ARTICLE

doi:10.1038/nature13007

Poly(A)-tail profiling reveals an embryonic switch in translational control

Alexander O. Subtelny[1,2,3,4]*, Stephen W. Eichhorn[1,2,3]*, Grace R. Chen[1,2,3], Hazel Sive[2,3] & David P. Bartel[1,2,3]

Poly(A) tails enhance the stability and translation of most eukaryotic messenger RNAs, but difficulties in globally measuring poly(A)-tail lengths have impeded greater understanding of poly(A)-tail function. Here we describe poly(A)-tail length profiling by sequencing (PAL-seq) and apply it to measure tail lengths of millions of individual RNAs isolated from yeasts, cell lines, *Arabidopsis thaliana* leaves, mouse liver, and zebrafish and frog embryos. Poly(A)-tail lengths were conserved between orthologous mRNAs, with mRNAs encoding ribosomal proteins and other 'housekeeping' proteins tending to have shorter tails. As expected, tail lengths were coupled to translational efficiencies in early zebrafish and frog embryos. However, this strong coupling diminished at gastrulation and was absent in non-embryonic samples, indicating a rapid developmental switch in the nature of translational control. This switch complements an earlier switch to zygotic transcriptional control and explains why the predominant effect of microRNA-mediated deadenylation concurrently shifts from translational repression to mRNA destabilization.

Most eukaryotic mRNAs end with poly(A) tails, which are added by a nuclear poly(A) polymerase following cleavage of the primary transcript during transcriptional termination[1]. These tails are then shortened by deadenylases[2,3], although in some contexts (for example, animal oocytes or early embryos, or at neuronal synapses), they can be re-extended by cytoplasmic poly(A) polymerases[4,5]. In the cytoplasm, the poly(A) tail promotes translation and inhibits decay[2,5].

Although poly(A) tails must exceed a minimal length to promote translation, an influence of tail length beyond this minimum is largely unknown. The prevailing view is that longer tails generally lead to increased translation[5,6]. This idea partly stems from the known importance of cytoplasmic polyadenylation in activating certain genes in specific contexts[4,5], and the increased translation observed in *Xenopus* oocytes and *Drosophila* embryos when appending synthetic tails of increasing length onto an mRNA[7,8]. Support for a more general coupling of tail length and translation comes from studies of yeast extracts[9] and yeast cells[10,11]. However, the general relationship between tail length and translational efficiency has not been reported outside of yeast, primarily because transcriptome-wide measurements have been unfeasible for longer-tailed mRNAs.

Poly(A)-tail length profiling by sequencing (PAL-seq)

We developed a high-throughput sequencing method that accurately measures individual poly(A) tails of any physiological length (Fig. 1a). After generating sequencing clusters and before sequencing, a primer hybridized immediately 3′ of the poly(A) sequence is extended using a mixture of dTTP and biotin-conjugated dUTP as the only nucleoside triphosphates and conditions that were optimized to yield full-length extension products without terminal mismatches (Extended Data Fig. 1a). This key step quantitatively marks each cluster with biotin in proportion to the length of the poly(A) tail (Fig. 1a). After sequencing the 36 nucleotides immediately 5′ of the poly(A) site, the flow cell is incubated with fluorophore-tagged streptavidin, which binds the biotin incorporated during primer extension to impart fluorescence intensity proportional to the poly(A)-tract length. To account for the density of each cluster, this raw intensity is normalized to that of the fluorescent bases added during sequencing by synthesis[12], thereby yielding a normalized fluorescence intensity for the poly(A) tail of each transcript, paired with a sequencing read that identifies its poly(A) site and thus the gene of origin.

Each starting sample was spiked with a cocktail of mRNA-like standards of known tail lengths (Extended Data Fig. 1b) to produce a standard curve for converting normalized fluorescence intensities to poly(A)-tail lengths (Fig. 1b). We refer to each of these tail-length measurements paired with its identifying sequence as a poly(A) tag. Although recovery of tags from the standards varied somewhat, it did not vary systematically with tail length, which indicated that length-related biases were not an issue (Extended Data Fig. 1c). Additional analyses indicated that mRNA degradation did not bias against longer poly(A) tails (Extended Data Fig. 2a).

Because alternative start sites or alternative splicing can generate different transcripts with the same poly(A) site, we considered our results with respect to unique gene models (abbreviated as 'genes') rather than to transcripts (even though polyadenylation occurs on transcripts, not genes). Moreover, tags for alternative poly(A) sites of the same gene were pooled, unless stated otherwise. With this pipeline, analysis of RNA from NIH3T3 mouse fibroblasts (3T3 cells) yielded at least one tag from 10,094 unique protein-coding genes (including 97% of the 9,976 genes with at least one mRNA molecule per cell, as determined by RNA-seq) and ≥ 100 tags from 2,873 genes, coverage typical of most samples (Supplementary Table 1).

Tail-length diversity within each species

Median tail lengths in mammalian cells (range, 67–96 nucleotides) exceeded those in *A. thaliana* leaves and *Drosophila melanogaster* S2 cells (51 and 50 nucleotides, respectively), which exceeded those in budding (*Saccharomyces cerevisiae*) and fission (*Schizosaccharomyces pombe*) yeast (27 and 28 nucleotides, respectively) (Fig. 2a). Similar differences between mammalian, fly, plant and yeast cells were observed when comparing tail-length averages for individual genes (Fig. 2b). For genes within each species, mean tail lengths varied, with the 10th and 90th

[1]Howard Hughes Medical Institute, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. [2]Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, Massachusetts 02142, USA. [3]Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. [4]Harvard-MIT Division of Health Sciences and Technology, Cambridge, Massachusetts 02139, USA.
*These authors contributed equally to this work.

©2014 Macmillan Publishers Limited. All rights reserved

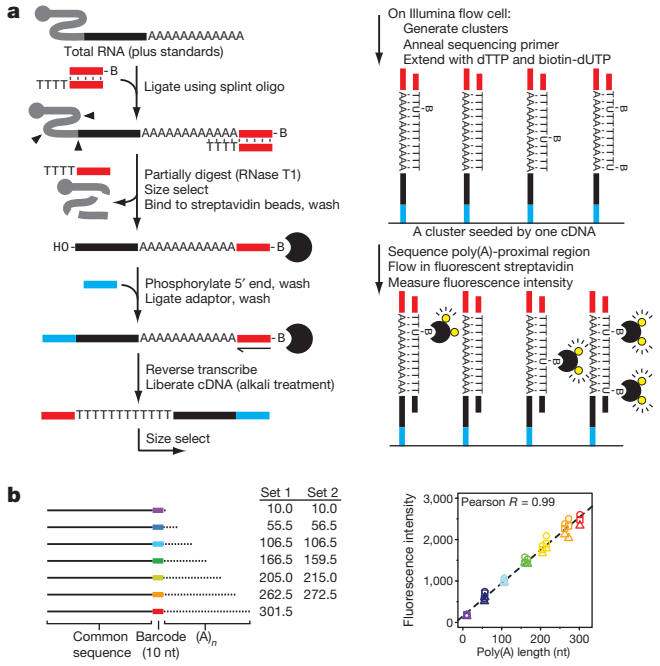


Figure 1 | Global measurement of poly(A)-tail lengths. **a**, Outline of PAL-seq. For each cluster, the fluorescence intensity reflects the tail length of the cDNA that seeded the cluster. Although the probability of incorporating a biotin-conjugated dU opposite each tail nucleotide is uniform, stochastic incorporation results in a variable number of biotins for each molecule within a cluster. **b**, Median streptavidin fluorescence intensities for two sets of mRNA-like molecules with indicated poly(A)-tail lengths, which were added to 3T3 (circle), HEK293T (triangle) and HeLa (square) samples for tail-length calibration.

percentiles differing by 1.4- to 1.6-fold. Variation was also observed for different mRNA transcripts from the same gene (Fig. 2c). For most genes the distributions were unimodal, with the mode approaching the mean (Fig. 2d). Poly(A)-tail lengths increased when progressing through the cleavage, blastula and gastrula stages of zebrafish embryonic development (2, 4 and 6 h post-fertilization (hpf), respectively) and the analogous stages of frog development (Fig. 2a, b, d). Processed data reporting tail lengths for all genes detected in each sample are provided in the Gene Expression Omnibus (accession number GSE52809).

Comparison of tail lengths for orthologous genes in human (HeLa and HEK293T) and mouse (3T3 and liver) cells revealed moderately strong correlations, indicating that tail lengths are conserved (Extended Data Table 1, Spearman R (R_s) as high as 0.46). When searching for gene classes that tended to have longer or shorter tails, the most striking and pervasive enrichment was for ribosomal protein and other 'housekeeping' genes among the short-tailed genes (Extended Data Table 2). This enrichment was strong in yeast, despite previous reports that ribosomal-protein genes tend to have long tails[10,11]. To address this and other discrepancies with previous yeast studies (Extended Data Fig. 3a, b), we used an independent method to measure the poly(A)-tail lengths of eight yeast genes, including four ribosomal protein genes. The results were much more consistent with our measurements than with the previous measurements (Extended Data Figs 3 and 4). Both previous reports used the polyadenylation state microarray (PASTA) method, which fractionates RNAs by stepwise thermal elution from poly(U)-Sepharose. Although studies have successfully used poly(U)-Sepharose fractionation to detect tail-length changes for the same genes in different contexts[13–15], detecting differences between different genes in the same context is more challenging. Our results suggest that PASTA, as previously implemented in yeasts[10,11], is less suitable than PAL-seq for intergenic comparisons, although we cannot exclude the possibility that the discrepancies arose from different growth conditions.

The types of genes with shorter or longer tails differed between the embryonic samples and the other samples (Extended Data Table 2). Genes in the early embryo might not have the same tail lengths as their orthologues do in other contexts because before the maternal-to-zygotic transition (MZT), which occurs at ~3 hpf in zebrafish[16] and at approximately stage 8 in *Xenopus laevis*[17], transcription is not yet active, and some maternal transcripts are masked for later use, whereas others are subject to cytoplasmic polyadenylation[5]. At 6 hpf in zebrafish, ribosomal protein mRNAs had switched from being enriched in shorter-tailed genes

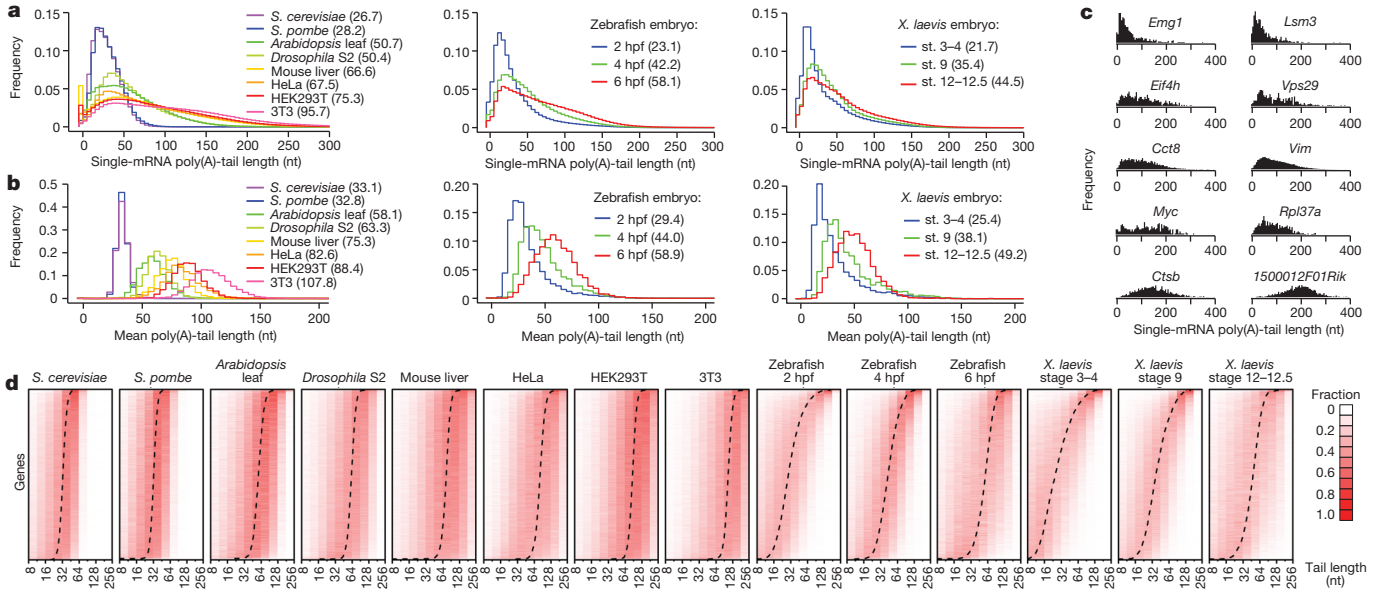


Figure 2 | Poly(A)-tail lengths in yeast, plant, fly and vertebrate cells. **a**, Bulk tail-length distributions. For each sample, histograms tally tail-length measurements for all poly(A) tags mapping to annotated 3′ UTRs (bin size = 5 nucleotides). Leftmost bin includes all measurements <0 nucleotides. Median tail lengths are in parentheses. **b**, Intergenic tail-length distributions. For each sample, histograms tally average tail lengths for protein-coding genes with ≥ 50 tags (yeasts, zebrafish and *Xenopus*) or ≥ 100 tags (other samples). Median average tail lengths are in parentheses. **c**, Intragenic tail-length distributions for 10 genes sampling the spectrum of average tail lengths in 3T3 cells. **d**, Intragenic tail-length distributions. Heat maps show the frequency distribution of tail lengths for each gene tallied in **b**. The colour intensity indicates the fraction of the total for the gene. Genes are ordered by average tail length (dashed line). Results from the *S. cerevisiae* total-RNA sample are reported in this figure.

©2014 Macmillan Publishers Limited. All rights reserved

to being enriched in longer-tailed genes (Extended Data Table 2), perhaps because these were mostly newly synthesized transcripts, which tended to have longer tails at this stage (Extended Data Fig. 5).

Because deadenylation is an important early step in eukaryotic mRNA decay[2,3,18], we examined the relationship between poly(A)-tail length and published mRNA stability values (Extended Data Table 1). Tail length and half-life were slightly negatively correlated in HeLa and 3T3 cells ($R_s = -0.048$ and -0.16, respectively) and variably correlated in yeast, depending on the source of the half-life measurements (R_s from -0.44 to 0.23). The weak relationships in HeLa and 3T3 cells would be expected if mRNAs with different half-lives have similar steady-state tail-length distributions, with the less stable mRNAs transiting through the distributions more quickly.

No strong, easily interpretable correlations between tail length and mRNA features (length of 3′ untranslated region (3′ UTR), length of open reading frame (ORF), total length, splice-site number, splice-site density) or expression (steady-state accumulation and nuclear-to-cytoplasmic ratio) were observed (Extended Data Table 1). Of these, the strongest correlations were between tail length and steady-state accumulation (R_s from -0.44 to 0.25), and between tail length and mRNA length (R_s from –0.12 to 0.36) or features related to mRNA length. Support for the latter relationship was also observed in intragenic comparisons, which revealed a weak positive relationship between tail length and the length of tandem 3′-UTR isoforms (Extended Data Fig. 6a). In early zebrafish embryos this relationship between 3′-UTR isoforms was even more pronounced when a predicted cytoplasmic polyadenylation element (CPE)[4,19] was present in the region unique to the longer isoform (Extended Data Fig. 6b).

Decoupling of tail length and translation

Most reports of increased translation of longer-tailed mRNAs have used oocytes and early embryos[4,5]. To examine whether this phenomenon reported in early embryos for a few genes applies transcriptome-wide, we performed ribosome footprint profiling and RNA-seq to measure translational efficiencies[20] from the embryonic samples used to measure tail lengths. We found that in early embryos (cleavage and blastula stages) of both fish and frog, mean poly(A)-tail length correlated strongly with translational efficiency (Fig. 3a, R_s from 0.62 to 0.77). No other mRNA feature has been reported to correlate so well with translational efficiency in any system.

