Comparative analysis of alternative polyadenylation in *S. cerevisiae* and *S. pombe*

Xiaochuan Liu,[1] Mainul Hoque,[1] Marc Larochelle,[2] Jean-François Lemay,[2] Nathan Yurko,[3] James L. Manley,[3] François Bachand,[2] and Bin Tian[1]

[1] Department of Microbiology, Biochemistry and Molecular Genetics, Rutgers New Jersey Medical School, Newark, New Jersey 07103, USA; [2] RNA Group, Department of Biochemistry, Université de Sherbrooke, Sherbrooke, Quebec J1E 4K8, Canada; [3] Department of Biological Sciences, Columbia University, New York, New York 10027, USA

Alternative polyadenylation (APA) is a widespread mechanism that generates mRNA isoforms with distinct properties. Here we have systematically mapped and compared cleavage and polyadenylation sites (PASs) in two yeast species, *S. cerevisiae* and *S. pombe*. Although >80% of the mRNA genes in each species were found to display APA, *S. pombe* showed greater 3′ UTR size differences among APA isoforms than did *S. cerevisiae*. PASs in different locations of gene are surrounded with distinct sequences in both species and are often associated with motifs involved in the Nrdl-Nab3-Senl termination pathway. In *S. pombe*, strong motifs surrounding distal PASs lead to higher abundances of long 3′ UTR isoforms than short ones, a feature that is opposite in *S. cerevisiae*. Differences in PAS placement between convergent genes lead to starkly different antisense transcript landscapes between budding and fission yeasts. In both species, short 3′ UTR isoforms are more likely to be expressed when cells are growing in nutrient-rich media, although different gene groups are affected in each species. Significantly, 3′ UTR shortening in *S. pombe* coordinates with up-regulation of expression for genes involved in translation during cell proliferation. Using *S. pombe* strains deficient for Pcf11 or Pab2, we show that reduced expression of 3′-end processing factors lengthens 3′ UTR, with Pcf11 having a more potent effect than Pab2. Taken together, our data indicate that APA mechanisms in *S. pombe* and *S. cerevisiae* are largely different: *S. pombe* has many of the APA features of higher species, and Pab2 in *S. pombe* has a different role in APA regulation than its mammalian homolog, PABPN1.

[Supplemental material is available for this article.]

Most protein-coding genes in eukaryotes contain multiple cleavage and polyadenylation sites (PASs), leading to expression of alternative polyadenylation (APA) isoforms (Shi 2012; Elkon et al. 2013; Mayr 2016; Proudfoot 2016; Tian and Manley 2017). The global APA profile is dynamically regulated in cell proliferation, differentiation, and development (Sandberg et al. 2008; Ji and Tian 2009; Ji et al. 2009; Shepard et al. 2011), as well as in response to extracellular cues and cell stresses (Flavell et al. 2008; Yoon and Brem 2010; Graber et al. 2013; Hollerer et al. 2016).

PASs are defined by surrounding sequence motifs (Graber et al. 1999; Hu et al. 2005). In vertebrates, prominent upstream motifs include the A[A/U]UAAA motif located within 40 nt upstream of the PAS, as well as UGUA and U-rich motifs that typically surround the A[A/U]UAAA motif. Downstream motifs include U-rich and UGUG motifs. The sequence motifs around the PAS vary in lower species. The A[A/U]UAAA was more degenerate in *Saccharomyces cerevisiae* (Graber et al. 1999), and the UGUG motif appears to be restricted to vertebrates (Tian and Graber 2012). The PAS motifs determine the strength of a PAS in a combinatory manner (Cheng et al. 2006; Weng et al. 2016) and are recognized by RNA-binding proteins (RBPs) in the cleavage and polyadenylation (PA) complex. The PA complex contains more than 20 proteins, some of which form subcomplexes (Shi and Manley 2015). In addition, many additional proteins are associated with the PA complex, potentially functioning to connect PAS usage with other cellular processes, such as transcription and splicing (Shi et al. 2009). Although largely conserved, PA complexes in mammals and *S. cerevisiae* differ substantially in organization of subcomplexes as well as constituent factors (Mandel et al. 2008; Tian and Graber 2012).

Although *S. cerevisiae* and *Schizosaccharomyces pombe* are remarkably useful as eukaryotic models for basic molecular mechanisms, the two species are distantly related, with the last divergence believed to be ~350 million years ago (Dujon 2010; Hoffman et al. 2015). It is generally believed that due to more rapid evolution, *S. cerevisiae* differs to a greater extent than *S. pombe* from the common ancestor with metazoans (Dujon 2010; Hoffman et al. 2015). As such, many aspects of the molecular mechanisms in higher species are more conserved in *S. pombe* than in *S. cerevisiae*. Studies indicated that APA is widespread in *S. cerevisiae* (Kim Guisbert et al. 2007; Yoon and Brem 2010; Graber et al. 2013; Moqtaderi et al. 2013; Wilkening et al. 2013), and it is frequently regulated under growth and stress conditions (Kim Guisbert et al. 2007; Yoon and Brem 2010; Graber et al. 2013; Moqtaderi et al. 2013; Wilkening et al. 2013). Although less studied in *S. pombe*, APA has also been reported to be widespread (Mata 2013; Schlackow et al. 2013). How *S. cerevisiae* and *S. pombe* compare with each other with respect to PAS motifs and APA properties has not been addressed. Here, we used 3′READS, a method

specialized in sequencing of the 3′ end of poly(A)+ RNA, to examine APA in *S. cerevisiae* and *S. pombe*. We compared the genomic features of and sequence motifs around the PASs in these two yeast species. Further, using cells growing in minimal or rich media as model conditions, we compared APA changes in these two species. Finally, we analyzed APA regulation in *S. pombe* strains deficient in certain PA factors.

