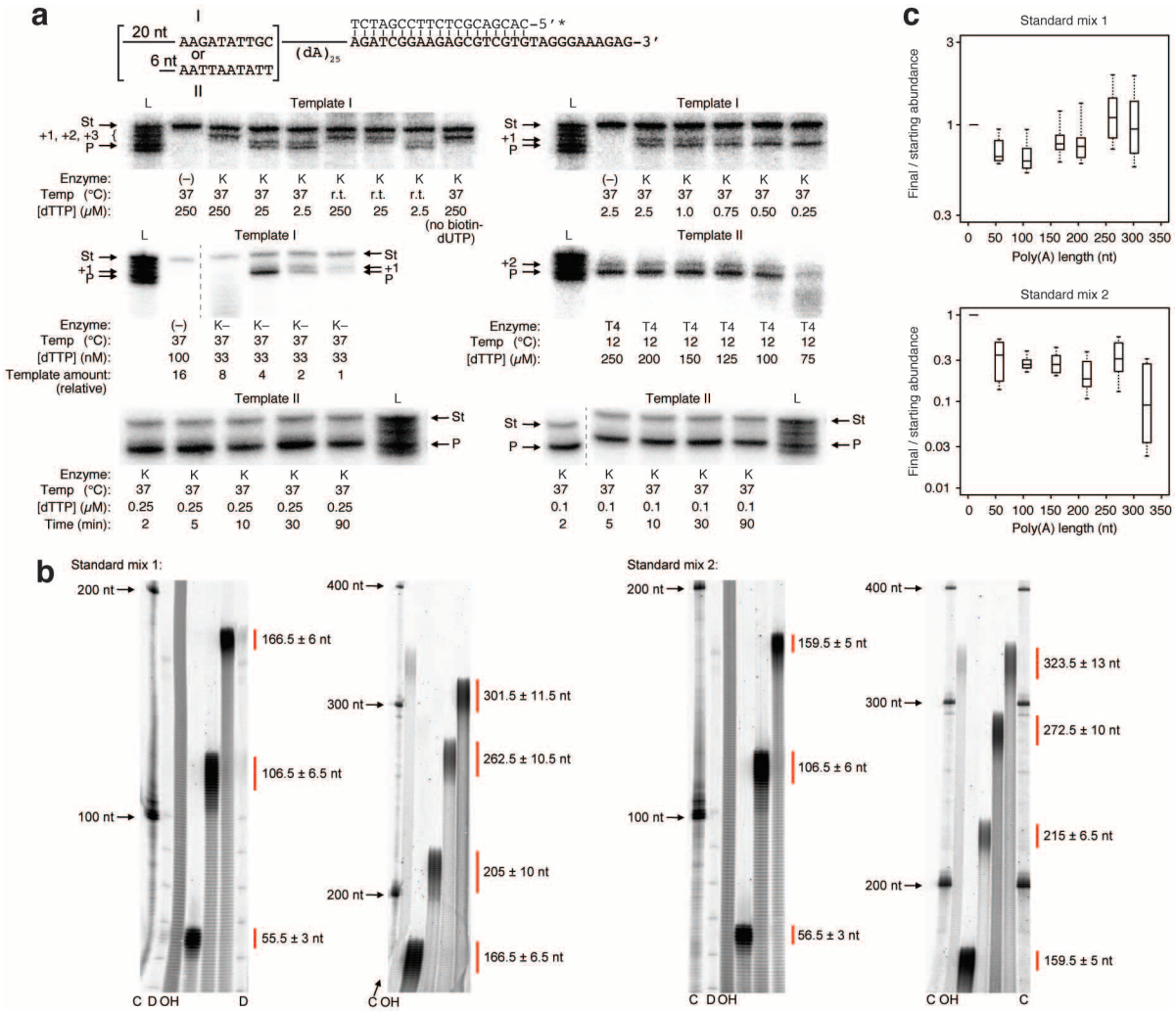
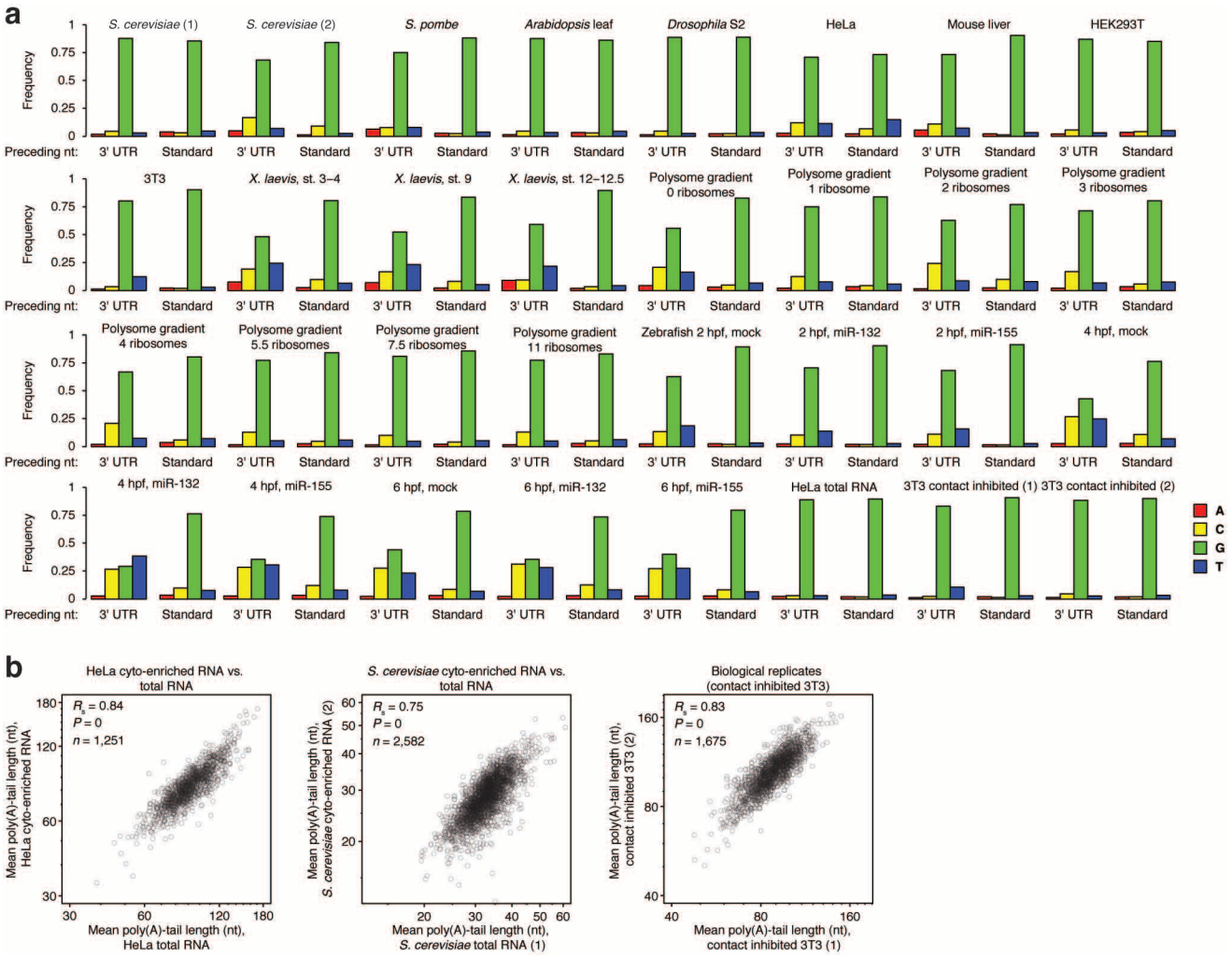
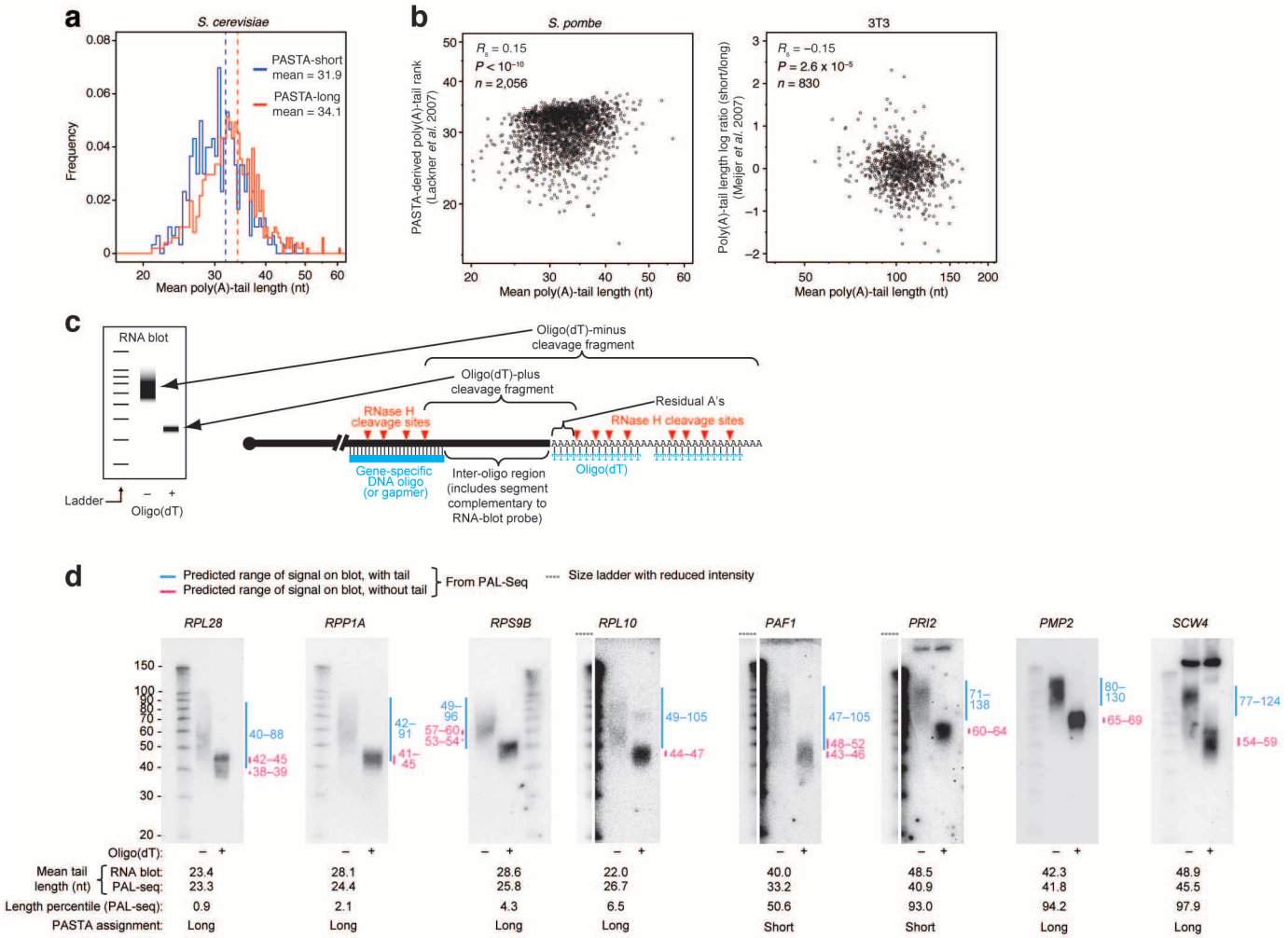


