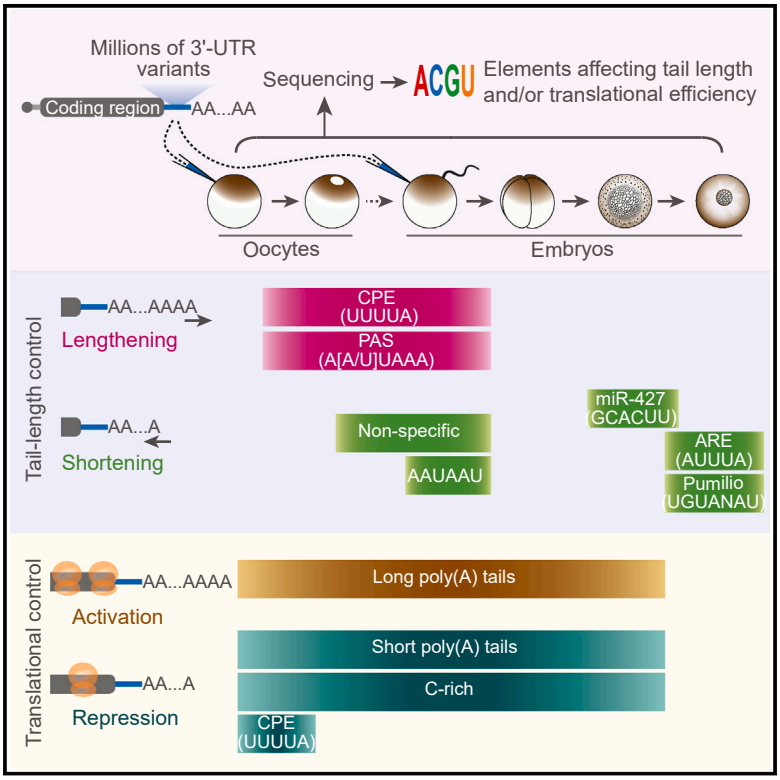
Article

Developmental Cell

Control of poly(A)-tail length and translation in vertebrate oocytes and early embryos

Graphical abstract



Highlights

- UUUUA, influenced by contextual features, specifies most cytoplasmic polyadenylation
- Stage-specific 3′ UTR elements drive waves of tail-length shortening in embryos
- UUUUA and C-rich motifs can direct tail-length-independent translational repression
- Tail-length control is conserved in oocytes of frogs, mice, and humans

Authors

Kehui Xiang, Jimmy Ly, David P. Bartel

Correspondence

dbartel@wi.mit.edu

In brief

Xiang et al. analyze millions of different mRNA reporters to identify 3′ UTR sequence elements and their contextual features that control poly(A)-tail length and translation during early development of frogs and fish. These insights provide a unified paradigm for the translational control of early development of vertebrate animals.

Xiang et al., 2024, Developmental Cell *59*, 1058–1074
April 22, 2024
https://doi.org/10.1016/j.devcel.2024.02.007

CellPress

CellPress
OPEN ACCESS

Developmental Cell

Article

Control of poly(A)-tail length and translation in vertebrate oocytes and early embryos

Kehui Xiang,[1,2,3] Jimmy Ly,[2,3] and David P. Bartel[1,2,3,4,*]
[1]Howard Hughes Medical Institute, Cambridge, MA 02142, USA
[2]Whitehead Institute for Biomedical Research, Cambridge, MA 02142, USA
[3]Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[4]Lead contact
*Correspondence: dbartel@wi.mit.edu
https://doi.org/10.1016/j.devcel.2024.02.007

SUMMARY

During oocyte maturation and early embryogenesis, changes in mRNA poly(A)-tail lengths strongly influence translation, but how these tail-length changes are orchestrated has been unclear. Here, we performed tail-length and translational profiling of mRNA reporter libraries (each with millions of 3′ UTR sequence variants) in frog oocytes and embryos and in fish embryos. Contrasting to previously proposed cytoplasmic polyadenylation elements (CPEs), we found that a shorter element, UUUUA, together with the polyadenylation signal (PAS), specify cytoplasmic polyadenylation, and we identified contextual features that modulate the activity of both elements. In maturing oocytes, this tail lengthening occurs against a backdrop of global deadenylation and the action of C-rich elements that specify tail-length-independent translational repression. In embryos, cytoplasmic polyadenylation becomes more permissive, and additional elements specify waves of stage-specific deadenylation. Together, these findings largely explain the complex tapestry of tail-length changes observed in early frog and fish development, with strong evidence of conservation in both mice and humans.

INTRODUCTION

In vertebrate animals, transcription ceases when oocytes become fully grown and stay quiescent throughout oocyte maturation and early embryogenesis.[1–3] To progress through these transcriptionally silent developmental stages, oocytes and early embryos rely on post-transcriptional control of a stockpile of maternally deposited mRNAs. In zebrafish and frogs, most of the maternally deposited mRNAs are stable until zygotic transcription ensues during the mid-blastula transition (MBT).[4,5] Precise temporal regulation of translation from these mRNAs is crucial for the oocyte-to-embryo transition and early embryogenesis.[6–8]

Much of this regulation occurs through changes in mRNA poly(A)-tail lengths. In oocytes and early embryos of flies, fish, frogs, and mice, mRNA translational efficiency (TE) and poly(A)-tail length are strongly coupled, such that mRNAs with long poly(A) tails are translated much more efficiently than those with short tails.[9–12] In frog prophase I-arrested oocytes, most mRNAs have short poly(A) tails and are translationally repressed.[13] As these oocytes mature and complete meiosis I, a group of mRNAs, including *mos* and *ccnb1*, which encode proteins essential for meiosis re-entry, undergo cytoplasmic polyadenylation.[14,15] This tail lengthening causes increased translation, thereby promoting oocyte maturation,[16–18] a process conserved in mammals.[19–21]

Cytoplasmic polyadenylation requires two sequence elements within mRNA 3′ UTRs: a polyadenylation signal (PAS) and a cytoplasmic polyadenylation element (CPE).[22,23] The PAS has the consensus sequence AAUAAA (top alternative variant, AUUAAA)—the same motif used for nuclear 3′ end cleavage and polyadenylation. It is recognized by the cleavage and polyadenylation specificity factor (CPSF), a protein complex shared with nuclear polyadenylation.[24] The CPE is recognized by the CPE-binding protein (CPEB1).[25] The most common CPEs are reported to be UUUUAU and UUUUAAU, although additional, non-consensus elements (UUUUACU, UUUUCAU, UUUUCCA, UUUUAAAU, and UUUUAAGU) have also been implicated.[21,23] Nonetheless, hundreds of mRNAs that lack any reported CPE undergo cytoplasmic polyadenylation during oocyte maturation, which suggests the activity of one or more yet-to-be-identified elements that direct tail lengthening. Although several RNA elements, including the Musashi-binding element (MBE),[26] the translational control sequence (TCS),[27] and the Pumilio-binding element (PBE)[28] are reported to contribute to cytoplasmic polyadenylation of specific mRNAs, whether these effects are direct, independent of CPE, and global are unclear.

The reported CPEs are each identified from only a few endogenous mRNAs, without the benefit of transcriptome-wide analyses. Recent studies have used high-throughput sequencing methods to measure mRNA poly(A)-tail lengths transcriptome-wide during oocyte maturation in frogs, mice, and humans.[29–32]

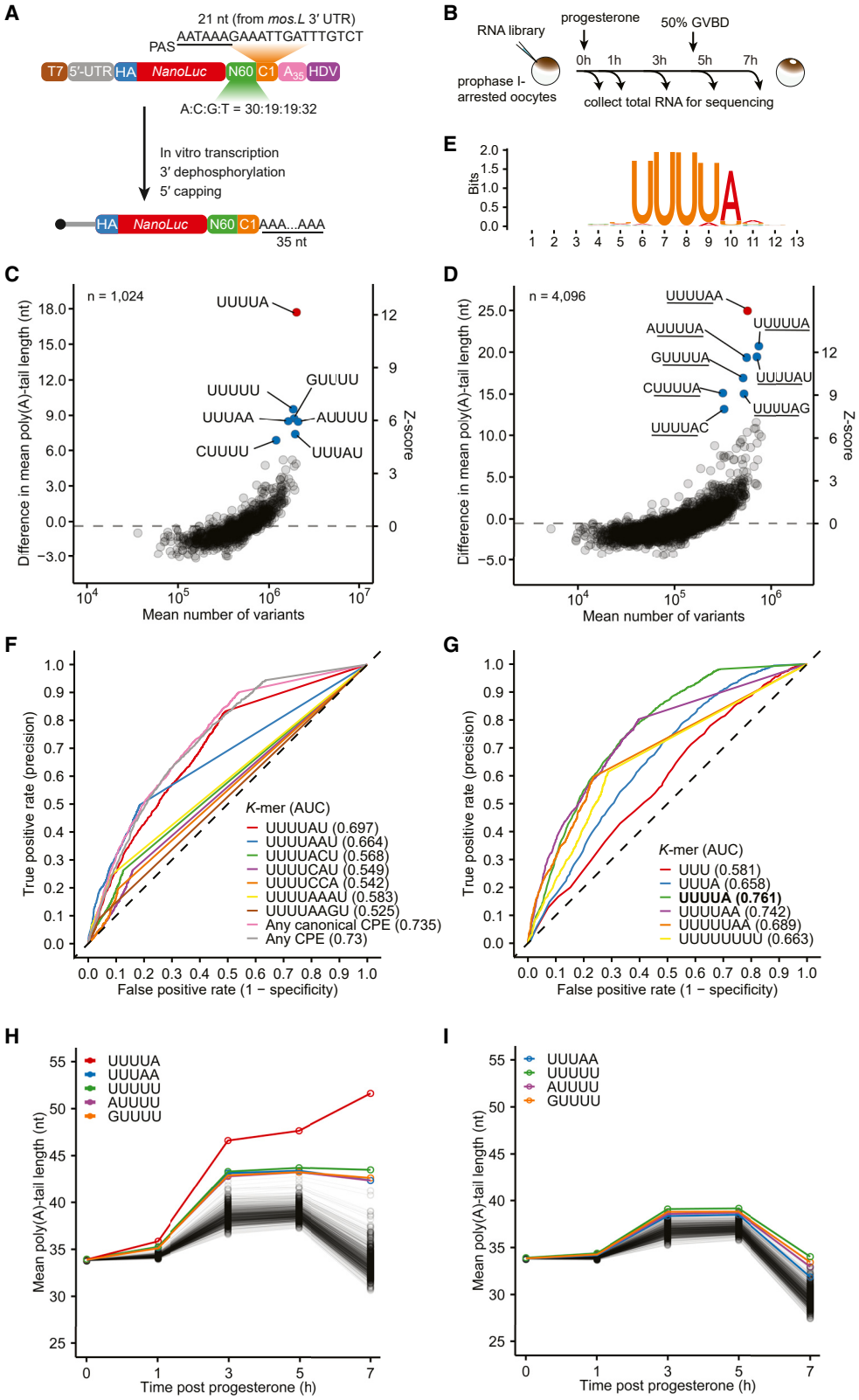


Figure 1. A core CPE controls cytoplasmic polyadenylation in frog oocytes

(A) The N60-PAS^{mos} mRNA library.

(B) Experimental scheme for mRNA library injection and sample collection.

(legend continued on next page)

However, little progress has been made to define sequence elements that modulate tail lengths or to quantitively determine the effects of such elements, in part due to (1) the limited sequence space represented by a few thousand endogenous mRNAs, and (2) difficulty in designating causal sequence elements from many co-occurring sequences within 3′ UTRs.

Cytoplasmic polyadenylation is also essential for early embryogenesis,[33,34] but the primary sequence elements that drive tail lengthening in the embryo are less well understood. In frog embryos, a few mRNAs are reported to employ elements that differ from the CPE used in oocytes, including a proposed embryonic CPE[35] and C-rich elements,[36] but the prevalence of these elements in tail lengthening of the transcriptome is unknown. Besides polyadenylation, mRNAs of early embryos are also subject to deadenylation through elements recognized by RNA-binding proteins or microRNAs.[37–41] Indeed, in fly embryos, poly(A)-tail lengthening occurs broadly, with specificity primarily imposed by tail shortening.[10,42] In flies, fish, and frogs, tail shortening is essential for translational repression and then clearance of most maternal mRNAs upon onset of zygotic transcription.[39–41] Nevertheless, how embryos control poly(A)-tail lengthening and shortening of different mRNAs in successive developmental stages is not well understood.

Here, we generated diverse mRNA libraries containing millions of 3′ UTR sequence variants and examined tail lengths and translation of library molecules after injection into developing oocytes or embryos. Analyses of these data provided insights into tail length and translation control in oocytes and early embryos.

RESULTS

A core CPE controls cytoplasmic polyadenylation

To acquire information useful for identifying the 3′ UTR features that specify cytoplasmic polyadenylation, we generated a large library of mRNA variants (Figure 1A), injected this mRNA library into prophase I-arrested frog oocytes, and both sequenced the 3′ UTR and measured the poly(A)-tail length for individual variants over the course of oocyte maturation (Figure 1B). Each member of the mRNA library contained the NanoLuc coding region, followed by a 3′ UTR containing a 60-nt random-sequence region, designed to match the mean nucleotide composition of frog 3′ UTRs (30% A, 19% C, 19% G, and 32% U). This random-sequence region was followed by a segment from the last 21 nt of *Xenopus laevis mos.L* 3′ UTR, which contained the PAS and was followed by a 35-nt poly(A) tail. We call this mRNA library N60-PAS^{mos}.

Analysis of these library molecules had several advantages over analysis of endogenous mRNAs. First, for each library molecule, the variable sequence was within a uniform context, which avoided potentially confounding effects of different mRNA lengths, different initial poly(A)-tail lengths, different 5′ UTR or coding sequences, or different nuclear or cytoplasmic histories. Second, a larger number of different mRNAs could be examined, with this number limited by sequencing depth (tens of millions) rather than by the number of unique expressed mRNAs (tens of thousands). Third, the use of reporters that presented each *k*-mer (sequence motif of length *k*) within a diversity of non-biological sequences enabled causal relationships between *k*-mers and tail length to be confidently established. In contrast, biological mRNAs contain *k*-mers that tend to co-occur with each other, which can confound causality. For example, if a motif that specified mRNA localization and a motif that specified polyadenylation tended to co-occur with each other within natural 3′ UTRs, both would correlate with extension of natural mRNA tails, even though only one functioned to specify polyadenylation.

As oocytes underwent maturation, the median tail length of the library gradually lengthened until after 5 h, the time of germinal vesical breakdown (GVBD) followed by meiosis I, after which point the median tail length substantially shortened (Figure S1A). By 7 h post-progesterone treatment, 30% of mRNAs had tails <20 nt (Figure S1A), concurring with the observation that short-tailed mRNAs are stable in maturing frog oocytes.[43,44]

To identify sequence motifs that mediated cytoplasmic polyadenylation, we evaluated all 5-mers, determining for each the average tail length of variants containing that 5-mer—with and without progesterone treatment. The greatest tail-length change was observed for the UUUUA 5-mer; variants with this motif had average tail lengths 17.7 nt longer in the matured oocytes compared with the uninjected sample (Figure 1C). In contrast, the next six strongest 5-mers were associated with weaker tail lengthening (6.9–9.5 nt). Moreover, they each resembled a partial UUUUA that if extended by a U on the 5′ terminus or an A on the 3′ terminus would form the full UUUUA.

Analyses of shorter and longer *k*-mers confirmed the strong association of UUUUA with tail lengthening. When all 4-mers were examined, UUUU and UUUA—the two 4-mers that constitute UUUUA, separated from the others (Figure S1B). Similarly, among all 6-mers, the eight that were extensions

(C) Tail-length changes associated with each 5-mer in the 3′ UTRs of the N60-PAS^{mos} library, comparing between 0 and 7 h post-progesterone treatment. Plotted for each 5-mer are differences in mean tail lengths observed for mRNAs with that 5-mer.
(D) As in (C), but for 6-mers.
(E) Sequence logo generated from the 180 8-mers most strongly associated with tail lengthening in the N60-PAS^{mos} library, comparing between 0 and 7 h post-progesterone treatment.
(F) ROC curves testing the ability of previously proposed CPEs to classify endogenous mRNAs as subject to cytoplasmic polyadenylation during frog oocyte maturation. A curve is shown for each of the indicated CPEs, including the two canonical CPEs (UUUUAU and UUUUAAU), as well as for the indicated combinations of elements.
(G) ROC curves testing the ability of the best 3-, 4-, 5-, 6-, 7-, and 8-mers to classify endogenous mRNAs as subject to cytoplasmic polyadenylation during frog oocyte maturation.
(H) The effects of 5-mers on tail lengths of mRNAs in the N60-PAS^{mos} library during frog oocyte maturation. Plotted for each 5-mer are the mean tail lengths of mRNAs containing that 5-mer.
(I) As in (H), except mRNAs that contained UUUUA were excluded.
See also Figure S1.

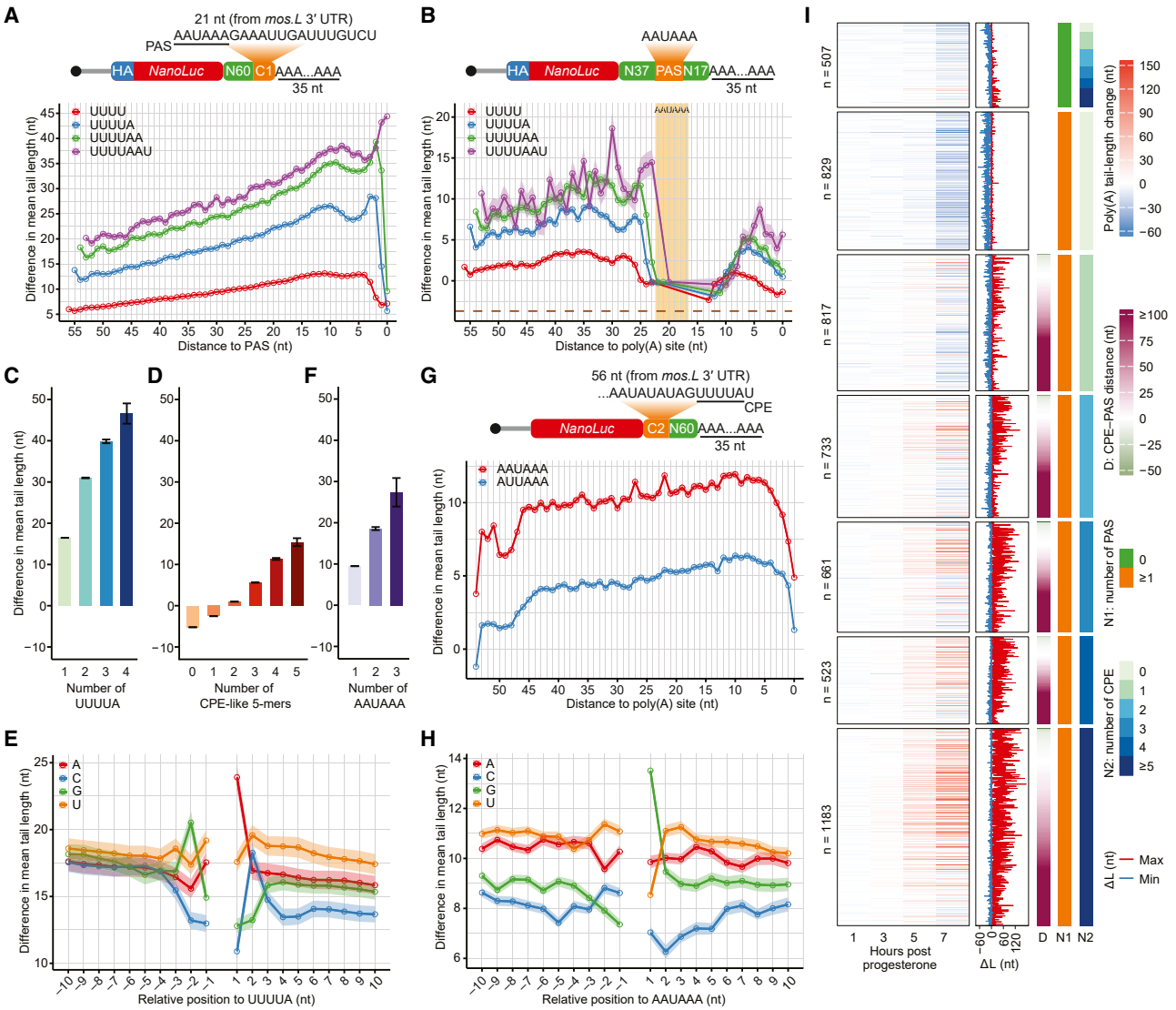


Figure 2. Context of the CPE and the PAS influences cytoplasmic polyadenylation in frog oocytes

(A) Effect of the distance between the CPE and the PAS. Plotted for each of the four motifs is the difference in mean tail length observed for mRNAs of the N60-PAS^{mos} library (schematic at top) containing that motif at each position along the variable region of the 3′ UTR, comparing between 0 and 7 h post-progesterone treatment. Shaded areas along lines indicate standard error of the difference between means.

