recruited to the mRNAs. We can envision two contrasting roles for ribosomes. In one case, ribosomes might protect mRNAs or promote mRNA stability by displacing decay factors during translation. If this is the case, we would expect ribosome occupancy and therefore translation efficiency to be lower in longer ORFs. Alternatively, ribosomes might recruit more decay factors to long ORFs, which can accommodate more ribosomes, resulting in a higher decay rate.

To distinguish between these models, we compared translation efficiencies of maternal mRNAs with long (>80th %) versus short (<20%) ORFs using publicly available ribosome profiling data[7] at 5 hpf (Figure 3D). mRNAs with long ORFs had significantly lower ribosome density or translation efficiency (KS test; $p = 9.1e–18$), supporting the model where ribosomes promote mRNA stability by protecting the mRNA or displacing decay factors. Next, we directly compared the translation of two reporter mRNAs with either

short (1×) or long (4×) ORFs followed by a self-cleaving 2A peptide and nanoluciferase sequence.[31] Co-injection of each reporter along with a firefly-luciferase control into one-cell-stage zebrafish embryos revealed that the mRNA with a short ORF had ~2-fold higher translational output than the long ORF after normalizing to mRNA levels (Figure 3E). This finding again supports our hypothesis that ribosomes protect or displace decay factors and do so less efficiently for long ORFs. In addition, the long ORF nanoluciferase reporter decayed faster than the short one based on qPCR analysis (Figure S2A). Together, these results indicate that mRNAs with long ORFs not only decay faster but also have lower protein output compared to ones with short ORFs.

Next, to test whether translation itself affects mRNA decay, we designed reporter mRNAs without start codons, hence called pseudo-ORFs, but with two different lengths (600 bp vs. 3 × 600 bp = 1,800bp). Neither had any alternative start codons such as AUG or CUG (weak start codon) in any reading frame. Upon injection of these mRNAs into one-cell-stage zebrafish embryos, we observed that both reporters decayed at a similar rate independent of their length (Figure 3F). However, upon injection of analogous ORF reporters (CAI = 0.87) in which a start codon was introduced, we found that the long ORF experienced faster decay compared to the short ORF (Figure 3G). These results suggest that mRNA protection could be achieved by a higher ribosome density or by effectively displacing decay factors. They also suggest that ribosomes are not mediating mRNA decay by introducing decay factors and argue against the second model. Taken together, our results indicate that, when ribosomes are scanning but not translating the mRNA, ORF length does not affect mRNA stability.

To further test the ribosome protection hypothesis, we manipulated mRNA translation initiation by modifying the Kozak sequence of our ORF reporters. Kozak sequences were based on zebrafish translation assays from Grzegorski et al.[32] As predicted, when both long and short ORF reporters shared the same strong Kozak sequence, the long ORF reporter was less stable than the short one (Figure 3H). However, when we weakened the Kozak sequence strength of the short ORF reporter, it decayed faster than the reporter with the long ORF and strong Kozak sequence (Figures 3I and 3J). Our results suggest that weakening the Kozak sequence slowed down translation initiation (Figure S2D), thereby reducing ribosome footprints and its protective influence.

In summary, we observe that mRNAs with long ORFs (Figures 3B, 3C, 3G, and 3H) and lower translation (Figures 3F, 3I, and 3J) correlate with higher rates of decay, and long ORFs have lower protein output (Figure 3E). Taken together, we hypothesize that unknown decay factors can bind more readily to poorly translated ORFs with fewer ribosomes protecting them, and ribosomes can displace these factors more efficiently from short ORFs.

Surveillance, deadenylation, and decapping machineries “sense” ORF length

mRNA ORF length correlates with decay in human cells and zebrafish, but this effect has also been observed in other eukaryotic systems, such as budding yeast.[27,28] Because low codon optimality exacerbated the effect of ORF length in mRNA decay in zebrafish (Figures 3B and 3C), we first assessed the role of codon optimality in mRNA decay in wild-type (WT) yeast for long and short ORFs (datasets[33,34]). We compared the decay rates of long vs.

short ORFs for mRNAs with high (>75th%) and low (<25th%) gene codon optimality and found that the effect of ORF length in mRNA decay was the most significant for genes with low codon optimality (Figure 4B, $p = 1.51e–5$; Figure S4, low codon optimality ORF length distribution). We therefore used this gene set for subsequent analysis. To identify factors responsible for regulating the stability of mRNAs with long ORFs, we analyzed mRNA decay rates in a yeast knockout (KO) screen.[33,35] We reasoned that the loss of function of factors that participate in the decay of long ORFs would lead to stabilization of these mRNAs, thus reducing the difference in the rate of decay between short- and long-ORF mRNAs and allowing us to identify genes involved in this pathway (Figure 4A). To this end we compared the decay rates of short- and long-ORF mRNAs in 47 *Saccharomyces cerevisiae* BY4741 KO strains using publicly available datasets.[33,35] The knocked-out genes comprised factors involved in RNA regulation, including mRNA deadenylation, decapping, exonucleases, endonuclease, exosome-associated factors, transcription termination, nuclear RNA surveillance, splicing, RBPs, mRNA export, and translation surveillance. Unlike in the WT strain, mRNA decay rates were independent of ORF length when knocked-out genes were associated with pathways for mRNA deadenylation (*CCR4*, *POP2*, *NOT3*), 5′ to 3′ decay (*XRN1*), decapping (*LSM1*, *EDC3*, *DHH1*, *PAT1*), and surveillance (*UPF1*, *UPF2*, *UPF3*) (Figures 4C–4G). However, in the remaining KO strains from Sun et al.,33 the difference in the decay rates between short- and long-ORF mRNAs remained significant (Figure S3). When deadenylation factors *PAN2* or *PAN3* were knocked out, the differences in decay rates were amplified (Figure 4E). Because mRNA deadenylation is carried out both by PAN2-PAN3 (initial rapid phase) and CCR4-NOT complex (bulk deadenylation),[36] CCR4-NOT activity is likely increased in the absence of PAN2 or PAN3. In agreement with Ccr4p and Pop2p acting as catalytic subunits of CCR4-NOT complex,[36] we observed a significant reduction in the overall decay rate of all mRNAs in respective mutants (Figure 4E). Notably, in *upf1Δ*, *upf2Δ*, *upf3Δ*, *not3Δ*, *lsm1Δ*, *edc3Δ*, and *dhh1Δ* strains, the decay rate of only long-ORF mRNAs was significantly decreased, suggesting target specificity. Taken together, we identify mRNA deadenylation, decapping, surveillance, and 5′ to 3′ exonuclease factors that are implicated in ORF-length-dependent mRNA decay (Figure 4H).

mRNA decay factors interact among each other and Upf1 may act as a core factor involved in the very first step in recognizing poorly translated long ORFs. In yeast, nonsense-mediated decay (NMD) is carried out by Upf1p, Upf2p, and Upf3p, which target mRNAs with premature translation termination for decay.[37,38] Previous reports also found that Upf1 binds and recruits decay factors such as Upf2p, Upf3p, Dcp2p, Edc3p, Dhh1p, Lsm1p, Ccr4, Pop2, Not3p, and Not1p.[39–47] Thus, we hypothesize a function for Upf1 protein beyond NMD, where the regulation of long-ORF decay could be initiated by Upf1 protein binding to mRNA and leading to the recruitment of deadenylation and the decapping factors. We propose that Upf1 functions in an exon-junction complex (EJC)-independent NMD pathway, which we term OMD.

Upf1 targets long ORFs, uORF mRNA, and ORF-like 3′ UTRs to promote decay

To test whether Upf proteins preferentially bind long vs. short ORFs, we performed cross-linking and immunoprecipitation (CLIP) with ORF length reporters (from Figure 3G) using N-terminal 3xFLAG-tagged zebrafish Upf1, Upf2, and Upf3a, as well as Celf1 as an RBP

control. Notably, only Upf1 had a significant binding preference for the long-ORF reporter (Figure 6A), while Upf2, Upf3a, and Celf1 had similar affinities for both long- and short ORF-reporters. We have previously shown that long ORFs have lower translation efficiency, and hypothesized that ribosomes might fail to displace a decay factor recruited to poorly translated ORFs. We therefore reasoned that Upf1 also binds poorly translated ORFs such as those translationally repressed by uORFs and oORFs.[7] To test this, we co-injected the N-terminal 3xFLAG-tagged zebrafish Upf1, Upf2, and Celf1 mRNAs with reporters containing uORFs or an oORF into one-cell-stage zebrafish embryos. We observed increased binding affinity of tagged Upf1 for reporters with three uORFs and an oORF, while Upf2 and Celf1 proteins did not exhibit binding specificity for these reporters (Figure 6B). Additionally, Upf1 bound a low-translation-initiation reporter containing a weak Kozak sequence ~3× more efficiently than one with a strong Kozak sequence (Figure S2B). Together, these experiments suggest that Upf1 binds to poorly translated long ORFs and ribosome-depleted mRNAs with uORFs or weak Kozak sequences.

Next, we investigated whether Upf1 preferentially affects gene expression of mRNAs with long ORFs and uORFs. To this end, we knocked down (KD) Upf1 using small interfering RNA (siRNA) in HEK293 cells (Figures 5A and 5B, $n = 2$) and compared gene expression to control RBP - G3bp1 siRNA KD. mRNAs with long ORFs (>90th %) were significantly stabilized (Figure 5C, $p = 5.2e{-}8$) in Upf1 KD cells, but not in control G3bp1 KD. Long ORFs were stabilized independent of 3′ UTR length or the presence of uORFs/oORFs (Figures S5A–S5G). mRNAs with uORFs were similarly stabilized in Upf1 KD cells (Figure 5D, $p = 7.7e{-}7$), highlighting the conserved role of Upf1 in regulating mRNA decay of poorly translated mRNAs.

We hypothesize that Upf1 promotes decay of mRNAs with long ORFs and poorly translated coding sequences because of inefficient ribosome displacement of Upf1 leading to OMD. If this model is correct, then we reasoned that untranslated coding sequence inserted into the 3′ UTR should also elicit increased decay and binding by Upf1, since these ORFs would not be translated. To test this hypothesis, we first developed a 3′ UTR MPRA using random fragments derived from the ORFs and 3′ UTRs of 350 zebrafish genes that were inserted downstream of GFP ORF.[14] These mRNAs were *in vitro* transcribed, capped, and had a polyadenylation signal, but they lack introns and are not subject to canonical EJC-enhanced NMD.[48] To investigate how these sequences influence mRNA decay, we injected this mRNA library into one-cell-stage zebrafish embryos and measured decay between 2 and 6 hpf. We found that mRNAs with untranslated sequences derived from ORFs exhibited a higher decay rate than those derived from 3′ UTRs (Figure 6C), suggesting that poorly or untranslated coding sequences are targeted for preferential decay.

Next, to investigate whether untranslated ORF sequence-directed mRNA decay correlates with Upf1 binding, we used N-terminal 3xFLAG-tagged zebrafish Upf1, Upf2, Upf3a, and Upf3b, as well as GFP as a control, to pull down bound fragments from the MPRA library. We observed that 3xFLAG-Upf1-bound regions were enriched in ORF fragments compared to input (Figure 6D), while Upf2, Upf3a, and Upf3b, as well as GFP did not exhibit preferential binding to ORF or 3′ UTR-derived fragments (Figures 6E and S6A–S6F). Thus, the untranslated ORF-derived sequences are preferentially bound by Upf1 and experience

faster mRNA decay. Taken together, these results support a role for Upf1 in OMD as a sensor of poorly translated or untranslated ORF-like sequences that leads to mRNA decay.

Upf1 phosphorylation is critical to bind GGAG motifs

Next, we asked whether Upf1 binds specific sequence motifs to mediate targeted decay. First, to investigate motifs leading to mRNA decay, we performed k-mer enrichment analysis across regions that undergo preferential decay using the MPRA 3′ UTR library (log2FC(6/2 hpf) < −0.3). Because the library is composed of randomly fragmented sequences, we can compare the sequence coverage at each nucleotide between the two time points providing near-nucleotide resolution (see STAR Methods). We identified k-mers that were significantly enriched in regions that undergo decay, including the miR-430 target seed (GCACTT, $p = 1.2e{-}41$), C-rich motif[16] (CCTCC, $p = 2.4e{-}18$), and GGAGC/GAGCA motifs ($p = 9.7e{-}7$ and $8e{-}6$, respectively) (Figure 6F, Fisher's exact test after Bonferroni correction). Next, to identify k-mers preferentially bound by Upf1, we identified a set of sequences enriched in the Upf1 pull-down compared to input (log2FC(pull-down/input) > 0.5). Among these Upf1-bound G-rich k-mers, we observed a shared GGAG motif that was also identified in the mRNA decay k-mer analysis (Figure 6G). Interestingly, Upf1 k-mers (AGGAG and GGACG) inserted into the 3′ UTRs of reporters significantly increased mRNA decay, further supporting the destabilizing effect of these k-mers (Figure S6J). Together, these findings suggest that GGAG motif-directed mRNA decay might function via Upf1.