Results

APA in *S. cerevisiae* and *S. pombe*

We were interested in studying APA in yeast genomes. We obtained total RNAs from the budding yeast *S. cerevisiae* and the fission yeast *S. pombe* grown in rich and minimal media (Fig. 1A; Methods) and subjected them to deep sequencing analysis using our 3′-end–based method 3′READS (Hoque et al. 2013). 3′READS is not affected by the internal priming problem in PAS identification, in which internal A-rich regions of mRNA are treated as poly(A) tails during oligo(dT)-based reverse transcription, leading to erroneous mapping of the PAS (Nam et al. 2002). This technical advantage may be particularly important for yeast genomes that contain abundant A-rich sequences in 3′ UTRs (Graber et al. 1999). We obtained more than 6 million PAS reads (reads mapped to PAS) (Methods) for each species (Supplemental Table S1). Using publicly annotated gene models coupled with 3′ end extension

using strand-specific RNA-seq data (Methods), we found that, in both species, most PAS reads were mapped to 3′ UTRs (>90%) (Fig. 1B, left), corresponding to ~60% of unique PAS read locations (Fig. 1B, right). About 24% and 23% of the unique PAS read locations were in coding sequences (CDS) and 5′ UTR sequences in *S. cerevisiae* and *S. pombe* genomes, respectively (Fig. 1B), accounting for ~4% and 2% of total PAS reads for these two species. About 0.2% of PAS reads (2% of the total PAS read locations) were additionally mapped in introns of *S. pombe* (Fig. 1B). Further, in each species, 4%–5% of PAS reads were mapped to unannotated intergenic regions, accounting for 14%–18% of unique locations. In summary, despite pervasive occurrences of PASs in the two genomes, most PA events take place in 3′ UTRs.

Heterogeneous cleavage by the PA machinery often leaves multiple cleavage sites (CSs) located close to one another (Fig. 1C). As such, CSs dependent on the same surrounding sequence motifs are typically clustered together to form a PAS. In higher species, because different PASs are well separated (Tian et al. 2005), heterogeneous cleavage can be readily distinguished from APA. For example, a distance of 24 nt was used to group adjacent CSs into a PAS cluster in mammalian genes (Tian et al. 2005). However, due to overlapping sequence motifs between adjacent PASs in yeast genomes (Geisberg et al. 2014), grouping CSs is not straightforward. We thus set out to model two types of distances between adjacent CSs: one corresponding to the distance between CSs within a PAS cluster (heterogeneous cleavage), and the other between adjacent PAS clusters (APA) (Fig. 1C). We used a maximum expectation method to identify two modes in the distribution of distance between adjacent CSs and used the crossover point of these two identified modes as the cutoff to separate PAS clusters (Methods). This method resulted in values of 7.1 and 7.4 nt, respectively, for clustering CSs in *S. cerevisiae* and *S. pombe* (Fig. 1C). As a control to validate this approach, we applied the same method on mouse CSs we previously identified (Hoque et al. 2013), which yielded 24.1 nt as the cutoff (Supplemental Fig. S1), a number in good agreement with the 24 nt previously identified for mammalian CSs (Tian et al. 2005). After clustering of heterogeneous CSs, we found that 84.7% and 82.4% of mRNA genes in *S. cerevisiae* and *S. pombe*, respectively, displayed APA when each APA isoform abundance was required to be >5% of all isoforms in a gene (Fig. 1D). The frequencies dropped as more stringent cutoffs were used. For example, the percentages of APA genes would decrease to 50%–60% if a minor APA isoform abundance was required to be >20% (Fig. 1D). Together, our data indicate widespread APA in both yeast species.

APA sites in 3′ UTRs lead to isoforms with different 3′ UTR lengths (Fig. 2A). We found that 78% and 71% of protein-coding genes in *S. cerevisiae* and *S. pombe*, respectively, expressed multiple
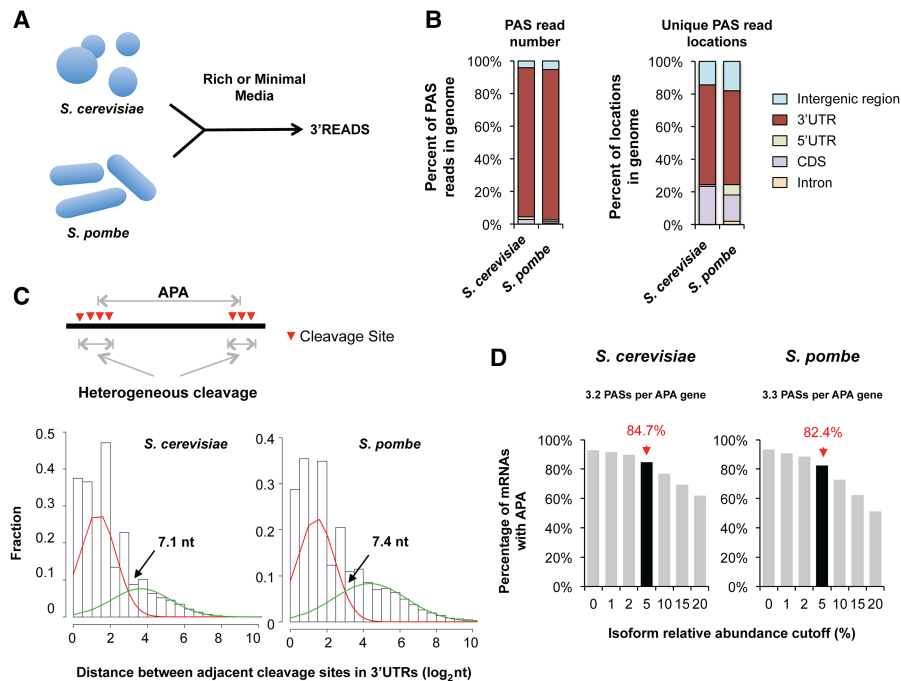

