nature research

Peer Review Information

Journal: Nature Structural and Molecular Biology
Manuscript Title: Decoding mRNA translatability and stability from 5'UTR
Corresponding author name(s): Prof. Shu-Bing Qian

Reviewer Comments & Decisions:

Decision Letter, initial version:

10th Mar 2020

Dear Shu-Bing,

Thank you again for submitting your manuscript "Decoding mRNA translatability and stability from 5'UTR". We have now received comments from the three reviewers who evaluated your paper (appended below). In light of those reports, we remain interested in your study and would like to invite you to respond to the comments of the referees, in the form of a revised manuscript.

You will see that while all reviewers appreciate the approach, reviewer #3, an expert in mRNA quality control, is hesitant about the advance provided and finds the conclusions mainly confirmatory. Given the more positive assessment of reviewer #1 and #2, we are willing to overrule this concern, but a more comprehensive screen for decay factors suggested by reviewer #1 should be considered and the work needs to be presented within the full context of the current literature, toning down some of the initial assertions that little is known about the role of 5' UTR sequences on mRNA translation and decay. In addition, all technical question and concerns raised by the three reviewers must be fully resolved and caveats should be addressed experimentally, or disclosed and discussed in the text, if experimental approaches are not feasible.

Please be sure to address/respond to all concerns of the referees in full in a point-by-point response and highlight all changes in the revised manuscript text file. If you have comments that are intended for editors only, please include those in a separate cover letter.

We are committed to providing a fair and constructive peer-review process. Do not hesitate to contact us if there are specific requests from the reviewers that you believe are technically impossible or unlikely to yield a meaningful outcome.

We expect to see your revised manuscript within 6-8 weeks. If you cannot send it within this time,

please contact us to discuss an extension; we would still consider your revision, provided that no similar work has been accepted for publication at NSMB or published elsewhere.

As you already know, we put great emphasis on ensuring that the methods and statistics reported in our papers are correct and accurate. As such, if there are any changes that should be reported, please submit an updated version of the Reporting Summary along with your revision.

Please follow the links below to download these files:

Reporting Summary:
https://www.nature.com/documents/nr-reporting-summary.pdf

Please note that the form is a dynamic 'smart pdf' and must therefore be downloaded and completed in Adobe Reader.

Please note that all key data shown in the main figures as cropped gels or blots should be presented in uncropped form, with molecular weight markers. These data can be aggregated into a single supplementary figure item. While these data can be displayed in a relatively informal style, they must refer back to the relevant figures. These data should be submitted with the final revision, as source data, prior to acceptance, but you may want to start putting it together at this point.

SOURCE DATA: we urge authors to provide, in tabular form, the data underlying the graphical representations used in figures. This is to further increase transparency in data reporting, as detailed in this editorial (http://www.nature.com/nsmb/journal/v22/n10/full/nsmb.3110.html). Spreadsheets can be submitted in excel format. Only one (1) file per figure is permitted; thus, for multi-paneled figures, the source data for each panel should be clearly labeled in the Excel file; alternately the data can be provided as multiple, clearly labeled sheets in an Excel file. When submitting files, the title field should indicate which figure the source data pertains to. We encourage our authors to provide source data at the revision stage, so that they are part of the peer-review process.

Data availability: this journal strongly supports public availability of data. All data used in accepted papers should be available via a public data repository, or alternatively, as Supplementary Information. If data can only be shared on request, please explain why in your Data Availability Statement, and also in the correspondence with your editor. Please note that for some data types, deposition in a public repository is mandatory - more information on our data deposition policies and available repositories can be found below:
https://www.nature.com/nature-research/editorial-policies/reporting-standards#availability-of-data

We require deposition of coordinates (and, in the case of crystal structures, structure factors) into the Protein Data Bank with the designation of immediate release upon publication (HPUB). Electron microscopy-derived density maps and coordinate data must be deposited in EMDB and released upon publication. Deposition and immediate release of NMR chemical shift assignments are highly encouraged. Deposition of deep sequencing and microarray data is mandatory, and the datasets must be released prior to or upon publication. To avoid delays in publication, dataset accession numbers must be supplied with the final accepted manuscript and appropriate release dates must be indicated at the galley proof stage.

While we encourage the use of color in preparing figures, please note that this will incur a charge to

partially defray the cost of printing. Information about color charges can be found at http://www.nature.com/nsmb/authors/submit/index.html#costs

Nature Structural & Molecular Biology is committed to improving transparency in authorship. As part of our efforts in this direction, we are now requesting that all authors identified as 'corresponding author' on published papers create and link their Open Researcher and Contributor Identifier (ORCID) with their account on the Manuscript Tracking System (MTS), prior to acceptance. This applies to primary research papers only. ORCID helps the scientific community achieve unambiguous attribution of all scholarly contributions. You can create and link your ORCID from the home page of the MTS by clicking on 'Modify my Springer Nature account'. For more information please visit please visit www.springernature.com/orcid.

Please use the link below to submit your revised manuscript and related files:

[REDACTED]

Note: This URL links to your confidential home page and associated information about manuscripts you may have submitted, or that you are reviewing for us. If you wish to forward this email to co-authors, please delete the link to your homepage.

We look forward to seeing the revised manuscript and thank you for the opportunity to review your work.

With kind regards,
Anke

Anke Sparmann, PhD
Senior Editor
Nature Structural and Molecular Biology
ORCID 0000-0001-7695-2049

Referee expertise:

Referee #1: epigenetic gene regulation, ncRNAs

Referee #2: gene regulation, single-cell sequencing and massively parallel reporter assays

Referee #3: mRNA quality control

Reviewers' Comments:

Reviewer #1:
Remarks to the Author:

In this study, the authors address several important questions about the links between mRNA translation and decay, and the role of uORFs. They use an innovative method involving a library of reporter mRNAs with a random 10-mer that can potentially drive uORF translation upstream of a canonical GFP ORF. By directly transfecting the mRNA library into cells, then measuring translation of the uORF or main ORF (via monosome/polysome fractionation) or mRNA stability, they are able to relate these two processes. This demonstrates that main ORF translation tends to stabilize mRNAs, whereas uORF translation is destabilizing. Furthermore, they discover sequence elements that destabilize mRNAs through inhibiting translation (GGC and RG4 motifs).

A major advantage of directly transfecting mRNAs (rather than reporter plasmids) is that chemically modified transcripts can be used. The authors exploit this to look for sequence elements that enable translation when the mRNA has a non-functional cap analogue (ApppG). They find that A-rich elements enable cap-independent translation. As the ApppG-capped mRNAs are not translated, the authors also use these mRNAs to investigate decay in the absence of translation. This reveals that A-rich elements also promote mRNA decay, which the authors suggest depends on PABP.

Overall, this is a powerful approach to explore links between mRNA translation and decay, that could potentially be extended to address many other questions (e.g. how do sequence elements behave in different 5'UTRs, or when placed at different distances from the main ORF; to what extend RNA modifications influence translation, etc). The authors go some way towards addressing the mechanisms by which A-rich and G-rich elements modulate translation/decay. There are some caveats to this approach (e.g. mRNAs directly transfected into cells may not behave the same as mRNAs that are synthesized by transcription in the cells). Addressing these experimentally or by discussing them more fully in the text would help further strengthen this study.

Specific points:
Figure 1B: The authors conclude that a suboptimal uORF translation initiation (TI) context increases GFP levels – however, the FACS plots do not obviously show an increase in GFP levels (just a reduction in 25D1 levels). The authors should show a histogram of GFP levels for optimal and suboptimal uORF TI context reporters if they wish to make this point.

Figure S2B: The authors should present a few more metrics describing the mRNA reporter library. In particular, how many unique 10-mers are actually present (either in the oligo pools, or in the sequenced RNAs)? How many copies of each 10-mer are typically detected? And how reproducible are these experiments (e.g. do the enrichments in Figure S2D or 1C look similar for two independent experiments?).

Figure 2A: The authors generally measure decay rates of mRNAs directly transfected into cells (rather than mRNAs that have been synthesized by transcription in the cells). They should show that the decay rates are equivalent, at least for a few examples. This could be done by measuring decay rates of mRNAs synthesized from plasmids following transcription shut-off (as in Fig S4D), and comparing to decay rates of mRNAs directly transfected into cells (potentially also after ActD treatment, to ensure conditions are comparable).

Figure 2A: The authors rely on the assumption that in vivo decay rates in the absence of translation can be measured from in vitro decay experiments. However, they do not show that these decay rates are equivalent. One way in which this could be done is to compare the in vivo and in vitro decay rates of sequence variants that are found in ribosome-free fraction (i.e. those in Fig 3A), as these should

not be translated. This would check that in vivo and in vitro decay rates are equivalent, when the effect of translation is removed.

Figure S3C: The authors validate the low vs high GFP translation of monosome and polysome associated sequence variants by comparing two with an ATG, and two without. However, they do not show a comparison for ATG variants in optimal versus suboptimal sequence contexts, which would help validate the ability of their approach (Figure S3A/B) to measure the contribution of features beyond the start codon.

Line 150: The authors use a non-functional cap analogue ApppG to measure mRNA decay in the absence of translation. Could they please discuss whether this cap analogue is also likely to interfere with mRNA decay (e.g. decapping?).

Line 168: The authors state that "many sequences bearing the GGC-motif coincided with computationally predicted RG4 structures" – please could they quantify this statement? Are GC-rich motifs not predicted to form RG4 structures also enriched in the ribosome-free fraction? It would also be helpful to show the sequence context into which the 10-mer is placed, as in order for the 10-mer to contribute to RG4s, the surrounding sequence must also contain at least one GG.

Line 178: The authors claim to have examined the role of decay factors systematically. This is not quite true, because not all possible decay pathways were considered. This statement should therefore be toned down and it should be discussed that other factors/pathways may have been missed.
In fact, a more comprehensive screen for decay factors would strengthen this paper and the authors may want to consider this for a revised version of the manuscript.

Figure 3F: The QUMA-1 FACS signal suggests that there are two cell populations – those with, and those without, RG4s. However, the cells will contain many copies of the reporter mRNA, and in addition, many endogenous transcripts that can also form RG4s. Can the authors offer any explanation why the QUMA-1 signal distribution is bimodal, rather than continuous? Furthermore, the high QUMA-1 signal seems to be heavily dependent on the reporter mRNA (P1 vs N1) – is this surprising, given that the cells will also contain many other quadruplex-forming mRNAs? Perhaps the authors could comment on this.

Figure 3I: It would be helpful to have non-DHX36-knockdown FACS plots alongside (or at least to know whether these data were obtained in the same experiment as those for Figure 3F).

Figure 3G: The authors suggest that QUMA-1 signal is inversely correlated to overall GFP intensity, and draw a straight line on the graph. As there are no intermediate GFP data points, this is a bit of a stretch.

Line 199: From this point on, the authors investigate the effect of 5'UTR sequence elements that affect GFP (main ORF) translation, presumably independent of uORF translation. This should be made clearer in the text. The authors make the assumption that A-rich sequence elements drive cap-independent main CDS translation, but might this also reflect translation of the uORFs? (perhaps the authors could compare ATG-containing and non-ATG-containing 10-mer reporters to address this).

Figure 4B: Given that the authors now focus on sequence promoting/inhibiting translation or decay through a mechanism other than acting as uORF start codons, it is not entirely clear why they

continue to analyze 3-mers in the recovered 10-mers? Are the same patterns evident if longer k-mers are assessed? Are there any sequence motifs that correlate with ribosome binding (for the ApppG-capped library), or is it simply A richness?

Figure S8B: Along the same lines, it is unclear why 10-mers found in unstable mRNA reporters are enriched for C or A at specific positions (i.e. position 7, 8 and 9 for A). Does this mean that the decay-promoting element might, in fact, be a motif such as CCCNNNAAA?