Upf1 is an RNA helicase that can get phosphorylated to repress translation initiation[49] and recruits other factors to initiate NMD.[50] To understand how Upf1 achieves its binding specificity, we mutated the 3xFLAG-Upf1 construct and then repeated the CLIP assay on the MPRA library. We compared three mutants: (1) 3xFLAG-Upf1-G495R/G497E, which lacks ATPase and helicase activities but retains the ability to bind RNA and can be hyperphosphorylated; (2) 3xFLAG-Upf1-R843C, which lacks helicase activity and cannot be phosphorylated; and (3) 3xFLAG-Upf1-C126S, which neither binds to Upf2 nor gets phosphorylated.[50] Similar to WT Upf1, k-mer enrichment analysis revealed that the Upf1-G495R/G497E mutant binds GGAG motifs (Figures 6I and S6H). Thus, Upf1's ATPase and helicase activities are not critical for its binding specificity. Strikingly, both Upf1-R843C and Upf1-C126S mutants lost their binding specificity to RNA, thus implicating an essential role for phosphorylation in the specificity of Upf1 (Figures 6J, 6K, S6G, and S6I). To further test the correlation between mRNA decay and Upf1's binding specificity, we centered our CLIP and decay datasets around the miR-430 (GCACUU), Upf1 (GGAGNG), and stabilizing poly(U) (UUUNUUU) motifs. Sequences preferentially bound to WT and mutant Upf1 (G495R/G497E) were associated with mRNA decay (Figure 6M). Neither protein had binding preference to the miR-430 decay motif (Figure 6L), and they were depleted from the poly(U) stabilizing motifs (Figure 6N). Thus, Upf1's phosphorylation sites are necessary for its binding specificity to GGAG motifs. Notably, the average GC richness surrounding Upf1-bound decay regions was ~45%, compared to ~34% for poly(U) stability regions (Figure 6O), and more generally the GC content in our MPRA library was higher for fragments derived from the coding sequences (51.4%) than 3′ UTRs (36.9%). This could explain the stronger Upf1 binding we observed for fragments derived from zebrafish

ORFs (Figure 6D). In fact, we and others have shown that ORFs have evolved across divergent species (*Drosophila*, zebrafish, medaka, *Xenopus*, and humans) to have higher GC content compared to their 3′ UTRs (Figures 7A and 7B).[18] Additionally, ORFs with higher GC content have relatively high-GC content 3′ UTRs and are enriched in genes involved in animal development and cell differentiation (Figures 7C–7E). Taken together, our data suggest that OMD via Upf1 provides a common mechanism to regulate ribosome-depleted long ORFs, mRNAs with untranslated GC-rich regions (e.g., in their 3′ UTR or after a premature stop codon), and mRNAs with low translation efficiency due to the presence of uORFs/oORFs or a weak Kozak.

DISCUSSION

Post-transcriptional gene regulation plays a key role in shaping gene expression in vertebrates. Here, we (1) define conserved *cis*-regulatory elements associated with mRNA decay and stability across vertebrates, (2) demonstrate that long ORFs have reduced translation efficiency and induce enhanced mRNA decay, and (3) identify multiple factors involved in sensing ORF length-mediated decay, with Upf1 playing a key role in targeting poorly translated ORF sequences. Importantly, we find that Upf1 preferentially targets mRNAs encoding long ORFs, mRNAs containing uORFs, and inefficiently translated ORF-like sequences to mediate mRNA decay, in an EJC-independent NMD pathway that we term OMD.

First, we find that Kozak sequence, uORFs, miRNAs, AU-rich elements, codon optimality, and ORF length are important features that regulate mRNA stability *in vivo*. These results are consistent with previous reports.[5–11,17,25,27–30] Our results reveal that these *cis*-elements function combinatorially to regulate mRNA stability in both zebrafish and HEK293 cells (Figures 1E and 1F), highlighting the conserved function of these elements across species. While codons have been linked to influencing mRNA decay in vertebrates,[9,10,17] we find that their corresponding amino acids had very similar effects on mRNA decay rate, illustrating the difficulty of parsing the specific contributions of these interdependent mRNA features in an endogenous context. Further experiments will be necessary to parse the contributions of codons, amino acids, and related RNA structures to the regulation of mRNA stability. Overall, our analysis identified conserved *cis*-elements that are most significantly associated with mRNA stability in steady and non-steady systems in vertebrates.

Second, we identified ORF length as a conserved decay feature that is anti-correlated with mRNA stability. This revealed a convergent mechanism that regulates mRNA decay for mRNAs with long ORFs but also for mRNAs containing uORFs/oORFs and poorly translated GC-rich sequences that are bound by Upf1 through a GGAG[G/A] motif leading to OMD. Previous reports have also found a negative correlation between ORF length and endogenous mRNA stability in yeast[27,28] and human cell lines,[29,30] but the mechanism underlying these observations remained poorly understood. Similar to our findings in zebrafish, ORF length is negatively correlated with translation efficiency in WT yeast as measured by ribosome footprinting[51] (Pearson correlation $r = -0.5$), indicating a conserved mechanism of mRNA decay. The analysis of mRNA stability across different yeast KO strains has been a useful resource to search for factors that target long-ORF mRNAs. We

observed that mutants in multiple components of the mRNA decay machinery exhibited generally slower rates of mRNA decay relative to WT, including *ccr4Δ*, *pop2Δ*, *caf40Δ*, *xrn1Δ*, *rrp6Δ*, *rrp47Δ*, *pat1Δ*, *scd6Δ*, and *thp2Δ* (Figure S3). Because these factors affect mRNA stability for both long- and short-ORF-containing mRNAs, we were unable to discern whether these factors target mRNAs with long ORFs specifically for decay as all mRNAs were significantly stabilized in these mutants. In contrast, *upf1Δ*, *upf2Δ*, *upf3Δ*, *not3Δ*, *lsm1Δ*, *edc3Δ*, and *dhh1Δ* strains elicited specific effects, suggesting that OMD functions through mRNA surveillance, deadenylation, and decapping. Additionally, KD of another factor required for EJC-independent NMD, Smg6, resulted in a significant stabilization of mRNAs with long ORFs (Figure S5H), but the effect was not significant for uORF-containing sequences (Figure S5I). This might be caused by the partial KD of Smg6. We propose that OMD likely falls under the umbrella of EJC-independent NMD, wherein long ORFs or poorly translated ORFs lead to mRNA decay independent of the EJC complex proteins. Indeed, given that premature termination codons (PTCs) leave part of the ORFs untranslated, we propose that OMD could also function to increase decay for mRNAs with PTCs, especially those that are mono-exonic and lack EJCs. Among the mutants identified, Not3 is involved in nuclear RNA surveillance and mRNA export,[52] suggesting that it may tag mRNAs with long ORFs more efficiently in the nucleus and target them for degradation in the cytoplasm. Similarly, Upf1 has been observed to routinely shuttle between the nucleus and cytoplasm via a mechanism that requires its RNA helicase activity and could tag specific sequences for decay.[53] While our data indicate that poorly translated long ORFs are less stable than short ones due to the recruitment of Upf1, longer ORFs could also increase ribosome pausing if they contain non-optimal codons. This could further contribute to lower mRNA stability. We propose that Upf1 binds mRNAs at specific GC-rich sequences to promote mRNA decay, and is removed by the ribosome to promote mRNA stability.

Third, our results reveal that untranslated GC-rich sequences derived from ORFs enhance mRNA decay (i.e., when inserted into 3′ UTRs) in zebrafish. We propose that detection of these sequences by Upf1 serves as a surveillance pathway to regulate mRNA stability. This recognition would exert evolutionary pressure such that, to avoid mRNA decay, endogenous 3′ UTRs have evolved to be GC-poor and U-rich, thereby enhancing mRNA stability.[15,16,19] This model is consistent with previous studies that have reported the stabilization of mRNAs containing uORFs in the absence of Upf1[54] and preferential Upf1 binding to mRNAs with long 3′ UTRs[55] and GC-rich 3′ UTRs.[56] Our study uses MPRA-CLIP to reveal that Upf1 binds preferentially GC-rich regions (including GGAG[G/A]) in a process that requires Upf1 phosphorylation. Unfortunately, zebrafish with a homozygous mutation in Upf1 die at 8 days post fertilization,[57] likely due to the temporary rescue of the zygotic function by the maternally inherited mRNA and protein, thus preventing us from analyzing its role on maternal mRNA decay. In the future, the generation of germline Upf1 KOs or degron mutants will be required to understand the full spectrum of Upf1's role in embryonic development and mRNA decay. Furthermore, previous reports have demonstrated that Upf1 binds to ORFs during puromycin-induced translation inhibition,[58] supporting a model where Upf1 binds coding sequences that are poorly translated and where the ribosome ejects Upf1 from its binding sites in the coding sequence. Conversely, ribosomes can trigger mRNA decay when a ribosome collides with a slower leading ribosome, forming a di-ribosome

structure recognized by the ubiquitin ligase ZNF598.[59] We propose that Upf1 not only recognizes exon-junction proteins due to PTCs but also targets for decay ribosome-depleted GC-rich sequences. Based on this model, the location of a PTC and the longer ribosome-depleted GC-rich ORFs could cooperatively affect the rate of mRNA decay. Consistent with this model, previous studies have suggested that the position of the PTC influences the rate of decay, with higher decay for early-occurring PTCs that consequently leave a longer portion of the ORF untranslated.[60,61] In summary, we propose that Upf1 provides a unifying mechanism to surveil mRNAs with low translation efficiency, including long ORFs, mRNAs with uORFs/oORFs or a weak Kozak, and ribosome-depleted GC-rich sequences, to regulate their stability through OMD.

Limitations of the study

While our results demonstrate a significant effect of ORF length on mRNA decay and how Upf1 recognizes poorly translated mRNAs and long ORFs, several limitations need to be considered. First, while the Upf1 CLIP assay using massively parallel reporters provided key insight into its binding-sequence preferences, the sequence space was limited to ~350 randomly fragmented zebrafish ORFs and 3′ UTRs. Future work will need to assess both 5′ UTRs and several endogenous sequences, as well as a randomly generated synthetic library. Second, ORF length influences mRNA structure. Our screen of *cis*-regulatory elements has not considered RNA structure as a feature. In addition, in order to test the regulatory effect of a given mRNA structure on ORF-length-dependent decay, the mRNA sequence would need to be altered in a way that would introduce additional confounding regulatory variables, such as varying codon optimality, new binding motifs for RBPs, and others. Despite the aforementioned limitations, we identify key *cis*-elements that function to regulate mRNA stability and uncover the mechanism behind ORF-length-dependent decay.

STAR★METHODS

RESOURCE AVAILABILITY

Lead contact—Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Antonio J. Giraldez (antonio.giraldez@yale.edu).

Materials availability—All unique/stable reagents generated in this study are available from the lead contact without restriction.

Data and code availability

- High throughput sequencing data have been deposited at the Sequence Read Archive and are publicly available as of the date of publication. Accession numbers are listed in the key resources table. All data reported in this paper will be shared by the lead contact upon request.
- This paper does not report original code.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Zebrafish maintenance and embryo collection—Zebrafish (Danio rerio) embryos were acquired through natural mating of adult fish with mixed wild-type backgrounds, including TU-AB and TLF strains, aged between 5 and 18 months. The wild-type adults were chosen randomly from a designated group of fish reserved for embryo collection. Embryos collected from various wild-type crosses were combined. All zebrafish and embryo-related experiments were conducted at a temperature of 28°C. Experimental samples were collected at specified time points as mentioned in the manuscript. The maintenance of fish lines adhered to the research guidelines of the International Association for Assessment and Accreditation of Laboratory Animal Care, and animal protocols received approval from the Yale University Institutional Animal Care and Use Committee (IACUC).