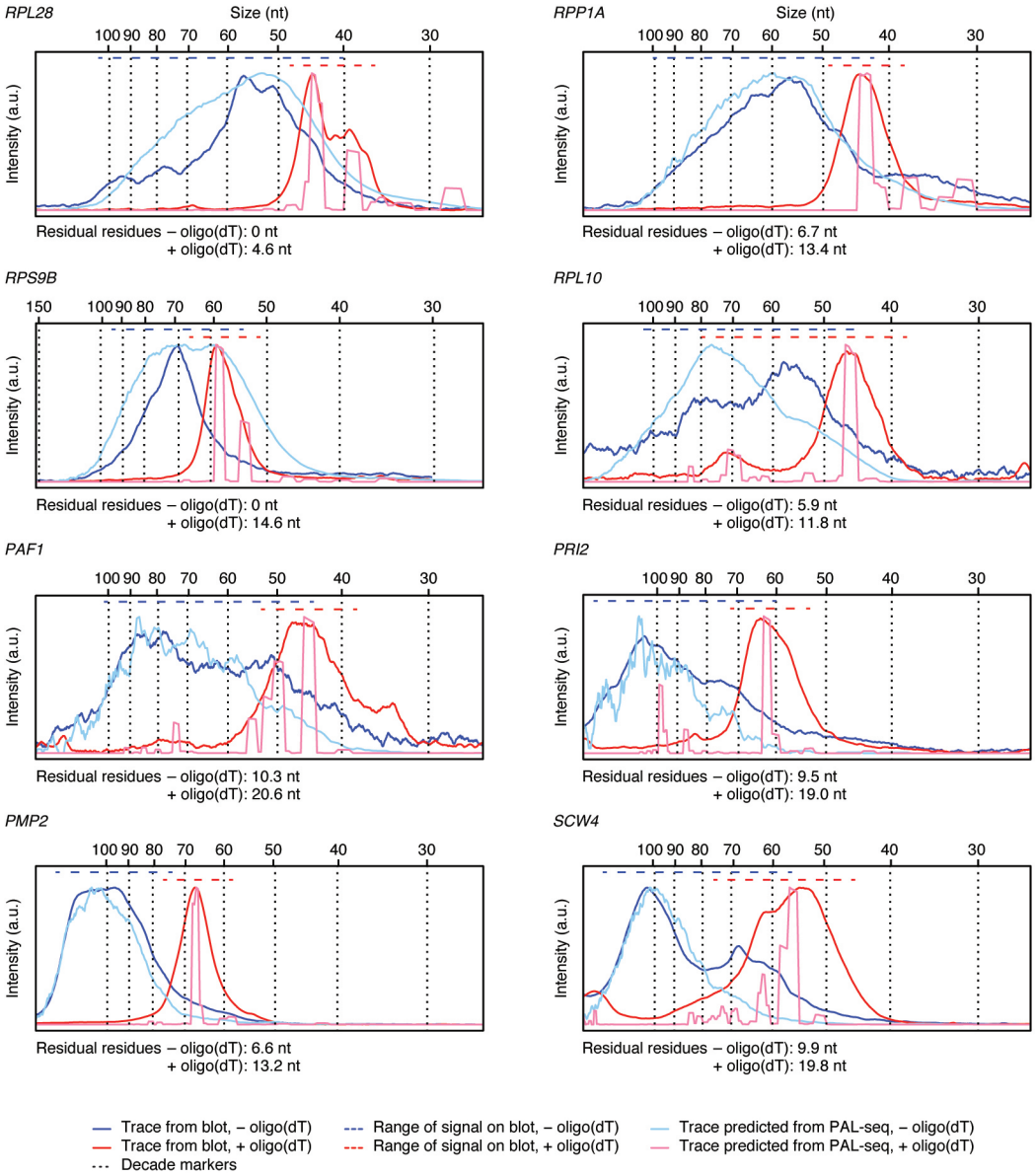
Extended Data Figure 1 | Development and characterization of the PAL-seq method. **a**, Optimization of the primer-extension reaction. A 5′-radiolabelled primer was annealed to a single-stranded DNA template containing a $(dA)_{25}$ tract immediately upstream of the primer-binding site (top schematic). 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).