Reviewer #2:
Remarks to the Author:
This is an elegant paper that uses a massively parallel reporter assay to characterize how sequence elements in the 5'UTR influence mRNA translation and stability. The authors create a dual reporter system by inserting the coding sequence for a short peptide into the 5'UTR of a fluorescent protein. The peptide sequence in turn is preceded by a 10nt random sequence to create a diverse 5'UTR library. The authors then use a fluorescence assays and polysome profiling to quantify translation of the upstream open reading frame (uORF) and of the GFP protein and to understand the interplay between the two. By combining these translation assays with time course mRNA stability experiments they identify novel determinants of mRNA stability and translation efficiency. For example, they find that G-quadruplex motifs reduce translation and mark mRNA for degradation.
This is a useful paper that can be published in NMSB provided the authors can address the following comments:

• Positional information is missing from the codon frequency analysis of the MPRA data. Effects of out-of-frame start codons are not discussed even though such out-of-frame starts should be common in the random sequence. It is possible that these effects are difficult to identify in the polysome profiling assay: assuming that there are some out of frame stop codons following SIINFEKL, the resulting peptide might be of similar length than the SIINFEKL peptide generated from an in-frame start. However, it is also possible that OOF uAUG effect may introduce a divergence between 25D1/GFP analysis and M/P analysis.

• Line 66: "Indeed, the presence of an optimal TIS favored uORF translation at the expense of GFP (Fig. 1b). Altering the sequence context flanking the AUG reduced the uORF-encoded 25D1 signals with a corresponding increase of GFP levels. The reciprocal relationship between SIINFEKL and GFP translation is consistent with the leaky scanning model. "
o It's not clear that GFP levels get reduced with increasing 25D1 signal. For the strongest Optimal AUG group (left top) in Fig. 1b, GFP levels appear high and the trend (from left to right or top to bottom) is not obvious given the way the data is presented.

• Line 95: "Nevertheless, many AUG-like codons are enriched in the 25D1H population, whereas GC-rich triplets are over-represented in the GFPH population (Supplementary Fig. 2e). "
o This figure is not very conclusive, AUG-like codons (red dots) are spanning the range for both 25D1 and GFP, also there is no notation to see which ones are GC-rich codons (the blue dots also looks like span the range of GFP).

• Line 104: "Indeed, mRNAs uncovered from the monosome show a prominent enrichment of an AUG codon within the insert, which conversely, is highly depleted from polysome-derived mRNAs (Fig. 1c).

NAU and UGN triplets are also overrepresented in monosome mRNAs, another indication of AUG codons with varied flanking sequences."
o How were these triplet frequencies calculated? If ATG is favored in monosome over polysome, should TGG also be favored since ATGG should be a very strong 4mer?

• Line 109: "To examine the sequence context of AUG, we scored the monosome/polysome (M/P) ratio (log2) for sequences with all permutations of NNNNAUGNNN (Supplementary Fig. 3a). A direct comparison of high and low M/P ratio revealed the importance of a purine (A or G) at −3 position and a G at +4 position (Supplementary Fig. 3b). To validate the above sequencing results, we chose several top hits from discrete ribosome fractions and examined their translational status by flow cytometry (Supplementary Fig. 3c)."
o AUGs could start at 8 positions in the random 10- mer but motifs are only computed for AUG at position -6. The reasoning behind this choice is missing.
o Constructs M1, and M2 were picked with different AUG positions why was this choice made.
o How were P1 and P2 were picked (without AUG)?
o How many reporters in the top hits of polysome contain AUGs?

• Line 116: "Notably, non-AUG codons, including near cognate codons, are poorly enriched in the monosome-associated mRNAs, regardless of the sequence context (Fig. 1d)."
o How do results from mRNA reporters in Fig.1d relate to those from plasmid reporters in Fig. 1b? On first sight, these results seem contradictory.

• Line 192: "Several helicases such as DHX36 are known to unwind RG4 structures, permitting mRNA translation. Consistently, knocking down DHX36 resulted in marked accumulation of QUMA-1 signals with further reduced GFP levels in transfected cells (Fig. 3i). "
o In Fig. 3i, both P1 and N1's GFP levels appear reduced, but only N1 has the RG4.

• Line 223: "An inspection of in vitro mRNA stability uncovered an enrichment of the poly(A) tract from mRNA reporters short-lived in cell lysates (Supplementary Fig. 8b)."
o This analysis seems to identify position-dependent motifs seem such as CCC showing at the most 5' positions. Why did the authors pick 10A or 10C for downstream analysis rather than selecting interesting library members?

• Line 225: "Consistently, 10A exhibits a much faster turnover rate than M1 and P1 in the absence of translation (Fig. 4f). "
o The right panel in this figure shows that 10A lies in between M1 and P1 while in vivo. But this result is not explicitly discussed in the text.

• Line 226: "The in vitro destabilizing effect of A-rich sequence is further supported by variants bearing different amount of A residues (Supplementary Fig. 8c). "
o How were the 5A and 4A sequences chosen, where are the As located in the 10-mer position matrix what is the rest of the sequence.
o A statistical analysis of the N-mer library might make it possible to obtain a better understanding of the relationship between the number and positions of As than is possible to obtain from a small set of reporters.

• Line 255: "Perhaps our most surprising finding is the diverse mechanistic connection between translation initiation and mRNA decay (Fig. 4i). "

o This figure is difficult to interpret. Elements should be explicitly written in the figure or caption. How the relative rank/degree of effects for both translatability and stability were determined/calculated is unclear.

Reviewer #3:
Remarks to the Author:
In this work, Jia et al. employ an in vitro transcribed RNA reporter library to monitor the impact of 5′ UTR sequence on mRNA translation and stability in HEK293 cells. Notably, a randomized 10 nt sequence was placed at the beginning of a short, upstream open reading frame (uORF) within the 5′ UTR of a GFP-encoding mRNA, and uORF and/or GFP expression was monitored by FACS analysis. Additionally, mRNA association with ribosomes was biochemically evaluated by polysome analysis (sucrose gradient centrifugation) and mRNA stability in cells (and cell lysates where translation is absent) was calculated. Based on uORF and/or GFP expression and mRNA co-sedimentation with one or more ribosomes, the authors were able to identify correlations between the randomized 5′ UTR sequence, translation, and mRNA stability. Data from these reporters led to a number of findings:

- translation initiation efficiency is both start codon and sequence context dependent
- efficient translation of the main ORF protects mRNA from degradation
- uORF translation leads to targeting of the mRNA to nonsense-mediated mRNA decay (NMD)
- G quadraplex structure within 5′ UTRs prevents translation and destabilizes mRNA
- unstructured A-rich sequences within a 5′ UTR can serve as internal ribosome entry - sites and promote cap-independent translation and stabilization of mRNA

The reporter system to monitor uORF usage and evaluate the relationship between translation initiation and mRNA decay is clever, allowing for the evaluation of cap-independent translation (by adding a non-functional ApppG cap analog) and avoiding pitfalls associated with plasmid-based reporter assays. However, the premise of the study that little is known about the role for 5′ UTR sequences and translation initiation on mRNA translation and decay is misleading. The relationship between mRNA translation and stability has long been appreciated, and the role for uORF translation in targeting mRNA to NMD is well known. Moreover, the role for G-rich and A-rich sequences in blocking ribosomes scanning and promoting cap-independent translation, respectively, is well documented. Unfortunately, findings from this study only confirm previous conclusions gathered over the years from reporters, individual mRNAs, or other parallel reporter assays. There are, in addition, a number of issues that need to be addressed.

1. While it is shown that G quadraplex-containing reporter mRNA localize to P bodies (DCP2-labeled foci), it is incorrect to conclude that decay occurs in these structures, as this function has never been experimentally ascribed for P bodies.
2. The authors should report information related to the efficiency of the in vitro capping and polyadenylation reactions, and the length distribution of the polyA tail in the final RNA library - as these impact the translatability and stability of the transfected mRNA.
3. Although the authors recognize the limitation, the inability of the study to assign translation efficiency to individual codons (due to the likelihood of multiple RNAs contributing to the total translation output per cell) is a significant downside.
4. Although the authors conclude that the modest enrichment in AUG codons in cells expressing the uORF is likely due to bypass by scanning ribosomes, it is more likely that these mRNAs are depleted

from the pool due to their rapid removal from the cells by NMD. Indeed, unstable mRNAs will always be underrepresented in this assay.
5. The author's advocate that a 10A sequence destabilizes reporter mRNA in cell lysates through degradation by deadenylases known to be involved in 3' polyA tail shortening (lines 231-233). Given that these enzymes are exoribonucleases, this is unlikely. Based on findings that PABP1 plays a role in the stability of these mRNAs in lysates, it is more likely that PABP1 serves to recruit these nucleases to the transcript but that decay ensues from the 3' end.
6. It is generally accepted that 'mRNA surveillance' pertains to the detection and destruction of aberrant mRNA (and mRNA encountering defects in translation). The modulation of uORF and GFP CDS translation seen here by the various 5' UTR sequences does not seem, to this reviewer, to apply to the concept of 'quality control'.

Author Rebuttal to Initial comments

nature research

We are very pleased to receive the unanimous agreement from all the Referees that our study using the massively parallel uORF reporter assay is *innovative*, *powerful*, *and clever*. The complex relationship between mRNA translatability and stability is *important* and *useful*. We have carefully considered all the Referees' concerns and thoroughly revised our manuscript accordingly. In particular, we have conducted experiments knocking down additional decay factors to achieve more comprehensive screening. We have also constructed new mRNA reporters to strengthen the original conclusion. As a result, a total of nine supplementary figures have been updated. We sincerely hope that the Referees and the editor will find the revised manuscript suitable for publication in ***NSMB.***

A detailed point-by-point response to Referees' comments is listed below in blue.

Referee #1:

In this study, the authors address several important questions about the links between mRNA translation and decay, and the role of uORFs. They use an innovative method involving a library of reporter mRNAs with a random 10-mer that can potentially drive uORF translation upstream of a canonical GFP ORF. By directly transfecting the mRNA library into cells, then measuring translation of the uORF or main ORF (via monosome/polysome fractionation) or mRNA stability, they are able to relate these two processes. This demonstrates that main ORF translation tends to stabilize mRNAs, whereas uORF translation is destabilizing. Furthermore, they discover sequence elements that destabilize mRNAs through inhibiting translation (GGC and RG4 motifs).

A major advantage of directly transfecting mRNAs (rather than reporter plasmids) is that chemically modified transcripts can be used. The authors exploit this to look for sequence elements that enable translation when the mRNA has a non-functional cap analogue (ApppG). They find that A-rich elements enable cap-independent translation. As the ApppG-capped mRNAs are not translated, the authors also use these mRNAs to investigate decay in the absence of translation. This reveals that A-rich elements also promote mRNA decay, which the authors suggest depends on PABP.

Overall, this is a powerful approach to explore links between mRNA translation and decay, that could potentially be extended to address many other questions (e.g. how do sequence elements behave in different 5'UTRs, or when placed at different distances from the main ORF; to what extend RNA modifications influence translation, etc). The authors go some way towards addressing the mechanisms by which A-rich and G-rich elements modulate translation/decay. There are some caveats to this approach (e.g. mRNAs directly transfected into cells may not behave the same as mRNAs that are synthesized by transcription in the cells). Addressing these experimentally or by discussing them more fully in the text would help further strengthen this study.

We thank the Referee for concisely summarizing our work. We are very pleased to receive the Referee's positive comments about the novelty and the significance of our study.

Specific points:

1. Figure 1B: The authors conclude that a suboptimal uORF translation initiation (TI) context increases GFP levels – however, the FACS plots do not obviously show an increase in GFP levels (just a reduction in 25D1 levels). The authors should show a histogram of GFP levels for optimal and suboptimal uORF TI context reporters if they wish to make this point.

The Referee is correct that, in Fig. 1b of the manuscript, the change of GFP levels is less evident than 25D1. This is mainly due to the use of plasmid reporters that often show saturated GFP signals. In fact, this is the main reason that we decided to employ mRNA reporters to overcome this problem. As mentioned in the main text of the manuscript (page 3), "mRNA reporters increased the 25D1/GFP ratio by approximately 10-fold in transfected cells" (Supplementary Fig. 1b).

Nevertheless, we agree with the Referee that a histogram of GFP may support the point we are making. We have followed the Referee's suggestion by analyzing the fluorescence intensity of GFP positive cells. As shown in **Figure 1** of this letter, the trend of GFP change is negatively correlated with 25D1 signals, albeit with smaller scales. In the revised manuscript, we have included this new data as Supplementary Fig. 1b.