Figure 1. Mapping PASs in *S. cerevisiae* and *S. pombe* genomes. (*A*) Schematic showing the experimental design. (*B*) Statistics of PAS reads mapped to different regions of *S. cerevisiae* and *S. pombe* genomes (*left*) and unique PAS read locations (*right*). See Methods for definition of PAS read. (*C, top*) Schematic showing the difference between heterogeneous cleavage and APA. Heterogeneous cleavage sites are merged to a PAS cluster. (*Bottom*) Distributions of distances between adjacent CSs in *S. cerevisiae* (*left*) and *S. pombe* (*right*). Two distribution modes were identified by the maximum expectation method (Methods), with the red line showing the distance between CSs within a PAS cluster and the green line the distance between CSs from different PAS clusters. The crossover point (indicated by arrow) was identified for each plot, which was used to group CSs into PAS clusters. (*D*) APA frequencies of mRNA genes in *S. cerevisiae* (*left*) and *S. pombe* (*right*). Different abundance cutoffs were used for calling APA sites, leading to different APA frequencies as indicated. The overall APA frequencies were 84.7% and 82.4% in *S. cerevisiae* and *S. pombe* genomes, respectively, when the relative abundance cutoff was set at 5%.
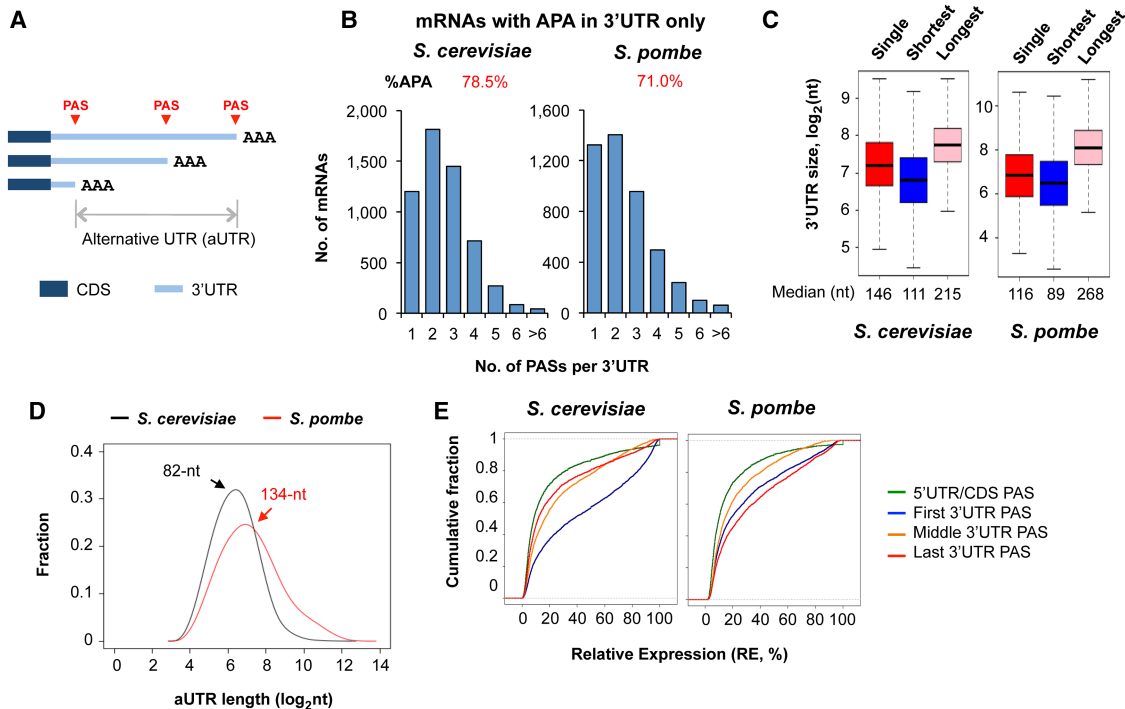
Figure 2. 3′ UTR regulation by APA. (*A*) Schematic showing 3′ UTR length control by APA. The region between the first PAS and last PAS is called alternative UTR, or aUTR. (*B*) Number of mRNA genes with 3′ UTR APA sites in *S. cerevisiae* (5583 genes in total) (*left*) and *S. pombe* (4579 genes in total) (*right*). The percentage of mRNA genes with 3′ UTR APA is 78.5% in *S. cerevisiae* and 71% in *S. pombe.* The average number of 3′ UTR PASs per gene is 2.6 in *S. cerevisiae* and 2.5 in *S. pombe*. (*C*) Box plots of 3′ UTR size for genes without APA (single) and for genes with APA (the sizes of shortest and longest 3′ UTR isoforms were plotted). (*D*) Distance between the first and last PASs of genes with 3′ UTR APA. The average values for *S. cerevisiae* and *S. pombe* are indicated. (*E*) Relative expression levels of APA isoforms. APA isoform types were based on PAS locations, as indicated.

3′ UTR isoforms (Fig. 2B). On average, an mRNA gene had 2.6 3′ UTR APA sites in *S. cerevisiae* and 2.5 in *S. pombe* (Fig. 2B). The shortest and longest 3′ UTRs of a gene were 111 and 215 nt, respectively, in *S. cerevisiae*; they were 89 and 268 nt in *S. pombe*. Thus, compared to the budding yeast, the shortest 3′ UTR isoform in the fission yeast is shorter, but the longest isoform is longer. Consequently, the averaged distance between the first and last PASs in 3′ UTR, also termed alternative 3′ UTR (aUTR), in the fission yeast is 63% larger than that in the budding yeast (134 nt versus 82 nt) (Fig. 2D).