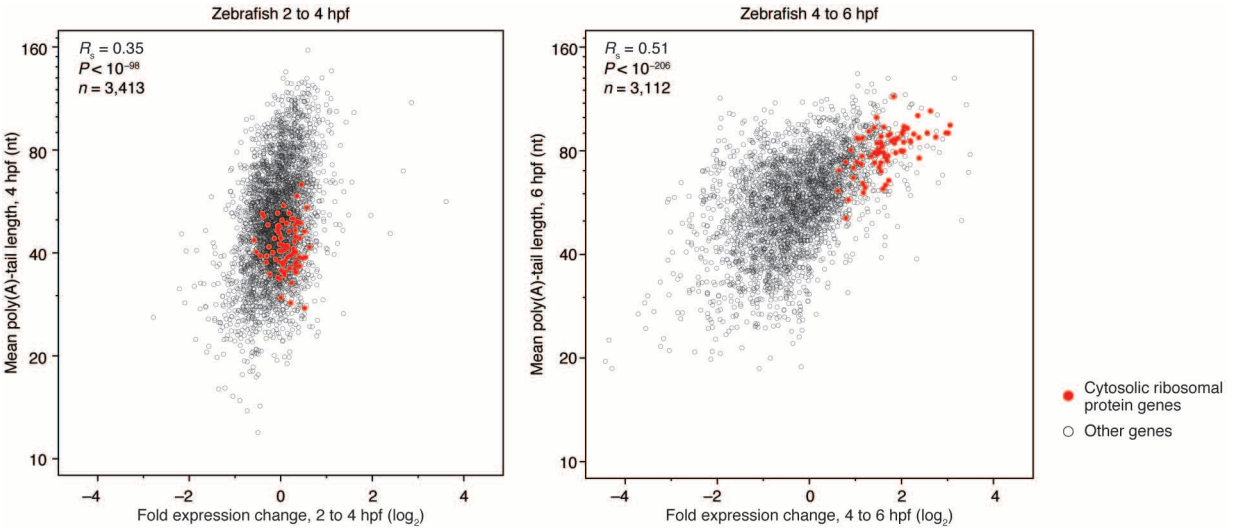
Extended Data Figure 2 | Validation of PAL-seq performance. **a**, Evidence against non-specific RNA degradation. Plotted are nucleotide identities at the positions immediately upstream of poly(A) tags that both mapped uniquely to the genome (or standards) and ranged from 22–30 nucleotides in length (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.