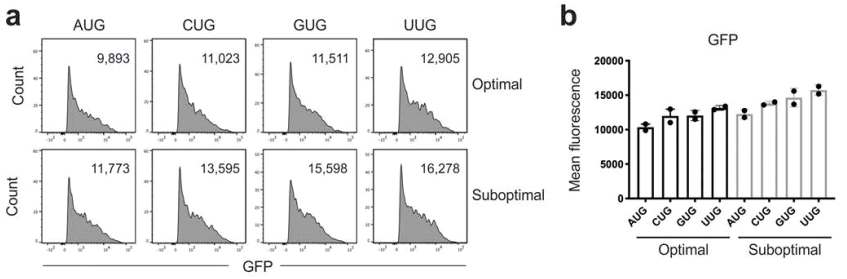

Figure 1. **(a)** Representative flow cytometry histograms of HEK293-K^b cells transfected with plasmids of uORF reporters with optimal context (top panel) or suboptimal context (bottom panel) of AUG, CUG, GUG or UUG codons. Values show the mean fluorescence of GFP. **(b)** A bar graph shows the GFP mean fluorescence of HEK293-K^b cells transfected with plasmids of uORF reporters as (a). *n* = 2 biological replicates, Error bars indicate SEM.

2. Figure S2B: The authors should present a few more metrics describing the mRNA reporter library. In particular, how many unique 10-mers are actually present (either in the oligo pools, or in the sequenced RNAs)? How many copies of each 10-mer are typically detected? And how reproducible are these experiments (e.g. do the enrichments in Figure S2D or 1C look similar for two independent experiments?)

We apologize for the lack of details in describing the mRNA reporter library. The entire random 10-mers have a total of 1,048,576 sequence combinations. Our reporter library contains >98% unique random sequences (1,033,666 for replicate #1 and 1,048,326 for replicate #2). Importantly, the majority of random sequences (80%) have >5 read counts (or copies) as shown in **Figure 2a** of this letter.

For many key experiments, we have conducted at least two biological replicates to validate the conclusions. For the library of original oligos, two replicates are correlated well (*Rho* = 0.79, $P < 2.2 \times 10^{-16}$, **Figure 2b** of this letter). Notably, for figures presented in the manuscript, we merged two replicates to increase the robustness of analysis. We have added this information in Supplementary Fig. 2c and 2d in the revised manuscript.

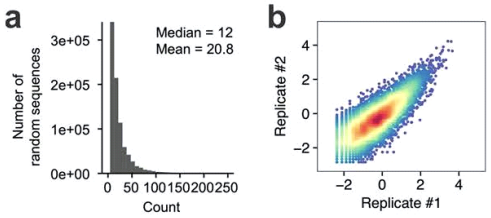

Figure 2. **(a)** A histogram shows the distribution of read count on individual unique random sequences. **(b)** A scatter plot shows the correlation of read counts between two biological replicates of original oligo sequences.

3. Figure 2A: The authors generally measure decay rates of mRNAs directly transfected into cells (rather than mRNAs that have been synthesized by transcription in the cells). They should show that the decay rates are equivalent, at least for a few examples. This could be done by measuring decay rates of mRNAs synthesized from plasmids following transcription shut-off (as in Fig S4D), and comparing to decay rates of mRNAs directly transfected into cells (potentially also after ActD treatment, to ensure conditions are comparable).

The Referee raised an important point in terms of the decay rates of mRNAs synthesized in vitro and in vivo. As pointed out by the Referee, we have partly addressed this issue in Supplementary Fig. 4d (now Fig. 5d) of the manuscript. One unique advantage of using mRNA reporters is that their intracellular decay can be measured in a non-invasive manner, i.e., without treating cells with transcription inhibitors like actinomycin D (ActD). Since the Referee's concern is highly relevant, we have expanded mRNA decay assays using more plasmid transfected cells. For mRNAs with poor translatability (N1, N2, N4, and N8), plasmid transfection resulted in the similar results as mRNA reporters (**Figure 3a** of this letter). We have now included this new result as Supplementary Fig. 6b in the revised manuscript.

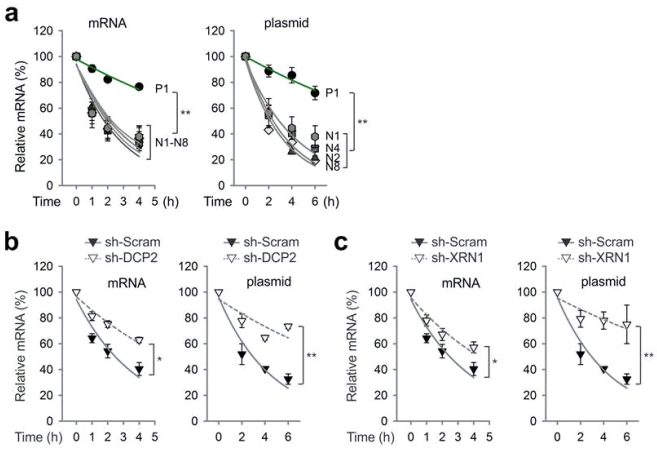

Figure 3. **(a)** HEK293-K^b cells were transfected with mRNA reporters (left) or plasmids (right) encoding sequence variants (N1 – N8), followed by mRNA decay assay. **(b)** HEK293-K^b cells with DCP2 knockdown were transfected with the N1 mRNA reporter (left) or plasmid (right), followed by mRNA decay assay. **(c)** HEK293-K^b cells with XRN1 knockdown were transfected with the N1 mRNA reporter (left) or plasmid (right), followed by mRNA decay assay. ($n = 3$ biological replicates; t test). Error bars indicate SEM. * $P < 0.05$; ** $P < 0.01$.

Using plasmid DNA encoding N1, we have also repeated the decay assay in cells lacking decay factors (XRN1 or DCP2). Consistent with mRNA reporter assays, the plasmid reporter was also stabilized upon XRN1 or DCP2 knockdown (**Figure 3b** and **3c** of this letter). Taken together, mRNAs synthesized in vitro mimic the behavior of endogenous transcripts.

4. Figure 2A: The authors rely on the assumption that in vivo decay rates in the absence of translation can be measured from in vitro decay experiments. However, they do not show that these decay rates are equivalent. One way in which this could be done is to compare the in vivo and in vitro decay rates of sequence variants that are found in ribosome-free fraction (i.e. those in Fig 3A), as these should not be translated. This would check that in vivo and in vitro decay rates are equivalent, when the effect of translation is removed.

We thank the Referee's insightful comment, but it is not our intention to claim that "in vivo decay rates in the absence of translation can be measured from in vitro decay experiments". For mRNA decay assay in the lysates, we were merely seeking a way to control mRNA turnover inside cells. In fact, we do not expect those rates are equivalent for two reasons. First, mRNAs in the ribosome-free fraction are not "absolutely" non-translatable. Second, in vivo and in vitro mRNA degradation processes are initiated by different mechanisms. To address the Referee's concern, we selected sequences enriched in the ribosome-free fraction. A comparison of their relative half-lives in vivo and in vitro revealed poor correlation (*Rho* = –0.07, **Figure 4** of this letter). This feature is likely due to the complex relationship between mRNA translatability and stability. For instance, RG4 prevents mRNA translation and triggers mRNA decay only inside cells. By contrast, the A-rich element has opposing effect on mRNA stability in vivo and in vitro. In the revised manuscript, we have further clarified this point in the main text.

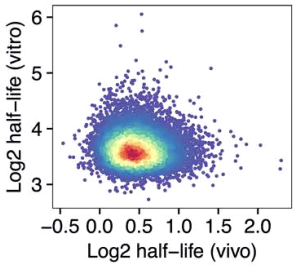

Figure 4. A scatter plot shows the correlation between half-lives of in vivo and in vitro for sequences enriched in the ribosome-free fraction.

5. Figure S3C: The authors validate the low vs high GFP translation of monosome and polysome associated sequence variants by comparing two with an ATG, and two without. However, they do not show a comparison for ATG variants in optimal versus suboptimal sequence contexts, which would help validate the ability of their approach (Figure S3A/B) to measure the contribution of features beyond the start codon.

We appreciate the Referee's effort to improve our manuscript. To follow the Referee's experimental suggestion, we selected additional candidates from monosome and polysome fractions. The two initial monosome variants (M1 and M2) are within the suboptimal context. We now included two new variants (M3 and M4) with optimal context. Indeed, M3 and M4 showed much higher 25D1 signals than M1 and M2 (**Figure 5** of this letter). By contrast, two new polysome variants (P3 and P4) exhibited substantial amount of GFP with background 25D1 signals. These results are in line with

the notion that both the start codon identity and sequence context contribute to translation efficiency. In the revised manuscript, these new results are presented in

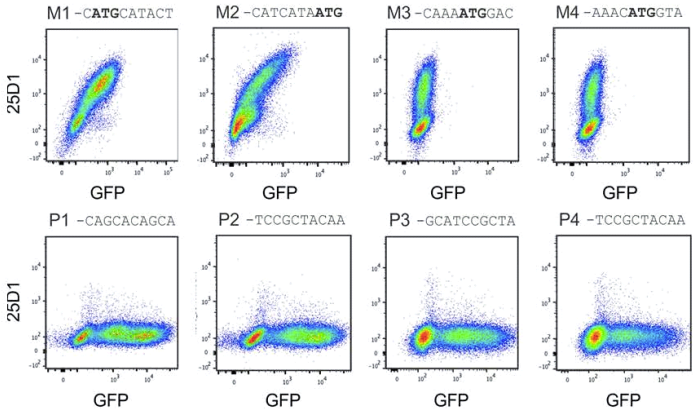

Figure 5. Representative flow cytometry of HEK 293-K^b cells transfected with mRNA reporters with sequence variants chosen from monosome (M1, M2, M3, M4) or polysome (P1, P2, P3, P4).

Supplementary Fig. 4d.

6. Line 150: The authors use a non-functional cap analogue ApppG to measure mRNA decay in the absence of translation. Could they please discuss whether this cap analogue is also likely to interfere with mRNA decay (e.g. decapping?).

If we understand the Referee's question correctly, the Referee is asking whether the cap analog ApppG prevents decapping. The widely used cap analog lacks methylation and therefore cannot bind to eIF4E. With structural similarity to m7G, we do not believe it will interfere with mRNA decay pathways. Since mRNAs capped with ApppG are short-lived (Fig. 4d of the manuscript), they are likely subjected to canonical decay pathways like decapping. However, we cannot rule out subtle difference in their functionality. We believe (and hope the Referee will concur) that the cap analog serves as a better control for m7G than non-capped mRNAs.

7. Line 168: The authors state that "many sequences bearing the GGC-motif coincided with computationally predicted RG4 structures" – please could they quantify this statement? Are GC-rich motifs not predicted to form RG4 structures also enriched in the ribosome-free fraction? It would also be helpful to show the sequence context into which the 10-mer is placed, as in order for the 10-mer to contribute to RG4s, the surrounding sequence must also contain at least one GG.

The Referee raised an excellent point in terms of the RG4sequence determinants. Potential RG4 motif is commonly described as G_X-N_{1-7}-G_X-N_{1-7}-G_X-N_{1-7}-G_X, where x is 3-6 (Kwok, *et al.* Cold Spring Harb Perspect Biol. 2018). From our 10-mer variants, however, it is difficult to compute the RG4 score because of the length limitation. Although it is possible that the surrounding sequence could contribute to the RG4 structure, we tend to favor the possibility of inter-molecular RG4 complex formation as shown in Fig. 3e of the manuscript. This possibility makes quantification of RG4 scores from 10-mer sequences less informative. Fortunately, QUMA-1 staining offers a more

straightforward way to monitor RG4 formation via flow cytometry or microscopy. Given the broad sequence variation of our library, we believe that the diversity of RG4 could be beyond what the algorithm "Quadparser" could predict. We thank the Referee for understanding.

8. Line 178: The authors claim to have examined the role of decay factors systematically. This is not quite true, because not all possible decay pathways were considered. This statement should therefore be toned down and it should be discussed that other factors/pathways may have been missed.
In fact, a more comprehensive screen for decay factors would strengthen this paper and the authors may want to consider this for a revised version of the manuscript.

