

# Decoupled degradation and translation enables noise-modulation by poly(A)-tails

Carmen Grandi<sup>1,2</sup>, Martin Emmaneel<sup>1,2</sup>, Frank H. T. Nelissen<sup>1,2</sup>, Laura W. M. Roosenboom<sup>1,3,4</sup>, Yoanna Petrova<sup>1</sup>, Omnia Elzokla<sup>1</sup>, Maïke M. K. Hansen<sup>1,2,5,\*</sup>.

## Affiliations:

<sup>1</sup>Institute for Molecules and Materials, Radboud University, Heyendaalseweg 135, 6525 AJ Nijmegen, the Netherlands.

<sup>2</sup>Oncode Institute, Nijmegen, The Netherlands

## Current affiliations:

<sup>3</sup>Laboratory of Immunoengineering, Department of Biomedical Engineering, Eindhoven University of Technology, Eindhoven, The Netherlands.

<sup>4</sup>Institute for Complex Molecular Systems (ICMS), Eindhoven University of Technology, Eindhoven, The Netherlands.

<sup>5</sup>Lead contact

\*Correspondence to: E-mail: [maïke.hansen@ru.nl](mailto:maïke.hansen@ru.nl) (M.M.K.H)

## SUMMARY

Poly(A)-tails are crucial for mRNA translation and degradation, but the exact relationship between tail length and mRNA kinetics remains unclear. Here we employ a small library of identical mRNAs that differ only in their poly(A)-tail length to examine their behavior in human embryonic kidney cells. We find that tail length strongly correlates with mRNA degradation rates, but is decoupled from translation. Interestingly, an optimal tail length of ~100 nucleotides displays the highest translation rate, which is identical to the average endogenous tail length measured by nanopore sequencing. Furthermore, poly(A)-tail length variability—a feature of endogenous mRNAs—impacts translation efficiency but not mRNA degradation rates. Stochastic modelling combined with single-cell tracking reveals that poly(A)-tails provide cells with an independent handle to tune gene expression fluctuations by decoupling mRNA degradation and translation. Together, this work contributes to the basic understanding of gene expression regulation and has potential applications in nucleic acid therapeutics.

**Keywords:** Poly(A)-tail length; mRNA degradation and translation; gene expression kinetics; stochastic noise; protein fluctuations.

# INTRODUCTION

The regulation of gene expression is essential for maintaining cellular function and involves numerous transcriptional and post-transcriptional steps that ultimately affect mRNA and protein levels<sup>1-3</sup>. Tuning transcription is an efficient way for cells to control gene expression, as it allows for the shutdown of non-essential genes<sup>4,5</sup>. However, both post-transcriptional and post-translational processes can provide more rapid mechanisms that regulate previously synthesized mRNAs and proteins, to quickly influence their fate<sup>6,7</sup>. One particularly important post-transcriptional process thought capable of impacting mRNA functionality that has gained increased attention in recent years is polyadenylation<sup>8-12</sup>.

Renewed interest in poly(A)-tails was sparked by improvements in sequencing of homopolymeric nucleic acid regions, which has allowed better characterization of poly(A)-tails<sup>13-16</sup>. This led to the discovery that poly(A)-tails do not have fixed lengths, but are rather non-uniform and dynamic elements. In fact, mRNAs in the cytoplasm of mammalian cells display high intergenic variability in poly(A)-tail length ranging from 50 to 100 adenosines, depending on the gene<sup>12,13</sup>. Since poly(A)-tails are involved in numerous processes such as nuclear export<sup>17</sup>, translational initiation<sup>18-21</sup>, and mRNA degradation<sup>12,22-24</sup>, tail-length is likely to impact a wide-range of kinetic properties of mRNA and protein biogenesis. Although mRNA translation and degradation rates have been shown to span a drastic 1000-fold range<sup>12</sup>, the direct effect of poly(A)-tail length on mRNA stability and protein expression remains elusive, in part due to the presence of other regulatory elements on endogenous mRNAs that obscure data interpretation<sup>25-28</sup>. For example, differences in stability and translation efficiency of mRNAs with different poly(A)-tail lengths could be confounded by differences in the 3'UTR<sup>27</sup> or codon optimality<sup>28-30</sup>, which are known to affect mRNA kinetics. Taken together, using endogenous mRNAs to study the relationship between poly(A)-tail length and mRNA degradation and translation kinetics can make it difficult to detangle cause and effect.

Interestingly, single-molecule quantification of poly(A)-tail length has demonstrated that mRNAs can display a high degree of intragenic variability in tail length (i.e., between different transcripts of the same gene)<sup>12,13</sup>. Surprisingly, the functional relevance of this variability remains unknown. In general, the 3' termini of eukaryotic mRNA contain binding sites for regulatory proteins as well as miRNAs that play critical roles in mRNA and protein biogenesis. Furthermore, these sequences often undergo changes both in physiological and pathological conditions<sup>31,32</sup>, as do poly(A)-tails<sup>33,34</sup>. Therefore, given that most elements of an mRNA are functionally conserved features, it is highly likely that intragenic variability in poly(A)-tail length has also evolved to play a functionally relevant role. Yet, what this role might be remains undefined.

Here, we develop a strategy to study the direct effect of poly(A)-tail length on mRNA translation and degradation rates in human embryonic kidney cells (HEK293T/17), through the synthesis of mRNAs with different, yet specific, poly(A)-tail lengths. In this cellular model, we find that poly(A)-tail length decouples mRNA degradation and translation rates. Intuitively, we further observe that the introduction of variability in poly(A)-tail lengths results in changes in the translation efficiency, due to the inclusion in the distribution of tail lengths with poorer translation rates. This might also provide cells an alternate handle to regulate protein fluctuations. This work contributes to the basic understanding of gene expression regulation, but could also find application in the context of mRNA-based technologies, as an emerging approach in the field of nucleic acid therapeutics (e.g., mRNA-based vaccines, cancer immunotherapies) to optimize mRNA translation efficiency.

## RESULTS

### Poly(A)-tail length negatively correlates with mRNA half-life while being decoupled from translation rates

To examine the role of poly(A)-tail length on mRNA translation and degradation without other confounding regulatory elements that obscure the true effect of tail length<sup>25,26</sup>, we generated a small library of synthetic GFP (Green Fluorescent Protein) coding mRNAs (Figure S1-S2). These mRNAs contain identical 5'UTR and 3'UTR sequences (Table S1), and a defined poly(A)-tail length spanning from 5 to 150 adenosines (Figure S1A-C). Each tail-specific mRNA was transfected into HEK293T/17 cells, as they provide consistent results due to their high transfection efficiency and low maintenance<sup>35</sup>. In this cellular model, the translation into GFP molecules mainly occurred between 5 and 15 hours post transfection (Figure 1A and Figure S1D). Interestingly, there appears to be an optimal tail length (100 nt) around which GFP is highly expressed (Figure 1B, in green). The GFP expression from the other poly(A)-tail lengths decreases almost symmetrically around this optimum. A very short tail (5 nt) shows almost no translation, but also the longest tail (150 nt) displays surprisingly low protein expression (Figure 1B, grey and blue respectively).

The final GFP levels are heavily impacted by different kinetic steps, such as mRNA degradation ( $k_{d1}$ ) and translation ( $k_{tl}$ ) (Figure 1C). Therefore, to determine the tail-specific mRNA translation rates, we first quantified the effect of poly(A)-tail length on mRNA degradation. To this end, we performed single molecule fluorescence *in situ* hybridization (smFISH)<sup>36</sup> at different time points post transfection (Figure 1D). The half-lives measured span from ~1 to ~4 hours, indicating that the poly(A)-tail length has a strong effect on mRNA stability (Figure 1E, Table S2). Specifically, there is a high positive correlation (Pearson's  $r = 0.95$ ) between poly(A)-tail length and mRNA degradation rate, meaning that mRNAs with longer tails are degraded faster (Figure 1F). The finding that the levels of mRNAs measured for all tails 3 hours post-transfection is comparable (Figure S2C) indicates that there are no discernible differences in mRNA degradation occurring within the first 3 hours post-transfection. While a slight positive correlation has been observed between tail length and mRNA degradation for endogenous mRNAs<sup>37-40</sup>, our results suggest that the strong influence of poly(A)-tail length on mRNA degradation was previously concealed by the presence of other regulatory elements, that likely dampen the observed effect of tail length on mRNA stability.

Next, to extract the  $k_{tl}$  values of each poly(A)-tail length, we used the measured  $k_{d1}$  values as an input and performed non-linear least square optimization on the tail-specific GFP expression curves (Figure 1B) with the following set of ordinary differential equations (ODEs):

$$\frac{dmRNA}{dt} = -mRNA_t * k_{d1} \quad (\text{Equation 1})$$

$$\frac{dpGFP}{dt} = k_{tl} * mRNA_t - k_m * pGFP \quad (\text{Equation 2})$$

$$\frac{dmGFP}{dt} = k_m * pGFP - k_{d2} * mGFP \quad (\text{Equation 3})$$

where the maturation rate ( $k_m = 0.490$ ,  $t_2^1 = 1.4$  hours) of the pGFP (premature GFP) into mGFP (mature GFP) and the mGFP degradation rate ( $k_{d2} = 0.039$ ,  $t_2^1 = 17.7$  hours) were measured with live cell imaging (Figure S1E) and assumed to be common parameters for all tails. Interestingly, the optimized  $k_{tl}$  values show a similar trend to the expression curves of Figure 1B, with the 100 nt long poly(A)-tail displaying the highest translation rates, and the shorter (30 nt) as well as the longer (150 nt) poly(A)-tails showing lower translation rates (Figure 1G). Strikingly, the 100 nt tail shows a ~2-fold higher translation rate than the 50 tail and ~6-fold higher translation rate than the 150 tail. In both PCR and IVT products we were unable to detect the presence of non-A nucleotides (Figure S2A-

B), indicating that the decrease in translation rate for longer tails is not due to non-A nucleotides in the poly(A) sequence. Although we cannot completely exclude the presence of non-A nucleotides in the poly(A)-tails, unspecific addition of low amounts of non-A nucleotides would likely be length dependent and can therefore not explain the observed peak in translation rates for the 100 tail. Furthermore, we did not observe any tail-specific or cell-size-specific differences in transfection efficiencies (Figure S2C-D). As GFP expression can show significant variation between different IVT libraries (Figure S2E), it is important to consider multiple RNA libraries per tail length. Lastly, the capping efficiency of Vaccinia Enzyme is not known to be affected by changes in the mRNA sequence and length<sup>41</sup>, and indeed different tail lengths did not seem to alter the optimal concentration of capping enzyme for GFP expression (Figure S2F). Taken together, the observed effect of poly(A)-tail length on translation rates is surprising as it demonstrates a clear decoupling between mRNA degradation and translation.

### Deadenylation kinetics are not the main determinant of tail-specific degradation rates

Next, we sought to identify the mechanism underlying the relationship between tail length and degradation rate. Previous reports indicate that deadenylation is a limiting step of mRNA degradation<sup>12,42,43</sup>. Therefore, tail-specific deadenylation rates might explain the positive correlation observed between mRNA tail length and degradation rate. We thus focused on the 50, 100 and 150 nt long poly(A)-tails and quantified poly(A)-tail lengths at different time points (3, 5, and 8 hours post-transfection) with direct RNA Nanopore sequencing<sup>16</sup>. The Tailfindr package<sup>44</sup> uses the raw ONT FAST5 data as input and estimates the poly(A)-tail length based on normalization with the read-specific nucleotide translocation rate (Figure 2A). To estimate the accuracy of the measurement, we spiked control mRNAs into untransfected samples. For all tails, the spike-in median tail length is around the expected size (56 nt for the 50 tail, 103 nt for the 100 tail and 142 nt for the 150 tail). If after transfection the deadenylation of mRNAs proceeds slowly, we would expect a gradual decrease of the median poly(A)-tail length over time, due to the accumulation of short-tailed isoforms. Instead, if deadenylation occurs very quickly, the mRNA body would be fully degraded without leaving any short-tailed intermediates, and the median poly(A)-tail length would remain similar to the spike-in.

Interestingly, both the 50 and 150 nt long poly(A)-tails do not show any significant change in the median tail length compared to the spike-ins (Figure 2B-C). Only the 100-tail variant shows a slight reduction in the median poly(A)-tail length compared to the spike-in (Figure 2C). It is however important to mention that Nanopore sequencing might be biased towards the selection of long poly(A)-tailed mRNA species, even if poly(A)+ mRNA enrichment was not performed during library preparation. We therefore used a PCR-based poly(A)-tail assay that estimates tail length<sup>45,46</sup> to confirm the absence of short tailed intermediates for the 100 tail (Figure S3A-B). This suggests that deadenylation proceeds quickly and with a comparable speed for all the analyzed tails. Furthermore, the data indicate that an initial shortening of the poly(A)-tail, previously observed when endogenous mRNAs exit the nucleus<sup>12</sup>, is likely not required to activate the translation of exogenous mRNAs. Therefore, previous observations linking stark differences in deadenylation rates to poly(A)-tail length are possibly impacted by other RNA regulatory elements that we are not including in our study, such as codon optimality, 3'UTR sequences, and different transcript lengths which have been reported to affect deadenylation rates<sup>27,28,47</sup>. Hence, our observed relationship between poly(A)-tail lengths and mRNA degradation rates (Figure 1F) cannot be explained by differences in deadenylation kinetics. Together, these data indicate that deadenylation is fast in mRNAs differing solely in tail length, and while we cannot exclude tail-specific deadenylation rates, these are likely too fast to be the cause of observed differences in degradation rates.

Since tail-specific deadenylation kinetics do not appear to underlie the observed differences in degradation rates, we proceeded to determine if tail length alters the number of cytoplasmic poly(A)-

binding proteins (PABPCs) that interact with an mRNA. It is known that PABPCs are present in excess compared to mRNAs in differentiated cells<sup>48–50</sup> and that their direct interaction with deadenylases occurs during poly(A)-tail shortening<sup>28,51,52</sup>. Therefore, differences in the number of PABPCs bound to each tail length could explain the observed correlation between mRNA half-life and poly(A)-tail length. For this reason, we performed *in vitro* binding of PABPC1 (the most abundant PABPC in mammals<sup>53</sup>) to mRNAs with 50, 100, and 150 nt poly(A)-tails. We visualized the binding of increasing concentrations of PABPC1 by capillary electrophoresis and quantified the band intensity (Figure 2D and Figure S3C). With increasing PABPC1 concentrations multiple peaks (i.e., bands) appear at high molecular weights, which represent single units of PABPC1 binding to the poly(A)-tails. In particular, the 50, 100, and 150 tail can respectively accommodate a maximum of 3, 4 and 5 proteins (Figure 2D). Considering that the PABPC footprint is thought to be ~25 nt<sup>54,55</sup>, the 50 tail was expected to bind only 2 units of PABPC1. The observation that the 50 tail binds 3 proteins could potentially be explained by the first protein partially binding to the 3'UTR. This explanation was further strengthened when performing the assay for the 5 tail (Figure S3D-E). The finding that the number of bound proteins correlates with tail length and mRNA degradation rate has been recently described in yeast as well<sup>37</sup>. This suggests the possibility that the observed relationship between tail length and mRNA degradation rates could be determined by the number of PABPCs that can bind to the poly(A)-tail. Since long tails accommodate more PABPCs, they are expected to trigger deadenylation more easily through their direct interaction with deadenylation complexes<sup>23,28,52,56</sup>. In particular, Schäfer et al. (2019)<sup>23</sup>, have shown a dependence of the Pan2-Pan3 complex affinity on the number of PABPCs. Consequently, mRNAs with long poly(A)-tails would likely be degraded earlier in time. On the other hand, short tails might escape degradation as the low number of PABPCs would only trigger deadenylation later, leading to longer half-lives. Together, the data show that short poly(A)-tails have slower degradation and a lower occupancy of PABPCs, whereas deadenylation kinetics appear comparable to longer tails.

Notably, deadenylation can trigger mRNA degradation through the direct interaction of the deadenylation complexes with Xrn1 or the exosome<sup>57–60</sup>. In the first case, deadenylation is followed by decapping and 5' to 3' degradation by Xrn1<sup>57,61</sup>, while in the second instance, the exosome degrades the mRNA with 3' to 5' directionality<sup>60,62</sup>. Mukherjee et al. (2002)<sup>62</sup> hypothesized a deadenylase dependent targeting of mRNAs for decay. The authors suggested that Ccr4-Not complex, which is known to deadenylate mRNAs with shorter poly(A)-tails<sup>63,64</sup> and low PABPC1 load<sup>28</sup>, could target mRNAs for decapping and 5' to 3' degradation pathways, while the Pan2/Pan3 complex, which is known to deadenylate mRNAs with longer poly(A)-tails<sup>63,64</sup> and high PABPC load<sup>23,65</sup>, could target mRNAs for 3' to 5' degradation by the exosome<sup>59,61</sup>. We therefore considered whether the long and short tails might be degraded with different directionality.