Gene Ontology (GO) analysis of genes with 3′ UTR APA identified distinct GO terms for *S. cerevisiae* and *S. pombe* (Table 1). In *S. cerevisiae*, the top enriched biological process (BP) terms were "organic cyclic compound biosynthesis process," "organic cyclic compound metabolic process," and "response to drug." In contrast, in *S. pombe*, the top BP terms were "Golgi vesicle transport," "vesicle-mediated transport," and "cytoplasmic transport." These results indicate different sets of genes tend to harbor 3′ UTR APA sites in these two species. Interestingly, GO terms related to intracellular transport was also significant for mouse genes with 3′ UTR APA (Supplemental Table S2), suggesting similarities between the fission yeast and higher species in APA functions.

We next examined the expression levels of different APA isoforms. APA isoforms were divided into different groups based on the relative location of PAS in the gene, namely, first (F), middle (M), and last (L) PASs in 3′ UTR and PASs located in upstream CDS or 5′ UTR (Fig. 2E). In both species, isoforms using CDS or 5′ UTR PASs were expressed at the lowest levels. Interestingly, distal PAS isoforms were generally expressed at lower levels than proxi-

mal PAS isoforms in *S. cerevisiae* (blue versus red lines in Fig. 2E, left). In contrast, the pattern was reversed in the fission yeast, with distal PAS isoforms being more expressed than proximal PAS isoforms (blue versus red lines in Fig. 2E, right), a trend similar to that in mammals (Hoque et al. 2013).

PAS motifs in two yeast genomes

We next examined and compared sequence motifs around PASs between the two yeast species. We first identified all significantly enriched 6-mers and 4-mers around the PASs in each genome, focusing on three regions around the PAS, i.e., −60 to −31 nt, −30 to −1 nt, and +1 to +30 nt (Fig. 3; Supplemental Fig. S2). Consistent with previous reports (Mata 2013; Schlackow et al. 2013), despite similar surrounding nucleotide frequencies (Fig. 3A), PASs in the two yeast species were associated with distinct upstream and downstream sequence motifs (Fig. 3B). PASs of *S. cerevisiae* were surrounded by upstream UAUA, U-rich and A-rich motifs, and downstream U-rich and UAUA motifs (Fig. 3B; Supplemental Fig. S2, top). In contrast, PASs of *S. pombe* were associated with upstream UAUA and AAU[A/G]AA motifs and downstream GUA, and U-rich and A-rich motifs (Fig. 3B; Supplemental Fig. S2, bottom). Thus, as summarized in Figure 3C, the PASs in the budding yeast are loosely defined by UAUA, U-rich, and A-rich motifs located both upstream of and downstream from the PAS, whereas PAS motifs in the fission yeast are more clearly defined in specific positions relative to the cleavage site.

The downstream GUA motif in *S. pombe* appears to be the binding site of Seb1, a homolog of Nrd1 in budding yeast, which

Table 1. Top Gene Ontology terms associated with yeast genes that contain 3′ UTR APA sites

Species	Category	GO Term	$-\log_{10}P$
S. cerevisiae	BP	Organic cyclic compound biosynthetic process	5.1
	BP	Organic cyclic compound metabolic process	4.9
	BP	Response to drug	2.0
	CC	Nucleus	2.3
	CC	Non-membrane-bounded organelle	2.0
	CC	Mitochondrial protein complex	2.0
S. pombe	BP	Golgi vesicle transport	3.6
	BP	Vesicle-mediated transport	3.2
	BP	Cytoplasmic transport	2.9
	CC	Golgi apparatus	2.9
	CC	Vacuolar membrane	2.5
	CC	Vacuole	2.5

(BP) Biological process, (CC) cellular component.

was recently found to be involved in the regulation of PAS selection (Lemay et al. 2016). Nrd1 is part of the Nrd1-Nab3-Sen1 (NNS) complex involved in 3′-end processing of short genes in a PAS-independent manner (Arndt and Reines 2015; Porrua and Libri 2015). However, its role as a fail-safe mechanism for termination of PAS-containing genes has also been demonstrated in the budding yeast (Rondon et al. 2009). Interestingly, some CU-rich motifs, such as UUCU, UCUU, CUUU, and UUUC, were found to be mildly enriched around the PASs of S. cerevisiae (Fig. 3B; Supplemental Fig. S2). These motifs resemble the binding sites of Nab3, another subunit of the NNS complex, suggesting a general role of NNS in 3′-end processing of mRNA genes in the budding yeast.

We also identified several AG-rich motifs around the PASs in S. cerevisiae (Fig. 3B; Supplemental Fig. S2). They were previously reported to be associated with PASs regulated in stress conditions (Yoon and Brem 2010; Graber et al. 2013). Further analysis of CDS PASs indicated that these motifs were highly enriched in the flanking regions of CDS PASs in both species (Supplemental Fig. S3). Conceivably, the higher fraction of CDS PASs in the budding yeast (Fig. 1B, right) makes these motifs more prominent than they were in the fission yeast. How these motifs play a role in recruiting the PA complex for PA, however, remains to be elucidated.