In these early embryonic stages, a twofold increase in tail length corresponded to a large increase in translational efficiency—greater than 6-fold when doubling the tail from 20 to 40 nucleotides in 2 hpf zebrafish (Fig. 3a). Although longer-tailed mRNAs were more likely to contain a CPE, the relationship between tail length and translational efficiency for CPE-containing mRNAs was no different from that of other mRNAs (Extended Data Fig. 7a). In theory, this coupling might not be causal, or it might be causal but strictly due to either translational inhibition causing tail shortening or translational activity preventing tail shortening. Alternatively, all or at least some of the coupling might result from longer tail length causing more efficient translation in the early embryo. We favour this last possibility because it agrees with the known importance of cytoplasmic polyadenylation for activating genes in maturing oocytes[8,21,22] and early embryos[23,24] of *Xenopus* and in certain other vertebrate contexts[25–30]. Even more importantly, it agrees with the increased translation observed in *Xenopus* oocytes when appending prosthetic poly(A) tails of increasing length onto an mRNA[8].

The strong coupling observed in the blastula largely disappeared in gastrulating embryos (Fig. 3a; $R_s = 0.13$ for both fish and frog). This disappearance was not because of the more restricted tail-length range observed at gastrulation (Extended Data Fig. 7b). Moreover, we observed no positive correlation of a meaningful magnitude between mean poly(A)-tail length and translational efficiency in HeLa cells, HEK293T cells, 3T3 cells, mouse liver, *S. cerevisiae* or *S. pombe* ($R_s = -0.10, 0.07, -0.04, 0.00, -0.12$ and -0.15, respectively) (Fig. 3b). Our results in yeasts differed from those reported earlier[10,11], which we again attribute to the limitations of previous methods. In 3T3 cells, metabolic labelling has been used to infer protein-synthesis rates[31], which correlated with our translational efficiencies ($R_s = 0.44$, $P < 10^{-158}$) and did not correlate positively with tail lengths ($R_s = -0.20$, $P < 10^{-16}$). Taken together, our results indicate that beginning at gastrulation, translational control undergoes a mechanistic change that uncouples translational efficiency from poly(A)-tail length.

Intragenic comparison of tail length and translation

The simplest interpretation of the weak or negative correlations we observed between tail length and translational efficiency in yeast and mammalian cells is that increasing average tail length over the physiological range does not enhance translation in these contexts. However, our comparisons of average tail length and average translational efficiency

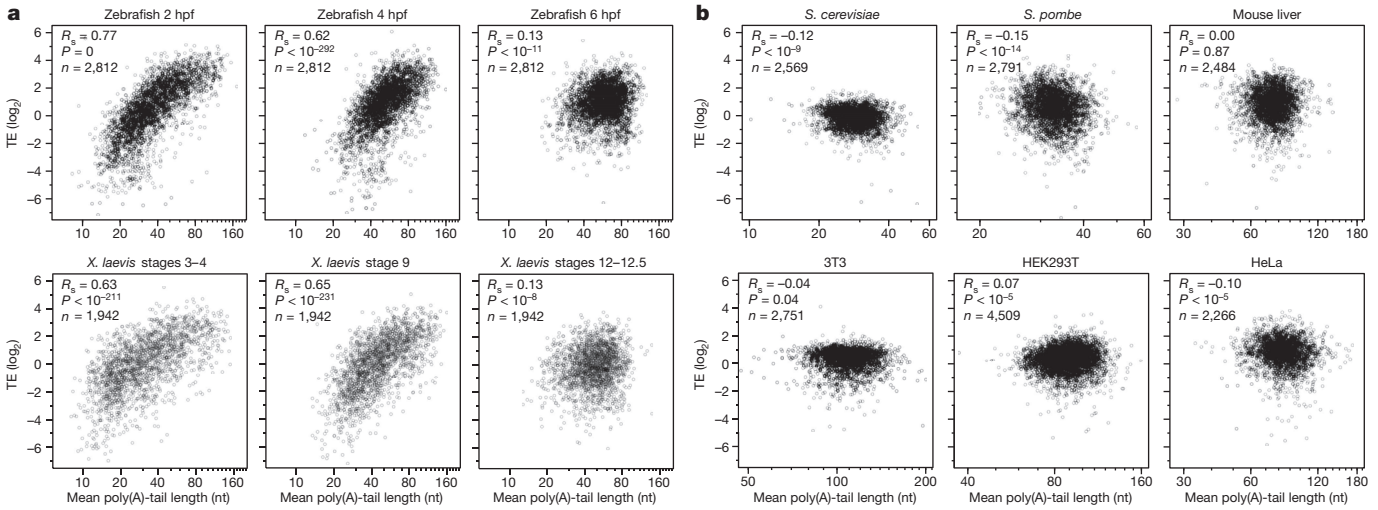


Figure 3 | Transient coupling between poly(A)-tail length and translational efficiency. **a**, Relationship between mean tail length and translational efficiency (TE) for genes with ≥ 50 poly(A) tags from embryonic samples at the indicated developmental stages. For each stage, tail lengths and translational efficiencies were obtained from the same sample. *MGC116473* and *DDX24* fell outside the plot for *X. laevis*, stages 3–4, and *LOC100049092* fell outside the plot for *X. laevis*, stages 12–12.5. **b**, Relationship between mean tail length and translational efficiency in the indicated cells, for genes with ≥ 50 (yeasts) or ≥ 100 (others) tags. With the exception of HeLa[35], tail lengths and translational efficiencies were from the same samples. *S. cerevisiae YBR196C*, *YLR355C* and *YDL080C*, *S. pombe SPCC63.04.1*, mouse liver *NM_007881* and *NM_145470*, HEK293T *NM_001007026*, *NM_021058* and *NM_003537*, and HeLa *NM_001007026* fell outside their respective plots.

©2014 Macmillan Publishers Limited. All rights reserved

between genes (Fig. 3b) might have missed a relationship that would be observed when looking at differentially translated mRNAs from the same gene. To address this possibility, we fractionated 3T3 cell lysate to isolate mRNAs associated with different numbers of ribosomes and measured the tail lengths in each fraction (Fig. 4a). To learn how poly(A)-tail length related to ribosome density for individual genes, we plotted mean tail-length values as a function of the number of bound ribosomes and fit the data for each gene with a straight line (Fig. 4b). The slopes of these lines were generally small, and most were slightly negative (Fig. 4b); positive slopes would have been expected if longer tails enhanced translation. Thus, the increase in median length observed between the lightest and heaviest fractions when considering bulk tail lengths (Fig. 4a; 66 and 82 nucleotides, respectively) did not indicate a relationship between longer tails and enhanced translation but instead might have reflected the positive correlation between ORF length and tail length observed in 3T3 cells (Extended Data Table 1; $R_s = 0.36$). The trend of mostly negative slopes prevailed even when excluding data from mRNA not associated with any ribosomes (Extended Data Fig. 7c), or when examining subsets of genes with longer or shorter mean tail lengths, or with higher or lower translational efficiencies (Extended Data Fig. 7d). This global intragenic analysis (Fig. 4b) supports the conclusion drawn from intergenic analyses (Fig. 3), that in all yeast and mammalian contexts examined (and presumably in most other cellular contexts), mRNAs with longer poly(A) tails are not more efficiently translated.

A shift in the ultimate effects of miRNAs

MicroRNAs (miRNAs) are small RNAs that pair to sites in mRNAs to target these messages for post-transcriptional repression[32]. Global measurements indicate that miRNA targeting causes mostly mRNA destabilization, with translational repression comprising a detectable but minor component of the overall repression[33–36]. The only known exception is the transient translational repression observed in early zebrafish embryos[36]. At 4 hpf miR-430 targeting causes mostly translational repression with very little mRNA destabilization, whereas by 6 hpf the outcome shifts to mostly mRNA destabilization[36]. Because miR-430 is induced only ~1.5 h before the 4-hpf stage, these results are interpreted as revealing the dynamics of miRNA action, in which an early phase of translational repression gives way to a later phase in which destabilization dominates[36]. When considering that miRNA targeting promotes poly(A)-tail shortening through the recruitment of deadenylase complexes[37], our results suggest an alternative mechanism for the shift in miRNA regulatory outcomes. In this mechanism, miRNAs mediate tail shortening at both 4 and 6 hpf, but because of the switch in the nature of translational control (as well as destabilization of short-tailed mRNAs at later stages), tail shortening has very different consequences in the two stages: at 4 hpf, tail shortening predominantly decreases translational efficiency, whereas at 6 hpf, it predominantly decreases mRNA stability.

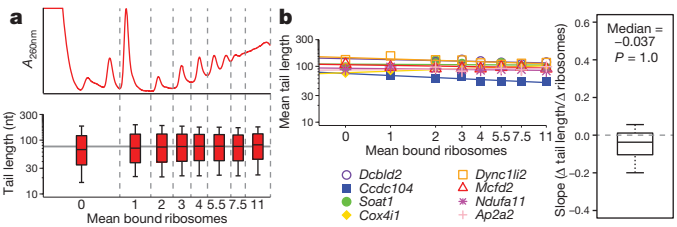


Figure 4 | No detectable intragenic coupling between poly(A)-tail length and translational efficiency. **a**, Global analysis of tail lengths across a polysome profile for 3T3 cells. The absorbance trace indicates mean number of ribosomes bound per mRNA for each fraction from the sucrose gradient (top, fractions demarcated with vertical dashed lines). Box plots show distributions of bulk tail lengths in each fraction for all tags mapping to annotated 3′ UTRs (bottom). Box plot percentiles are line, median; box, 25th and 75th percentiles; whiskers, 10th and 90th percentiles. The horizontal line indicates the overall median of the median tail lengths. **b**, Relationship between tail lengths and ribosomes bound per mRNA for mRNAs from the same gene. For each gene, the data from **a** were used to plot the mean tail length as a function of bound ribosomes. Log-log plots for 8 randomly selected genes with ≥ 50 poly(A) tags in ≥ 6 fractions are shown (left), with lines indicating linear least-squares fits to the data (adding a pseudocount of 0.5 ribosomes to the fraction with 0 ribosomes). The box plot shows the distribution of slopes for all genes with ≥ 50 poly(A) tags in ≥ 4 fractions (right; $n = 4{,}079$; one-sided, one-sample Wilcoxon test; box plot percentiles as in **a**).

To integrate miRNA-mediated repression with effects on tail length, we injected one-cell zebrafish embryos with miRNAs that are normally not present in the early embryo and examined the influence of these injected miRNAs on ribosome-protected fragments, mRNA levels and poly(A)-tail lengths at 2, 4 and 6 hpf. Injecting miR-155 caused ribosome-protected fragments from many of its predicted targets to decrease relative to ribosome-protected fragments from no-site control mRNAs (Fig. 5a). Despite the decrease in ribosome-protected fragments, target mRNA levels did not change relative to the controls at 2 and 4 hpf, indicating that at these stages miR-155 targeting caused mostly translational repression. In contrast, decreases in ribosome-protected fragments were accompanied by nearly commensurate mRNA reductions at 6 hpf, indicating that by this stage the outcome of repression had shifted to mostly mRNA destabilization (Fig. 5a). Thus, the shift in miRNA regulatory outcome that occurs between 4 and 6 hpf is not specific to miR-430 or its targets. With respect to mechanism, the observation of this shift between 4 and 6 hpf, even though the injected miR-155 was present and active much earlier than was miR-430, indicated that the shift reflected a transition from the unusual regulatory regime operating in pre-gastrulation embryos (in which translational efficiency is sensitive to tail length) more than it reflected the dynamics of miRNA action.

The tail-length results further supported a mechanism involving shifting consequences of tail-length shortening. Predicted miR-155 targets had shortened tails at 2 and 4 hpf (Fig. 5b), which explained most of the miRNA-induced translational repression observed at these stages (Fig. 5c). By 6 hpf, the tail-length decreases observed at 4 hpf had mostly abated for predicted miR-155 targets (Fig. 5b), and these mRNAs were instead less abundant (Fig. 5a), in concordance with their extent of deadenylation at 4 hpf (Extended Data Fig. 8a). These observations agreed with the idea that tail shortening destabilizes mRNAs at later developmental stages and indicated that the miRNA-mediated deadenylation occurring during the earlier developmental stages promotes decay later. With shorter tails no longer associated with reduced translation (Fig. 3a) and instead associated with reduced mRNA levels, the ultimate consequence of miRNA-mediated repression shifted from translational repression to mRNA destabilization (Fig. 5c). Analogous results were obtained after injecting a different miRNA, miR-132 (Fig. 5c, Extended Data Fig. 8).

Because tail length was no longer strongly coupled with translational efficiency (Fig. 3a), tail-length changes did not explain the decrease in mean translational efficiency observed at 6 hpf for predicted miR-132 targets (Fig. 5c). We conclude that when poly(A)-tail length is uncoupled from translational efficiency, the translational repression often detected as a minor component of the overall repression[33–35] arises from a mechanism different from the one that dominates pre-gastrulation.

Our results provide a compelling explanation for miRNA-mediated translational repression in the pre-gastrulation zebrafish embryo: miRNAs induce poly(A) shortening, which decreases translational efficiency at this developmental period. They also explain why the pre-gastrulation zebrafish embryo is the only known context for which translational repression is the dominant outcome of miRNA-mediated regulation; in all other contexts examined, tail-length shortening causes mRNA destabilization with little or no effect on translational efficiency.

Two gene-regulatory regimes

Our results from yeast, cultured mammalian cells and mouse liver refute the prevailing view that poly(A)-tail length broadly influences

©2014 Macmillan Publishers Limited. All rights reserved

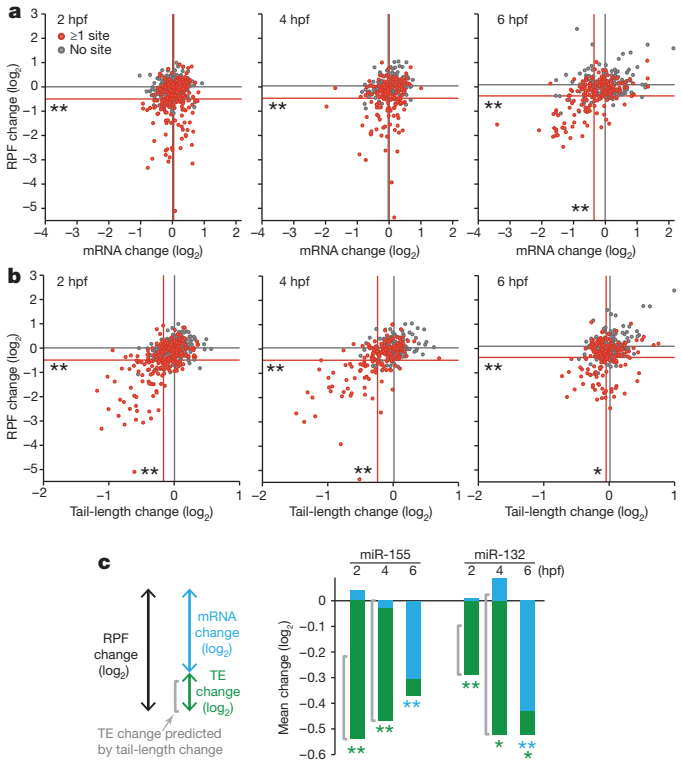


Figure 5 | The influence of miR-155 on ribosomes, mRNA abundance and tails in the early zebrafish embryo. **a**, Relationship between changes in ribosome-protected fragments (RPFs) and changes in mRNA levels after injecting miR-155. Changes observed between miRNA- and mock-injected embryos are plotted at the indicated stages for predicted miR-155 target genes (red, genes with ≥ 1 miR-155 site in their 3′ UTR) and control genes (grey, genes that have no miR-155 site, yet resemble the predicted targets with respect to 3′-UTR length). To ensure that differences observed between 4 and 6 hpf were not the result of examining different genes, only site-containing genes and no-site control genes detected at both 4 and 6 hpf are shown for these stages. Lines indicate mean changes for the respective gene sets, with statistically significant differences between the sets indicated (*$P \leq 0.05$; **$P < 10^{-4}$, one-tailed Kolmogorov–Smirnov test). Because injected miRNAs partially inhibited miR-430-mediated repression, genes with miR-430 sites were not considered. Data were normalized to the median changes observed for the controls. **b**, Relationship between changes in ribosome-protected fragments and changes in mean tail lengths after injecting miR-155. Tail lengths were determined using PAL-seq, otherwise as in **a**. **c**, A developmental switch in the dominant mode of miRNA-mediated repression. The schematic (left) depicts the components of the bar graphs, showing how the changes in ribosome-protected fragments (RPFs) comprise both mRNA and translational efficiency (TE) changes. The compound bar graphs show the fraction of repression attributed to mRNA degradation (blue) and translational efficiency (green) for the indicated stage, depicting the overall impact of miR-155 (centre; plotting results from **a** and **b** for genes with sites) and miR-132 (right, plotting results from Extended Data Fig. 8b for genes with sites). Slight, statistically insignificant increases in mRNA for predicted targets resulted in blue bars extending above the axis. For samples from stages at which tail length and translational efficiency were strongly coupled, a bracket adjacent to the compound bar indicates the fraction of repression attributable to shortened tails. Significant changes for each component are indicated with asterisks of the corresponding colour (*$P \leq 0.05$; **$P < 10^{-4}$, one-tailed Kolmogorov–Smirnov test).

translational efficiency. In doing so, they add to the known differences between the regulatory regime operating in these cells and that operating in early metazoan embryos.