(B) Effect of the relative position of the CPE and the PAS. Plotted for each of the four motifs is the difference in mean tail length observed for mRNAs of the N37-PAS-N17 library containing that motif at each position along the 3′ UTR. The shaded region marks the PAS. The dashed brown line indicates the difference in mean tail length of all mRNAs in the library. Otherwise, this panel is as in (A).

(C) Influence of the number of the CPEs. Plotted is the difference in mean tail length of mRNAs in the N60-PAS^{mos} library containing the indicated numbers of CPEs, comparing between 0 and 7 h post-progesterone treatment. Error bars indicate standard error of the difference between means.

(D) Influence of the number of CPE-like motifs. This panel is as (C), but for CPE-like motifs, UUUUU, UGUUU, GUUUU, UUUGU, and AUUUU, which were the top five motifs associated with tail lengthening, after excluding all variants containing UUUUA.

(E) Effect of CPE flanking nucleotides. Plotted is the difference in mean tail length observed for mRNAs in the N60-PAS^{mos} library containing one CPE and the indicated nucleotide at the indicated position, comparing between 0 and 7 h post-progesterone treatment. Shaded areas along the lines indicate standard error of the difference between means.

(F) Influence of the number of PASs. Plotted is the difference in mean tail length of mRNAs in the CPE^{mos}-N60 library containing the indicated numbers of PASs, comparing between 0 and 5 h post-progesterone treatment; otherwise, as in (C).

(G) Effect of the distance between the PAS and the mRNA 3′ end. Plotted is the difference in mean tail length of mRNAs in the CPE^{mos}-N60 library containing the indicated PAS element at each position along the variable region of the 3′ UTR; otherwise as in (A).

(H) Effect of PAS flanking nucleotides. Plotted is the difference in mean tail length observed for mRNAs in the CPE^{mos}-N60 library containing one PAS (AAUAAA) and the indicated nucleotide at the indicated position relative to the PAS, comparing between 0 and 5 h post-progesterone treatment; otherwise, as in (E).

(I) Association of 3′ UTR sequence features with tail-length changes of endogenous mRNAs of frog oocytes. The heatmaps on the left compare median tail lengths of frog oocyte mRNAs collected at the indicated time after progesterone treatment with those in oocytes not treated with progesterone, with each row

(legend continued on next page)

of UUUUA were associated with the longest tail-length differences (Figure 1D).

These results indicated that UUUUA is the CPE of frog oocytes. This conclusion was somewhat different than previous results, in that previously identified CPEs (UUUUAU, UUUUAAU, UUUUACU, UUUUCAU, UUUUCCA, UUUUAAAU, and UUUUAAGU) are all longer than 5 nt, such that those that contain UUUUA are each extended by one or more additional nucleotides. To eliminate the spillover signal from a stronger motif to a weaker motif, we re-examined the effects of 6-mers by iteratively removing all mRNAs containing the 6-mer with the largest association with tail lengthening and then re-calculating the average tail lengths of all remaining 6-mers. After excluding the top two 6-mers (UUUUAA and UUUUAU), other 6-mer motifs, including UUUUUA, AUUUUA, UUUUAG, and UUUUAC, which would not have previously been recognized as a CPE, were associated with substantial tail lengthening (Figure S1C), thereby supporting the designation of the shorter, UUUUA motif as the CPE. Moreover, a search for sequence logos from 8-mers associated with elongated tails in matured oocytes identified only one consensus logo, which matched the UUUUA motif, with little information contributed by flanking nucleotides, further supporting the conclusion that UUUUA is the CPE of frog oocytes (Figure 1E). The close match between the UUUUA CPE and the binding site for human CPEB1 *in vitro*[45] and for fly CPEB1 *in vivo*[46] suggested that these results extend to diverse species and further suggested that binding of CPEB1 protein confers the specificity of CPE recognition without help from additional factors.

To examine if UUUUA also acts in the context of endogenous mRNAs, we analyzed endogenous mRNAs collected alongside the library during frog oocyte maturation. Globally, tails of these mRNAs followed a trend resembling that of the injected mRNA library, with median lengths increasing before GVBD and decreasing afterward (Figure S1D). Because *X. laevis* 3′ UTRs were poorly annotated, we used our sequencing data to identify mRNA poly(A) sites and thereby infer 3′ UTRs. With these improved annotations and their median tail-length values, we evaluated the performance of different *k*-mers in classifying endogenous mRNA substrates for cytoplasmic polyadenylation using receiver operating characteristic (ROC) curves. UUUUA outperformed each of the previously reported CPEs (Figures 1F and 1G). Moreover, it had greater area under the curve (AUC) than any combination of previously reported CPEs (Figures 1F and 1G). Indeed, among all *k*-mers with lengths of 3–8 nt, the UUUUA CPE had the greatest AUC value (Figure 1G), regardless of the threshold used for categorizing true positives (Figure S1E).

Because cytoplasmic polyadenylation occurs in different phases of frog oocyte maturation,[28] we returned to analysis of library molecules to examine the dynamics of tail-length changes. When plotting mean tail lengths of mRNAs containing each of the 5-mers during the course of maturation, the CPE was consistently associated with the longest tails (Figure 1H). From 1 to 3 h, CPE-dependent tail lengthening was accompanied by a modest overall increase in tail lengths, whereas after 5 h (upon GVBD), CPE-dependent tail lengthening opposed an overall decrease in tail lengths. This latter behavior concurred with reports that endogenous mRNAs lacking a CPE undergo non-sequence-specific deadenylation in mature oocytes.[47,48]

During 1–3 h, some U-rich motifs were also associated with modest tail lengthening that did not fully abate when all CPE-containing variants were excluded, which indicated that some preferential tail lengthening could be achieved without a perfect match to the CPE (Figure 1I). These U-rich motifs each matched four contiguous nucleotides of the UUUUA CPE, which suggested that their association might have been due to the promiscuous binding of CPEB1 to CPE-like motifs. Nonetheless, these degenerate CPEs were typically insufficient to overcome the global deadenylation occurring during the 5–7 h time interval, such that after GVBD, the UUUUA CPE was the only motif to robustly support net polyadenylation among 5-mers (Figure 1I) or even 8-mers (Figure S1F).

Despite distinct global tail-length trajectories before and after GVBD, motif-associated changes in early (1–3 h) and late (5–7 h) stages of polyadenylation were highly correlated (Figure S1G), suggesting that the sequence-specific component of tail-length regulation is consistent throughout frog oocyte maturation. Moreover, no 5-mers had evidence of maturation-stage-specific synergistic effects with the CPE (Figure S1H). Notably, for 3′ UTRs without a CPE, no substantial lengthening was associated with the MBE, the TCS, or the PBE, indicating that these elements do not mediate cytoplasmic polyadenylaiton (Figure S1F). Perhaps the previous observations are attributable to effects on CPE context within the previously analyzed mRNAs.[49]

Together, our results indicate that sequence-specific tail-length control during frog oocyte maturation is primarily mediated by cytoplasmic polyadenylation that requires a CPE: UUUUA. This specific tail lengthening occurs against a backdrop of non-sequence-specific deadenylation after GVBD.

CPE and PAS contexts explain differences in cytoplasmic polyadenylation

Although the average tail length of CPE-containing library variants increased by 17.7 nt in matured oocytes, tail lengths for individual variants varied substantially (Figure S2A). We thus sought to identify contextual features that contributed to this variation.

When examining the CPE position, the magnitude of cytoplasmic polyadenylation increased as the CPE approached the PAS (Figure 2A). This relationship was almost linear until the distance between the CPE and the PAS reached approximately 6 nt, at which point it dipped slightly, before dramatically declining at 2 nt. We attribute the dip to inhibitory pairing between the CPE and the PAS, and the dramatic decline to a steric clash between CPEB1 and CPSF when the CPE and the PAS

representing a unique 3′ UTR of an mRNA with a defined poly(A) site. Also indicted for each of these UTRs is the CPE-PAS distance and the minimal and maximal tail-length change over the 7 h treatment. UTRs are grouped based on the presence of a canonical PAS (within 150 nt of the 3′ end) and the number of CPEs (within 1,000 nt of the 3′ end). Only UTRs with poly(A) sites that had more than 50 poly(A) tags in all datasets were analyzed. For UTRs that contained more than one PAS, the one closest to the 3′ end was used. For UTRs that contained more than one CPE, the one closest to the PAS was used to calculate the CPE-PAS distance. See also Figure S2.

were too close. Similar positional relationships with polyadenylation were observed for different-lengthed *k*-mers with partial or extended CPEs (Figure 2A) and for many CPE-like motifs (Figure S2B). This observed effect of CPE-PAS proximity was somewhat at odds with a report that varying the distance between the CPE and the PAS within a 26 nt range does not affect cytoplasmic polyadenylation,[50] a difference attributable to the lower resolution of the previous study, which examined far fewer mRNAs.

To interrogate the positional effect of the CPE downstream of the PAS, we repeated our experiments with another mRNA library referred to as N37-PAS-N17, in which the PAS was flanked by random-sequence regions on both sides (Figure 2B). Upstream of the PAS, the CPE had a positional effect resembling that observed for the N60-PAS^{mos} library (Figure 2B), which indicated that the interplay between these two elements is not restricted to mRNAs modeled from frog *mos.L* mRNA. Downstream of the PAS, the extent of polyadenylation gradually increased until the distance between the PAS and the CPE reached 6 nt, after which it started to decline (Figure 2B). Thus, the relative position of the CPE on either side of the PAS influences cytoplasmic polyadenylation.

The number of the CPEs within the 3′ UTR also impacted cytoplasmic polyadenylation, with more CPEs leading to larger tail-length increases (Figure 2C). In most scenarios, two CPEs on the same mRNA had less effect on polyadenylation than the summed effect of individual CPEs at the same positions, the exceptions being when one CPE was very close to the PAS (<2 nt) or when both CPEs were far away from the PAS (>30 nt), in which case the two CPEs appeared to act synergistically (Figure S2C). For variants without any CPE, multiple copies of CPE-like motifs supported some polyadenylation, but approximately five CPE-like motifs were required to approach the tail-length extension imparted by a single CPE (Figures 2C, 2D, and S2D).

With respect to sequence context, a U-rich context promoted stronger polyadenylation (Figure 2E), presumably due to secondary CPE-like motifs, which are generally U rich (Figure S2B). By contrast, a C-rich context promoted weaker polyadenylation (Figure 2E). The nucleotide identity at −2, −1, +1, and +2 positions had the strongest average impact on polyadenylation, with polyadenylation enhanced by A and U at both the −1 and +1 positions, G at the −2 position, and C as well as U at the +2 position. The impact of these nucleotides at positions flanking the CPE explained why some of the longer motifs were originally designated as CPEs. Nevertheless, the contributions to polyadenylation by these flanking nucleotides were much weaker than those by the first and the last nucleotides of the 5-mer CPE (Figures S2E and S2F), which supported designation of the 5-mer as the CPE.

The flanking dinucleotide effects on polyadenylation could not be explained by the predicted structural accessibility of the CPE, calculated as the predicted likelihood of the CPE not pairing to other nucleotides within each mRNA (Figure S2G). Nonetheless, structural accessibility can modulate cytoplasmic polyadenylation for select frog mRNAs,[49] and we found that tail length modestly correlated with the predicted minimum free energy of 3′ UTR folding even after GC-content was matched (Figure S2H), implying that 3′ UTR structure might negatively impact CPE or PAS recognition.

Next, we assessed the function of the PAS and its contextual features during cytoplasmic polyadenylation, injecting a third reporter library, designated CPE^{mos}-N60, in which the 3′ UTR of each mRNA had a 56-nt fragment from *mos.L* that contained a CPE but not a PAS, followed by a 60-nt random-sequence region. The 6-mer from the random-sequence region associated with the strongest tail lengthening was the canonical PAS, AAUAAA, and the next six were AUUAAA and five other near matches to the canonical element (Figure S2I). Interestingly, more PAS elements resulted in greater tail-length increases (Figure 2F). Moreover, for a single PAS element, the magnitude of cytoplasmic polyadenylation increased as the distance between the PAS and the 3′ end became shorter, until ∼5 nt, at which point it began to drop (Figure 2G). This positional effect occurring for distances of 45–5 nt defied the diminished lengthening effect expected due to the increased distance between the CPE and the PAS (Figure 2A), indicating that the distance between the PAS and the 3′ end has an even stronger influence than the distance between the PAS and the CPE.

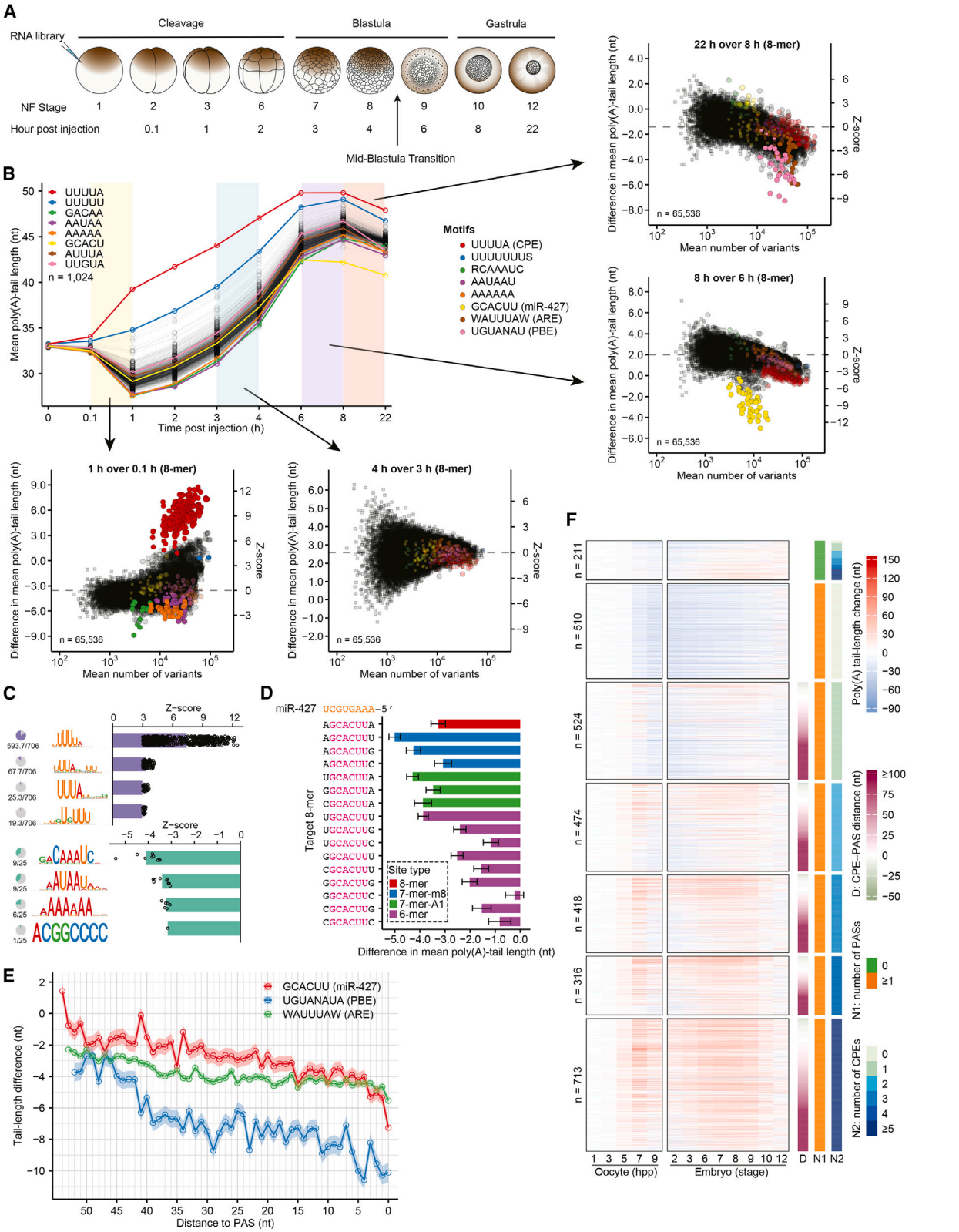
The flanking nucleotides of the PAS also impacted cytoplasmic polyadenylation. At most flanking positions, a U or A enhanced polyadenylation (Figure 2H). The main exception was at position +1, where a G strongly increased polyadenylation, whereas a C or U was more detrimental than at most other positions (Figure 2H).

To examine if principles elucidated from analyses of mRNA reporter libraries also applied to endogenous mRNAs, we evaluated tail-length changes of endogenous mRNAs during frog oocyte maturation as a function of the number of CPEs, the presence of a PAS, and the distance between the CPE and the most proximal PAS (Figure 2I). Consistent with the results from analyses of reporters, tail lengthening correlated with more CPEs, the presence of a PAS, and CPE-PAS proximity. For 3′ UTRs with only one or two CPEs, tail lengthening was minimal if these CPEs were distant from the PAS (≥100 nt). Interestingly, mRNAs that had no canonical PAS (neither AAUAAA nor AUUAAA) within the last 150 nt of their 3′ UTR but did have ≥5 CPEs frequently underwent detectible tail lengthening, albeit to a lesser extent than those that had a PAS, which implied that binding of multiple CPEB1s can partially compensate for weak association between CPSF and a non-canonical PAS.

In summary, cytoplasmic polyadenylation in frog oocytes is influenced by contextual features, including nucleotides flanking the CPE and the PAS, the number of CPEs and PASs, the CPE-PAS distance, the PAS-poly(A) distance, and to a lesser extent, the structural accessibility.

The CPE and deadenylation elements successively control tail length during frog early embryogenesis

To extend our investigation beyond oocytes to early embryogenesis, we injected our N60-PAS^{mos} and N37-PAS-N17 libraries into *in-vitro*-fertilized frog eggs at the one-cell stage (stage 1) and harvested total-RNA samples at successive developmental stages until gastrulation (stage 12) (Figure 3A). We did the same for a third library, referred to as N60, in which the 3′ UTR was composed of 60 random-sequence nucleotides (Figure S3C). For each library, we calculated the average tail lengths of mRNAs with each 5-mer at each sampled stage (Figures 3B, S3D, and S3E). When comparing adjacent stages, we also calculated



(legend on next page)

changes in mean tail lengths of mRNAs with each 8-mer (Figure 3B).

In very early embryogenesis (within 1 h post-injection [hpi]), results resembled those observed during late stages of oocyte maturation (Figure 1H), suggesting that early embryos inherit the tail-length regulatory regime observed in matured eggs. In this period, median tail lengths rapidly shortened, indicating that deadenylation was the default (Figures S3A–S3C). Nonetheless, this tail shortening could be countered by the PAS and either a CPE or long stretch of U (Figures 3B, 3C, and S3D–S3H). Although U_{12} is reported to act as an embryonic CPE,[35] it did not appear to promote polyadenylation as effectively as the CPE (Figure S3I).

We also found several sequence motifs associated with greater-than-average tail-length shortening (Figures 3B, 3C, and S3F–S3H). One of them, AAUAAU, was identified in analyses of all three libraries. This element, although rich in A and U, differed from the canonical AU-rich element (ARE) AUUUA known to promote deadenylation at later stages[37] (Figure 3B), implying that an unknown protein or mechanism is involved.