We appreciate the Referee's obvious care in reviewing our manuscript. We agree that the list of decay factors we have tested in the original manuscript is far from complete. To achieve a more comprehensive screen, we have now included PARN, EXOSC4, and DIS3L2, in addition to UPF1, DCP2, XRN1, CNOT1, and PAN3. Those decay factors cover 5' → 3' and 3' → 5' RNA degradation pathways involving nonsense-mediated decay, decapping, deadenylation, and exoribonucleases. For the non-translatable mRNA (N1), we found that knocking down DCP2 or XRN1 stabilized the N1 reporter. This result is consistent with the notion that degradation of non-translatable mRNA follows the 5' → 3' pathway. Intriguingly, silencing EXOSC4 also slightly stabilized the N1 reporter (**Figure 6a** of this letter). Since EXOSC4 participates in multiple cellular RNA processing and degradation events, it is possible that Exosome also participates in the degradation of mRNAs in the absence of translation. We have now re-organized these results in Supplementary Fig. 7a in the revised manuscript.

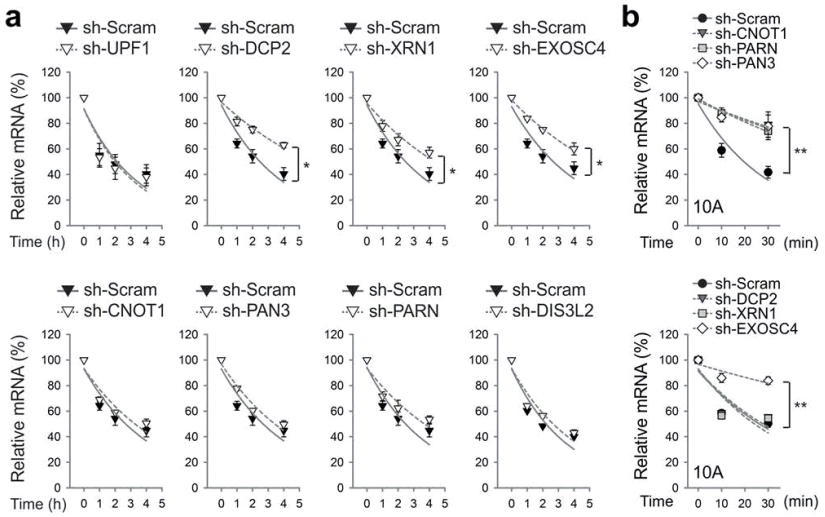

Figure 6. (**a**) HEK293-K^0 cells with specific decay factor knockdown were transfected with the N1 mRNA reporter, followed by RT-qPCR at indicated time points. (**b**) The in vitro stability of 10A mRNA reporters in the lysates of HEK293-K^0 cells lacking specific decay factors was determined by RT-qPCR at indicated time points. (n = 3 biological replicates; t test). Error bars indicate SEM. * $P < 0.05$; ** $P < 0.01$.

We further took this opportunity to examine the role of those decay factors in the in vitro degradation of the 10A reporter. Silencing EXOSC4, but not DCP2 or XRN1, stabilized 10A in the whole cell lysates (**Figure 6b** of this letter). This is rather expected because EXOSC4 forms the RNA exosome complex that has the 3' → 5' RNA exoribonuclease activity. Notably, we have already observed stabilizing effects in cell lysates lacking Ccr4-Not and Pan2-Pan3 complexes, which participate in the 3' end poly(A) tail shortening. These results collectively support the notion that the in vitro degradation of A rich element in 5'UTR follows the 3' → 5' decay pathway. We thank the Referee's insightful comments that led us to delve deeper into different mRNA decay pathways.

9. Figure 3F: The QUMA-1 FACS signal suggests that there are two cell populations – those with, and those without, RG4s. However, the cells will contain many copies of the reporter mRNA, and in addition, many endogenous transcripts that can also form RG4s. Can the authors offer any explanation why the QUMA-1 signal distribution is bimodal, rather than continuous? Furthermore, the high QUMA-1 signal seems to be heavily dependent on the reporter mRNA (P1 vs N1) – is this surprising, given that the cells will also contain many other quadruplex-forming mRNAs? Perhaps the authors could comment on this.

The Referee raised an interesting question about different QUMA-1 signals in to cell populations, which also bewildered us to some extent. We believe there is a threshold within individual cells controlling RG4 stability, which is reflected by QUMA-1 staining. In this regard, we agree with the Referee that the copy number of the reporter mRNA could contribute to RG4 formation as well as stability. To address the Referee's concern regarding endogenous transcripts, we measured QUMA-1 signals in cells without transfection. As shown in **Figure 7** of this letter, control cells also showed a minor population with high QUMA-1 signals, which is similar to cells transfected with the P1 reporter. Therefore, we believe that the increased QUMA-1 signals in the presence of N1 mRNA reporter is specific to RG4 sequences, rather than an artifact of reporter mRNA. In the revised manuscript, we have discussed this point in the main text.

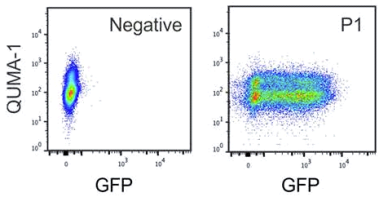

Figure 7. Representative flow cytometry of HEK 293-K^b cells transfected with P1 mRNA reporters (right) or no transfection (bottom). Transfected cells were stained with QUMA-1.

10. Figure 3I: It would be helpful to have non-DHX36-knockdown FACS plots alongside (or at least to know whether these data were obtained in the same experiment as those for Figure 3F).

The Referee's point is well-taken. In fact, the non-DHX36-knockdown plots were shown in Fig. 3f, just above Fig. 3I. Both data were obtained in the same experiment. We have clarified this in the figure legend of the revised manuscript.

11. Figure 3G: The authors suggest that QUMA-1 signal is inversely correlated to overall GFP intensity, and draw a straight line on the graph. As there are no intermediate GFP data points, this is a bit of a stretch.

We appreciate the Reviewer's attention to detail. We agree that the data points in Fig. 3g are not ideal, partly because we chose top hits from the ribosome-free fraction. As a result, all the N reporters (N1 – N8) exhibited low GFP levels. By contrast, the P1 reporter has the strongest GFP. Regardless, we feel that the inverse correlation clearly supports the notion that RG4 negatively regulates translation. We thank the Referee for understanding.

12. Line 199: From this point on, the authors investigate the effect of 5'UTR sequence elements that affect GFP (main ORF) translation, presumably independent of uORF translation. This should be made clearer in the text. The authors make the assumption that A-rich sequence elements drive cap-independent main CDS translation, but might this also reflect translation of the uORFs? (perhaps the authors could compare ATG-containing and non-ATG-containing 10-mer reporters to address this).

We apologize for not being clearer in the original description of the reporters containing the A-rich element. Since many A-rich elements do not specify start codons, the uORF translation is negligible. We have now clarified this confusion in the main text of the revised manuscript.

With respect to the question whether the A-rich element in 5'UTR enables cap-independent uORF translation, we have followed the Referee's experimental suggestion by inserting a 10-mer containing (AAAAAUGAAA). In the presence of the normal m7G cap, it drives the translation of both SIINFEKL and GFP (**Figure 8** of this letter). This result is quite expected because the AUG is within a suboptimal context. When the m7G cap is replaced with the non-functional cap analog ApppG, both 25D1 and GFP signals were reduced but still above the background. Therefore, the A-rich element in 5'UTR enables cap-independent uORF translation and the efficiency relies on the start codon optimality. We have now included this interesting result as Supplementary Fig. 9c in the revised manuscript.

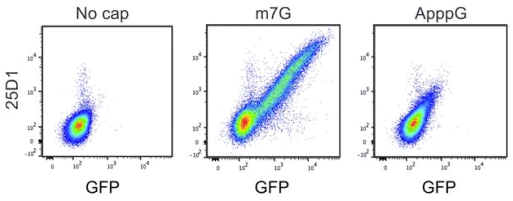

Figure 8. Representative flow cytometry of HEK 293-K^b cells transfected with mRNA reporters containing AAAAAUGAAA capped with m^{7}G or ApppG.

13. Figure 4B: Given that the authors now focus on sequence promoting/inhibiting translation or decay through a mechanism other than acting as uORF start codons, it is not entirely clear why they continue to analyze 3-mers in the recovered 10-mers? Are the same patterns evident if longer k-mers are assessed? Are there any sequence motifs that correlate with ribosome binding (for the ApppG-capped library), or is it simply A richness?

We appreciate the Referee's insightful comments. To address the Referee's concern, we analyzed other k-mers for the ApppG-capped library and found the length of k-mers does not significantly affect our original conclusion. As shown in **Figure 9** of this letter, the frequency of A-rich 4-mers in the groups of ribosome-free and ribosome-bound mRNAs resembles that of 3-mers. Notably, the majority of ribosome-bound mRNAs

have 5 As. However, we observed neither conserved motifs nor specific positions of A-cluster, indicating that the A-rich sequence element *per se* enables cap-independent translation. We have further discussed this point in the revised manuscript.

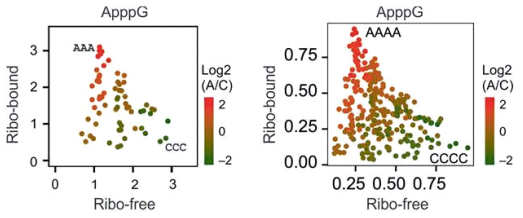

Figure 9. A scatter plot shows the distribution of triplets between ribosome free and ribosome-bound fractions. Each point refers a 3-mers (left) or 4-mers (right) sequence, and all points are color-encoded based on the ratio of A/C.

14. Figure S8B: Along the same lines, it is unclear why 10-mers found in unstable mRNA reporters are enriched for C or A at specific positions (i.e. position 7, 8 and 9 for A). Does this mean that the decay-promoting element might, in fact, be a motif such as CCCNNNAAA?

The Referee brought up an interesting point and we were equally curious about the positional effect of the A-rich element in controlling the mRNA stability in vitro. To address this question, we have followed the Referee's experimental suggestion by constructing new mRNA reporters with 4A and 5A at different positions. As shown in **Figure 10** of this letter, 4A-2 and 5A-2 reporters exhibited comparable half-lives as 10A. In addressing the similar question raised by the Referee #2, we have analyzed whether the number and position of A-clusters contribute to the mRNA stability in vitro (**Page 15** of this letter). It is interesting that mRNAs with longer As within the insert tend to have shorter half-lives (**Figure 10** of this letter). Additionally, the A-cluster located at the 3' end of the 10-mers has stronger effect on the stability, although the individual reporter assays only showed subtle differences. We have discussed these results in the main text of the revised manuscript.

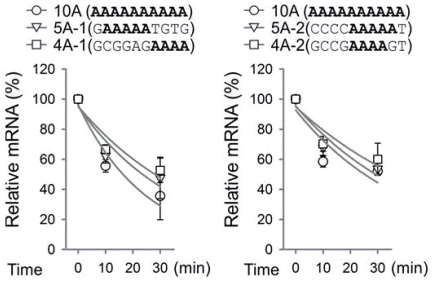

Figure 10. The in vitro decay of mRNA reporters (10A, 5A, and 4A) in the lysates of HEK293-Kb cells was determined by RT-qPCR at indicated time points. (n = 3 biological replicates; t test). Error bars indicate SEM.

Referee #2:

This is an elegant paper that uses a massively parallel reporter assay to characterize how sequence elements in the 5'UTR influence mRNA translation and stability. The authors create a dual reporter system by inserting the coding sequence for a short peptide into the 5'UTR of a fluorescent protein. The peptide sequence in turn is preceded by a 10nt random sequence to create a diverse 5'UTR library. The authors then use a fluorescence assays and polysome profiling to quantify translation of the upstream open reading frame (uORF) and of the GFP protein and to understand the interplay between the two. By combining these translation assays with time course mRNA stability experiments they identify novel determinants of mRNA stability and translation efficiency. For example, they find that G-quadruplex motifs reduce translation and mark mRNA for degradation.
This is a useful paper that can be published in NMSB provided the authors can address the following comments:

We are very pleased and grateful to receive the Referee's positive comments about the significance of our study.

1. Positional information is missing from the codon frequency analysis of the MPRA data. Effects of out-of-frame start codons are not discussed even though such out-of-frame starts should be common in the random sequence. It is possible that these effects are difficult to identify in the polysome profiling assay: assuming that there are some out of frame stop codons following SIINFEKL, the resulting peptide might be of similar length than the SIINFEKL peptide generated from an in-frame start. However, it is also possible that OOF uAUG effect may introduce a divergence between 25D1/GFP analysis and M/P analysis.