We next asked whether proximal and distal PASs in yeast genes were surrounded with different motifs, as in mammals (Legendre and Gautheret 2003; Hu et al. 2005; Tian et al. 2005). To this end, we directly compared the first and last 3′ UTR PASs for their surrounding motifs in each species. Interestingly, in S. cerevisiae, UAUA motifs were much more significantly enriched for the first PAS (Fig. 3D). In contrast, in S. pombe, the upstream AAUAAA and UA-rich motifs were more enriched for the last PAS, and the first PAS tended to be surrounded by U-rich motifs, especially in the downstream region (Fig. 3E). Remarkably, the last PAS in both species tended to be surrounded by putative NNS binding motifs (Fig. 3D,E), i.e., CU-rich motifs in the budding yeast and the GUA motif in the fission yeast, further implicating a general role of the NNS pathway in termination of mRNA genes. These results indicate that, due to differential placements of favorable sequence motifs, proximal PASs in S. cerevisiae were stronger than distal ones, whereas distal PASs in S. pombe were stronger than proximal ones (Fig. 3F). This finding is consistent with the observation that distal PAS isoforms tend to be expressed at higher levels than proximal PASs in S. pombe but not in S. cerevisiae (Fig. 2E).

We further examined sequence conservation levels around the first and last 3′ UTR PASs in the two yeast species. Using PhastCons scores based on closely related yeast species (Methods), we found that the upstream region of PASs in both yeast species had higher conservation levels than the downstream region (Fig. 3G; Supplemental Fig. S4A), indicating the importance of upstream motifs for PAS usage in these species. However, the conservation profiles of yeast PASs were quite different than those of mouse PASs, which showed a conspicuous increase of conservation in the region containing the AAUAAA motif (Supplemental Fig. S4). This result highlights the diffusive nature of functional motifs around yeast PASs. Interestingly, although the upstream region of the first 3′ UTR PAS in S. cerevisiae was substantially more conserved that of the last PAS, this difference was much milder in S. pombe (Fig. 3G), indicating greater negative selection on distal PASs in S. pombe than in S. cerevisiae. Presumably, distal PASs in S. pombe are more important for gene expression than those in S. cerevisiae.

PASs of overlapping transcripts

Expression of antisense RNA is pervasive in S. cerevisiae (Xu et al. 2011). We next wanted to compare PASs of both sense and antisense transcripts in the two yeast species. Using the annotated transcript end site (TES) as a reference point, we found that although both species showed a similar PAS peak around the TES on the sense strand (Fig. 4A), they appeared significantly different on the antisense strand (Fig. 4A). There was an obvious antisense PAS peak located before the TES in S. cerevisiae (highlighted by an arrow, Fig. 4A, left), indicating substantial overlap between sense and antisense transcripts at the 3′ end. In contrast, the sense and antisense peaks in S. pombe overlapped to a much lesser extent (Fig. 4A, right). Interestingly, the antisense PAS peak in S. cerevisiae coincided almost perfectly with the stop codon (Fig. 4B, left), which was not observed in S. pombe (Fig. 4B, right). Taken together, these results indicate that, due to differential PAS placements on the antisense strand near the 3′ end of gene, two yeast genomes have starkly different antisense transcript landscapes.

We also examined PASs around the transcription start site (TSS). Both yeast species displayed a sense PAS peak upstream of the TSS (Fig. 4C). However, the peak in S. cerevisiae was much more conspicuous than the one in S. pombe (Fig. 4C), and more PASs in the budding yeast were located immediately downstream from the TSS than in the fission yeast (Fig. 4C, arrows). Similar results were obtained when the start codon was used as the reference point (Fig. 4D). These data indicate that transcription in S. cerevisiae is more likely to run into the downstream gene, causing transcriptional interference (Lemay and Bachand 2015; Proudfoot 2016).

Regulation of APA in different growth conditions

Cells growing in nutrient-restricted conditions are more likely to withdraw from cell cycle and reach a quiescent state (Hoffman et al. 2015). We next asked whether APA profiles change in cells growing in rich versus minimal media. Using the top two most abundant 3′ UTR PAS isoforms of each gene, we examined the global 3′ UTR APA trend. We found that genes with up-regulated proximal PAS isoforms in the rich media as compared to minimal media outnumbered those with the opposite trend by 1.7-fold in S. cerevisiae (359 versus 206) (Fig. 5A). S. pombe displayed a similar but much more conspicuous trend, with a bias of 2.7-fold (546 versus 206) (Fig. 5B). Therefore, yeast genes in general prefer expression of
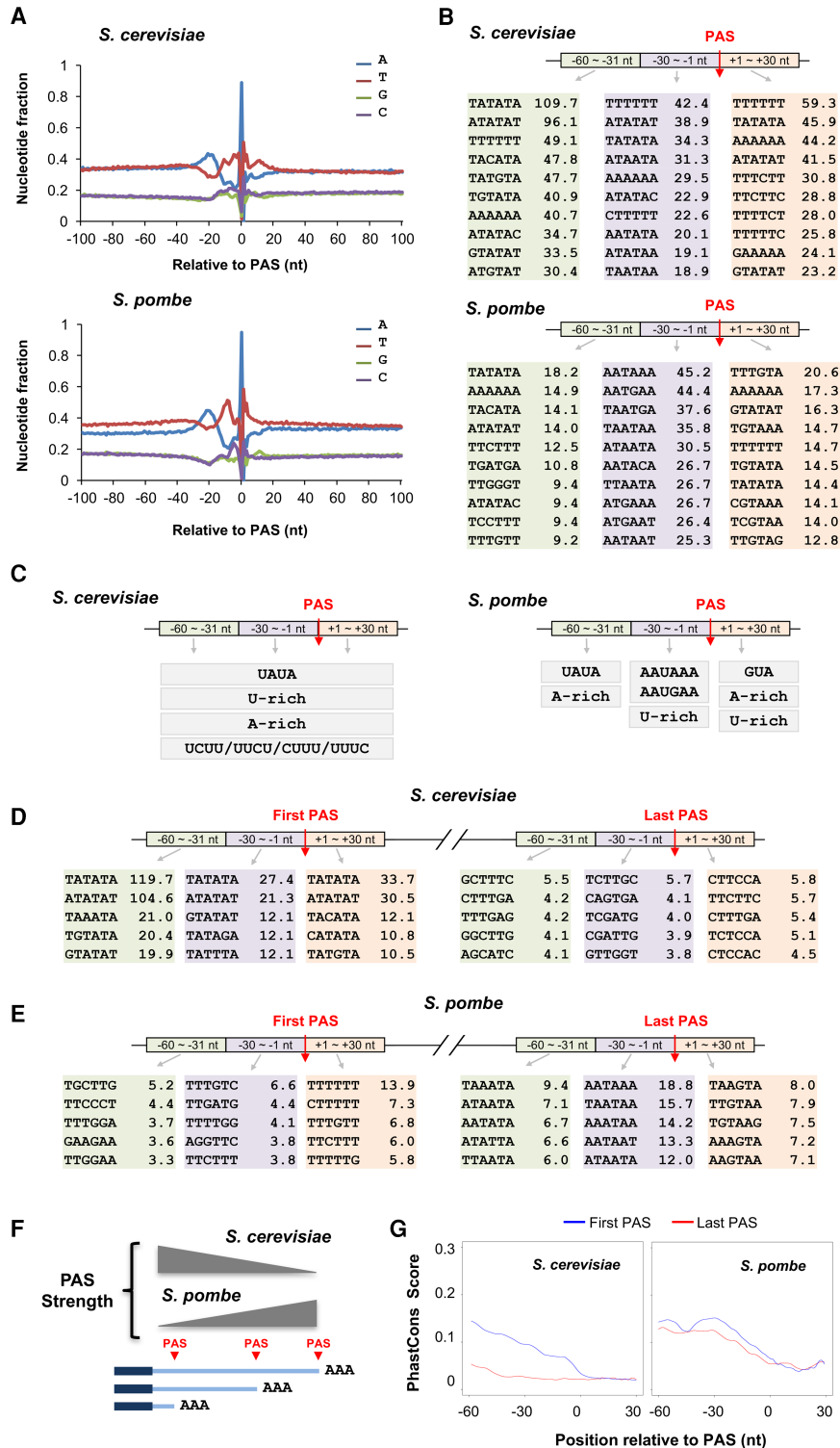
Figure 3 (A–G)