During late-cleavage and early-blastula stages (1–4 hpi; stages 3–8), the median tail length of the N60-PAS^{mos} library gradually increased, whereas that of the N37-PAS-N17 and the N60 libraries did not (Figures S3A–S3C). This increase was attributed to CPE-like motifs within the constant region of the N60-PAS^{mos} library (Figure S3J), an effect not observed in maturing oocytes, implying that either global deadenylation became less prevalent at the embryonic stages or cytoplasmic polyadenylation became more prevalent. During this period, no motifs other than the CPE were strongly associated with tail-length changes in any of the three libraries (Figures 3B, S3D, and S3E). The deadenylation occurring before and during this period resulted in substantial accumulation of tail-less mRNAs in the N37-PAS-N17 and the N60 libraries (Figures S3B and S3C), which concurred with the report that poly(A) tails are not required for mRNA stability during frog early embryonic development.[37] In contrast, these deadenylated mRNAs and other short-tailed mRNAs appeared to be degraded after 4 hpi (stage 8), causing large increases in the median tail lengths of all three libraries (Figures 3B and S3A–S3E). These results reflected the shift in the function of poly(A)-tail length—from enhancing translation efficiency to enhancing mRNA stability—during frog embryonic development.[9]

During and immediately after the MBT (4–8 hpi; stages 8–10), mRNAs that contained miR-427 sites were specifically deadenylated, which concurred with accumulation of miR-427 after zygotic genome activation.[41] The effects of different miR-427 site types and flanking nucleotides corresponded well to relative binding affinities measured for microRNA–Ago2 complexes and their targets (Figure 3D).[52] The miR-427 sites also tended to be more effective when located closer to the mRNA 3′ ends (Figure 3E).

In gastrulating embryos, miR-427-associated deadenylation dampened, giving way to preferential deadenylation of mRNAs that contained either an ARE or a PBE (Figure 3B), both of which also had 3′ end positional efficacy resembling that of miR-427 target sites (Figure 3E). These element-specific deadenylation activities coincided with increased ARE-binding protein Zfp36 (TTP in mammals) and Pumilio proteins,[53] the latter of which are essential for embryonic development in flies[54] and mice,[55] but their roles during gastrulation of frog embryos had not been reported.

To investigate whether our library-elucidated sequence elements play similar roles in endogenous mRNAs, we measured endogenous tail lengths at embryonic stages at which our injected libraries had been sampled. At stages when transcription was still silent (stages 2–8), tail-length extension resembled that observed in matured oocytes, although dependence on the PAS was more relaxed, with many mRNAs containing $\geq$2 CPEs but no canonical PAS undergoing tail lengthening (Figure 3F). Most mRNAs with tails that had been extended earlier, during oocyte maturation, either maintained their long tails or underwent further tail lengthening (Figure 3F). Overall, mRNA tail lengths at each stage during this period of embryonic development correlated well with each other (Pearson correlation coefficient, R_p = 0.82–0.98, Figure S3K) and also well with those in matured oocytes (9 h post progesterone, R_p = 0.7–0.86, Figure S3K).

Effects of individual elements on endogenous mRNA tail lengths were difficult to detect due to different starting tail lengths and various confounding motifs within 3′ UTRs of

Figure 3. The CPE and deadenylation elements successively control tail length during frog early embryogenesis

(A) Experimental scheme for mRNA library injection and sample collection. NF stage: developmental stage according to Nieuwkoop and Faber.[51]

(B) Effects of different 3′ UTR sequence motifs on tail lengths of the N60-PAS^{mos} library during frog early embryogenesis. Plotted at the top left, for each 5-mer over the course of early development, are mean tail lengths of library mRNAs containing that 5-mer. Shaded areas highlight the stages between which tail-length differences are assessed in the peripheral scatter plots. Each scatter plot shows mean tail-length changes associated with each 8-mer in the 3′ UTRs of library mRNAs, plotted as in Figure 1C, except circles represent 8-mers with significant differences in mean tail lengths (Welch's t test against the global average, adjusted $p < 0.01$), and squares represent 8-mers without significant differences. Circles and squares representing notable motifs are colored (key, S represents C or G; R represents A or G; W represents A or U; N represents A, C, G, or U).

(C) Sequence motifs associated with tail-length changes during early frog embryogenesis. This panel is as in Figure 1E but examines embryos between 0.1 and 1 hpi, with each pie chart (left) indicating the fraction of 8-mers aligning to the sequence logo and each bar (right) showing the mean Z score, with points representing scores of individual 8-mers.

(D) Effects of nucleotides flanking the miR-427 seed match on deadenylation. Shown are mean tail-length changes of mRNAs in the N60-PAS^{mos} library for 16 different 8-mer sites, each centered on the 6-mer miR-427 seed match (pink), comparing between 6 and 8 hpi (stages 9 and 10, respectively). Results for 8-mers that match canonical 7- and 8-nt target sites are colored (key). Error bars indicate standard error of the difference between means.

(E) Positional effects of elements that promote deadenylation. Plotted for each element are the differences in mean tail length of mRNAs in the N60-PAS^{mos} library containing the element at the indicated distances to the PAS, comparing between 6 and 8 hpi (stages 9 and 10, respectively) for the miR-427 target site and between 8 and 22 hpi (stages 10 and 12, respectively) for the PBE and ARE; otherwise, as in Figure 2A.

(F) Effect of 3′ UTR sequence features on tail-length changes for frog mRNAs in oocytes and early embryos. The heatmaps on the left were split between oocytes at different times during maturation (hpp, hours post progesterone) and embryos at different developmental stages; otherwise as in Figure 2I.

See also Figure S3.

different mRNAs. Nevertheless, when these factors were matched between cohorts of mRNAs either containing or not containing an element, we observed statistically significant support for impact of most elements identified from analyses of libraries, at embryonic stages matching those at which we observed impact on reporters (Figures S3L–S3P). The only exception was the preferential deadenylation of PBE-containing transcripts, observed for reporter mRNAs between stages 10–12 but not for endogenous mRNAs (Figures 3B and S3Q). This lack of signal for PBE activity in endogenous mRNAs might have been due to the confounding effects of newly transcribed mRNAs with long tails. These results illustrated advantages of using injected mRNAs to identify regulatory elements that are otherwise difficult to identify within endogenous mRNAs.

In summary, we found that frog early embryos achieved temporal regulation of mRNA tail lengths by combining more permissive, oocyte-inherited, CPE-mediated cytoplasmic polyadenylation and stage-specific deadenylation driven by 3′ UTR elements.

Fish embryos use a more permissive CPE and undergo more global polyadenylation

To examine tail-length regulation in fish, we injected the N37-PAS-N17 and N60 libraries into one-cell zebrafish embryos and monitored poly(A)-tail lengths as embryos developed to gastrulation (shield stage, Figure 4A). In some respects, tail-length control in zebrafish embryos resembled that of frog embryos. For example, most injected mRNAs underwent deadenylation within the first hour, with stronger effects observed for those that contained poly(A) motifs, the AAUAAU element, or AREs (Figures 4B, 4C, S4A, and S4B). Moreover, after zygotic genome activation, miR-430, the fish homolog of frog miR-427, directed deadenylation of its target mRNAs[39] (Figures 4B, 4C, S4A, and S4B).

In other respects, tail-length control differed in zebrafish embryos. First, the 5-mer associated with the longest tails was UUUUU, not UUUUA (Figure 4B). When controlling for flanking nucleotides, UUUUA was favored over UUUUU, but only slightly, whereas in frog oocytes and embryos, it was strongly favored (Figure 4D). Thus, we designate UUUUW (where W = A or U) as the core CPE in zebrafish embryos. Second, after the initial deadenylation phase, tails of most injected mRNAs were extended. This polyadenylation was not driven by a detectable sequence element; the mean tail lengths of mRNAs with nearly every 5-mer increased to a similar extent (12 nt on average) during this period (from the 32-cell to the high stage). Moreover, the tail-length increase occurred for both the N37-PAS-N17 and N60 libraries (Figures 4B and S4A) and persisted even when all mRNAs that contained either a UUUU or a canonical PAS (AWUAAA) were excluded (Figure S4C), showing that tail lengthening during these stages did not require either a CPE or PAS. Nonetheless, mRNAs containing a CPE or an AAUUGG underwent slightly more polyadenylation (Figures 4C and S4B).

These observations extended to endogenous mRNAs of fish embryos, in that most of these mRNAs (>90%) underwent tail lengthening before the high stage, regardless of whether they had a CPE, although lengthening was substantially greater for those that had a UUUUW CPE (Figure 4E). As with our injected libraries, we also observed greater lengthening for fish mRNAs that contained AAUUGG (Figure S4D), the mechanism of which merits further investigation.

Frog oocytes and embryos control translation using both tail-length-dependent and tail-length-independent mechanisms

Because poly(A)-tail length strongly influences mRNA TE in frog oocytes and early embryos,[9,29,44] we asked how 3′ UTR sequences that control tail length impact translation in these developmental contexts. The N60-PAS^{mos} library was injected into oocytes, and translating ribosomes were pelleted from lysates prepared at different developmental stages (Figure 5A). Enrichment for a *k*-mer in the mRNA of the pellet compared with that of the input indicated its effect on TE.

In frog oocytes undergoing maturation, TE gradually increased for mRNAs with CPE-containing *k*-mers (Figure 5B), a trajectory matching their tail-length changes (Figure 1H). Indeed, the TE associated with 6-mers strongly correlated with mean poly(A)-tail length (R_p = 0.89, Figure 5C). Controls confirmed that these results depended on progesterone treatment and were sensitive to EDTA disruption of 80S ribosomes (Figures 5B, 5C, and S5A). The high correlation confirmed strong coupling between poly(A)-tail length and TE for endogenous mRNAs in frog oocytes[29,44] and indicated that most translational control mediated by 3′ UTR sequences occurs through effects on poly(A)-tail length.

Some 6-mers containing multiple contiguous Cs were associated with translational repression without a corresponding decrease in tail length, which defied the global trend (Figures 5B and 5C). This repression was observed at all time points in both progesterone-treated and control oocytes (Figure 5B). Stronger repression was observed when 3′ UTRs contained more C-rich elements, again without significant differences in tail lengths (Figure S5B). Stronger repression was also observed when the C-rich element was closer to the PAS (Figure S5C). To rule out non-translation-related depletion artifacts caused by the sucrose cushion, we measured TE with a method resembling NaP-TRAP,[56] in which actively translated mRNAs are preferentially immunoprecipated based on association with the nascent chain (Figure S5D). mRNAs with C-rich elements were also depleted in this dataset (Figure S5E). These results indicated that translation of mRNAs containing C-rich elements is repressed in frog oocytes in a manner that is independent of tail length.

Interestingly, although in matured oocytes TE and poly(A)-tail length of injected mRNAs with CPEs was highly correlated, in control oocytes not undergoing maturation, the TEs of these mRNAs were substantially lower than expected based on their longer-than-average poly(A) tails, implying that they were translationally repressed (Figure 5C). Although translational repression of CPE-containing mRNAs had been reported in prophase I-arrested oocytes,[57–59] whether such repression depended on poly(A)-tail length had not been determined. Our results demonstrated that the CPE-mediated translational repression was tail-length independent. This repression substantially increased when more CPEs were present within the 3′ UTR (Figure 5D), consistent with previous observations.[12,59,60] Moreover, in prophase I-arrested oocytes, one CPE had strong synergistic translational repression with a second CPE but not with any other 5-mer (Figure S5F). The synergistic repression by multiple

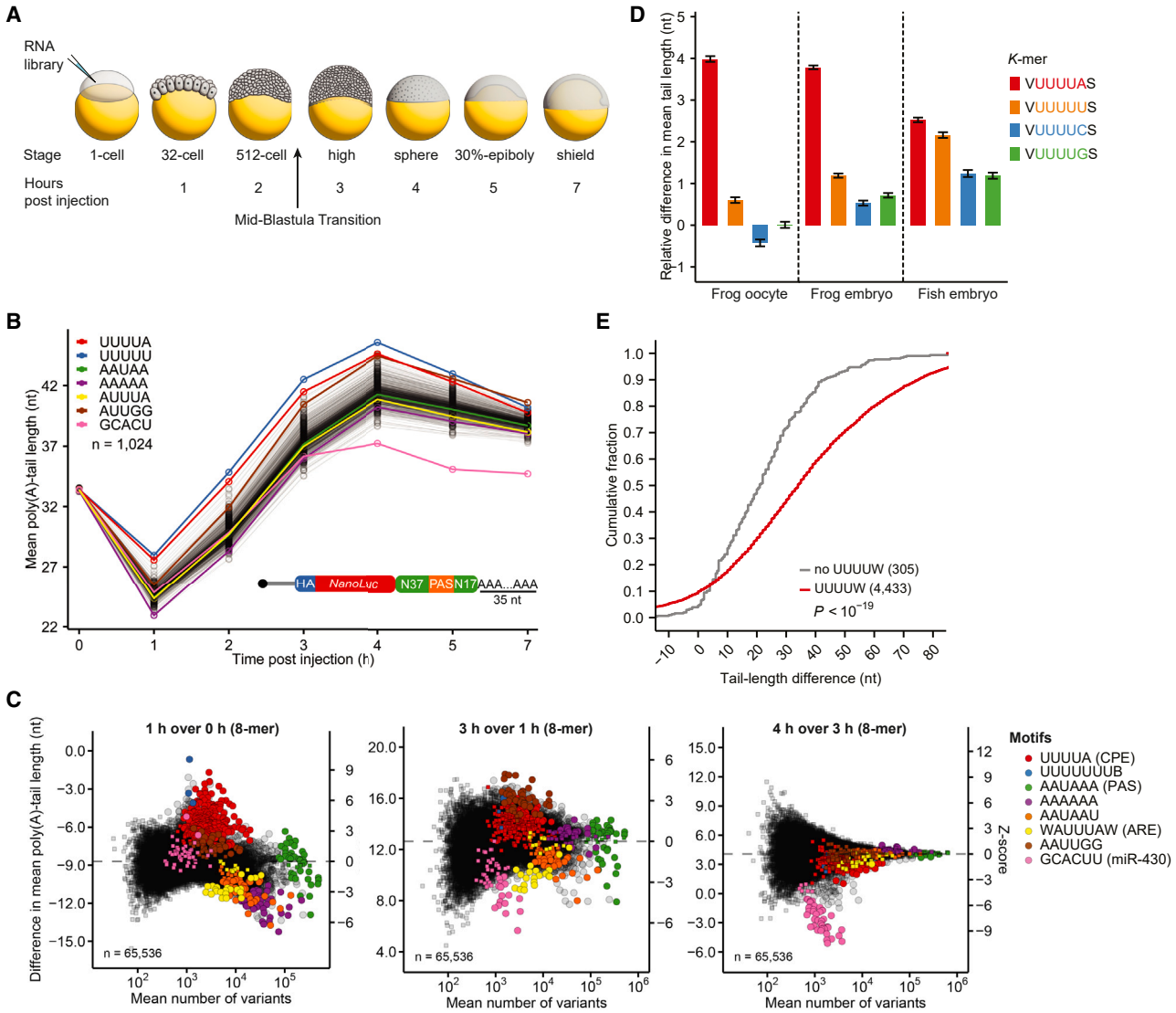


Figure 4. Early fish embryos use a more permissive CPE and undergo more global polyadenylation

(A) Experimental scheme for mRNA library injection and sample collection.

(B) Effects of 5-mers on tail lengths of mRNAs in the N37-PAS-N17 library during zebrafish early embryogenesis. Plotted for each 5-mer are the mean tail lengths of mRNAs containing that 5-mer within their 3′ UTRs. The inset shows the schematic of the injected N37-PAS-N17 library.

(C) Mean tail-length changes associated with each 8-mer in the 3′ UTRs of mRNAs of the N37-PAS-N17 library, plotted as in Figure 3B.

(D) Effects of the CPE and related motifs on poly(A)-tail length in different biological contexts. Shown are mean tail-length changes of mRNAs in the N37-PAS-N17 library containing the indicated sequence motifs (plotting changes relative to mean tail lengths of all variants; V represents A, C, or G; S represents C or G), comparing between 0 and 7 h post-progesterone treatment in frog oocytes, between 0.1 and 1 hpi in frog embryos, and between 0 and 1 hpi in zebrafish embryos (right). Error bars indicate standard error of the difference between means.

(E) Preferential tail lengthening of CPE-containing mRNAs. Plotted are cumulative distributions of tail-length changes for endogenous zebrafish mRNAs with 3′ UTRs that either contained or did not contain UUUUW, comparing between 1-cell and high-stage embryos. The numbers of unique mRNAs in each group are in parentheses; p value, Mann-Whitney U test.

See also Figure S4.

CPEs was highly specific because mRNAs that contained multiple copies of a CPE-like motif UUUUU were not repressed (Figures 5D and S5F).

To validate tail-length-independent translational repression by the C-rich element and CPE, we measured luciferase activities of individual reporters injected into frog oocytes. We tested two 3′ UTR sequences: one from the endogenous mRNA *krtcap2.L* and the other from a variant in the N60-PAS^{mos} mRNA library, both of which contained two C-rich elements (Figure 5E). The reporters ended with 3′-dA to inhibit tail-length changes. In both sequence contexts, translation of reporters containing C-rich elements was repressed compared with that of reporters in which the C-rich elements were mutated (Figure 5E). When the mutated elements were changed to CPEs, translation was again repressed,

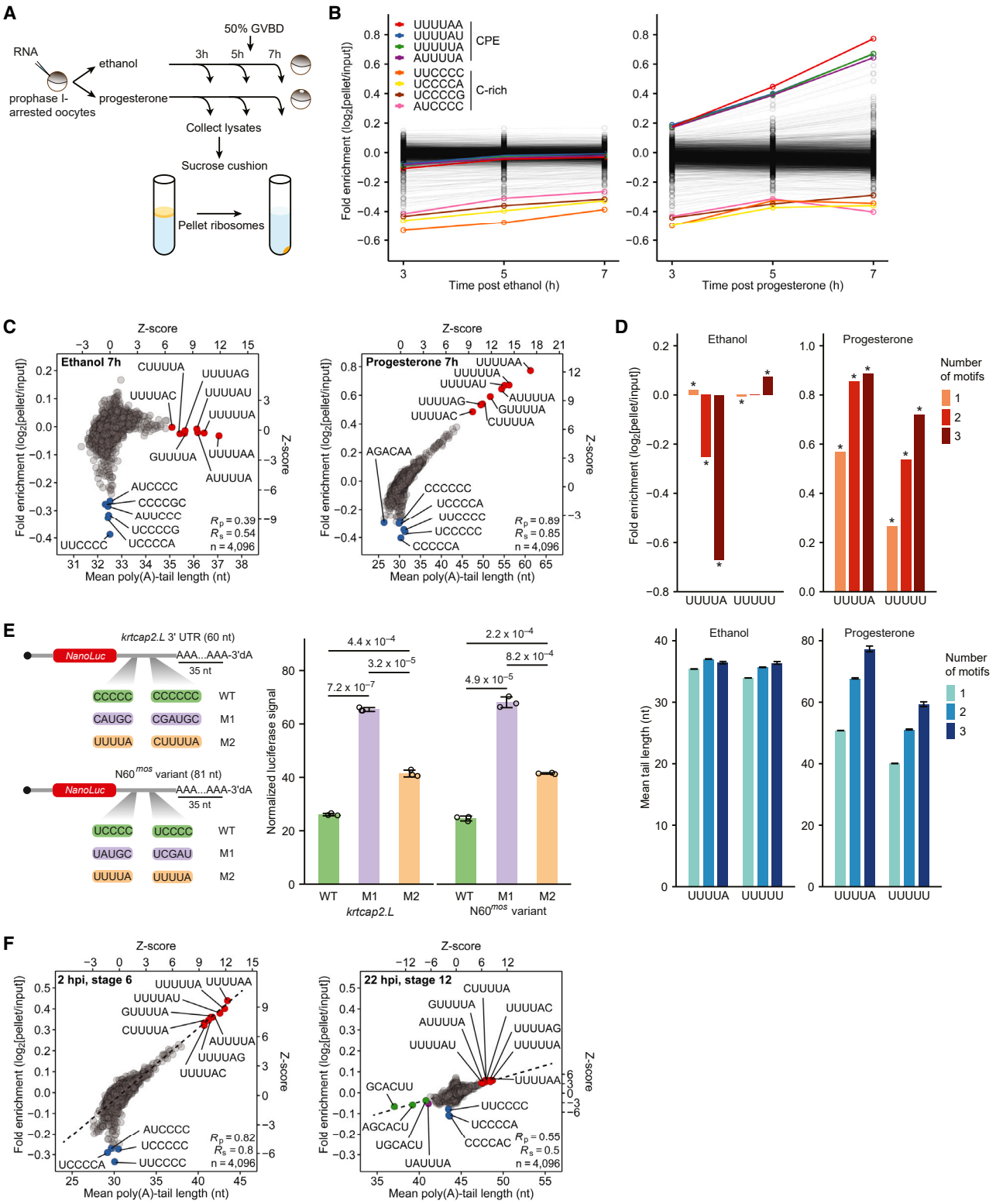


Figure 5. Translation in frog oocytes and early embryos is controlled by both tail-length-dependent and tail-length-independent mechanisms

(A) Experimental scheme for translational profiling of mRNAs in frog oocytes by ribosome pelleting. For each mRNA, TE is measured as the fold enrichment observed between the pellet and the input.