In order to check the directionality of degradation of our synthetic mRNAs, we used a two-color smFISH based approach in which the two ends of the GFP mRNA are targeted with two different sets of probes (Figure S3F, 5' end in blue and 3' end in red). If mRNAs are being degraded from 5' to 3' we expected to observe an accumulation of 3' associated signal, whilst degradation from 3' to 5' would result in accumulation of 5' associated signal (Figure 2E, 5' end in blue and 3' end in red). Using this approach, we counted the spots associated with each end in single-cells, for the 30, 50, 100 and 150 poly(A)-tails (Figure S3G). To determine directionality, we calculated the difference between the number of spots associated with the 3' end signal and the spots associated with the 5' end signal per cell (Figure 2F). Interestingly, samples transfected with mRNAs with 30 nt long poly(A)-tail show a skewness towards the 3' associated signal, suggesting 5' to 3' degradation. The 50 and 100 poly(A)-tails progressively show less skewness compared to the 30 tail. On the other hand, the 150 poly(A)-tails show a slight skewness towards the 5' associated signal, suggesting 3' to 5' degradation (Figure 2F). It is important to highlight that this kind of analysis does not consider the concomitant presence



of individual 5' and 3' spots in the same cell, but only the difference in the total number of spots. Therefore, we cannot rule out that mRNAs with short (i.e., 30 tail) and long (i.e., 150 tail) poly(A)-tails could potentially be degraded in both directions, with one of the two events being more prevalent and leading to the observed distributions. These results are in line with the proposed hypothesis that Pan2/Pan3 complex—which processes longer poly(A)-tails (>110 nt<sup>66</sup>)—triggers the exosome for 3' to 5' degradation. Conversely, shorter tails—deadenedylated by the Ccr4-Not complex—are targeted for decapping and subsequent 5' to 3' degradation. The findings that the 100 tail shows a more symmetrical distribution, suggests that both deadenylation complexes might be competing for the processing of intermediate tail lengths. Although we cannot exclude endonucleolytic cleavage, the data indicate that shorter tails are bound by fewer PABCs, are enriched in 3' signal (implying 5' to 3' degradation), and display slow degradation. On the other hand, longer tails are bound by more PABCs, are slightly enriched in 5' signal (implying 3' to 5' degradation), and display fast degradation.

### **Altering tail length impacts the fraction of actively translated mRNAs while ribosomal distribution remains unchanged**

To determine if the observed peak in translation rate for the 100 nt long poly(A)-tail (Figure 1G) is a result of changes in the ribosomal distribution along the mRNA sequence, we performed ribosome-sequencing on the 50, 100 and 150 tail 5 hours post transfection. The term distribution used here purely refers to the density or RPFs (Ribosome Protected Fragments) on the GFP CDS<sup>67</sup>, and not to the translation efficiency, measured as the percentage of mRNAs associated with one or more ribosomes. Interestingly, at this time point all tails show very similar profiles (Figure 3A-B) comparable to an endogenous control gene (Figure S4A-B), indicating that differences in the ribosomal distribution alone are unlikely to explain the 2-, to 3-fold differences in GFP levels observed already 5 hours post transfection (see Figure 1B). Furthermore, the analysis of subcodon ribosome footprint profiles (Figure 3C) does not reveal striking differences in the P-site occupancy associated with the different frames, suggesting that no poly(A)-tail length is enriched in translation initiation from downstream or out-of-frame starting codons. We next applied the previously described two-color smFISH approach to quantify the ratio of actively translated mRNAs. Other studies that used the same technique previously demonstrated that for actively translated mRNAs, the 3' and 5' ends are further apart compared to non-translated mRNAs, most likely because translating ribosomes maintain the mRNAs in a more linear state (Figure 3D and Figure S4C)<sup>68,69</sup>. In line with this, 5h post-transfection the 100 tail shows the highest percentage (69%) of actively translated mRNAs (i.e., with ends further apart), whilst the 50 and 150 tail show a lower percentage (65.4% and 56.3% respectively) of actively translated mRNAs (Figure 3E-F and S4C). Lastly, the 30 tail has a very similar percentage (55.6%) of actively translated mRNAs to the 150 tail, in agreement with the translation rates measured through live-imaging (Figure 1G).

A previously proposed model suggests that increased ribosome flux leads to faster mRNA degradation<sup>70</sup>. However, our ribosome profiling data shows that there are no striking differences in the distribution of ribosome footprints between identical mRNAs with different poly(A)-tail lengths. While the performed analysis does not provide information on the translation efficiency, tail length seems to impact the % of mRNAs that are actively translating. Therefore, one explanation that reconciles previous work and our own is that because poly(A)-tail length does not alter ribosome density it does not cause a coupling of degradation and translation. Therefore, another mechanism might be responsible for the previously observed coordination between these two processes.

### **Altering translation rates does not impact mRNA stability**

We next sought to confirm the observed decoupling of mRNA degradation and translation. The lack of correlation between mRNA degradation and translation rates implies that the two processes are not

mechanistically linked through changes in poly(A)-tail length. To assess this, we aimed to determine if translation itself could influence mRNA stability. We thus synthesized GFP coding mRNAs with a 100 nt long poly(A)-tail, but different Kozak sequences at the 5' end—motifs that function as translation initiation sites<sup>71</sup>. We inserted mutations in the consensus sequence to create weaker Kozak sequences (Figure 4A and Table S1) that lead to lower GFP expression (indicated with *kz1*, *kz2* and *kz3* from the strongest to the weakest)<sup>72,73</sup>. Changes in the Kozak sequence cause changes in translation initiation rates, but not in the elongation rates, as the Open Reading Frame (ORF) and therefore the codon usage, remains identical<sup>74</sup>. Reduction of the initiation rate should be reflected in lower translation rates. Indeed, the quantification of GFP levels upon mRNA transfection confirmed the expected trend, as *kz1* shows a much higher protein expression compared to *kz2* and *kz3* sequences (Figure 4B). Despite these differences in mRNA translation, we did not observe any significant difference in their degradation (Figure 4C, Table S2).

As it has been shown that mRNAs containing weak Kozak sequences are often associated with increased translation initiation from downstream nested ORFs<sup>75</sup>, we performed ribosome-sequencing (5 hours post transfection) to check if the same is occurring for the GFP mRNA containing the *kz2* sequence. The profiles in the RPFs distribution of *kz1* and *kz2* mRNAs show slight differences (Figure 4D-E) compared to an endogenous control gene (Figure S4D-E). These dissimilarities seem to be oriented within the first 100 nt of the CDS sequence, also evident in the P-site occupancy of frame 1 (Figure 4F, top). Conversely, there does not seem to be a prominent increase in translation initiation from downstream (Figure 4F, top) or out-of-frame (Figure 4F, bottom) AUGs. This suggests a reduction in the translation initiation from the first AUG codon in the mRNA containing the *kz2* sequence, that could in part explain the differences observed in the GFP expression (Figure 4B). However, there does not appear to be an obvious increase in translational initiation either from downstream or out-of-frame AUGs that could account for the stark reduction in GFP expression observed even 5 hours post transfection. Overall, these results show that changes in the Kozak sequence affect protein expression but do not alter mRNA stability, in agreement with other studies<sup>76</sup>. Taken together, this enforces a mechanistic decoupling of mRNA degradation from translation.

## Introducing variability in the poly(A)-tail length

To determine the physiological relevance of our observations, we extended our study to endogenous mRNAs and proceeded to quantify their tail length by nanopore sequencing (Figure S5A-D). Specifically, we quantified the mean poly(A)-tail length and intragenic poly(A)-tail length variability (Fano factor =  $\sigma^2/\mu$ ) of ~700 endogenous genes. Interestingly, the analysis revealed an average endogenous poly(A)-tail length of 101 nt, which is strikingly close to the tail length we determined to have the highest translation efficiency (i.e., 100 nt). The quantification of the intragenic poly(A)-tail length variability showed that most endogenous mRNAs display a broad steady-state distribution of tail lengths (Figure 5A and S5E-F), as previously reported<sup>12,13</sup>. However, the approach we have followed thus far assumes that each gene is present in the cytoplasm with a single poly(A)-tail length.

Therefore, to mimic more closely the behavior of endogenous mRNAs, and study the effect of this variability on protein expression and mRNA stability, we recreated the endogenous distribution of poly(A)-tail lengths by combining different ratios of synthetic mRNAs with defined poly(A)-tail length (Figure 5B and Figure S5G). The mean tail length of the synthetic distribution is 92 nt and the theoretical  $k_{d1}$  is 0.379 (obtained from the weighted average of  $k_{d1}$  of each tail, see STAR Methods for details). Since these values are similar to those of the 100 tail ( $k_{d1} = 0.363$ ), this tail length was used for comparison. As expected, the mixed population shows lower expression than the 100 tail (Figure 5C, in grey), because of the greater fractional abundance of more suboptimal sequences compare to the 100 tail (Figure 1B). The  $k_{d1}$  measured experimentally with smFISH is close to the predicted one ( $k_{d1} = 0.310$ , see STAR Methods) and similar to the degradation rate of the 100 tail

(Figure 5D). These results indicate once more that mRNA translation and degradation are regulated in an independent manner through changes in poly(A)-tail length. In fact, the introduction of variability in the poly(A)-tail length decreases the translation rate (Figure 5C, right panel) without affecting the average degradation rate of mRNAs. Notably, these results provide a possible explanation for why it has thus far been difficult to link poly(A)-tail length to translational efficiency of endogenous mRNAs. Previous studies assumed that mRNAs are present in the cytoplasm with a single poly(A)-tail length<sup>39,49,77</sup>, however, two genes with similar mean tail length but different variability are characterized by strikingly different translation rates (Figure 5C). Therefore, variability in tail length needs to be considered when studying the effect of the poly(A)-tail on mRNA and protein biogenesis of endogenous genes. In summary, because of the decoupled effect of the poly(A)-tail length on mRNA degradation and translation (i.e., monotonic versus peaked function), the translation rate of the mixed tail population decreases, while the degradation rate stays the same.

### Poly(A)-tails impact both amplitude and frequency of protein fluctuations

With the findings that tail length independently alters mRNA translation and degradation, and knowing that kinetic parameters influence gene expression noise<sup>78,79</sup>, we next sought to identify the impact of poly(A)-tail length on noise. To this end, we performed single-cell tracking combined with time lapse microscopy (Figure 6A and Figure S6, left panel) of cells transfected with mRNAs corresponding to discrete specific tail lengths as well as the synthetic distribution of tails (Figure 5B). Single-cells were tracked for 5h after mRNA transfection and tracks were selected by applying a set of filters (see STAR Methods for details). From each single-cell GFP expression trajectory we quantified the following three parameters per tail: i) the amplitude of fluctuations in the GFP levels over time (measured as the variance after detrending each trajectory); ii) the frequency of these fluctuations (measured as value  $t_{1/2}$ , i.e., the time at which the autocorrelation of the detrended track is equal to half of its initial value); and iii) the cell-to-cell variability in the translation rate (measured as the Fano factor – i.e. variance of the  $k_{tl}$  over the mean  $k_{tl}$  of each specific poly(A)-tail (Figure 6B)<sup>80–82</sup>.

When analyzing the experimental single-cell tracks, it is important to consider the following two aspects that can conceal the real trend of the data: i) the technical noise arising from imaging, cell-tracking and cell-segmentation; ii) and the biological extrinsic noise arising, for instance, from the variability in the number of mRNAs transfected into each cell (Figure S2C). These two elements might result in noisier protein fluctuations and would attribute higher  $k_{tl}$  values to cells that received more mRNAs. Therefore, to identify solely intrinsic effects of poly(A)-tail lengths, we performed stochastic simulations (based on Figure 1C and equations 1, 2 and 3) that capture our system using the previously measured rate constants as input parameters<sup>83,84</sup> and that do not include any source of extrinsic noise, as they start with a defined number of mRNAs at  $t_0$  (Figure S6, right panel, see STAR Methods).

Interestingly, the GFP fluctuations that show the highest amplitude (i.e., variance in the detrended trajectories) are expressed from mRNAs with a poly(A)-tail length of 100 nucleotides. Further, there is an almost symmetrical decrease in fluctuation amplitude for the longer and shorter tails (Figure 6C-D). This implies that the amplitude of the protein fluctuations is primarily influenced by the  $k_{tl}$ , where higher  $k_{tl}$  are associated with larger fluctuations. To explore this relationship, we expanded our simulations beyond the experimentally measured parameters for each tail, testing all possible combinations of  $k_{tl}$  and  $k_{dl}$ , while still remaining in the range of the experimentally measured values. The heatmap in Figure 6E confirms the dependence of the amplitude of protein fluctuations on  $k_{tl}$ , and does not seem to be considerably influenced by changes in  $k_{dl}$ . Surprisingly, the frequency of protein fluctuations decreases (i.e., the autocorrelation  $t_{1/2}$  increases) with the poly(A)-tail length (and therefore with  $k_{deg}$ ), implying that longer tails—subject to faster degradation—are associated with



less frequent protein fluctuations compared to shorter and more stable tails (Figure 6F-H). While the trend of the experimental and simulated data is similar, the fluctuation frequency and amplitude values differ. This indicates that although the model is likely a simplification that does not generate the true range of protein fluctuations, the effect of tail length on these fluctuations is nevertheless captured.

We next performed non-linear least square optimization on the experimental and simulated single-cell trajectories with the same set of ODEs used for the bulk expression data (Equations 1, 2, 3). We extracted cell-specific  $k_{dt}$  values and cell-to-cell variability in  $k_{dt}$  (represented by the Fano factor) for each poly(A)-tail length. The amount of extrinsic variability present in the experimental data can be appreciated by comparing the experimental (Figure S7A, green) and simulated  $k_{dt}$  distributions in the presence (Figure S7A, black) and absence (Figure S7A, brown) of mRNA transfection variability. However, the aforementioned time-resolved fluctuations are not drastically affected by this extrinsic variability, as the simulated fluctuations with and without variability in mRNA numbers are characterized by similar amplitudes and frequencies (black dot and green dot respectively in Figure 6D and G). Furthermore, a correlation between mean  $k_{dt}$  and cell-to-cell variability in  $k_{dt}$  is observed in both the experimental data (Pearson's  $r=0.79$ , Figure S7B) and simulated data, modelled in the absence of extrinsic noise (Pearson's emerges (Figure S7C-D).

Having established that the poly(A)-tail can provide cells with the ability to fine-tune protein fluctuations through decoupling of mRNA degradation and translation, we sought to decipher how variability in tail length might impact protein fluctuations. When comparing more closely the 100 tail to the mixed population of tails, for similar mean poly(A)-tail length (i.e., 100 in both cases), the mRNA degradation is comparable while translation rate is very different (Figure 5C-D). Therefore, a distribution of tail lengths centered around 100 nucleotides generates lower amplitude yet similar frequency protein fluctuations compared to a fixed 100 nucleotide tail length (Figure S7F-I). Hence, when regulating the amplitude of protein fluctuations cells could either shorten or lengthen poly(A)-tails so that the translational rate is decreased (i.e., towards a tail length of 30 or 150 nucleotides). However, as changes in the average tail length would be associated with changes in the degradation rate, the frequency of protein fluctuations would also be impacted. Instead, replacing a single poly(A)-tail length with a distribution of tail lengths centered around the same mean would allow cells to regulate the amplitude of fluctuations without affecting their frequency, as the average degradation rate would remain unchanged.

The observation that higher degradation rates result in lower frequency fluctuations and that this frequency appears to plateau around intermediate  $k_{dt}$  levels (Figure 6F-H) is surprising because it is more intuitive that high degradation rates result in higher frequency fluctuations. This suggests that our experimental system may not fully capture the behavior of endogenous genes, where mRNAs are also actively transcribed. Therefore, we adapted our model to include a two-state transcription model<sup>85</sup>, where the promoter toggles between an OFF and ON state (defined by  $k_{off}$  and  $k_{on}$ ) and transcribes mRNAs at a rate defined by  $k_t$  (Figure 6I, see STAR Methods). As expected, the amplitude of protein fluctuations still shows a similar dependence on  $k_{dt}$  (Figure 6J), however tails with a higher degradation rate are characterized by more frequent fluctuations (i.e., the autocorrelation  $t_{1/2}$  is lower) when transcription is occurring (Figure 6K). Finally, the additional simulations including the two-state transcription model (Figure 6I), confirm that the Fano in  $k_{dt}$  increases with increased translation rate (Figure S7D). In conclusion, these simulations indicate that when mRNAs are endogenously expressed, longer poly(A)-tails might generate protein fluctuations with higher frequency than shorter tails, while higher amplitude fluctuations are instead associated with intermediate tail lengths that exhibit high  $k_{dt}$ . Taken together, these data suggest that by decoupling mRNA degradation and translation cells could tune amplitude and frequency of protein fluctuations through changes in the poly(A)-tail length.

## 350 DISCUSSION

The involvement of poly(A)-tails in mRNA translation and degradation has long been known and has been extensively reviewed<sup>86–88</sup>. However, to date there are still ambiguities in defining the effect of poly(A)-tail length on mRNA degradation and translation kinetics. This arises because endogenous mRNAs' fate is established by synergistic (or opposing) effects of multiple regulatory elements and not solely by the poly(A)-tail. Therefore, there might be many forms of compensation that conceal the real role of poly(A)-tail length in gene expression kinetics. For example, Lima et al. (2018)<sup>39</sup> observed that transcripts with high percentages of optimal codons have relatively short poly(A)-tails, while transcripts with lower codon optimality had longer, more diffuse tail sizes. Furthermore, 360 alternative polyadenylation sites of the same gene are linked to different poly(A)-tail lengths<sup>13</sup> and the consequent changes in the 3'UTR sequences can directly affect deadenylation rates<sup>27,89</sup>. This is further complicated by the fact that most endogenous mRNAs are present in the cytoplasm with a distribution of poly(A)-tail lengths, rather than a single, defined length (Figure 5A)<sup>12,13</sup>. As intragenic variability in poly(A)-tail length has been observed only recently and its role is still unknown, this could have influenced previous efforts to link poly(A)-tail length to mRNA translation and degradation kinetics.