This absence or presence of coupling between poly(A)-tail length and translational efficiency can be rationalized in light of the potential interplay among regulatory options available in the two regulatory regimes. Our yeast, mammalian and gastrulation-stage cells were transcriptionally active, which offers ample opportunities for nuclear control of gene expression. Moreover, active transcription enables unstable mRNAs to be replaced if required, thereby expanding the contexts in which differential mRNA stability can be exploited for gene control. Thus, an additional layer of control in which translational efficiency depends on poly(A)-tail length is dispensable. More importantly, because this type of coupling would lower output from older mRNA molecules that, in the absence of cytoplasmic polyadenylation, would often have shorter poly(A) tails, the utility of gene regulation through mRNA stability would be compromised. In this conventional regulatory regime, long-lived mRNAs would have less value if they were translated less efficiently because of their shorter tails.

For fish and frog embryos at the cleavage stage, the regulatory regime was very different. These embryos were transcriptionally inactive, which not only precludes the use of transcription and other nuclear processes to alter gene expression programs but also limits the use of differential mRNA stability, because degraded mRNAs cannot be replaced until zygotic transcription begins. Perhaps as a consequence, many mRNAs with short tails were observed (Fig. 2a), consistent with the known stability of short-tailed mRNAs in early embryos[19,38]. In these circumstances, early embryonic cells apparently harness differential tail length for global gene control. This result expands the known behaviour of individual genes in *Xenopus* embryos[23,24] and the observation that early embryonic cells have robust cytoplasmic polyadenylation[4], which increases the utility of a tail-length regulatory mechanism. Compared to metazoan cells subject to the standard regulatory regime (for example, 6-hpf zebrafish embryos and the mammalian cells examined), cleavage-stage embryos had more uniform intragenic tail lengths and more variable intergenic lengths (Fig. 2d), as required for efficient harnessing of the tail-length regulatory regime. With their tail-length distribution also shifted towards shorter tails (Fig. 2b), cleavage-stage embryos can most efficiently exploit the tail-length differences with the greatest impact (Fig. 3a).

The transition between these two very different gene-regulatory regimes was rapid but not immediate. Despite their zygotic transcription, late-blastula embryos still coupled tail length with translation. Indeed, to the extent that newly transcribed zygotic mRNAs tended to have longer tails than did the maternally inherited mRNAs (Extended Data Fig. 5), the continued coupling observed in this hybrid regime would act to increase the relative output from these newly minted mRNAs, thereby sharpening the MZT.

We suspect that the tail-length regulatory regime observed in early embryos operates in other systems in which transcription is repressed (or occurs at a distant location) and cytoplasmic polyadenylation is active, such as early embryos of other metazoan species, maturing oocytes and neuronal synapses[5]. The ability to measure poly(A)-tail lengths at single-mRNA resolution should provide important insights in these systems.

METHODS SUMMARY

Cytoplasmically enriched lysates were prepared from HEK293T, 3T3, mouse liver, *X. laevis*, *S. pombe* and zebrafish samples, as well as one of the *S. cerevisiae* samples, and divided into three portions, one each for PAL-seq, RNA-seq and ribosome profiling. PAL-seq was performed as outlined in Fig. 1a. RNA-seq and ribosome profiling were performed essentially as described previously[35]. Poly(A) tags were mapped to a reference genome (or transcriptome) of the species, carrying forward those that mapped uniquely to the genome (or transcriptome) and also overlapped with the 3′ UTR of a transcript model chosen to represent a gene. Ribosome-protected fragments and RNA-seq tags were mapped to ORFs, as described previously[35], except tags mapping within the first 50 nucleotides of an ORF were discarded (to exclude signal from ribosomes that might have initiated after cycloheximide was added). Each mRNA with a 3′ UTR that had at least one 7-nucleotide site matching the miRNA seed region[32] was predicted to be a target of that miRNA. Genes that had no 6-nucleotide seed match anywhere within their mature transcript were classified as no-site genes, from which a set of no-site control genes was selected such that its 3′-UTR length distribution matched that of the predicted targets.

©2014 Macmillan Publishers Limited. All rights reserved

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 25 July; accepted 23 December 2013.
Published online 29 January 2014.

1. Moore, M. J. & Proudfoot, N. J. Pre-mRNA processing reaches back to transcription and ahead to translation. *Cell* **136,** 688–700 (2009).
2. Goldstrohm, A. C. & Wickens, M. Multifunctional deadenylase complexes diversify mRNA control. *Nature Rev. Mol. Cell Biol.* **9,** 337–344 (2008).
3. Chen, C. Y. & Shyu, A. B. Mechanisms of deadenylation-dependent decay. *Wiley Interdiscip. Rev. RNA* **2,** 167–183 (2011).
4. Richter, J. D. Cytoplasmic polyadenylation in development and beyond. *Microbiol. Mol. Biol. Rev.* **63,** 446–456 (1999).
5. Weill, L., Belloc, E., Bava, F. A. & Mendez, R. Translational control by changes in poly(A) tail length: recycling mRNAs. *Nature Struct. Mol. Biol.* **19,** 577–585 (2012).
6. Eckmann, C. R., Rammelt, C. & Wahle, E. Control of poly(A) tail length. *Wiley Interdiscip. Rev. RNA* **2,** 348–361 (2011).
7. Sallés, F. J., Lieberfarb, M. E., Wreden, C., Gergen, J. P. & Strickland, S. Coordinate initiation of *Drosophila* development by regulated polyadenylation of maternal messenger RNAs. *Science* **266,** 1996–1999 (1994).
8. Barkoff, A., Ballantyne, S. & Wickens, M. Meiotic maturation in *Xenopus* requires polyadenylation of multiple mRNAs. *EMBO J.* **17,** 3168–3175 (1998).
9. Preiss, T., Muckenthaler, M. & Hentze, M. W. Poly(A)-tail-promoted translation in yeast: implications for translational control. *RNA* **4,** 1321–1331 (1998).
10. Beilharz, T. H. & Preiss, T. Widespread use of poly(A) tail length control to accentuate expression of the yeast transcriptome. *RNA* **13,** 982–997 (2007).
11. Lackner, D. H. *et al.* A network of multiple regulatory layers shapes gene expression in fission yeast. *Mol. Cell* **26,** 145–155 (2007).
12. Nutiu, R. *et al.* Direct measurement of DNA affinity landscapes on a high-throughput sequencing instrument. *Nature Biotechnol.* **29,** 659–664 (2011).
13. Rosenthal, E. T., Tansey, T. R. & Ruderman, J. V. Sequence-specific adenylations and deadenylations accompany changes in the translation of maternal messenger RNA after fertilization of *Spisula* oocytes. *J. Mol. Biol.* **166,** 309–327 (1983).
14. Palatnik, C. M., Wilkins, C. & Jacobson, A. Translational control during early *Dictyostelium* development: possible involvement of poly(A) sequences. *Cell* **36,** 1017–1025 (1984).
15. Paynton, B. V., Rempel, R. & Bachvarova, R. Changes in state of adenylation and time course of degradation of maternal mRNAs during oocyte maturation and early embryonic development in the mouse. *Dev. Biol.* **129,** 304–314 (1988).
16. Kane, D. A. & Kimmel, C. B. The zebrafish midblastula transition. *Development* **119,** 447–456 (1993).
17. Newport, J. & Kirschner, M. A major developmental transition in early *Xenopus* embryos: II. Control of the onset of transcription. *Cell* **30,** 687–696 (1982).
18. Decker, C. J. & Parker, R. A turnover pathway for both stable and unstable mRNAs in yeast: evidence for a requirement for deadenylation. *Genes Dev.* **7,** 1632–1643 (1993).
19. Aanes, H. *et al.* Zebrafish mRNA sequencing deciphers novelties in transcriptome dynamics during maternal to zygotic transition. *Genome Res.* **21,** 1328–1338 (2011).
20. Ingolia, N. T., Ghaemmaghami, S., Newman, J. R. & Weissman, J. S. Genome-wide analysis *in vivo* of translation with nucleotide resolution using ribosome profiling. *Science* **324,** 218–223 (2009).
21. McGrew, L. L., Dworkin-Rastl, E., Dworkin, M. B. & Richter, J. D. Poly(A) elongation during *Xenopus* oocyte maturation is required for translational recruitment and is mediated by a short sequence element. *Genes Dev.* **3,** 803–815 (1989).
22. Paris, J. & Richter, J. D. Maturation-specific polyadenylation and translational control: diversity of cytoplasmic polyadenylation elements, influence of poly(A) tail size, and formation of stable polyadenylation complexes. *Mol. Cell. Biol.* **10,** 5634–5645 (1990).
23. Paris, J. & Philippe, M. Poly(A) metabolism and polysomal recruitment of maternal mRNAs during early *Xenopus* development. *Dev. Biol.* **140,** 221–224 (1990).
24. Simon, R., Tassan, J. P. & Richter, J. D. Translational control by poly(A) elongation during *Xenopus* development: differential repression and enhancement by a novel cytoplasmic polyadenylation element. *Genes Dev.* **6,** 2580–2591 (1992).
25. Vassalli, J. D. *et al.* Regulated polyadenylation controls mRNA translation during meiotic maturation of mouse oocytes. *Genes Dev.* **3,** 2163–2171 (1989).
26. Gebauer, F., Xu, W., Cooper, G. M. & Richter, J. D. Translational control by cytoplasmic polyadenylation of *c-mos* mRNA is necessary for oocyte maturation in the mouse. *EMBO J.* **13,** 5712–5720 (1994).
27. Wu, L. *et al.* CPEB-mediated cytoplasmic polyadenylation and the regulation of experience-dependent translation of α-CaMKII mRNA at synapses. *Neuron* **21,** 1129–1139 (1998).
28. Oh, B., Hwang, S., McLaughlin, J., Solter, D. & Knowles, B. B. Timely translation during the mouse oocyte-to-embryo transition. *Development* **127,** 3795–3803 (2000).
29. Burns, D. M. & Richter, J. D. CPEB regulation of human cellular senescence, energy metabolism, and p53 mRNA translation. *Genes Dev.* **22,** 3449–3460 (2008).
30. Novoa, I., Gallego, J., Ferreira, P. G. & Mendez, R. Mitotic cell-cycle progression is regulated by CPEB1 and CPEB4-dependent translational control. *Nature Cell Biol.* **12,** 447–456 (2010).
31. Schwanhäusser, B. *et al.* Global quantification of mammalian gene expression control. *Nature* **473,** 337–342 (2011).
32. Bartel, D. P. MicroRNAs: target recognition and regulatory functions. *Cell* **136,** 215–233 (2009).
33. Baek, D. *et al.* The impact of microRNAs on protein output. *Nature* **455,** 64–71 (2008).
34. Hendrickson, D. G. *et al.* Concordant regulation of translation and mRNA abundance for hundreds of targets of a human microRNA. *PLoS Biol.* **7,** e1000238 (2009).
35. Guo, H., Ingolia, N. T., Weissman, J. S. & Bartel, D. P. Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature* **466,** 835–840 (2010).
36. Bazzini, A. A., Lee, M. T. & Giraldez, A. J. Ribosome profiling shows that miR-430 reduces translation before causing mRNA decay in zebrafish. *Science* **336,** 233–237 (2012).
37. Braun, J. E., Huntzinger, E. & Izaurralde, E. A molecular link between miRISCs and deadenylases provides new insight into the mechanism of gene silencing by microRNAs. *Cold Spring Harb. Perspect. Biol.* **4,** a012328 (2012).
38. Audic, Y., Omilli, F. & Osborne, H. B. Postfertilization deadenylation of mRNAs in *Xenopus laevis* embryos is sufficient to cause their degradation at the blastula stage. *Mol. Cell. Biol.* **17,** 209–218 (1997).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank D. Weinberg, V. Auyeung, I. Ulitsky, C. Jan, J.-W. Nam, A. Shkumatava, S.-J. Hong, Y. Erlich and the Whitehead Genome Technology Core (V. Dhanapal, L. Francis, S. Gupta and T. Volkert) for discussions; J.-W. Nam, I. Ulitsky and D. Weinberg for assistance with transcript annotation; C. Bresilla, X. Guo, S.-J. Hong and A. Rothman for experimental assistance; and D. Weinberg for comments on the manuscript. Supported by NIH grant GM067031 (D.P.B.) and NIH Medical Scientist Training Program fellowship T32GM007753 (A.O.S.). D.P.B. is an investigator of the Howard Hughes Medical Institute.

Author Contributions A.O.S. developed PAL-seq, generated tail-length measurements, and performed associated analyses. S.W.E. performed ribosome profiling, RNA-seq and associated analyses. G.R.C. performed zebrafish injections and assisted with staging. D.P.B. supervised with help from H.S. All authors helped to design the study and write the manuscript.

Author Information Sequencing data and the processed data for each gene are available at the Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo) under accession number GSE52809. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.P.B. (dbartel@wi.mit.edu).

©2014 Macmillan Publishers Limited. All rights reserved

METHODS

PAL-seq. Total RNA or RNA from cytoplasmically enriched lysate (~1–50 μg) was supplemented with two mixes of tail-length standards and trace marker RNA containing an internal ^{32}P-label (*) (5′-ugagguaguagguuguauagu*caauccuaauccauuccaauccuaaucauucaaaaaaaaaa-3′, IDT), which was used to monitor subsequent ligation, partial-digestion and capture steps. Polyadenylated ends in the mixture of cellular RNA and standards were ligated to a 3′-biotinylated adaptor DNA oligonucleotide (5′-pAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGACA CATAC-biotin-3′, IDT) in the presence of a splint DNA oligonucleotide (5′-TT CCGATCTTTTTTTTTT-3′, IDT) using T4 Rnl2 (NEB) in an overnight reaction at 18 °C. The RNA was partially digested with RNase T1 (Ambion) as described[39], extracted with phenol-chloroform, ethanol precipitated and then purified on a denaturing polyacrylamide gel (selecting 104–750-nucleotide fragments), which removed residual unreacted 3′ adaptor. Splinted-ligation products were captured on streptavidin M-280 Dynabeads (Invitrogen) and, when still bound to the beads, 5′ phosphorylated with 3′-phosphatase–deficient T4 polynucleotide kinase (NEB) and ligated to a 5′ adaptor oligonucleotide (5′-C3.spacer-CAAGCAGAAGACGG CATACGAGTTCAGAGTTCTAcaguccgacgauc-3′, IDT; uppercase, DNA; lowercase, RNA) using T4 Rnl1 (NEB) in an overnight reaction at 22 °C. Following reverse transcription using SuperScript II (Invitrogen) and a primer oligonucleotide (5′-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACG-3′, IDT), complementary DNA (cDNA) was liberated from the beads by base hydrolysis and purified on a denaturing polyacrylamide gel (selecting 166–790-nucleotide length DNA). Purified cDNA was denatured at room temperature in 5–100 mM NaOH, neutralized with addition of HT1 hybridization buffer (Illumina) and applied to an Illumina flow cell (at a typical concentration of 1.0–1.2 pM). Standard cluster generation, linearization, 3′-end blocking and primer hybridization were performed on a cBot cluster generation system (Illumina).