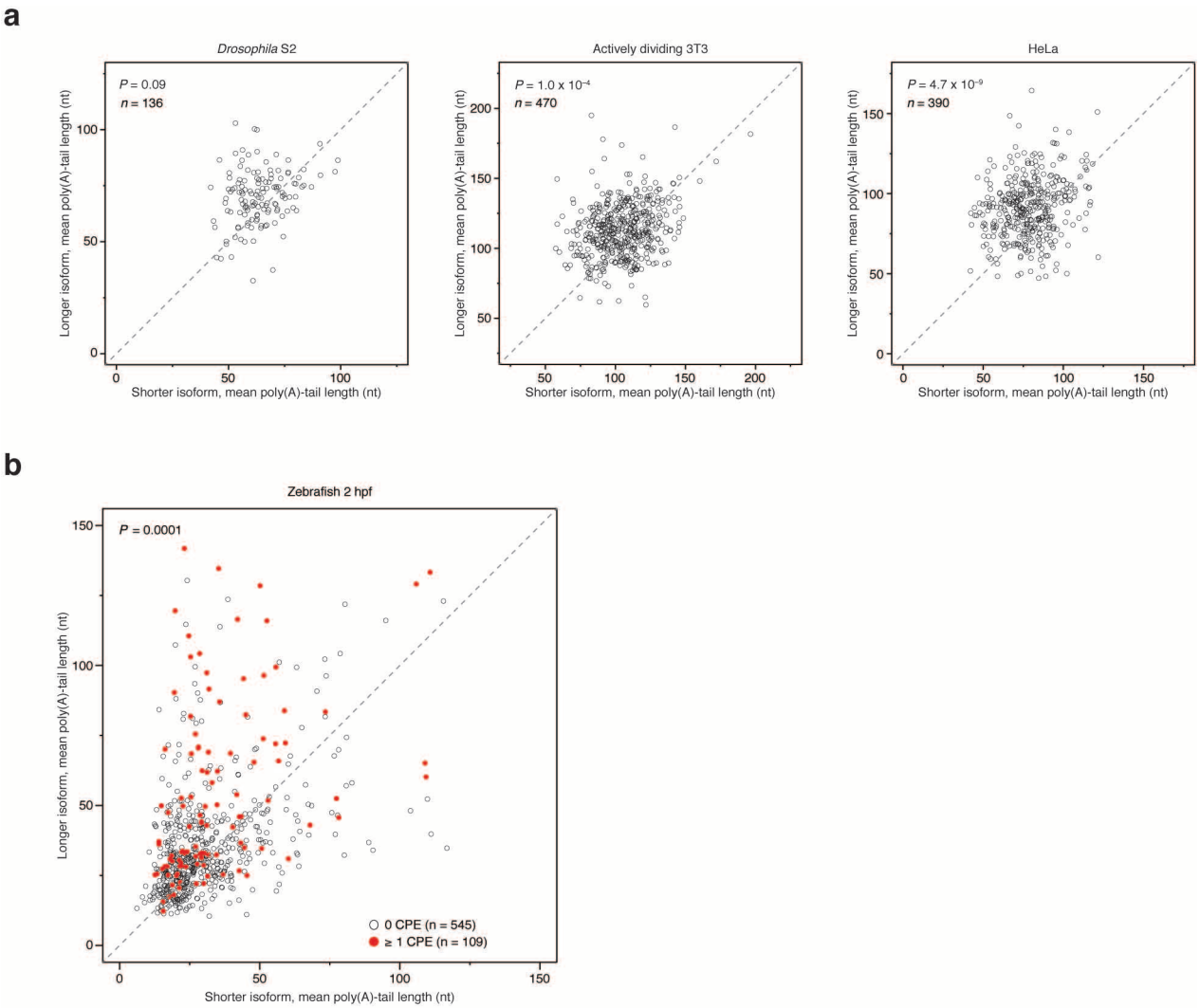
Extended Data Figure 3 | Discrepancies between the results of PAL-seq and those of previous methods. **a**, Comparison of *S. cerevisiae* poly(A)-tail lengths measured by PAL-seq on total RNA to the previous results from PASTA analysis[10]. Plotted are mean poly(A)-tail lengths measured by PAL-seq for 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.