In order to discern the role of poly(A)-tail length and its variability in protein expression regulation, we synthesized identical mRNAs that only differ in tail length. We first studied the effect of single 370 poly(A)-tail lengths on mRNA degradation and translation kinetics (Figure 1), and found that in the cell model used (HEK293T/17) there is a strong negative correlation between poly(A)-tail length and mRNA half-life, as previously reported<sup>13,38–40,90</sup>. Strikingly, we observed a very different relationship between poly(A)-tail length and protein levels, where intermediate tails (i.e., 100 nt) show the highest protein production. The obtained results are surprising because for identical mRNAs one would expect mRNA half-life to correlate with protein levels, since long-lived mRNAs would have more time to be translated into proteins. Therefore, these results suggest the presence of an independent regulation of mRNA translation and degradation through changes in poly(A)-tail length. This decoupling was further confirmed by synthesizing mRNAs with variable Kozak sequences but identical poly(A)-tails (Figure 4). Indeed, these mRNAs were characterized by different levels of 380 protein expression, due to differences in translation initiation rates, but identical degradation rates. Together, these data indicate that the poly(A)-tail length is the main determinant of the stability of otherwise identical mRNAs.

For Nanopore-sequenced tail lengths (50, 100 and 150 nt), we observed no significant modification in tail length over time (Figure 2), indicating rapid deadenylation across all tails. It therefore appears that the main determinant of mRNA half-life might be the moment in which initiation of degradation is triggered, rather than the speed of the poly(A)-tail shortening. Therefore, we propose that the number of PABPCs bound to each tail could influence mRNA stability<sup>37</sup>. We show that the number of PABPCs that can be accommodated on the tails correlates with tail length and therefore mRNA 390 degradation rate (Figure 7A). Furthermore, we identified different directionalities in mRNA degradation depending on tail length (Figure 7B). These results fit a previously proposed model<sup>62</sup>, proposing that the Ccr4-Not complex could target mRNAs for decapping and 5' to 3' degradation pathways, while the Pan2/Pan3 complex could target mRNAs for 3' to 5' degradation by the exosome<sup>59,61</sup>. Our findings are consistent with short tails—known to be deadenylated by the Ccr4-Not complex<sup>63,64</sup>—being targeted for 5' to 3' degradation by Xrn1, while long tails—known to be deadenylated by the Pan2/Pan3 complex<sup>63,64</sup>—are targeted for 3' to 5' degradation by the exosome. This is further reinforced by the known preferential activity of Ccr4-Not and Pan2/Pan3 on tails with low and high PABPCs occupancy respectively<sup>28,24,65</sup>. Taken together, our study provides insights into

the correlation between poly(A)-tail length, PABPCs binding, and mRNA degradation directionality.

Notably, there is an optimal tail length (~100 nt) around which mRNAs are highly translated and this tail length shows the highest % of mRNAs in an actively translating state (Figure 7C). Interestingly, the optimal tail length (100 nt) is almost identical to the mean endogenous poly(A)-tail length we measured in HEK293T/17 cells (101 nt). On the other hand, both very short as well as very long tails are poorly translated (i.e., 5, 30 and 150 tails). Markedly, these tail lengths that show lower translation efficiencies also tend to be less common in our cellular model<sup>90</sup> (Figure 5). It is therefore possible that translation efficiency is an evolutionary pressure that has selected for many transcripts to have this optimum tail length of approximately 100 nt. This however raises the question of why endogenous mRNAs show variability in their poly(A)-tail length (Figure 5). Our results show that a distribution of tail lengths (centered around a mean of ~100 nt) has a similar degradation rate to the 100 tail but much lower protein expression. This confirms the decoupling between translation and degradation observed for mRNAs with uniform poly(A)-tail length. Furthermore, it could explain why previous studies have failed to observe a direct link between poly(A)-tail length and translation efficiency, since mRNAs with identical mean poly(A)-tail lengths can show drastic differences in translation rates depending on the variability in the tail length.

We further explore how poly(A)-tail length impacts gene expression noise (Figure 6). We find that the decoupling of mRNA degradation and translation through the poly(A)-tail allows for independent tuning of protein fluctuation amplitude and frequency. Specifically, by exploiting a two-state transcription model we simulated mRNA transcription, and found that long poly(A)-tails, which are degraded faster, tend to be associated with high frequency protein fluctuations (Figure 7D). Instead, the amplitude of this fluctuations increases with the translation rate (Figure 7D), and is therefore higher for intermediate tail lengths. Further, the introduction of variability in the poly(A)-tail length allows the independent tuning of fluctuations amplitude without affecting their frequency. While gene expression noise is heavily implicated in physiology and pathology<sup>91-94</sup>, post-transcriptional noise regulatory mechanisms still remain scarce<sup>95</sup>. It is possible that changes in poly(A)-tail length and variability throughout disease progression<sup>33,34</sup> or developmental processes<sup>90</sup> is a strategy cells implement to fine-tune gene expression fluctuations.

In summary, our study highlights the crucial role of poly(A)-tail length in protein expression regulation, and advances the current understanding by demonstrating the existence of decoupled mRNA translation and degradation through changes in poly(A)-tail length and variability. We do not exclude the presence of different behaviors in other cellular systems. In particular, it would be interesting to use the same approach to study poly(A)-tail length behavior in undifferentiated cells, where endogenous poly(A)-tails have already been reported to have different effects on mRNA translation and degradation kinetics, compared to differentiated cells<sup>49</sup>.

## ACKNOWLEDGMENTS

We thank Hans Heus, Tom de Greef and members of the Hansen lab for the thoughtful discussions and suggestions; Quan Ta, Femke Michels and Louise van Hoof for initial help with the project; Jelle Postma and Aigars Piruska for the assistance with microscope imaging. M.M.K.H acknowledges generous support from Radboud University, the Christine Mohrmann fellowship, the Netherlands Organization for Scientific Research (NWO) [OCENW.XS3.055, OCENW.XS21.2.050, VI.Vidi.223.065], and Oncode Institute, which is partly financed by the Dutch Cancer Society.

## AUTHOR CONTRIBUTIONS

450 C.G. and M.M.K.H. conceived and designed the study. C.G., M.E., F.H.T.N., L.W.M.R., Y.P. and O.E. performed the experiments. C.G., M.E. and M.M.K.H analyzed and interpreted the cell-imaging data. C.G. performed the stochastic simulations and the bioinformatic analysis. C.G. and M.M.K.H analyzed and interpreted the stochastic simulations. C.G. and M.M.K.H. wrote the paper.

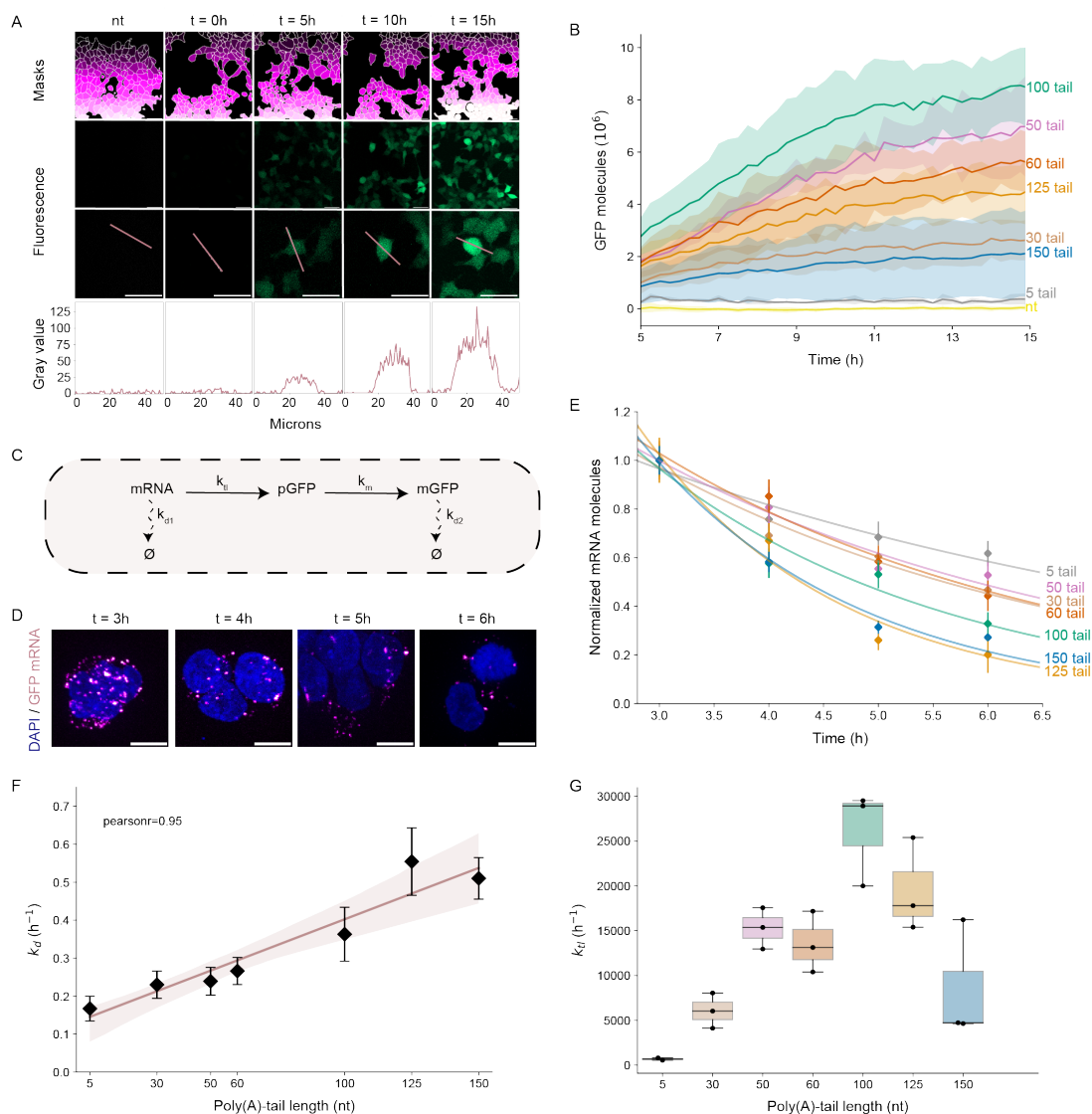
# **CONFLICT OF INTEREST**

Authors declare no competing interests.



## 460 FIGURES AND LEGENDS

Figure 1



**Figure 1. The poly(A)-tail decouples mRNA degradation and translation.**

(A) Synthetic GFP coding mRNAs are translated upon cell-transfection and the increase in GFP intensity is quantified over time (scale bar = 50  $\mu\text{m}$ ).

(B) Direct measurement of average protein numbers expressed in cells transfected with synthetic GFP coding mRNAs reveals the presence of an optimal tail length (100 nt) around which mRNAs are highly translated. 500-750 cells are considered for each biological replicate, where  $n=2$  for the 5 tail and  $n=3$  for all the other tails. The standard deviation of the biological replicates of each tail is represented by the shaded areas.

(C) Schematic of cytoplasmic mRNA and GFP protein kinetics used to define the set of ODEs (equations 1, 2, 3) needed to extract  $k_{dt}$  values from the GFP expression curves in B.

(D) Representative images of smFISH performed on transfected cells between 3 and 6 hours after mRNA transfection (scale bar = 10  $\mu\text{m}$ ).

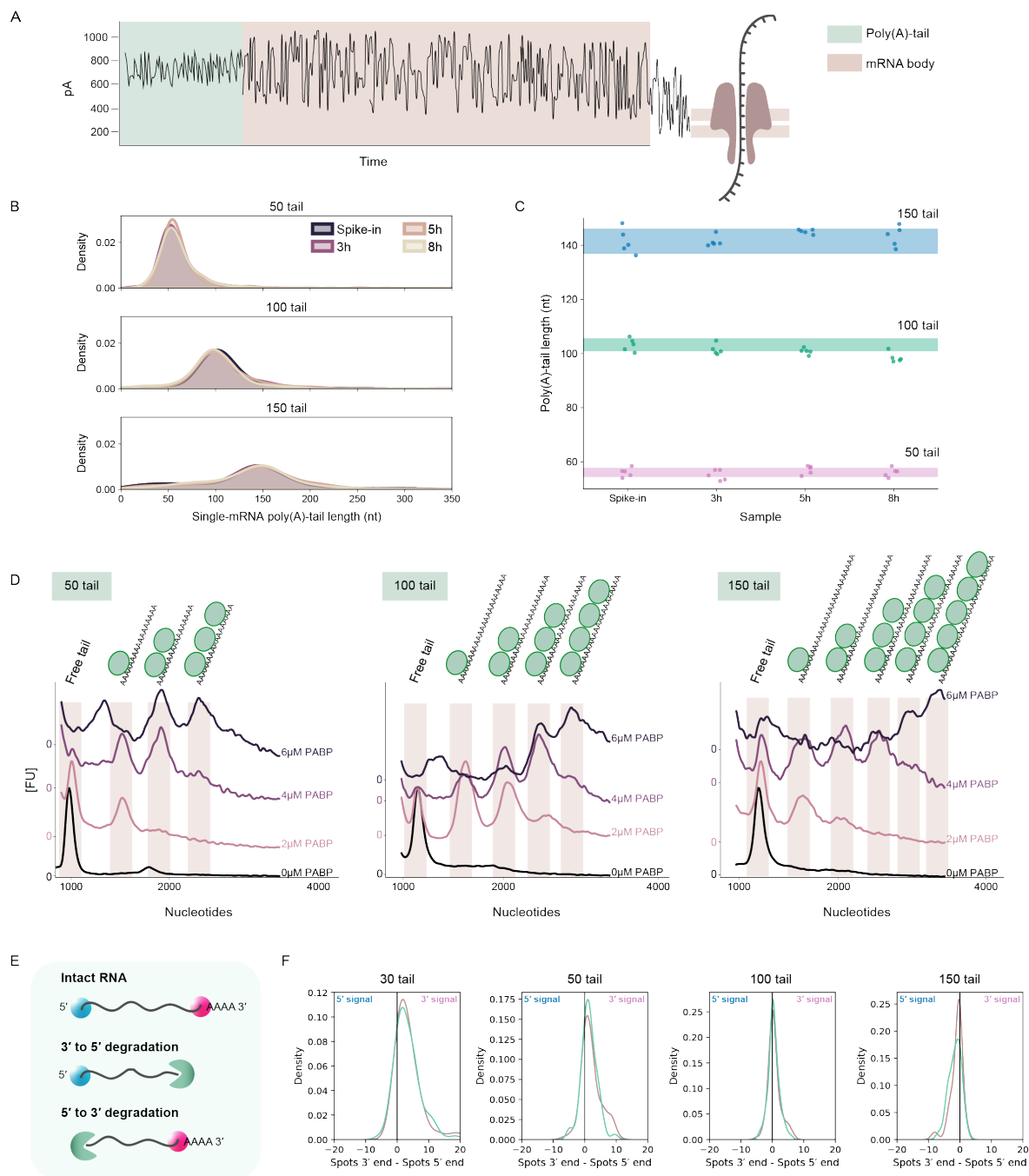
(E) Fitted exponential decay curves of mRNAs with different poly(A)-tail lengths. Each timepoint shows the normalized mean mRNA numbers of  $\sim 100$  cells ( $n=1$  for each time point). The error bars represent the standard error of the mean.

(F) Poly(A)-tail length positively correlates with mRNA degradation rate (Pearson's  $R = 0.95$ ). The error bars represent the standard error of the optimized value of  $k_d$  ( $n = 1$ ), obtained from the fitted

curves of panel E.

480 **(G)** Fitted  $k_{tl}$  values of the experimental GFP expression curves shown in panel B. Data points represent biological replicates (n=2 for the 5 tail and n=3 for all the other tails). The box plot shows the median and the interquartile range, and the whiskers represent the dispersion of the data.

Figure 2



**Figure 2. Determinants of the relationship between poly(A)-tail length and mRNA degradation**

**(A)** Quantification of poly(A)-tail length from ONT FAST5 reads. Poly(A)-tails are identified by their distinct shape compared to the mRNA body.

490 **(B)** Distributions of poly(A)-tail lengths 3-8 hours post transfection quantified with Nanopore sequencing (n=1 per sample). Spike-in mRNAs are used to define the accuracy in the poly(A)-tail length measurement.

**(C)** Changes in the median poly(A)-tail length over time from subsampled reads. Each point represents one subsampled dataset (50 reads subsampled 5 times from the initial dataset). The shaded area represents the standard deviation of the subsampled datasets from the spike-in poly(A)-tails (50 reads subsampled 5 times).

500

**(D)** *In vitro* binding of PABPC1 to identical mRNAs with different poly(A)-tail length, measured with capillary electrophoresis (n=1 per concentration of PABPC1). Each peak corresponds to a single PABPC1 unit binding to the tail. Peaks are normalized to the highest value of each sample and plotted for comparison. Zero values on the y axis correspond to the minimum value of each assay (0  $\mu$ M, 2  $\mu$ M, 4  $\mu$ M and 6  $\mu$ M PABPC1).