After transferring the flow cell to a Cluster Station designed for an Illumina Genome Analyzer, the sequencing primer was extended using a reaction mix containing 100 U ml^{-1} Klenow polymerase (NEB), 200 nM dTTP and either 10 nM (yeast samples) or 4 nM (other samples) biotin-16-dUTP (Roche). Extension was for 30 min at 37 °C, flowing a fresh aliquot (≥ 50 μl) of reaction mix every 2 min to replenish dNTPs. Following primer extension, the flow cell was placed on a Genome Analyzer II sequencer (Illumina) for 36 cycles of standard sequencing-by-synthesis. After three additional cycles of cleavage to remove any residual sequencing fluorophores, the flow cell was washed with buffer (40.25 mM phosphate buffered saline (PBS), pH 7.4, 0.1% Tween), blocked with streptavidin-binding buffer (300 μg ml^{-1} bovine serum albumin (NEB), 40.25 mM PBS, pH 7.4, 0.1% Tween), washed with buffer again, and then imaged, as carried out previously[12]. This cycle of wash, block, wash, image was then repeated with a binding step inserted after the blocking step, in which the flow cell was incubated with 30 nM Alexa Fluor 532 streptavidin (Invitrogen) in streptavidin-binding buffer for 10 min at 20 °C. This expanded cycle was then repeated two more times, but with 100 nM streptavidin included at the binding step. Fluorescence was captured in the T and G channels because the wavelength of the excitation laser for these channels (532 nm) was identical to the fluorophore excitation wavelength. The sequential imaging confirmed that the second 100 nM streptavidin incubation did not increase mean cluster intensity (monitored in real time as part of the 'first base report'), which indicated saturation of available biotin.

Calculation of poly(A)-tail lengths. Raw images taken during sequencing-by-synthesis and after binding of fluorescent streptavidin were processed with Firecrest image-analysis and Bustard base-calling software (Illumina, version 1.9.0, using default parameters) to generate a read (FASTQ) file and another file containing the position of each cluster, the read sequence, the quality score for each base, and the base intensities in all four channels for every cycle of sequencing and streptavidin binding. Reads were aligned to a reference genome (hg18 for human, mm9 for mouse, dm3 for fly, danRer7 for fish, TAIR10 for *A. thaliana*, Spombe1 for *S. pombe*, and sacCer3 for *S. cerevisiae*) or a reference transcriptome (curated from Unigene mRNA sequences for *X. laevis*) using the Bowtie program for short-read mapping and the parameters '–l 25 –n 2 –m 1 –3 *z*', in which *z* was the number of streptavidin-binding cycles plus one. Reads containing ambiguous base calls (as indicated by characters 'N' or '.') at any position in the first 36 nucleotides were discarded, as were reads mapping to multiple genomic loci. Reads that did not map to the genome were aligned to Bowtie indexes corresponding to the tail-length standards. For the remaining reads, mapping to the genome and standards was repeated, accounting for the possibility that the read failed to map because the sequence extended past the poly(A)-proximal fragment of the transcript and into the 5′ adaptor. This mapping was reiterated for 16 rounds (to capture tags of ≥ 20 nucleotides, with each round considering previously unmapped reads in which the 5′ adaptor sequence started a nucleotide closer to the beginning of the read (requiring a perfect match to only the final 6 nucleotides of the adaptor after the fifth round)). Before each round of mapping, the adaptor sequence was stripped by adjusting the Bowtie '−3' parameter. For the *A. thaliana* sample, the first sequenced base was of low quality, and raw images from the first cycle of sequencing were excluded for image analysis and base calling. Consequently, sequence reads were 35 nucleotides long, and only 15 rounds of iterative adapter trimming/sequence mapping were performed.

For each read carried forward as mapping to a single locus (of either the genome or the length standards), the cluster fluorescence intensity in the T channel after the first 100 nM streptavidin flow-in was recorded as the raw streptavidin fluorescence intensity. From this raw intensity, the intensity after the 0 nM flow-in was subtracted as background, and the resulting background-subtracted intensity was divided by the relative cluster intensity observed during sequencing-by-synthesis, which normalized for the density of molecules within the cluster. The relative cluster intensity was calculated by first dividing the fluorescence intensity of every sequenced base in the read by the median intensity for that base among all clusters with the same base at the same position, and then taking the average of the resulting values over the length of the read. Normalized streptavidin intensities were transformed to poly(A)-tail lengths using linear regression parameters derived from the median intensities of the standards and their mode poly(A)-tail lengths. For yeast, *Arabidopsis*, *Drosophila*, *Xenopus* and zebrafish samples, only the standards with tails of 10, 50 and 100 nucleotides were used in the linear regression. For the other samples, all of the standards were used except for one with a 324-nucleotide tail (barcode sequence = 5′-CUCACUAUAC-3′), which was typically not sufficiently abundant for accurate measurement of its tail length. Each tail length was then paired with the genomic (or standard) coordinates to yield a poly(A) tag.

Assigning poly(A) tags to genes. Reference transcript annotations were downloaded (in refFlat format) from the UCSC Genome Browser or another database (Ensembl for zebrafish, Unigene for *X. laevis*, TAIR for *A. thaliana* and PomBase for *S. pombe*). For human, mouse, zebrafish and fly, transcript 3′ ends were re-annotated using poly(A) sites identified by 3P-seq[39] and the workflow described previously[40]. For each gene, a representative transcript model was chosen as the one that had the longest ORF and the longest 3′ UTR corresponding to that ORF. These reference transcript databases and a file with the sequences of the internal standards are available for anonymous download (http://web.wi.mit.edu/bartel/pub/publication.html). *S. cerevisiae* representative transcript models were from Gene Expression Omnibus accession GSE53268. Poly(A) tags that overlapped the 3′ UTR of the representative transcript model by at least one nucleotide were assigned to that gene. Tags with tail-length measurements < -50 and $> 1{,}000$ nucleotides (which included < 0.0009% of the tags in any sample) were excluded from all analyses. Mean tail-length measurements < 1 nucleotide (which included measurements from 11 analysed genes) were replaced with a value of 1.0 nucleotide in the intragenic analysis across the polysome gradient (Fig. 4b). When considering the depth of a representative PAL-seq data set from 3T3 cells, we considered 1.0 reads per kilobase per million reads (RPKM) as the RNA-seq level indicating an average of one mRNA molecule per cell. This estimate was conservative, in that a comparison to published mRNA abundances in 3T3 cells[31] indicated that 1.0 RPKM from our experiment corresponded to about 0.2 mRNA molecules per 3T3 cell.

RNA preparation for PAL-seq. For libraries made from *S. pombe*, HEK293T, 3T3, mouse liver, *X. laevis*, and mock-, miR-132- and miR-155-injected zebrafish samples, as well as the *S. cerevisiae* sample analysing cytoplasmically enriched RNA, RNA was extracted from a portion of the lysate prepared for ribosome profiling and RNA-seq. These cleared lysates were enriched in cytoplasm. For libraries made from HeLa and polysome-gradient samples, RNA was extracted from similar cytoplasmically enriched lysates. For the polysome gradient fractionation (Fig. 4a), lysate preparation and centrifugation were performed as for ribosome profiling, but without nuclease digestion before fractionation. For other libraries, total RNA was used. The correlation observed when comparing PAL-seq results from HeLa cytoplasmically enriched RNA and total RNA resembled that observed between biological replicates (Extended Data Fig. 2b; $R_s = 0.84$ and 0.83, respectively). The measured lengths in both types of RNA preparation were similar, despite the possibility that total RNA might have included more long-tailed mRNAs due to a population of nascent mRNAs that had full-length tails and were awaiting export to the cytoplasm. However, not all nuclear mRNAs are expected to have full-length tails (as some would still be in the process of being polyadenylated at the time of sample collection), and the nuclear population of mRNAs awaiting export presumably comprised a small fraction of the cellular mRNAs.

Tail-length standards. The common 5′ region of each standard and the unique 3′ region, consisting of the standard-specific barcode and poly(A)-tail (Fig. 1b), were synthesized separately and then ligated together to make full-length standards. To generate each 3′ region, a 5′-phosphate-bearing RNA oligonucleotide (IDT) consisting of the barcode segment followed by a 10-nucleotide poly(A) segment was extended with *E. coli* poly(A) polymerase (NEB), with ATP concentration and reaction time adjusted to yield tails of the desired length. To narrow the tail-length distribution, extension products were sequentially purified on two

©2014 Macmillan Publishers Limited. All rights reserved

denaturing polyacrylamide gels, excising products with tails of the desired length range and reducing the variability of tailed RNA to be mostly within ~5–25 nucleotides, depending on the length of tail added (Extended Data Fig. 1b). The 5′ region of the standards was synthesized by *in vitro* transcription of a template containing *Renilla* luciferase sequence followed by that of a modified HDV ribozyme[41]. After gel purification of the 5′ product of HDV self-cleavage, the 2′,3′-cyclic phosphate at its 3′ end was removed with T4 polynucleotide kinase (NEB; 3,000 µl reaction containing 30,000 U of enzyme and 100 mM MES-NaOH, pH 5.5, 10 mM $MgCl_2$, 10 mM β-mercaptoethanol, 300 mM NaCl, 37 °C, 6 h). After another gel purification, the dephosphorylated product was joined to the poly(A)-tailed barcode oligonucleotide by splinted ligation using T4 Rnl2 (NEB) and a DNA bridge oligonucleotide with 10 nucleotides of complementarity to each side of the ligation junction. Ligation products were gel purified and mixed in desired ratios before being added to RNA samples for PAL-seq.

Ribosome footprint profiling. Immediately before sample collection, cultured mammalian cells were incubated with media containing 100 µg ml^{-1} cycloheximide for 10 min at 37 °C to stop translation elongation. Cells were washed twice with ice-cold 9.5 mM PBS, pH 7.3, containing 100 µg ml^{-1} cycloheximide, and lysed by adding lysis buffer (10 mM Tris-HCl, pH 7.4, 5 mM $MgCl_2$, 100 mM KCl, 2 mM dithiothreitol, 100 µg ml^{-1} cycloheximide, 1% Triton X-100, 500 U ml^{-1} RNasin Plus, and protease inhibitor (1x complete, EDTA-free, Roche)) and triturating four times with a 26-gauge needle. After centrifuging the crude lysate at 1,300*g* for 10 min at 4 °C, the supernatant was removed and flash-frozen in liquid nitrogen. Cultured *S. pombe* cells were grown to mid-log phase and then harvested (without cycloheximide pre-treatment) by filtering off the media and flash freezing the remaining paste, which was then manually ground into a fine powder with a mortar and pestle while being bathed in liquid nitrogen. The powder was thawed on ice, resuspended in lysis buffer and processed as described for the other lysates. Zebrafish embryos were enzymatically dechorionated and then incubated in 100 µg ml^{-1} cycloheximide in E3 buffer (5 mM NaCl, 0.17 mM KCl, 0.33 mM $CaCl_2$, 0.33 mM $MgSO_4$) for 5 min at room temperature. The embryos were then transferred into lysis buffer and flash frozen. Soon after fertilization the jelly membranes of *X. laevis* embryos were chemically removed, and at the desired stages, embryos were flash frozen in lysis buffer without cycloheximide pre-treatment. Once thawed, these samples were clarified as above and then processed in the same manner as other lysates. Prior to dissecting liver, a 6-week-old, male C57BL/6 mouse was killed by cervical dislocation. The liver was excised, flash frozen, and manually ground and processed as described for *S. pombe*. Ribosome profiling and RNA-seq were performed on cleared lysates essentially as described[35], using RiboMinus-treated RNA for the *S. pombe* RNA-seq sample, and poly(A)-selected RNA for all others, with a detailed protocol available at http://bartellab.wi.mit.edu/protocols.html. *S. cerevisiae* RPF and RNA-seq data were from GSE53268 and were derived from the same sample as the *S. cerevisiae* PAL-seq sample analysing cytoplasmically enriched RNA.

RPF and RNA-seq tags were mapped to the ORFs, as described previously[35] (using the assemblies and transcript models used for PAL-seq), except reads overlapping the first 50 nucleotides of each ORF were disregarded. This was done to minimize a bias from ribosomes accumulating at or shortly after the start codon, which results from translation initiation events continuing in the face of cycloheximide-inhibited elongation[20]. Because of this bias, genes with shorter ORFs have artefactually higher translational efficiencies if all the bound ribosomes are considered (as in conventional polysome gradient analysis). This cycloheximide effect might have distorted the translational efficiency measurements in studies that calculated ribosome densities using polysome gradient fractionation followed by microarray analysis (including those reporting a positive correlation between ribosome density and poly(A)-tail length[10,11]), but could not have influenced the conclusions of our polysome-gradient experiment, because our analysis focused on intragenic comparisons (Fig. 4b). Translational efficiencies were considered only for genes exceeding a cutoff of 10 RPM (reads per million uniquely mapped reads) in the RNA-seq library. When calculating sequencing depth (the 'M' of RPM), all uniquely mapped reads that overlapped the mRNA primary or mature transcript were counted for all samples except the *X. laevis* samples; only the uniquely mapped reads overlapping ORFs were counted for *X. laevis*. For the analysis of miRNA effects, only genes exceeding a cutoff of 10 RPM in the mock-injected RNA-seq and RPF libraries, and ≥ 50 PAL-seq tags in the mock-injected and miRNA-injected samples were considered.

Statistics, reagents and animal models. All statistical tests were two-sided unless indicated otherwise. No power testing was done to anticipate the sample size needed for adequate statistical power. No randomization or blinding was used for miRNA injection experiments. Features of mRNAs (for example, poly(A)-tail length, mRNA length, expression level, and so on) were not normally distributed, nor were changes in expression due to miRNA-mediated repression. Therefore, non-parametric measures or tests were used when making comparisons involving such quantities, and these tests do not make assumptions about equal variance between groups. Mammalian cell lines were obtained from ATCC, and S2 cells were the same as in ref. 42 (that is, adapted to growth in serum-free media). The BY4741 strain was used for *S. cerevisiae*, 972 for *S. pombe,* Columbia for *A. thaliana*, and AB for zebrafish. All animal experiments were performed in accordance with a protocol approved by the MIT Committee on Animal Care.

Zebrafish injections. Zebrafish embryos were injected at the one-cell stage with 1 nl of 10 µM miRNA duplex (miR-155 or miR-132) or buffer alone using a PLI-100 Plus Pico-Injector. Duplexes were made by combining RNAs (IDT) corresponding to either miR-132 (5′-uaacagucuacagccauggucg-3′) and miR-132* (5′-accguggcauuagauuguuacu-3′) or miR-155 (5′-uuaaugcuaaucgugauaggggu-3′) and miR-155* (5′-accuaugcuguuagcauuaauc-3′) in annealing buffer (30 mM Tris-HCl, pH 7.5, 100 mM NaCl, 0.1 mM EDTA), heating to 90 °C for 1 min, and slow cooling to room temperature over several hours. Injected embryos were incubated in E3 buffer at 28 °C until time of sample collection.