METHOD DETAILS

RNA-sequencing analysis—Zebrafish mRNA sequencing datasets for embryonic developmental time course were obtained from previously published sources, namely SRA: SRP189512,[16] SRA: SRP149556,[62] and SRA: SRP072296[9] datasets. The ribosome profiling dataset generated in zebrafish embryos was from *Johnstone* et al. and the reads from start and stop positions were excluded to calculate gene ribosome density.[7] The gene counts, datasets, and genome browser tracks have been updated and can be accessed at https://data.giraldezlab.org. LabxDB was utilized to store, manage, and integrate the data into the bioinformatics pipeline for all high throughput sequencing experiments.[63] The half-life dataset (GEO: GSE153258) for HEK293 cells ($N = 2$) was generated using the TimeLapse-seq method.[22,23] We used Upf1 (Thermo Fisher Scientific, siRNA ID: s11928), G3bp1 (Thermo Fisher Scientific, siRNA ID: s19754), Smg6 (Thermo Fisher Scientific, siRNA ID: s23489) and negative control (Thermo Fisher Scientific, Catalog number:4390843) siRNAs in HEK293 cells and performed poly(A) selected RNA sequencing 2 days post transfection.

***Cis*-element selection and fold change analysis—**We used a custom R script for mRNA stability analysis (Figure 1). Briefly, we annotated all zebrafish and HEK293 mRNAs for frequency of *cis*-elements. In the 5′UTR we looked at Kozak score,[32] 5′UTR length, uORF or oORF (length, frequency, Kozak strength, count), and GC-richness. In the ORF we counted nucleotide/codon/amino acid frequencies, repeats of codons/amino acids, and ORF length. In the 3′UTR we searched for counts of k-mers of 6–8nt, 3′UTR length, and GC-richness. We further narrowed down the list of zebrafish mRNAs to only maternal mRNAs,[20,21] while for HEK293 we focused on genes with calculated half-lives. In zebrafish, we used lab generated developmental time-course mRNA-seq dataset (poly(A) selected) and picked several timepoints after fertilization: 0.2, 1.5, 2, 2.5, 3, 3.7, 4, 4.3, 5.3, and 6 h postfertilization (hpf). To assess time dependent mRNA decay in zebrafish we further divided genes into 4 groups based on the time points between which decay occurred: 0.2–6hpf (Group 1), 2–6hpf (Group 2), 2.5–6hpf (Group 3), and 4–6hpf (Group 4). Within each group we filtered *cis*-elements based on an enrichment of at least 20%, absolute Spearman correlation of *cis*-element occurrence to mRNA decay >0.01, and occurrence of *cis*-element in a group vs. all remaining mRNAs >1. Exact *cis*-element filtering was applied to the HEK293 dataset, except the whole dataset was divided into quartiles based on half-life values. To test what *cis*-elements affect mRNA stability we compared mRNA decay

differences (log2 Fold Change(late/early) for zebrafish, half-life for HEK293) between genes enriched in (>80th %) and depleted of (<20th %) a *cis*-element of interest. To calculate p values we used the unpaired two-samples Wilcoxon test with Bonferroni correction. We only considered *cis*-elements whose p value was <0.01. For zebrafish we highlighted the time interval (late/early) during which the difference in mRNA decay was the most significant. We merged the resulting *cis*-elements with mean differences in mRNA decay, removed duplicates and built a summary mRNA stability plot (Figures 1C and 1D).

Random forest regression model—Random forest regression models were used to predict mRNA decay. We used the output from the *cis*-element selection step as features (Figure 2). For both the zebrafish and HEK293 datasets we built tables with *cis*-element occurrence for every mRNA (maternal mRNAs from zebrafish) and the corresponding mRNA decay values (log2 Fold Change(6/2hpf) for zebrafish, half-life for HEK293).

To avoid the problem of model overfitting, the data was randomly divided into two groups: a test set and a training set, containing 20% and 80% of the mRNAs, respectively. The training data was divided into five equal-sized groups. The model was trained on four of these groups, using the fifth group as a test set to assess the accuracy of the model. This process was repeated five times, with each group serving as the test set once. To evaluate the performance of the model, a linear regression analysis was conducted, comparing the model's predictions with the experimental data from the test set. The mean prediction across all the test sets was calculated to provide a measure of the model's performance.

Yeast data analysis—To investigate factors involved in ORF length decay, we used an *S. cerevisiae* knockout screen from *Sun* et al.,[33] where the mRNA decay rate was measured. To narrow down genes based on codon optimality, we used percent optimal codons (cTE) values for *S. cerevisiae* from *Harigaya* et al.[34] To compare the difference in decay rates between genes with long (>90th %) and short (<10th %) ORFs we performed an unpaired, nonparametric Mann-Whitney test. Custom python script was used to graph violin plots. Similar analysis was performed on the Upf1 KO RNA-seq dataset from *Celik* et al.,[35] where log2 fold change between Upf1 and WT was compared.

MPRA decay assay—The transcriptome-based reporter library was created using overlap-extension PCR, where primers were designed to map to the SP6 promoter and downstream of the SV40 polyadenylation site.[14] Size-selected fragments of 56–129 nt were inserted downstream of GFP ORF using overlap-extension PCR. These libraries were then injected into zebrafish embryos at the one-cell stage. Following library preparation[14] and sequencing, 3′UTR MPRA profiles were generated for each UTR in the targeted library. These profiles represented the read coverage at different positions, normalized to counts per million (CPM), using the total counts of UTR profiles in each sample. The MPRA decay sequencing dataset was previously published SRA: SRP090954[14] and SRA: SRP189389.[16]

Reporter design/generation—ORF length reporters were generated as follows: a synthetic ORF with randomly positioned codons of specific Kozak sequence, length and optimality (Codon Adaptation Index: http://genomes.urv.es/CAIcal/) were ordered as gBlocks from IDT (Integrated DNA Technologies) and inserted into pCS2+ vector with

In-Fusion HD enzyme (Takara, 639642). Reporters had shared beta-globin 5′UTR and prrg2 or beta-globin 3′UTR followed by an SV40 poly(A) site. To make the 3x or 6x versions of ORFs, we cut the short ORF with a restriction enzyme (EcoRV or Sma1) and inserted the same ORF with In-Fusion HD enzyme. uORF/oORF reporters had GFP as the main ORF. 3xFLAG reporters were amplified from zebrafish cDNA and corresponding mutations for Upf1 were introduced with PCR. Mutagenized sites were conserved between human and zebrafish Upf1. Final constructs were confirmed by Sanger sequencing. Plasmids were linearized with NotI. The resulting PCR products were gel-purified and used as template for *in vitro* transcription using the SP6 mMessage mMachine kit (ThermoFisher, AM1340) to generate reporter mRNA. The resulting mRNA was purified using the RNeasy RNA extraction kit (QIAGEN, 74104). Zebrafish embryos were injected with 100 pg of reporter mRNA total. Thirty embryos were collected for each condition at different time points for RNA extraction and subsequent Northern blot analysis.

Luciferase assay—Zebrafish embryos were co-injected with 10pg each of nano and firefly luciferase mRNAs into 1-cell embryos. Reporters with short (1x titin domain) or long (4x titin domain) ORFs followed by a nanoluciferase sequence. A self-cleaving 2A peptide, P2A, was included between the titin ORF and nanoluciferase to allow efficient cleavage of the translated product in zebrafish.[31] 10 embryos were collected every hour, snap-frozen in liquid nitrogen and used for qPCR or dual luciferase assay. Nano and firefly luciferase activities were measured using the Nano-Glo Dual-Luciferase Reporter Assay System (Promega, N1610). Firefly luciferase activity was utilized as a control to standardize the measurements of nano luciferase. qPCR measurements for nano and firefly luciferase were used to further correct differential decay of reporters.

Reporter pulldown procedure—3xFLAG tagged pulldown experiments on UV crosslinked 3′UTR MPRA library and reporters were performed according to the RESA-CLIP procedure from *Yartseva* et al. *2017.*[14] First, 3xFLAG-tagged constructs were cloned and *in vitro* transcribed to capped mRNA. We injected either 50 or 150 embryos for single reporter or 3′UTR MPRA library, respectively, and pulled down UV crosslinked RNA-protein complexes on 3hpf embryos. Our input was 5% of the lysate's total volume that was used to quantify pulldown efficiency. We performed either qPCR or high-throughput sequencing on a NovaSeq 6000 for the single reporter or 3′UTR library, respectively.

K-mer enrichment analysis—The log2 ratio of normalized coverage between two experimental conditions was used to identify position-specific differential regulation/binding. Raw counts in each experiment were normalized by dividing by the total counts for that experiment (sum of all counts in all positions across all UTRs), and then the log2 ratio was calculated. We defined the list of sequences for k-mer analysis as those regions with a log2 ratio of < -0.3 for the MPRA decay assay or a log2 ratio of >0.5 for MPRA library pulldown assays with a minimum read length of 8nt. For both target and background sequences we searched for 5nt k-mers. To calculate k-mer enrichment we first counted k-mer occurrences in the target and background sequences. Next, the enrichment for the k-mer was computed as the ratio of target count to expected count. Significance was calculated using Fisher's exact test with the Bonferroni corrected p value.

Western blot—HEK293 cells were collected and heated at 70°C for 10 min in 19.5 μL of water, 7.5 μL of 4x NuPAGE LDS Sample Buffer (ThermoFisher Scientific), 3 μL DTT (Sigma-Aldrich). Samples were separated 1.0mm on 4–12% polyacrylamide NuPAGE Bis-Tris gels (ThermoFisher Scientific) for 50 min at 180 V and wet electrotransferred onto a nitrocellulose membrane (GE LifeSciences) for 70 min at 20 V. Membranes were incubated in blocking solution (5% milk in PBS-T) for 2 h. Then primary antibodies were diluted in blocking solution and incubated with the membrane overnight at 4°C with constant rotation in Anti-Upf1 1:5,000 (PA554031; Thermo Fisher Scientific), or anti-Actin 1:5,000 (A5060, Sigma). Secondary antibody Goat Anti-Rabbit IgG Antibody, (H + L) HRP conjugate (AP307P; Millipore) was incubated with the membrane at 1:10,000 for 1 h at room temperature. Membranes were washed three times for 5 min after each incubation and developed by chemiluminescent detection using ECL western blotting substrate (Thermo Fisher, 34095) and X-ray film (E3012, Denville Scientific).

Northern blot—Total RNA was extracted from zebrafish embryos with TRIZOL (Invitrogen, 15596–018). Briefly, 3 μg of total RNA was resuspended in 7.5μL of formamide and 7.5μL of 2x tracking dye (1mM EDTA, 60 mM triethanolamine, 60 mM tricine, 0.0004% bromophenol blue, 0.025% formaldehyde) and heated at 65°C for 10 min to denature the RNA. Samples were separated by electrophoresis using a 1% agarose/1.3% formaldehyde gel in 1x Tri/Tri buffer (30 mM triethanolamine, 30 mM tricine). The gel was capillary transferred to a Nytran SPC membrane (Whatman, 10416294) overnight. RNA was crosslinked to the membrane with 254 nm UV light at 1200 mJ. Membranes were prehybridized with 5 mL of ExpressHyb hybridization solution (Clontech, 686831) for 1 h at 37°C with constant rotation. RNA species were detected by oligonucleotide probes hybridized at 42°C, overnight with 5 mL of ExpressHyb solution and 5,000,000 cpm/ml of either the reporter or the U6 control radiolabeled probes (Table S5). Oligo probes were radiolabeled by T4 PNK end labeling (New England BioLabs, M0201S) with γ-P32-ATP. Radiolabeled probes were purified using ProbeQuant G-50 Micro Columns (GE Healthcare, 28903408). Membranes were washed once with 2x SSC/0.05% SDS for 10 min at room temperature and once with 0.1x SSC/0.1% SDS for 2 min at 42°C. Northern blots were quantitated using a phosphorimager (Bio-Rad Personal Molecular Imager) and analyzed with Fiji ImageJ. Levels of reporter mRNA were normalized to U6 controls.

qRT-PCR measurements of RNA abundance—Total RNA was extracted from 30 embryos per experimental condition and DNAse treated. cDNA was synthesized from 3 μg of total RNA using reverse transcription with random hexamers and the Superscript III reverse transcriptase kit (Invitrogen, 18080093). cDNA was diluted 1:20 and 10 μL reactions for PCR were prepared with 5 μL of Power SYBR Green PCR Master Mix (Applied Biosystems, 4368706), 4.5 μL of 1:100 diluted cDNA, and 0.5 μL of 10 mM forward and reverse primer mix. At least two biological and two technical replicates were performed for each sample. Relative expression was measured with ViiA 7 software v1.2.2 using the ΔΔCT method. Oligonucleotides used for qRT-PCR are listed in Table S5.