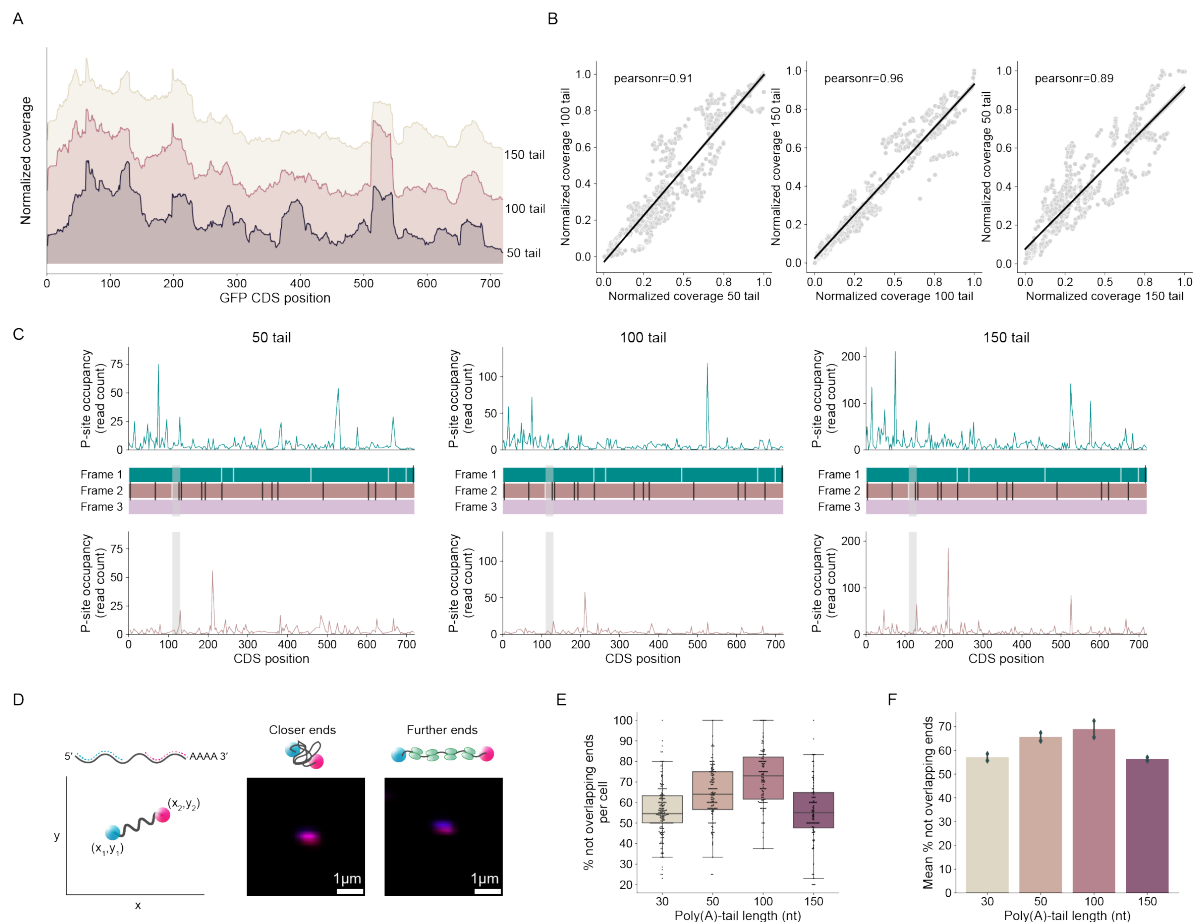
**(E)** Schematics of the two-color smFISH approach used to identify the directionality of degradation. 3' to 5' degradation would cause accumulation of the 5' associated signal (blue), and 5' to 3' degradation would cause accumulation of the 3' associated signal (red).

**(F)** Density distributions obtained by calculating the difference between the 3' and 5' signal counted in single cells (~100-150 cells considered, the green and pink curves refer to 2 biological replicates). A negative skewness (towards the left) indicates accumulation of the 5' associated signal, and a positive skewness (towards the right) indicates accumulation of the 3' associated signal.

510



Figure 3



**Figure 3. High translation rates are associated with high ratios of actively translated mRNAs.**

(A) Normalized read coverage of ribosome protected fragments along the GFP CDS of mRNAs with different poly(A)-tail lengths (50, 100 and 150 nt; n=1 per sample).

(B) Correlation between read coverage along the GFP CDS of mRNAs with 50, 100 and 150 nt long poly(A)-tails (n=1 per sample).

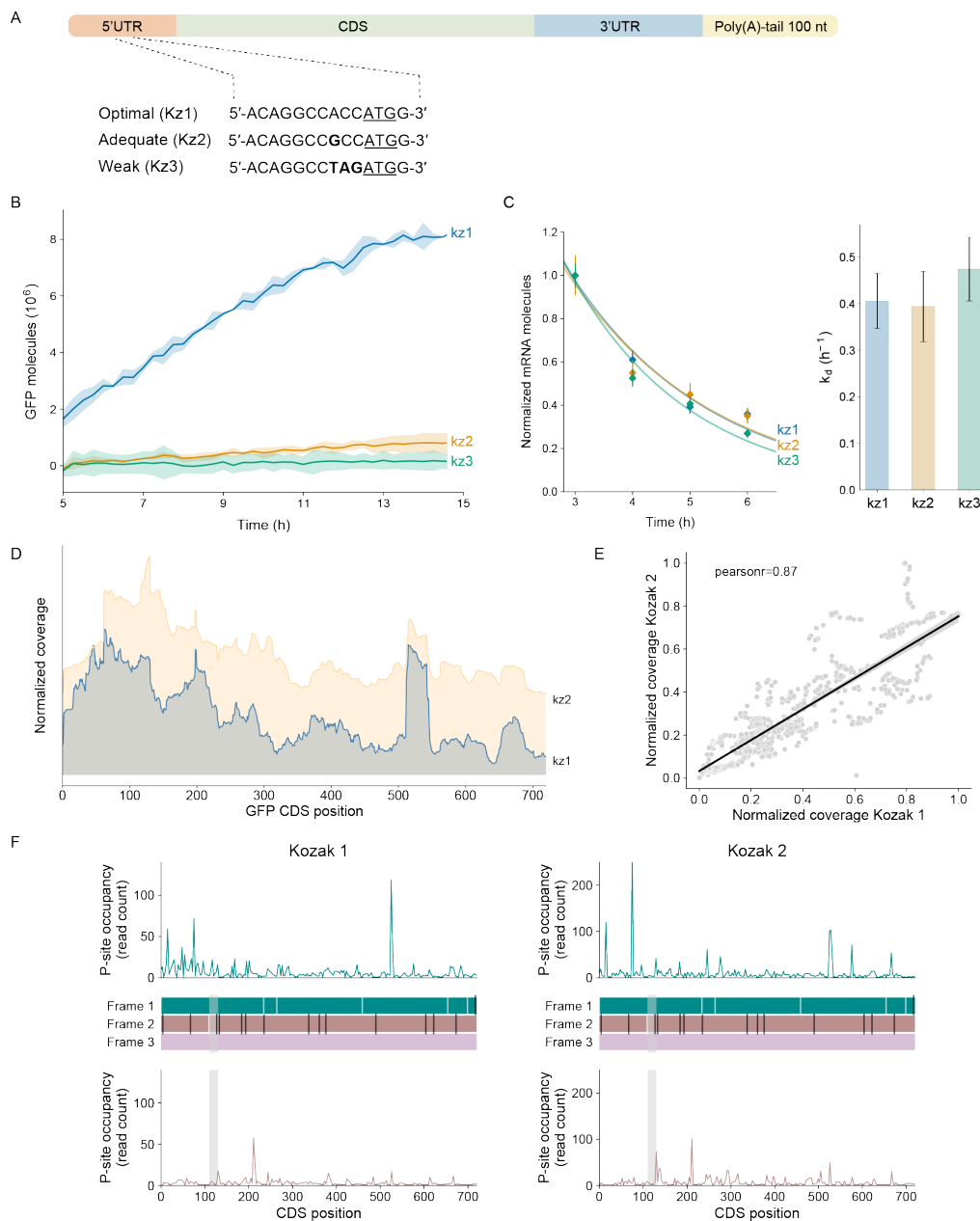
(C) Subcodon ribosome footprint profiles for GFP mRNAs with different poly(A)-tail lengths (50, 100 and 150 nt; n=1 per sample). (*Top*) Density profile of footprints translating the CDS Frame 1; (*middle*) ORF plot where the white dashes indicate AUG codons and black dashes indicate stop codons; (*bottom*) Density profile of footprints translating the CDS Frame 2. Nested ORF identified in Frame 2 is highlighted by the grey shaded area.

(D) Illustration depicting the measurement of the distance between the 3' and 5' ends. Each end is associated with a pair of xy coordinates obtained by identifying the local maxima from the max projection of the z-stack. Closer ends result in overlapping signals and indicate non-translating mRNAs, whilst ends that are further apart exhibit distinguishable signals and indicate translating mRNAs<sup>68,69</sup>.

(E) Percentage of not overlapping ends measured in single cells, for a single biological replicate of the 30, 50, 100 and 150 poly(A)-tails (100-150 cells). The box plot shows the median and the interquartile range, and the whiskers represent the dispersion of the data.

(F) Mean percentage of not overlapping ends measured for the 30, 50, 100 and 150 poly(A)-tails in two biological replicates. The error bars indicate the standard deviation between two replicates (n=2), where each replicate consists of 100-150 cells.

Figure 4



**Figure 4. Mutating the Kozak sequence does not impact mRNA stability.**

(A) Variation of the Kozak sequence in the 5'UTR displayed from the most adequate to the weakest. (B) Mutations in the Kozak sequence of identical mRNAs affect protein expression. Around 500-750 cells are considered for each biological replicate ( $n=2$ ). The standard deviation of the biological replicates of each Kozak sequence is represented by the shaded areas.

(C) Left: fitted exponential decay curves of mRNAs with different Kozak sequences but identical poly(A)-tail length, where each timepoint shows the normalized mean mRNA numbers of  $\sim 100$  cells ( $n=1$  for each time point), and the error bars represent the standard error of the mean. Right: degradation rates for the different Kozak sequences analyzed, where the error bars represent the standard error of the optimized value of  $k_d$  ( $n=1$ ) obtained from the fitted curves (left).

(D) Normalized read coverage of ribosome protected fragments along the GFP CDS of mRNAs containing the optimal (Kozak1) and adequate (Kozak2) Kozak sequences and 100 nt long poly(A)-

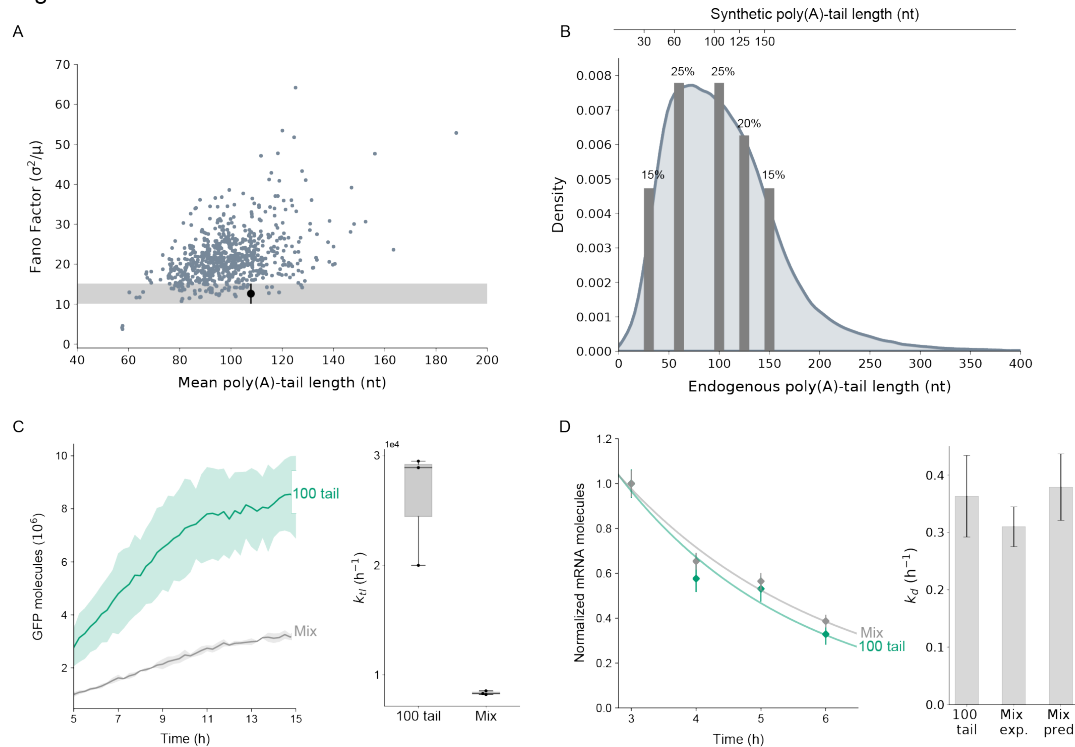
550 tail length (100 nt; n=1 per sample).

**(E)** Correlation between read coverage along the GFP CDS of mRNAs containing the optimal (Kozak1) and adequate (Kozak2) Kozak sequences and 100 nt long poly(A)-tail length (n=1 per sample).

**(F)** Subcodon ribosome footprint profiles for GFP mRNAs with different Kozak sequence and 100 nt long poly(A)-tail length (n=1 per sample). (*Top*) Density profile of footprints translating the CDS Frame 1; (*middle*) ORF plot where the white dashes indicate AUG codons and black dashes indicate stop codons; (*bottom*) Density profile of footprints translating the CDS Frame 2. Nested ORF identified in Frame 2 is highlighted by the grey shaded area.

560

Figure 5



### Figure 5. Introducing variability in the poly(A)-tail length.

(A) Most of the endogenous mRNAs of HEK293T/17 cells show variability in their poly(A)-tail length. The black dot represents the 100 tail spike-in, used to define the technical noise in the quantification of the poly(A)-tail length as after its synthesis it should contain only a single isoform with defined tail length. The shaded area and the error bar represent the standard deviation, obtained by subsampling 5 times 50 reads. Each green dot represents a single mRNA species with at least 50 reads.

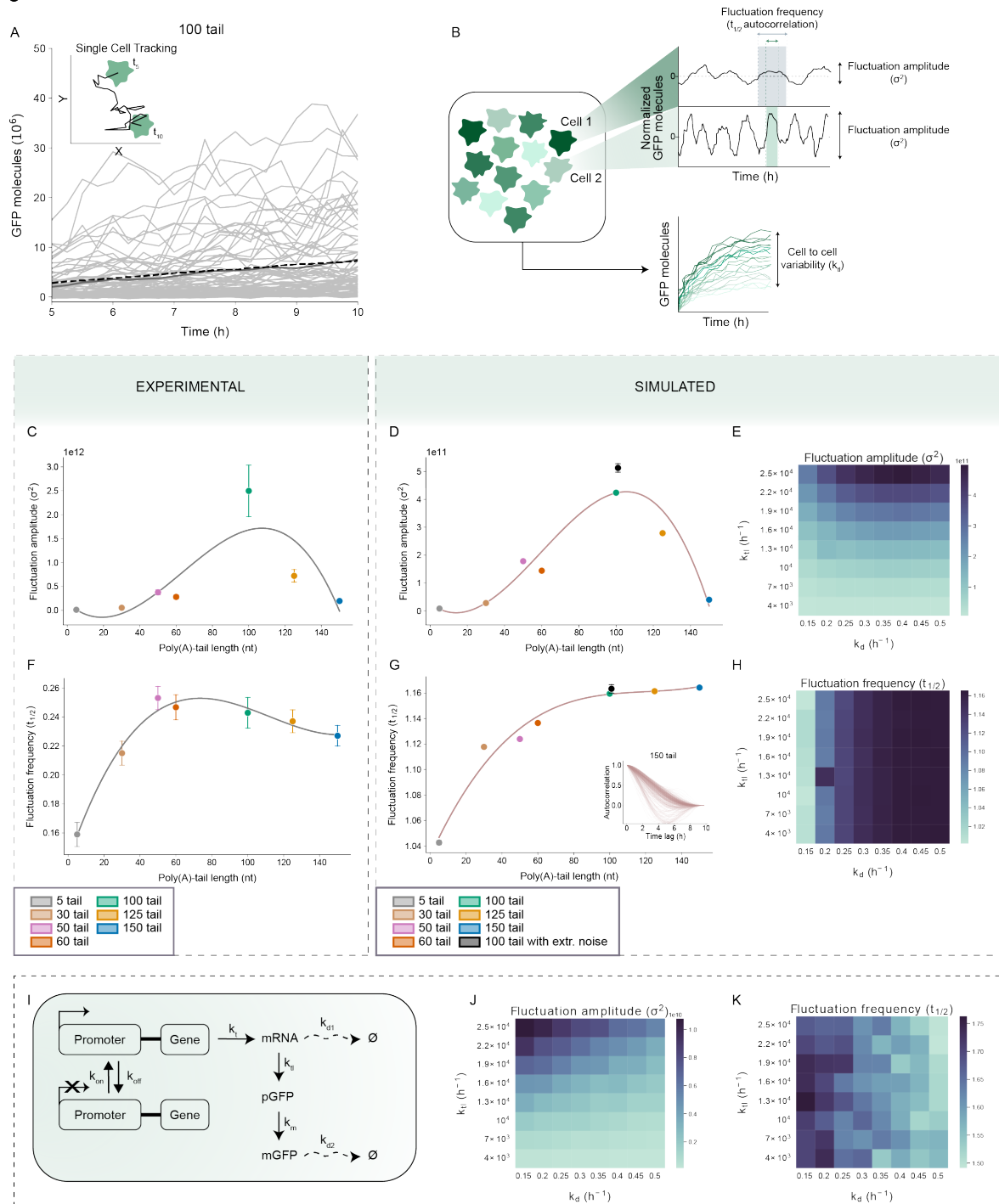
570 (B) Recreation of the endogenous poly(A)-tail length distribution of HEK293T/17 cells by mixing the synthetic mRNAs with defined tail length in different ratios. The artificial distribution has a mean tail length centered around 92 nt.