The Referee raised an excellent point! We are also interested in dissecting out-of-frame translation from the random 10-mers. Although 25D1/GFP analysis has the potential to reveal positional information of start codons, M/P analysis showed much better reading frame information. This is mainly because polysome profiling separates individual mRNAs and the uORF encoding SIINFEKL occupies one ribosome only. As shown in **Figure 11** of this letter, we clearly see an enrichment of AUG in the frame 0. To our surprise, there are substantial amount of out-of-frame translation events that cannot be explained by different open reading frames with similar length. Using individual reporter assays, we confirmed that these are true out-of-frame translation initiated from the in-

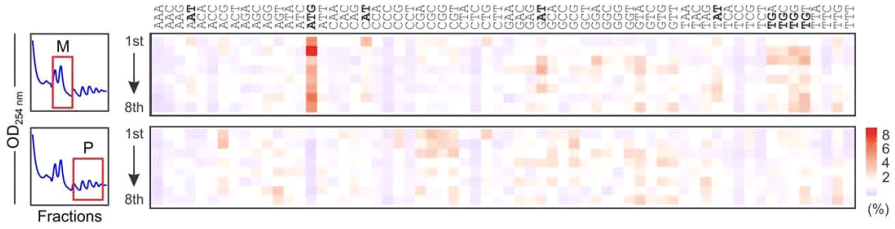

Figure 11. HEK293-K^b cells were transfected with mRNA reporters followed by ribosome fractionation. The original frequency of triplets in monosome (M) and polysome (P) fractions is shown as a heat map. The triplets ATG, NAT and TGN are highlighted.

frame AUG. Currently we are actively investigating the underlying mechanism and hoping to wrap up another exciting story focusing on start codon-associated frameshifting. We feel that including the incomplete data about frameshifting makes the current manuscript less focused. We sincerely thank the Referee for understanding.

2. Line 66: "Indeed, the presence of an optimal TIS favored uORF translation at the expense of GFP (Fig. 1b). Altering the sequence context flanking the AUG reduced the uORF-encoded 25D1 signals with a corresponding increase of GFP levels. The reciprocal relationship between SIINFEKL and GFP translation is consistent with the leaky scanning model. "
o It's not clear that GFP levels get reduced with increasing 25D1 signal. For the strongest Optimal AUG group (left top) in Fig. 1b, GFP levels appear high and the trend (from left to right or top to bottom) is not obvious given the way the data is presented.

We appreciate the Referee's effort to improve our manuscript. We agree that the change of GFP levels in Fig. 1b is less evident than 25D1. This is mainly due to the use of plasmid reporters that show accumulated GFP signals. In fact, this is the main reason that we decided to employ mRNA reporters to overcome this problem. As mentioned in the main text of the manuscript (page 3), "mRNA reporters increased the 25D1/GFP ratio by approximately 10-fold in transfected cells" (Supplementary Fig. 1b).

Since the Referee #1 raised the similar concern (**Page 2** of this letter), we presented a histogram of GFP that supports the point we are making (**Figure 1** of this letter). The trend of GFP change is negatively correlated with 25D1 signals, albeit with smaller scales. In the revised manuscript, we have added the mean fluorescence of GFP as Supplementary Fig. 1b.

3. Line 95: "Nevertheless, many AUG-like codons are enriched in the 25D1H population, whereas GC-rich triplets are over-represented in the GFPH population (Supplementary Fig. 2e). "
o This figure is not very conclusive, AUG-like codons (red dots) are spanning the range for both 25D1 and GFP, also there is no notation to see which ones are GC-rich codons (the blue dots also looks like span the range of GFP).

We apologize for the lack of clarity in describing the data shown in Supplementary Fig. 2e. In general, flow-based 25D1/GFP analysis has lower resolution than sucrose gradient-based M/P analysis. This is mainly because a single cell could be transfected with multiple mRNA molecules with different sequence variants. To address the Referee's concern, we compared the codon frequency between 25D1/GFP and M/P analysis. As shown in **Figure 12** of this letter, different codons are separated much better in M/P than 25D1/GFP, although the overall distribution is comparable. We have clarified this point in the revised manuscript.

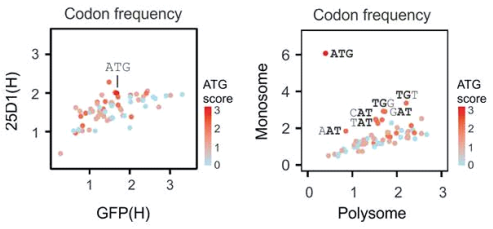

Figure 12. Correlation of triplet frequencies within the sequence variants enriched in 25D1 vs GFP (left) or monosome vs polysome (right). All points are color-encoded based on the similarity to ATG.

4. Line 104: "Indeed, mRNAs uncovered from the monosome show a prominent enrichment of an AUG codon within the insert, which conversely, is highly depleted from polysome-derived mRNAs (Fig. 1c). NAU and UGN triplets are also overrepresented in monosome mRNAs, another indication of AUG codons with varied flanking sequences."
o How were these triplet frequencies calculated? If ATG is favored in monosome over polysome, should TGG also be favored since ATGG should be a very strong 4mer?

We thank the Referee for careful reading of our manuscript. We counted the frequency of 3-mers in the inserted sequences enriched in monosome or polysome. The heatmap in Fig. 1c of the manuscript shows the original frequencies of 3-mers, and the barplot shows the ratio of 3-mers in monosome over polysome. The Referee is absolutely correct that TGG is also enriched in the monosome fraction (Fig. 1c of the manuscript), because +4G is within the optimal sequence context. To substantiate this point further, we computed the frequency of 4-mers and it is clear that ATGG is the most favorable one enriched in the monosome fraction (**Figure 13** of this letter). We have included this result in Supplementary Fig. 4c and emphasized this notion in the revised manuscript.

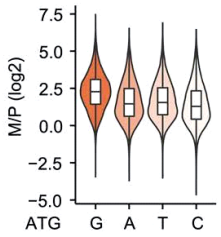

Figure 13. A violin plot shows the ratio of relative frequencies of inserted sequences in monosome over polysome, when ATG triplet is followed by different nucleotides.

5. Line 109: "To examine the sequence context of AUG, we scored the monosome/ polysome (M/P) ratio (log2) for sequences with all permutations of NNNNAUGNNN (Supplementary Fig. 3a). A direct comparison of high and low M/P ratio revealed the importance of a purine (A or G) at –3 position and a G at +4 position (Supplementary Fig. 3b). To validate the above sequencing results, we chose several top hits from discrete ribosome fractions and examined their translational status by flow cytometry (Supplementary Fig. 3c)."
o AUGs could start at 8 positions in the random 10- mer but motifs are only computed for AUG at position -6. The reasoning behind this choice is missing.
o Constructs M1, and M2 were picked with different AUG positions why was this choice made.
o How were P1 and P2 were picked (without AUG)?
o How many reporters in the top hits of polysome contain AUGs?

We apologize for not being clearer in the original description of the AUG sequence context analysis. With respect to the AUG positions, the Referee is correct that the in-frame AUG could start at 2 and 8 positions in addition to the position 5. However, only the AUG triplet at the position 5 has enough flanking sequences that enable us to capture –3 and +4 positions in order to analyze the effect of sequence context.

Regarding how to select candidate reporters for validation, we were primarily guided by the ranking of M/P ratio as well as the sequence variation. Since the Referee #1 asked more reporters with different AUG positions for validation (**Page 4** of this letter), we have now included additional M and P reporters. With a total of 8 mRNA reporters (**Figure 5** of this letter), our original conclusion has been further strengthened.

With respect to the number of sequence variants containing AUGs, we counted the fraction of sequence variants containing at least one AUG triplet in monosome and polysome fractions (**Figure 14** of this letter). When different thresholds are applied to define polysome-enriched and monosome-enriched variants, we obtained similar results with > 90% monosome sequences containing at least one AUG triplets. By contrast, AUG triplet is almost depleted from the polysome fraction.

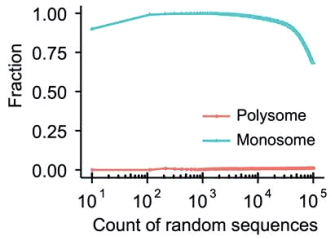

Figure 14. Fractions of sequence variants containing at least one AUG were counted in monosome and polysome fractions. X-axis shows the different thresholds to define monosome-enriched and polysome-enriched variants. For example, 100 refers to the top 100 sequences with the highest M/P ratio, and the bottom 100 sequences defined as mRNAs with the lowest M/P ratio. Y-axis shows the fraction of random sequences that contain at least one AUG triplet.

6. Line 116: "Notably, non-AUG codons, including near cognate codons, are poorly enriched in the monosome-associated mRNAs, regardless of the sequence context (Fig. 1d)."
o How do results from mRNA reporters in Fig.1d relate to those from plasmid reporters in Fig. 1b? On first sight, these results seem contradictory.

The Referee brought up an astute point in terms of the discrepancy between Fig. 1b and Fig. 1d, which reflects differences between mRNA and plasmid reporters, as well as 25D1/GFP and M/P analysis. For plasmids reporters shown in Fig. 1b, there is an accumulative effect after 24 h transfection. As a result, highly translated products are saturated, as in the case of GFP, while poorly translated products from non-AUG codons tend to be amplified. For mRNA reporters in Fig. 1d, the M/P ratio captures mRNA populations with uORF translation only (monosome), GFP only (polysome), or both (polysome). For mRNA reporters bearing non-AUG start codons, leaky scanning likely relocates those reporters into polysome. Therefore, M/P analysis is not suitable to capture the mRNA group undergoing both uORF and GFP translation. In the revised manuscript, we have clarified this confusion in the main text.

7. Line 192: "Several helicases such as DHX36 are known to unwind RG4 structures, permitting mRNA translation. Consistently, knocking down DHX36 resulted in marked accumulation of QUMA-1 signals with further reduced GFP levels in transfected cells (Fig. 3i). "
o In Fig. 3i, both P1 and N1's GFP levels appear reduced, but only N1 has the RG4.

We appreciate the Referee's attention to detail in reviewing our manuscript. We agree that knocking down DHX36 reduced global translation, possibly as a result of stress response. However, the GFP level of P1 is evidently higher than that of N1 as shown in Fig. 3i. Given the dramatic increase of QUMA-1 signals after DHX36 knockdown, it is clear that DHX36 is the major helicase in resolving RG4. We feel that the different GFP levels between P1 and N1 can be appreciated in Fig. 3i. We thank the Referee for understanding.

8. Line 223: "An inspection of in vitro mRNA stability uncovered an enrichment of the poly(A) tract from mRNA reporters short-lived in cell lysates (Supplementary Fig. 8b)."
o This analysis seems to identify position-dependent motifs seem such as CCC showing at the most 5' positions. Why did the authors pick 10A or 10C for downstream analysis rather than selecting interesting library members?

We agree with the Referee that position of the A-rich motif within the 10-mers is an interesting issue. To this end, we have constructed 5A and 4A reporters in addition to 10A. Since the Referee #1 raised the similar concern **(Page 9** of this letter), we have constructed additional 5A and 4A reporters with different positions of the poly(A) tract. For in vitro degradation, they all behaved like 10A (**Figure 10** of this letter). In addition, we monitored their potential in cap-independent translation using flow cytometry. It appears that 5A and 4A enables cap-independent translation (**Figure 15** of this letter). We have now included these new results as Supplementary Fig. 9b in the revised manuscript.

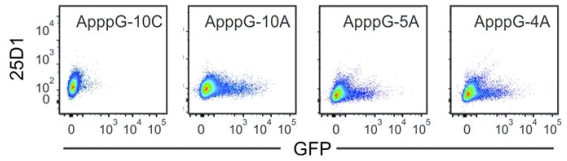

Figure 15. Representative flow cytometry scatterplots of HEK 293-K^b cells transfected with 10C, 10A, 5A, or 4A mRNA reporters capped with non-functional ApppG.

9. Line 225: "Consistently, 10A exhibits a much faster turnover rate than M1 and P1 in the absence of translation (Fig. 4f). "
o The right panel in this figure shows that 10A lies in between M1 and P1 while in vivo. But this result is not explicitly discussed in the text.