QUANTIFICATION AND STATISTICAL ANALYSIS

All analyses were performed with custom scripts written in Python 3 or R. For violin plots, solid lines represent the median. For yeast ORF length data comparisons, statistical significance was computed using the Mann-Whitney U test. Bar plots present mean values with standard deviation as the error bars. To test significance for cumulative plots, we performed a Kolmogorov-Smirnov test.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

We thank C. Castaldi, E. Sykes, B. Sullivan, and B. De Kumar from the Yale Center for Genome Analysis for sequencing support; S. Dube, T. Gerson, and D. Karpel for technical help; J.A. Schofield and M.D. Simon for providing the half-life dataset for HEK293 cells; S. Krishna and H. Lee for feedback on the manuscript; W.V. Gilbert for providing titin repeat plasmids; Y. Mishima for providing long 3′ UTR plasmids; reviewers for critical feedback that allowed us to improve the manuscript; and all members of the Giraldez lab for feedback and support.

Funding:

National Institutes of Health grant R01 HD100035 (A.J.G.) and National Institutes of Health grant R35 GM122580 (A.J.G.).

REFERENCES

1. Glisovic T, Bachorik JL, Yong J, and Dreyfuss G. (2008). RNA-binding proteins and post-transcriptional gene regulation. FEBS Lett. 582, 1977–1986. 10.1016/j.febslet.2008.03.004. [PubMed: 18342629]
2. Gerstberger S, Hafner M, and Tuschl T. (2014). A census of human RNA-binding proteins. Nat. Rev. Genet. 15, 829–845. 10.1038/nrg3813. [PubMed: 25365966]
3. Tadros W, and Lipshitz HD (2009). The maternal-to-zygotic transition: a play in two acts. Development 136, 3033–3042. 10.1242/dev.033183. [PubMed: 19700615]
4. Yartseva V, and Giraldez AJ (2015). The Maternal-to-Zygotic Transition During Vertebrate Development: A Model for Reprogramming. Curr. Top. Dev. Biol 113, 191–232. 10.1016/bs.ctdb.2015.07.020. [PubMed: 26358874]
5. Nyikó T, Sonkoly B, Mérai Z, Benkovics AH, and Silhavy D. (2009). Plant upstream ORFs can trigger nonsense-mediated mRNA decay in a size-dependent manner. Plant Mol. Biol 71, 367–378. 10.1007/s11103-009-9528-4. [PubMed: 19653106]
6. Barbosa C, Peixeiro I, and Romão L. (2013). Gene expression regulation by upstream open reading frames and human disease. PLoS Genet. 9, e1003529. 10.1371/journal.pgen.1003529.
7. Johnstone TG, Bazzini AA, and Giraldez AJ (2016). Upstream ORFs are prevalent translational repressors in vertebrates. EMBO J. 35, 706–723. 10.15252/embj.201592759. [PubMed: 26896445]
8. Presnyak V, Alhusaini N, Chen YH, Martin S, Morris N, Kline N, Olson S, Weinberg D, Baker KE, Graveley BR, and Coller J. (2015). Codon optimality is a major determinant of mRNA stability. Cell 160, 1111–1124. 10.1016/j.cell.2015.02.029. [PubMed: 25768907]
9. Bazzini AA, Del Viso F, Moreno-Mateos MA, Johnstone TG, Vejnar CE, Qin Y, Yao J, Khokha MK, and Giraldez AJ (2016). Codon identity regulates mRNA stability and translation efficiency during the maternal-to-zygotic transition. EMBO J. 35, 2087–2103. 10.15252/embj.201694699. [PubMed: 27436874]
10. Wu Q, Medina SG, Kushawah G, DeVore ML, Castellano LA, Hand JM, Wright M, and Bazzini AA (2019). Translation affects mRNA stability in a codon-dependent manner in human cells. Elife 8, e45396. 10.7554/eLife.45396.

11. Giraldez AJ, Mishima Y, Rihel J, Grocock RJ, Van Dongen S, Inoue K, Enright AJ, and Schier AF (2006). Zebrafish MiR-430 promotes deadenylation and clearance of maternal mRNAs. Science 312, 75–79. 10.1126/science.1122689. [PubMed: 16484454]
12. Lund E, Liu M, Hartley RS, Sheets MD, and Dahlberg JE (2009). Deadenylation of maternal mRNAs mediated by miR-427 in Xenopus laevis embryos. RNA 15, 2351–2363. 10.1261/rna.1882009. [PubMed: 19854872]
13. Bushati N, Stark A, Brennecke J, and Cohen SM (2008). Temporal reciprocity of miRNAs and their targets during the maternal-to-zygotic transition in Drosophila. Curr. Biol 18, 501–506. 10.1016/j.cub.2008.02.081. [PubMed: 18394895]
14. Yartseva V, Takacs CM, Vejnar CE, Lee MT, and Giraldez AJ (2017). RESA identifies mRNA-regulatory sequences at high resolution. Nat. Methods 14, 201–207. 10.1038/nmeth.4121. [PubMed: 28024160]
15. Rabani M, Pieper L, Chew GL, and Schier AF (2017). A Massively Parallel Reporter Assay of 3' UTR Sequences Identifies In Vivo Rules for mRNA Degradation. Mol. Cell 68, 1083–1094.e5. 10.1016/j.molcel.2017.11.014. [PubMed: 29225039]
16. Vejnar CE, Abdel Messih M, Takacs CM, Yartseva V, Oikonomou P, Christiano R, Stoeckius M, Lau S, Lee MT, Beaudoin JD, et al. (2019). Genome wide analysis of 3' UTR sequence elements and proteins regulating mRNA stability during maternal-to-zygotic transition in zebrafish. Genome Res. 29, 1100–1114. 10.1101/gr.245159.118. [PubMed: 31227602]
17. Mishima Y, and Tomari Y. (2016). Codon Usage and 3' UTR Length Determine Maternal mRNA Stability in Zebrafish. Mol. Cell 61, 874–885. 10.1016/j.molcel.2016.02.027. [PubMed: 26990990]
18. Zhang L, Kasif S, Cantor CR, and Broude NE (2004). GC/AT-content spikes as genomic punctuation marks. Proc. Natl. Acad. Sci. USA 101, 16855–16860. 10.1073/pnas.0407821101. [PubMed: 15548610]
19. Litterman AJ, Kageyama R, Le Tonqueze O, Zhao W, Gagnon JD, Goodarzi H, Erle DJ, and Ansel KM (2019). A massively parallel 3' UTR reporter assay reveals relationships between nucleotide content, sequence conservation, and mRNA destabilization. Genome Res. 29, 896–906. 10.1101/gr.242552.118. [PubMed: 31152051]
20. Lee MT, Bonneau AR, Takacs CM, Bazzini AA, DiVito KR, Fleming ES, and Giraldez AJ (2013). Nanog, Pou5f1 and SoxB1 activate zygotic gene expression during the maternal-to-zygotic transition. Nature 503, 360–364. 10.1038/nature12632. [PubMed: 24056933]
21. Heyn P, Kircher M, Dahl A, Kelso J, Tomancak P, Kalinka AT, and Neugebauer KM (2014). The earliest transcribed zygotic genes are short, newly evolved, and different across species. Cell Rep. 6, 285–292. 10.1016/j.celrep.2013.12.030. [PubMed: 24440719]
22. Schofield JA, Duffy EE, Kiefer L, Sullivan MC, and Simon MD (2018). TimeLapse-seq: adding a temporal dimension to RNA sequencing through nucleoside recoding. Nat. Methods 15, 221–225. 10.1038/nmeth.4582. [PubMed: 29355846]
23. Luo Y, Schofield JA, Na Z, Hann T, Simon MD, and Slavoff SA (2021). Discovery of cellular substrates of human RNA-decapping enzyme DCP2 using a stapled bicyclic peptide inhibitor. Cell Chem. Biol 28, 463–474.e7. 10.1016/j.chembiol.2020.12.003.
24. Subtelny AO, Eichhorn SW, Chen GR, Sive H, and Bartel DP (2014). Poly(A)-tail profiling reveals an embryonic switch in translational control. Nature 508, 66–71. 10.1038/nature13007. [PubMed: 24476825]
25. Fan XC, Myer VE, and Steitz JA (1997). AU-rich elements target small nuclear RNAs as well as mRNAs for rapid degradation. Genes Dev. 11, 2557–2568. 10.1101/gad.11.19.2557. [PubMed: 9334320]
26. Medina-Muñoz SG, Kushawah G, Castellano LA, Diez M, DeVore ML, Salazar MJB, and Bazzini AA (2021). Crosstalk between codon optimality and cis-regulatory elements dictates mRNA stability. Genome Biol. 22, 14. 10.1186/s13059-020-02251-5. [PubMed: 33402205]
27. Geisberg JV, Moqtaderi Z, Fan X, Ozsolak F, and Struhl K. (2014). Global analysis of mRNA isoform half-lives reveals stabilizing and destabilizing elements in yeast. Cell 156, 812–824. 10.1016/j.cell.2013.12.026. [PubMed: 24529382]

28. Chan LY, Mugler CF, Heinrich S, Vallotton P, and Weis K. (2018). Non-invasive measurement of mRNA decay reveals translation initiation as the major determinant of mRNA stability. Elife 7, e32536. 10.7554/eLife.32536.
29. Duan J, Shi J, Ge X, Dölken L, Moy W, He D, Shi S, Sanders AR, Ross J, and Gejman PV (2013). Genome-wide survey of interindividual differences of RNA stability in human lymphoblastoid cell lines. Sci. Rep 3, 1318. 10.1038/srep01318. [PubMed: 23422947]
30. Narula A, Ellis J, Taliaferro JM, and Rissland OS (2019). Coding regions affect mRNA stability in human cells. RNA 25, 1751–1764. 10.1261/rna.073239.119. [PubMed: 31527111]
31. Kim JH, Lee SR, Li LH, Park HJ, Park JH, Lee KY, Kim MK, Shin BA, and Choi SY (2011). High cleavage efficiency of a 2A peptide derived from porcine teschovirus-1 in human cell lines, zebrafish and mice. PLoS One 6, e18556. 10.1371/journal.pone.0018556.
32. Grzegorski SJ, Chiari EF, Robbins A, Kish PE, and Kahana A. (2014). Natural variability of Kozak sequences correlates with function in a zebrafish model. PLoS One 9, e108475. 10.1371/journal.pone.0108475.
33. Sun M, Schwalb B, Pirkl N, Maier KC, Schenk A, Failmezger H, Tresch A, and Cramer P. (2013). Global analysis of eukaryotic mRNA degradation reveals Xrn1-dependent buffering of transcript levels. Mol. Cell 52, 52–62. 10.1016/j.molcel.2013.09.010. [PubMed: 24119399]
34. Harigaya Y, and Parker R. (2016). Analysis of the association between codon optimality and mRNA stability in Schizosaccharomyces pombe. BMC Genom. 17, 895. 10.1186/s12864-016-3237-6.
35. Celik A, Baker R, He F, and Jacobson A. (2017). High-resolution profiling of NMD targets in yeast reveals translational fidelity as a basis for substrate selection. RNA 23, 735–748. 10.1261/rna.060541.116. [PubMed: 28209632]
36. Wahle E, and Winkler GS (2013). RNA decay machines: deadenylation by the Ccr4-not and Pan2-Pan3 complexes. Biochim. Biophys. Acta 1829, 561–570. 10.1016/j.bbagrm.2013.01.003. [PubMed: 23337855]
37. He F, and Jacobson A. (2015). Nonsense-Mediated mRNA Decay: Degradation of Defective Transcripts Is Only Part of the Story. Annu. Rev. Genet 49, 339–366. 10.1146/annurev-genet-112414-054639. [PubMed: 26436458]
38. Fourati Z, Roy B, Millan C, Coureux PD, Kervestin S, van Tilbeurgh H, He F, Usón I, Jacobson A, and Graille M. (2014). A highly conserved region essential for NMD in the Upf2 N-terminal domain. J. Mol. Biol 426, 3689–3702. 10.1016/j.jmb.2014.09.015. [PubMed: 25277656]
39. Xing W, Muhlrad D, Parker R, and Rosen MK (2020). A quantitative inventory of yeast P body proteins reveals principles of composition and specificity. Elife 9, e56525. 10.7554/eLife.56525.
40. Sheth U, and Parker R. (2003). Decapping and decay of messenger RNA occur in cytoplasmic processing bodies. Science 300, 805–808. 10.1126/science.1082320. [PubMed: 12730603]
41. Sheth U, and Parker R. (2006). Targeting of aberrant mRNAs to cytoplasmic processing bodies. Cell 125, 1095–1109. 10.1016/j.cell.2006.04.037. [PubMed: 16777600]
42. Mugler CF, Hondele M, Heinrich S, Sachdev R, Vallotton P, Koek AY, Chan LY, and Weis K. (2016). ATPase activity of the DEAD-box protein Dhh1 controls processing body formation. Elife 5, e18746. 10.7554/eLife.18746.
43. Carroll JS, Munchel SE, and Weis K. (2011). The DExD/H box ATPase Dhh1 functions in translational repression, mRNA decay, and processing body dynamics. J. Cell Biol. 194, 527–537. 10.1083/jcb.201007151. [PubMed: 21844211]
44. Coller J, and Parker R. (2005). General translational repression by activators of mRNA decapping. Cell 122, 875–886. 10.1016/j.cell.2005.07.012. [PubMed: 16179257]
45. Fischer N, and Weis K. (2002). The DEAD box protein Dhh1 stimulates the decapping enzyme Dcp1. EMBO J. 21, 2788–2797. 10.1093/emboj/21.11.2788. [PubMed: 12032091]
46. Sweet T, Kovalak C, and Coller J. (2012). The DEAD-box protein Dhh1 promotes decapping by slowing ribosome movement. PLoS Biol. 10, e1001342. 10.1371/journal.pbio.1001342.
47. Swisher KD, and Parker R. (2011). Interactions between Upf1 and the decapping factors Edc3 and Pat1 in Saccharomyces cerevisiae. PLoS One 6, e26547. 10.1371/journal.pone.0026547.