(C) Left: protein expression of the mixed population of poly(A)-tails compared to the 100 nt long tail. Right: optimized values of  $k_d$  of the experimental GFP expression curves. 500-750 cells are considered for each biological replicate ( $n=3$ ). The standard deviation of the biological replicates of each tail is represented by the shaded areas.

580 (D) Left: fitted exponential decay curves of the 100 nt long poly(A)-tail and the artificial distribution of poly(A)-tails (green and grey respectively, left panel). Each timepoint shows the normalized mean mRNA numbers of ~100 cells ( $n=1$  for each time point), and the error bars represent the standard error of the mean. Right: the degradation rate of the mixed population of poly(A)-tails measured experimentally (Mix exp.) compared to the predicted one (Mix pred.) and to the one of the 100 tail, where the error bars represent the standard error of the optimized value of  $k_d$  ( $n = 1$ ).



Figure 6



**Figure 6. Deducing the role of poly(A)-tail length in noise regulation.**

(A) Tracking of single-cells from a population of cells transfected with mRNAs with 100 nt long poly(A)-tail, obtained from time-lapse microscopy experiments. Each line represents the GFP levels that increase over time in each single cell that was tracked. The solid line represents the mean of the trajectories left after filtering (see STAR Methods), while the dashed line represents the measurement from the whole population of Figure 1B.

(B) A population of cells translating identical mRNAs can show variability in the translation rates and protein levels, due to the stochastic nature of gene expression. Inside individual cells, protein levels fluctuate over time with a particular frequency (autocorrelation time,  $t_{1/2}$ ) and these fluctuations in turn vary in amplitude ( $\sigma^2$ ).

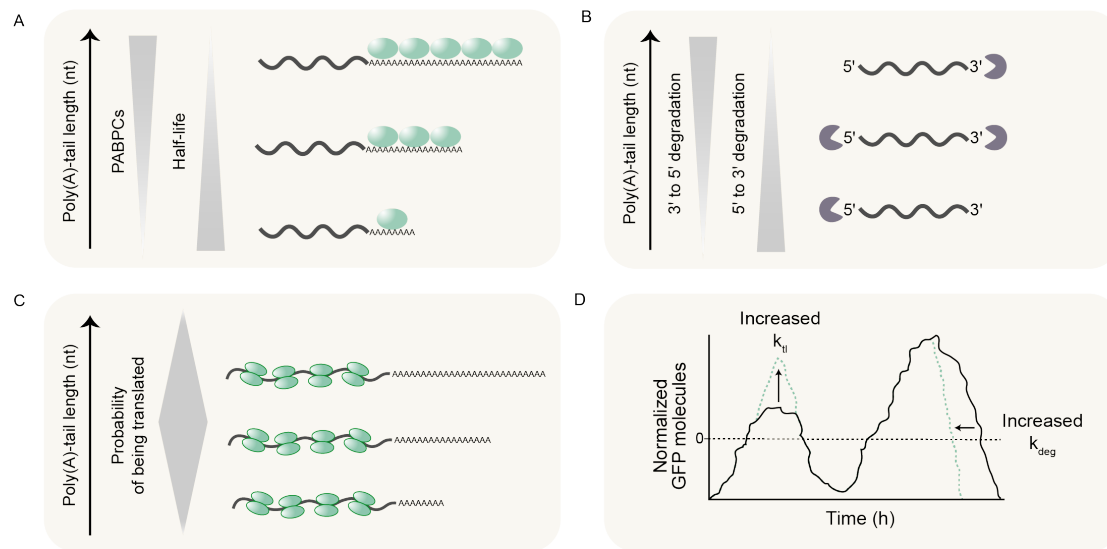
600 **(C-E)** The mean amplitude (measured as  $\sigma^2$ ) in protein fluctuations versus the poly(A)-tail length for both experimental (left) and simulated (central and right) data. The error bars in C and D represent the standard error of the mean. (C) The following number of cells were considered per tail length: 5 tail – 35, 30 tail – 148, 50 tail – 204, 60 tail – 170, 100 tail – 79, 125 tail – 157, 150 tail – 175. (D) 500 iterations (i.e., simulated cells) were considered for all tails except for the 100 tail with extrinsic noise where 1000 iterations were considered. (E) To save computational time, only 300 iterations (i.e., simulated cells) were considered for each combination of rates.

610 **(F-H)** The mean protein fluctuation frequency (measured as  $t_{1/2}$  of the autocorrelation time) versus poly(A)-tail length, both for experimental (left) and simulated (central and right) data. The error bars in F and G represent the standard error of the mean. (F) The following number of cells were considered per tail length: 5 tail – 35, 30 tail – 148, 50 tail – 204, 60 tail – 170, 100 tail – 79, 125 tail – 157, 150 tail – 175. (G) 500 iterations (i.e., simulated cells) were considered for all tails except for the 100 tail with extrinsic noise where 1000 iterations were considered. (H) To save computational time, only 300 iterations (i.e., simulated cells) were considered for each combination of rates. Inset in G shows the changes in autocorrelation values through the time lags considered for the simulated 150 nt long poly(A)-tail.

**(I)** Schematic of the two-state transcription model used to simulate GFP expression from different poly(A)-tails with the same transcription rate  $k_{tl}$ .

**(J-K)** Mean fluctuation amplitude and frequency values quantified for all simulated combinations of  $k_{tl}$  and  $k_{dl}$ , using the two-state transcription model described in I. 300 iterations (i.e., simulated cells) were considered for each combination of rates.

Figure 7



# **Figure 7. Poly(A)-tail length effects mRNA degradation, translation, and protein fluctuations.**

**(A)** Schematic showing that poly(A)-tail length impacts mRNA half-life. Longer tails can accommodate more PABPCs compared to short ones. The number of PABPCs that can be accommodated negatively correlates with mRNA half-life.

**(B)** Schematic showing that identical mRNAs with different poly(A)-tail lengths likely undergo degradation with different directionality.

**(C)** Schematic showing that identical mRNAs with different poly(A)-tail lengths are associated with similar ribosomal distribution along the CDS, but differ in the % of actively translated mRNAs.

**(D)** Schematic showing that increase in translation rate is associated to increase in the fluctuation amplitude, while increase in degradation rate causes increase in fluctuation frequency.

## STAR★Methods

### Key Resource Table

REAGENT or RESOURCE	SOURCE	IDENTIFIER
<b>Experimental models: Cell lines</b>		
HEK293T/17	ATCC	CRL-11268
<b>Recombinant DNA</b>		
pEF-GFP	Addgene	Cat#11154
<b>Critical commercial assays</b>		
MEGAscript™ T7 Transcription Kit	Invitrogen	Cat#AMB13345
RNA High Sensitivity Kit	Invitrogen	Cat#Q32852
Poly(A) Tail-Length Assay Kit	Invitrogen	Cat#764551KT
Agilent High Sensitivity DNA Kit	Agilent Technologies	Cat#5067-4626
Agilent Nano 6000 RNA Kit	Agilent Technologies	Cat#5067-1511
QIAquick PCR Purification Kit	QIAGEN	Cat#28104
QIAquick gel extraction kit	QIAGEN	Cat#28704
One-Step Capping and 2'-O496 Methylation Kit	NEB	Cat#M0366
JetMESSENGER® mRNA Transfection Kit	Westburg	Cat#PO 150-01
Direct RNA Sequencing kit	Oxford Nanopore Technologies	Cat#SQK-RNA002
Dynabeads MyOne Streptavidin C1	Invitrogen	Cat#65001
<b>Chemicals, peptides, and recombinant proteins</b>		
Poly-L-Lysine 0.01%	Sigma-Aldrich	Cat#P4707
GFP	Abcam	Cat#ab51992
PABPC1	Biorbyt	Cat#ORB244425
Turbo DNase	ThermoFisher	Cat#AM2238
Ambion RNase I, cloned, 100 U/μL	Invitrogen	Cat#AM2294
T4 RNA Ligase 2, truncated K227Q	NEB	Cat#M0351S
Superscript III	Invitrogen	Cat#18080093
T4 RNA Ligase	ThermoFisher	Cat#EL0021
CircLigase ssDNA Ligase	LGC Biosearch Technologies	Cat#CL4111K
SYBR™ Gold Nucleic Acid Gel Stain	Invitrogen	Cat#S11494
TRIzol Reagent	Ambion	Cat#15596026
DMEM	Gibco	Cat#10313021
FBS	Gibco	Cat#A3160401
<b>Oligonucleotides</b>		
Primers for PCR, poly(A)-tail length assay and Sanger sequencing, see Table S3	This paper	N/A
Probes for smFISH, see Table S5 and Table S6	This paper	N/A
<b>Softwares and algorithms</b>		



Python v3.8	Python Software Foundation	<a href="https://www.python.org/">https://www.python.org/</a>
Numpy (Python package)		
RSudio v1.4	The R Project for Statistical Computing	<a href="https://cran.r-project.org/">https://cran.r-project.org/</a>
TailFindr v1.3 (R package)	Krauze et al. <sup>44</sup>	<a href="https://github.com/adnaniazi/tailfindr">https://github.com/adnaniazi/tailfindr</a>
ImageJ	Fiji Distribution	<a href="https://fiji.sc/">https://fiji.sc/</a>
Labkit (ImageJ plugin)	Arzt et al. <sup>96</sup>	<a href="https://imagej.net/plugins/labkit/">https://imagej.net/plugins/labkit/</a>
Trackmate (ImageJ plugin)	Tinevez et al., Ershov et al. <sup>97,98</sup>	<a href="https://imagej.net/plugins/trackmate/">https://imagej.net/plugins/trackmate/</a>
PSF generator (ImageJ plugin)	Sage et al., Kirshner et al. <sup>99,100</sup>	<a href="http://bigwww.epfl.ch/algorithms/psfgenerator/">http://bigwww.epfl.ch/algorithms/psfgenerator/</a>
DeconvolutionLab2 (ImageJ plugin)	Sage et al. <sup>101</sup>	<a href="https://bigwww.epfl.ch/deconvolution/deconvolutionlab2/">https://bigwww.epfl.ch/deconvolution/deconvolutionlab2/</a>
Cellpose 2D	Stringer et al. <sup>102</sup>	<a href="https://github.com/HenriquesLab/ZeroCostDL4Mic">https://github.com/HenriquesLab/ZeroCostDL4Mic</a>
FISH-quant	Mueller et al. <sup>103</sup>	<a href="https://fish-quant.github.io/">https://fish-quant.github.io/</a>
Minimap2	Li <sup>104</sup>	<a href="https://github.com/lh3/minimap2">https://github.com/lh3/minimap2</a>
ImJoy	Ouyang et al. <sup>105</sup>	<a href="https://imjoy.io/#/">https://imjoy.io/#/</a>
<b>Others</b>		
Minion Mk1b system	Oxford Nanopore Technologies	Cat#MIN-101B
Flow Cell (R9.4.1)	Oxford Nanopore Technologies	Cat#FLO-MIN106D
2100 Bioanalyzer Instrument	Agilent Technologies	Cat#G2939BA
Qubit 4 Fluorometer	Invitrogen	Cat#Q33238
8-well #1.5 polymer coverslip coated with Collagen IV	Ibidi	Cat#80802
ibiTreat 8-well #1.5 polymer coverslips	Ibidi	Cat#80806

## CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by Maïke M. K. Hansen (maïke.hansen@ru.nl).

## 640 EXPERIMENTAL MODEL AND SUBJECT DETAILS

### *In vitro* synthesis of GFP-coding mRNAs with defined poly(A)-tail

650 *Amplification of pEF-GFP plasmid by PCR.* To obtain the DNA template used in the *in vitro* transcription (IVT) reactions, the CDS of interest was amplified by PCR from the pEF-GFP plasmid<sup>106</sup> (Addgene). The forward primer was designed to contain the upstream spacer, T7 promoter, downstream spacer, Kozak sequence of interest and start codon, followed by a gene specific sequence of 22 nt. The addition of these sequences at the 5'UTR is needed to enhance ribosome binding, transcription initiation and translation initiation. The reverse primer used contains a gene specific sequence of 20 nt, followed by an oligo(dT) sequence of defined length (5, 30, 50, 60, 100, 125 or 150) at the 5'end (see Table S3 for primer sequences). This design ensures the binding of the primer just after the  $\beta$ -globin poly(A) signal. As a result, the generated 3'UTR is 225 nucleotides in length,

which is equivalent to what would be obtained by transfecting the pEF-GFP plasmid directly into cells, as described by Matsuda et al.<sup>106</sup>. The 3'UTR sequences have been analyzed for potential miRNA enrichment using the miTEA online tool to rule out any potential off-target effects<sup>107,108</sup>. PCR reactions were assembled into 100 µL of volume and contained a final concentration of 1X Pfu DNA Polymerase buffer, 0.2 mM dNTPs mix, 1 µM forward primer, 1 µM reverse primer, 150 pM plasmid template, and 2.5 units of Pfu DNA polymerase (in-house purified). PCR conditions are reported in Table S4. PCR products were subsequently purified using the QIAquick PCR Purification kit (QIAGEN). The quality of each DNA template was assessed on an agarose gel and by capillary electrophoresis (Bioanalyzer, Agilent High Sensitivity DNA Kit).

*In Vitro Transcription.* The DNA products with pure poly(A)-tails were *in vitro* transcribed with the MEGAscript™ T7 Transcription Kit (Invitrogen) to obtain mRNAs with defined poly(A)-tail lengths. Reactions were assembled according to manufacturer specifications using 15 nM PCR products as the DNA template and incubated at 37°C for 4 hours. DNA template was then removed by adding 2U of Turbo Dnase (ThermoFisher) and incubating at 37°C for 15 minutes. The products of the IVT reactions were purified by the addition of LiCl, precipitation at -20°C for 60 minutes followed by centrifugation for 15 minutes at 4°C at maximum speed. The pellet was washed with 70% ethanol and re-centrifuged at maximum speed to maximize the removal of unincorporated nucleotides. The obtained pellet was then resuspended in nuclease-free water, and the concentration was determined by Qubit™ using the RNA High Sensitivity kit (Invitrogen™). The quality of the obtained mRNAs was assessed by capillary electrophoresis (Bioanalyzer, Agilent RNA 6000 Nano Kit). 10 µg of the RNA products were *in vitro* capped with the One-Step Capping and 2'-O-Methylation kit (NEB). With this kit, capping is nearly 100% efficient and all capped structures are added in the proper orientation (as indicated by NEB). Reactions were assembled in 20 µL and incubated for 60 minutes at 37°C. The final GFP mRNA sequence is reported in Table S1. Two to three replicates of the IVT libraries were synthesized for each poly(A)-tail length and used to perform the experiments to avoid technical variation.

*Purity of tailed PCR products.* An additional (100T)-tailed DNA product containing 5 randomly interspersed As was synthesized by PCR using a specific tailed primer, as previously described (see Table S3 for primer sequence). Tailed PCR products were sequenced by mixing 15 ng/1000 bp DNA with 25 pmol Forward primer (see Table S3 for sequence) in a final volume of 20 µL. Samples were sequenced by Sanger sequencing by Baseclear B.V., Leiden, The Netherlands.

*Purity of tailed mRNAs.* A 5'PO4 and 3'NH2 modified DNA oligonucleotide (linker, see Table Sx for sequence) was ligated to the 3' end of the GFP mRNA with a 50, 100 or 150 nt long poly(A)-tail using T4 RNA ligase. Each 25 µL ligation reaction contained 50 ng RNA, 1 µM of linker, 1X T4 RNA ligase buffer, 1 mM ATP, 12.5% PEG8000 and 10 Units of T4 RNA ligase (ThermoFisher). The reaction was incubated at 25°C for 3 hours at 500 rpm in a thermomixer. The ligation reaction was quenched by the addition of EDTA to 10 mM.

A DNA primer complementary to the linker was annealed to 2.5 µL of the ligation mixture and subsequently reverse transcribed in a 20 µL reaction for 40 minutes at 48°C using Superscript III (Invitrogen) according to the manufacturer's instructions. Subsequently, three PCRs of 20 µL, each containing 2.5 µL of reverse transcription mixture, were carried out using Pfu DNA polymerase (in house produced) and a primer set covering the poly(A)-tail until 300 nt upstream of the tail. Reactions were cycled for 15 rounds (95°C for 30 seconds, 49°C for 30 seconds, 72°C for 60 seconds). Reaction mixtures were loaded on a 1% agarose gel, the amplified DNA product was excised from gel under UV light and purified using a QIAquick gel extraction kit. Sanger sequencing was finally performed by Baseclear BV, Leiden, The Netherlands. See Table S3 for all primer sequences.