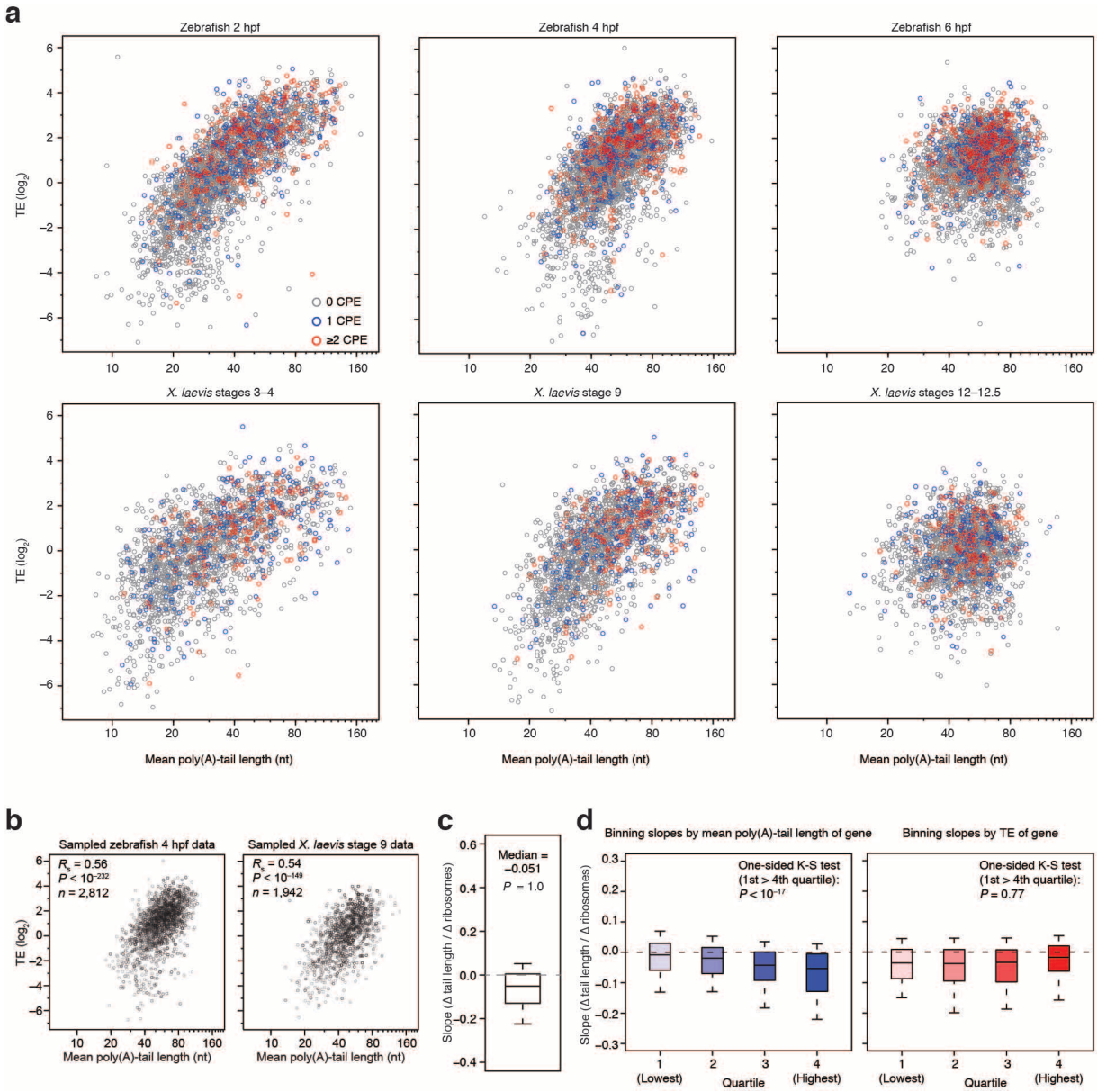
Extended Data Figure 4 | The signal distributions for the RNA blots (Extended Data Fig. 3d) compared with those predicted using PAL-seq. Predicted traces from PAL-seq accounted for the estimated number of residual nucleotides flanking the inter-oligo region after RNase H cleavage, as described (Extended 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.