Predicted miRNA targets. MicroRNA target genes were predicted using the reference transcript database used to assign zebrafish poly(A) tags. Each gene with a 3′ UTR that had at least one 7-nucleotide site matching the miRNA seed region[32] was predicted to be a target of that miRNA. Genes that had no 6-nucleotide miRNA seed match anywhere within their transcript were classified as no-site genes, from which a set of no-site control genes was selected such that its 3′-UTR length distribution matched that of the predicted targets.

Calculation of the relationship between poly(A)-tail length and translational efficiency. For experiments in which zebrafish embryos were mock-injected or injected with miR-132 or miR-155, least-squares second-order polynomial regression was performed to determine the change in $\log_2$ translational efficiency for each change in $\log_2$ poly(A)-tail length. To prevent microRNA effects on translational efficiency and/or tail length from influencing any relationship, the regression analyses were performed after excluding genes for which the mRNAs contained a perfect match to either the seed (nucleotides 2–7 of the miRNA) of miR-430 (the predominant endogenous miRNA at 4 and 6 hpf) or the seed of the injected miRNA. These regression results were used to estimate the translational efficiency change attributable to tail-length change for each gene.

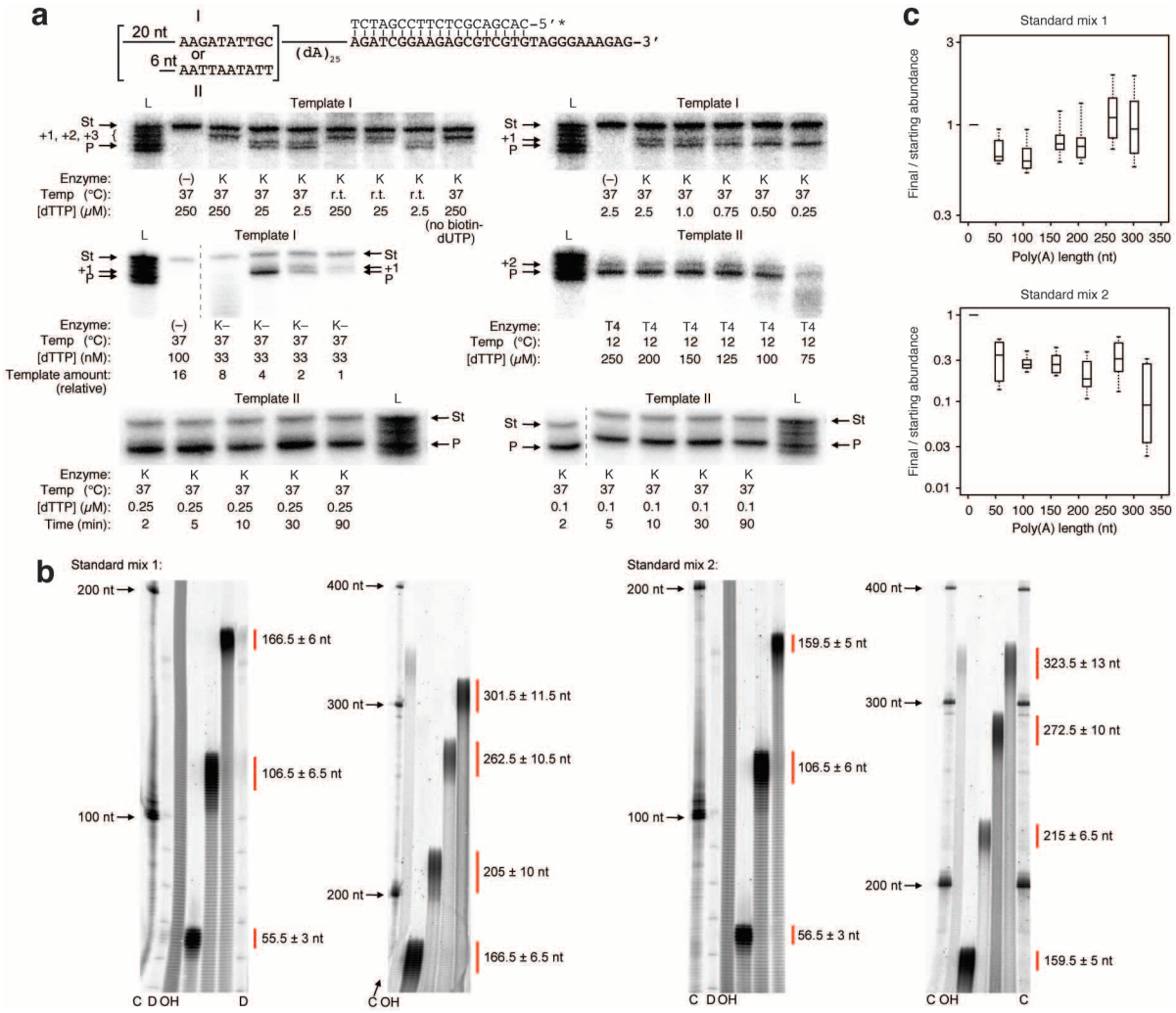
Poly(A)-tail measurements on RNA blots. Single-gene poly(A)-tail lengths were measured on RNA blots after directed RNase H cleavage of the interrogated mRNA. Standard methods[43] were modified to enable higher resolution for shorter tails (<50 nucleotides), such as those found on yeast mRNAs. Total RNA (3–20 µg) was heat-denatured for 5 min at 65 °C in the presence or absence of a $(dT)_{18}$ oligonucleotide (IDT, 33 pmol oligonucleotide µg^{-1} total RNA), and in the presence of 25 pmol of a DNA oligonucleotide (or gapmer oligonucleotide, which had 16 DNA nucleotides flanked on each side by five 2′-*O*-methyl RNA nucleotides) that was complementary to a segment within the 3′-terminal region of the interrogated mRNA. After snap-cooling on ice, the RNA was treated with RNase H (Invitrogen) for 30 min at 37 °C in a 20 µl reaction according to the manufacturer's instructions. The reaction was stopped by addition of gel loading buffer (95% formamide, 18 mM EDTA, 0.025% SDS, dyes) and then analysed on RNA blots resembling those used for small-RNA detection[44] (detailed RNA blot protocol available at http://bartellab.wi.mit.edu/protocols.html). Briefly, after separation of the RNA on a denaturing polyacrylamide gel and transfer onto a Hybond-NX membrane (GE Healthcare), the blot was treated with EDC (*N*-(3-dimethylaminopropyl)-*N*′-ethylcarbodiimide; Sigma-Aldrich), which crosslinked the 5′ phosphate of the 3′-terminal RNase H cleavage product to the membrane[45]. The blot was then hybridized to a probe designed to pair to the region spanning the RNase H cleavage site and the poly(A) site. Comparison of these 3′-terminal fragments with and without poly(A) tails revealed the length of the tails.

39. Jan, C. H., Friedman, R. C., Ruby, J. G. & Bartel, D. P. Formation, regulation and evolution of *Caenorhabditis elegans* 3′UTRs. *Nature* **469,** 97–101 (2011).
40. Ulitsky, I. *et al.* Extensive alternative polyadenylation during zebrafish development. *Genome Res.* **22,** 2054–2066 (2012).
41. Schürer, H., Lang, K., Schuster, J. & Morl, M. A universal method to produce in vitro transcripts with homogeneous 3′ ends. *Nucleic Acids Res.* **30,** e56 (2002).
42. Ruby, J. G., Jan, C. H. & Bartel, D. P. Intronic microRNA precursors that bypass Drosha processing. *Nature* **448,** 83–86 (2007).
43. Sallés, F. J., Richards, W. G. & Strickland, S. Assaying the polyadenylation state of mRNAs. *Methods* **17,** 38–45 (1999).
44. Lau, N. C., Lim, L. P., Weinstein, E. G. & Bartel, D. P. An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. *Science* **294,** 858–862 (2001).
45. Pall, G. S., Codony-Servat, C., Byrne, J., Ritchie, L. & Hamilton, A. Carbodiimide-mediated cross-linking of RNA to nylon membranes improves the detection of siRNA, miRNA and piRNA by northern blot. *Nucleic Acids Res.* **35,** e60 (2007).

©2014 Macmillan Publishers Limited. All rights reserved

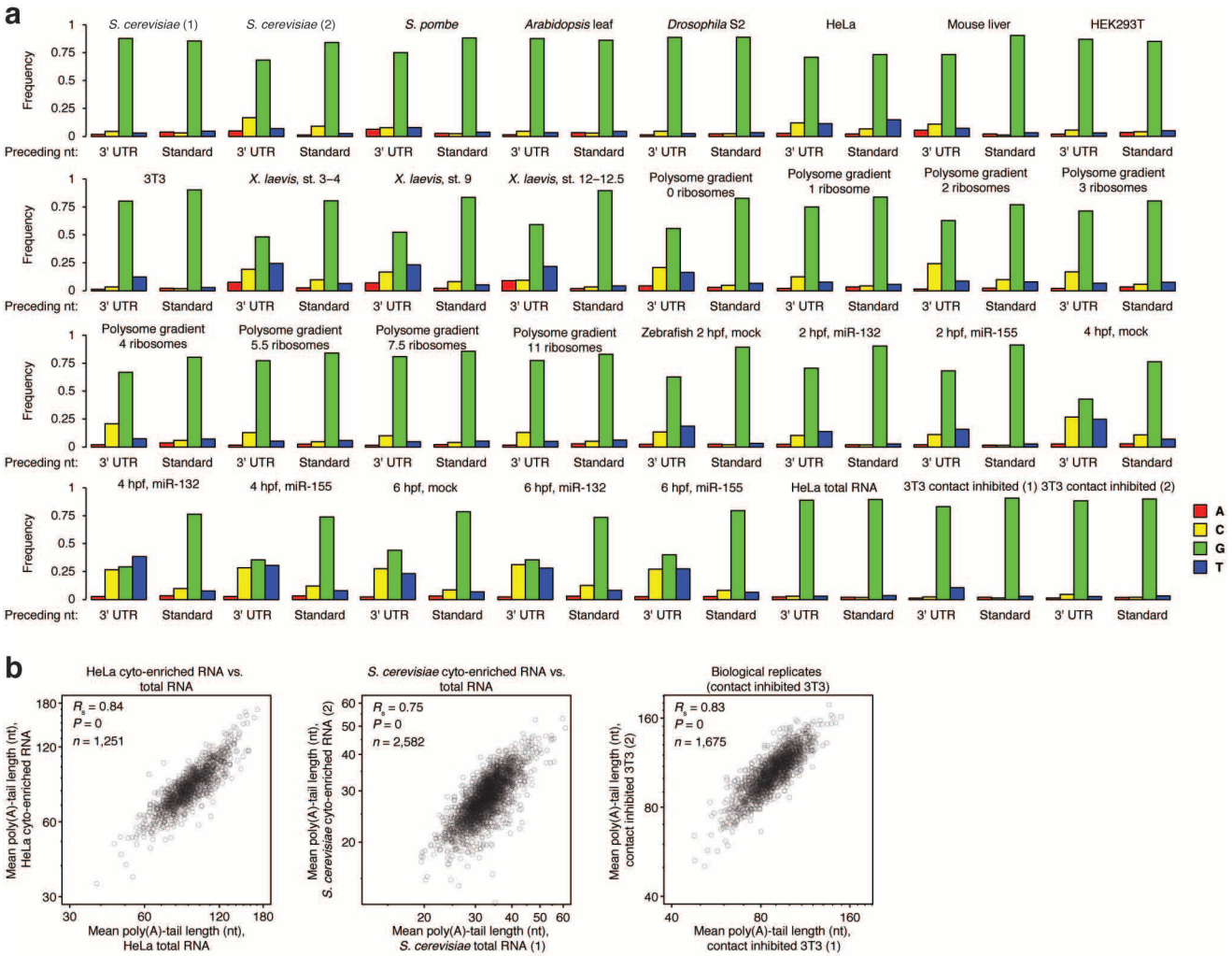
46. Meijer, H. A. *et al.* A novel method for poly(A) fractionation reveals a large population of mRNAs with a short poly(A) tail in mammalian cells. *Nucleic Acids Res.* **35,** e132 (2007).
47. Djebali, S. *et al.* Landscape of transcription in human cells. *Nature* **489,** 101–108 (2012).
48. Holstege, F. C. *et al.* Dissecting the regulatory circuitry of a eukaryotic genome. *Cell* **95,** 717–728 (1998).
49. Wang, Y. *et al.* Precision and functional specificity in mRNA decay. *Proc. Natl Acad. Sci. USA* **99,** 5860–5865 (2002).
50. Grigull, J., Mnaimneh, S., Pootoolal, J., Robinson, M. D. & Hughes, T. R. Genome-wide analysis of mRNA stability using transcription inhibitors and microarrays reveals posttranscriptional control of ribosome biogenesis factors. *Mol. Cell. Biol.* **24,** 5534–5547 (2004).
51. Shalem, O. *et al.* Transient transcriptional responses to stress are generated by opposing effects of mRNA production and degradation. *Mol. Syst. Biol.* **4,** 223 (2008).
52. Munchel, S. E., Shultzaberger, R. K., Takizawa, N. & Weis, K. Dynamic profiling of mRNA turnover reveals gene-specific and system-wide regulation of mRNA decay. *Mol. Biol. Cell* **22,** 2787–2795 (2011).
53. Sun, M. *et al.* Comparative dynamic transcriptome analysis (cDTA) reveals mutual feedback between mRNA synthesis and degradation. *Genome Res.* **22,** 1350–1359 (2012).
54. Haimovich, G. *et al.* Gene expression is circular: factors for mRNA degradation also foster mRNA synthesis. *Cell* **153,** 1000–1011 (2013).
55. Sun, M. *et al.* Global analysis of eukaryotic mRNA degradation reveals Xrn1-dependent buffering of transcript levels. *Mol. Cell* **52,** 52–62 (2013).
56. Larsson, E., Sander, C. & Marks, D. mRNA turnover rate limits siRNA and microRNA efficacy. *Mol. Syst. Biol.* **6,** 433 (2010).
57. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102,** 15545–15550 (2005).