48. Metze S, Herzog VA, Ruepp MD, and Mühlemann O. (2013). Comparison of EJC-enhanced and EJC-independent NMD in human cells reveals two partially redundant degradation pathways. RNA 19, 1432–1448. 10.1261/rna.038893.113. [PubMed: 23962664]
49. Isken O, Kim YK, Hosoda N, Mayeur GL, Hershey JWB, and Maquat LE (2008). Upf1 phosphorylation triggers translational repression during nonsense-mediated mRNA decay. Cell 133, 314–327. 10.1016/j.cell.2008.02.030. [PubMed: 18423202]
50. Kurosaki T, Li W, Hoque M, Popp MWL, Ermolenko DN, Tian B, and Maquat LE (2014). A post-translational regulatory switch on UPF1 controls targeted mRNA degradation. Genes Dev. 28, 1900–1916. 10.1101/gad.245506.114. [PubMed: 25184677]
51. Thompson MK, Rojas-Duran MF, Gangaramani P, and Gilbert WV (2016). The ribosomal protein Asc1/RACK1 is required for efficient translation of short mRNAs. Elife 5, e11154. 10.7554/eLife.11154.
52. Kerr SC, Azzouz N, Fuchs SM, Collart MA, Strahl BD, Corbett AH, and Laribee RN (2011). The Ccr4-Not complex interacts with the mRNA export machinery. PLoS One 6, e18302. 10.1371/journal.pone.0018302.
53. Singh AK, Choudhury SR, De S, Zhang J, Kissane S, Dwivedi V, Ramanathan P, Petric M, Orsini L, Hebenstreit D, and Brogna S. (2019). The RNA helicase UPF1 associates with mRNAs co-transcriptionally and is required for the release of mRNAs from gene loci. Elife 8, e41444. 10.7554/eLife.41444.
54. Hurt JA, Robertson AD, and Burge CB (2013). Global analyses of UPF1 binding and function reveal expanded scope of nonsense-mediated mRNA decay. Genome Res. 23, 1636–1650. 10.1101/gr.157354.113. [PubMed: 23766421]
55. Hogg JR, and Goff SP (2010). Upf1 senses 3'UTR length to potentiate mRNA decay. Cell 143, 379–389. 10.1016/j.cell.2010.10.005. [PubMed: 21029861]
56. Imamachi N, Salam KA, Suzuki Y, and Akimitsu N. (2017). A GC-rich sequence feature in the 3' UTR directs UPF1-dependent mRNA decay in mammalian cells. Genome Res. 27, 407–418. 10.1101/gr.206060.116. [PubMed: 27940950]
57. Lawir DF, Sikora K, O'Meara CP, Schorpp M, and Boehm T. (2020). Pervasive changes of mRNA splicing in upf1-deficient zebrafish identify rpl10a as a regulator of T cell development. Proc. Natl. Acad. Sci. USA 117, 15799–15808. 10.1073/pnas.1917812117. [PubMed: 32571908]
58. Zünd D, Gruber AR, Zavolan M, and Mühlemann O. (2013). Translation-dependent displacement of UPF1 from coding sequences causes its enrichment in 3' UTRs. Nat. Struct. Mol. Biol 20, 936–943. 10.1038/nsmb.2635. [PubMed: 23832275]
59. Juszkiewicz S, Chandrasekaran V, Lin Z, Kraatz S, Ramakrishnan V, and Hegde RS (2018). ZNF598 Is a Quality Control Sensor of Collided Ribosomes. Mol. Cell 72, 469–481.e7. 10.1016/j.molcel.2018.08.037. [PubMed: 30293783]
60. Nagy E, and Maquat LE (1998). A rule for termination-codon position within intron-containing genes: when nonsense affects RNA abundance. Trends Biochem. Sci 23, 198–199. 10.1016/s0968-0004(98)01208-0. [PubMed: 9644970]
61. Lykke-Andersen S, and Jensen TH (2015). Nonsense-mediated mRNA decay: an intricate machinery that shapes transcriptomes. Nat. Rev. Mol. Cell Biol. 16, 665–677. 10.1038/nrm4063. [PubMed: 26397022]
62. Beaudoin JD, Novoa EM, Vejnar CE, Yartseva V, Takacs CM, Kellis M, and Giraldez AJ (2018). Analyses of mRNA structure dynamics identify embryonic gene regulatory programs. Nat. Struct. Mol. Biol 25, 677–686. 10.1038/s41594-018-0091-z. [PubMed: 30061596]
63. Vejnar CE, and Giraldez AJ (2020). LabxDB: versatile databases for genomic sequencing and lab management. Bioinformatics 36, 4530–4531. 10.1093/bioinformatics/btaa557. [PubMed: 32502232]

Highlights

- ORF length, uORFs, Kozak sequence, codon optimality are conserved mRNA regulatory *cis*-elements
- mRNAs with long ORFs are poorly translated and unstable
- Upf1 binds mRNAs with long ORFs, uORFs, untranslated ORFs, and GC-rich mRNA to promote decay
- Upf1's phosphorylation is critical for binding its targets

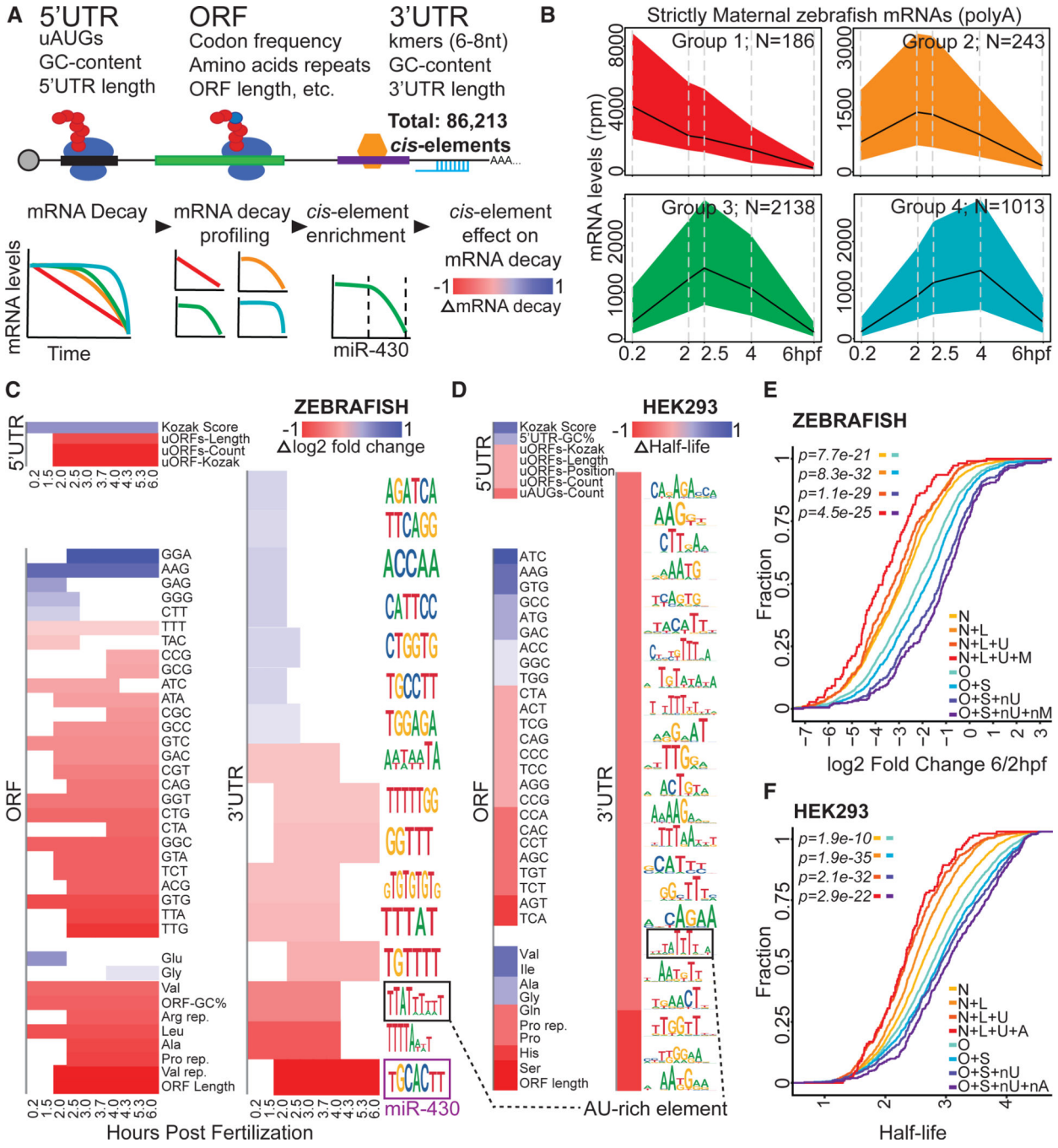


Figure 1. Select mRNA cis-elements affect mRNA stability in zebrafish early development

(A) Schematics showing the locations of the 86,213 mRNA *cis*-elements that were searched for in zebrafish and HEK293 cells, and how *cis*-element enrichment and the effect on mRNA decay were defined based on mRNAs that were grouped by decay profile.

(B) Strictly maternal zebrafish mRNAs, representing the poly(A)-selected transcriptome, grouped according to their mRNA decay profiles. The relative increase in zebrafish strictly maternal mRNA levels over early embryonic development in some clusters is due to poly(A) tail lengthening[24] (Ribo 0 data: Figure S1A).

(C) Heatmaps showing the effect on mRNA stability due to enrichment in the identified *cis*-elements for strictly maternal zebrafish genes. Values represent the mean log2 fold change difference between the 80th % and 20th % of genes enriched for a given *cis*-element. The most significantly different (t test) time interval was plotted for each *cis*-element. Red, decay; blue, stability.

(D) Mean half-life difference in mRNA stability between the 80th % and 20th % of genes enriched for *cis*-elements in HEK293 cells. Red, decay; blue, stability.

(E) The additive effects of mRNA *cis*-elements. Cumulative plots representing the log2 fold change (6/2 hpf) of zebrafish maternal mRNAs that are enriched (top vs. bottom 50%) in the indicated combinations of elements. N, non-optimal; L, long ORF; U, uORFs; M, miR-430; O, optimal; S, short ORF; nU, no uORFs; nM, no miR-430; N (n = 1,520 genes), N + L (n = 760), N+ L+ U (n = 341), N + L + U + M (n = 109), O (n = 1,520), O + S (n = 758), O + S + nU (n = 294), O + S + nU + nM (n = 221).

(F) Cumulative plot representing the half-lives of HEK293 mRNAs that are enriched (top vs. bottom 50%) in the indicated combinations of elements. N, non-optimal; L, long ORF; U, uORFs; A, AU-rich element, TATTTAT; O, optimal; S, short ORF; nU, no uORFs; nA, no AU-rich element; N (n = 2,103 genes), N + L (n = 1051), N + L + U (n = 420), N+L+U+A (n = 128), O (n = 2,103), O + S (n = 1,051), O + S + nU (n = 556), O + S + nU + nA (n = 425). To test statistical significance, we used a Mann-Whitney test.