## Cell handling

*Culture of HEK293T/17 cells.* HEK293T/17 cells (ATCC) were cultured in Dulbecco's modified Eagle's medium (DMEM, ThermoFisher) supplemented with 4.5 g/L D-Glucose, L-glutamine, Sodium Pyruvate, 10% (v/v) foetal bovine serum (FBS, ThermoFisher) and antibiotic solution (50 U/mL Penicillin and 50 µg/mL Streptomycin) at 37°C, in a humidified 5% CO<sub>2</sub> atmosphere, until reaching a confluency of 70-80%.

*mRNA transfection.* For live-cell imaging experiments, cells were seeded at a concentration of  $5 \times 10^4$  cells/mL in an 8-well #1.5 polymer coverslip coated with Collagen IV (Ibidi) two days before imaging. For smFISH experiments, ribosome profiling and Nanopore sequencing,  $1.5 \times 10^5$  cells/mL were seeded respectively in 6-well (smFISH and ribosome profiling) and 12-well plates (Nanopore sequencing) the day before transfection.

Transfection of the synthetic GFP coding mRNAs was performed using the JetMESSENGER® mRNA Transfection kit (Westburg). Immediately prior to transfection, GFP coding mRNAs were diluted in jetMESSENGER mRNA buffer and JetMESSENGER reagent (mRNA/JetMESSENGER reagent ratio 1:2). The mRNA solution was incubated for 15 minutes at room temperature and then added to the cells in standard growth media to a final concentration of 2.6 nM.

## smFISH

*Probes design.* Stellaris probes were designed using the designer tool from BioSearch Technologies (<https://www.biosearchtech.com>). For one-color smFISH a set of probes was designed to detect the GFP coding sequence (Table S1), using a masking level of 5 and a minimum spacing length of 2 nt between each probe. A total of 30 probes of 18 nt in length were conjugated with TAMRA. See Table S5 for probes sequence. For the two-color smFISH two sets of probes were designed to detect 401 nt on each end of the GFP sequence, using a masking level of 3 and a minimum of spacing length of 0 nt between each probe. A total of 15 and 18 probes of 18 nt in length were respectively conjugated with Quasar for 3' end and with TAMRA for the 5' end. See Table S6 for sequence.

*Sample preparation.* Cells were trypsinized two hours after transfection and the total sample volume was split into aliquots of 250 µL that were seeded in ibiTreat 8-well #1.5 polymer coverslips (Ibidi) previously coated with Poly-L-Lysine 0.01% (Sigma-Aldrich). Cells were fixed 3h, 4h, 5h and 6h post transfection, by incubating in fixation solution (PBS in 4% formaldehyde) for 10 minutes at room temperature, followed by two washing steps with PBS. Cells were then incubated in 70% ethanol, allowing membrane permeabilization for 1 hour at 4°C, followed by two washes with wash buffer (2x SSC and 10% formamide). Probes were diluted in a buffer composed by 0.1 g/ml of dextran sulphate, 2x SSC and 1% formamide to a final concentration of 25 nM, and were let hybridize overnight at 37°C. The following day, cells were washed with wash buffer and shortly incubated with DAPI (0.5 µg/ml in wash buffer, 15-20 minutes at 37°C) and washed with 2x SSC. Cells were imaged in PBS.

*Image acquisition.* Cells were imaged with an Andor spinning disk confocal with FRAP-PA (bleaching, photoactivation) equipped with an Andor iXon 897 EMCCD camera, using a 60x/1.40 NA, oil objective. For each XY location of the one-color smFISH, a z-stack of 21 steps, 0.9 µm each, was taken. DAPI and TAMRA were excited by 405 nm (12% intensity) and 561 nm (10% intensity) lasers respectively, with 300 ms of exposure time. For each XY location of the two-color smFISH, a z-stack of 42 steps, 0.175 µm each, was taken. DAPI, TAMRA and QUASAR were excited by 405 (12% intensity, 300 ms), 561 (20% intensity, 500 ms) and 670 (22% intensity, 500 ms) lasers respectively.

*Image processing and data analysis.* Cell masks were obtained using the background of the DAPI signal in the deep-learning method Cellpose 2D<sup>102</sup>. The pre-trained models from the notebook jointly developed by the Jacquemet (<https://cellmig.org/>) and Henriques (<https://henriqueslab.github.io/>) laboratories were used for this purpose (freely available on GitHub: [HenriquesLab/ZeroCostDL4Mic: ZeroCostDL4Mic: A Google Colab based no-cost toolbox to explore Deep-Learning in Microscopy \(github.com\)](https://github.com/HenriquesLab/ZeroCostDL4Mic)). Clumped cells and wrongly segmented cells were manually excluded from the analysis. Fluorescent spots corresponding to GFP coding mRNAs were detected using the plugin FISH-quant<sup>103</sup> in the computing platform ImJoy<sup>105</sup>. Dense areas were decomposed to avoid under detection of clustered mRNAs. This detection returned as output csv files containing the XYZ coordinates of each mRNA. Spots were then assigned to the corresponding cell mask and counted using in-house Python programs (available upon request).

in the one-color smFISH the mean number of mRNA molecules was calculated for each time point and an exponential decay curve was fitted to the different data points to extract degradation rates and half-life values for the different poly(A)-tails. The theoretical degradation rate of the synthetic distribution of poly(A)-tails was obtained by calculating the weighted mean of the degradation rates of the poly(A)-tails that compose the distribution, where the weights are the percentages of mRNA transfected:

$$k_{d1} = (k_{d1}30tail * 0.15) + (k_{d1}60tail * 0.25) + (k_{d1}100tail * 0.25) + (k_{d1}125tail * 0.20) + (k_{d1}150tail * 0.15) \quad (\text{Equation 4})$$

In the two-color smFISH the number of mRNA molecules was measured for both channels (561 and 670) at the single cell level using the FISH-quant plugin.

To measure the distance between 3' and 5' ends of mRNAs, the two-color smFISH images were further processed. First, a PSF image was generated for both channels using the PSF generator plugin<sup>99,100</sup> in FIJI with the Richards & Wolf 3D optical Model<sup>109</sup>. Images were then deconvolved with the Deconvolution Lab2 plugin<sup>101</sup>, using the Richardson-Lucy algorithm<sup>110,111</sup> (n iterations = 20). Max projections of the deconvolved images were used as input to identify local maxima and the corresponding xy coordinates returned for each channel were analyzed using in-house Python programs. In specific, the distance in pixels between 3' and 5' ends was calculated as the following:

$$\text{Distance in pixels} = \sqrt{(x_{5'end} - x_{3'end})^2 + (y_{5'end} - y_{3'end})^2} \quad (\text{Equation 5})$$

Only ends with distance below 2.5 pixels were considered as part of the same mRNA. Finally, a threshold of  $\geq 1$  pixel was set to define non-overlapping ends (i.e., translating mRNAs).

The intensity analysis of the 3' and 5' associated signal (Figure S4E) was performed in Fiji and data were plotted in Python (intensity was normalized to 1 for each channel).

## Live-Cell Imaging

*Image acquisition.* The live-cell imaging was performed with a SP8x AOBS-WLL confocal microscope (Leica-microsystem), inside a chamber closed with a lid allowing the control of temperature, airflow, CO<sub>2</sub> and relative humidity levels. Since laser intensity varied between experiments, a 100 tail 24h transfected sample and a 50 tail transfected sample were always included to set up the microscope (laser power and exposure time) and normalize the data respectively. Cells were imaged for 10 hours, starting from 5 hours post transfection in order to minimize photobleaching and phototoxicity. Images were acquired every 15 minutes, with a monochrome DFC365FX camera and using a 40x/0.60 NA air objective (3.3 mm long distance). Approximately 5 XYZ positions were imaged in each well, where the z-plane consisted of 7 steps of 3.8  $\mu$ m. Cells were excited at 488 nm with a pulsed White Light Laser (WLL) and the emitted fluorescence was detected with a High Sensitivity Detector (HyD). A normal transmitted light PMT detector was used to acquire bright-field images.



800 *Image processing and data analysis.* Maximum intensity projections of the fluorescence images were obtained and the background was subtracted by using negative control images from non-transfected cells, in order to eliminate autofluorescence effects.

The acquired bright-field images were used to segment cells in the deep-learning method Cellpose 2D<sup>102</sup>. A training dataset was first created by manually labelling a set of images with Labkit in Fiji (<https://imagej.net/plugins/labkit/>). The dataset was then used to train the model in the Cellpose 2D notebook<sup>102</sup> (freely available on GitHub: [HenriquesLab/ZeroCostDL4Mic](https://github.com/HenriquesLab/ZeroCostDL4Mic): [ZeroCostDL4Mic: A Google Colab based no-cost toolbox to explore Deep-Learning in Microscopy \(github.com\)](https://colab.research.google.com/github/HenriquesLab/ZeroCostDL4Mic/blob/main/ZeroCostDL4Mic.ipynb)) jointly developed by the Jacquemet (<https://cellmig.org/>) and Henriques (<https://henriqueslab.github.io/>) laboratories.

810 Images were automatically segmented with the trained model and used in Fiji to quantify the mean GFP intensity of each cell. Masks with an area smaller than 100  $\mu\text{m}^2$ , or bigger than 450  $\mu\text{m}^2$  were assumed to be wrongly segmented cells and excluded. For single-cell analysis, the obtained masks were used to track the cells for the first 5 hours with TrackMate in Fiji ([TrackMate \(imagej.net\)](https://imagej.net/trackmate/))<sup>97,98</sup>, using the LAP tracker algorithm. The data retrieved from Fiji were further analyzed with in-house Python scripts.

Average whole-population tracks and single-cell tracks of each experiment were normalized between the minimal value of non-transfected cells and the maximal value of the average 50-tail GFP expression track used in that same experiment. The whole experiment was discarded if the transfection efficiency was too low (i.e., < 80%), where the transfection efficiency was measured at 15h as the percentage of cells showing higher fluorescence than the non-transfected cells. Wrong single-cell tracks were filtered out by applying the following criteria: i) cells had been tracked for less than 5 hours; ii) cells divided and the tracks split; iii) tracks showed a change in intensity from  $t_n$  to  $t_{n+1}$  that was  $> \pm 50\%$  of  $t_n$  intensity; iv) the last time point at 10h had lower intensity than the first one at 5h; v) the entire track had a negative slope; and vi) the last time point was lower than the mean intensity of non-transfected cells. To convert grey values into numbers of GFP molecules a calibration curve was performed by making serial dilutions of purified GFP (Abcam) in cell culture media (Figure S1D). Using this calibration curve, the average cells intensity of a 50 tail expression curve—acquired at the moment of calibration—was converted into GFP concentration and finally GFP molecules with the following equations:

$$\text{GFP } \mu\text{M} = \frac{\text{GFP intensity}(t_n) - \text{intensity NT cells } (t_n)}{\text{Slope calibration curve}} \quad (\text{Equation 5})$$

$$\text{GFP molecules} = \text{GFP } \mu\text{M} * V_{\text{cell}} * \text{NA} \quad (\text{Equation 6})$$

830 where the volume of each single cell was obtained from the area of each single-cell mask, and NA is Avogadro's number.

To avoid inconsistencies due to changes in the microscope lasers over time, each experiment included a 50 tail sample used for normalization of all the other tails. The normalized curves of the other tails were converted into GFP concentration by multiplying them by the GFP expression curve of the 50 tail control generated at the moment of calibration, and finally into GFP molecules using Equation 6.

840 *Defining mRNA<sub>0</sub>.* A 3h delay in fluorescence onset was considered due to timing needed for endosomal uptake and release of the mRNA molecules in the cells<sup>112,113</sup>. At this time point, the average number of mRNAs per cell quantified experimentally with smFISH was between 100 and 200 molecules (Figure S2C). The cell-to-cell variability in mRNA molecules after transfection is considered to be the same and is not affected by the poly(A)-tail length. We therefore chose 150 as initial number of mRNAs for curve fitting (mRNA<sub>0</sub>).

*Defining k<sub>m</sub> and k<sub>d2</sub>.* The GFP expression curve of the 50 nt tail was extended to include 3 additional timepoints (21, 28 and 44 hours) and the data between these time points are interpolated (Figure S1E). All GFP expression curves were fitted by performing non-linear least square optimization with the

set of ODEs defined by equations 4, 5, and 6 to determine the GFP maturation ( $k_m$ ) and degradation rate ( $k_{decay}$ ). The mRNA degradation rates used as input for the fitting are reported in Table S2, the initial number of mRNAs ( $mRNA_{t0}$ ) was set to 150 (see above), while the tail-specific  $k_{tl}$  values are kept as variable parameters. The maturation and degradation rates extracted of GFP were  $kd_2=0.039$ ,  $k_m=0.49$ .

*Fitting for tail-specific  $k_{tl}$ .* The defined parameters (see above) were used to perform non-linear least square optimization on the tail-specific GFP expression curves of each replicate and of single-cell trajectories (Figure 1B and Figure S6). The parameters used are reported in the Table S7.

## Nanopore sequencing

*Library preparation.* Cell pellets were collected 1h, 3h, 5h and 8h after transfection. Total RNA was extracted using TRIzol reagent (Invitrogen) and its quality was assessed with the Bioanalyzer (Agilent RNA 6000 Nano Kit). The library preparation was performed using the Direct RNA Sequencing kit (Nanopore) according to manufacturer's instructions. Samples sequencing was carried out on a Minion Mk1b system with flow cells FLO-MIN106.

*Data analysis and poly(A)-tail measurement.* FAST5 files were converted into FASTQ files by basecalling raw sequencing data with Guppy and aligned to the reference genome using Minimap2<sup>104</sup>. Quality control of reads and sequencing performance was performed in Python (Figure S3). Poly(A)-tail lengths were estimated from raw FAST5 files using the Tailfindr package in R<sup>44,114</sup>. This analysis returned a list of the estimated tail lengths which can be assigned to transcript IDs by using the SAM file obtained in the alignment step. Since datasets had different sizes, 100 reads were randomly subsampled 5 times from the total amount of reads of each sample and the median tail length determined. For these 5 sub-datasets of each tail, the mean and standard deviation was calculated. The variability in the poly(A)-tail length of mRNAs was measured with the Fano factor ( $\sigma^2/\mu$ ) for genes that had a number of reads >50. This threshold was set by decreasing the number of GFP reads included in a subsampled dataset (i.e.,  $n=500/50/30/20/10$ ) and randomly subsampling 5 times for each  $n$ . The standard deviation between subsamples increases drastically when the size of the subsampled dataset decreases below 50 reads (Figure S4B).

The theoretical poly(A)-tail length variability of the mixed population was calculated as the Fano factor ( $\sigma^2/\mu$ ) (black cross, Figure S4C). This value was then corrected by adding the technical noise measured for the 100-tail spike-in (red cross, Figure S4C).

## PABPCs binding assay

Reactions were assembled into 5  $\mu$ L of volume and contained a final concentration of 1X Binding Buffer (20 mM Tris-HCl, pH 7.5, 20 mM  $MgCl_2$ , 0.2 M KCl), 0.5  $\mu$ M mRNA, and 2, 4 or 6  $\mu$ M PABPC1 (Bio-Connect). The mixed reactions were incubated for 60 minutes at 37°C and then immediately analyzed by capillary electrophoresis (Bioanalyzer, Agilent RNA 6000 Nano Kit). Data were normalized to the highest (FU) value of each sample and plotted in python to allow sample comparison.

## Poly(A)-tail length assay kit

The poly(A)-tail length assay was performed according to manufacturer's instructions (ThermoFisher), using a two-step PCR amplification. PCR products were detected on a 2.5% agarose TBE gel. Primer sequences are reported in Table S3.



## Ribosome profiling

Cells were supplemented with Cycloheximide 100 µg/ml 5h after transfection, and harvested in 400 µl lysis buffer and lysed by triturating the samples 10 times through a 27G needle. Cell debris was removed by centrifugation at 20000 g at 4°C. 375 units of RNase I (Ambion) were added to the cleared supernatants and tubes were incubated horizontally at room temperature in a rotator for 45 minutes. Ribosomes were pelleted by ultracentrifugation at 173500 g in a Beckman Ti-90 rotor for 3 hours at 4°C on a 1 M sucrose cushion in polysome buffer. RNA was extracted from the ribosomes using Trizol (Ambion) and 5 µg of RNA was loaded onto a 15% PAA gel containing 8 M urea flanked by 28 nt + 33 nt RNA size markers. Bands were stained with SYBR Gold (Invitrogen) and fragments in the 28nt-33nt range were excised from the gel, eluted in sterile TE buffer and recovered by isopropanol precipitation. The RNA was subsequently dephosphorylated using T4 PNK (NEB) and a 5'-preadenylated linker was ligated to the 3'-end of the fragments using T4 RNA ligase 2, truncated K227Q (NEB). Ligated fragments were purified over 15% denaturing PAA gel, recovered and reverse transcribed using a primer containing library and index primer landing sequences that are separated by two internal C18 spacers. RNA was hydrolysed by alkaline treatment and the first strand DNA was purified over 15% denaturing PAA gel and circularized using CircLigase (LGC Biosearch). Circularized products were depleted from sequences originating from ribosomal RNA using biotinylated depletion oligonucleotides (IDT) and MyOne Streptavidin C1 Dynabeads (Invitrogen). Depleted circular DNAs were purified by isopropanol precipitation and indexed with barcodes suitable for Illumina Next Generation Sequencing during 12-14 cycles of PCR amplification with Pfu DNA polymerase (in-house purified). Amplified DNA libraries were purified from 8% native PAA gel, eluted in TE, recovered by isopropanol precipitation and quantified using by Qubit™ using the RNA High Sensitivity kit (Invitrogen™). Sequencing of the libraries was performed on an Illumina NovaSeq6000 by Genomescan B.V., Leiden, The Netherlands. Reads were paired and clipped using the Galaxy webserver ([www.usegalaxy.org](http://www.usegalaxy.org)) and aligned to the GRCh38 genome assembly or GFP sequence with the STAR aligner<sup>115</sup>. The analysis of the ribosome-protected fragments and P-sites was performed with the R Bioconductor package ribosomeProfilingQC<sup>116</sup>, where the P-site of each read was defined as the single position of the 13<sup>th</sup> nt from the 5'end of the read. Nested ORFs were identified with ORF Finder<sup>117</sup>. Refer to Ingolia et al<sup>118</sup> for a detailed protocol and buffers composition.