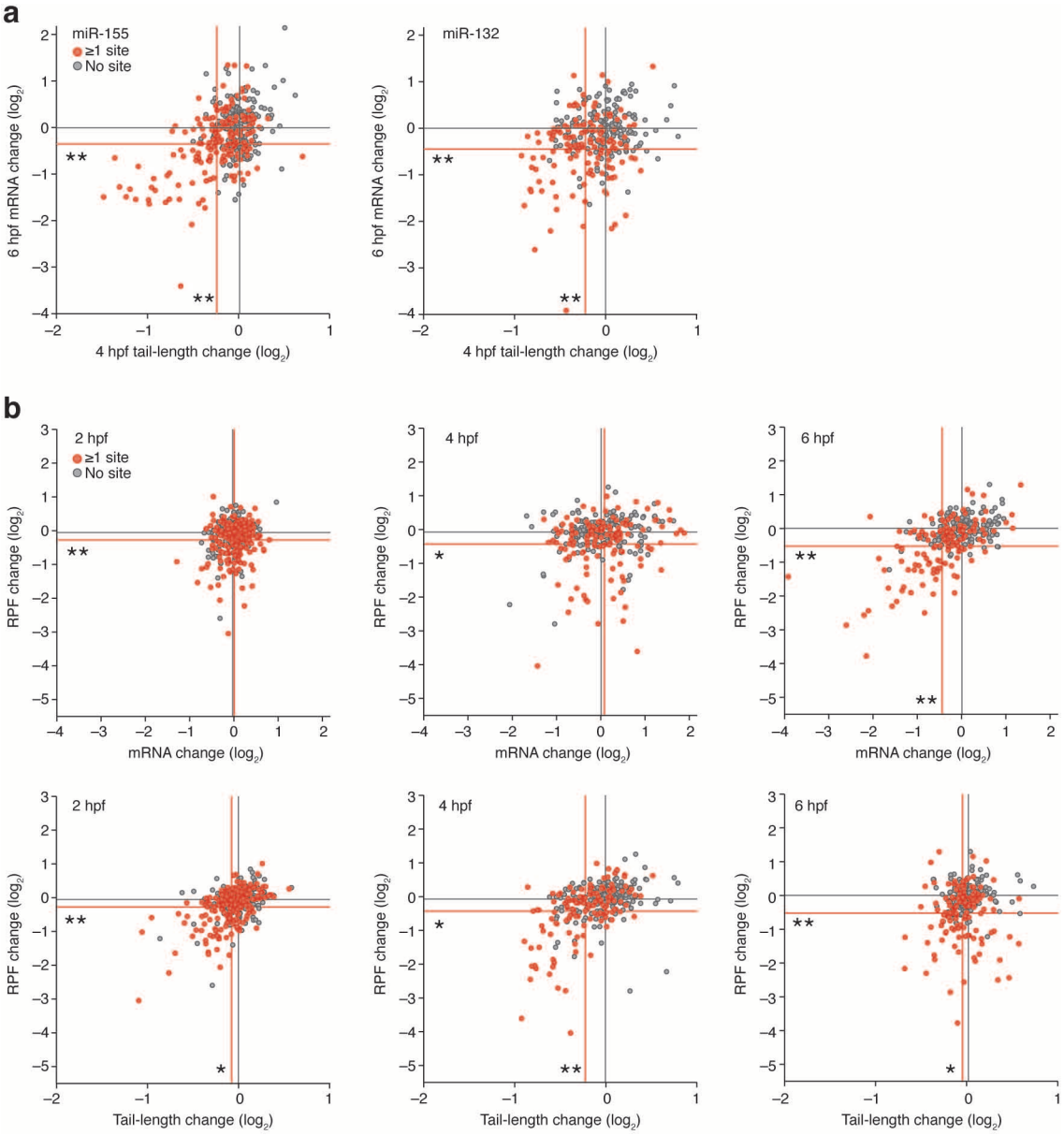
Extended Data Figure 5 | Relationship between poly(A)-tail length and changes in gene expression during zebrafish embryogenesis. Changes in gene expression between the indicated embryonic stages, as measured by RNA-seq, are plotted in relation to the mean poly(A)-tail length at the latter stage.



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.



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.



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.

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.

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.