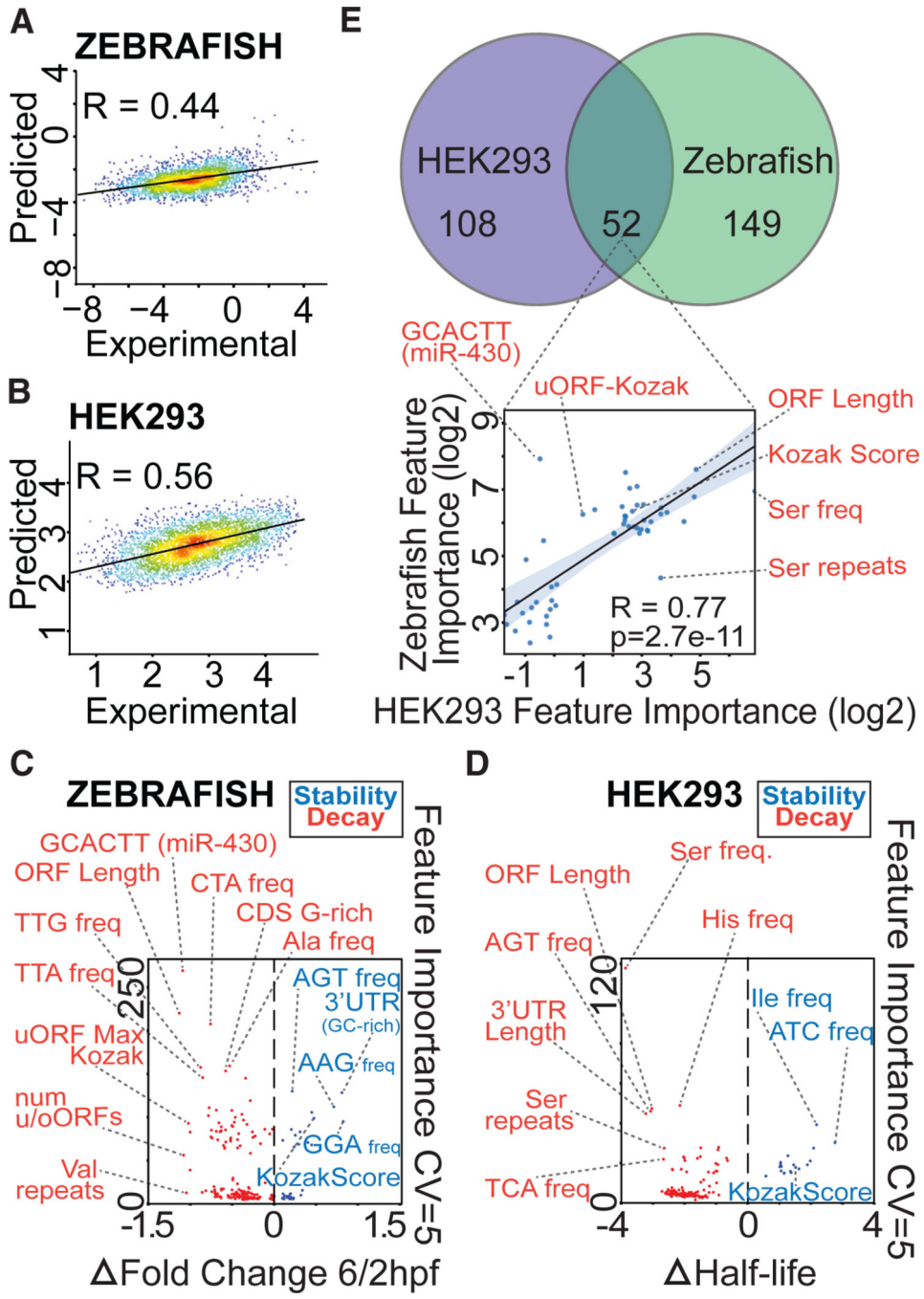


Figure 2. Conservation of mRNA cis-elements and their predictive power

(A and B) Scatterplot comparing each model's prediction to experimentally derived decay datasets in zebrafish and HEK293 cells (Pearson's R). Zebrafish log2 fold change (6/2 hpf) and HEK293 half-lives were used for experimental values.

(C and D) Five-fold cross-validated feature average importance was generated using each of the random-forest machine-learning models and plotted against the difference in the stabilities of mRNAs enriched in (80th %) or depleted of (20th %) elements of interest. mRNA stability was calculated as the log2 fold change (6/2 hpf) for zebrafish mRNAs (C)

and mean half-life for HEK293 mRNAs (D). Decay elements are in red; stability elements are in blue.

(E) *cis*-elements that are shared between HEK293 cells and zebrafish. Feature importance correlation between zebrafish and HEK293 cells.

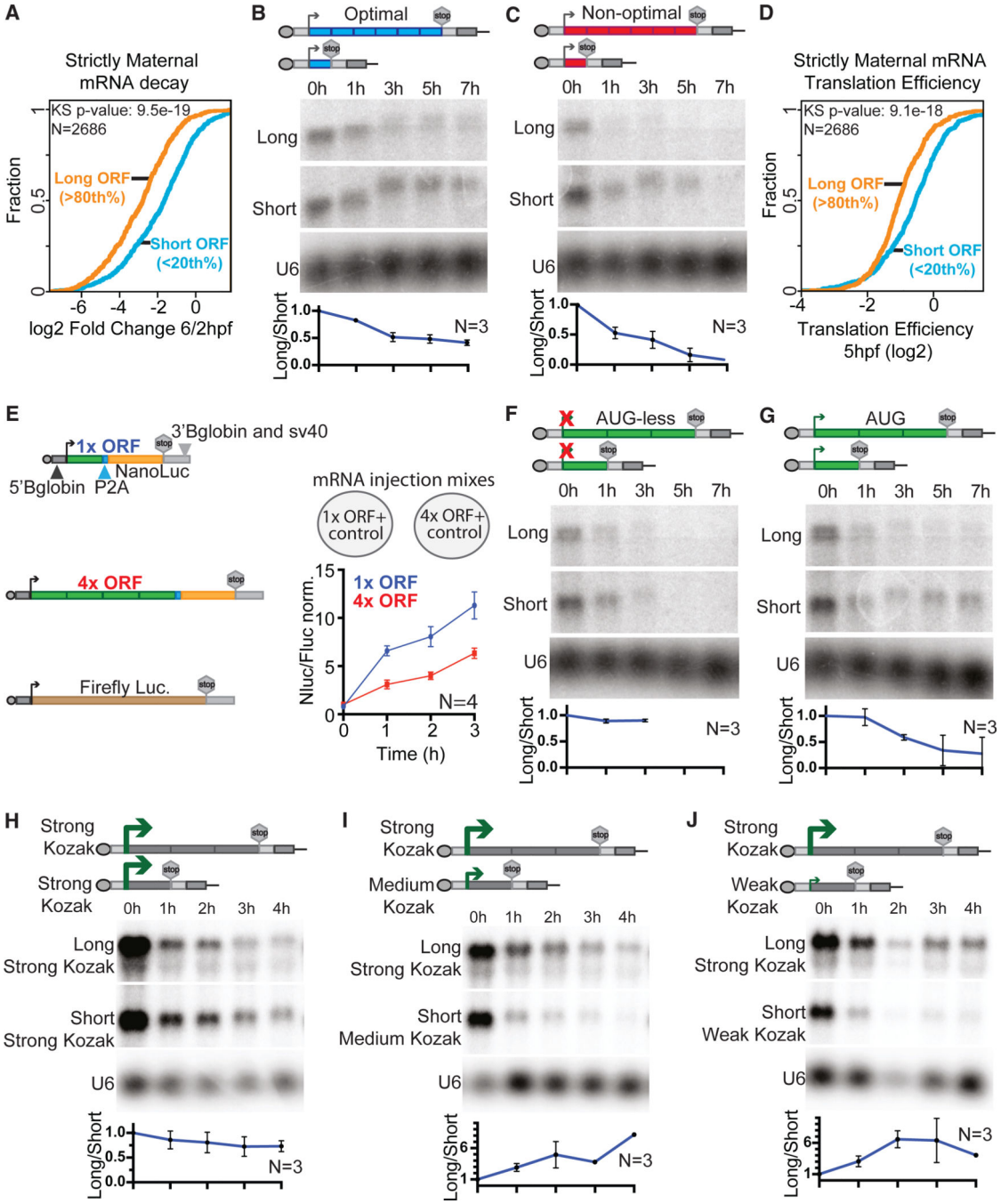


Figure 3. ORF length correlates with mRNA decay due to differences in translation efficiency

(A–C) (A) In zebrafish, the decay rate of long ORF mRNAs (>80th % in length) is faster than the decay rate of short ORF mRNAs (<20th % in length). Data are shown for the log2 fold change (6/2 hpf) of zebrafish strictly maternal genes. To test significance, we performed a Kolmogorov-Smirnov test; (B and C) Northern blot of the long (1,800 bp) ORF construct decaying faster than the short (600 bp) ORF construct. Constructs have optimality CAI scores of 0.91 in (B) and 0.6 in (C). Ratios of signals from long ORFs to short ORFs are plotted below. mRNA bands shift due to injected mRNA polyadenylation.[11,17,24]

(D) The translation efficiency (TE) of 5-hpf zebrafish long-ORF mRNAs is lower compared to short-ORF mRNAs. To test significance, we performed a Kolmogorov-Smirnov test.
(E) A dual-luciferase assay normalized to mRNA levels (Figure S2A; long ORFs decay faster) shows that short-ORF mRNAs have higher TE ($n = 4$). ORF length with Nano luciferase: 13 = 915 bp, CAI = 0.811; 43 = 1734 bp, CAI = 0.811. Firefly-luciferase ORF = 1,653 bp.
(F) 600-bp short-ORF and 1,800-bp long-ORF mRNAs with deleted start codons decay at a similar rate. Neither construct has AUG or CUG in any of the three frames.
(G) The 600-bp short ORF with AUG is more stable than the 1,800-bp long-ORF mRNA with AUG. Both have CAI scores of 0.87.
(H–J) Reduced translation initiation due to the introduction of weaker Kozak sequences in short-ORF mRNAs led to faster decay of short ORFs compared to the long-ORF construct with a strong Kozak sequence. Strong Kozak, GCAAAC**ATG**G; medium Kozak, CTTTCT**ATG**C; weak Kozak, CGTTGT**ATG**C. For all northern blots, the U6 loading control is shown, and mean values with ±1 standard deviation (SD) error bars are plotted for RNA quantification.