We apologize for not being clearer in the original description of the result shown in Fig. 4f. It is not surprising to find that, for in vivo half-life, 10A lies in between M1 and P1. For instance, 10A is expected to be more stable than M1 because of GFP translation. However, 10A has a faster turnover than P1 when not engaged in translation. In the revised manuscript, we have further discussed this notion in the main text.

10. Line 226: "The in vitro destabilizing effect of A-rich sequence is further supported by variants bearing different amount of A residues (Supplementary Fig. 8c). "
o How were the 5A and 4A sequences chosen, where are the As located in the 10-mer position matrix what is the rest of the sequence.
o A statistical analysis of the N-mer library might make it possible to obtain a better understanding of the relationship between the number and positions of As than is possible to obtain from a small set of reporters.

We appreciate the Referee's effort to improve our manuscript. With regard to the 5A and 4A sequences, they were chosen from top hits of variants capable of cap-independent translation. We agree that the position of A-rich element cannot be ignored. In address the Referee #1's similar concern **(Page 9** of this letter), we have constructed additional 5A and 4A reporters with different positions of the poly(A) tract.

With respect to the number and positions of A residues, the Referee's analysis suggestion is very well taken. To this end, we divided all mRNAs into different groups based on the length of maximum continuous As. In line with our original conclusion, mRNAs with longer As within the insert show a significantly shorter half-life (**Figure 16a** of this letter). When the position of A-cluster is considered, mRNAs with A-clusters located at the 3' end of the insert appear to have shorter half-life than those with A-clusters at the 5' end (**Figure 16b** of this letter). We thank the Referee's insightful comments that led us to obtain a better understanding of A-rich elements in 5'UTR. We have now presented these new results as Supplementary Fig. 10d of the revised manuscript.

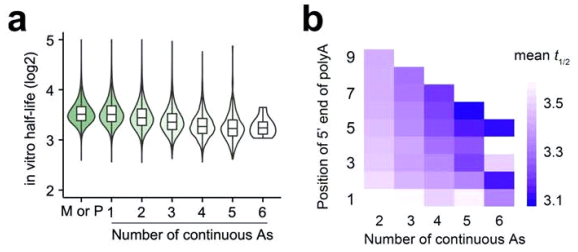

Figure 16. Effects of A-rich elements on in vitro mRNA turnover. (**a**) A violin plot shows half-life of mRNAs groups with different number of continuous As in random sequences. (**b**) A heat map shows the effect of A-cluster length and position on in intro half-life.

11. Line 255: "Perhaps our most surprising finding is the diverse mechanistic connection between translation initiation and mRNA decay (Fig. 4i). "
o This figure is difficult to interpret. Elements should be explicitly written in the figure or caption. How the relative rank/degree of effects for both translatability and stability were determined/calculated is unclear.

We apologize for the confusion in our original model that was meant to show the complex relationship between mRNA translatability and stability. Given the multiple layers of regulation, we realize that it is challenging to integrate all the proposed mechanisms into one graph. Color codes for translatability and stability are arbitrarily determined according to our data sets. In the revised manuscript, we have added more detailed description in the figure legend.

Referee #3:

In this work, Jia et al. employ an in vitro transcribed RNA reporter library to monitor the impact of 5' UTR sequence on mRNA translation and stability in HEK293 cells. Notably, a randomized 10 nt sequence was placed at the beginning of a short, upstream open reading frame (uORF) within the 5' UTR of a GFP-encoding mRNA, and uORF and/or GFP expression was monitored by FACS analysis. Additionally, mRNA association with ribosomes was biochemically evaluated by polysome analysis (sucrose gradient centrifugation) and mRNA stability in cells (and cell lysates where translation is absent) was calculated. Based on uORF and/or GFP expression and mRNA co-sedimentation with one or more ribosomes, the authors were able to identify correlations between the randomized 5' UTR sequence, translation, and mRNA stability. Data from these reporters led to a number of findings:

- translation initiation efficiency is both start codon and sequence context dependent
- efficient translation of the main ORF protects mRNA from degradation
- uORF translation leads to targeting of the mRNA to nonsense-mediated mRNA decay (NMD)
- G quadraplex structure within 5' UTRs prevents translation and destabilizes mRNA
- unstructured A-rich sequences within a 5' UTR can serve as internal ribosome entry - sites and promote cap-independent translation and stabilization of mRNA

The reporter system to monitor uORF usage and evaluate the relationship between translation initiation and mRNA decay is clever, allowing for the evaluation of cap-independent translation (by adding a non-functional ApppG cap analog) and avoiding pitfalls associated with plasmid-based reporter assays. However, the premise of the study that little is known about the role for 5' UTR sequences and translation initiation on mRNA translation and decay is misleading. The relationship between mRNA translation and stability has long been appreciated, and the role for uORF translation in targeting mRNA to NMD is well known. Moreover, the role for G-rich and A-rich sequences in blocking ribosomes scanning and promoting cap-independent translation, respectively, is well documented. Unfortunately, findings from this study only confirm previous conclusions gathered over the years from reporters, individual mRNAs, or other parallel reporter assays. There are, in addition, a number of issues that need to be addressed.

We thank the Referee for concisely summarizing our work. We agree that the putative relationship between mRNA translation and degradation has been extensively pursued. However, the majority of those studies are limited to individual mRNAs with known sequences. To our knowledge, few attempts have been made to systemically establish the relationship using completely randomized sequences in 5'UTR. Our results, for the first time, offer a holistic view of mRNA fates imposed by 10 nucleotide differences in 5'UTR. As pointed out by the Referee #1, it sets off an important platform to explore sequence elements in different positions as well as with modifications. Nevertheless, we are in full agreement with the Referee that we should avoid "misleading" statement. In the revised manuscript, we have paid great attention to tone down "overstated" assertions in the main text.

1. While it is shown that G quadraplex-containing reporter mRNA localize to P bodies (DCP2-labeled foci), it is incorrect to conclude that decay occurs in these structures, as this function has never been experimentally ascribed for P bodies.

We all due respect, we disagree with the Referee that P bodies cannot be the site of RNA degradation. A recent study reported a dual role of P-bodies in mRNA degradation and storage (doi.org/10.7554/eLife.29815.001). In our extended screen of decay factors for non-translatable mRNAs (i.e., N1 reporter), we found that the RG4 mRNA follows the 5' → 3' decay pathway because they are sensitive to DCP2 and XRN1. Given the P-body localization of both decay factors, it is reasonable to connect P bodies with RG4 mRNA degradation. Nevertheless, we cannot exclude the involvement of other organelles in RG4 mRNA degradation. Therefore, we replaced this statement with a more suggestive sentence in the revised manuscript.

2. The authors should report information related to the efficiency of the in vitro capping and polyadenylation reactions, and the length distribution of the polyA tail in the final RNA library - as these impact the translatability and stability of the transfected mRNA.

We appreciate the Referee's concern in terms of the quality of in vitro synthesized mRNA reporters. We used commercial kits for in vitro capping and polyadenylation, which have been widely used in the field. It has been estimated that an average length of poly(A) tail is ~150 nucleotides, which resembles endogenous transcripts. We completely agree with the Referee that the poly(A) tail length will affect both translatability and stability. However, our massively parallel mRNA reporters are synthesized in the same batch and the variation is equally distributed. We believe (and hope the Referee will concur) that our mRNA reporter system is well-controlled to minimize technical bias.

3. Although the authors recognize the limitation, the inability of the study to assign translation efficiency to individual codons (due to the likelihood of multiple RNAs contributing to the total translation output per cell) is a significant downside.

We believe that the Referee refers to the flow cytometry-based 25D1/GFP assay, which we are fully aware of the limitation. We have stated in the main text that "a single cell could be transfected with multiple mRNA molecules with different sequence variants, individual codons are not segregated well in distinct cell populations". In fact, this is the main reason that we employed ribosome fractionation to achieve single mRNA separation. We hope the Referee would appreciate that a combination of two independent approaches (25D1/GFP and M/P) greatly improved our data analysis.

4. Although the authors conclude that the modest enrichment in AUG codons in cells expressing the uORF is likely due to bypass by scanning ribosomes, it is more likely that these mRNAs are depleted from the pool due to their rapid removal from the cells by NMD. Indeed, unstable mRNAs will always be underrepresented in this assay.

The Referee raised an important point regarding uORF translation and NMD. We agree that many underrepresented variants are likely due to accelerated decay. Notably, the most underrepresented group contains the A-rich element, which prompted us to

investigate the role of 5'UTR poly(A) in mRNA degradation in vivo and in vitro. We have extended this discussion in the revised manuscript.

5. The author's advocate that a 10A sequence destabilizes reporter mRNA in cell lysates through degradation by deadenylases known to be involved in 3' polyA tail shortening (lines 231-233). Given that these enzymes are exoribonucleases, this is unlikely. Based on findings that PABP1 plays a role in the stability of these mRNAs in lysates, it is more likely that PABP1 serves to recruit these nucleases to the transcript but that decay ensues from the 3' end.

We appreciate the Referee's insightful comment and apologize for the overstatement. We are glad to have this opportunity to re-phrase those sentences in the main text of the revised manuscript.

6. It is generally accepted that 'mRNA surveillance' pertains to the detection and destruction of aberrant mRNA (and mRNA encountering defects in translation). The modulation of uORF and GFP CDS translation seen here by the various 5' UTR sequences does not seem, to this reviewer, to apply to the concept of 'quality control'.

The Referee raised an interesting point regarding the concept of mRNA "quality control". Compared to typical "faulty" messengers (i.e., damaged or incorrectly processed RNAs), our mRNA reporters seem to be "normal". It is conceivable that cells have to evolve multiple mechanisms to deal with different types of mRNA "quality". The best way to monitor mRNA "quality" is by ribosomes, which has been well-established for NMD. We tend to believe that cells use the similar, if not exactly the same, surveillance system to monitor little translation (uORF) or no translation (RG4), which has been minimized during evolution. Regardless, the Referee's concern is well-taken and it is not our intention to challenge the well-accepted concept. In the revised manuscript, we have taken the opportunity to clarify this concept in the discussion.

Closing words

We again thank all the Referees for their time in reading our manuscript and providing expert commentary. We have thoroughly revised our manuscript and addressed all of the concerns. In particular, we have performed additional experiments to strengthen our original conclusions. We believe the revised manuscript has been much improved. We wish to convey our most sincere thanks to all the Referees and the Editor for considering our work for publication in ***NSMB.***

Decision Letter, first revision:

27th May 2020

Dear Shu-Bing,

Thank you again for submitting the revision of your manuscript "Decoding mRNA translatability and stability from 5'UTR". We have now received the reports of the referees (appended below). Based on these comments, we are happy to accept your paper, in principle, for publication as an Article in Nature Structural & Molecular Biology, on the condition that you revise your manuscript in response to our editorial requirements.

Within a few days, we will send you detailed instructions for the final revision, along with information on these editorial and formatting requirements.

Data availability: this journal strongly supports public availability of data. Please place the data used in your paper into a public data repository, or alternatively, present the data as Supplementary Information. If data can only be shared on request, please explain why in your Data Availability Statement, and also in the correspondence with your editor. Please note that for some data types, deposition in a public repository is mandatory - more information on our data deposition policies and available repositories can be found below:
https://www.nature.com/nature-research/editorial-policies/reporting-standards#availability-of-data

TRANSPARENT PEER REVIEW

Nature Structural & Molecular Biology offers a transparent peer review option for new original research manuscripts submitted from 1st December 2019. We encourage increased transparency in peer review by publishing the reviewer comments, author rebuttal letters and editorial decision letters if the authors agree. Such peer review material is made available as a supplementary peer review file. Please state in the cover letter 'I wish to participate in transparent peer review' if you want to opt in, or 'I do not wish to participate in transparent peer review' if you don't. Failure to state your preference will result in delays in accepting your manuscript for publication.

Please note: we allow redactions to authors' rebuttal and reviewer comments in the interest of confidentiality. If you are concerned about the release of confidential data, please let us know specifically what information you would like to have removed. Please note that we cannot incorporate redactions for any other reasons. Reviewer names will be published in the peer review files if the reviewer signed the comments to authors, or if reviewers explicitly agree to release their name. For more information, please refer to our FAQ page.

If you have any questions at this stage, please do not hesitate to contact me.

