AMERICAN SOCIETY FOR MICROBIOLOGY | Molecular and Cellular Biology®

RESEARCH ARTICLE

3′ Untranslated Regions Are Modular Entities That Determine Polyadenylation Profiles

Kai Hin Lui,[a] Joseph V. Geisberg,[a] Zarmik Moqtaderi,[a] Kevin Struhl[a]

[a]Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts, USA

ABSTRACT The 3′ ends of eukaryotic mRNAs are generated by cleavage of nascent transcripts followed by polyadenylation, which occurs at numerous sites within 3′ untranslated regions (3′ UTRs) but rarely within coding regions. An individual gene can yield many 3′-mRNA isoforms with distinct half-lives. We dissect the relative contributions of protein-coding sequences (open reading frames [ORFs]) and 3′ UTRs to polyadenylation profiles in yeast. ORF-deleted derivatives often display strongly decreased mRNA levels, indicating that ORFs contribute to overall mRNA stability. Poly(A) profiles, and hence relative isoform half-lives, of most (9 of 10) ORF-deleted derivatives are very similar to their wild-type counterparts. Similarly, in-frame insertion of a large protein-coding fragment between the ORF and 3′ UTR has minimal effect on the poly(A) profile in all 15 cases tested. Last, reciprocal ORF/3′-UTR chimeric genes indicate that the poly(A) profile is determined by the 3′ UTR. Thus, 3′ UTRs are self-contained modular entities sufficient to determine poly(A) profiles and relative 3′-isoform half-lives. In the one atypical instance, ORF deletion causes an upstream shift of poly(A) sites, likely because juxtaposition of an unusually high AT-rich stretch directs polyadenylation closely downstream. This suggests that long AT-rich stretches, which are not encountered until after coding regions, are important for restricting polyadenylation to 3′ UTRs.

KEYWORDS 3′ UTR, 3′-end formation, gene expression, mRNA stability, polyadenylation

Eukaryotic messenger RNAs (mRNAs) exhibit a remarkable degree of physical and functional heterogeneity. Transcriptional and posttranscriptional processing of a protein-coding gene typically gives rise to numerous mRNA isoforms that can vary in exon composition as well as in the 5′ and 3′ untranslated regions (5′ UTRs and 3′ UTRs) that flank the coding sequence (1–4). Varied exon composition plays a critical role in metazoan proteome diversity, while variation in 5′ and 3′ UTRs has important regulatory consequences.

3′ UTRs play critical roles in numerous physiological pathways (5). Alternative polyadenylation within 3′ UTRs yields numerous 3′-mRNA isoforms (1, 6–8) that can differ in many biological properties, including translation, subcellular localization, mRNA stability, and macromolecular complex assembly (5). Sequence elements within 3′ UTRs can stabilize or destabilize mRNAs, thereby differentially affecting 3′-isoform half-lives and steady-state levels (4). Metazoan 3′ UTRs often contain microRNA binding sites that typically reduce mRNA stability and translation in a manner that is isoform specific and physiologically regulated (9). Global upstream shifting of poly(A) sites within 3′ UTRs is a driver of oncogenic transformation (10–12), and broad-scale downstream shifting is associated with certain developmental stages (13–15). Mutations in 3′ UTRs have been associated with a wide variety of malignancies, including immune disorders such as immune dysregulation, polyendocrinopathy, enteropathy, X-linked syndrome (IPEX) and lupus (9).

The 3′ ends of eukaryotic mRNAs are generated during the process of transcriptional elongation by cleavage of the nascent transcript downstream of the coding region, followed by addition of a poly(A) tail (14, 16–18). For a typical yeast gene, this process generates ~50

Address correspondence to Kevin Struhl, kevin@hms.harvard.edu.

The authors declare no conflict of interest.

Received 22 June 2022
Returned for modification 19 July 2022
Accepted 1 August 2022
Published 16 August 2022

mRNA isoforms with poly(A) sites located primarily in a subregion of the 3′ UTR called the end zone (1, 6, 7). Steady-state levels of these 3′ isoforms can range over several orders of magnitude, while 3′-isoform endpoint positions (i.e., proximity to stop codons) can vary by hundreds of nucleotides.

The poly(A) profile of an individual gene is defined by the relative steady-state expression levels of all 3′-mRNA isoforms, and it exhibits little variation across a variety of growth conditions (e.g., osmotic stress, alternate carbon source, and nutrient-poor medium) (19). However, the RNA polymerase (Pol) II elongation rate regulates the poly(A) profile, with slow Pol II mutants shifting poly(A) profiles upstream, closer to the open reading frame (ORF), and fast Pol II mutants causing a downstream shift (19). Both upstream and downstream shifts involve differential utilization of poly(A) sites, not the generation of new sites. Yeast cells undergoing the diauxic response show an upstream shift that is extremely similar to that observed in Pol II mutants, presumably due to reduced Pol II speed under this condition (19).

Paradoxically, and for mechanistically unknown reasons, polyadenylation occurs at numerous sites within the 3′ UTR, but it rarely occurs within the much larger coding regions (6) and introns (20). In yeast, polyadenylation is associated with a degenerate sequence motif consisting of a long AU-rich stretch, followed by short U-rich sequences that flank several A residues immediately downstream of the cleavage site (6, 21). It has been suggested that the long AU-rich stretch is permissive for polyadenylation (perhaps by changing the speed or nature of the Pol II elongation complex) and that selection of specific poly(A) sites reflects intrinsic sequence preferences of the cleavage/polyadenylation machinery (6). However, it is unknown whether the 3′ UTR is an independent domain sufficient to specify the poly(A) profile. In addition, the 3′ UTR plays a key role in the stability of 3′-mRNA isoforms (4), but it is unclear whether protein-coding sequences are also involved.

Here, we dissect the relative contributions of protein-coding sequences and 3′ UTRs to the polyadenylation profile and mRNA stability in yeast. Analysis of deletion, insertion, and chimeric derivatives demonstrates that poly(A) profiles are (with one notable exception) minimally affected by ORFs and other upstream sequences. Thus, 3′ UTRs are self-contained modular entities that are sufficient to determine poly(A) profiles and relative half-lives of 3′ isoforms. At the atypical *BYE1* locus, deletion of the ORF juxtaposes an especially AT-rich region next to the 3′ UTR and results in an upstream shift in the poly(A) profile. ORF-deleted derivatives often cause significant differences in steady-state transcript levels, suggesting that protein-coding sequences contribute to overall mRNA stability in an isoform-independent fashion.

RESULTS

Complete deletion of ORFs generally does not affect polyadenylation profiles. To assess the contribution of ORFs to poly(A) profiles, we used CRISPR to make precise coding sequence deletions at 10 yeast loci (1 ORF deletion/strain) of various lengths (330 to 4,407 nucleotide [nt]; Fig. 1A). Using 3′READS, we mapped poly(A) sites at nucleotide-level resolution for all 10 strains. Biological replicates of both the wild-type (WT) and ORF-deleted loci are very highly correlated on the isoform levels across entire 3′ UTRs (Fig. S1 and Table S1).