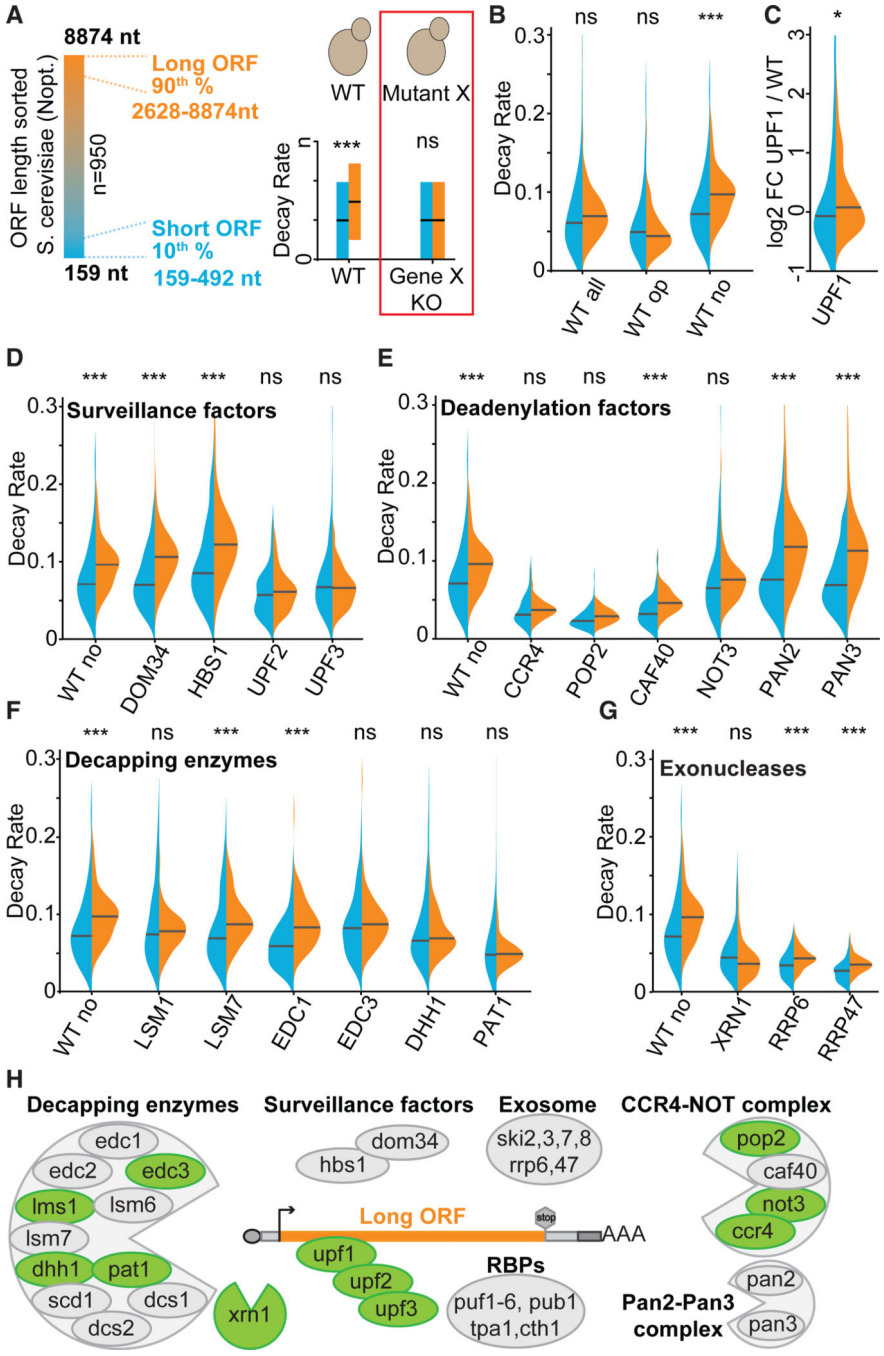


Figure 4. Yeast screen identifies genes in the ORF length decay pathway

(A) Long-ORF mRNAs in *S. cerevisiae* (>90th % in length) have faster decay rates than short-ORF mRNAs (<10th % in length). A starting subset of genes representing the top quartile of non-optimal coding sequences was used (n = 950). Mutants for genes with non-significant differences in decay rates between these two groups of ORF mRNAs potentially participate in the ORF length decay pathway. Short ORF, n = 95; long ORF, n = 95.

(B) Decay rates of all genes, top 25th % of optimal ORF genes, and top 25th % of non-optimal ORF genes. In each dataset, genes with ORF length >90th % and <10th % are

compared. All subsequent analyses compared this subset of genes among non-optimal ORF genes.

(C and D) Among mRNA surveillance genes, Upf1, Upf2, and Upf3 sense mRNAs with long ORFs. We used log2 fold change of mRNA expression between *UPF1* mutant and WT (C) and mRNAs with long ORFs get relatively stabilized. The decay rates of only long ORF mRNAs are significantly decreased in *UPF2* and *UPF2* mutants (D).

(E) Among mRNA deadenylation genes, Ccr4, Pop2, and Not3 sense mRNAs with long ORFs.

(F) Among mRNA decapping genes, Lsm1, Edc3, Dhh1, and Pat1 sense mRNAs with long ORFs.

(G) Xrn1, an exonuclease, senses mRNAs with long ORFs.

(H) Summarized depiction of the factors involved in the pathway for sensing long ORFs. For the highlighted factors, when mutated, the decay rate of long ORFs was found to be comparable to that of short ORFs. However, for the remaining depicted factors, there remained significant differences in the decay rates between long ORFs and short ORFs when mutated. We performed Mann-Whitney U test with Bonferroni-corrected p values: *$p \leq 0.0005$, **$p \leq 0.0001$, ***$p \leq 0.00001$. The decay rate datasets from KO strains are from Sun et al.,[33] with the exception of (C) (from Celik et al.[35]).

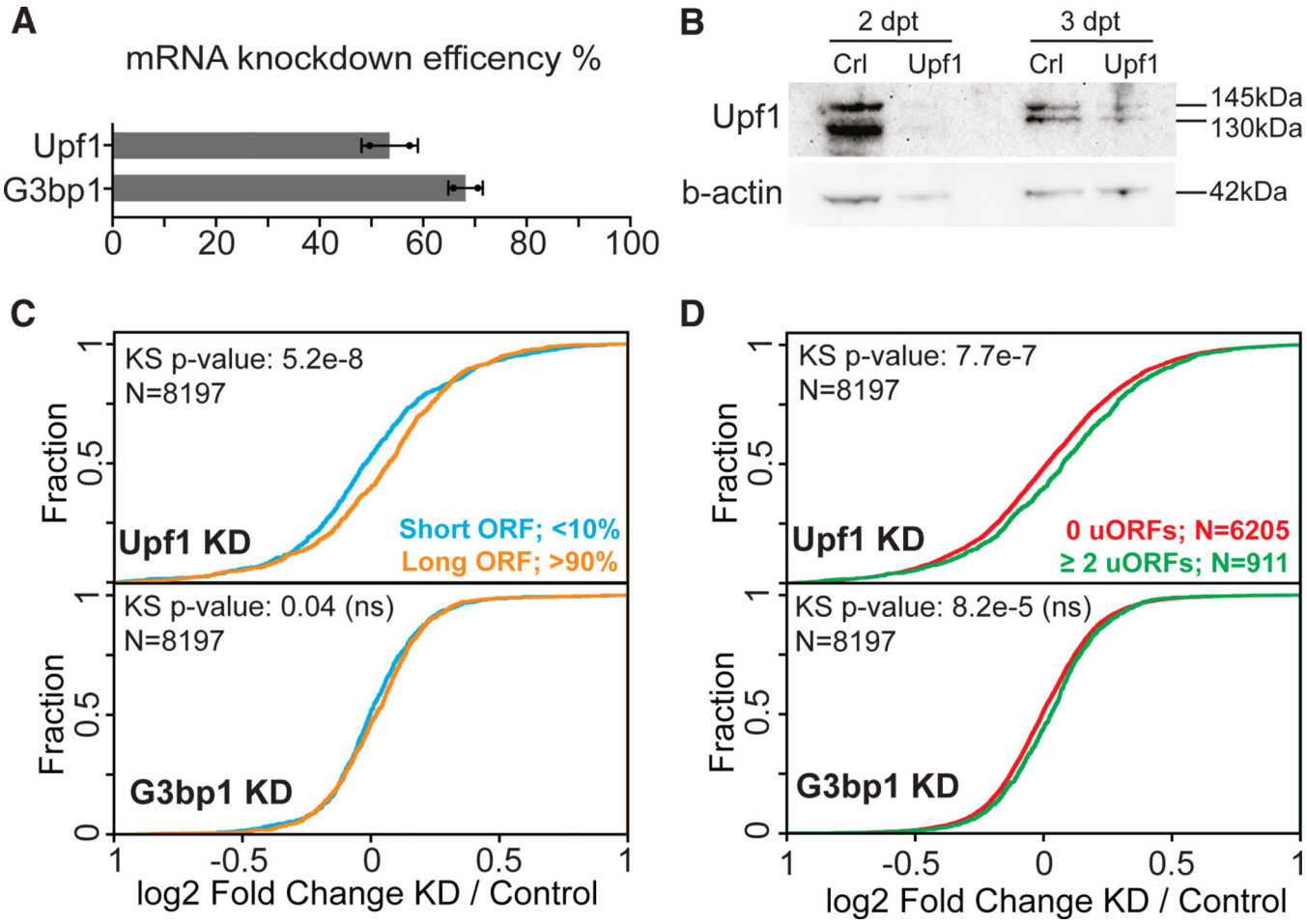


Figure 5. mRNAs with long ORFs and uORFs are stabilized in Upf1 KD HEK293 cells

(A) KD efficiency of Upf1 in Upf1 KD cells ($n = 2$) and G3bp1 in G3bp1 KD cells ($n = 2$) compared to negative control.

(B) Western blot of Upf1 protein 2 and 3 days post transfection (dpt) with Upf1 siRNA. B-actin used as a loading control.

(C and D) mRNAs with long ORFs (C) and uORFs

(D) are significantly stabilized upon Upf1 KD, but not in G3bp1 KD cells. Log2 fold change relative to negative control is used as a measure of transcriptome expression changes when Upf1 or G3bp1 are KD. To test significance, we performed a Kolmogorov-Smirnov test. Bonferroni-corrected p value threshold = 8e–6.

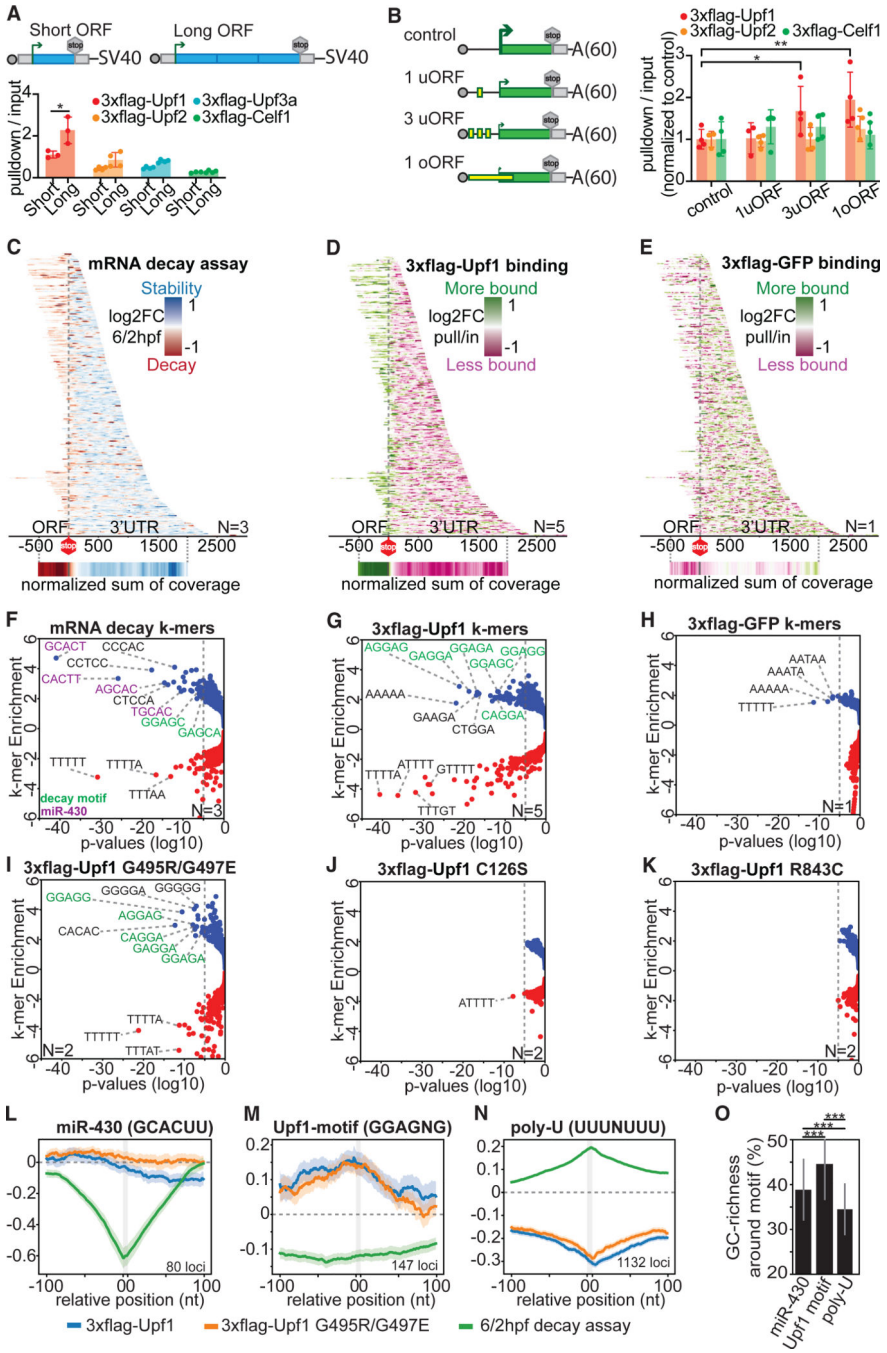


Figure 6. Upf1 binds mRNAs with reduced translation initiation in zebrafish

(A) Upf1 binds to mRNAs with longer ORFs. Pull-down vs. input was normalized to the endogenous gene PRPF39. qPCR analysis of reporter pull-down using 3xFLAG-Upf1, 3xFLAG-Upf2, 3xFLAG-Upf3a, and control 3xFLAG-Celf1 was performed on embryos collected at 3 h post injection.