(legend continued on next page)

although to a lesser extent than that observed for the reporters with C-rich elements.

An experiment examining the tail lengths and TEs of N60-PAS-^{mos} library injected into one-cell embryos yielded results resembling those observed for mature oocytes (Figures 5C and 5F), suggesting that early embryos inherit the regulatory regime operating in mature oocytes. The strong coupling between TE and poly(A)-tail length observed in oocytes and early embryos was substantially reduced at gastrulation (22 hpi, stage 12), (Figure 5F), as expected from the switch of gene-regulatory regimes that begins at this developmental period.[9]

Together, these results from our translational profiling of injected RNAs revealed an exceedingly tight relationship between CPE-dependent polyadenylation and enhanced translation during oocyte maturation and early embryonic development. Our results also showed that the CPE and the C-rich element can cause tail-length-independent translational repression.

Tail-length control during oocyte maturation is conserved among frogs, mice, and humans

To extend our analyses to mammals, we measured poly(A)-tail lengths of mouse mRNAs isolated from germinal vesicle (GV) oocytes and metaphase II (MII) oocytes that matured *in vivo*. In mouse GV oocytes, median tail lengths of most mRNAs were >50 nt, in contrast to both frog and human oocytes, where those of homologous mRNAs were significantly shorter (Figures S6A and S6B). As in frog oocytes, mRNAs in mouse oocytes underwent substantial tail-length changes during maturation[32,61] (Figure 6A). Such changes strongly correlated with TE changes[32,61] (Figure 6B), as observed for both fly[10,11] and frog oocytes.[29] As in frog oocytes, ROC analyses indicated that UUUUA had the greatest AUC value among all *k*-mers with lengths 3–8 nt (Figures S6C and S6D), and it outperformed any combination of previously reported CPEs (Figure 6C). Analysis of published tail length and TE datasets from human oocytes[31,62] yielded similar results (Figures 6D, 6E, S6E, and S6F). Moreover, the contextual features of the UUUUA element found in frog oocytes also appeared to function in both mouse and human oocytes (Figures 6F and 6G). Thus, UUUUA specifies cytoplasmic polyadenylation in mouse and human oocytes, and the principles of tail-length control are conserved among maturing oocytes of frogs, mice, and humans.

Finally, we asked if conserved tail lengthening of homologous genes might shed light on their function in oocyte maturation and early embryogenesis. We identified 19 genes whose mRNA poly(A) tails underwent substantial lengthening ($\geq$15 nt) in frog, mouse, and human oocytes (Figure 6H). Of these 19, nine had $\geq$2-fold TE increases in both mouse and human oocytes (Figure 6H). Among these nine were those that encoded proteins with known functions in meiosis (*MAD2L1*, *SUN*1, and *MOS*). Others were involved in cell cycle regulation (*PRC1*, *BOD1*, and *GMNN*) and mRNA-splicing regulation (*MAGOH*). We suggest that timely upregulation of these proteins during oocyte maturation, by tail-length-mediated translational activation, supports meiosis and early embryonic development.

DISCUSSION

Although cytoplasmic polyadenylation during oocyte maturation has been studied for decades, the CPE responsible for poly(A)-tail lengthening had been only loosely defined, based on a handful of frog oocyte mRNAs. We identified UUUUA as the primary CPE in frog oocytes and early embryos. This motif is shorter than anticipated, which explains why it had been previously missed in many mRNAs that undergo cytoplasmic polyadenylation. Yet despite its short length, UUUUA outperformed any combination of previously proposed longer CPEs in predicting tail lengthening for both injected and endogenous mRNAs.

We also found that UUUUA specifies cytoplasmic polyadenylation in mouse and human oocytes—with similar contextual features as in frog oocytes. However, the CPE identified in zebrafish embryos differed, in that it permitted either A or U at the last position of the 5-mer (UUUUW). This degeneracy might have resulted from the binding of two alternative CPEB paralogs: CPEB1 (cpeb1a/b in fish) and CPEB4 (cpeb4a/b in fish). As indicated from binding assays[45,63] and crosslinking sites,[46,64] CPEB1 prefers to bind sites ending in A, whereas CPEB4 favors sites ending in U. Moreover, in fish embryos, production of CPEB1 is only ~3-fold[9,41] more than that of CPEB4, whereas in frog oocytes and embryos the difference is 80-fold,[9,44] implying that in fish embryos but not frog oocytes and embryos expression of CPEB4 protein might reach a level that substantially impacts cytoplasmic polyadenylation.

Identification of the actual CPE enabled key contextual features of this element to be identified. Interestingly, the flanking nucleotides that favorably influenced CPE activity extended beyond the sequence-specific interactions observed in the structure of CPEB1 and its RNA ligand,[65] and the same was

(B) Effects of 6-mers on the TE of mRNAs in the N60-PAS^{mos} library during frog oocyte maturation. Results are shown for control prophase I-arrested oocytes that had been treated with ethanol (left) and for maturing oocytes that had been treated with progesterone (right).

(C) TE and tail lengths associated with each 6-mer in the 3′ UTRs of the N60-PAS^{mos} library. Shown for each 6-mer is the mean TE observed for mRNAs with that 6-mer, plotted as a function of mean tail length for prophase I-arrested oocytes (7 h post-ethanol treatment, left) and matured oocytes (7 h post-progesterone treatment, right). Colored points indicate 6-mers containing either a CPE (red) or C-rich element (blue).

(D) Tail-length-dependent and -independent translational regulation of CPE-containing mRNAs in frog oocytes. Shown at the top for the CPE and a related motif are TEs of mRNAs in the N60-PAS^{mos} library that contained the indicated number of motifs. Shown at the bottom are mean tail lengths of mRNAs in the N60-PAS^{mos} library that contained the indicated number of motifs. Increased tail length and translation associated with UUUUU were presumably from UUUUU motifs that preceded an A. Results are shown for prophase I-arrested oocytes (7 h post-ethanol treatment, left) and matured oocytes (7 h post-progesterone treatment, right); *adjusted $p < 0.05$, one-sided binomial test. Error bars indicate standard error.

(E) Translational repression by C-rich elements and the CPE in frog oocytes. On the left are schematics of reporters that examined the effects of elements in two different 3′ UTR sequence contexts. On the right are NanoLuc signals for each reporter after normalization to a co-injected firefly luciferase reporter. Error bars indicate the standard deviation of biological triplicates; p values are from Student's t tests between indicated groups; WT, wild type.

(F) Translation and tail-length regulation by 3′ UTR sequence in frog embryos. This panel is as in (C), but for embryos at stages 6 (left) and 12 (right), and 6-mers containing either a miR-427 site (green) or an ARE (purple) are also colored.

See also Figure S5.

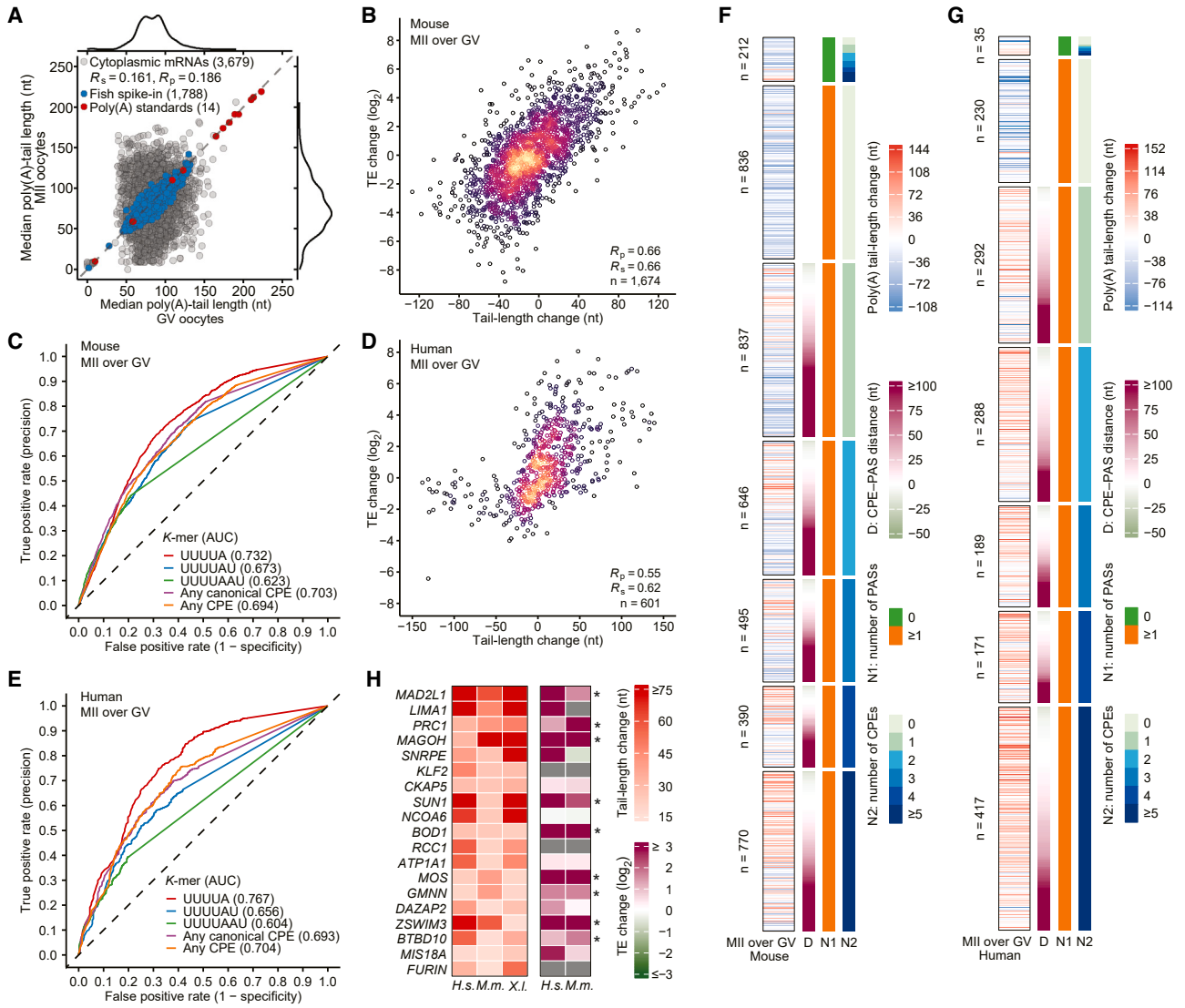


Figure 6. Tail-length control is conserved in mouse and human oocytes

(A) Poly(A) tail-length change during mouse oocyte maturation. Shown is the plot comparing median tail lengths in mouse MII oocytes to those in mouse GV oocytes. Results are shown for mouse cytoplasmic mRNAs (gray), zebrafish spike-in mRNAs (blue), and poly(A) standards (red). Numbers of mRNAs are indicated in parentheses. Only RNAs with ≥50 tags in each sample were analyzed. On the sides are density distributions for each sample.

(B) Coupling of TE with tail length in mouse oocytes. Shown is the relationship between differences in TE and differences in poly(A)-tail length for mouse oocyte mRNAs, comparing values measured in GV oocytes with those measured in MII oocytes.

(C) ROC curves testing the ability of the UUUUA and previously proposed CPEs to classify endogenous mRNAs as subject to cytoplasmic polyadenylation during mouse oocyte maturation; otherwise, as in Figure 1F.

(D) As in (B), but for human oocytes.

(E) As in (C), but for human oocytes.

(F) Effect of 3′ UTR sequence features on tail-length changes of mRNAs as mouse GV oocytes mature to MII oocytes; otherwise, as in Figure 2I.

(G) As in (E), but for human oocytes.

(H) Genes with substantial mRNA tail lengthening (≥15 nt) in human (*H.s.*), mouse (*M.m.*), and frog (*X.l.*) oocytes. Heatmaps show tail-length changes (left) and TE changes (right), comparing between human GV and MII oocytes, mouse GV and MII oocytes, and frog oocytes 0 and 7 h post-progesterone treatment. Gray indicates values not available. Asterisks indicate genes with ≥2-fold TE changes.

See also Figure S6.

true for the PAS, as observed in the structure of human CPSF bound to its RNA ligand.[66,67] Perhaps the flanking nucleotides that influenced activity facilitate either early recognition of these elements at a step preceding the final bound state or cooperative binding of both elements. With identification of the CPE and contextual features that influence its activity and that of the PAS, we propose a simplified paradigm of cytoplasmic polyadenylation specificity in early development, in which only two

context-modulated elements explain why poly(A) tails of some mRNAs are extended much more than those of others.

Complementing cytoplasmic polyadenylation, deadenylation also played a role in temporally sculpting poly(A)-tail lengths. This deadenylation occurred in one major wave followed by several smaller ones. The first wave took place after GVBD and lasted until the first hour after fertilization. During this period, deadenylation activity was strong and largely independent of sequence motifs, such that not only mRNAs lacking either a CPE or a PAS but also those containing CPEs in suboptimal sequence contexts underwent net deadenylation. This enabled maturing oocytes to concentrate translational resources on the relatively few mRNAs with long tails, a strategy seemingly also used in fish, mice, and humans.

The subsequent waves of deadenylation occurred shortly after fertilization and continued through early embryonic development, during a period in which tail lengthening became more permissive. Unlike the first wave, during these subsequent waves, tail-length shortening was driven by sequence elements within mRNA 3′ UTRs, which each worked at specific stages to shorten tails of mostly different mRNAs, thereby directing phased translational repression and clearance of different sets of maternal mRNAs at different stages. The waves of deadenylation likely involved multiple deadenylases, as both PARN[68,69] and CCR4-NOT[70] have been implicated. Moreover, mutations of BTG4, an adaptor protein for CCR4-NOT are associated with mouse[71,72] and human infertility,[73,74] supporting the significance of deadenylation during these developmental stages.

These global and sequence-specific effects all work together to produce the complex tapestry of tail-length changes observed for endogenous mRNA as follows: The context-modulated CPE and PAS specify sequence-specific tail lengthening against a backdrop of global tail shortening in the maturing oocytes and more permissive tail lengthening in the embryo, sculpted by the subsequent action of additional sequence elements that specify smaller waves of tail shortening.

Adding another layer of regulation, the CPE and C-rich motif mediated tail-length-independent translational repression. In prophase-I-arrested oocytes, the CPE binding of CPEB1 is proposed to recruit both the poly(A) polymerase Gld2 and the deadenylase PARN, with the stronger activity of PARN causing a net shortening of poly(A) tails of CPE-containing mRNAs.[75] However, in prophase-I-arrested oocytes, we did not observe preferential deadenylation of library mRNAs that contained a CPE; on the contrary, their poly(A) tails were slightly longer than average, indicating that repression of CPE-containing mRNAs does not require either short tails or the act of deadenylation. The mechanism of repression presumably instead involves other CPEB1-binding proteins, such as Maskin and eIF4ENIF1 (4E-T), both of which inhibit formation of the eIF4E-eIF4G complex required for efficient translation.[76–78] Then, when oocytes are maturing, CPEB1 is phosphorylated at Ser-174, which increases its association with CPSF and Gld2, thereby elongating poly(A) tails and transforming the CPE from a repressing motif to an activating motif.[79] This dual regulatory role of the CPE enables more pronounced translational changes for CPE-containing mRNAs, many of which are important for meiotic progression.[80]

With respect to the C-rich element, C nucleotides in 3′ UTRs provide an unfavorable context for the CPE, which could, in principle, lead to shorter tails and lower TE of mRNAs that contain C-rich elements. However, our results indicated that C-rich elements could repress TE without relying on tail shortening. A similar 3′ UTR element, which is CU-rich, recruits poly(C)-binding proteins HNRNPK and PCBP1 to inhibit 60S ribosomal subunit joining on *LOX* mRNA in erythroid cells.[81,82] Perhaps these or other poly(C)-binding proteins similarly recognize the C-rich element to mediate repression of oocyte mRNAs. Further investigation will be required to identify proteins that repress mRNAs containing the C-rich element and examine how this repression might contribute to oocyte maturation and early embryonic development.

Limitations of the study

Our mRNA libraries varied only within the 3′ UTR. Although this design had the benefit of eliminating confounding effects outside the 3′ UTR, it also prevented us from detecting influences of the 5′ UTR or the coding region on poly(A)-tail length or translation, as well as any interplay of these other regions with the 3′ UTR. Furthermore, the initial tail length of 35 nt in our mRNA libraries was designed to match the mode of the tail-length distribution of frog oocyte mRNAs. Although this relatively short tail length facilitated the detection of tail lengthening, it reduced the sensitivity of detecting tail shortening. The current sequencing depth also imposed a constraint, in that we were only able to assess tail-length changes associated with *k*-mers of a limited size (i.e., a contiguous linear *k*-mer of up to 10 nt), whereas larger elements with more informational complexity were too rare in our libraries to have been analyzed at our sequencing depth.

STAR★METHODS

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

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Animal husbandry
- METHOD DETAILS
  - Construction of DNA templates for in vitro transcription
  - Preparation of mRNAs for injections
  - Frog oocytes and embryos and fish embryos
  - Oocyte and embryo injections
  - Sample collection for poly(A) tail-length and translational profiling
  - Mouse oocyte and egg collection
  - Sucrose cushion for translational profiling
  - Nascent-chain pulldown of ribosome-associated RNAs
  - Luciferase assays
  - 3'-UTR variant and tail-length sequencing
  - PAL-seq
  - Genome references and gene annotations
  - Annotation of mRNA 3'-end isoforms

- 3'-UTR annotations and sequences
- PAL-seq data analysis
- Reporter 3'-UTR variant and tail-length analysis
- *K*-mer-associated tail length and tail-length change of reporter mRNAs
- Sequence motifs as binary predictors of endogenous cytoplasmic polyadenylation targets
- Motif logos associated with tail-length changes
- Structural accessibility and tail lengths
- Motif-associated tail-length change for endogenous mRNAs
- Enrichment analysis of *k*-mer-associated translational efficiency
- Measurement of poly(A)-tail lengths in mouse and human oocytes
- Tail-length changes of homologous genes in frogs, mice, and humans
- Relationship between poly(A)-tail length and TE in human and mouse oocytes

● QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.devcel.2024.02.007.

ACKNOWLEDGMENTS

We thank Thy Pham, Sean McGeary, and other current and former members of the Bartel lab for helpful discussions, A. Lokapally and H. Sive for experiments with frog embryos, D. Tomasello and O. Paugois for experiments with fish embryos, and the Whitehead Institute Genome Technology Core for sequencing. D.P.B. is an investigator of the Howard Hughes Medical Institute.