The poly(A) profiles of the ORF-deleted (ΔORF) loci and their wild-type counterparts (WT) are remarkably similar (examples in Fig. 2A), with most (6 out of 10) WT versus ΔORF Pearson correlation coefficients being very high ($R > 0.9$; Table S1). In three cases, the WT versus ΔORF correlation is slightly lower ($R = \sim 0.7$), but this is probably due to fewer reads and more error in one of the biological replicates. Overall, the pairwise WT versus ΔORF correlations are slightly lower than the replicate correlations of the wild-type strain (Fig. 3A; $P = 4.8 \times 10^{-4}$, Mann-Whitney U test), but indistinguishable from the ΔORF biological replicates (Fig. 2B; $P > 0.05$, Mann-Whitney U test). Thus, complete removal of the protein-coding sequences tested, even those that comprise > 80% of a mature mRNA, generally has little influence on selection of poly(A) sites, although very subtle effects are sometimes observed. Furthermore, the 3′ UTR is sufficient to generate the poly(A) profile even when the juxtaposed

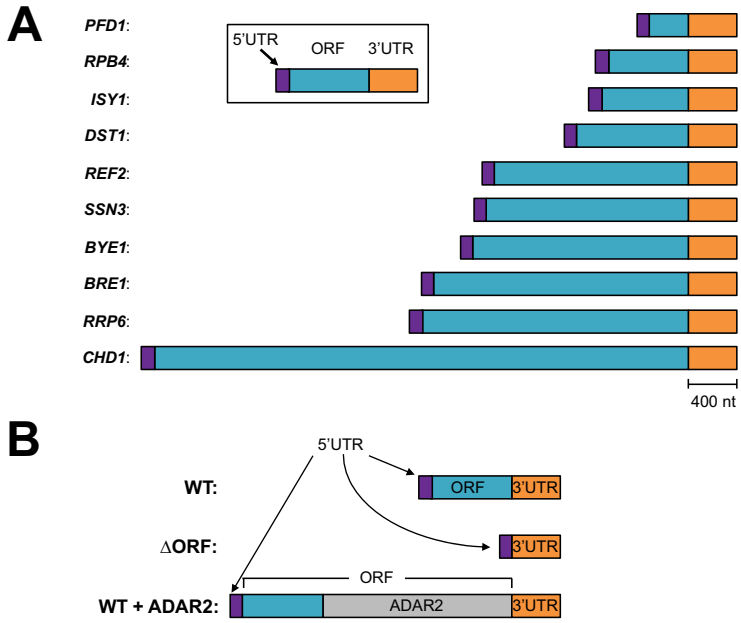


FIG 1 Schematic of wild-type and mutant genes. (A) Diagrams of the 10 wild-type genes used in the analysis of poly(A) profiles upon complete ORF deletion. (B) Diagrams of a representative wild-type (WT), ORF-deletion, and ADAR2-insertion allele. All drawings are to scale with 5′ UTRs (purple), ORFs of widely different lengths (blue or gray for ADAR2), and 3′ UTRs (orange) indicated.

5′ UTRs of the RNAs expressed from the various ΔORF loci are very short (typically 20 to 100 nt).

Deletion of the *BYE1* ORF causes an upstream shift in the polyadenylation profile. In 1 of the 10 genes tested (*BYE1*), the WT versus ΔORF correlation is very low ($R = 0.23$) and far below the correlations of biological replicates for either the WT or *bye1-ΔORF* strain (Fig. 3A). In this exceptional case, polyadenylation occurs at the same sites within the 3′ UTR, but these poly(A) sites are used at different relative frequencies. Specifically, the poly(A) profile in the *bye1-ΔORF* strain is an upstream-shifted version of the wild-type *BYE1* profile (Fig. 3A) reminiscent of upstream shifts caused by Pol II derivatives with slow elongation rates (19). However, the upstream shift in the *bye1-ΔORF* strain is much more dramatic than observed at the wild-type *BYE1* gene in strains with slow Pol II derivatives or under diauxic conditions (Fig. 3A).

The ΔORF alleles not only remove protein-coding sequences, but they also create novel junctions between the 5′ UTRs and 3′ UTRs that could have functional consequences. We therefore considered the possibility that the unusual behavior of the *bye1-ΔORF* allele might be due to juxtaposition of the 5′ UTR to the 3′ UTR. Interestingly, the region between −40 and −1 of the *bye1-ΔORF* allele has an AT content of 82%, whereas the average of all other wt and ΔORF alleles is only 67% (range, 59 to 74%) (Fig. 3B and C). In addition, the coding (top) strand of the −40 to −1 region of the *bye1-ORF* allele has an exceptionally high content of T residues (65%), compared to all other WT and ΔORF alleles (range, 15 to 40%). As high AT-rich stretches occur upstream of poly(A) sites (6), artificial juxtaposition of especially AT-rich stretches upstream of the 3′ UTR may cause the upstream shift in the poly (A) profile.

A large protein-coding insertion upstream of 3′ UTRs does not affect poly(A) profiles. As a complementary approach, we inserted a 1.5-kb region containing the catalytic domain of *Drosophila* ADAR2 (22, 23) in frame between the C-terminal end of the ORF and the beginning of the 3′ UTR (Fig. 1B; 1 target gene per strain, 15 in total). These 15 genes with the large ADAR2 insertion are distinct from the 10 genes used for complete ORF deletions. As expected, poly(A) profiles of biological replicates are highly correlated for both the wild-type and insertion alleles (Fig. 4 and Table S2). The pairwise correlation of wild-type versus insertion derivatives (Fig. 4 and Table S2) is also extremely strong (median, $R = 0.97$; range, 0.80