With kind regards,
Anke

Anke Sparmann, PhD
Senior Editor
Nature Structural and Molecular Biology
ORCID 0000-0001-7695-2049

Reviewer #1 (Remarks to the Author):

The authors have done a great job of addressing all of our comments. The extra experiments have helped further validate their approach (e.g. testing the effect of AUG and surrounding context in Fig S4d; comparing the stability of transfected vs transcribed mRNAs in Fig S6b) and add more mechanistic insight (e.g. more decay factors tested for decay promoted by RG4 and A-rich UTR sequences). The text now reads more clearly, and they have explained/addressed some of the puzzling observations (e.g. biomodal QUMA-1 staining; position of As within the destabilizing sequences). We are pleased that their additional work supports and further strengthens their conclusions, and we would be happy to see this work published.

Reviewer #2 (Remarks to the Author):

This looks good now and can be published as is.

28th May 2020

Dear Shu-Bing,

Thank you for your patience as we've prepared the guidelines for final submission of your Nature Structural & Molecular Biology manuscript, "Decoding mRNA translatability and stability from 5'UTR" (NSMB-A42901A). Please follow the instructions provided here and in the attached files, as the formal acceptance of your manuscript will be delayed if these issues are not addressed.

When you upload your final materials, please include a concise point-by-point response to the points below.

POLICY ISSUES

1. Data availability: this journal strongly supports public availability of data. Please place the data used in your paper into a public data repository, or alternatively, present the data as supplementary information. If data can only be shared on request, please explain why in your Data Availability Statement, and also in the correspondence with your editor. Please note that for some data types, deposition in a public repository is mandatory (see Pt. 2 below).

2. DATA DEPOSITION: Deposition of deep sequencing data is mandatory, and the datasets must be released prior to or upon publication. Accession codes must be provided in your final submission for acceptance, and entries must be accessible at the galley proof stage.

3. Nature Research is taking an active approach to improving our transparency standards and increasing the reproducibility of all of our published results. Detailed information on experimental design and reagents is now collected on our Life Sciences Reporting Summary, which will be published alongside your paper. Please provide an updated version of the Reporting Summary (which will be published with the paper) with your final files.

https://www.nature.com/authors/policies/ReportingSummary.pdf

Please also upload a revised Editorial policy checklist.

https://www.nature.com/authors/policies/Policy.pdf

GENERAL FORMATTING

4. Please make sure all references are cited in numerical order and place Methods-only references after the Methods section, following the numbering of the main reference list (i.e. do not start at 1).

5. References: the reference list should contain papers that have been published or accepted by a named publication or recognized preprint server. Published conference abstracts, numbered patents and research datasets that have been assigned a digital object identifier may also be included in the reference list. Unpublished meeting abstracts, personal communications and manuscripts under consideration (and not formally accepted) may be cited only within the text and should not be added to the reference list. Please provide names of all authors of unpublished data. If you cite personal communications or unpublished data of any individuals who are not authors of your manuscript, you must supply copies of written permission from the primary investigator of each group cited.

FIGURES & TABLES

6. Please make sure all figures and tables, including Extended Data Figures, are cited in the text in numerical order.

7. Our article format allows up to 8 main figures. Currently, Fig. 3 and Fig. 4 are fairly data-dense; splitting the information into different figures might improve the final manuscript layout. You could also consider moving the most relevant supplementary Figures to the main text, if you wish.

SUPPLEMENTARY INFORMATION

All Supplementary Information must be submitted in accordance with the instructions in the attached Inventory of Supporting Information, and should fit into one of three categories:

1. EXTENDED DATA FIGURES: Extended Data Figures are an integral part of the paper and only data that directly contribute to the main message should be presented. These figures will be integrated into the full-text HTML version of your paper and will be appended to the online PDF. There is a limit of 10 Extended Data Figures, and each must be referred to in the main text. Each Extended Data Figure should be of the same quality as the main figures, and should be supplied at a size that will allow both the figure and legend to be presented on a single legal-sized page. Each figure should be submitted as an individual .jpg, .tif or .eps file with a maximum size of 10 MB each. All Extended Data figure legends must be provided in the attached Inventory of Accessory Information, not in the figure files themselves.
All Extended Data Figure must be called out in order as Extended Data Fig. 1, Extended Data Fig 2, etc.

2. SUPPLEMENTARY INFORMATION: Supplementary Information is material that is essential background to the study but which is not practical to include in the printed version of the paper (for example, large figures, video files, large data sets and calculations). Each item must be referred to in the main manuscript and detailed in the attached Inventory of Accessory Information. Supplementary Tables containing large data sets should be in Excel format, with the table number and title included within the body of the table. All textual information and any additional Supplementary Figures (which should be presented with the legends directly below each figure) should be provided as a single, combined PDF. Please note that we cannot accept resupplies of Supplementary Information after the paper has been formally accepted unless there has been a critical scientific error.

Supplementary items (such as Supplementary Tables, Videos, Notes, and additional Supplementary Figures if permitted), should be numbered and called out in main article, as Supplementary Figure 1 (not SI1) and so on.

3. SOURCE DATA: We encourage you to provide source data for your figures. Full-length, unprocessed gels and blots must be provided as source data for any relevant figures, and should be provided as individual PDF files for each figure containing all supporting blots and/or gels with the linked figure noted directly in the file. Statistical source data (i.e., data behind graphs; here e.g. for Fig. 1d, 2b-f, 3d, g, j, 4e, f, h) should be provided in Excel format, one file for each relevant figure, with the linked figure noted directly in the file. For imaging source data, we encourage deposition to a relevant repository, such as figshare (https://figshare.com/) or the Image Data Resource (https://idr.openmicroscopy.org).
Source data should be cited in the legend text (e.g., "Uncropped images for panels a-c are available as source data online" or "Source data for graphs in d-f are available online").

STATISTICS & REPRODUCIBILITY

8. GRAPHS: Wherever statistics have been derived (e.g. error bars, box plots, statistical significance), the legend needs to provide and define the n number (i.e. the sample size used to derive statistics) as a precise value (not a range), using the wording "n=X biologically independent samples/cell cultures/animals/independent experiments" etc. as applicable. All error bars need to be defined (e.g., s.d. or s.e.m.) together with a measure of center (e.g., mean or median) and should be accompanied by their precise n number, defined as noted above.

9. All box plots need to be defined in terms of minima, maxima, center, and percentiles, and should be accompanied by their precise n number defined as noted above.

10. Wherever statistical significance has been derived, precise P values should be provided if possible and appropriate. The type of statistical test used needs to be defined in the legend, whether they were one-sided or two-sided or whether adjustments were made for multiple comparisons.

11. When representative experiments are shown, you should state in the legends how many times each experiment was repeated independently with similar results. Please indicate number of times experiments were repeated, number of images collected, etc.

12. If applicable, the Methods should include a statistics section, listing statistical tests used, whether the test was one- or two-tailed, exact values for both significant and non-significant P values where relevant; F values and degrees of freedom for all ANOVAs; and t-values and degrees of freedom for t-tests.

13. Cell lines: the Methods should include a section with cell lines used, origin, whether they were tested for mycoplasma and, where relevant, whether they were authenticated or not.

14. Competing interests statement: Please include a competing interests statement as a separate section after the Author Contributions, under the heading "Competing interests", and enumerate any such circumstances there, or read: The authors declare no competing interests.

15. Reporting Summary statement: This should be placed after Online Methods section and read: Further information on experimental design is available in the Nature Research Reporting Summary linked to this article.

16. Data Availability statement: This should be placed after Code Availability / Reporting Summary statement (before Methods-only references).

We suggest that you list in this order:

- data deposited in public repositories, with accession codes or DOIs.

- data available as Source Data (e.g. "Source data for figure 3d, 4b and 4c are available with the paper online.")

- if any data can only be shared upon request, please specify what those data are and explain why.

More information and examples can be found at
http://www.nature.com/authors/policies/data/data-availability-statements-data-citations.pdf

TRANSPARENT PEER REVIEW

17. Nature Structural & Molecular Biology offers a transparent peer review option for new original research manuscripts submitted from 1st December 2019. We encourage increased transparency in peer review by publishing the reviewer comments, author rebuttal letters and editorial decision letters

if the authors agree. Such peer review material is made available as a supplementary peer review file. Please state in the cover letter 'I wish to participate in transparent peer review' if you want to opt in, or 'I do not wish to participate in transparent peer review' if you don't. Failure to state your preference will result in delays in accepting your manuscript for publication.

Please note: we allow redactions to authors' rebuttal and reviewer comments in the interest of confidentiality. If you are concerned about the release of confidential data, please let us know specifically what information you would like to have removed. Please note that we cannot incorporate redactions for any other reasons. Reviewer names will be published in the peer review files if the reviewer signed the comments to authors, or if reviewers explicitly agree to release their name. For more information, please refer to our FAQ page.

OTHER REQUIREMENTS

18. Ensure that all required forms found in the Policy Worksheet are uploaded to our Journal Processing system as "Supplementary Materials".

Nature Research journals encourage authors to share their step-by-step experimental protocols on a protocol sharing platform of their choice. Nature Research's Protocol Exchange is a free-to-use and open resource for protocols; protocols deposited in Protocol Exchange are citable and can be linked from the published article. More details can found at www.nature.com/protocolexchange/about.

In addition to addressing these points, please refer to the attached policy and rights worksheet, which contains information on how to comply with our legal guidelines for publication and describes the files that you will need to upload prior to final acceptance. You must initial the relevant portions of this checklist, sign it and return it with your final files. I have also attached a formatting guide for you to consult as you prepare the revised manuscript. Careful attention to this guide will ensure that the production process for your paper is more efficient.

Please use the following link for uploading these materials:
[REDACTED]

We ask that you aim to return your revised paper within 7-10 days. If you have any further questions, please feel free to contact me.

With kind regards,
Anke

Anke Sparmann, PhD
Senior Editor

Nature Structural and Molecular Biology
ORCID 0000-0001-7695-2049

Reviewer #1:
Remarks to the Author:
The authors have done a great job of addressing all of our comments. The extra experiments have helped further validate their approach (e.g. testing the effect of AUG and surrounding context in Fig S4d; comparing the stability of transfected vs transcribed mRNAs in Fig S6b) and add more mechanistic insight (e.g. more decay factors tested for decay promoted by RG4 and A-rich UTR sequences). The text now reads more clearly, and they have explained/addressed some of the puzzling observations (e.g. biomodal QUMA-1 staining; position of As within the destabilizing sequences). We are pleased that their additional work supports and further strengthens their conclusions, and we would be happy to see this work published.

Reviewer #2:
Remarks to the Author:
This looks good now and can be published as is.

Author Rebuttal, first revision:

We are excited to receive the decision of acceptance in principle for our manuscript. A detailed point-by-point response to the editor's comments is listed below in blue.

POLICY ISSUES

1. Data availability: this journal strongly supports public availability of data. Please place the data used in your paper into a public data repository, or alternatively, present the data as supplementary information. If data can only be shared on request, please explain why in your Data Availability Statement, and also in the correspondence with your editor. Please note that for some data types, deposition in a public repository is mandatory (see Pt. 2 below).

Yes, our deep sequencing data sets have been deposited into GEO with the accession number GSE145046.

2. DATA DEPOSITION: Deposition of deep sequencing data is mandatory, and the datasets must be released prior to or upon publication. Accession codes must be provided in your final submission for acceptance, and entries must be accessible at the galley proof stage.

Yes, our deep sequencing data sets have been deposited into GEO with the accession number GSE145046.

3. Nature Research is taking an active approach to improving our transparency standards and increasing the reproducibility of all of our published results. Detailed information on experimental design and reagents is now collected on our Life Sciences Reporting Summary, which will be published alongside your paper. Please provide an updated version of the Reporting Summary (which will be published with the paper) with your final files.
https://www.nature.com/authors/policies/ReportingSummary.pdf

Please also upload a revised Editorial policy checklist.
https://www.nature.com/authors/policies/Policy.pdf

Yes, both Reporting Summary and Editorial Policy Checklist have been updated.

GENERAL FORMATTING

4. Please make sure all references are cited in numerical order and place Methods-only references after the Methods section, following the numbering of the main reference list (i.e. do not start at 1).

Yes, the reference after the Methods section has been updated.