AUTHOR CONTRIBUTIONS

K.X. and D.P.B. conceived the project, designed the study, and wrote the manuscript. J.L. isolated RNA from mouse oocytes and discussed and commented on the manuscript. K.X. performed all the other experiments. K.X. analyzed all the data.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: September 29, 2023
Revised: December 28, 2023
Accepted: February 16, 2024
Published: March 8, 2024

REFERENCES

1. Almouzni, G., and Wolffe, A.P. (1995). Constraints on transcriptional activator function contribute to transcriptional quiescence during early Xenopus embryogenesis. EMBO J. *14*, 1752–1765.
2. Li, L., Zheng, P., and Dean, J. (2010). Maternal control of early mouse development. Development *137*, 859–870.
3. Vastenhouw, N.L., Cao, W.X., and Lipshitz, H.D. (2019). The maternal-to-zygotic transition revisited. Development *146*, dev161471.
4. Duval, C., Bouvet, P., Omilli, F., Roghi, C., Dorel, C., Leguellec, R., Paris, J., and Osborne, H.B. (1990). Stability of Maternal mRNA in Xenopus Embryos: Role of Transcription and Translation. Mol. Cell. Biol. *10*, 4123–4129.
5. Yartseva, V., and Giraldez, A.J. (2015). The Maternal-to-Zygotic Transition During Vertebrate Development: A Model for Reprogramming. Curr. Top. Dev. Biol. *113*, 191–232.
6. Richter, J.D., and Lasko, P. (2011). Translational control in oocyte development. Cold Spring Harbor Perspect. Biol. *3*, a002758.
7. Teixeira, F.K., and Lehmann, R. (2019). Translational Control during Developmental Transitions. Cold Spring Harb. Perspect. Biol. *11*, a032987.
8. Conti, M., and Kunitomi, C. (2024). A genome-wide perspective of the maternal mRNA translation program during oocyte development. Semin. Cell Dev. Biol. *154*, 88–98.
9. Subtelny, A.O., Eichhorn, S.W., Chen, G.R., Sive, H., and Bartel, D.P. (2014). Poly(A)-tail profiling reveals an embryonic switch in translational control. Nature *508*, 66–71.
10. Eichhorn, S.W., Subtelny, A.O., Kronja, I., Kwasnieski, J.C., Orr-Weaver, T.L., and Bartel, D.P. (2016). mRNA poly(A)-tail changes specified by deadenylation broadly reshape translation in Drosophila oocytes and early embryos. eLife *5*, e16955.
11. Lim, J., Lee, M., Son, A., Chang, H., and Kim, V.N. (2016). mTAIL-seq reveals dynamic poly(A) tail regulation in oocyte-to-embryo development. Genes Dev. *30*, 1671–1682.
12. Xiong, Z., Xu, K., Lin, Z., Kong, F., Wang, Q., Quan, Y., Sha, Q.Q., Li, F., Zou, Z., Liu, L., et al. (2022). Ultrasensitive Ribo-seq reveals translational landscapes during mammalian oocyte-to-embryo transition and pre-implantation development. Nat. Cell Biol. *24*, 968–980.
13. Radford, H.E., Meijer, H.A., and de Moor, C.H. (2008). Translational control by cytoplasmic polyadenylation in Xenopus oocytes. Biochim. Biophys. Acta *1779*, 217–229.
14. Sheets, M.D., Fox, C.A., Hunt, T., Vande Woude, G.V., and Wickens, M. (1994). The 3′-untranslated regions of c-mos and cyclin mRNAs stimulate translation by regulating cytoplasmic polyadenylation. Genes Dev. *8*, 926–938.
15. de Moor, C.H., and Richter, J.D. (1997). The Mos pathway regulates cytoplasmic polyadenylation in Xenopus oocytes. Mol. Cell. Biol. *17*, 6419–6426.
16. McGrew, L.L., Dworkin-Rastl, E., Dworkin, M.B., and Richter, J.D. (1989). Poly(A) elongation during Xenopus oocyte maturation is required for translational recruitment and is mediated by a short sequence element. Genes Dev. *3*, 803–815.
17. Sheets, M.D., Wu, M., and Wickens, M. (1995). Polyadenylation of c-mos mRNA as a control point in Xenopus meiotic maturation. Nature *374*, 511–516.
18. Barkoff, A., Ballantyne, S., and Wickens, M. (1998). Meiotic maturation in Xenopus requires polyadenylation of multiple mRNAs. EMBO J. *17*, 3168–3175.
19. Gebauer, F., Xu, W., Cooper, G.M., and Richter, J.D. (1994). Translational control by cytoplasmic polyadenylation of c-mos mRNA is necessary for oocyte maturation in the mouse. EMBO J. *13*, 5712–5720.
20. Tay, J., Hodgman, R., and Richter, J.D. (2000). The Control of Cyclin B1 mRNA Translation during Mouse Oocyte Maturation. Dev. Biol. *221*, 1–9.
21. Reyes, J.M., and Ross, P.J. (2016). Cytoplasmic polyadenylation in mammalian oocyte maturation. Wiley Interdiscip. Rev. RNA *7*, 71–89.
22. Ivshina, M., Lasko, P., and Richter, J.D. (2014). Cytoplasmic Polyadenylation Element Binding Proteins in Development, Health, and Disease. Annu. Rev. Cell Dev. Biol. *30*, 393–415.
23. Charlesworth, A., Meijer, H.A., and de Moor, C.H. (2013). Specificity factors in cytoplasmic polyadenylation. Wiley Interdiscip. Rev. RNA *4*, 437–461.
24. Dickson, K.S., Bilger, A., Ballantyne, S., and Wickens, M.P. (1999). The Cleavage and Polyadenylation Specificity Factor in Xenopus laevis Oocytes Is a Cytoplasmic Factor Involved in Regulated Polyadenylation. Mol. Cell. Biol. *19*, 5707–5717.

25. Hake, L.E., and Richter, J.D. (1994). CPEB is a specificity factor that mediates cytoplasmic polyadenylation during Xenopus oocyte maturation. Cell *79*, 617–627.

26. Charlesworth, A., Wilczynska, A., Thampi, P., Cox, L.L., and MacNicol, A.M. (2006). Musashi regulates the temporal order of mRNA translation during Xenopus oocyte maturation. EMBO J. *25*, 2792–2801.

27. Charlesworth, A., Yamamoto, T.M., Cook, J.M., Silva, K.D., Kotter, C.V., Carter, G.S., Holt, J.W., Lavender, H.F., MacNicol, A.M., Ying Wang, Y.Y., et al. (2012). Xenopus laevis zygote arrest 2 (zar2) encodes a zinc finger RNA-binding protein that binds to the translational control sequence in the maternal Wee1 mRNA and regulates translation. Dev. Biol. *369*, 177–190.

28. Piqué, M., López, J.M., Foissac, S., Guigó, R., and Méndez, R. (2008). A Combinatorial Code for CPE-Mediated Translational Control. Cell *132*, 434–448.

29. Yang, F., Wang, W., Cetinbas, M., Sadreyev, R.I., and Blower, M.D. (2020). Genome-wide analysis identifies cis-acting elements regulating mRNA polyadenylation and translation during vertebrate oocyte maturation. RNA *26*, 324–344.

30. Liu, Y., Nie, H., Liu, H., and Lu, F. (2019). Poly(A) inclusive RNA isoform sequencing (PAIso−seq) reveals wide-spread non-adenosine residues within RNA poly(A) tails. Nat. Commun. *10*, 5292.

31. Liu, Y., Zhao, H., Shao, F., Zhang, Y., Nie, H., Zhang, J., Li, C., Hou, Z., Chen, Z.-J., Wang, J., et al. (2023). Remodeling of maternal mRNA through poly(A) tail orchestrates human oocyte-to-embryo transition. Nat. Struct. Mol. Biol. *30*, 200–215.

32. Lee, K., Cho, K., Morey, R., and Cook-Andersen, H. (2023). An extended wave of global mRNA deadenylation sets up a switch in translation regulation across the mammalian oocyte-to-embryo transition. https://doi.org/10.1101/2023.03.21.533564.

33. Aoki, F., Hara, K.T., and Schultz, R.M. (2003). Acquisition of transcriptional competence in the 1-cell mouse embryo: Requirement for recruitment of maternal mRNAs. Mol. Reprod. Dev. *64*, 270–274.

34. Aanes, H., Winata, C.L., Lin, C.H., Chen, J.P., Srinivasan, K.G., Lee, S.G.P., Lim, A.Y.M., Hajan, H.S., Collas, P., Bourque, G., et al. (2011). Zebrafish mRNA sequencing deciphers novelties in transcriptome dynamics during maternal to zygotic transition. Genome Res. *21*, 1328–1338.

35. Simon, R., and Richter, J.D. (1994). Further analysis of cytoplasmic polyadenylation in Xenopus embryos and identification of embryonic cytoplasmic polyadenylation element-binding proteins. Mol. Cell. Biol. *14*, 7867–7875.

36. Vishnu, M.R., Sumaroka, M., Klein, P.S., and Liebhaber, S.A. (2011). The poly(rC)-binding protein alphaCP2 is a noncanonical factor in X. laevis cytoplasmic polyadenylation. RNA *17*, 944–956.

37. Voeltz, G.K., and Steitz, J.A. (1998). AUUUA Sequences Direct mRNA Deadenylation Uncoupled from Decay during Xenopus Early Development. Mol. Cell. Biol. *18*, 7537–7545.

38. Paillard, L., Omilli, F., Legagneux, V., Bassez, T., Maniey, D., and Osborne, H.B. (1998). EDEN and EDEN-BP, a cis element and an associated factor that mediate sequence-specific mRNA deadenylation in Xenopus embryos. EMBO J. *17*, 278–287.

39. Giraldez, A.J., Mishima, Y., Rihel, J., Grocock, R.J., Van Dongen, S.V., Inoue, K., Enright, A.J., and Schier, A.F. (2006). Zebrafish MiR-430 promotes deadenylation and clearance of maternal mRNAs. Science *312*, 75–79.

40. Bushati, N., Stark, A., Brennecke, J., and Cohen, S.M. (2008). Temporal reciprocity of miRNAs and their targets during the maternal-to-zygotic transition in Drosophila. Curr. Biol. *18*, 501–506.

41. Lund, E., Liu, M., Hartley, R.S., Sheets, M.D., and Dahlberg, J.E. (2009). Deadenylation of maternal mRNAs mediated by miR-427 in Xenopus laevis embryos. RNA *15*, 2351–2363.

42. Coll, O., Villalba, A., Bussotti, G., Notredame, C., and Gebauer, F. (2010). A novel, noncanonical mechanism of cytoplasmic polyadenylation operates in Drosophila embryogenesis. Genes Dev. *24*, 129–134.

43. McCrae, M.A., and Woodland, H.R. (1981). Stability of non-polyadenylated viral mRNAs injected into frog oocytes. Eur. J. Biochem. *116*, 467–470.

44. Xiang, K., and Bartel, D.P. (2021). The molecular basis of coupling between poly(A)-tail length and translational efficiency. eLife *10*, e66493.

45. Dominguez, D., Freese, P., Alexis, M.S., Su, A., Hochman, M., Palden, T., Bazile, C., Lambert, N.J., Van Nostrand, E.L.V., Pratt, G.A., et al. (2018). Sequence, Structure, and Context Preferences of Human RNA Binding Proteins. Mol. Cell *70*, 854–867.e9.

46. Stepien, B.K., Oppitz, C., Gerlach, D., Dag, U., Novatchkova, M., Krüttner, S., Stark, A., and Keleman, K. (2016). RNA-binding profiles of Drosophila CPEB proteins Orb and Orb2. Proc. Natl. Acad. Sci. USA *113*, E7030–E7038.

47. Varnum, S.M., and Wormington, W.M. (1990). Deadenylation of maternal mRNAs during Xenopus oocyte maturation does not require specific cis-sequences: a default mechanism for translational control. Genes Dev. *4*, 2278–2286.

48. Fox, C.A., and Wickens, M. (1990). Poly(A) removal during oocyte maturation: a default reaction selectively prevented by specific sequences in the 3′ UTR of certain maternal mRNAs. Genes Dev. *4*, 2287–2298.

49. Weill, L., Belloc, E., Castellazzi, C.L., and Méndez, R. (2017). Musashi 1 regulates the timing and extent of meiotic mRNA translational activation by promoting the use of specific CPEs. Nat. Struct. Mol. Biol. *24*, 672–681.

50. McGrew, L.L., and Richter, J.D. (1990). Translational control by cytoplasmic polyadenylation during Xenopus oocyte maturation: characterization of cis and trans elements and regulation by cyclin/MPF. EMBO J. *9*, 3743–3751.

51. Nieuwkoop, P.D., and Faber, J. (1994). Normal Table of Xenopus laevis (Daudin): a Systematical and Chronological Survey of the Development from the Fertilized Egg Till the end of Metamorphosis (New York: Garland Pub).

52. McGeary, S.E., Lin, K.S., Shi, C.Y., Pham, T.M., Bisaria, N., Kelley, G.M., and Bartel, D.P. (2019). The biochemical basis of microRNA targeting efficacy. Science *366*, eaav1741.

53. Peshkin, L., Wühr, M., Pearl, E., Haas, W., Jr., Freeman, R.M., Gerhart, J.C., Klein, A.M., Horb, M., Gygi, S.P., and Kirschner, M.W. (2015). On the Relationship of Protein and mRNA Dynamics in Vertebrate Embryonic Development. Dev. Cell *35*, 383–394.

54. Wreden, C., Verrotti, A.C., Schisa, J.A., Lieberfarb, M.E., and Strickland, S. (1997). Nanos and pumilio establish embryonic polarity in Drosophila by promoting posterior deadenylation of hunchback mRNA. Development *124*, 3015–3023.

55. Lin, K., Zhang, S., Shi, Q., Zhu, M., Gao, L., Xia, W., Geng, B., Zheng, Z., and Xu, E.Y. (2018). Essential requirement of mammalian Pumilio family in embryonic development. Mol. Biol. Cell *29*, 2922–2932.

56. Strayer, E.C., Krishna, S., Lee, H., Vejnar, C., Beaudoin, J.-D., and Giraldez, A.J. (2023). NaP-TRAP, a novel massively parallel reporter assay to quantify translation control. https://doi.org/10.1101/2023.11.09.566434.

57. Barkoff, A.F., Dickson, K.S., Gray, N.K., and Wickens, M. (2000). Translational Control of Cyclin B1 mRNA during Meiotic Maturation: Coordinated Repression and Cytoplasmic Polyadenylation. Dev. Biol. *220*, 97–109.

58. Stutz, A., Conne, B., Huarte, J., Gubler, P., Völkel, V., Flandin, P., and Vassalli, J.-D. (1998). Masking, unmasking, and regulated polyadenylation cooperate in the translational control of a dormant mRNA in mouse oocytes. Genes Dev. *12*, 2535–2548.

59. de Moor, C.H., and Richter, J.D. (1999). Cytoplasmic polyadenylation elements mediate masking and unmasking of cyclin B1 mRNA. EMBO J. *18*, 2294–2303.

60. Dai, X.-X., Jiang, J.-C., Sha, Q.-Q., Jiang, Y., Ou, X.-H., and Fan, H.-Y. (2019). A combinatorial code for mRNA 3′-UTR-mediated translational control in the mouse oocyte. Nucleic Acids Res. *47*, 328–340.

61. Liu, Y., Nie, H., Zhang, C., Hou, Z., Wang, J., and Lu, F. (2021). Poly(A) tail length is a major regulator of maternal gene expression during the mammalian oocyte-to-embryo transition. https://doi.org/10.1101/2021.08.29.458052.

62. Zou, Z., Zhang, C., Wang, Q., Hou, Z., Xiong, Z., Kong, F., Wang, Q., Song, J., Liu, B., Liu, B., et al. (2022). Translatome and transcriptome co-profiling reveals a role of TPRXs in human zygotic genome activation. Science *378*, abo7923.

63. Ray, D., Kazan, H., Cook, K.B., Weirauch, M.T., Najafabadi, H.S., Li, X., Gueroussov, S., Albu, M., Zheng, H., Yang, A., et al. (2013). A compendium of RNA-binding motifs for decoding gene regulation. Nature *499*, 172–177.

64. Poetz, F., Lebedeva, S., Schott, J., Lindner, D., Ohler, U., and Stoecklin, G. (2022). Control of immediate early gene expression by CPEB4-repressor complex-mediated mRNA degradation. Genome Biol. *23*, 193.

65. Afroz, T., Skrisovska, L., Belloc, E., Guillén-Boixet, J., Méndez, R., and Allain, F.H.-T. (2014). A fly trap mechanism provides sequence-specific RNA recognition by CPEB proteins. Genes Dev. *28*, 1498–1514.

66. Sun, Y., Zhang, Y., Hamilton, K., Manley, J.L., Shi, Y., Walz, T., and Tong, L. (2018). Molecular basis for the recognition of the human AAUAAA polyadenylation signal. Proc. Natl. Acad. Sci. USA *115*, E1419–E1428.

67. Clerici, M., Faini, M., Muckenfuss, L.M., Aebersold, R., and Jinek, M. (2018). Structural basis of AAUAAA polyadenylation signal recognition by the human CPSF complex. Nat. Struct. Mol. Biol. *25*, 135–138.

68. Moraes, K.C.M., Wilusz, C.J., and Wilusz, J. (2006). CUG-BP binds to RNA substrates and recruits PARN deadenylase. RNA *12*, 1084–1091.

69. Körner, C.G., Wormington, M., Muckenthaler, M., Schneider, S., Dehlin, E., and Wahle, E. (1998). The deadenylating nuclease (DAN) is involved in poly(A) tail removal during the meiotic maturation of Xenopus oocytes. EMBO J. *17*, 5427–5437.

70. Sha, Q.Q., Yu, J.L., Guo, J.X., Dai, X.X., Jiang, J.C., Zhang, Y.L., Yu, C., Ji, S.Y., Jiang, Y., Zhang, S.Y., et al. (2018). CNOT6L couples the selective degradation of maternal transcripts to meiotic cell cycle progression in mouse oocyte. EMBO J. *37*, e99333.

71. Yu, C., Ji, S.-Y., Sha, Q.-Q., Dang, Y., Zhou, J.-J., Zhang, Y.-L., Liu, Y., Wang, Z.-W., Hu, B., Sun, Q.-Y., et al. (2016). BTG4 is a meiotic cell cycle–coupled maternal-zygotic-transition licensing factor in oocytes. Nat. Struct. Mol. Biol. *23*, 387–394.

72. Liu, Y., Lu, X., Shi, J., Yu, X., Zhang, X., Zhu, K., Yi, Z., Duan, E., and Li, L. (2016). BTG4 is a key regulator for maternal mRNA clearance during mouse early embryogenesis. J. Mol. Cell Biol. *8*, 366–368.

73. Zheng, W., Zhou, Z., Sha, Q., Niu, X., Sun, X., Shi, J., Zhao, L., Zhang, S., Dai, J., Cai, S., et al. (2020). Homozygous Mutations in BTG4 Cause Zygotic Cleavage Failure and Female Infertility. Am. J. Hum. Genet. *107*, 24–33.

74. Liu, R., Zhou, Y., Li, Q., Chen, B., Zhou, Z., Wang, L., Wang, L., Sang, Q., and Jin, L. (2021). A novel homozygous missense variant in BTG4 causes zygotic cleavage failure and female infertility. J. Assist. Reprod. Genet. *38*, 3261–3266.

75. Kim, J.H., and Richter, J.D. (2006). Opposing polymerase-deadenylase activities regulate cytoplasmic polyadenylation. Mol. Cell *24*, 173–183.

76. Stebbins-Boaz, B., Cao, Q., de Moor, C.H., Mendez, R., and Richter, J.D. (1999). Maskin Is a CPEB-Associated Factor that Transiently Interacts with eIF-4E. Mol. Cell *4*, 1017–1027.

77. Minshall, N., Reiter, M.H., Weil, D., and Standart, N. (2007). CPEB Interacts with an Ovary-specific eIF4E and 4E-T in Early Xenopus Oocytes. J. Biol. Chem. *282*, 37389–37401.