## Modelling of GFP expression

*Model without transcription.* A model of the chemical master equation (CME) of Figure 1C was constructed to simulate GFP expression from different poly(A)-tails. Chemical reaction schemes were coded in Python and simulated using the Gillespie algorithm<sup>83,84</sup>. The  $k_{tl}$  values used as input for the stochastic simulations correspond to the mean translation rate of each tail. All the parameters used as input for the stochastic models are reported in Table S8 and S9. Initial conditions for all species were set to 0, except for mRNA<sub>10</sub> which was set to 150 for the simulations without extrinsic noise, or randomly picked from a normal distribution centred at 150 for the simulations accounting extrinsic noise. Simulations were run for time = 10 (simulated hours).

*Two-state transcription model.* A simplified two-state transcription model of the CME of Figure 6I was constructed to simulate GFP expression from different poly(A)-tails with the same transcription rate constant ( $k_t$ ). Chemical reaction schemes were coded in Python and simulated using the Gillespie algorithm<sup>83,84</sup>.  $k_m$  and  $k_{d2}$  were increased compared to the simulations without transcription, in order to decrease the computational power. Initial conditions for all species were set to 0, except for Promoter OFF which was set to 1. Simulations were run for time = 20 (simulated hours). All the parameters used as input for the stochastic models are reported in Table S9.

## CV<sup>2</sup>, t<sub>1/2</sub> and Fano $k_{tl}$ calculation for experimental and simulated single-cell trajectories.

For the calculation of the fluctuation amplitude and frequency, experimental single-cell tracks were first normalized to 0 and both experimental and simulated tracks were detrended by linear fitting. The mean fluctuation amplitude was calculated as the variance ( $\sigma^2$ ) from each detrended single-cell

trajectory. The mean fluctuation frequency was calculated as the  $t_{1/2}$  of the autocorrelation using Numpy packages in Python to extract the time lag value at which the autocorrelation is equal to  $0.5^{80-82}$ . To obtain the cell-to-cell variability (Fano  $k_{fl}$ ), first non-linear least square optimization was performed on each individual (experimental and simulated) raw trajectory to extract the  $k_{fl}$  value of each (experimental and simulated) cell. The Fano  $k_{fl}$  was finally measured from the experimental and simulated single-cell  $k_{fl}$  distribution of each tail as the  $\sigma^2/\mu$ .

## REFERENCES

1. Vogel, C., and Marcotte, E.M. (2012). Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nat Rev Genet* 13, 227. 10.1038/NGR3185.
2. Vogel, C., de Sousa Abreu, R., Ko, D., Le, S.Y., Shapiro, B.A., Burns, S.C., Sandhu, D., Boutz, D.R., Marcotte, E.M., and Penalva, L.O. (2010). Sequence signatures and mRNA concentration can explain two-thirds of protein abundance variation in a human cell line. *Mol Syst Biol* 6. 10.1038/MSB.2010.59.
3. Franks, A., Airolidi, E., and Slavov, N. (2017). Post-transcriptional regulation across human tissues. *PLoS Comput Biol* 13, e1005535. 10.1371/JOURNAL.PCBI.1005535.
4. Rolfe, D.F.S., and Brown, G.C. (1997). Cellular energy utilization and molecular origin of standard metabolic rate in mammals. *Physiol Rev* 77, 731–758. 10.1152/PHYSREV.1997.77.3.731.
5. Buttgerit, F., and Brand, M.D. (1995). A hierarchy of ATP-consuming processes in mammalian cells. *Biochem J* 312 ( Pt 1), 163–167. 10.1042/BJ3120163.
6. Schwanhüusser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., Chen, W., and Selbach, M. (2011). Global quantification of mammalian gene expression control. *Nature* 2011 473:7347 473, 337–342. 10.1038/nature10098.
7. Zhao, B.S., Roundtree, I.A., and He, C. (2017). Post-transcriptional gene regulation by mRNA modifications. *Nat Rev Mol Cell Biol* 18, 31. 10.1038/NRM.2016.132.
8. Jurado, A.R., Tan, D., Jiao, X., Kiledjian, M., and Tong, L. (2014). Structure and function of pre-mRNA 5'-end capping quality control and 3'-end processing. *Biochemistry* 53, 1882–1898. 10.1021/BI401715V.
9. Edmonds, M., Vaughan, M.H., and Nakazato, H. (1971). Polyadenylic acid sequences in the heterogeneous nuclear RNA and rapidly-labeled polyribosomal RNA of HeLa cells: possible evidence for a precursor relationship. *Proc Natl Acad Sci U S A* 68, 1336–1340. 10.1073/PNAS.68.6.1336.
10. Edmonds, M., and Abrams, R. (1960). Polynucleotide Biosynthesis: Formation of a Sequence of Adenylate Units from Adenosine Triphosphate by an Enzyme from Thymus Nuclei. *Journal of Biological Chemistry* 235, 1142–1149. 10.1016/S0021-9258(18)69494-3.
11. Lim, L., and Canellakis, E.S. (1970). Adenine-rich polymer associated with rabbit reticulocyte messenger RNA. *Nature* 227, 710–712. 10.1038/227710A0.
12. Eisen, T.J., Eichhorn, S.W., Subtelny, A.O., Lin, K.S., McGeary, S.E., Gupta, S., and Bartel, D.P. (2020). The Dynamics of Cytoplasmic mRNA Metabolism. *Mol Cell* 77, 786-799.e10. 10.1016/j.molcel.2019.12.005.
13. Legnini, I., Alles, J., Karaikos, N., Ayoub, S., and Rajewsky, N. (2019). FLAM-seq: full-length mRNA sequencing reveals principles of poly(A) tail length control. *Nat Methods* 16, 879–886. 10.1038/s41592-019-0503-y.
14. Ibrahim, F., Oppelt, J., Maragkakis, M., and Mourelatos, Z. (2021). TERA-Seq: true end-to-end sequencing of native RNA molecules for transcriptome characterization. *Nucleic Acids Res*, 1–18. 10.1093/nar/gkab713.
15. Yu, F., Zhang, Y., Cheng, C., Wang, W., Zhou, Z., Rang, W., Yu, H., Wei, Y., Wu, Q., and Zhang, Y. (2020). Poly(A)-seq: A method for direct sequencing and analysis of the transcriptomic poly(A)-tails. *PLoS One* 15, 1–21. 10.1371/journal.pone.0234696.
16. Jain, M., Abu-Shumays, R., Olsen, H.E., and Akeson, M. (2022). Advances in nanopore direct RNA sequencing. *Nature Methods* 2022 19:10 19, 1160–1164. 10.1038/s41592-022-01633-w.

17. Fuke, H., and Ohno, M. (2008). Role of poly (A) tail as an identity element for mRNA nuclear export. *Nucleic Acids Res* 36, 1037–1049. 10.1093/NAR/GKM1120.
18. Gallie, D.R. (1991). The cap and poly(A) tail function synergistically to regulate mRNA translational efficiency. *Genes Dev* 5, 2108–2116. 10.1101/GAD.5.11.2108.
19. Wakiyama, M., Imataka, H., and Sonenberg, N. (2000). Interaction of eIF4G with poly(A)-binding protein stimulates translation and is critical for *Xenopus* oocyte maturation. *Current Biology* 10, 1147–1150. 10.1016/S0960-9822(00)00701-6.
- 1010 20. Kahvejian, A., Svitkin, Y. v., Sukarieh, R., M'Boutchou, M.N., and Sonenberg, N. (2005). Mammalian poly(A)-binding protein is a eukaryotic translation initiation factor, which acts via multiple mechanisms. *Genes Dev* 19, 104–113. 10.1101/GAD.1262905.
21. Wells, S.E., Hillner, P.E., Vale, R.D., and Sachs, A.B. (1998). Circularization of mRNA by eukaryotic translation initiation factors. *Mol Cell* 2, 135–140. 10.1016/S1097-2765(00)80122-7.
22. Prieto, S., de La Cruz, B.J., and Scheffler, I.E. (2000). Glucose-regulated Turnover of mRNA and the Influence of Poly(A) Tail Length on Half-life\*. *Journal of Biological Chemistry* 275, 14155–14166. 10.1074/jbc.275.19.14155.
- 1020 23. Schäfer, I.B., Yamashita, M., Schuller, J.M., Schüssler, S., Reichelt, P., Strauss, M., and Conti, E. (2019). Molecular Basis for poly(A) RNP Architecture and Recognition by the Pan2-Pan3 Deadenylation. *Cell* 177, 1619. 10.1016/J.CELL.2019.04.013.
24. Wiederhold, K., and Passmore, L.A. (2010). Cytoplasmic deadenylation: regulation of mRNA fate. *Biochem Soc Trans* 38, 1531–1536. 10.1042/BST0381531.
25. Batey, R.T. (2006). Structures of regulatory elements in mRNAs. Preprint, 10.1016/j.sbi.2006.05.001 10.1016/j.sbi.2006.05.001.
26. Tanguay, R.L., and Gallie, D.R. (1996). Translational efficiency is regulated by the length of the 3' untranslated region. *Mol Cell Biol* 16, 146–156. 10.1128/mcb.16.1.146.
27. Lowell, J.E., Rudner, D.Z., and Sachs, A.B. (1992). 3'-UTR-dependent deadenylation by the yeast poly(A) nuclease. *Genes Dev* 6, 2088–2099. 10.1101/GAD.6.11.2088.
- 1030 28. Webster, M.W., Chen, Y.H., Stowell, J.A.W., Alhusaini, N., Sweet, T., Graveley, B.R., Collier, J., and Passmore, L.A. (2018). mRNA Deadenylation Is Coupled to Translation Rates by the Differential Activities of Ccr4-Not Nucleases. *Mol Cell* 70, 1089–1100.e8. 10.1016/J.MOLCEL.2018.05.033.
29. Presnyak, V., Alhusaini, N., Chen, Y.-H., Martin, S., Morris, N., Kline, N., Olson, S., Weinberg, D., Baker, K.E., Graveley, B.R., et al. (2015). Codon Optimality Is a Major Determinant of mRNA Stability. *Cell* 160, 1111–1124. 10.1016/j.cell.2015.02.029.
30. Wu, Q., Medina, S.G., Kushawah, G., DeVore, M.L., Castellano, L.A., Hand, J.M., Wright, M., and Bazzini, A.A. (2019). Translation affects mRNA stability in a codon-dependent manner in human cells. *Elife* 8. 10.7554/eLife.45396.
- 1040 31. Gruber, A.J., and Zavolan, M. (2019). Alternative cleavage and polyadenylation in health and disease. *Nat Rev Genet* 20, 599–614. 10.1038/S41576-019-0145-Z.
32. Dharmalingam, P., Mahalingam, R., Yalamanchili, H.K., Weng, T., Karmouty-Quintana, H., Guha, A., and A. Thandavarayan, R. (2022). Emerging roles of alternative cleavage and polyadenylation (APA) in human disease. *J Cell Physiol* 237, 149–160. 10.1002/JCP.30549.
33. Scriver, A. (2002). Polyadenylate polymerase (PAP) and 3' end pre-mRNA processing: function, assays, and association with disease. *Crit Rev Clin Lab Sci* 39, 193–224. 10.1080/10408360290795510.
34. Curinha, A., Braz, S.O., Pereira-Castro, I., Cruz, A., and Moreira, A. (2014). Implications of polyadenylation in health and disease. *Nucleus* 5, 508–519. 10.4161/NUCL.36360.
- 1050 35. Tan, E., Chin, C.S.H., Lim, Z.F.S., and Ng, S.K. (2021). HEK293 Cell Line as a Platform to Produce Recombinant Proteins and Viral Vectors. *Front Bioeng Biotechnol* 9, 1288. 10.3389/FBIOE.2021.796991/BIBTEX.
36. Raj, A., van den Bogaard, P., Rifkin, S.A., van Oudenaarden, A., and Tyagi, S. (2008). Imaging individual mRNA molecules using multiple singly labeled probes. *Nat Methods* 5, 877. 10.1038/NMETH.1253.
37. Na Audébert, L., Féuérbach, F., Décourty, L., Namané, A., Pérmal, E., Badis, G.L., and Savéanu, C. Deadenylation rate is not a major determinant of RNA degradation in yeast. 10.1101/2023.01.16.524186.

38. Chan, L.Y., Mugler, C.F., Heinrich, S., Vallotton, P., and Weis, K. (2018). Non-invasive measurement of mRNA decay reveals translation initiation as the major determinant of mRNA stability. *Elife* 7. 10.7554/ELIFE.32536.
39. Lima, S.A., Chipman, L.B., and Nicholson, A.L. (2018). Short Poly(A) Tails are a Conserved Feature of Highly Expressed Genes. *Physiol Behav* 176, 139–148. 10.1038/nsm.3499.Short.
40. Shapiro, R.A., Herrick, D., Manrow, R.E., Blinder, D., and Jacobson, A. (1988). Determinants of mRNA Stability in *Dictyostelium discoideum* Amoebae: Differences in Poly(A) Tail Length, Ribosome Loading, and mRNA Size Cannot Account for the Heterogeneity of mRNA Decay Rates. *Mol Cell Biol* 8, 1957–1969. 10.1128/mcb.8.5.1957-1969.1988.
41. Fuchs, A.-L., Neu, A., and Sprangers, R. (2016). A general method for rapid and cost-efficient large-scale production of 5' capped RNA. *RNA* 22, 1454–1466. 10.1261/rna.056614.116.
42. Cao, D., and Parker, R. (2001). Computational modeling of eukaryotic mRNA turnover. *RNA* 7, 1192. 10.1017/S1355838201010330.
43. Muhrad, D., Decker, C.J., and Parker, R. (1994). Deadenylation of the unstable mRNA encoded by the yeast MFA2 gene leads to decapping followed by 5'→3' digestion of the transcript. *Genes Dev* 8, 855–866. 10.1101/GAD.8.7.855.
44. Krause, M., Niazi, A.M., Labun, K., Torres Cleuren, Y.N., Müller, F.S., and Valen, E. (2019). TailFindR: Alignment-free poly(A) length measurement for Oxford Nanopore RNA and DNA sequencing. *RNA* 25, 1229–1241. 10.1261/RNA.071332.119/-/DC1.
45. Martin, G., and Keller, W. (1998). Tailing and 3'-end labeling of RNA with yeast poly(A) polymerase and various nucleotides. *RNA* 4, 226.
46. Kusov, Y.Y., Shatirishvili, G., Dzagurov, G., and Gauss-Müller, V. (2001). A new G-tailing method for the determination of the poly(A) tail length applied to hepatitis A virus RNA. *Nucleic Acids Res* 29, e57. 10.1093/NAR/29.12.E57.
47. Wiener, D., Antebi, Y., and Schwartz, S. (2021). Decoupling of degradation from deadenylation reshapes poly(A) tail length in yeast meiosis. *Nat Struct Mol Biol* 28, 1038–1049. 10.1038/s41594-021-00694-3.
48. Görlach, M., Burd, C.G., and Dreyfuss, G. (1994). The mRNA Poly(A)-Binding Protein: Localization, Abundance, and RNA-Binding Specificity. *Exp Cell Res* 211, 400–407. 10.1006/EXCR.1994.1104.
49. Xiang, K., and Bartel, D.P. (2021). The molecular basis of coupling between poly(A)-tail length and translational efficiency. *Elife* 10. 10.7554/ELIFE.66493.
50. Gu, S., Jeon, H.M., Nam, S.W., Hong, K.Y., Rahman, M.S., Lee, J.B., Kim, Y., and Jang, S.K. (2022). The flip-flop configuration of the PABP-dimer leads to switching of the translation function. *Nucleic Acids Res* 50, 306–321. 10.1093/NAR/GKAB1205.
51. Bernstein, P., Peltz, S.W., and Ross, J. (1989). The poly(A)-poly(A)-binding protein complex is a major determinant of mRNA stability in vitro. *Mol Cell Biol* 9, 659–670. 10.1128/mcb.9.2.659-670.1989.
52. Yi, H., Park, J., Ha, M., Lim, J., Chang, H., and Kim, V.N. (2018). PABP Cooperates with the CCR4-NOT Complex to Promote mRNA Deadenylation and Block Precocious Decay. *Mol Cell* 70, 1081–1088.e5. 10.1016/j.molcel.2018.05.009.
53. Xie, J., Wei, X., and Chen, Y. (2021). Loss of PABPC1 is compensated by elevated PABPC4 and correlates with transcriptome changes. *bioRxiv*, 2021.02.07.430165. 10.1101/2021.02.07.430165.
54. Baer, B.W., and Kornberg, R.D. (1983). The Protein Responsible for the Repeating Cytoplasmic Poly(A)-Ribonucleoprotein Structure of. *J Cell Biol* 96, 717–721.
55. Sachs, A.B., Davis, R.W., and Kornberg, R.D. (1987). A single domain of yeast poly(A)-binding protein is necessary and sufficient for RNA binding and cell viability. *Mol Cell Biol* 7, 3268. 10.1128/MCB.7.9.3268.
56. Siddiqui, N., Mangus, D.A., Chang, T.C., Palermino, J.M., Shyu, A. bin, and Gehring, K. (2007). Poly(A) nuclease interacts with the C-terminal domain of polyadenylate-binding protein domain from poly(A)-binding protein. *J Biol Chem* 282, 25067–25075. 10.1074/JBC.M701256200.
57. Raisch, T., Bhandari, D., Sabath, K., Helms, S., Valkov, E., Weichenrieder, O., and Izaurralde, E. (2016). Distinct modes of recruitment of the CCR4-NOT complex by *Drosophila* and vertebrate Nanos. *EMBO J* 35, 974–990. 10.15252/embj.201593634.