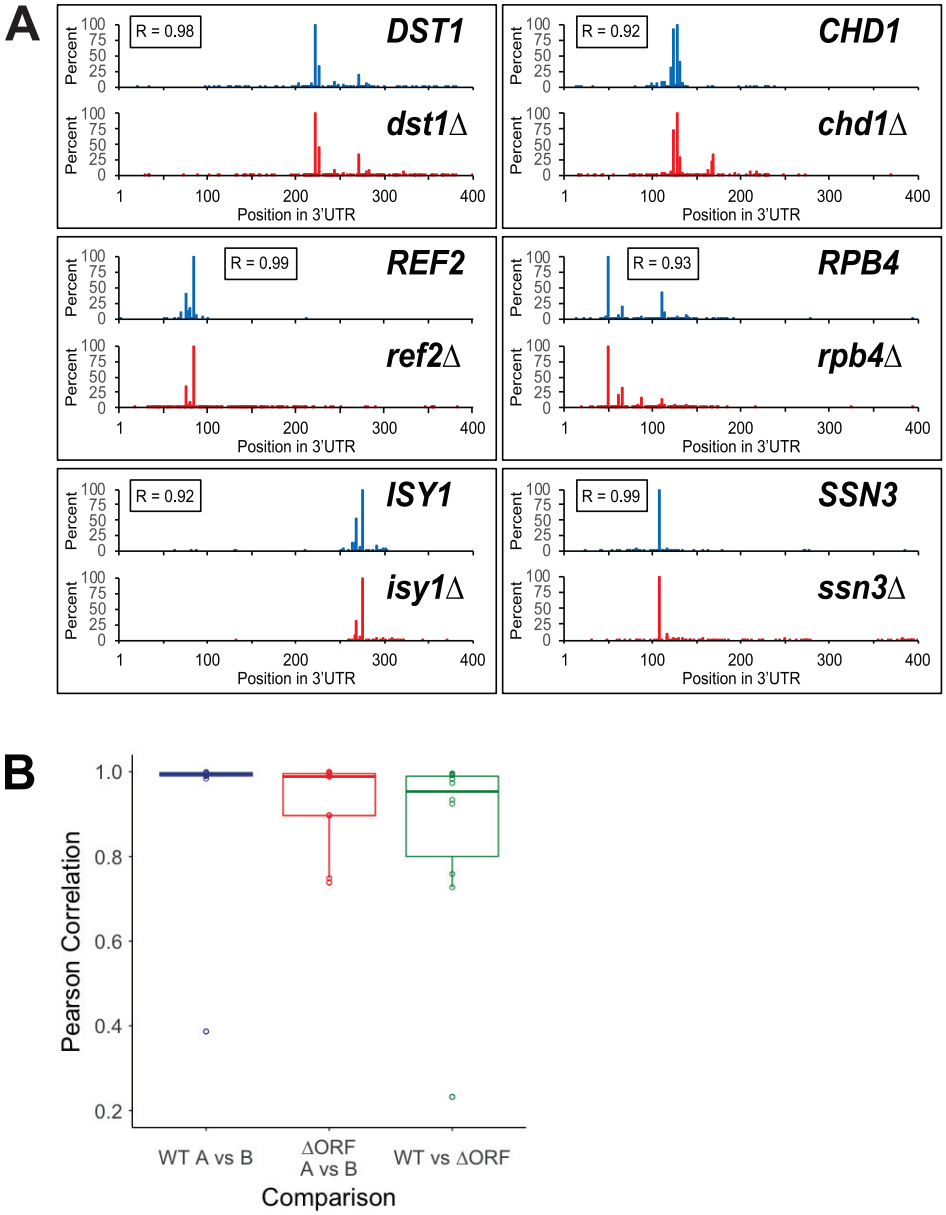


FIG 2 High similarity of poly(A) profiles of wild-type and ORF-deleted genes. (A) Poly(A) profiles for the indicated loci with Pearson correlation coefficients for each comparison highlighted in boxes. (B) Box plots of WT versus ΔORF and replicate correlations (10 loci). Some ΔORF replicate and WT:ΔORF correlations are slightly lower due to fewer sequence reads in the ΔORF samples.

to >0.99) and barely different from biological replicates ($P = 0.04$, Mann-Whitney U test). Thus, introduction of the same large heterologous region directly upstream of 15 different yeast 3′ UTRs has little effect on polyadenylation patterns of the native genes and does not cause their 3′-isoform profiles to become more similar to one another.

Transplanting a 3′ UTR to a different locus does not affect the poly(A) profile. To independently confirm the above results, we investigated whether transplantation of a 3′ UTR to a heterologous protein-coding region is sufficient to confer the donor 3′-UTR poly (A) profile to the recipient locus. The 3′ UTRs of *DPL1* and *YPS7* (which are located on different chromosomes and are distinct from the 25 genes analyzed above) were interchanged within one strain and poly(A) profiles of the chimeric alleles were compared to their wild-type counterparts (Fig. 5; biological replicates in Fig. S2). The poly(A) profiles of both chimeric alleles exhibit very strong correlation with wild-type alleles containing the same 3′