78. Lorenzo-Orts, L., Strobl, M., Steinmetz, B., Leesch, F., Pribitzer, C., Roehsner, J., Schutzbier, M., Dürnberger, G., and Pauli, A. (2024). eIF4E1b is a non-canonical eIF4E protecting maternal dormant mRNAs. EMBO Rep. *25*, 404–427.

79. Mendez, R., Murthy, K.G.K., Ryan, K., Manley, J.L., and Richter, J.D. (2000). Phosphorylation of CPEB by Eg2 Mediates the Recruitment of CPSF into an Active Cytoplasmic Polyadenylation Complex. Mol. Cell *6*, 1253–1259.

80. Meneau, F., Dupré, A., Jessus, C., and Daldello, E.M. (2020). Translational Control of Xenopus Oocyte Meiosis: Toward the Genomic Era. Cells *9*, 1502.

81. Ostareck, D.H., Ostareck-Lederer, A., Wilm, M., Thiele, B.J., Mann, M., and Hentze, M.W. (1997). mRNA Silencing in Erythroid Differentiation: hnRNP K and hnRNP E1 Regulate 15-Lipoxygenase Translation from the 3′ End. Cell *89*, 597–606.

82. Ostareck, D.H., Ostareck-Lederer, A., Shatsky, I.N., and Hentze, M.W. (2001). Lipoxygenase mRNA Silencing in Erythroid Differentiation. Cell *104*, 281–290.

83. Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol. Cell *38*, 576–589.

84. Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842.

85. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal 17, 10–12.

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

87. Putri, G.H., Anders, S., Pyl, P.T., Pimanda, J.E., and Zanini, F. (2021). Analysing high-throughput sequencing data in Python with HTSeq 2.0. https://doi.org/10.48550/arXiv.2112.00939.

88. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al. (2011). Scikit-learn: Machine Learning in Python. J. Mach. Learn. Res. *12*, 2825–2830.

89. Wagih, O. (2017). ggseqlogo: a versatile R package for drawing sequence logos. Bioinformatics *33*, 3645–3647.

90. Lorenz, R., Bernhart, S.H., Höner Zu Siederdissen, C.H., Tafer, H., Flamm, C., Stadler, P.F., and Hofacker, I.L. (2011). ViennaRNA Package 2.0. Algorithms Mol. Biol. *6*, 26.

91. Ho, D.E., Imai, K., King, G., and Stuart, E.A. (2011). MatchIt: Nonparametric Preprocessing for Parametric Causal Inference. J. Stat. Softw. *42*, 1–28.

92. Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis (Springer).

93. Cock, P.J.A., Antao, T., Chang, J.T., Chapman, B.A., Cox, C.J., Dalke, A., Friedberg, I., Hamelryck, T., Kauff, F., Wilczynski, B., et al. (2009). Biopython: freely available Python tools for computational molecular biology and bioinformatics. Bioinformatics *25*, 1422–1423.

94. R Core Team (2021). R: A Language and Environment for Statistical Computing (R Foundation for Statistical Computing).

95. Sive, H.L., Grainger, R.M., and Harland, R.M. (2000). Early Development of Xenopus laevis: A Laboratory Manual (Cold Spring Harbor Laboratory Press).

96. Zahn, N., James-Zorn, C., Ponferrada, V.G., Adams, D.S., Grzymkowski, J., Buchholz, D.R., Nascone-Yoder, N.M., Horb, M., Moody, S.A., Vize, P.D., et al. (2022). Normal Table of Xenopus development: a new graphical resource. Development *149*, dev200356.

97. Kimmel, C.B., Ballard, W.W., Kimmel, S.R., Ullmann, B., and Schilling, T.F. (1995). Stages of embryonic development of the zebrafish. Dev. Dyn. *203*, 253–310.

98. Medina-Muñoz, S.G., Kushawah, G., Castellano, L.A., Diez, M., DeVore, M.L., Salazar, M.J.B., and Bazzini, A.A. (2021). Crosstalk between codon optimality and cis -regulatory elements dictates mRNA stability. Genome Biol. *22*, 14.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological Samples		
Xenopus laevis ovaries	Nasco	LM00935 (discontinued)
Defolliculated oocytes	Xenopus1	12005
Chemicals, Peptides, and Recombinant Proteins		
Progesterone	Millipore Sigma	P0130
Dithiothreitol	GoldBio	DTT25
3′-dATP	Jena Bioscience	NU-1123L
Gentamicin	Thermo Fisher	15750060, 15710064
human chorionic gonadotropin (hCG)	Millipore Sigma	CG5-1VL
pregnant mare's serum gonadotropin (PMSG)	Lee Biosolutions	493-10-2.5
polyvinylpyrrolidone	Millipore Sigma	P2307
milrinone	Millipore Sigma	M4659
hyaluronidase	Millipore Sigma	H4272
cycloheximide	Millipore Sigma	C7698
Critical Commercial Assays		
Nano-Glo Dual-Luciferase Reporter Assay System	Promega	N1630
Deposited Data		
Sequencing data from this paper	Gene Expression Omnibus	GSE241107
Raw data for Figure 5E	Mendeley Data	https://doi.org/10.17632/k35r7fm5d3.1
PAIso-seq for human oocytes	GSA-Human	HRA001911
RNA-seq and Ribo-seq data for human oocytes	Gene Expression Omnibus	GSE197265
RNA-seq and Ribo-seq data for mouse oocytes	Gene Expression Omnibus	GSE165782
RNA-seq data for prophase I-arrested oocytes (stage VI)	Gene Expression Omnibus	GSE166544
RNA-seq data for fish (Danio rerio) embryos	Gene Expression Omnibus	GSE148391
Human genomic sequences and gene annotations	GENCODE	release 25, GRCh38.p7, primary assembly
Mouse genomic sequences and gene annotations	GENCODE	release 10, GRCh38.p4, primary assembly
Frog genomic sequences and gene annotations	Xenbase	v10.1 assembly
Frog mitochondrial genomic sequences and gene annotations	NCBI	NC_001573.1
Fish genomic sequences and gene annotations	Ensembl	GRCz11.v97
Oligonucleotides		
Oligonucleotides	See Table S1	NA
Recombinant DNA		
C071	Addgene	174343
Software and Algorithms		
PAL-seq data analysis pipeline	This paper	https://github.com/coffeebond/PAL-seq
Massively parallel reporter tail-length analysis pipeline	This paper	https://github.com/coffeebond/MPRA_tail_seq
HOMER	Heinz et al.[83]	v4.11.1 (http://homer.ucsd.edu/homer/)
bedtools	Quinlan and Hall[84]	v2.26.0 (http://bedtools.readthedocs.io)
cutadapt	Martin et al.[85]	v3.7 (http://cutadapt.readthedocs.io)
STAR	Dobin et al.[86]	v2.7.1a (https://github.com/alexdobin/STAR)
HTSeq	Putri et al.[87]	1.99.2 (https://htseq.readthedocs.io/en/latest/)
ghmm	N/A	Beta (https://ghmm.sourceforge.net)
scikit-learn	Pedregosa et al.[88]	0.22.2 (https://scikit-learn.org/stable/)
ggseqlogo	Wagih[89]	0.1 (https://github.com/omarwagih/ggseqlogo)

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
RNAfold	Lorenz et al.[90]	2.6.4 (https://www.tbi.univie.ac.at/RNA/RNAfold.1.html)
MatchIt	Ho et al.[91]	4.5.5 (https://cran.r-project.org/web/packages/MatchIt/vignettes/MatchIt.html)
ggplot2	Wickham[92]	3.4.4 (https://ggplot2.tidyverse.org)
Biopython	Cock et al.[93]	1.7.9 (https://biopython.org)
python	N/A	v2.7, v3.6 (https://www.python.org)
R	R Core Team[94]	v4.1 (https://www.r-project.org)

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, David Bartel (dbartel@wi.mit.edu).

Materials availability

This study did not generate new unique reagents. Materials are available from the lead contact, David Bartel (dbartel@wi.mit.edu) upon request.

Data and code availability

- All standard sequencing data are available in the Gene Expression Omnibus under the accession number GSE241107. Raw intensity data for PAL-seq and reporter mRNA tail-length sequencing cannot be deposited in public databases due to large sizes and are available from the lead contact, David Bartel (dbartel@wi.mit.edu) upon request. Raw data from Figure 5E were deposited on Mendeley at https://doi.org/10.17632/k35r7fm5d3.1.
- PAL-seq data analyses were performed using a custom script written in Python 2.7 and available at https://github.com/coffeebond/PAL-seq. Reporter mRNA tail-length sequencing data analyses were performed using a custom script written in Python 2.7 and available at https://github.com/coffeebond/MPRA_tail_seq.
- 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

Animal husbandry

The C57BL/6J inbred mouse line was used in this study. Mice were housed in a 12/12 h light/dark cycle with constant temperature and food at the Whitehead Institute for Biomedical Research. The Massachusetts Institute of Technology Department of Comparative Medicine provided daily cage maintenance and veterinarian health checks. Wild-type frogs (*Xenopus laevis*) were kept in dechlorinated water tanks at the Whitehead animal facility on a 12/12 h, light/dark cycle. Embryos were obtained and staged as described.[95,96] Wild-type AB line adult zebrafish (*Danio rerio*) were maintained at 28°C on a 12/12 h, light/dark cycle. Embryos were obtained and staged as described.[97] All animal experiments performed in this study were approved by the Massachusetts Institute of Technology Committee on Animal Care.

METHOD DETAILS

Construction of DNA templates for in vitro transcription

For the N60-PAS^{mos} library, the DNA template was assembled from four DNA fragments by overlapping PCR. Three out of four fragments were amplified using the KAPA HiFi HotStart Kit (Roche, KK2502) following the manufacturer's suggested protocol, from the plasmid C071 with the following primer sets: KXU024 and KXR303 (for the T7 promoter and the 5' UTR, fragment 1 [F1]), KXF303 and KXU236 (for the *NanoLuc* coding region, fragment 2 [F2]), KXU110 and KXU068 (for the HDV ribozyme, fragment 4 [F4]). The other fragment (F3), which contained a 60-nt random-sequence region (A:C:G:T = 30:19:19:32), a 21-nt region based on the 3' end of frog *mos.L* mRNA, and a 35-nt poly(A) stretch was constructed by Klenow extension of oligos KXS050 and KXS051. The two oligos (200 pmol each) were annealed in water in a 49-μl reaction, which was incubated at 85°C for 5 min and then slowly cooled to 25°C at 0.1°C/sec. A mixture containing 6 μl NEBuffer 2 (New England Biolabs, B7002S), 2 μl Klenow enzyme (5 units/μl, New England Biolabs, M0210S), and 3 μl 10 mM dNTP (KAPA HiFi Kits from Roche, KK2502) were added to the annealed oligos. The reaction was incubated at 25°C for 15 min and stopped by adding 1.2 μl 500 mM EDTA and incubating at 75°C for 20 min. Overlapping PCR reactions were performed to join F1 and F2, and F3 and F4 separately. The joined products were then joined to form the final DNA

product. A standard overlapping PCR was performed in a 100-μl reaction containing 10 pmol of each fragment, 2 μl KAPA HiFi enzyme, 3 μl 10 mM dNTP, and 20 μl 5x KAPA HiFi buffer for 15 cycles, with an annealing temperature of 66°C. All PCR products and double-stranded DNAs were purified with agarose gels (Lonza, 50004) and the GeneJet Gel Extraction Kit (Thermo Fisher, K0692).

DNA templates for the N60 and the N37-PAS-N17 libraries were constructed similarly except for F3. For this fragment, a splint ligation was carried out with an acceptor oligo containing the variable region(s) (KXS088 for the N60 library and KXS087 for the N37-PAS-N17 library), a donor oligo KXS053, and the splint oligo KXS055. The oligos (1 nmol each) were mixed in water in a 50-μl reaction, heated to 95°C for 5 min, and then slowly cooled to 25°C at 0.1°C/sec. 24 μl was taken from the annealed product, mixed with 3 μl T4 DNA ligase and 3 μl T4 DNA ligase buffer (New England Biolabs, M0202S), and incubated at 25°C overnight (16 hr). The ligated product was mixed with an equal volume of Gel Loading Buffer II (Thermo Fisher, AM8547), heated for 5 min at 95°C, and resolved on an 8% urea-acrylamide denaturing gel (National Diagnostics, EC-828). The band migrating at the size of 138 nt was identified by UV-shadowing, excised, macerated, and eluted in a buffer containing 10 mM HPEPS pH 7.5 and 300 mM NaCl at 70°C for 30 min. The gel pieces were removed using Spin-X columns (Corning, 8160). DNAs were then precipitated with isopropanol and resuspended in water for downstream Klenow extension with oligo KXU237 in the same way as F3 in the N60-PAS^{mos} library.

The DNA template for the CPE^{mos}-N60 library was constructed in a similar way to that for the N60 library, except for the following differences: 1) F1 and F2 were amplified as one fragment from the plasmid C071 with primers KXU024 and KXS054 (thus this library lacks the HA tag); 2) KXS052 was used as the acceptor oligo for the splint ligation when F3 was generated.

For individual mRNA reporters used for luciferase assays, DNA templates were assembled with fragments F1 and F2 as in the DNA template for the N60-PAS^{mos} library, except that F2 was amplified with a different reverse primer (one of KXU333–KXU338). Each reverse primer contained a different 3'-UTR sequences and a 35-nt poly(A) tail.

Preparation of mRNAs for injections

For mRNA libraries, in vitro transcription was performed in a 100-μl reaction containing 40 mM Tris pH 8.0, 21 mM $MgCl_2$, 2 mM Spermidine (Sigma, 85558-1G), 1 mM dithiothreitol (GoldBio, DTT25), 5 mM NTP Set (Thermo Fisher, R0481), 0.2 units yeast inorganic pyrophosphatase (New England Biolabs, M2403L), 80 units SUPERase•In (Thermo Fisher, AM2694), 2 μg DNA template, and T7 RNA Polymerase (purified in-house and used at final concentration of 6.4 ng/μl). After incubation at 37°C for 3 hr, 2 units of DNase I (New England Biolabs, M0303S) were added, followed by another 20 min incubation at 37°C. To enhance HDV ribozyme cleavage, thermal cycling was performed (65°C for 90 sec and 37°C for 5 min, four cycles, 50 μl of reaction per tube). Before gel loading, 2 μl 0.5 M EDTA pH 8.0 and 100 μl Gel Loading Buffer II were added. After incubation at 65°C for 5 min, RNAs were separated on 5% urea-acrylamide denaturing gels. Desired RNA bands were identified by UV-shadowing, excised, macerated, and eluted in a buffer containing 10 mM HPEPS pH 7.5 and 300 mM NaCl at 23°C overnight (> 16 hr) on a thermomixer, shaking at 1400 rpm (15 sec on and 105 sec off). The gel pieces were removed using Spin-X columns, and RNAs were precipitated with isopropanol and resuspended in water for downstream reactions.

Capping of RNAs was carried out with the Vaccinia Capping System (New England Biolabs, M2080S) following the manufacturer's protocol, except that the Vaccinia capping enzyme was used at 2 units/μl. RNAs were then purified by phenol/chloroform extraction and ethanol precipitation. Water-resuspended RNAs were applied to Micro Bio-Spin P-30 columns (Bio-Rad, 7326250) for desalting.

2',3'-cyclic phosphates generated by HDV ribozyme cleavage were removed in a 100-μl reaction containing up to 100 μg capped RNAs, 50 units T4 Polynucleotide Kinase (PNK, New England Biolabs, M0201S), 1x T4 PNK buffer, and 25 units SUPERase•In. After incubation at 37°C for 1 hr, RNAs were purified by phenol/chloroform extraction and ethanol precipitation. RNAs were then resuspended in 1x Gel Loading Buffer II and purified with urea-acrylamide gels in the same way as after in vitro transcription. All RNAs were checked for integrity by visualization on formaldehyde-agarose denaturing gels as described[44] before being stored at –80°C.

Individual mRNAs used for luciferase assays were in vitro transcribed the same way as mRNA libraries but in 50-μl reactions, and no HDV cycling was performed after the DNase-I digestion. RNAs were supplemented with 2 μl 0.5 M EDTA pH 8.0 and applied to Micro Bio-Spin P-30 columns for desalting. RNAs were then purified by phenol/chloroform extraction and ethanol precipitation. Capping was performed the same as for the mRNA libraries. 3'-dATP was added to RNAs with Yeast Poly(A) Polymerase (ThermoFisher, 74225Z25KU) in a 40-μl reaction containing 1x Yeast Poly(A) Polymerase Reaction Buffer, 4 μM RNA, 500 μM 3'-dATP (Jena Bioscience, NU-1123L), 1200 units Yeast Poly(A) Polymerase, and 20 units SUPERase•In. After incubation at 37°C for 1 h, RNAs were applied to Micro Bio-Spin P-30 columns, and the flowthrough was further purified by phenol/chloroform extraction and ethanol precipitation. To confirm that the 3' ends were blocked by 3'-dATP, poly(A) tailing assays were performed with purified RNAs in a 10-μl reaction containing 1x Yeast Poly(A) Polymerase Reaction Buffer, 500 ng RNA, 500 μM ATP (New England Biolabs, P0756S), and 300 units Yeast Poly(A) Polymerase. After 20 min incubation at 37°C, RNAs were resolved on formaldehyde-agarose denaturing gels as described.[44]

Frog oocytes and embryos and fish embryos

Frog oocytes used for the injection of the N60-PAS^{mos} and the CPE^{mos}-N60 libraries and measurements of endogenous mRNA poly(A)-tail lengths were obtained from frog (*X. laevis*) ovaries purchased from Nasco (LM00935) as described.[44] Due to the discontinuation of frog ovaries from Nasco, frog oocytes used for injections of the N60 and N37-PAS-N17 libraries as well as individual mRNA reporters were obtained from Xenopus1 (12005). Because these oocytes were defolliculated by the vendor, healthy Stage V and VI oocytes were hand-picked and transferred to OR-2 buffer (5 mM HEPES pH 7.6, 82.5 mM NaCl, 2.5 mM KCl, 1 mM

$MgCl_2$, 1 mM $CaCl_2$, 1 mM Na_2HPO4) supplemented with 100 μg/ml Gentamicin (Thermo Fisher, 15750060) incubated at 18°C overnight (> 16 h) for recovery before injections.

Frog oocytes were matured in vitro in OR-2 buffer supplemented with 10 μg/ml progesterone (Millipore Sigma, P0130) from a 10 mg/ml stock in ethanol. For controls, oocytes were incubated in OR-2 buffer with ethanol (0.1%). The time required for frog oocytes to mature varied between different batches. In most cases, 50% of oocytes reached GVBD (displaying a white spot in the animal pole) between 3 and 5 h, and 100% of oocytes reached GVBD by 7 h after incubation with progesterone.

For frog embryos, female frogs (*X. laevis*) were injected with 700 units of human chorionic gonadotropin (hCG) into the dorsal lymph sac. After 12–16 h, eggs were harvested and fertilized in vitro with testis pieces cut out from a sacrificed male frog in 0.1x MBS (8.8 mM NaCl, 0.1 mM KCl, 0.1 mM $MgSO_4$, 0.5 mM HEPES, 0.25 mM $NaHCO_3$, 0.07 mM $CaCl_2$, pH 7.8). The fertilized eggs were de-jellied in 2% (w/v) cysteine (pH 8.0), thoroughly rinsed with 0.1x MBS, and transferred to 1x MBS for injections. After injections, frog embryos were kept in 1x MBS at 23°C until stage 10, when they were moved to 15°C until they reached stage 12. Frog embryos were staged as described.[51,96]

Zebrafish (*Danio rerio*) embryos were obtained from the wide-type AB line, incubated in E3 medium (5 mM NaCl, 0.17 mM KCl, 0.33 mM $CaCl_2$, 0.33 mM $MgSO_4$, 0.1% w/v Methylene Blue) at 28°C, and staged as described.[97]

All animal use was in accordance with a protocol approved by the Massachusetts Institute of Technology Committee on Animal Care.