©2014 **Macmillan Publishers Limited. All rights reserved**



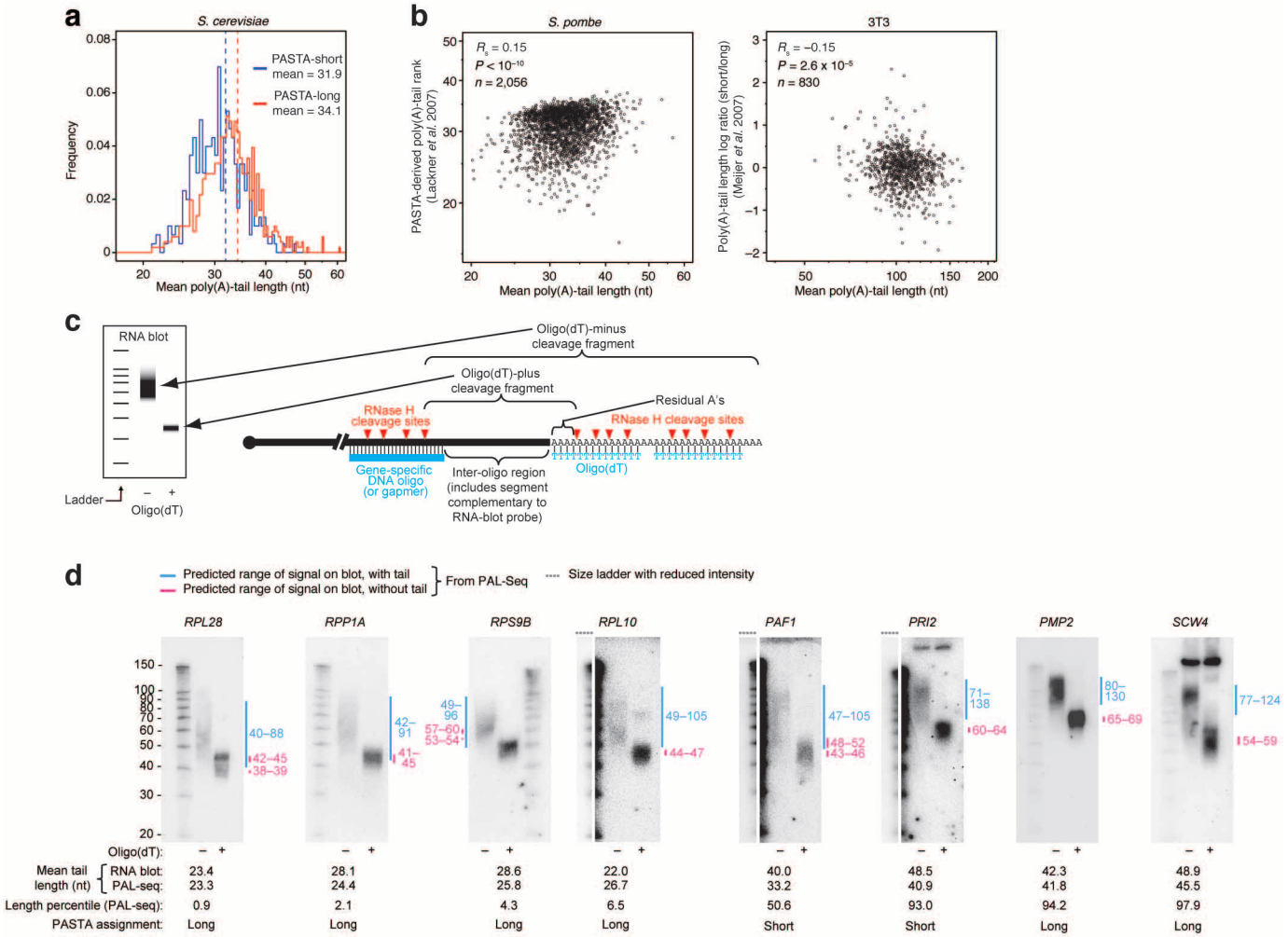
Extended Data Figure 1 | Development and characterization of the PAL-seq method. a, Optimization of the primer-extension reaction. A 5′-radiolabelled primer was annealed to a single-stranded DNA template containing a $(dA)_{25}$ tract immediately upstream of the primer-binding site (top schematic). Two templates (I and II), which differed at the segment immediately 5′ of the poly(dA) tract, were used in the experiments shown. Primer extension was performed with either Klenow fragment (K, NEB), Klenow fragment lacking 3′-to-5′ exonuclease activity (K−, NEB), or T4 DNA polymerase (T4, NEB). Reactions contained the recommended buffer and enzyme concentrations and a 50:1 molar mixture of dTTP:biotin-16-dUTP at the dTTP concentrations indicated. In one experiment (centre left), the dTTP concentration was kept constant, and the concentration of the primer-template duplex was varied instead. Reactions were incubated for 5 min, unless stated otherwise (bottom two panels), at the indicated temperature (temp; room temperature, r.t.), then stopped and in most cases supplemented with a gel-mobility standard (St), which was a ^{32}P-labelled synthetic oligonucleotide that had four extra dT residues appended to the intended full-length primer-extension product (P). Products were resolved on denaturing polyacrylamide gels, alongside a size ladder (L), which was a mixture of ^{32}P-labelled oligonucleotides that differed from the full-length primer-extension product by −1, 0, +1, +2, +3 and +4 dTs (three of these are indicated as +1, +2, +3), and visualized using a phosphorimager. Full-length extension without additional untemplated nucleotides was favoured by using Klenow fragment at 37 °C with very low dTTP concentrations (upper right panel and bottom two panels). Under these conditions the product did not change with prolonged reaction times (bottom). **b**, Poly(A)-tail lengths of the synthetic standards. Poly(A) tails >10 nucleotides retained some length heterogeneity generated during their enzymatic synthesis. To determine the actual poly(A) lengths of the barcode-poly(A) RNAs used to generate the standards, each RNA was ^{33}P-labelled at its 5′ terminus and analysed on denaturing polyacrylamide gels under conditions that enabled single-nucleotide resolution. The values to the right of each panel indicate the modes and approximate ranges of the poly(A) tail lengths (after accounting for the 10-nucleotide barcodes). Also shown are marker lanes with ^{33}P-labelled Century Plus ladder (C, Ambion), ^{33}P-labelled Decade ladder (D, Ambion) and a partial base-hydrolysis ladder of the labelled barcode-poly(A) RNA used to make the 324-nucleotide standard of mix 2 (OH). **c**, The relative PAL-seq yield of each poly(A)-length standard. For each standard in the indicated mix, the yield of poly(A) tags relative to that of the A_{10} standard is plotted, after normalizing to the starting ratio determined from analysis of 5′-labelled mix on a denaturing polyacrylamide gel. Box plots show the distribution of yields for 32 PAL-seq libraries (line, median; box, 25th and 75th percentiles; whiskers, 10th and 90th percentiles).

©2014 Macmillan Publishers Limited. All rights reserved



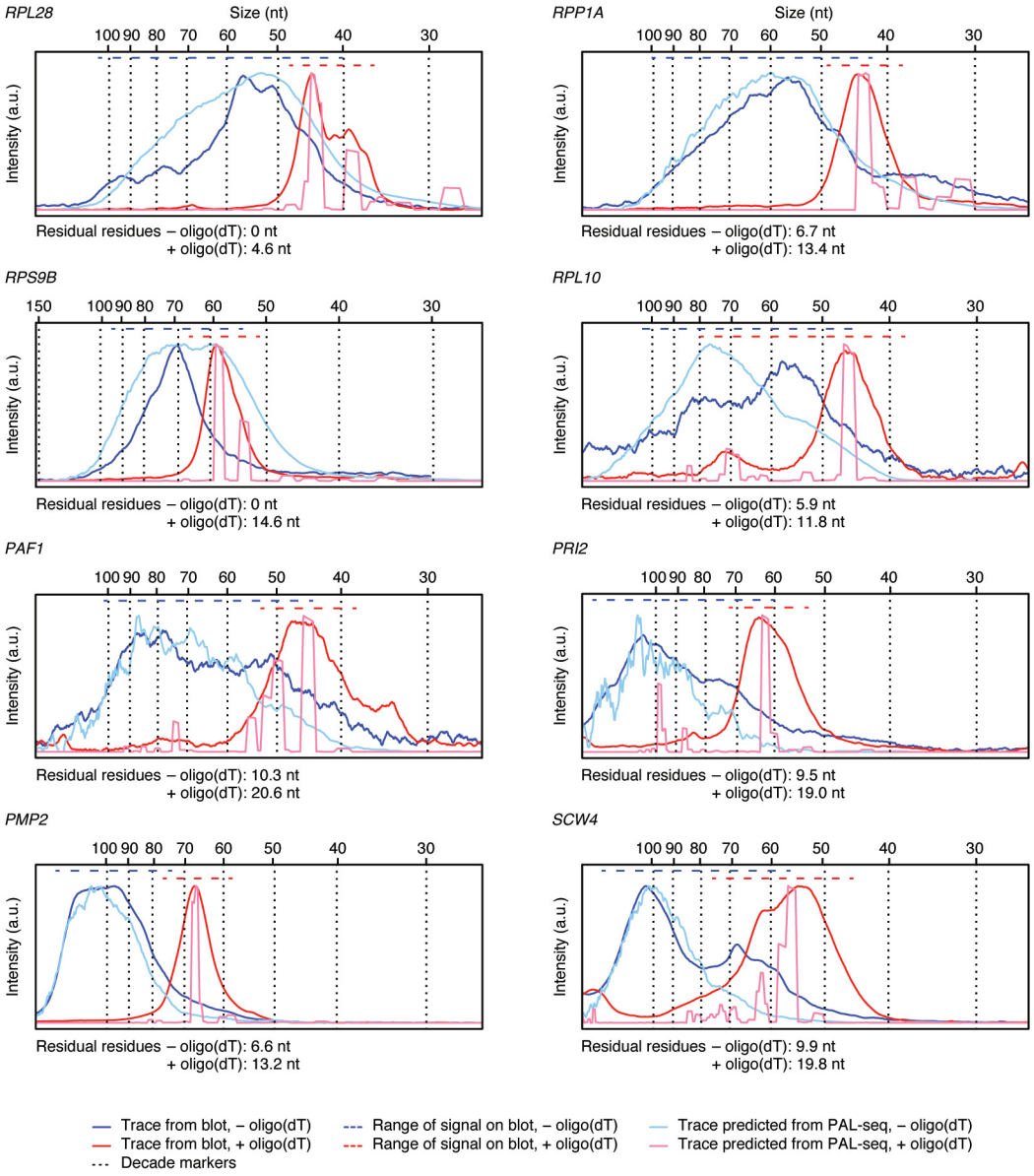
Extended Data Figure 2 | Validation of PAL-seq performance. **a**, Evidence against non-specific RNA degradation. Plotted are nucleotide identities at the positions immediately upstream of poly(A) tags that both mapped uniquely to the genome (or standards) and ranged from 22–30 nucleotides in length (a range chosen to be long enough to enable mostly unique mapping to the genome, yet short enough to include enough 5′ adaptor nucleotides in a 36-nucleotide read to clearly identify the 5′ end of the tag). Frequencies were normalized to the aggregate nucleotide composition of positions 23–31 in either uniquely genome- or standard-mapping tags that extended the full length of the reads (36 nucleotides). Because RNase T1 cuts after Gs, the nucleotide preceding each 22–30-nucleotide tag was expected to be G, unless the mRNA had been cut for some other reason. The high frequency of G indicated that most mRNA fragments had not been cut for other reasons, which also implied that for these samples the poly(A) tails had also remained intact. We are unable to explain the high signal for an upstream U or C in some samples. Nonetheless, the frequency of an upstream A was low, which indicated that there had been little cleavage after As, again implying that the poly(A) tails had remained intact. In the *A. thaliana* leaf analysis, for which the raw reads had the first base removed, estimation of RNA integrity was performed with length ranges shortened by one nucleotide (for example, informative poly(A) tags were 21–29 nucleotides long). **b**, Consistent results from similar samples or biological replicates. Plotted are the relationships between average poly(A)-tail lengths generated using HeLa total RNA or RNA from a cytoplasmically enriched lysate (left), between average poly(A)-tail lengths generated using *S. cerevisiae* total RNA or RNA from a cytoplasmically enriched lysate (sample 1 and 2, respectively; middle), and between average poly(A)-tail lengths generated using cytoplasmically enriched lysates from two different 3T3 cell lines (right). Although the 3T3 lines were each engineered to express a miRNA (either miR-1 or miR-155), the miRNA was not induced in the cells used for this comparison. *NM_001007026* fell outside the plot for HeLa, and *YDL080C* fell outside the plot for yeast.

©2014 Macmillan Publishers Limited. All rights reserved



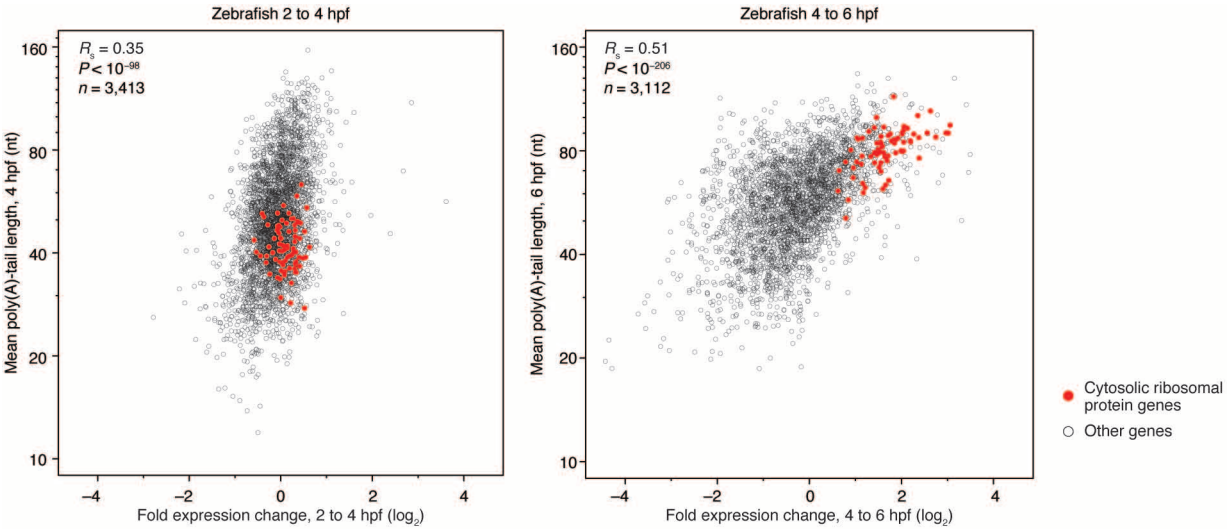
Extended Data Figure 3 | Discrepancies between the results of PAL-seq and those of previous methods. **a**, Comparison of *S. cerevisiae* poly(A)-tail lengths measured by PAL-seq on total RNA to the previous results from PASTA analysis[10]. Plotted are mean poly(A)-tail lengths measured by PAL-seq for genes previously classified as having either short or long tails (PASTA-short and PASTA-long, respectively)[10]. The vertical dashed lines indicate the mean for each group as measured by PAL-seq. **b**, Comparison between PAL-seq measurements and either PASTA-derived poly(A)-tail ranks in fission yeast[11] (left), or results of a related method reporting log ratios of short- and long-tail fractions in actively dividing 3T3 cells[46] (right). **c**, Schematic of tail-length measurements using RNA blots. A DNA oligonucleotide or a gapmer (chimaeric oligonucleotide with DNA flanked by 2′-*O*-methyl-RNA) was designed to pair near the 3′ end of the mRNA. This oligonucleotide directed RNase H cleavage, thereby generating 3′-terminal mRNA fragments with lengths suitable for high-resolution analysis on RNA blots. Some of each sample was also incubated with oligo(dT), which directed RNase H removal of most of the poly(A) tail. Cleavage fragments were resolved on RNA blots and detected by probing for the inter-oligo region of the mRNA. The average poly(A)-tail length was calculated as the difference in the average sizes of the oligo(dT)-minus and oligo(dT)-plus fragments, plus the average number of residual adenosine residues that remained because of incomplete digestion of the poly(A) tail (residual As). For each reaction guided by a gene-specific DNA oligo, the average number of residual adenosines was estimated as half the difference between the known length of the inter-oligo region and the observed length of the oligo(dT)-plus fragment. For the two reactions guided by a gene-specific gapmer (*RPL28* and *RPS9B*), the inter-oligo region extended through the residues pairing to one of the 2′-*O*-methyl-RNA segments, and cleavage was assumed to occur across from the most poly(A)-proximal DNA residue. Thus, the average number of residual adenosines was estimated as the difference between the length of the inter-oligo region and the observed length of the oligo(dT)-plus fragment. **d**, RNA blots used to measure poly(A)-tail lengths, as described in panel **c**, with the length information determined by PAL-seq (on total RNA) and PASTA[10] indicated below each blot for comparison. For each lane, the range of high signal predicted based on PAL-seq results (Extended Data Fig. 4) is shown as a line next to the blot (with and without oligo(dT), red and blue, respectively). These predicted sizes took into account the residual nucleotides flanking the inter-oligo region, using the migration of the oligo(dT)-plus fragment to estimate the residual nucleotides on one or both ends as described in panel **c**. Genes chosen for analysis were required to be adequately expressed and to have a relatively homogeneous cleavage and poly(A) site, as determined by 3P-seq (data not shown). Nonetheless, some genes, such as *RPL28*, had frequently used alternative cleavage and poly(A) sites, as reflected by the two ranges marked in red. A preference was also given to ribosomal protein genes and genes with contradictory poly(A)-tail lengths when comparing the results of PAL-seq and PASTA.