58. Lebreton, A., and Seraphin, B. (2008). Exosome-mediated quality control: Substrate recruitment and molecular activity. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* 1779, 558–565. 10.1016/j.bbagr.2008.02.003.
59. Houseley, J., and Tollervey, D. (2009). The Many Pathways of RNA Degradation. *Cell* 136, 763–776. 10.1016/j.cell.2009.01.019.
- 1120 60. Decker, C.J., and Parker, R. (2002). mRNA decay enzymes: Decappers conserved between yeast and mammals. *Proceedings of the National Academy of Sciences* 99, 12512–12514. 10.1073/pnas.212518099.
61. Nagarajan, V.K., Jones, C.I., Newbury, S.F., and Green, P.J. (2013). XRN 5'→3' exoribonucleases: Structure, mechanisms and functions. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* 1829, 590–603. 10.1016/j.bbagr.2013.03.005.
62. Mukherjee, D. (2002). The mammalian exosome mediates the efficient degradation of mRNAs that contain AU-rich elements. *EMBO J* 21, 165–174. 10.1093/emboj/21.1.165.
63. Zhang, G., Luo, H., Li, X., Hu, Z., and Wang, Q. (2023). The Dynamic Poly(A) Tail Acts as a Signal Hub in mRNA Metabolism. *Cells* 12, 572. 10.3390/cells12040572.
- 1130 64. Jolles B., and Jean-Jean O. (2021). Poly(A) tail degradation in human cells: ATF4 mRNA as a model for biphasic deadenylation. *Biochimie*, 128–134. 10.1016/j.biochi.2021.03.013.
65. Wolf, J., Valkov, E., Allen, M.D., Meineke, B., Gordiyenko, Y., McLaughlin, S.H., Olsen, T.M., Robinson, C. V, Bycroft, M., Stewart, M., et al. (2014). Structural basis for Pan3 binding to Pan2 and its function in mRNA recruitment and deadenylation. *EMBO J* 33, 1514–1526. 10.15252/emboj.201488373.
66. Bresson, S., and Tollervey, D. (2018). Tailing Off: PABP and CNOT Generate Cycles of mRNA Deadenylation. *Mol Cell* 70, 987–988. 10.1016/J.MOLCEL.2018.06.009.
67. Brar, G.A., and Weissman, J.S. (2015). Ribosome profiling reveals the what, when, where and how of protein synthesis. *Nat Rev Mol Cell Biol* 16, 651–664. 10.1038/nrm4069.
- 1140 68. Khong, A., and Parker, R. (2018). mRNP architecture in translating and stress conditions reveals an ordered pathway of mRNP compaction. *Journal of Cell Biology* 217, 4124–4140. 10.1083/jcb.201806183.
69. Adivarahan, S., Livingston, N., Nicholson, B., Rahman, S., Wu, B., Rissland, O.S., and Zenklusen, D. (2018). Spatial Organization of Single mRNPs at Different Stages of the Gene Expression Pathway. *Mol Cell* 72, 727–738.e5. 10.1016/j.molcel.2018.10.010.
70. Dave, P., Roth, G., Griesbach, E., Mateju, D., Hochstoeger, T., and Chao, J.A. (2023). Single-molecule imaging reveals translation-dependent destabilization of mRNAs. *Mol Cell* 83, 589–606.e6. 10.1016/j.molcel.2023.01.013.
- 1150 71. Kozak, M. (1987). An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Res* 15, 8125–8148. 10.1093/NAR/15.20.8125.
72. Blanco, N., Williams, A.J., Tang, D., Zhan, D., Misaghi, S., Kelley, R.F., and Simmons, L.C. (2020). Tailoring translational strength using Kozak sequence variants improves bispecific antibody assembly and reduces product-related impurities in CHO cells. *Biotechnol Bioeng* 117, 1946–1960. 10.1002/BIT.27347.
73. Volkova, O.A., and Kochetov, A. v. (2012). Interrelations between the Nucleotide Context of Human Start AUG Codon, N-end Amino Acids of the Encoded Protein and Initiation of Translation. <http://dx.doi.org/10.1080/07391102.2010.10508575> 27, 611–618. 10.1080/07391102.2010.10508575.
- 1160 74. Li, J., Liang, Q., Song, W., and Marchisio, M.A. (2017). Nucleotides upstream of the Kozak sequence strongly influence gene expression in the yeast *S. cerevisiae*. *J Biol Eng* 11, 25. 10.1186/s13036-017-0068-1.
75. Benitez-Cantos, M.S., Yordanova, M.M., O'Connor, P.B.F., Zhdanov, A. V., Kovalchuk, S.I., Papkovsky, D.B., Andreev, D.E., and Baranov, P. V. (2020). Translation initiation downstream from annotated start codons in human mRNAs coevolves with the Kozak context. *Genome Res* 30, 974–984. 10.1101/gr.257352.119.
76. Asrani, K.H., Farelli, J.D., Stahley, M.R., Miller, R.L., Cheng, C.J., Subramanian, R.R., and Brown, J.M. (2018). Optimization of mRNA untranslated regions for improved expression of therapeutic mRNA. <https://doi.org/10.1080/15476286.2018.1450054> 15, 756–762. 10.1080/15476286.2018.1450054.

- 1170 77. Eisen, T.J., Li, J.J., and Bartel, D.P. (2022). The interplay between translational efficiency, poly(A) tails, microRNAs, and neuronal activation. *RNA* 28, 808–831. 10.1261/RNA.079046.121/-/DC1.
78. Munsky, B., Neuert, G., and van Oudenaarden, A. (2012). Using Gene Expression Noise to Understand Gene Regulation. *Science* (1979) 336, 183–187. 10.1126/science.1216379.
79. Hansen, M.M.K., Desai, R. V., Simpson, M.L., and Weinberger, L.S. (2018). Cytoplasmic Amplification of Transcriptional Noise Generates Substantial Cell-to-Cell Variability. *Cell Syst* 7, 384–397.e6. 10.1016/j.cels.2018.08.002.
80. Dar, R.D., Razooky, B.S., Singh, A., Trimeloni, T. v., McCollum, J.M., Cox, C.D., Simpson, M.L., and Weinberger, L.S. (2012). Transcriptional burst frequency and burst size are equally modulated across the human genome. *Proc Natl Acad Sci U S A* 109, 17454–17459. 10.1073/pnas.1213530109.
- 1180 81. Weinberger, L.S., Dar, R.D., and Simpson, M.L. (2008). Transient-mediated fate determination in a transcriptional circuit of HIV. *Nat Genet* 40, 466–470. 10.1038/NG.116.
82. Austin, D.W., Allen, M.S., McCollum, J.M., Dar, R.D., Wilgus, J.R., Sayler, G.S., Samatova, N.F., Cox, C.D., and Simpson, M.L. (2006). Gene network shaping of inherent noise spectra. *Nature* 439, 608–611. 10.1038/NATURE04194.
83. Gillespie, D.T. (2007). Stochastic simulation of chemical kinetics. *Annu Rev Phys Chem* 58, 35–55. 10.1146/ANNUREV.PHYSICHEM.58.032806.104637.
84. Gillespie, D.T. (1977). Exact stochastic simulation of coupled chemical reactions. *Journal of Physical Chemistry* 81, 2340–2361. 10.1021/j100540a008.
- 1190 85. Alon, U. (2007). Network motifs: theory and experimental approaches. *Nat Rev Genet* 8, 450–461. 10.1038/nrg2102.
86. Passmore, L.A., and Collier, J. (2021). Roles of mRNA poly(A) tails in regulation of eukaryotic gene expression. *Nat Rev Mol Cell Biol* 0123456789. 10.1038/s41580-021-00417-y.
87. Nicholson, A.L., and Pasquinelli, A.E. (2019). Tales of Detailed Poly(A) Tails. *Trends Cell Biol.* 176, 139–148. 10.1016/j.physbeh.2017.03.040.
88. Eckmann, C.R., Rammelt, C., and Wahle, E. (2011). Control of poly(A) tail length. *Wiley Interdiscip Rev RNA* 2, 348–361. 10.1002/wrna.56.
- 1200 89. Geissler, R., and Grimson, A. (2016). A position-specific 3'UTR sequence that accelerates mRNA decay. <http://dx.doi.org/10.1080/15476286.2016.1225645> 13, 1075–1077. 10.1080/15476286.2016.1225645.
90. Subtelny, A.O., Eichhorn, S.W., Chen, G.R., Sive, H., and Bartel, D.P. (2014). Poly(A)-tail profiling reveals an embryonic switch in translational control. *Nature* 508, 66–71. 10.1038/NATURE13007.
91. Desai, R. v., Chen, X., Martin, B., Chaturvedi, S., Hwang, D.W., Li, W., Yu, C., Ding, S., Thomson, M., Singer, R.H., et al. (2021). A DNA repair pathway can regulate transcriptional noise to promote cell fate transitions. *Science* 373. 10.1126/SCIENCE.ABC6506.
92. Hansen, M.M.K., Wen, W.Y., Ingerman, E., Razooky, B.S., Thompson, C.E., Dar, R.D., Chin, C.W., Simpson, M.L., and Weinberger, L.S. (2018). A Post-Transcriptional Feedback Mechanism for Noise Suppression and Fate Stabilization. *Cell* 173, 1609–1621.e15. 10.1016/J.CELL.2018.04.005.
- 1210 93. Li, Y., Shan, Y., Desai, R. v., Cox, K.H., Weinberger, L.S., and Takahashi, J.S. (2020). Noise-driven cellular heterogeneity in circadian periodicity. *Proc Natl Acad Sci U S A* 117, 10350–10356. 10.1073/PNAS.1922388117.
94. Shaffer, S.M., Dunagin, M.C., Torborg, S.R., Torre, E.A., Emert, B., Krepler, C., Beqiri, M., Sproesser, K., Brafford, P.A., Xiao, M., et al. (2017). Rare cell variability and drug-induced reprogramming as a mode of cancer drug resistance. *Nature* 546, 431–435. 10.1038/NATURE22794.
- 1220 95. Hansen, M.M.K., and Weinberger, L.S. (2019). Post-Transcriptional Noise Control. *BioEssays* 41, e1900044. 10.1002/bies.201900044.
96. Arzt, M., Deschamps, J., Schmied, C., Pietzsch, T., Schmidt, D., Tomancak, P., Haase, R., and Jug, F. (2022). LABKIT: Labeling and Segmentation Toolkit for Big Image Data. *Front Comput Sci* 4, 10. 10.3389/FCOMP.2022.777728/BIBTEX.



97. Tinevez, J.Y., Perry, N., Schindelin, J., Hoopes, G.M., Reynolds, G.D., Laplantine, E., Bednarek, S.Y., Shorte, S.L., and Eliceiri, K.W. (2017). TrackMate: An open and extensible platform for single-particle tracking. *Methods* 115, 80–90. 10.1016/J.YMETH.2016.09.016.
98. Ershov, D., Phan, M.-S., Pylvänäinen, J.W., Rigaud, S.U., Blanc, L. Le, Charles-Orszag, A., Conway, J.R.W., Laine, R.F., Roy, N.H., Bonazzi, D., et al. (2021). Bringing TrackMate in the era of machine-learning and deep-learning. *bioRxiv*, 2021.09.03.458852. 10.1101/2021.09.03.458852.
99. Sage, D., Kirshner, H., Pengo, T., Stuurman, N., Min, J., Manley, S., and Unser, M. (2015). Quantitative evaluation of software packages for single-molecule localization microscopy. *Nat Methods* 12, 717–724. 10.1038/nmeth.3442.
100. Kirshner, H., Aguet, F., Sage, D., and Unser, M. (2013). 3-D PSF fitting for fluorescence microscopy: implementation and localization application. *J Microsc* 249, 13–25. 10.1111/j.1365-2818.2012.03675.x.
101. Sage, D., Donati, L., Soulez, F., Fortun, D., Schmit, G., Seitz, A., Guiet, R., Vonesch, C., and Unser, M. (2017). DeconvolutionLab2: An open-source software for deconvolution microscopy. *Methods* 115, 28–41. 10.1016/j.ymeth.2016.12.015.
102. Stringer, C., Wang, T., Michaelos, M., and Pachitariu, M. (2020). Cellpose: a generalist algorithm for cellular segmentation. *Nature Methods* 2020 18:1 18, 100–106. 10.1038/s41592-020-01018-x.
103. Mueller, F., Senecal, A., Tantale, K., Marie-Nelly, H., Ly, N., Collin, O., Basyuk, E., Bertrand, E., Darzacq, X., and Zimmer, C. (2013). FISH-quant: automatic counting of transcripts in 3D FISH images. *Nature Methods* 2013 10:4 10, 277–278. 10.1038/nmeth.2406.
104. Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34, 3094–3100. 10.1093/BIOINFORMATICS/BTY191.
105. Ouyang, W., Mueller, F., Hjelmare, M., Lundberg, E., and Zimmer, C. (2019). ImJoy: an open-source computational platform for the deep learning era. *Nature Methods* 2019 16:12 16, 1199–1200. 10.1038/s41592-019-0627-0.
106. Matsuda, T., and Cepko, C.L. (2004). Electroporation and RNA interference in the rodent retina in vivo and in vitro. *Proc Natl Acad Sci U S A* 101, 16–22. 10.1073/PNAS.2235688100.
107. Eden, E., Navon, R., Steinfeld, I., Lipson, D., and Yakhini, Z. (2009). GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics* 10, 48. 10.1186/1471-2105-10-48.
108. Steinfeld, I., Navon, R., Ach, R., and Yakhini, Z. (2013). miRNA target enrichment analysis reveals directly active miRNAs in health and disease. *Nucleic Acids Res* 41, e45–e45. 10.1093/nar/gks1142.
109. Electromagnetic diffraction in optical systems, II. Structure of the image field in an aplanatic system (1959). *Proc R Soc Lond A Math Phys Sci* 253, 358–379. 10.1098/rspa.1959.0200.
110. Lucy, L.B. (1974). An iterative technique for the rectification of observed distributions. *Astron J* 79, 745. 10.1086/111605.
111. Richardson, W.H. (1972). Bayesian-Based Iterative Method of Image Restoration\*. *J Opt Soc Am* 62, 55. 10.1364/JOSA.62.000055.
112. Chan, C.L., Majzoub, R.N., Shirazi, R.S., Ewert, K.K., Chen, Y.J., Liang, K.S., and Safinya, C.R. (2012). Endosomal escape and transfection efficiency of PEGylated cationic liposome-DNA complexes prepared with an acid-labile PEG-lipid. *Biomaterials* 33, 4928–4935. 10.1016/J.BIOMATERIALS.2012.03.038.
113. Leonhardt, C., Schwake, G., Stögbauer, T.R., Rappl, S., Kuhr, J.T., Ligon, T.S., and Rädler, J.O. (2014). Single-cell mRNA transfection studies: delivery, kinetics and statistics by numbers. *Nanomedicine* 10, 679–688. 10.1016/J.NANO.2013.11.008.
114. Niazi, A.M., Krause, M., and Valen, E. (2021). Transcript Isoform-Specific Estimation of Poly(A) Tail Length by Nanopore Sequencing of Native RNA. *Methods in Molecular Biology* 2284, 543–567. 10.1007/978-1-0716-1307-8\_30.
115. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21. 10.1093/bioinformatics/bts635.
116. Ou, J., and Hoye, M. (2023). ribosomeProfilingQC: Ribosome Profiling Quality Control. 10.18129/B9.bioc.ribosomeProfilingQC.

117. Stothard, P. (2000). The Sequence Manipulation Suite: JavaScript Programs for Analyzing and Formatting Protein and DNA Sequences. *Biotechniques* 28, 1102–1104. 10.2144/00286ir01.
118. Ingolia, N.T., Brar, G.A., Rouskin, S., McGeachy, A.M., and Weissman, J.S. (2012). The ribosome profiling strategy for monitoring translation in vivo by deep sequencing of ribosome-protected mRNA fragments. *Nat Protoc* 7, 1534–1550. 10.1038/nprot.2012.086.