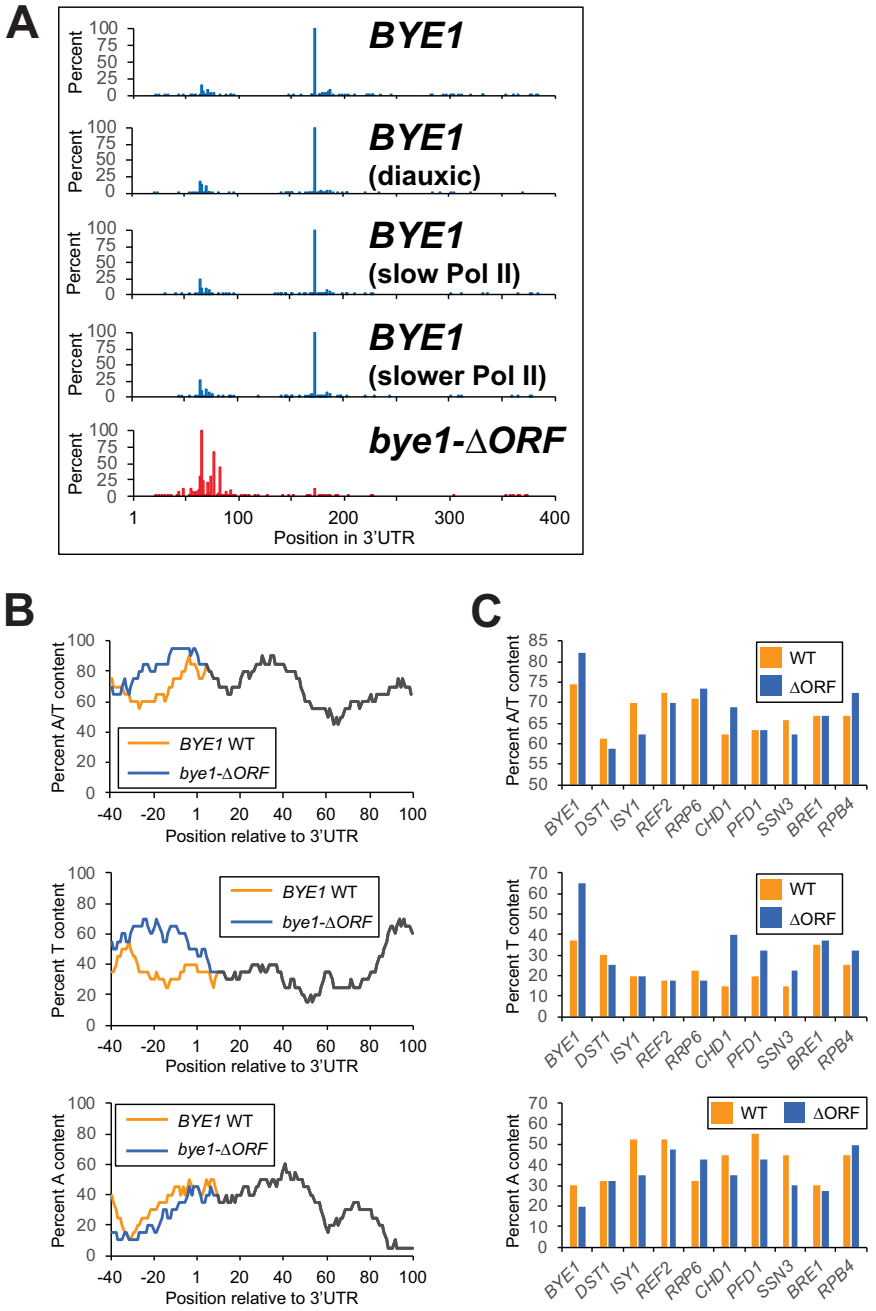


FIG 3 Deletion of the *BYE1* ORF causes an upstream shift in the poly(A) profile. (A) Poly(A) profiles of the wild-type *BYE1* locus in strains containing wild-type, slow, and slower Pol II derivatives as well as the *bye1-*ΔORF locus in the otherwise wild-type strain. (B) Percent AT, T, and A content of the coding strand in the region between −40 and +100 in the WT (orange) and *bye1-ΔORF* (blue) loci. Values represent a moving average (± 10 nt) at a given position, and gray shading indicates identical AT, T, or A content. The region between +1 and +100 is the same in both loci and shown in black. (C) Percent AT, T, and A content of the coding strand in the region between −40 and −1 in the WT and ΔORF loci.

UTR ($R = 0.9$ for the *YPS7* 3′ UTR; $R = 0.95$ for the *DPL1* 3′ UTR). Conversely, there is no correlation between wild-type alleles possessing identical ORFs but different 3′ UTRs ($R = -0.02$ for the *DPL1* ORF; $R = -0.01$ for the *YPS7* ORF). Thus, 3′ UTRs are sufficient to determine poly(A) profiles independently of upstream sequences and genomic location.

High similarity of poly(A) site usage in chimeric and native loci is independent of site location within the 3′ UTR. Polyadenylation sites of most yeast genes are located 50 to 250 nt downstream of translational termination codons, and nearly 80% of all 3′-UTR reads are found within this region. However, some yeast genes have a substantial proportion

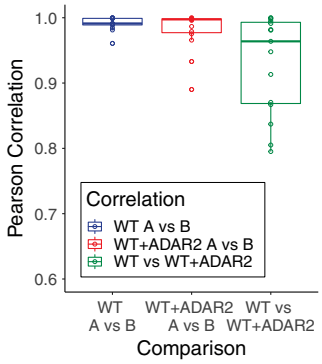
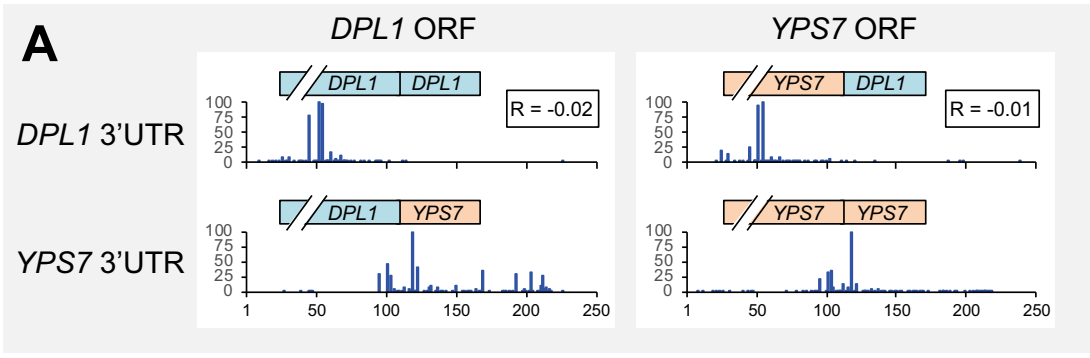


FIG 4 Box plot of poly(A) correlations of WT versus ADAR2 insertion and replicates (15 loci).

of 3′-mRNA isoforms that terminate either upstream or downstream of this window. We considered the possibility that ORF-proximal poly(A) site usage is preferentially affected when sequences immediately upstream of the 3′ UTR are altered. To address this, we computed poly(A) profile correlations in 50-nt subregions of chimeric and native loci with the same 3′ UTRs. With occasional exceptions and some variation among individual examples, median correlations in the 50-nt windows are uniformly high (median $R > 0.9$) for all subregions, including the most proximal (<50 nt from the stop codon) and most distal (>250 nt from the stop codon) positions (Fig. 6). Moreover, individual loci with significant reads within the first ~25 nt of the 3′ UTR show comparable usage in wild-type and ADAR2 insertion derivatives (*IMD3* and *CBC2*; Fig. S3). Thus, sequences outside the 3′ UTR, even those immediately adjacent to it, generally have very little influence on poly(A) site utilization, including poly(A) sites that are located just downstream of stop codons.



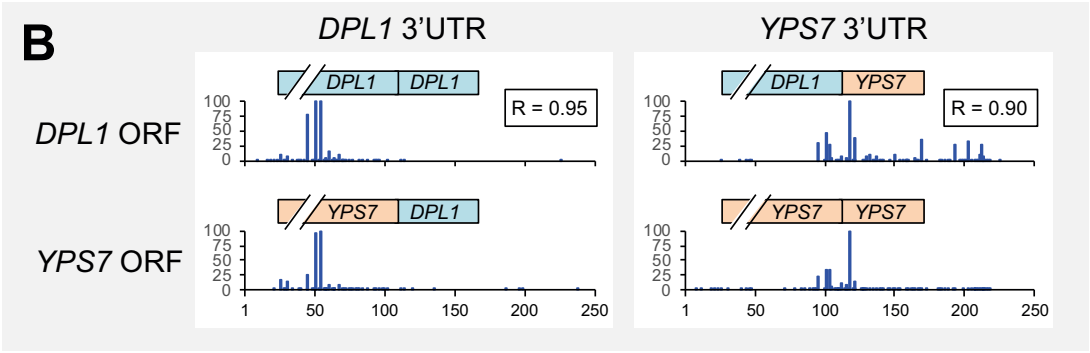


FIG 5 Poly(A) profile correlations in the 3′ UTRs of *DPL1*, *YPS7*, and their corresponding chimeric loci with interchanged 3′ UTRs. (A) Lack of correlation between poly(A) profiles of loci containing identical ORFs but different 3′ UTRs. As *DPL1* and *YPS7* 3′-UTR sequences differ, correlations were assayed across all identically positioned non-A residues. (B) High degree of poly(A) profile correlation at identical 3′ UTRs irrespective of upstream ORF.