(B) Upf1 binds to mRNAs with u/oORFs ($n = 4$). We used u/oORF reporters from Johnstone et al.[7] The diagram shows reporter designs with and without uORFs and oORFs represented by yellow boxes upstream of the main GFP ORF (861 nt, green box). 5′ UTR length (104

nt), 3′ UTR length (216 nt), and poly(A) length (60 nt) were constant for all reporters and 5′ UTRs differed by a single nucleotide change. uORFs were 18 nt each, while the oORF was 90 nt and overlapped with the first 35 nt of the main ORF. The size of the green arrow represents translation potential. qPCR analysis of reporter pull-down done as in (A). Pull-down enrichment over input was normalized to the control reporter and plotted using CT values normalized to the endogenous gene PRPF39. To test statistical significance, we used a Mann-Whitney test. *$p \leq 0.05$; **$p \leq 0.01$. Mean values with ±1 SD error bars are plotted.

(C) Heatmap representing the log2 fold change of profile coverage (6/2 hpf) at single-base resolution. ORF-derived sequences correlate with mRNA decay based on the MPRA library decay assay. Each of the 350 horizontal lines represents a region of an endogenous gene centered at a stop codon. Sum of regions of coverage normalized to nonzero coverage indicated higher mRNA decay from ORF-derived 3′ UTR fragments. $n = 3$.

(D and E) Upf1 binds to fragments derived from ORFs. The pull-down profiles of the MPRA reporter with (D) 3xFLAG-Upf1 and (E) 3xFLAG-GFP are shown. The library was pulled down from zebrafish embryos at 3 h post injection. The log2 fold change of the pull-down profile coverage compared to the input is plotted at single-base resolution. 3xFLAG-Upf1 $n = 5$, 3xFLAG-GFP $n = 1$.

(F–K) Upf1 binds decay motifs and phosphorylation is critical for binding specificity. 5-bp k-mer enrichment and depletion from the list of decay sequences (F) or 3xFLAG-Upf1 (G), 3xFLAG-GFP (H), 3xFLAG-Upf1 G495R/G497E (I), 3xFLAG-Upf1 C126S (J), and 3xFLAG-Upf1 R843C (K) binding regions relative to all sequences. Significance was calculated using a Fisher's exact test, with the Bonferroni-corrected p value threshold indicated by the dashed line. An A-rich k-mer was significantly enriched in CLIP experiments, likely due to non-specific binding also seen in the control GFP CLIP (H).

(L–N) Iceberg plots centered at miR-430 (L), Upf1 (M), or poly(U) (N) motifs. The log2 values (pull-down/input) of 3xFLAG-Upf1 and 3xFLAG-Upf1-G495R/G497E, along with log2 values of 6/2hpf 3′ UTR decay, are shown.

(O) Mean GC richness in regions ±100 bp around the Upf1 ($n = 147$), poly(U) ($n = 1,132$), or miR-430 ($n = 80$) motif. Error bars represent 1 SD. Mann-Whitney U test: miR-430 vs. Upf1, $p = 1.04e{-11}$; miR-430 vs. poly(U), $p = 2.42e{-14}$; Upf1 vs. poly(U), $p = 3.09e{-65}$.

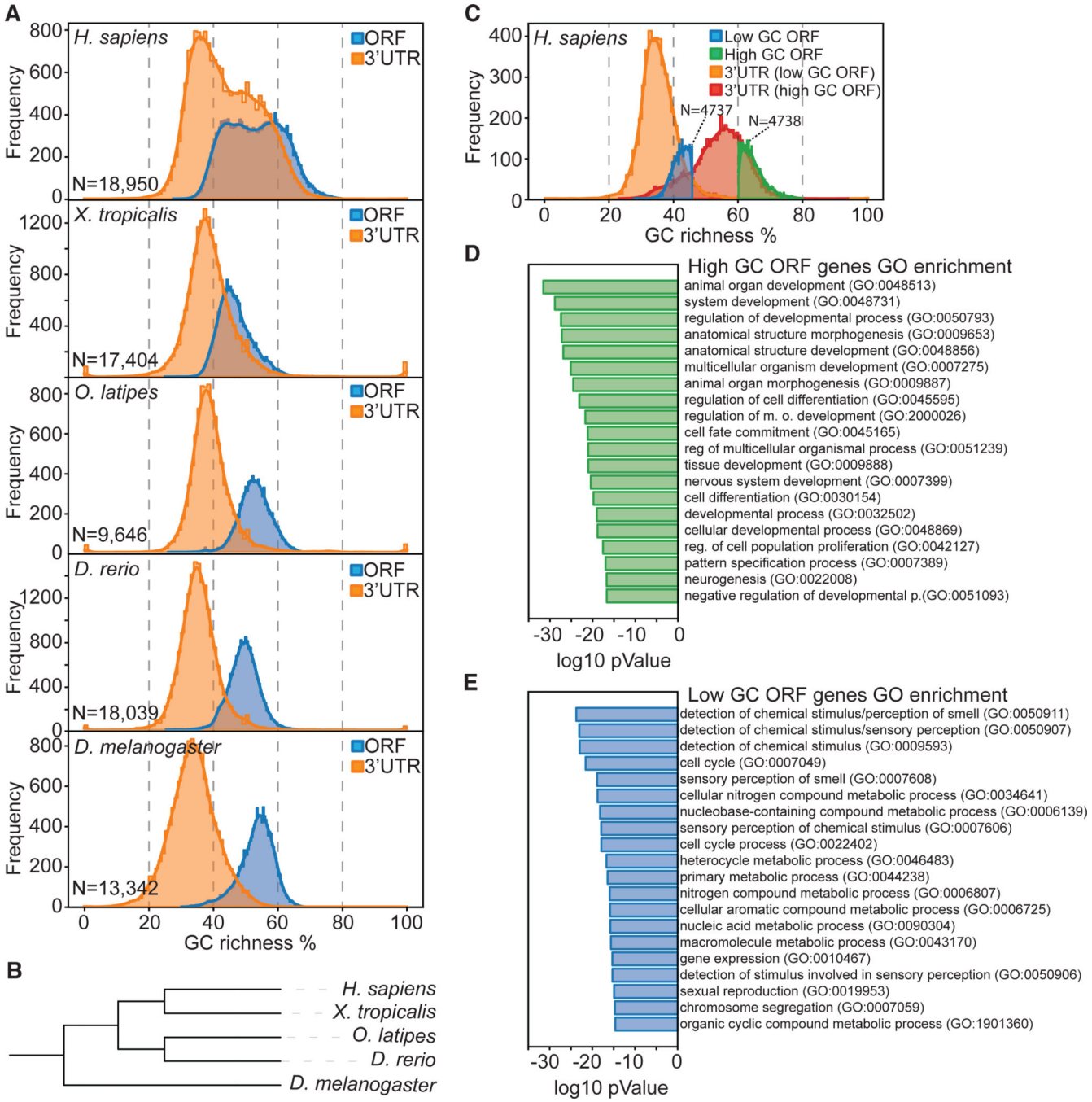


Figure 7. GC-rich ORFs are enriched in developmental genes

(A) GC richness of ORFs and 3′ UTRs in human (*Homo sapiens*), clawed frog (*Xenopus tropicalis*), zebrafish (*Danio rerio*), medaka (*Oryzias latipes*), and fruit fly (*Drosophila melanogaster*) protein-coding genes. Longest transcripts from Ensembl v108 were used.

(B) Species phylogenetic tree generated with https://phylot.biobyte.de.)

(C) High-GC-content ORF genes are enriched in GC-rich 3′ UTRs. Human ORFs were separated into <25th and >75th percentile GC-content groups. 3′-UTR GC content of the same group of genes was plotted.

(D) Human high-GC-content ORFs are enriched in developmental genes. We performed a GO PANTHER overrepresentation test, where genes with high-GC-content ORFs were compared to those with low GC content.

(E) Human low-GC-content ORFs are enriched in sensory, cell-cycle, and metabolic genes. We performed a GO PANTHER overrepresentation test, where genes with low-GC-content ORFs were compared to those with high GC content.

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Anti-Upf1 antibody	Thermo Fisher Scientific	Cat# PA5–54031; RRID:AB_2649284
Anti-Actin antibody	Sigma-Aldrich	Cat# A5060; RRID:AB_476738
Goat Anti-Rabbit IgG Polyclonal antibody, Horseradish peroxidase Conjugated	Millipore	Cat# AP307P; RRID:AB_92641
Bacterial and virus strains		
Stellar™ Competent Cells	Takara	Cat# 636766
Critical commercial assays		
Power SYBR Green PCR Master Mix	Applied Biosystems	Cat# 4368706
Superscript III Reverse Transcriptase kit	Invitrogen	Cat# 18080093
TRIzol reagent	Invitrogen	Cat# 15596–018
mMessage mMachine SP6 Transcription Kit	ThermoFisher	Cat# AM1340
Nano-Glo® Dual-Luciferase Reporter Assay System	Promega	Cat# N1610
Deposited data		
mRNA-sequencing of Upf1, Smg6 and G3bp1 siRNA knockdown in HEK293 cells	This manuscript	SRA: SRP495209
3xFlag-Upf proteins and control pulldowns of 3′UTR MPRA library	This manuscript	SRA: SRP495203
Experimental models: Organisms/strains		
Zebrafish: TU-AB and TLF strains	Zebrafish International Resource Center	N/A
Oligonucleotides		
Table S5	This manuscript	N/A
Recombinant DNA		
pCS2+3xFlag-Upf1	This study	N/A
pCS2+3xFlag-Upf1-R843C	This study	N/A

REAGENT or RESOURCE	SOURCE	IDENTIFIER
pCS2+3xFlag-Upf1- C126S	This study	N/A
pCS2+3xFlag-Upf1-G495R/G497E	This study	N/A
pCS2+3xFlag-Upf2	This study	N/A
pCS2+3xFlag-Upf3a	This study	N/A
pCS2+3xFlag-Upf3b	This study	N/A
pCS2+3xFlag-GFP	This study	N/A
pCS2+3xFlag-Celf1	This study	N/A
pCS2+1x-300bp-ORF-Optimal	This study	N/A
pCS2+6x-1800bp-ORF-Optimal	This study	N/A
pCS2+1x-300bp-ORF-Non-optimal	This study	N/A
pCS2+6x-1800bp-ORF-Non-optimal	This study	N/A
pCS2+1x-600bp-ORF-Strong-Kozak	This study	N/A
pCS2+1x-600bp-ORF-Medium-Kozak	This study	N/A
pCS2+1x-600bp-ORF-Weak-Kozak	This study	N/A
pCS2+3x-1800bp-ORF-Strong-Kozak	This study	N/A
pCS2+1x-600bp-ORF-Strong-Kozak-Nano-Luciferase	This study	N/A
pCS2+1x-600bp-ORF-Medium-Kozak-Nano-Luciferase	This study	N/A
pCS2+1x-600bp-ORF-Weak-Kozak-Nano-Luciferase	This study	N/A
pCS2+3x-1800bp-ORF-Strong-Kozak-Nano-Luciferase	This study	N/A
pCS2+1x-titin-ORF-P2A-Nano-Luciferase	This study	N/A
pCS2+4x-titin-ORF-P2A-Nano-Luciferase	This study	N/A
pCS2+firefly-Luciferase	This study	N/A
pCS2+1x-AUG-less-600bp-ORF	This study	N/A
pCS2+3x-AUG-less-1800bp-ORF	This study	N/A
pCS2+1x-AUG-600bp-ORF	This study	N/A
pCS2+3x-AUG-1800bp-ORF	This study	N/A
pCS2+3x-AUG-1800bp-ORF-long3′UTR	This study	N/A
uORF/oORF reporters	Johnstone et al. [7]	N/A
Software and algorithms		
Python (data analysis)	Python 3.11.6	https://www.python.org

REAGENT or RESOURCE	SOURCE	IDENTIFIER
R (data analysis)	R 4.3.2	https://www.r-project.org
Fiji ImageJ	Fiji 2.15.0	https://imagej.net/software/fiji/