©2014 Macmillan Publishers Limited. All rights reserved



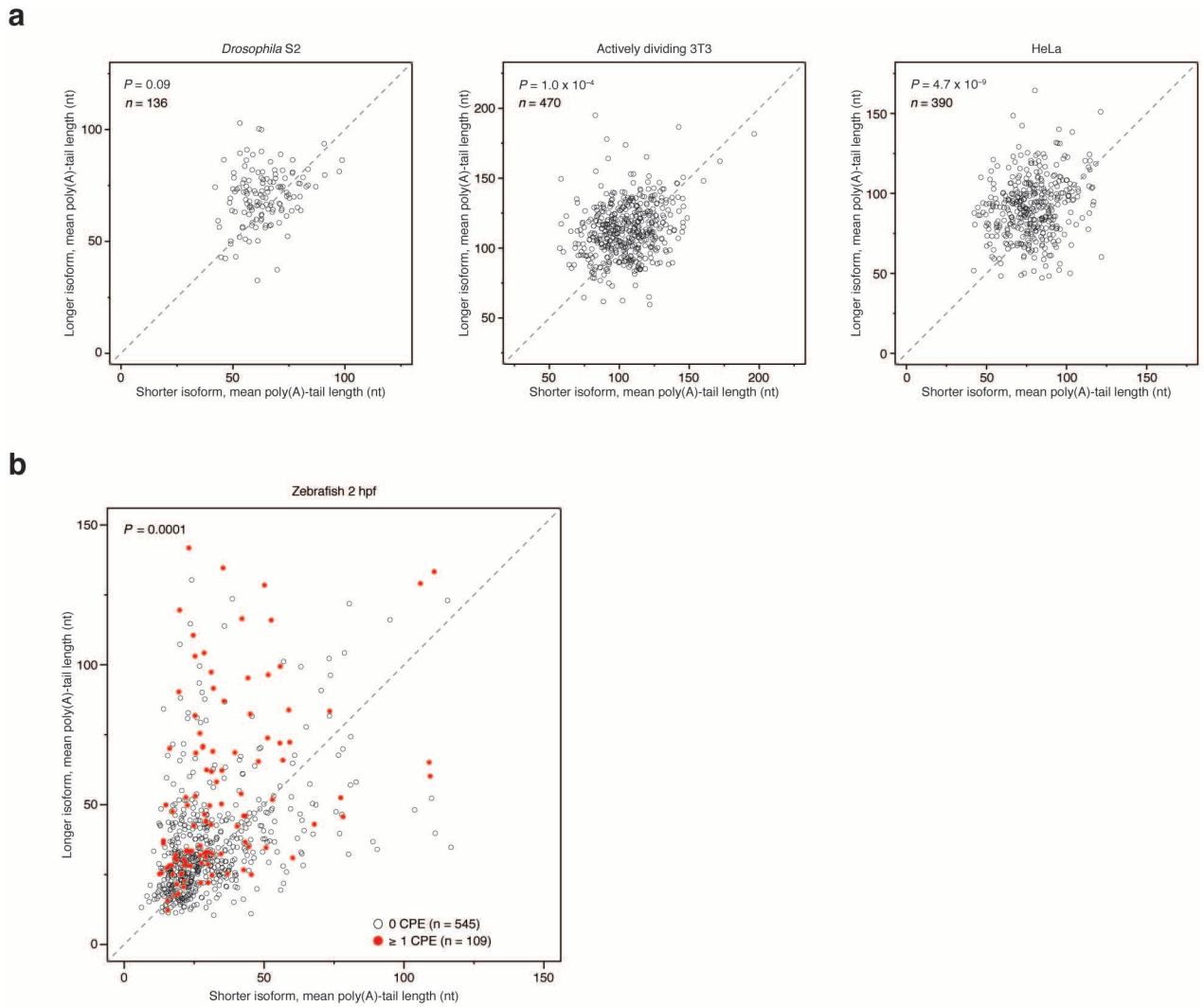
Extended Data Figure 4 | The signal distributions for the RNA blots (Extended Data Fig. 3d) compared with those predicted using PAL-seq. Predicted traces from PAL-seq accounted for the estimated number of residual nucleotides flanking the inter-oligo region after RNase H cleavage, as described (Extended Data Fig. 3c). The offsets added to account for these residual nucleotides are indicated below each plot. The horizontal dashed lines above each plot indicate the range of the signal determined by visual inspection of the RNA blots in Extended Data Fig. 3d (oligo(dT)-plus and minus, red and blue, respectively). Vertical dashed lines indicate the migration of Decade markers (Ambion). The vertical axes are in arbitrary units (a.u.). The range of the high signal predicted based on PAL-seq data (signal exceeding 33% of the maximum) was determined using these plots and shown on Extended Data Fig. 3d as vertical lines next to the RNA blots. For some genes, poly(A)-site heterogeneity caused the signal exceeding 33% to map to noncontiguous segments.

©2014 Macmillan Publishers Limited. All rights reserved



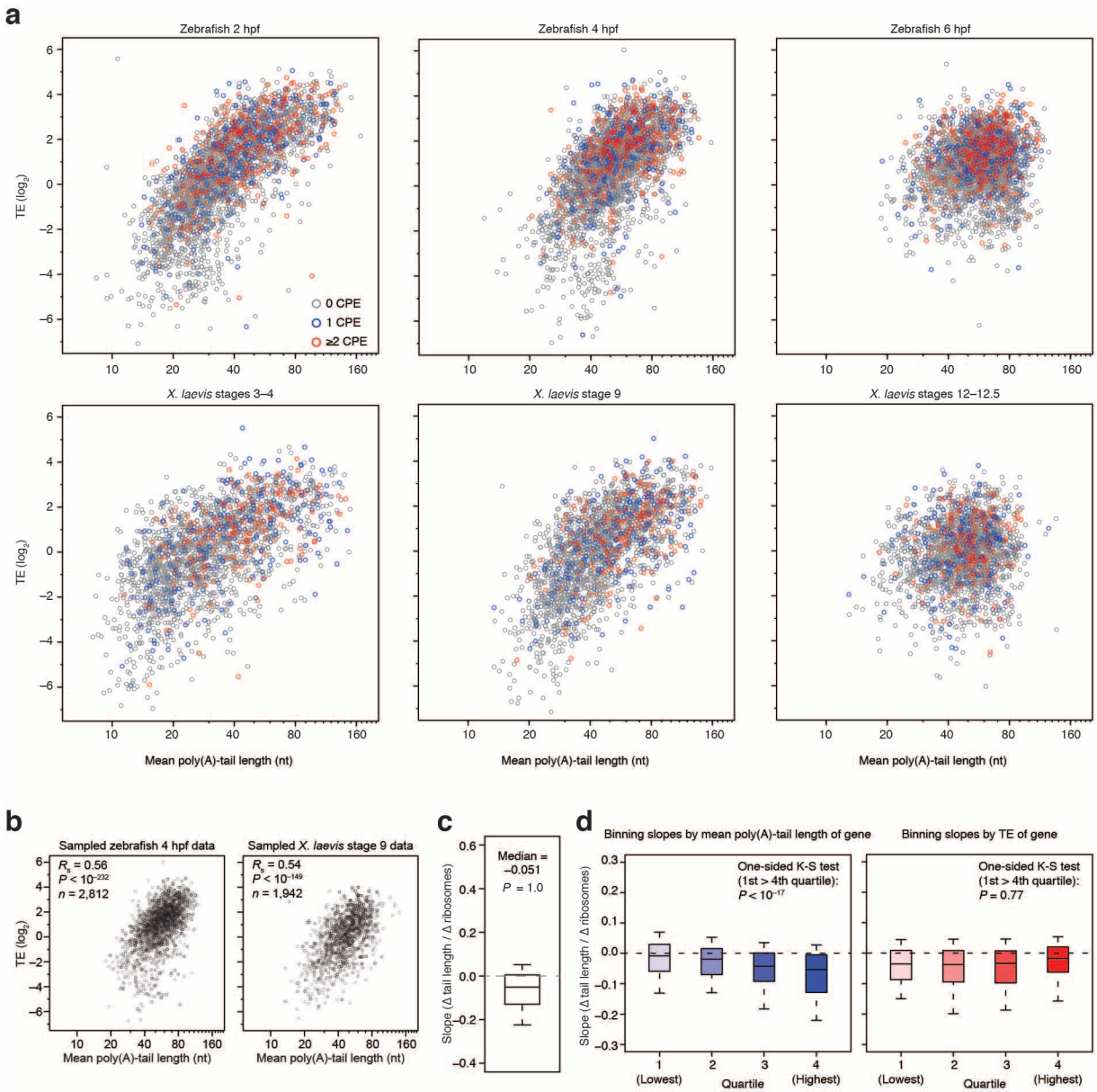
Extended Data Figure 5 | Relationship between poly(A)-tail length and changes in gene expression during zebrafish embryogenesis. Changes in gene expression between the indicated embryonic stages, as measured by RNA-seq, are plotted in relation to the mean poly(A)-tail length at the latter stage.

©2014 Macmillan Publishers Limited. All rights reserved



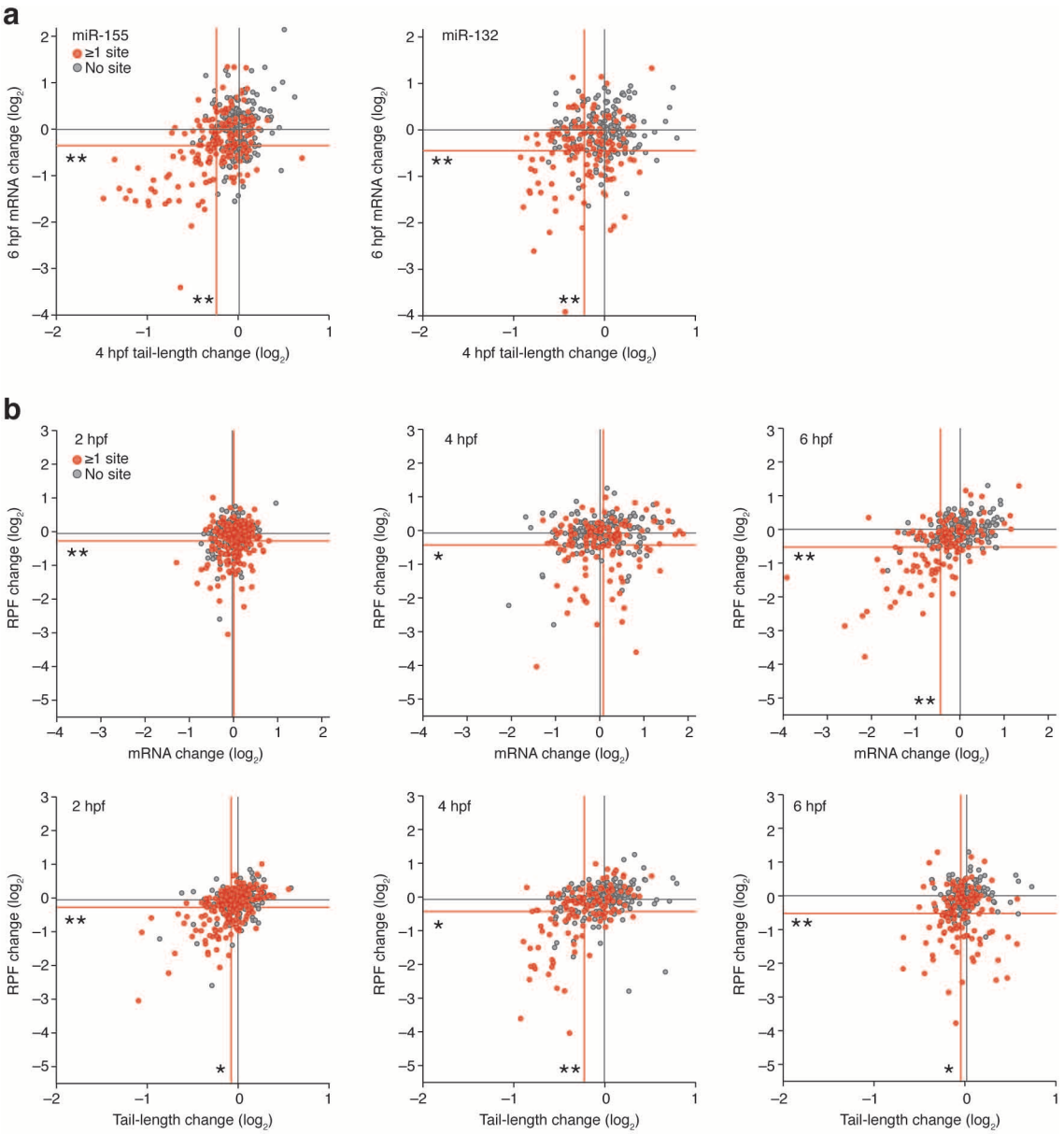
Extended Data Figure 6 | Poly(A)-tail lengths of tandem alternative 3′-UTR isoforms. **a**, Comparison of average poly(A)-tail lengths for proximal (short) and distal (long) isoforms in the indicated cell lines. Results are plotted for isoforms that were each represented by ≥ 25 poly(A) tags and had alternative poly(A) sites ≥ 500 nucleotides apart. For genes with more than one isoform pair meeting these criteria, the pair with poly(A) sites farthest apart was selected. Points for *NM_001007026* and *NM_003913* fell outside the boundaries of the plot for HeLa. *P* values, χ^2 test evaluating whether the relationship between isoform length and tail length differs from that expected by chance. **b**, Average poly(A)-tail lengths for proximal and distal 3′-UTR isoforms in 2 hpf zebrafish embryos, comparing results for genes that either contain (red circles), or do not contain (open circles), a CPE anywhere within the region unique to the distal isoform. A CPE was defined as U_{12}, permitting a single non-U anywhere within the 12 nucleotides[19]. For a CPE found in the unique region to be counted as present, a canonical poly(A) signal (AAUAAA) also had to exist in the last 30 nucleotides of the distal isoform[19,24]. For each gene with a CPE within the region unique to the distal isoform, five genes with unique distal regions of comparable length (± 10%) but lacking a CPE are also shown. Poly(A) tags from three zebrafish 2 hpf PAL-seq libraries (mock-, miR-132–, and miR-155–injected) were combined before calculating average tail length for each isoform. Tandem isoform pairs with a target site for miR-132 or miR-155 in the region unique to the distal isoform were not considered. Only genes for which both tandem isoforms had ≥ 25 poly(A) tags, and for which the alternative poly(A) sites were 50–500 nucleotides apart, are plotted. For genes for which isoform choice affected inclusion of a CPE, the isoform pair representing that gene was chosen as the two isoforms with the most 5′-proximal poly(A) sites that flanked a CPE and satisfied the above criteria. For the pool of genes from which controls were chosen, two adjacent isoforms were picked randomly. *P* value, Fisher's exact test, comparing genes with a CPE in the unique region to controls.

©2014 Macmillan Publishers Limited. All rights reserved



Extended Data Figure 7 | Relationship between poly(A)-tail length and translational efficiency, classifying genes based on CPE content, tail length or translational efficiency. **a**, The same data as in Fig. 3a, except genes were classified based on whether their 3′ UTR contained no CPE (grey), one CPE (blue), or two or more non-overlapping CPEs (red). **b**, Evidence that the more restricted tail-length range observed at gastrulation did not substantially impact the coupling between tail length and translational efficiency. The zebrafish 4 hpf data from Fig. 3a were sampled with replacement so as to have the same distribution of tail lengths observed at 6 hpf (left). Likewise, the *X. laevis* stage 9 data were sampled with replacement so as to have the same distribution of tail lengths observed at stage 12–12.5 (right). **c**, Box plot as in Fig. 4b for the same set of genes, with slopes calculated omitting data from the fraction without bound ribosomes. **d**, Box plots as in Fig. 4b, creating four equal bins of genes based on either overall mean poly(A)-tail length (left) or translational efficiency (right). The same slopes were used as in Fig. 4b, but considering only genes with a determined translational efficiency value and ≥ 100 poly(A) tags in the actively dividing 3T3 sample.