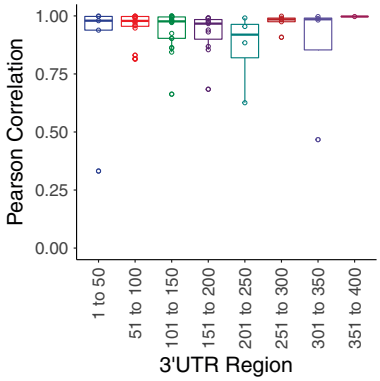


FIG 6 High poly(A) profile correlation within all 3′-UTR subregions. (A) Poly(A) correlations of WT versus ΔORF and WT versus ADAR2 insertion in the indicated windows (distance in nt from the stop codon and requires a minimum of 100 reads in both samples of a pair for each window). The number of points in each window mirrors the genome-wide read distribution; hence, there are more points in the regions between 50 to 200 nt and fewer points at more proximal and more distal locations.

Isoform half-life variation is encoded by 3′ UTRs. Many yeast genes (30 to 60%) exhibit 3′-isoform half-life variation, a phenomenon in which the half-lives of a gene's most and least stable isoforms vary by ≥2-fold (4, 24, 25). Thirteen of 27 genes examined here exhibit such variation, 7 show relative uniformity in same-gene isoform turnover, and the remaining 7 genes have insufficient sequence reads to generate reliable half-life data (Tables S1 and S2). Poly(A) profiles of genes that do or do not exhibit isoform stability variation are equally unaffected by changes in sequences upstream of 3′ UTRs (with variation median $R = 0.99$; no variation median $R = 0.99$; $P > 0.05$, Mann-Whitney U test). Thus, 3′-isoform half-life variation is essentially encoded by 3′ UTRs.

Protein-coding sequences increase mRNA stability of most genes but do not affect the relative stability of 3′ isoforms within a 3′ UTR. Pairwise comparisons of WT versus ΔORF loci not only show that 3′ UTRs determine poly(A) profiles, but they also provide information on the effect of protein-coding sequences on steady-state mRNA levels. As individual WT and ΔORF pairs have the identical promoter and 5′ UTR, changes in steady-state levels almost certainly reflect changes in mRNA stability (25). Of the nine genes in which the wild-type and ORF-deleted loci have extremely similar poly(A) profiles, six have ~2-fold higher levels of RNA in the WT strains relative to the ΔORF strain, and one (*RRP6*) has a 9.6-fold higher RNA level in the WT strain (Fig. 7A and Table S1). Thus, for seven out of these nine genes, it appears that the protein-coding sequences stabilize the mRNA in a 3′-isoform-independent manner. In striking contrast, the ORF-deleted *REF2* gene has 34 times higher RNA levels than the wild-type *REF2* gene. Possible explanations, not mutually exclusive, for this remarkably discordant effect might be strong destabilizing elements in the wild-type *REF2* gene, artifactual high stability of the *ref2-ΔORF* RNA, or an autoregulatory phenomenon in which loss of *REF2* function causes increased transcriptional initiation from the *REF2* promoter.

We initially generated 26 ORF-deleted genes but could only analyze poly(A) profiles for the 10 genes described in Fig. 1. For the remaining 16 genes, RNA levels in either the wild-type or ORF-deleted strain were too low for assessing the poly(A) profile. In most (13 out of 16) of these cases, the WT:ΔORF RNA ratio is very high (>5-fold; median, 26.4; range, 5.5 to 360; Fig. 7B and Table S2) and highly significant ($P = 1.4 \times 10^{-7}$, Mann-Whitney U test), indicating that most of these protein-coding regions exert a very strong mRNA-stabilizing effect. There is no relationship between the stabilizing effect of the coding region (WT:ΔORF RNA ratio) and translation efficiency of the wild-type gene as determined by ribosome profiling (26) (Pearson $R = 0.05$, $P > 0.05$).

Comparison of steady-state RNA levels of WT versus ADAR2 insertion loci provides information on how a heterologous coding region can influence mRNA stability. Interestingly, the WT:ADAR2-insertion ratios of RNA levels vary between 0.2 and 2.4, but most genes are minimally affected (<2-fold change in expression ratio) by ADAR2 insertion (Fig. 7C and Table S2). However, mRNA levels of five genes are altered by >2-fold, indicating that the

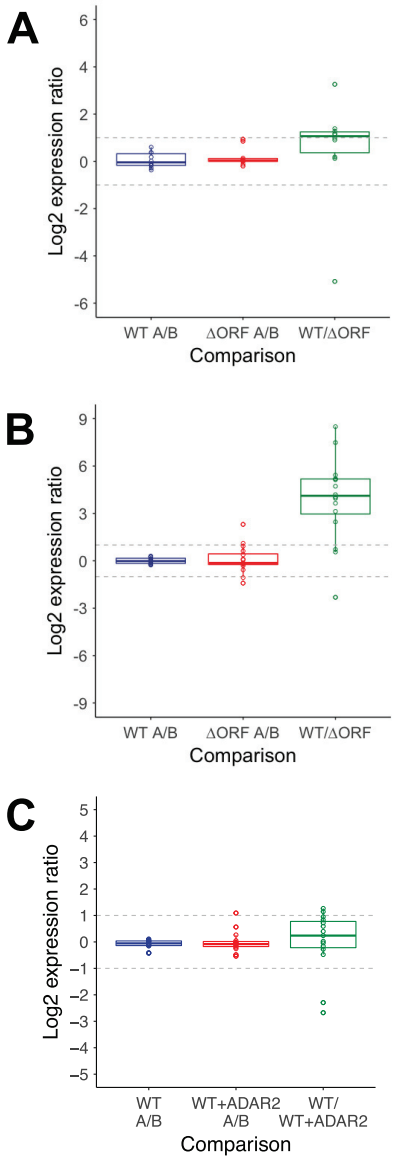


FIG 7 Protein-coding sequences typically stabilize steady-state mRNA levels. (A) Expression ratios (log_2) of steady-state levels (sum of all reads in each 3′ UTR) of the ΔORF allele with respect the corresponding WT allele of the 10 genes for which poly(A) profiles can be compared. (B) Expression ratios (log_2) of steady-state levels of the ΔORF allele with respect to the corresponding WT allele of the 16 genes for which poly(A) profiles cannot be assessed due to insufficient sequence reads in at least one sample. (C) Expression ratios (log_2) of steady-state levels of the ADAR2 insertion allele with respect the corresponding wild-type alleles.

ADAR2 coding region can either stabilize or destabilize mRNA levels, depending on the gene. This observation suggests that mRNA stability effects mediated by the ADAR2 and native yeast coding regions are not independent.

DISCUSSION

The 3′ UTR is sufficient to determine the polyadenylation profile for the vast majority of yeast genes. We examined the relative contributions of 3′ UTRs and ORF sequences to poly(A) profiles by performing three types of experiments across 27 yeast loci: (i) complete ORF deletion in 10 genes; (ii) a 1.5-kb in-frame insertion of ADAR2 coding sequences immediately upstream of 15 3′ UTRs; and (iii) reciprocal 3′-UTR swaps between two loci. Results from all three approaches demonstrate that the 3′ UTR is the primary driver of poly(A) profiles.