A — Nucleotide frequencies

S. cerevisiae

S. pombe

(Legend: A, T, G, C; x-axis: Relative to PAS (nt); y-axis: Nucleotide fraction)

B — Enriched 6-mers

S. cerevisiae

PAS					
−60 ~ −31 nt		−30 ~ −1 nt		+1 ~ +30 nt	
TATATA	109.7	TTTTTT	42.4	TTTTTT	59.3
ATATAT	96.1	ATATAT	38.9	TATATA	45.9
TTTTTT	49.1	TATATA	34.3	AAAAAA	44.2
TACATA	47.8	ATAATA	31.3	ATATAT	41.5
TATGTA	47.7	AAAAAA	29.5	TTTCTT	30.8
TGTATA	40.9	ATATAC	22.9	TTCTTC	28.8
AAAAAA	40.7	CTTTTT	22.6	TTTTCT	28.0
ATATAC	34.7	AATATA	20.1	TTTTTT	25.8
GTATAT	33.5	ATATAA	19.1	GAAAAA	24.1
ATGTAT	30.4	TAATAA	18.9	GTATAT	23.2

S. pombe

PAS					
−60 ~ −31 nt		−30 ~ −1 nt		+1 ~ +30 nt	
TATATA	18.2	AATAAA	45.2	TTTGTA	20.6
AAAAAA	14.9	AATGAA	44.4	AAAAAA	17.3
TACATA	14.1	TAATGA	37.6	GTATAT	16.3
ATATAT	14.0	TAATAA	35.8	TGTAAA	14.7
TTCTTT	12.5	ATAATA	30.5	TTTTTT	14.7
TGATGA	10.8	AATACA	26.7	TATAAA	14.5
TTGGGT	9.4	TTAATA	26.7	TATATA	14.4
ATATAC	9.4	ATGAAA	26.7	CGTAAA	14.1
TCCTTT	9.4	ATGAAT	26.4	TCGTAA	14.0
TTTGTT	9.2	AATAAT	25.3	TTGTAG	12.8

C — Summary of significant motifs

S. cerevisiae

PAS		
−60 ~ −31 nt	−30 ~ −1 nt	+1 ~ +30 nt
UAUA		
U-rich		
A-rich		
UCUU/UUCU/CUUU/UUUC		

S. pombe

PAS		
−60 ~ −31 nt	−30 ~ −1 nt	+1 ~ +30 nt
UAUA	AAUAAA	GUA
A-rich	AAUGAA	A-rich
	U-rich	U-rich

D — *S. cerevisiae*

First PAS						Last PAS					
−60 ~ −31 nt		−30 ~ −1 nt		+1 ~ +30 nt		−60 ~ −31 nt		−30 ~ −1 nt		+1 ~ +30 nt	
TATATA	119.7	TATATA	27.4	TATATA	33.7	GCTTTC	5.5	TCTTGC	5.7	CTTCCA	5.8
ATATAT	104.6	ATATAT	21.3	ATATAT	30.5	CTTTGA	4.2	CAGTGA	4.1	TTCTTC	5.7
TAAATA	21.0	GTATAT	12.1	TACATA	12.1	TTTGAG	4.2	TCGATG	4.0	CTTTGA	5.4
TGTATA	20.4	TATAGA	12.1	CATATA	10.8	GGCTTG	4.1	CGATTG	3.9	TCTCCA	5.1
GTATAT	19.9	TATTTA	12.1	TATGTA	10.5	AGCATC	4.1	GTTGGT	3.8	CTCCAC	4.5