©2014 Macmillan Publishers Limited. All rights reserved



Extended Data Figure 8 | The influence of miRNAs on ribosomes, mRNA abundance and tails in the early zebrafish embryo. **a**, The relationship between changes in tail length at 4 hpf (as determined by PAL-seq) and changes in mRNA abundance at 6 hpf (as determined by RNA-seq), after injecting miR-155 (left) or miR-132 (right). Changes observed between miRNA- and mock-injected embryos are plotted for predicted miRNA target genes (red, genes with ≥ 1 cognate miRNA site in their 3′ UTR) and control genes (grey, genes that have no cognate miRNA site yet resemble the targets with respect to 3′-UTR length). Lines indicate mean changes for the respective gene sets; statistically significant differences between the gene sets for each of the two parameters are indicated (*$P \leq 0.05$; **$P < 10^{-4}$, one-tailed Kolmogorov–Smirnov test). Because injected miRNAs partially inhibited miR-430–mediated repression, genes with a site complementary to nucleotides 2–7 of miR-430 were not considered. All data were normalized to the median changes observed for the controls. **b**, The relationship between changes in ribosome-protected fragments (RPFs) and changes in mRNA levels (top), and between changes in RPFs and changes in tail lengths (bottom) after injecting miR-132. At 2, 4 and 6 hpf, embryos were analysed using ribosome profiling, RNA-seq and PAL-seq. Plots are as in Fig. 5a.

©2014 Macmillan Publishers Limited. All rights reserved

Extended Data Table 1 | Relationships between poly(A)-tail lengths of orthologous genes in samples from different species (or the same gene, when the samples are from the same species), and relationships between poly(A)-tail length and the indicated mRNA features

Tail length conservation

Samples	*n*	R_s	*P* value
HeLa to HEK293T	1620	0.74	$< 10^{-276}$
HeLa to 3T3	1259	0.46	$< 10^{-67}$
HeLa to mouse liver	1095	0.21	$< 10^{-11}$
HeLa to S2	1087	0.16	$< 10^{-6}$
HeLa to *S. cerevisiae*	671	0.056	0.14
HEK293T to 3T3	1907	0.40	$< 10^{-74}$
HEK293T to mouse liver	1815	0.22	$< 10^{-20}$
HEK293T to S2	1877	0.16	$< 10^{-11}$
HEK293T to *S. cerevisiae*	1221	0.078	0.0063
3T3 to mouse liver	1548	0.37	$< 10^{-51}$
3T3 to S2	1194	0.20	$< 10^{-11}$
3T3 to *S. cerevisiae*	737	0.11	0.0034
mouse liver to S2	1238	–0.068	0.016
mouse liver to *S. cerevisiae*	784	0.0028	0.94
S2 to *S. cerevisiae*	959	0.094	0.0038
S. pombe to *S. cerevisiae*	1379	0.22	$< 10^{-15}$

mRNA expression level

Sample	*n*	R_s	*P* value
S. cerevisiae	3199	–0.44	$< 10^{-149}$
S. pombe	2791	–0.31	$< 10^{-61}$
HeLa[35]	2266	–0.053	0.012
Mouse liver	2484	0.085	2.2×10^{-5}
HEK293T	4509	–0.23	$< 10^{-53}$
3T3	2751	–0.23	$< 10^{-32}$
Zebrafish 2 hpf	3693	–0.13	$< 10^{-15}$
Zebrafish 4 hpf	3413	–0.011	0.53
Zebrafish 6 hpf	3112	0.25	$< 10^{-43}$

3'-UTR length

Sample	*n*	R_s	*P* value
S. cerevisiae	3265	0.063	0.00033
S. pombe	2911	0.24	$< 10^{-37}$
Arabidopsis leaf	1425	0.035	0.18
Drosophila S2	3488	0.28	$< 10^{-61}$
HeLa	2362	0.23	$< 10^{-29}$
Mouse liver	3415	–0.15	$< 10^{-18}$
HEK293T	4773	0.24	$< 10^{-62}$
3T3	2873	0.24	$< 10^{-36}$
Zebrafish 2 hpf	6749	0.17	$< 10^{-41}$
Zebrafish 4 hpf	5692	0.084	$< 10^{-9}$
Zebrafish 6 hpf	4594	–0.035	0.016

ORF length

Sample	*n*	R_s	*P* value
S. cerevisiae	3265	0.022	0.21
S. pombe	2911	0.031	0.093
Arabidopsis leaf	1425	0.28	$< 10^{-27}$
Drosophila S2	3488	0.23	$< 10^{-42}$
HeLa	2362	0.29	$< 10^{-45}$
Mouse liver	3415	–0.016	0.34
HEK293T	4773	0.25	$< 10^{-69}$
3T3	2873	0.36	$< 10^{-90}$
Zebrafish 2 hpf	6749	0.14	$< 10^{-28}$
Zebrafish 4 hpf	5692	0.021	0.11
Zebrafish 6 hpf	4594	0.029	0.046

mRNA length

Sample	*n*	R_s	*P* value
S. cerevisiae	3265	0.039	0.025
S. pombe	2911	0.12	$< 10^{-10}$
Arabidopsis leaf	1425	0.30	$< 10^{-29}$
Drosophila S2	3488	0.31	$< 10^{-77}$
HeLa	2362	0.32	$< 10^{-58}$
Mouse liver	3415	–0.12	$< 10^{-11}$
HEK293T	4773	0.31	$< 10^{-105}$
3T3	2873	0.36	$< 10^{-87}$
Zebrafish 2 hpf	6749	0.19	$< 10^{-52}$
Zebrafish 4 hpf	5692	0.068	$< 10^{-6}$
Zebrafish 6 hpf	4594	0.013	0.37

Splice-site number

Sample	*n*	R_s	*P* value
S. cerevisiae	3265	–0.093	$< 10^{-6}$
S. pombe	2911	0.021	0.25
Arabidopsis leaf	1425	0.39	$< 10^{-51}$
Drosophila S2	3488	0.30	$< 10^{-73}$
HeLa	2362	0.22	$< 10^{-27}$
Mouse liver	3415	0.0097	0.57
HEK293T	4773	0.23	$< 10^{-55}$
3T3	2873	0.31	$< 10^{-65}$
Zebrafish 2 hpf	6749	0.12	$< 10^{-23}$
Zebrafish 4 hpf	5692	0.066	$< 10^{-6}$
Zebrafish 6 hpf	4594	0.077	$< 10^{-6}$

Splice-site density

Sample	*n*	R_s	*P* value
S. cerevisiae	3265	–0.041	0.018
S. pombe	2911	–0.074	6.0×10^{-5}
Arabidopsis leaf	1425	0.24	$< 10^{-18}$
Drosophila S2	3488	–0.0021	0.90
HeLa	2362	–0.098	1.8×10^{-6}
Mouse liver	3415	0.13	$< 10^{-13}$
HEK293T	4773	–0.079	$< 10^{-7}$
3T3	2873	–0.055	0.0030
Zebrafish 2 hpf	6749	–0.062	$< 10^{-6}$
Zebrafish 4 hpf	5692	0.0070	0.60
Zebrafish 6 hpf	4594	0.086	$< 10^{-8}$

mRNA nuclear-to-cytoplasmic ratio

Sample	*n*	R_s	*P* value
HeLa[47]	2340	0.28	$< 10^{-43}$

mRNA half-life

Sample	*n*	R_s	*P* value
S. cerevisiae[48]	3027	0.048	0.0079
S. cerevisiae[49]	2592	–0.094	1.5×10^{-6}
S. cerevisiae[50]	1802	0.044	0.063
S. cerevisiae[51]	2320	–0.11	$< 10^{-7}$
S. cerevisiae[52]	3168	0.045	0.011
S. cerevisiae[53]	3256	–0.44	$< 10^{-149}$
S. cerevisiae[54]	1096	0.23	$< 10^{-13}$
S. cerevisiae[55]	2824	–0.35	$< 10^{-81}$
HeLa[56]	642	–0.048	0.23
3T3[31]	1780	–0.16	$< 10^{-11}$

When calculating splice-site density, a pseudocount of one was added to the number of splice sites in an mRNA. For the comparisons between poly(A)-tail length and expression level, mRNA abundances were measured by RNA-seq; data for HeLa were from ref. 35. For the relationship between poly(A)-tail length and mRNA nuclear-to-cytoplasmic abundance ratio, measurements of nuclear and cytoplasmic mRNA abundance in HeLa cells were from ref. 47. mRNA half-lives for *S. cerevisiae*, HeLa and 3T3 mRNAs were from refs 48–55, ref. 56 and ref. 31, respectively.

©2014 Macmillan Publishers Limited. All rights reserved

Extended Data Table 2 | Gene ontology (GO) categories enriched in shorter- or longer-tail genes, as determined by gene set enrichment analysis (GSEA)[57]

S. cerevisiae
cytosolic small ribosomal subunit (–4.358, 0)
cytosolic large ribosomal subunit (–4.1807, 0)
oxidoreductase activity (–3.1264, 0)
actin binding (–2.6333, 0.0013)
hydrogen ion transmembrane transporter activity (–2.6072, 0.0014)
protein refolding (–2.5312, 0.0028)
glycolysis (–2.4683, 0.0047)
microsome (–2.347, 0.0106)
proteasome complex (–2.3188, 0.0126)
aminoacyl-tRNA ligase activity (–2.1318, 0.0379)
mating projection tip (–2.1125, 0.0427)
ATP biosynthetic process (–2.1002, 0.0461)
kinetochore (2.28, 0.0369)
RNA Pol II core promoter proximal region sequence-specific DNA binding transcription factor activity involved in positive regulation of transcription (2.6171, 0.0021)
transmembrane transport (2.7216, 0.0007)
sequence-specific DNA binding (3.2069, 0)

S. pombe
cytosolic large ribosomal subunit (–4.8841, 0)
cytosolic small ribosomal subunit (–3.9765, 0)
cytoplasmic translational elongation (–3.5414, 0)
glycolysis (–2.5283, 0.0027)
translation initiation factor activity (–2.3317, 0.0121)
RNA Pol II core promoter proximal region sequence-specific DNA binding (2.8398, 0)

Arabidopsis leaf
structural constituent of ribosome (–2.4958, 0.0188)
lipid binding (–2.4453, 0.0143)
cysteine biosynthetic process (–2.2569, 0.0341)
rRNA modification (–2.2477, 0.0253)
response to absence of light (2.2135, 0.0279)
response to auxin stimulus (2.2338, 0.0263)
abscisic acid mediated signaling pathway (2.2619, 0.0249)
endoplasmic reticulum unfolded protein response (2.2861, 0.0246)
negative regulation of programmed cell death (2.601, 0.0015)
response to chitin (2.6803, 0.0006)
response to mechanical stimulus (2.9409, 0)

Drosophila S2
mitochondrial large ribosomal subunit (–4.7635, 0)
cytosolic small ribosomal subunit (–4.0894, 0)
cytosolic large ribosomal subunit (–3.7405, 0)
mitochondrial small ribosomal subunit (–3.1553, 0)
mitochondrial respiratory chain complex I (–2.9286, 0.0001)
ATP synthesis coupled proton transport (–2.5931, 0.001)
pyruvate metabolic process (–2.2117, 0.024)
ubiquinol-cytochrome-c reductase activity (–2.1951, 0.026)
tRNA binding (–2.1493, 0.0346)
small nuclear ribonucleoprotein complex (–2.1231, 0.038)
proteasome regulatory particle, lid subcomplex (–2.0899, 0.0435)
proteinaceous extracellular matrix (2.5315, 0.0175)

HeLa
cytosolic large ribosomal subunit (–2.7068, 0.0004)
mitochondrial ribosome (–2.6985, 0.0006)
chaperonin-containing T-complex (–2.2595, 0.0143)
proteasome complex (–2.2316, 0.0166)
exosome (RNase complex) (–2.2024, 0.0199)
nitric oxide metabolic process (–2.0896, 0.0399)
cell-cell junction (2.1558, 0.0471)
positive regulation of transcription from RNA Pol II promoter (2.205, 0.0377)
extracellular matrix (2.242, 0.0351)
glycosphingolipid metabolic process (2.2986, 0.025)
melanosome (2.5133, 0.0045)
antigen processing and presentation of exogenous peptide antigen via MHC class II (2.5749, 0.0039)
lysosome (2.9548, 0)

HEK293T
structural constitutent of ribosome (–3.26, 0)
cytosolic large ribosomal subunit (–2.4247, 0.0118)
proteasome complex (–2.2871, 0.0263)
epidermal growth factor receptor signaling pathway (2.3589, 0.0417)
clathrin coat assembly (2.3693, 0.0475)
branched chain family amino acid catabolic process (2.3775, 0.0499)
cell-cell junction (2.7562, 0.0065)

3T3
cytosolic large ribosomal subunit (–3.7571, 0)
proteasome complex (–2.9927, 0.0001)
electron transport chain (–2.6717, 0.0016)
translation initiation factor activity (–2.4636, 0.0052)
cytosolic small ribosomal subunit (–2.3661, 0.0098)
transmembrane receptor protein tyrosine kinase signaling pathway (2.1827, 0.0391)
transcription, DNA-dependent (2.2119, 0.0352)
protein serine/threonine kinase activity (2.2641, 0.0296)
proteinaceous extracellular matrix (2.3521, 0.0178)
collagen (2.6353, 0.0048)

Liver
cellular response to oxidative stress (2.557, 0.0119)
proton transport (2.5639, 0.0163)

zebrafish 2 hpf
ribosome (–3.3771, 0)
ATP synthesis coupled proton transport (–3.1102, 0)
ATP hydrolysis coupled proton transport (–3.0592, 0)
glycolysis (–2.611, 0.0012)
proteasome core complex (–2.3519, 0.0097)
cytochrome-c oxidase activity (–2.2381, 0.0186)
lipid binding (–2.0649, 0.0492)
convergent extension involved in gastrulation (2.1808, 0.0396)
ATP-dependent helicase activity (2.2225, 0.0344)
double-stranded RNA binding (2.2407, 0.0352)
RNA splicing (2.4446, 0.0101)
protein kinase activity (2.5718, 0.0049)

zebrafish 4 hpf
ribosome (–3.2965, 0)
ATP hydrolysis coupled proton transport (–2.5236, 0.0044)
spindle pole (–2.3207, 0.0171)
ATP synthesis coupled proton transport (–2.2413, 0.0297)
glycolysis (–2.1842, 0.0394)
fin regeneration (2.1141, 0.0497)
convergent extension involved in axis elongation (2.1437, 0.0425)
nuclear-transcribed mRNA catabolic process, nonsense-mediated decay (2.1683, 0.0383)
cell-cell adherens junction (2.1891, 0.0351)
double-stranded RNA binding (2.2892, 0.0186)
ATP-dependent helicase activity (2.4816, 0.0042)
nuclear pore (2.6084, 0.0025)
RNA splicing (2.6176, 0.0025)

zebrafish 6 hpf
cell redox homeostasis (–2.2444, 0.0393)
transferase activity, transferring glycosyl groups (–2.1834, 0.0434)
vesicle-mediated transport (–2.132, 0.0489)
ATP-dependent helicase activity (2.2093, 0.0187)
DNA-dependent ATPase activity (2.3181, 0.0094)
cell adhesion (2.3641, 0.0065)
regulation of cell cycle (2.4049, 0.0057)
proteasome complex (2.4547, 0.0042)
nucleosomal DNA binding (2.4653, 0.0042)
structural constituent of ribosome (4.8594, 0)

For each sample, GSEA was performed on genes ranked based on their mean poly(A)-tail length. The normalized enrichment score (NES) and false-discovery rate (Q value) are indicated in parentheses next to each enriched GO category. A negative NES indicates a category enriched in the shorter-tail genes, whereas a positive value indicates a category enriched in the longer-tail genes. Enriched GO categories were manually curated to eliminate redundant or uninformative categories.

©2014 Macmillan Publishers Limited. All rights reserved