Oocyte and embryo injections

All injections were performed at 23°C with a PLI-100 Plus Pico-Injector. Stage V or VI frog oocytes that were healthy after overnight recovery were selected for injection. For the N60-PAS^{mos} and CPE^{mos}-N60, 4 nl mRNA (0.5 pmol/μl) was injected per oocyte. The N60 and N37-PAS-N17 libraries were mixed at a 1:3 molar ratio, and 4 nl (0.5 pmol/μl) of the mix was injected per oocyte. For luciferase assays, each oocyte was injected with 2 nl of an RNA mixture, containing NanoLuc luciferase reporter mRNA with a unique 3'-UTR sequence (0.05 pmol/μl) and a firefly luciferase mRNA with 120-nt poly(A) region followed by a mutant mouse Malat1 3' end (0.1 pmol/μl), which was used for normalization.[44]

Frog embryos were injected at stage 1. For the N60-PAS^{mos} library, 4 nl mRNA (0.5 pmol/μl) was injected per embryo. The N60 and N37-PAS-N17 libraries were mixed at a 1:3 molar ratio, and 4 nl (0.1 pmol/μl) was injected per embryo. Fish embryos were injected at the 1-cell stage. The N60 and N37-PAS-N17 libraries were mixed at a 1:3 molar ratio in a buffer containing 0.025% phenol red and 100 mM KCl. Each embryo was injected with 2 nl of the RNA library mix (0.1 pmol/μl).

Sample collection for poly(A) tail-length and translational profiling

Frog oocytes (50–100) were collected at the indicated time after either ethanol or progesterone treatment. OR-2 buffer was removed and oocytes were washed three times with ice-cold buffer RL (20 mM HEPES pH 7.5, 100 mM KCl, 5 mM $MgCl_2$, 1% [v/v] Triton X-100, 100 μg/ml cycloheximide, cOmplete protease inhibitor cocktail [MilliporeSigma, 11836170001, 1 tablet per 10 ml buffer], 10 μl/oocyte). After removing all wash buffer, oocytes were lysed in buffer RL supplemented with 200 units/ml SUPERase•In (10 μl/oocyte) by vigorous shaking and pipetting. Lysates were cleared by centrifugation at 5000*g* at 4°C for 10 min. A portion of each supernatant (50 μl, equivalent to 5 oocytes) was mixed with 500 μl Tri Reagent (Thermo Fisher, AM9738). The rest of each supernatant was either used for nascent-chain pulldown immediately or flash-frozen in liquid nitrogen in aliquots stored at –80°C for analysis on a sucrose cushion. Note that for reporter mRNAs, the "0 h post progesterone treatment" sample was prepared by collecting the supernatant of the lysate of untreated oocytes and combining it with the uninjected mRNA library before the Tri Reagent was added to the mix. For endogenous mRNAs, the "0 h post progesterone treatment" sample was prepared by collecting the supernatant of the lysate of oocytes not treated with any reagents.

Frog embryos (~30) were collected at indicated developmental stages the same way as frog oocytes. Fish embryos (10–15) were collected at indicated developmental stages, E3 medium was removed, and embryos were washed once with 1 ml E3 medium and incubated in 1 ml E3 medium with 2 mg/ml pronase (Roche Diagnostics, 10165921001) for 4 min for de-chorionation. After removing all E3 medium, 1 ml Tri Reagent was added to the embryos, followed by vigorous shaking and pipetting.

All RNA isolations with the Tri Reagent were performed with Phasemaker tubes (Thermo Fisher, A33248) according to the manufacturer's protocol.

Mouse oocyte and egg collection

For the collection of prophase I-arrested (GV) oocytes, 5- to 8-week-old female mice were superovulated by intraperitoneal injection of 5–10 I.U. of pregnant mare's serum gonadotropin (PMSG) (Lee Biosolutions, 493-10-2.5). At 48 h post injection, the mice were euthanized, and the ovaries were dissected into MEM (Sigma Aldrich, M0268), containing 25 mM HEPES pH 7.3, 0.1 mg/mL sodium pyruvate (Sigma Aldrich, P4562), 3 mg/mL polyvinylpyrrolidone (Sigma Aldrich, P2307), 10 μg/mL gentamicin (Life Technologies, 15710064), and 2.5 μM milrinone (Sigma Aldrich, M4659). Oocytes were collected in 37°C MEM in the presence of milrinone to prevent maturation. Cumulus cells were removed from cumulus-oocyte-complexes by repeated aspiration through a glass pipette. GV oocytes were collected in TRI reagent, and total RNA was isolated according to manufacturer's instructions. Approximately 100 oocytes from multiple animals were collected per biological replicate.

For the collection of metaphase II (MII) oocytes, 5- to 8-week-old female mice were injected with 5–10 I.U. of PMSG. At 48 h after PMSG injection, the animals were injected with 5–10 I.U. of Human chorionic gonadotropin (hCG) (Millipore Sigma, CG5-1VL). At 16 h

post hCG injection, cumulous-enclosed egg complexes were collected from the oviduct and cultured in 37°C MEM with hyaluronidase (3 mg/ml, Sigma Aldrich, H4272) for 2 min. Denuded eggs were then washed free of hyaluronidase, collected in TRI Reagent, and RNA was isolated according to manufacturer's instructions. Approximately 100 MII oocytes from multiple animals were collected per biological replicate.

Sucrose cushion for translational profiling

Frog oocyte or embryo lysate (~200 μl) was laid on top of an ice-cold 1.8-ml sucrose cushion (20 mM HEPES pH 7.5, 100 mM KCl, 5 mM $MgCl_2$, 1% [v/v] Triton X-100, 1 M sucrose, 100 μg/ml cycloheximide, 20 units/ml SUPERase•In) in a polycarbonate tube (Beckman Coulter, 362305). Ribosomes were pelleted by centrifugation at 110,000 rpm (657,000*g*) at 4°C in a Beckman TLA-100 rotor for 1 h. After centrifugation, the supernatant was removed, and 1 ml Tri Reagent was added to the pellet.

Nascent-chain pulldown of ribosome-associated RNAs

Frog oocyte or embryo lysate was incubated with anti-HA magnetic beads (Thermo Fisher, 88837) using 20 μl slurry per 500 μl lysate. The bead mixture was incubated at 4°C for 1 h with end-to-end rotation. Beads were then immobilized using a magnetic stand, and the supernatant was removed. Beads were washed three times with 0.6 ml buffer RL, resuspended in 100 μl buffer RL, mixed with 1 ml Tri Reagent and saved for RNA isolation.

Luciferase assays

Oocyte lysates for luciferase assays were prepared as described.[44] Luciferase assays were performed with Nano-Glo Dual-Luciferase Reporter Assay System (Promega, N1630) in a GloMax Discover plate reader according to the manufacturer's protocol.

3'-UTR variant and tail-length sequencing

In some cases, total RNA containing reporter mRNA libraries was ligated to a pre-adenylated 3' adapter directly, but for the N60 library injected into fish embryos and frog oocytes, the N37-PAS-N17 library injected into fish embryos and frog oocytes, and the CPE^{mos}-N60 library injected into frog oocytes, reporter library mRNAs were first enriched by hybridization with biotinylated oligos. RNA isolated from the oocyte or embryo lysate was mixed with 8 pmol KXSH009, 8 pmol KXSH010, and 2x SSC (0.3 M NaCl, 30 mM sodium citrate pH 7.0) in a total of 50 μl. The RNA and oligos were annealed by incubation at 70°C for 5 min and then slowly cooling to 23°C at 0.1°C/sec. The annealed mixture was combined with 40 μl MyOne Streptavidin C1 beads (Thermo Fisher, 65002) and incubated for 20 min at 23°C on a thermal mixer, shaking with 15 sec on and 1 min 45 sec off. The supernatant was separated from the beads with a magnetic rack and removed. The beads were washed twice with 300 μl 1xB&W buffer (5 mM Tris-HCl pH 7.5, 0.5 mM EDTA, 1 M NaCl) and once with 300 μl 2x SSC. The RNA was eluted from the beads first with 100 μl 10 mM HEPES pH 7.5 at 65°C for 3 min and then with 100 μl water at 65°C for 3 min. The eluates were combined, precipitated with ethanol, and resuspended in 6.5 μl water.

The enriched or unenriched RNA isolated from oocyte or embryo lysates was ligated to a pre-adenylated 3' adapter in a 10 μl reaction containing 5 μM 3' adapter KXS330, 50 mM HEPES pH 7.5, 10 mM $MgCl_2$, 10 mM dithiothreitol, 1 unit/μl T4 RNA ligase 1 (New England Biolabs, M0204S) and incubated at 23°C for 150 min. After ligation, RNA was extracted with phenol/chloroform, precipitated with ethanol, and resuspended in 11.4 μl water. The ligated RNA was mixed with 0.6 μl 100 μM reverse transcription primer KXS037 in a total volume of 12 μl, incubated at 65°C for 5 min, and cooled on ice for 1 min. The annealed RNA was reverse transcribed in a 20 μl reaction containing 1x First-Strand Buffer, 500 μM dNTPs, 5 mM dithiothreitol, 1 unit/μl SUPERase•In, and 200 units SuperScript III (Thermo Fisher, 18080044) incubated at 50°C for 1 h. RNA was then hydrolyzed with addition of 3.3 μl 1 M NaOH and incubation at 90°C for 10 min, followed by neutralization with 36.7 μl 1 M HEPES pH 7.5, and the cDNA was collected by desalting with a Micro Bio-Spin P-30 column. The cDNA library was amplified in a 50 μl PCR reaction with KXS037 and a barcoded primer (KXS057) that hybridized to a constant region of the library, using the KAPA HiFi HotStart Kit following the manufacturer's suggested protocol for 10–15 cycles. The PCR-amplified library was cleaned up twice with AMPure XP beads (Beckman Coulter, A63881) with a beads-to-sample ratio of 1.2.

Sequencing was performed on an Illumina HiSeq 2500, with a standard single-read run of 300 cycles and sequencing primer KXS067 (for the N60-PAS^{mos}, N60, and N37-PAS-N17 libraries) or KXS056 (for the CPE^{mos}-N60 library).

PAL-seq

Sequencing of endogenous mRNA poly(A)-tail lengths was performed with PAL-seq v3 (for frog oocytes) or PAL-seq v4 (for frog embryos, fish embryos, and mouse oocytes) as described.[44] When preparing the sequencing libraries of mRNAs from fish embryos, a different 3' adapter (KXS013) was used for 3'-end ligation, and poly(A)-selected mRNA from HeLa cells was used as spike-in, replacing poly(A)-selected mRNA from zebrafish ZF4 cell line normally used as the spike-in. After the first round of sequencing revealed that many of the reads from the fish mRNA libraries were for 5.8S rRNA, the cDNA of 5.8S rRNA was depleted from the cDNA libraries with an antisense oligo as follows. The seven cDNA libraries made from mRNAs of zebrafish embryos at different stages were mixed at roughly equal molar ratios in a total of 5 fmol, and then 50 pmol KXSH015 and 2x SSC were added in a total of 100 μl. The cDNAs and the oligo were annealed by incubation at 65°C for 5 min and then slowly cooled to 23°C at 0.1°C/sec. The annealed mixture was combined with 100 μl MyOne Streptavidin C1 beads and incubated for 20 min at 23°C on a thermal mixer, shaking with 15 sec on and 1 min 45 sec off. The supernatant was separated from the beads with a magnetic rack. The beads were washed once with 200 μl

1xB&W buffer. The supernatant and the wash were combined, precipitated with ethanol, and resuspended in 30 μl water. After this 5.8S depletion, libraries were sequenced again as technical replicates. Sequencing data of replicates were merged for each sample after confirming consistency.

Genome references and gene annotations

Human (release 25, GRCh38.p7, primary assembly) and mouse (release 10, GRCh38.p4, primary assembly) genomic sequences were downloaded from the GENCODE website (https://www.gencodegenes.org). Sequences of mitochondrial pseudogenes in human and mouse genomes were masked as described.[44] Frog (*X. laevis*) genomic sequences (v10.1 assembly) were downloaded from the Xenbase website (www.xenbase.org). The frog mitochondrial genomic sequence (NC_001573.1) was obtained from the NCBI website (https://www.ncbi.nlm.nih.gov/) and appended to the frog genome. Fish (*D. rerio*) genomic sequences (GRCz11) were downloaded from the Ensembl website (www.ensembl.org).

Human (release 25, GRCh38.p7, primary assembly) and mouse (release 10, GRCh38.p4, primary assembly) gene annotations were downloaded from the GENCODE website. Frog (*X. laevis*) gene annotations (v10.1 assembly) were downloaded from the Xenbase website. Frog mitochondrial gene annotations were curated based on the information obtained from the NCBI website (NC_001573.1) and appended. Fish (*D. rerio*) gene annotations (GRCz11.v97) were downloaded the Ensembl website.

For each species, the gene annotation file was processed for downstream analysis with a custom script in the following steps: 1) non-protein-coding annotations were removed; 2) annotation entries without "gene_id" or "transcript_id" were removed; 3) annotation entries (either exon or CDS) of a transcript that overlapped were merged; 4) for each gene, transcript isoforms without CDS annotations were removed if at least one other isoform had CDS annotations, and transcript isoforms without exon annotations were removed if at least one other isoform had exon annotations. The processed gene annotations were referred to as "multi-isoform gene annotations". The multi-isoform gene annotations for each species were further processed to generate a file referred to as "single-isoform gene annotations", in which a "main" transcript isoform was annotated for each gene by the following tie-breaker: 1) the isoform with both exon and CDS annotations; 2) the isoform with the most number of exons (only for frog annotations); 3) the isoform that was the longest (summing all exons); 4) the isoform with the longest CDS.

Annotation of mRNA 3'-end isoforms

PAL-seq tags were used to annotate mRNA 3' ends because they provided information on cleavage and polyadenylation sites. Software HOMER[83] was used to call peaks in the "tss" style, with PAL-seq data as the "Test" group and RNA-seq data as the "Control" group, and the following parameters "-style tss -strand separate -fdr 0.001 -ntagThreshold 10 -size 41 -center". Uniquely mapped reads in same cell type of the same species were merged for both PAL-seq data and RNA-seq data. Specifically, the following datasets were used for the input: 1) frog oocytes/embryos: PAL-seq data for prophase I-arrested oocytes (stage VI) at 0, 1, 3, 5, 7, and 9 h post progesterone treatment (Test) and RNA-seq data for prophase I-arrested oocytes (stage VI) oocytes[44] (Control); 2) fish embryos: PAL-seq for embryos at 0, 1, 2, 3, 4, 5, and 6 h post fertilization (Test) and RNA-seq data for embryos at 0, 1, 2, 3, 4, 5, 6, 7, and 8 h post fertilization[98] (Control); 3) mouse oocytes: PAL-seq data for GV and MII oocytes (Test) and RNA-seq data for GV, MI, and MII oocytes[12] (Control). 4) human oocytes: the 3'-end coverage of the PacBio-based long-read sequencing data from PAIso-seq in human GV oocytes[31] (Test) and RNA-seq data from GV oocytes[62] (Control).

A custom script was used to filter peaks and associate each of them with an mRNA isoform of a gene in the following steps: 1) the center position of each peak was extracted as the UTR 3' end; 2) peaks on the mitochondrial chromosome were removed; 3) *bedtools* (v2.26.0)[84] *intersect* was used with the "single-isoform gene annotations" to find the mRNA isoform that uniquely overlapped with each peak, and peaks that overlapped with more than one mRNA were removed; 3) for peaks that didn't overlap with any mRNAs in step 2, *bedtools intersect* was used with the "multi-isoform gene annotations" to find the gene that uniquely overlapped each peak, and peaks that overlapped with more than one gene were removed. In cases in which a peak overlapped with more than one mRNA isoform of the same gene, the mRNA isoform with the longest exon that overlapped with the peak was chosen; 4) for peaks that had not overlapped with any mRNAs in previous steps, *bedtools closest* was used with the "single-isoform gene annotations" to find the closest exon of an mRNA isoform upstream of each peak.

After filtering peaks and associating each of them with an mRNA isoform, a "high" or "low" confidence tag was assigned to each peak for further filtering in downstream analysis based on the following criteria: 1) if a peak overlapped with the last exon of its associated gene, it was labeled as "high" confidence; 2) if a peak was downstream the last exon of its associated gene, a distance between the peak and the last exon was calculated. If this distance was smaller than 500 nt, the peak was labeled as "high" confidence. If this distance was larger than 10,000 nt, the peak was labeled as "low" confidence. 3) for the remaining peaks, a density-ratio test was performed to determine the confidence level of each peak. To perform this test, sequencing-depth-normalized RNA-seq read densities were calculated for two regions of each peak. The first region was the exon immediately upstream of the peak (partial exon if the peak and the exon overlapped or full exon if they didn't overlap). The second region was 150 nt immediately upstream of the peak. The ratio of the read densities of these two regions was calculated for all peaks (including those with a labeled confidence tag). A reference distribution of the ratio values for all peaks that overlapped the last exons of genes was obtained. For each remaining peak without a labeled confidence tag, if the ratio value for the peak was more than 1.5 inter-quantile range (IQR, difference between values at 25th percentile and 75th percentile) above the 75th percentile or less than 1.5 IQR below the 25th percentile of the ratio values of the reference distribution, it was labeled as "low" confidence. Otherwise, it was labeled as "high" confidence. Annotations of mRNA 3'-end isoforms used in this study are provided in Table S2.

3'-UTR annotations and sequences

For each organism, mRNA 3'-UTR annotations were curated with a custom script from the "muti-isoform gene annotations" and the "mRNA 3'-end isoform annotations" generated in this study. For each mRNA 3' end in the "mRNA 3'-end isoform annotations", if the 3'-UTR annotation of the 3'-end-associated mRNA isoform was available in the "multi-isoform gene annotations", the 3'-UTR annotation was extracted and the genomic coordinate of the 3' end was updated from the information in the "mRNA 3'-end isoform annotations". If the 3'-UTR annotation of the 3'-end-associated mRNA isoform was unavailable, the 3'-UTR annotation was derived in one of the following ways: 1) if both the exon and CDS annotations were available, the exons downstream of the last nucleotide of the CDS were extracted as the 3' UTR and the genomic coordinate of the 3' end was updated from the information in the "mRNA 3'-end isoform annotations"; 2) if the CDS annotation was available but the exon annotation was unavailable, the genomic region between the last nucleotide of the CDS and the 3 ' end in the "mRNA 3'-end isoform annotations" was used as the 3' UTR. If the 3' end of the curated 3' UTR had a confidence label "low" in the "mRNA 3'-end isoform annotations," then it was assigned the label "uncertain," which was used for filtering low-confidence 3' UTRs in downstream analyses. For mRNA 3' ends with 3' UTRs that could not be annotated as described above because either 1) the 3' end was upstream of all 3'-UTR annotations; 2) the 3' end was upstream of the last nucleotide of the CDS, or 3) neither CDS nor 3'-UTR annotations was available, a 2000-nt genomic region upstream the 3' end was used as the putative 3' UTR and this 3' UTR was assigned the label "uncertain". The curated mRNA 3'-UTR annotations are provided in Table S3.

Sequences corresponding to the mRNA 3'-UTR annotations were extracted from genomic sequences using *bedtools getfasta*. Sequence fragments of each mRNA 3' UTR were stitched together with a custom script. Unless indicated, all "uncertain" 3' UTRs were excluded from downstream analyses.

PAL-seq data analysis

PAL-seq v3 reads were trimmed with *cutadapt* (v3.7)[85] with the parameters "-m 15 –quality-base=64 -q 20,20 –match-read-wildcards -e 0.05 -a NNNNATCTCGTATGCCGTCTTCTGCTTG -O 7". The trimmed reads were mapped using *STAR* (v2.7.1a)[86] to the reference database containing the genomic sequences of the organism from which the mRNAs were obtained, the genomic sequences of humans or fish, depending on which spike-in RNAs were used, and the sequences of the RNA poly(A)-tail-length internal standards[9] that had also been added to each sample, with the parameters "–runThreadN 16 –runMode alignReads –outFilterMultimapNmax 1 –outReadsUnmapped Fastx –outFilterType BySJout –outSAMattributes All –outSAMtype BAM Unsorted SortedByCoordinate". Uniquely mapped reads were intersected with a database containing mRNA 3'-end annotations of the organism from which the mRNAs were obtained, the single-isoform gene annotations of humans or fish, depending on which spike-in RNAs were used, and 3'-end annotations of the poly(A) standards, using *bedtools intersect* with the parameters "-wa -wb -bed -s" and retaining reads that corresponded to a unique mRNA isoform. For each library, 10% of reads that intersected with the spike-in RNA annotations (but no more than 50,000 and no less than 5,000) were randomly picked as the training set for determining the Hidden Markov Model (HMM).