5. References: the reference list should contain papers that have been published or accepted by a named publication or recognized preprint server. Published conference abstracts, numbered patents and research datasets that have been assigned a digital object identifier may also be included in the reference list. Unpublished meeting abstracts, personal communications and manuscripts under consideration (and not formally accepted) may be cited only within the text and should not be added to the reference list. Please provide names of all authors of unpublished data. If you cite personal communications or unpublished data of any individuals who are not authors of your manuscript, you must supply copies of written permission from the primary investigator of each group cited.

Yes, we have doubled checked the main reference list.

FIGURES & TABLES

6. Please make sure all figures and tables, including Extended Data Figures, are cited in the text in numerical order.

Yes, we have doubled checked the main text.

7. Our article format allows up to 8 main figures. Currently, Fig. 3 and Fig. 4 are fairly data-dense; splitting the information into different figures might improve the final manuscript layout. You could also consider moving the most relevant supplementary Figures to the main text, if you wish.

We have followed the editor's suggestion by splitting the whole set of figures into a total of 7, which improved the manuscript layout.

SUPPLEMENTARY INFORMATION

All Supplementary Information must be submitted in accordance with the instructions in the attached Inventory of Supporting Information, and should fit into one of three categories:

1. EXTENDED DATA FIGURES: Extended Data Figures are an integral part of the paper and only data that directly contribute to the main message should be presented. These figures will be integrated into the full-text HTML version of your paper and will be appended to the online PDF. There is a limit of 10 Extended Data Figures, and each must be referred to in the main text. Each Extended Data Figure should be of the same quality as the main figures, and should be supplied at a size that will allow both the figure and legend to be presented on a single legal-sized page. Each figure should be submitted as an individual .jpg, .tif or .eps file with a maximum size of 10 MB each. All Extended Data figure legends must be provided in the attached Inventory of Accessory Information, not in the figure files themselves.

All Extended Data Figure must be called out in order as Extended Data Fig. 1, Extended Data Fig 2, etc.

The revised manuscript has a total of 10 Extended Data Figures and we have called them out as "Extended Data Fig" in the main text.

2. SUPPLEMENTARY INFORMATION: Supplementary Information is material that is essential background to the study but which is not practical to include in the printed version of the paper (for example, large figures, video files, large data sets and calculations). Each item must be referred to in the main manuscript and detailed in the attached Inventory of Accessory Information. Supplementary Tables containing large data sets should be in Excel format, with the table number and title included within the body of the table. All textual information and any additional Supplementary Figures (which should be presented with the legends directly below each figure) should be provided as a single, combined PDF. Please note that we cannot accept resupplies of Supplementary Information after the paper has been formally accepted unless there has been a critical scientific error.

Supplementary items (such as Supplementary Tables, Videos, Notes, and additional Supplementary Figures if permitted), should be numbered and called out in main article, as Supplementary Figure 1 (not SI1) and so on.

We have numbered the Supplementary Table and called it out in the main text. We also provided this information in the Inventory of Accessory Information.

3. SOURCE DATA: We encourage you to provide source data for your figures. Full-length, unprocessed gels and blots must be provided as source data for any relevant figures, and

should be provided as individual PDF files for each figure containing all supporting blots and/or gels with the linked figure noted directly in the file. Statistical source data (i.e., data behind graphs; here e.g. for Fig. 1d, 2b-f, 3d, g, j, 4e, f, h) should be provided in Excel format, one file for each relevant figure, with the linked figure noted directly in the file. For imaging source data, we encourage deposition to a relevant repository, such as figshare (https://figshare.com/) or the Image Data Resource (https://idr.openmicroscopy.org).

Source data should be cited in the legend text (e.g., "Uncropped images for panels a-c are available as source data online" or "Source data for graphs in d-f are available online").

We don't have blots or gels in our manuscript. However, we provided Statistical Source data in Excel format, one file for each figure. Also, we added Source data information in Inventory of Accessory Information.

STATISTICS & REPRODUCIBILITY

8. GRAPHS: Wherever statistics have been derived (e.g. error bars, box plots, statistical significance), the legend needs to provide and define the n number (i.e. the sample size used to derive statistics) as a precise value (not a range), using the wording "n=X biologically independent samples/cell cultures/animals/independent experiments" etc. as applicable. All error bars need to be defined (e.g., s.d. or s.e.m.) together with a measure of center (e.g., mean or median) and should be accompanied by their precise n number, defined as noted above.

We have defined the *n* number in each related legend. All error bars are defined as s.e.m. together with mean in each related legend.

9. All box plots need to be defined in terms of minima, maxima, center, and percentiles, and should be accompanied by their precise n number defined as noted above.

All box and violin blots have been clearly defined.

10. Wherever statistical significance has been derived, precise P values should be provided if possible and appropriate. The type of statistical test used needs to be defined in the legend, whether they were one-sided or two-sided or whether adjustments were made for multiple comparisons.

We have showed precise P values and the type of statistical test in the legend.

11. When representative experiments are shown, you should state in the legends how many times each experiment was repeated independently with similar results. Please indicate number of times experiments were repeated, number of images collected, etc.

We have indicated how many times each experiment was repeated and the number of images collected in the legend.

12. If applicable, the Methods should include a statistics section, listing statistical tests used, whether the test was one- or two-tailed, exact values for both significant and non-significant P values where relevant; F values and degrees of freedom for all ANOVAs; and t-values and degrees of freedom for t-tests.

We have provided a statistics section in the Methods section.

13. Cell lines: the Methods should include a section with cell lines used, origin, whether they were tested for mycoplasma and, where relevant, whether they were authenticated or not.

We included a section with cell line information and also provided detailed information in the Reporting Summary.

14. Competing interests statement: Please include a competing interests statement as a separate section after the Author Contributions, under the heading "Competing interests", and enumerate any such circumstances there, or read: The authors declare no competing interests.

We have included a statement of competing interests as a separate section in the manuscript.

15. Reporting Summary statement: This should be placed after Online Methods section and read: Further information on experimental design is available in the Nature Research Reporting Summary linked to this article.

We have provided this statement after the online Methods section.

16. Data Availability statement: This should be placed after Code Availability / Reporting Summary statement (before Methods-only references).

We suggest that you list in this order:

- data deposited in public repositories, with accession codes or DOIs.

- data available as Source Data (e.g. "Source data for figure 3d, 4b and 4c are available with the paper online.")

- if any data can only be shared upon request, please specify what those data are and explain why.

More information and examples can be found at

http://www.nature.com/authors/policies/data/data-availability-statements-data-citations.pdf.

We have followed the editor's suggestion and updated the Data Availability Statement.

TRANSPARENT PEER REVIEW

17. Nature Structural & Molecular Biology offers a transparent peer review option for new original research manuscripts submitted from 1st December 2019. We encourage increased transparency in peer review by publishing the reviewer comments, author rebuttal letters and editorial decision letters if the authors agree. Such peer review material is made available as a supplementary peer review file. Please state in the cover letter 'I wish to participate in transparent peer review' if you want to opt in, or 'I do not wish to participate in transparent peer review' if you don't. Failure to state your preference will result in delays in accepting your manuscript for publication.

Please note: we allow redactions to authors' rebuttal and reviewer comments in the interest of confidentiality. If you are concerned about the release of confidential data, please let us know specifically what information you would like to have removed. Please note that we cannot incorporate redactions for any other reasons. Reviewer names will be published in the peer review files if the reviewer signed the comments to authors, or if reviewers explicitly agree to release their name. For more information, please refer to our FAQ page.

We have stated in the cover letter that 'I wish to participate in transparent peer review'.

18. Ensure that all required forms found in the Policy Worksheet are uploaded to our Journal Processing system as "Supplementary Materials".

We have uploaded the Policy Worksheet as "Supplementary Materials".

Final Decision Letter:

16th Jun 2020

Dear Shu-Bing,

We are happy to accept your revised paper "Decoding mRNA translatability and stability from 5'UTR" for publication as an Article in Nature Structural & Molecular Biology.

Acceptance is conditional on the manuscript's not being published elsewhere and on there being no announcement of this work to the newspapers, magazines, radio or television until the publication date in Nature Structural & Molecular Biology.

Before the manuscript is sent to the printers, we shall make any detailed changes in the text that may be necessary either to make it conform with house style or to make it intelligible to a wider readership. If the changes are extensive, we will ask for your approval before the manuscript is laid out for production. Once your manuscript is typeset you will receive a link to your electronic proof via email within 20 working days, with a request to make any corrections within 48 hours. Please read proofs with great care to make sure that the sense has not been altered. If you have queries at any point during the production process then please contact the production team

at rjsproduction@springernature.com. Once your paper has been scheduled for online publication, the Nature press office will be in touch to confirm the details.

Please note that due to tight production schedules, proofs should be returned as quickly as possible to avoid delaying publication. If you anticipate any limitations to your availability over the next 2-4 weeks (such as vacation or traveling to conferences, etc.), please e-mail rjsproduction@springernature.com as soon as possible. Please provide specific dates that you will be unavailable and provide detailed contact information for an alternate corresponding author if necessary.

To assist our authors in disseminating their research to the broader community, our SharedIt initiative provides all co-authors with the ability to generate a unique shareable link that will allow anyone (with or without a subscription) to read the published article. Recipients of the link with a subscription will also be able to download and print the PDF.

As soon as your article is published, you can generate your shareable link by entering the DOI of your article here: <a href="http://authors.springernature.com/share">http://authors.springernature.com/share<a>. Corresponding authors will also receive an automated email with the shareable link

Note the policy of the journal on data deposition: http://www.nature.com/authors/policies/availability.html.

Your paper will be published online soon after we receive proof corrections and will appear in print in the next available issue. You can find out your date of online publication by contacting the production team shortly after sending your proof corrections. The embargo is set at 16:00 London time (GMT) / 11:00 am US Eastern time (EST), on the Monday of publication. Now is the time to inform your Public Relations or Press Office about your paper, as they might be interested in promoting its publication. This will allow them time to prepare an accurate and satisfactory press release. Include your manuscript tracking number (NSMB-A42901B) and our journal name, which they will need when they contact our press office.

About one week before your paper is published online, we shall be distributing a press release to news organizations worldwide, which may very well include details of your work. We are happy for your institution or funding agency to prepare its own press release, but it must mention the embargo date and Nature Structural & Molecular Biology. If you or your Press Office have any enquiries in the meantime, please contact press@nature.com.

You can now use a single sign-on for all your accounts, view the status of all your manuscript submissions and reviews, access usage statistics for your published articles and download a record of your refereeing activity for the Nature journals.

If you have not already done so, we strongly recommend that you upload the step-by-step protocols used in this manuscript to the Protocol Exchange. Protocol Exchange is an open online resource that allows researchers to share their detailed experimental know-how. All uploaded protocols are made freely available, assigned DOIs for ease of citation and fully searchable through nature.com. Protocols can be linked to any publications in which they are used and will be linked to from your article. You can also establish a dedicated page to collect all your lab Protocols. By uploading your Protocols to

Protocol Exchange, you are enabling researchers to more readily reproduce or adapt the methodology you use, as well as increasing the visibility of your protocols and papers. Upload your Protocols at www.nature.com/protocolexchange/. Further information can be found at www.nature.com/protocolexchange/about.

An online order form for reprints of your paper is available at https://www.nature.com/reprints/author-reprints.html. Please let your coauthors and your institutions' public affairs office know that they are also welcome to order reprints by this method.

Please note that we encourage the authors to self-archive their manuscript (the accepted version before copy editing) in their institutional repository, and in their funders' archives, six months after publication. Nature Research Group recognizes the efforts of funding bodies to increase access of the research they fund, and strongly encourages authors to participate in such efforts. For information about our editorial policy, including license agreement and author copyright, please visit www.nature.com/nsmb/ about/ed_policies/index.html

The Author's Accepted Manuscript (the accepted version of the manuscript as submitted by the author) may only be posted 6 months after the paper is published, consistent with our self-archiving embargo. Please note that the Author's Accepted Manuscript may not be released under a Creative Commons license. For Nature Research Terms of Reuse of archived manuscripts please see: http://www.nature.com/authors/policies/license.html#terms
If you have posted a preprint on any preprint server, please ensure that the preprint details are updated with a publication reference, including the DOI and a URL to the published version of the article on the journal website.

With kind regards,
Anke

Anke Sparmann, PhD
Senior Editor
Nature Structural and Molecular Biology
ORCID 0000-0001-7695-2049

Click here if you would like to recommend Nature Structural & Molecular Biology to your librarian: http://www.nature.com/subscriptions/recommend.html#forms