E — *S. pombe*

First PAS						Last PAS					
−60 ~ −31 nt		−30 ~ −1 nt		+1 ~ +30 nt		−60 ~ −31 nt		−30 ~ −1 nt		+1 ~ +30 nt	
TGCTTG	5.2	TTTGTC	6.6	TTTTTT	13.9	TAAATA	9.4	AATAAA	18.8	TAAGTA	8.0
TTCCCT	4.4	TTGATG	4.4	CTTTTT	7.3	ATAATA	7.1	TAATAA	15.7	TTGTAA	7.9
TTTGGA	3.7	TTTTGG	4.1	TTTGTT	6.8	AATATA	6.7	AAATAA	14.2	TGTAAG	7.5
GAAGAA	3.6	AGGTTC	3.8	TTCTTT	6.0	ATATTA	6.6	AATAAT	13.3	AAAGTA	7.2
TTGGAA	3.3	TTCTTT	3.8	TTTTTG	5.8	TTAATA	6.0	ATAATA	12.0	AAGTAA	7.1

F — PAS Strength: *S. cerevisiae*, *S. pombe*; PAS ... PAS ... PAS; AAA / AAA / AAA

G — PhastCons Score (First PAS, Last PAS); *S. cerevisiae* (left), *S. pombe* (right); x-axis: Position relative to PAS (nt)

Figure 3. Sequence motifs around PASs. (*A*) Nucleotide frequencies around the PASs of *S. cerevisiae* (*top*) and *S. pombe* (*bottom*). (*B*) Enriched 6-mers around the PASs of *S. cerevisiae* (*top*) and *S. pombe* (*bottom*). Enrichment values are *Z*-scores (Methods). Three regions around the PAS were analyzed separately, as indicated. (*C*) Summary of significant motifs around the PASs of *S. cerevisiae* and *S. pombe* based on 6-mer (*B*) and 4-mer (Supplemental Fig. S2) results. (*D*) Enriched 6-mers around the first and last 3′ UTR PASs of genes in *S. cerevisiae*. Values are −log$_{10}$(*P*), where *P* is based on the Fisher's exact test comparing first and last PAS sets. (*E*) As in *D*, except that *S. pombe* data are shown. (*F*) Summary of PAS strength versus location based on motif analyses. (*G*) PhastCons scores of the flanking regions of the first and last 3′ UTR PASs in *S. cerevisiae* (*left*) and *S. pombe* (*right*).

short 3′ UTR isoforms when grown in a nutrient-rich condition. This pattern is similar to that in higher species, in which proliferating cells tend to express shorter 3′ UTRs than quiescent, differentiated cells (Sandberg et al. 2008; Ji et al. 2009).

We next examined the Gene Ontology terms associated with genes with shortened and lengthened 3′ UTRs in cell growth. Interestingly, in *S. cerevisiae*, genes associated with translation, e.g., "translation," "ribosome," tended to have lengthened 3′ UTRs in rich media, and those associated with nitrogen metabolism, e.g., "organonitrogen compound biosynthetic process" and "cellular nitrogen compound biosynthetic process," tended to have 3′ UTRs either shortened or lengthened (Fig. 5C). Notably, genes associated with the nitrogen metabolism GO terms were also down-regulated at the expression levels (Supplemental Table S3), suggesting an interplay between APA and gene expression.

Several GO terms related to translation, such as "cytoplasmic translation" and "cytosolic ribosome," were significantly associated with genes with shortened 3′ UTRs in *S. pombe* (Fig. 5D). Interestingly, these GO terms were also the most significant for up-regulated genes at the expression levels (Supplemental Table S4), suggesting significant coupling between transcriptional regulation and APA. We thus directly compared genes with shortened 3′ UTRs and those with up-regulated gene expression. Indeed, a significant overlap was found between these two groups of genes (*P* < 4.9 × 10^{-17}, hypergeometric test) (Fig. 5E). Importantly, the common genes were significantly enriched for GO terms related to translation and ribosome (Fig. 5E), indicating that genes involved in protein synthesis utilize both transcriptional control and APA for gene regulation in cell growth.