To determine the poly(A)-tail length, a T-signal value $S_{i,j}$ was calculated for each sequencing cycle i of read 2 of each read cluster j by the following formula:

$$S_{i,j} = \max\left(-5, min\left(5, \log_2 \frac{I'^{T}_{i,j}}{I'^{A}_{i,j}+I'^{C}_{i,j}+I'^{G}_{i,j}}\right)\right)$$

where $I'^{N}_{i,j}$ is the normalized intensity of channel N (A, C, G, or T) at cycle i of read 2 of cluster j defined by the following formula:

$$I'^{N}_{i,j} = \frac{f\left(I^{N}_{i,j,2}\right)}{\bar{I}^{N}_{j,1}}$$

where $I^{N}_{i,j,2}$is the raw intensity of channel N at cycle i of read 2 of cluster j and the function f is defined by the following:

$$f(x) = \begin{cases} \min\left(|x|, \left|\frac{1}{x}\right|\right), x<0 \\ x, x \geq 0 \end{cases}$$

and $\bar{I}^{N}_{j,1}$ is the average intensity of channel N of read 1, defined by

$$\bar{I}^{N}_{j,1} = \frac{\sum_{k \in k'} I^{N}_{k,j,1}}{\sum_{k \in k'} 1}$$

where $I^{N}_{k,j,1}$is the raw intensity of channel N at cycle k of read 1 of cluster j, and k' indicates a set of cycles in the range of 11 to 35 where a base called by the sequencer is the same as the identity of the channel N in read 1 of cluster j. For cases in which $S_{i,j}$ could not be calculated because the normalization of a channel could not be performed or $I'^{T}_{i,j}$ was 0, $S_{i',j}$ from neighboring cycles $(i-5 \leq i' \leq i+5)$ were averaged to infer $S_{i,j}$. If a cluster had more than 5 cycles for which $S_{i,j}$ values were not available, it was discarded.

A five-state mixed Gaussian Hidden Markov model (from the Python *ghmm* package) was trained on the T-signal values of the training set, and the trained model was used to decode the sequence of the states from the T-signal values of the entire dataset. The initialization parameters and the process of determining poly(A)-tail lengths from decoded states were as described in previous PAL-seq data analyses.[44]

PAL-seq v4 data analysis was performed similarly to PAL-seq v3, with the following differences. First, the k'set used to obtain normalized intensity $\bar{I}^N_{j,1}$ included cycles in the range of 21–50, due to the longer length of read 1 in PAL-seq v4. Second, the poly(A)-tail length was determined by the length of the longest consecutive states that started before cycle 3 and were in the state 1 or 2. Third, for the fish embryo datasets, a different transition matrix was initialized for the HMM to allow backward state transitions:

$$\begin{bmatrix} 0.04 & 0.93 & 0.02 & 0.01 & 0.0 \\ 0.0 & 0.87 & 0.1 & 0.02 & 0.01 \\ 0.0 & 0.05 & 0.6 & 0.3 & 0.05 \\ 0.0 & 0.01 & 0.3 & 0.6 & 0.09 \\ 0.0 & 0.01 & 0.01 & 0.1 & 0.88 \end{bmatrix}$$

Reporter 3'-UTR variant and tail-length analysis

For the N60-PAS^{mos} library, sequencing reads from the FASTQ file were first examined for the expected constant sequence AATAAAGAAATTGATTTGTCT at position 61–81, allowing positional offsets between –3 and +3 nt. Reads with a 21-nt sequence that had no more than six mismatches to the constant sequence at any allowed positions were retained. For each retained read, the sequence preceding the constant sequence was recorded as the 3' UTR, and the region after this segment was used to determine the poly(A)-tail length. The retained reads were examined for the presence of the 3'-adapter sequence TCGTATGCCGTCTTC TGCTTG with 4 allowed mismatches. If the adapter sequence was identified for a read and the position of the adapter was before the starting position of the poly(A) tail, this read was discarded. If the 3' UTR of a retained contained an "N" nucleotide or if the lowest quality score of the 3' UTR was not larger than 10, this read was discarded. One percent of the retained reads (no more than 50,000) were randomly picked as the training set for determining the Hidden Markov Model (HMM).

The poly(A)-tail length was determined similarly to that in the PAL-seq v4, with the following differences: 1) an A-signal value $S_{i,j}$ was calculated for each sequencing cycle i of each read cluster j by swapping the A and T channels in the formula for the T-signal value; 2) the 3' UTR replaced read 1 in the PAL-seq analysis and the entire region was used to determine the average intensity of each channel; 3) the poly(A)-tail region defined by the constant-sequence segment replaced read 2 in the PAL-seq analysis (cycle 1 was the first cycle after the constant-sequence segment); 4) if a cluster had more than 50 cycles in which $S_{i,j}$ values were not available, it was discarded; 5) the transition matrix was initialized as it was used for the fish embryo datasets, allowing backward transitions between HMM states; 6) the poly(A)-tail length was determined by the length of the longest consecutive states that started before cycle 5 and were in the state 1 or 2. Due to the high sequence complexity of each library, most 3'-UTR sequences of these complex libraries were measured only once; for those that were measured more than once, the median value of poly(A)-tail lengths was used.

Analysis of the CPE^{mos}-N60 library was performed similarly to that of the N60-PAS^{mos} library, except no read-filtering was performed based on the constant region, because in contrast to the N60-PAS^{mos} library, this library did not contain a constant region in the sequencing read. Instead, the entire region of the first 60 nt was used as the 3' UTR, and for determining the poly(A)-tail length, the start of the poly(A) region was set at cycle 61. Sequences of some 3'-UTR variants of this library had strong similarities to the last 50 nt of the coding region of the mRNA reporter or to the 5.8S rRNA. These were presumably sequencing artifacts or contaminants. To exclude them, all 3'-UTR sequences of this library were aligned to the last 50 nt of the *NanoLuc* sequence and the 5.8S rRNA using the Python package *Biopython*,[93] and those that had an alignment score higher than 25 with the *NanoLuc* sequence or an alignment score higher than 30 with the 5.8S RNA were removed.

Analysis of the N60 and N37-PAS-N17 libraries was performed similarly to that of the CPE^{mos}-N60 library. Because these two libraries were mixed for co-injection and had the same 3'-UTR length, their sequencing results were processed together until the last step, at which the N60 and N37-PAS-N17 libraries were separated based on whether the sequence at positions 38–43 matched a PAS (AATAAA).

K-mer-associated tail length and tail-length change of reporter mRNAs

To obtain statistics (mean, median, and standard deviation) of poly(A)-tail length associated with a *k*-mer in an mRNA reporter library dataset, all 3'-UTR variants that contained the *k*-mer within their variable region were counted, and each unique 3'-UTR variant contributed equally to tail-length statistics. When examining the positional effect of a *k*-mer, only variants with a *k*-mer at a specific position were considered. The position of a *k*-mer was determined by its position relative to the 3' end of the variable region for the N60-PAS^{mos} library, or to the 5' end of the variable region for the other three libraries. When examining the effect of the co-presence of two or more *k*-mers, only variants containing the specified *k*-mers at non-overlapping positions were considered. When examining the effect of flanking nucleotides of a *k*-mer, only variants containing the *k*-mer and an indicated flanking nucleotide at the indicated position relative to the *k*-mer regardless of the position of the kmer, were considered.

To examine the tail-length change associated with a *k*-mer between two datasets, the difference was calculated between the two mean values of the tail length associated with the *k*-mer in each dataset. When indicated, a one-sided Welch's t-test was performed with the calculated means, standard deviations, and sample sizes associated with the *k*-mer against a reference value, either 0 or the

difference of mean tail lengths of all 3'-UTR variants in the two datasets compared. A Z-score of tail-length change was calculated for each *k*-mer with respect to all *k*-mers of the same length.

In some cases, *k*-mer-associated tail-length differences (*k*-mer of the same length) were examined in an iterative exclusion process. In each round of the iteration, mRNA variants with the variable region containing any *k*-mer in the *k*-mer exclusion list (starting with an empty list in the first round) were excluded from the analysis. The average tail-length difference was then calculated for all *k*-mers that were not in the exclusion list. The result of the *k*-mer with the largest tail-length increase (or decrease) was recorded, and this motif was added to the exclusion list before starting the next round. The iteration ended when either there was no *k*-mer motif associated with statistically significant increased (or decreased) tail length (Bonferroni correction-adjusted $P < 0.05$ in a Welsh's *t*-test) or no *k*-mer motif had a Z-score of the tail-length difference higher than 2 (or lower than –2 when tail-length decrease was examined).

Sequence motifs as binary predictors of endogenous cytoplasmic polyadenylation targets

The tail-length change was compared between the median tail length at 0 h post progesterone and that at 7 h post progesterone for endogenous mRNAs in frog oocytes. Only mRNA isoforms that met the following criteria were included in the analysis: 1) the length of the 3' UTR was longer than 10 nt; 2) the 3' UTR was not labeled as ''uncertain'' in the annotations; 3) the 3' UTR contained a PAS (AAUAAA or AUUAAA) within the last 150 nt of the 3' end, and 4) the mRNA isoform had more than 50 poly(A) tags in both datasets that were compared. For a given threshold of tail-length difference (ranging from 0–10 nt), mRNA isoforms with tail-length changes larger than the threshold were treated as true positives. The classification threshold was the negative value of the distance *d* to the 3' end. When a *k*-mer or a *k*-mer group was examined as the binary predictor for tail-length changes, for a given classification threshold $-d$ $(d > 0)$, an mRNA isoform was considered positive if this *k*-mer or one of the *k*-mers in the *k*-mer group was found within the last *d* nt of the 3′ UTR of this mRNA isoform ($\geq -d$). Values of true positive rate (TPR) and false positive rate (FPR) were calculated at all possible classification thresholds using the Python *scikit-learn* package[88] and were used to make the receiver operating characteristic curve (ROC). The area-under-curve (AUC) values were also calculated with the Python *scikit-learn* package.[88]

Motif logos associated with tail-length changes

Motif logos were made with 8-mers associated with tail-length changes. *k*-mers were filtered with indicated cutoff values for the tail-length difference, the Z-score of the tail-length difference, and the Bonferroni correction-adjusted *P* value (Welsh's *t*-test for tail-length changes). The weights used for producing motif logos were either tail-length changes or Z-scores, as indicated. Retained *k*-mers were ranked by the weight in a descending (when motifs associated with tail-length increase were examined) or ascending order (when motifs associated with tail-length decrease were examined). In an iterative procedure starting with the top-ranked *k*-mer, each *k*-mer was aligned to all *k*-mers in a seed list (starting with an empty list), except for the first *k*-mer, which was added to the seed list without any alignments. All alignments were performed without allowing gaps. Alignment scores were calculated for all possible alignment positions between the *k*-mer and all seeds, adopting the following scoring scheme: 1 for a match, –1.5 for a mismatch, –1 for a position offset. The alignment with the highest score (must be higher than –0.1) was kept and the weight was assigned to this *k*-mer-seed alignment. If the highest score was tied among multiple alignments, either between the *k*-mer and the same seed at different positions or between the *k*-mer and different seeds, all tied alignments were kept and the weight was evenly distributed among these alignments. If none of the alignments had a score higher than –0.1, this *k*-mer was considered not aligned and it was added to the seed list. This iterative procedure stopped when all retained *k*-mers had been added to the seed list or had been aligned.

A position weight matrix (PWM) was calculated for each *k*-mer seed in the seed list based on the weights of the seed and all the *k*-mers aligned to it. For each position of the alignments, the weight of each nucleotide contributed by each *k*-mer was summed. If a *k*-mer had no nucleotide aligned at this position, the weight was evenly distributed between A, C, G, and U nucleotides. The summed weights of nucleotides at each position were normalized to the total weights of that position. Motif logos were generated from the resulting PWMs with the R package ggseqlogo.[89]

For the N60-PAS^{mos} library in frog oocytes, *k*-mer-associated tail-length difference was calculated in an iterative exclusion process. We found this process was computationally heavy and the motif logos generated from this result were similar to those generated from *k*-mer-associated tail-length difference not calculated in an iterative exclusion process. As a result, for all other datasets, *k*-mer-associated tail-length differences used to generate motif logos were not calculated in an iterative exclusion process.

In Figure 1E, tail-length changes were examined in an iterative process. A cutoff of Z-score ≥ 3 and tail-length difference ≥ 2 was applied. Tail-length changes were used as weights. In Figures 3C, S3G, and S3H, A cutoff of Z-score ≥ 3 (for tail lengthening) or Z-score ≤ -3 (for tail shortening) was applied. Z-scores were used as weights.

Structural accessibility and tail lengths

The structural accessibility was predicted for each 3'-UTR variant with the entire 3'-UTR sequence (including the variable region and the constant region) using RNAfold[90] with the parameters ''–no PS –filename-full -p1''. Both the minimum of free energy (MFE) of the 3' UTR and the base-pairing probability of each base were obtained from the output. The unpairing probability of a motif was the geometric mean of the unpairing probability of each nucleotide of this motif, defined by

$$P_{unpairing,motif} = \sum_{N \in motif}^{1} \sqrt{\left(\prod_{N \in motif} (1 - P_{pairing,N}) \right)}$$

where $P_{pairing,N}$ is the base-pairing probability of the nucleotide N in the motif.

When examining the effect of structural accessibility on tail-length changes for the N60-PAS^{mos} library during frog oocyte maturation, only variants containing a single CPE (UUUUA) were considered. Tail lengths of variants at 7 h post-progesterone treatment were used as a proxy for assessing the tail-length changes from 0 h to 7 h, because most variants were not observed in both time points due to the sequence complexity of the library. In addition, because most variants had initial tail lengths of 35 nt at 0 h post-progesterone treatment, the end-point tail lengths reflected tail-length changes. To minimize the potential confounding effect of GC-content of the variant sequence, variants containing a CPE (UUUUA) at indicated positions were divided into 20 equal-sized groups, based on their GC-contents, and the correlation coefficients between the tail length and a structural accessibility metric of interest (MFE or unpairing probability of a motif) were calculated for all the variants within each group.

Motif-associated tail-length change for endogenous mRNAs

Tail-length changes were calculated for all mRNA isoforms with unique 3' ends by comparing their median tail lengths measured at two different developmental stages, requiring no fewer than the indicated number of poly(A) tags at both stages. To examine the tail-length change associated with a motif, mRNAs were divided into two groups based on whether they contained the motif of interest within their 3' UTRs. When specified, the two groups were subsetted to match the initial tail-length distribution and the 3'-UTR length distribution using the R software package MatchIt[91] with parameters "distance = 'glm', method = 'cem', k2k = TRUE". Tail-length changes of mRNAs from the two groups or the two subsetted groups were used to make cumulative distribution plots.

Enrichment analysis of *k*-mer-associated translational efficiency

k-mer-associcated translational efficiency E^k is defined as:

$$E^k = log_2\left(\frac{R_p^k}{R_i^k}\right)$$

where R_p^k is the fraction of mRNA variants in the library that contained the *k*-mer (or multiple *k*-mers when their co-occurrence was evaluated) in the pellet (for sucrose cushion) or eluate (in the case of nascent-chain pulldown), and R_i^k is the fraction of mRNA variants in the library that contained the *k*-mer in the input. R_p^k and R_i^k are defined as

$$R_p^k = \frac{C_p^k + 1}{C_p^{all} + 1}$$

$$R_i^k = \frac{C_i^k + 1}{C_i^{all} + 1}$$

where C_i^k , C_p^k represent the numbers of mRNA variants that contained the *k*-mer in the input and the pellet respectively. C_i^{all} and C_p^{all} represent the number of total mRNA variants in the input and the pellet respectively. When indicated, a binomial test was performed between R_p^k and R_i^k by assuming R_i^k as the expected success probability, $C_p^{all} + 1$ as the total number of trials, and $C_p^k + 1$ as the number of successes.

Measurement of poly(A)-tail lengths in mouse and human oocytes

Poly(A) tail-length measurements by PAL-seq in mouse GV oocytes and MII oocytes were merged for each pair of replicates, after confirming reproducibility. The tail-length change for each mRNA 3'-UTR isoform was calculated between the median tail length in GV oocytes and that in MII oocytes, requiring $\geq$ 50 poly(A) tags in each of the two merged datasets.

For human oocytes, PAIso-seq[31] bam files were downloaded from Genome Sequence Archive for Human (GSA-Human: https://ngdc.cncb.ac.cn/gsa-human/) with the accession number HRA001911. The reverse-complement sequence of the soft-clipped region at the 5' end of each read was extracted from the bam file. The poly(A)-tail length was determined by the length of the longest consecutive As (allowing individual non-A nucleotides within the stretch) that started before the third nucleotide of the extracted sequence. The tail-length change for each mRNA 3'-UTR isoform was calculated as the median tail length in GV oocytes compared to that in MII oocytes, requiring $\geq$ 50 poly(A) tags in each dataset.

Tail-length changes of homologous genes in frogs, mice, and humans

The frog-to-human gene ID conversion table was downloaded from Xenbase. The mouse-to-human gene ID conversion table was downloaded from the Ensembl website. Measured tail-length changes were compiled for human MII and GV oocytes, mouse MII and GV oocytes, and frog oocytes 0 and 7 h post-progesterone treatment. If a gene had multiple mRNA isoforms, the dominant isoform (> 50% by tag count in GV oocytes of humans and mice or frog oocytes 0 h post-progesterone treatment) was chosen to represent

that gene. If more than one homolog of a human gene was identified in another species (mice or frogs), the mean tail-length change of the homologs was used.

Relationship between poly(A)-tail length and TE in human and mouse oocytes

TEs measured by ribosome footprint profiling and mRNA-seq were obtained for mouse oocytes from a published study,[12] and for human oocytes from another published study.[62] Reads were trimmed as described.[12,62] The trimmed reads were mapped using *STAR* (v2.7.1a)[86] to the reference database containing the genomic sequences of mouse or human, with the parameters –runThreadN 20 –runMode alignReads –outFilterMultimapNmax 1 –outFilterType BySJout –outSAMattributes All –outSAMtype BAM SortedByCoordinate –outReadsUnmapped None''. Mapped reads were counted for each gene with htseq-count (1.99.2).[87] For both ribosome profiling and mRNA-seq, only reads that uniquely mapped to coding regions of annotated genes (excluding the first 15 codons and the last five codons) were included in downstream analyses. Data from technical replicates were merged. For TE analyses, an expression cutoff of 30 mRNA-seq reads was applied for each gene, with no cutoff for ribosome-footprint reads. TEs from biological replicates were averaged. For both mouse and human oocytes, mRNA 3'-UTR isoforms were required to have $\geq$ 50 poly(A) tags in each dataset for calculating tail-length changes. If a gene had more than one 3'-UTR isoform, then any isoform supported by $\geq$ 90% of the poly(A) tags mapping to that gene was chosen to represent that gene, whereas if no isoform was supported by $\geq$ 90% of the poly(A) tags, then no isoform of that gene was carried forward for analysis.

QUANTIFICATION AND STATISTICAL ANALYSIS

Graphical representation and statistical analysis were performed in *R* (v4.1, https://www.r-project.org). Welch's *t*-tests were used to compare the average tail-length change of mRNAs in a library containing a *k*-mer against the average tail-length change of all mRNAs. Binomial tests were used to compare fractions of mRNAs containing a *k*-mer in two different samples. Mann–Whitney U tests were used to compare tail-length-change distributions of two groups of endogenous mRNAs. Student's *t*-tests were used to compare two means of normalized luciferase signals. For sequencing analysis, standard errors were obtained from all independent mRNA variants in a group of interest in one experiment. For luciferase assays, a group of 6 injected oocytes was used as a biological replicate. Details of number of replicates, dispersion and precision measures and exact statistical tests can be found in the main text, figure legends, and STAR Methods.