Specifically, pairwise correlation coefficients of wild-type versus mutant poly(A) profiles are extremely high for a given 3′ UTR, indicating that the relative utilization of poly(A) sites is nearly identical. While the number of genes assayed is limited, the near-uniform result (26 out of 27 cases tested) indicates that poly(A) profiles for the vast majority of yeast genes are determined by the 3′ UTR. After the work here was completed, experiments involving highly rearranged genomes suggested that isoform profiles are affected by larger transcriptional neighborhoods (27). This apparent inconsistency with our results remains to be resolved.

mRNA 3′ ends are generated during Pol II elongation by the cleavage/polyadenylation complex that cleaves the nascent transcript and then adds a poly(A) tail (14, 16–18). Our results indicate that the 3′ UTR is a functionally autonomous and modular entity that generates the same poly(A) profile in a variety of different contexts, including when transplanted to different chromosomal locations. Thus, both the Pol II elongation and cleavage/polyadenylation complexes must distinguish between protein-coding regions and 3′ UTRs. This distinction is particularly important because poly(A) sites are largely restricted to 3′ UTRs even though protein-coding regions are larger.

Evidence that the long AT- and/or AU-rich stretch is important for distinguishing 3′ UTRs from coding regions. The distinction and functional independence of ORFs and 3′ UTRs are ultimately due to broad sequence differences at the DNA and/or RNA level. Unlike ORFs, yeast 3′ UTRs have relatively long AT- (or AU)-rich stretches, leading to our previous speculation that these stretches were permissive for polyadenylation (6). The striking upstream shift in the poly(A) profile conferred by the *bye1-ΔORF* allele is consistent with this idea and further suggests that regions of high AT (or AU) richness play a role in the distinction between ORFs and 3′ UTRs. In this allele, the 5′-UTR sequence immediately juxtaposed to the *BYE1* 3′ UTR has an abnormally high AT content compared all the other wild-type and ORF-deleted genes tested. In essence, this exceptional situation shifts the AT-rich stretch further upstream, thereby establishing a functional connection between this stretch and poly(A) site utilization.

Although ~80% of all polyadenylation events in yeast take place in a 200-nt window located 50 to 250 nt downstream from the stop codon, some poly(A) sites are located immediately downstream of ORFs. The corresponding 3′-mRNA isoforms likely have shorter AU-rich sequences and fewer specificity elements, yet their relative utilization is not generally affected by sequences at the end of ORFs. One explanation for this observation is that AU-rich regions and positioning elements are dispensable for ORF-proximal poly(A) sites.

However, cleavage and polyadenylation require that the nascent RNA exit the Pol II elongation complex, which occurs at least 20 nt beyond the actual cleavage site (28). Thus, the strong preference for polyadenylation within 3′ UTRs is likely to involve Pol II elongation through the AT-rich region, which is significantly larger than the AU-rich region in the mRNA isoform. An attractive, but clearly speculative, model is that elongation through these AT-rich regions results in compositional and/or conformational changes in the Pol II elongation machinery that facilitate association and subsequent activity of the cleavage/polyadenylation complex. In this regard, the especially AT-rich region of the *bye1-ΔORF* allele has a very strong T bias in the coding strand, and T stretches are associated with transcriptional termination by bacterial (29) and archaeal (30) RNA polymerases and by eukaryotic RNA polymerase III (31). In addition, the position weight matrix of the 50 nt AT-rich stretch in natural yeast genes (6) shows 33 positions where T is favored over A in the coding strand and only 17 positions where A is favored over T.

mRNA stability elements in protein-coding regions and 3′ UTRs function independently. Many yeast genes have mRNA stabilizing or destabilizing elements within their 3′ UTRs such that same-gene 3′ isoforms can have different half-lives (4). Same-gene 3′-mRNA isoforms can have extensive structural differences, often mediated by RNA-binding proteins, that are linked to association with the differential poly(A)-binding protein binding and with mRNA stability (32). Although we have not directly measured 3′-isoform half-lives in this paper, the near-identical poly(A) profiles of paired loci with or without ORF sequences indicate that variation in stability of 3′ isoforms is mediated primarily by 3′ UTRs.

In contrast, most of the ORF-deletion derivatives exhibit reduced expression of all 3′ isoforms within the 3′ UTR, indicating that protein-coding sequences can contain elements

that stabilize mRNAs. Thus, mRNA stability elements within protein-coding regions and 3′ UTRs function independently. This suggests that stability-influencing RNA:RNA and RNA:protein interactions between ORFs and 3′ UTRs are likely to be minimal.

The mechanism(s) by which protein-coding regions typically stabilize mRNAs is unknown. However, ORF sequences within mRNAs are frequently structured, and secondary structure within coding sequences is positively correlated with enhanced mRNA stability (33, 34). RNAs transcribed from ΔORF alleles might be rapidly degraded due to the absence of secondary structure, or they might be subject to the nonsense-mediated decay pathway, although they lack extended untranslated coding regions that are important for this pathway (35, 36). Another possibility is that RNA sequences within protein-coding regions, but not 5′ UTRs or 3′ UTRs, are preferentially bound by proteins that protect the mRNA from degradation activities. Finally, RNA sequences within protein-coding regions could facilitate nuclear export, which could remove the RNA from nuclear degrading activities and/or stabilize the RNA by association with cytoplasmic entities. Any explanation must account for why most protein-coding regions strongly stabilize mRNAs, whereas a significant minority does not.

Modularity of protein-coding regions and 3′ UTRs. Protein-coding regions and 3′ UTRs are, by definition, autonomous and functionally distinct with respect to translation via sense and nonsense codons. Our results indicate that ORFs and 3′ UTRs are also independent and functionally distinct modules with respect to polyadenylation profiles (and subsequent transcriptional termination dependent on the RNA cleavage step) and 3′-mRNA isoforms stability variation. The molecular basis of this functional distinction is incompletely understood, but the behavior of the ORF-deleted *BYE1* allele suggests that long stretches of high AT content and mechanistic properties of the Pol II elongation machinery may play functional roles. RNA structures and differential interactions with RNA-binding proteins may also contribute. This modularity also suggests that chimeric protein-coding genes with a heterologous 3′ UTR will generally have regulatory properties associated with that 3′ UTR. This might include not only alternative poly(A) sites and 3′-isoform stability variation but also properties such as intracellular localization, microRNA regulation, and developmental regulation. Thus, the modular nature of 3′ UTRs could control gene function in a manner analogous to the way transplanted promoters can regulate gene expression under specific environmental or developmental conditions.

MATERIALS AND METHODS

Strains. Deletions of the protein-coding ORFs and, and transplantations of the 3′ UTR were introduced into the JGY2000 strain (*MATa*, *his3Δ0*, *leu2Δ0*, *met15Δ0*, *ura3Δ0*, *rpb1::RPB1–FRB*, *rpl13::RPL13–FK512*) by CRISPR using derivatives of pML104 (37) to supply Cas9 and guide RNA. Insertion of the large *ADAR2-E465Q-3V5* sequence (referred to as *ADAR2*) upstream of the 3′ UTR was accomplished by PCR-mediated integration. All strains were confirmed by PCR and Sanger sequencing. Oligonucleotides and strains used in this work are listed in Table S3.

RNA analysis. All strains were grown in duplicate in 50 ml of YPD to optical density at 600 nm of 0.3 to 0.4 at 30°C. Total RNA was isolated and purified from 15 to 20 ml of cells using the hot acid phenol method followed by Qiagen RNeasy as described (19). 3′ READS was performed with 25 μg of purified total RNA with 18 cycles of amplification as described previously (38). Barcoded libraries were quantified on an Agilent Bioanalyzer 2100, pooled, and sequenced on the Illumina NextSeq 500 platform.

Data analysis. Sequencing data were processed as described previously (19) using Python 3 (www.python.org). After separating sequence reads from multiplexed libraries by barcode into output from individual samples, we removed adapter sequences from read ends and discarded reads that contained ambiguous bases or did not start with a T (corresponding to an A at the mRNA 3′ end, potentially from polyadenylation). We counted and deleted consecutive Ts at the beginning of each read, after which we mapped the first 17 nt of remaining sequence for each read were mapped to the *Saccharomyces cerevisiae* genome (version Sac cer3) using Bowtie (39), allowing no mismatches and excluding nonunique matches. To ensure that we were working with posttranscriptionally adenylated RNA, we examined the genomic sequence immediately downstream of each mapped read, keeping only those reads for which the initial T count exceeded the number of consecutive A residues in the adjacent genomic sequence. Last, we scaled the remaining mapped reads for each replicate to a total of 25 million.

End zone parameters. We assigned reads to a gene's 3′ UTR if they mapped to the 400-nt window immediately downstream of its ORF (i.e., +1 represents the first nucleotide downstream of the stop codon). End zone profiles (poly(A) profiles) were constructed by setting the maximally expressed isoform (max isoform) in each gene's 3′ UTR to 100 and linearly scaling expression levels of all other isoforms for that gene relative to this maximal value. All loci analyzed in this work possess a minimum 3′-UTR total of 500 scaled reads in each context (wild-type and ORF deletion, ADAR2 insertion or 3′-UTR transplantation).

Poly(A) profile correlation analysis across subregions and whole 3′ UTRs. In most cases, correlations of poly(A) profiles were performed by comparing scaled reads of wild-type versus mutant (ORF deletion, ADAR2 insertion, or 3′-UTR transplantation) loci at all 3′-UTR positions. For the 3′-UTR transplantation experiments (identical ORFs fused to different 3′ UTRs; Fig. 5), all nucleotide positions that were an A residue in either 3′ UTR were eliminated from consideration, and poly(A) profile correlations were computed across all remaining non-A positions.

For the analysis of poly(A) site usage similarity at wild-type and mutant loci as a function of location within 3′ UTRs (Fig. 6), we subdivided the 3′ UTR into eight 50-nt windows commencing with +1 to +50 and ending with +351 to +400. For each window, we computed a single correlation score across the 50 possible positions for any locus that contained a minimum of 100 reads in both wild-type and mutant (ORF deletion, ADAR2 insertion, or 3′-UTR transplantation) contexts. Because most poly(A) reads are located between positions +50 and +250, upstream (+17 to +50) and downstream (+251 to + 300, +301 to +350 and +351 to +400) windows contain fewer data points.

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

SUPPLEMENTAL FILE 1, PDF file, 0.2 MB.

SUPPLEMENTAL FILE 2, XLSX file, 14.4 MB.

ACKNOWLEDGMENTS

We thank Catherine Maddox for excellent technical assistance.

This work was supported by grants to K.S. from the National Institutes of Health (GM30186 and GM131801).

REFERENCES

1. Pelechano V, Wei W, Steinmetz LM. 2013. Extensive transcriptional heterogeneity revealed by isoform profiling. Nature 497:127–131. https://doi.org/10.1038/nature12121.
2. Berkovits BD, Mayr C. 2015. Alternative 3' UTRs act as scaffolds to regulate membrane protein localization. Nature 522:363–367. https://doi.org/10.1038/nature14321.
3. Floor SN, Doudna JA. 2016. Tunable protein synthesis by transcript isoforms in human cells. Elife 5:e10921. https://doi.org/10.7554/eLife.10921.
4. Geisberg JV, Moqtaderi Z, Fan X, Ozsolak F, Struhl K. 2014. Global analysis of mRNA isoform half-lives reveals stabilizing and destabilizing elements in yeast. Cell 156:812–824. https://doi.org/10.1016/j.cell.2013.12.026.
5. Mayr C. 2016. Evolution and biological roles of alternative 3'UTRs. Trends Cell Biol 26:227–237. https://doi.org/10.1016/j.tcb.2015.10.012.
6. Moqtaderi Z, Geisberg JV, Jin Y, Fan X, Struhl K. 2013. Species-specific factors mediate extensive heterogeneity of mRNA 3' ends in yeasts. Proc Natl Acad Sci U S A 110:11073–11078. https://doi.org/10.1073/pnas.1309384110.
7. Ozsolak F, Kapranov P, Foissac S, Kim SW, Fishilevich E, Monaghan AP, John B, Milos PM. 2010. Comprehensive polyadenylation site maps in yeast and human reveal pervasive alternative polyadenylation. Cell 143:1018–1029. https://doi.org/10.1016/j.cell.2010.11.020.
8. Sherstnev A, Duc C, Cole C, Zacharaki V, Hornyik C, Ozsolak F, Milos PM, Barton GJ, Simpson GG. 2012. Direct sequencing of Arabidopsis thaliana RNA reveals patterns of cleavage and polyadenylation. Nat Struct Mol Biol 19:845–852. https://doi.org/10.1038/nsmb.2345.
9. Gruber AJ, Zavolan M. 2019. Alternative cleavage and polyadenylation in health and disease. Nat Rev Genet 20:599–614. https://doi.org/10.1038/s41576-019-0145-z.
10. Mayr C, Bartel DP. 2009. Widespread shortening of 3'UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells. Cell 138:673–684. https://doi.org/10.1016/j.cell.2009.06.016.
11. Li J, Lu X. 2013. The emerging roles of 3' untranslated regions in cancer. Cancer Lett 337:22–25. https://doi.org/10.1016/j.canlet.2013.05.034.
12. Masamha CP, Xia Z, Yang J, Albrecht TR, Li M, Shyu AB, Li W, Wagner EJ. 2014. CFIm25 links alternative polyadenylation to glioblastoma tumour suppression. Nature 510:412–416. https://doi.org/10.1038/nature13261.
13. Elkon R, Ugalde AP, Agami R. 2013. Alternative cleavage and polyadenylation: extent, regulation and function. Nat Rev Genet 14:496–506. https://doi.org/10.1038/nrg3482.
14. Tian B, Manley JL. 2013. Alternative cleavage and polyadenylation: the long and short of it. Trends Biochem Sci 38:312–320. https://doi.org/10.1016/j.tibs.2013.03.005.
15. Weill L, Belloc E, Bava FA, Mendez R. 2012. Translational control by changes in poly(A) tail length: recycling mRNAs. Nat Struct Mol Biol 19:577–585. https://doi.org/10.1038/nsmb.2311.
16. Proudfoot NJ, Furger A, Dye MJ. 2002. Integrating mRNA processing with transcription. Cell 108:501–512. https://doi.org/10.1016/S0092-8674(02)00617-7.
17. Tian B, Manley JL. 2017. Alternative polyadenylation of mRNA precursors. Nat Rev Mol Cell Biol 18:18–30. https://doi.org/10.1038/nrm.2016.116.
18. Kumar A, Clerici M, Muckenfuss LM, Passmore LA, Jinek M. 2019. Mechanistic insights into mRNA 3'-end processing. Curr Opin Struct Biol 59:143–150. https://doi.org/10.1016/j.sbi.2019.08.001.
19. Geisberg JV, Moqtaderi Z, Struhl K. 2020. The transcriptional elongation rate regulates alternative polyadenylation in yeast. Elife 9:e59810. https://doi.org/10.7554/eLife.59810.
20. Berg MG, Singh LN, Younis I, Liu Q, Pinto AM, Kaida D, Zhang Z, Cho S, Sherrill-Mix S, Wan L, Dreyfuss G. 2012. U1 snRNP determines mRNA length and regulates isoform expression. Cell 150:53–64. https://doi.org/10.1016/j.cell.2012.05.029.
21. Guo Z, Sherman F. 1996. 3'-end-forming signals of yeast mRNA. Trends Biochem Sci 21:477–481. https://doi.org/10.1016/S0968-0004(96)10057-8.
22. Keegan LP, Gerber AP, Brindle J, Leemans R, Gallo A, Keller W, O'Connell MA. 2000. The properties of a tRNA-specific adenosine deaminase from Drosophila melanogaster support an evolutionary link between pre-mRNA editing and tRNA modification. Mol Cell Biol 20:825–833. https://doi.org/10.1128/MCB.20.3.825-833.2000.
23. Palladino MJ, Keegan LP, O'Connell MA, Reenan RA. 2000. dADAR, a Drosophila double-stranded RNA-specific adenosine deaminase is highly developmentally regulated and is itself a target for RNA editing. RNA 6:1004–1018. https://doi.org/10.1017/S1355838200000248.
24. Gupta I, Clauder-Munster S, Klaus B, Jarvelin AI, Aiyar RS, Benes V, Wilkening S, Huber W, Pelechano V, Steinmetz LM. 2014. Alternative polyadenylation diversifies post-transcriptional regulation by selective RNA-protein interactions. Mol Syst Biol 10:719. https://doi.org/10.1002/msb.135068.
25. Moqtaderi Z, Geisberg JV, Struhl K. 2022. A compensatory link between cleavage/polyadenylation and mRNA turnover regulates steady-state mRNA levels in yeast. Proc Natl Acad Sci U S A 119:e2121488119. https://doi.org/10.1073/pnas.2121488119.
26. Ingolia NT, Ghaemmaghami S, Newman JR, Weissman JS. 2009. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. Science 324:218–223. https://doi.org/10.1126/science.1168978.
27. Brooks AN, Hughes AL, Clauder-Munster S, Mitchell LA, Boeke JD, Steinmetz LM. 2022. Transcriptional neighborhoods regulate transcript isoform lengths and expression levels. Science 375:1000–1005. https://doi.org/10.1126/science.abg0162.

28. Bernecky C, Herzog F, Baumeister W, Plitzko JM, Cramer P. 2016. Structure of transcribing mammalian RNA polymerase II. Nature 529:551–554. https://doi.org/10.1038/nature16482.
29. Ray-Soni A, Bellecourt MJ, Landick R. 2016. Mechanisms of bacterial transcription termination: all good things must end. Annu Rev Biochem 85:319–347. https://doi.org/10.1146/annurev-biochem-060815-014844.
30. Maier LK, Marchfelder A. 2019. It's all about the T: transcription termination in archaea. Biochem Soc Trans 47:461–468. https://doi.org/10.1042/BST20180557.
31. Arimbasseri AG, Maraia RJ. 2015. Mechanism of transcription termination by RNA Polymerase III utilizes a non-template strand sequence-specific signal element. Mol Cell 58:1124–1132. https://doi.org/10.1016/j.molcel.2015.04.002.
32. Moqtaderi Z, Geisberg JV, Struhl K. 2018. Extensive structural differences of closely related 3′ mRNA isoforms: links to Pab1 binding and mRNA stability. Mol Cell 72:849–861. https://doi.org/10.1016/j.molcel.2018.08.044.
33. Leppek K, Byeon GW, Kladwang W, Wayment-Steele HK, Kerr CH, Xu AF, Kim DS, Topkar VV, Choe C, Rothschild D, Tiu GC, Wellington-Oguri R, Fujii K, Sharma E, Watkins AM, Nicol JJ, Romano J, Tunguz B, Diaz F, Cai H, Guo P, Wu J, Meng F, Shi S, Participants E, Dormitzer PR, Solorzano A, Barna M, Das R. 2022. Combinatorial optimization of mRNA structure, stability, and translation for RNA-based therapeutics. Nat Commun 13:1536. https://doi.org/10.1038/s41467-022-28776-w.
34. Mauger DM, Cabral BJ, Presnyak V, Su SV, Reid DW, Goodman B, Link K, Khatwani N, Reynders J, Moore MJ, McFadyen IJ. 2019. mRNA structure regulates protein expression through changes in functional half-life. Proc Natl Acad Sci U S A 116:24075–24083. https://doi.org/10.1073/pnas.1908052116.
35. He F, Jacobson A. 2015. Nonsense-mediated mRNA decay: degradation of defective transcripts Is only part of the story. Annu Rev Genet 49:339–366. https://doi.org/10.1146/annurev-genet-112414-054639.
36. Kurosaki T, Popp MW, Maquat LE. 2019. Quality and quantity control of gene expression by nonsense-mediated mRNA decay. Nat Rev Mol Cell Biol 20:406–420. https://doi.org/10.1038/s41580-019-0126-2.
37. Laughery MF, Hunter T, Brown A, Hoopes J, Ostbye T, Shumaker T, Wyrick JJ. 2015. New vectors for simple and streamlined CRISPR-Cas9 genome editing in Saccharomyces cerevisiae. Yeast 32:711–720. https://doi.org/10.1002/yea.3098.
38. Jin Y, Geisberg JV, Moqtaderi Z, Ji Z, Hoque M, Tian B, Struhl K. 2015. Mapping 3' mRNA isoforms on a genomic scale. Curr Protoc Mol Biol 110:4.23.1–24.23.1 7.
39. Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10:R25. https://doi.org/10.1186/gb-2009-10-3-r25.