A previous study found that genes involved in protein synthesis had significantly higher expression levels in proliferating fission yeast cells than quiescent cells (Marguerat et al. 2012). We thus asked how genes with 3′ UTR APA regulation in our data were regulated between proliferating and quiescent cells. Consistent with the result in Figure 5E, we found that genes with shortened 3′ UTRs in rich media had a much greater up-regulation of gene expression in cell proliferation (*P* = 2.8 × 10^{-8}, Kolmogorov–Smirnov [K-S] test) (Fig. 5F). Thus,
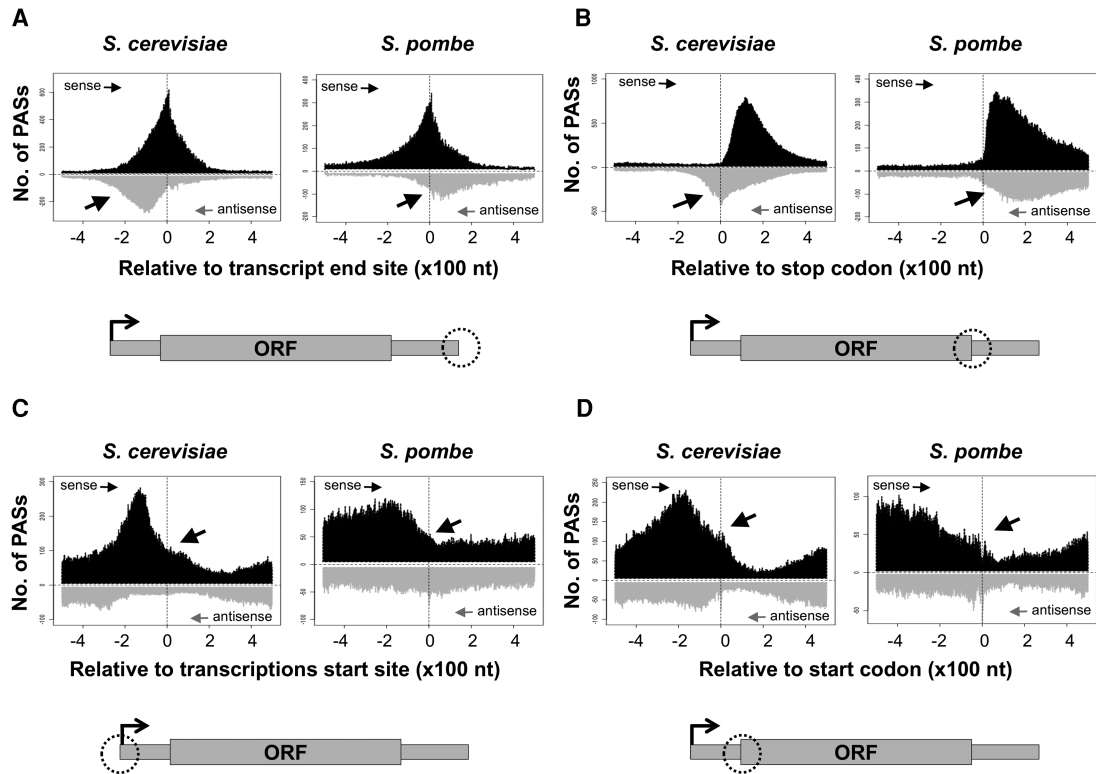
Figure 4. Metagene analysis of PASs around gene ends. Sense and antisense PASs around the transcript end site (*A*), stop codon (*B*), transcription start site (*C*), or start codon (*D*) in *S. cerevisiae* (*left*) and *S. pombe* (*right*). Significant differences between the two yeast species are highlighted by arrows.

3′ UTR shortening in *S. pombe* is coupled with up-regulation of gene expression during cell proliferation, especially for those involved in translation.

Regulation of APA by PA factors in *S. pombe*

PA factors play important roles in APA regulation in mammalian cells (Tian and Manley 2017). We next wanted to examine how expression level changes of PA factors may regulate APA in yeast cells. We focused on *S. pombe* strains deficient in either Pab2 or Pcf11. Pab2 is an RNA-binding protein that interacts with the poly(A) tail (Perreault et al. 2007). Ablation of its mammalian homolog PABPN1 leads to global 3′ UTR shortening (de Klerk et al. 2012; Jenal et al. 2012; Li et al. 2015), as well as increased abundance of antisense transcripts upstream of gene promoters (called uaRNA or PROMPTs) (Li et al. 2015). Pcf11 is a highly conserved PA factor across all eukaryotes (Mandel et al. 2008), although there is a substantial difference in protein size between yeast and mammal homologs. Pcf11 proteins in all species contain a Pol II CTD interaction domain, which is believed to be important for cotranscriptional 3′-end processing (Meinhart and Cramer 2004; Luo et al. 2006; Zhang and Gilmour 2006). We used a *pab2*-deleted strain (*pab2Δ*) grown in the rich media for 3′READS analysis. Because the *pcf11* gene is essential for cell viability in *S. pombe*, a conditional P_{nmt1}-*pcf11* strain in which the expression of *pcf11* is under the control of the thiamine-sensitive *nmt1* promoter was used. Accordingly, for 3′READS analysis of Pcf11-deficient cells, wild-type and P_{nmt1}-*pcf11* strains were grown in minimal media supplemented with thiamine to repress the expression of Pcf11 in P_{nmt1}-*pcf11* cells.

We found that the number of genes with lengthened 3′ UTRs was significantly greater than those with shortened 3′ UTRs in both P_{nmt1}-*pcf11* and *pab2Δ* strains, by 4.8- and 4.7-fold, respectively (Fig. 6A). In addition, we found that the APA sites consistently regulated in the two mutant strains significantly outnumbered those that were oppositely regulated (Fig. 6B). This result indicates that reduced expression of either Pcf11 or Pab2 led to regulation of a large common set of APA events. Notably, although repression of *pcf11* expression showed a similar effect on 3′ UTR length changes to that of *PCF11* knockdown in mammalian cells (Li et al. 2015), the result of *pab2* ablation in *S. pombe* is opposite to that of *PABPN1* knockdown in mammalian cells (de Klerk et al. 2012; Jenal et al. 2012; Li et al. 2015).

We also found that P_{nmt1}-*pcf11* and *pab2Δ* strains each had a unique set of regulated APA events (Fig. 6B, Venn diagram). We thus compared the extent of 3′ UTR lengthening in the two strains. We first separated APA events into five groups based on the distance between proximal and distal PASs (Fig. 6C), a parameter we previously found to be important for APA regulation (Li et al. 2015). Presumably, the distance, also called aUTR size, affects the competition for PA between two competing PASs. In both strains, the level of 3′ UTR lengthening, as determined by the difference in $\log_2$(dPAS/pPAS), was a function of aUTR size (Fig. 6C), i.e., the longer the aUTR size the more usage of distal PAS. This result is consistent with the notion that the mutant strains have decreased PA activity. Interestingly, the extent of 3′ UTR lengthening in P_{nmt1}-*pcf11* was substantially greater than that in *pab2Δ* across all aUTR size groups (green line consistently above orange line in Fig. 6C), indicating that inhibition of *pcf11* expression leads to a more severe reduction of PA activity than that of ablation of
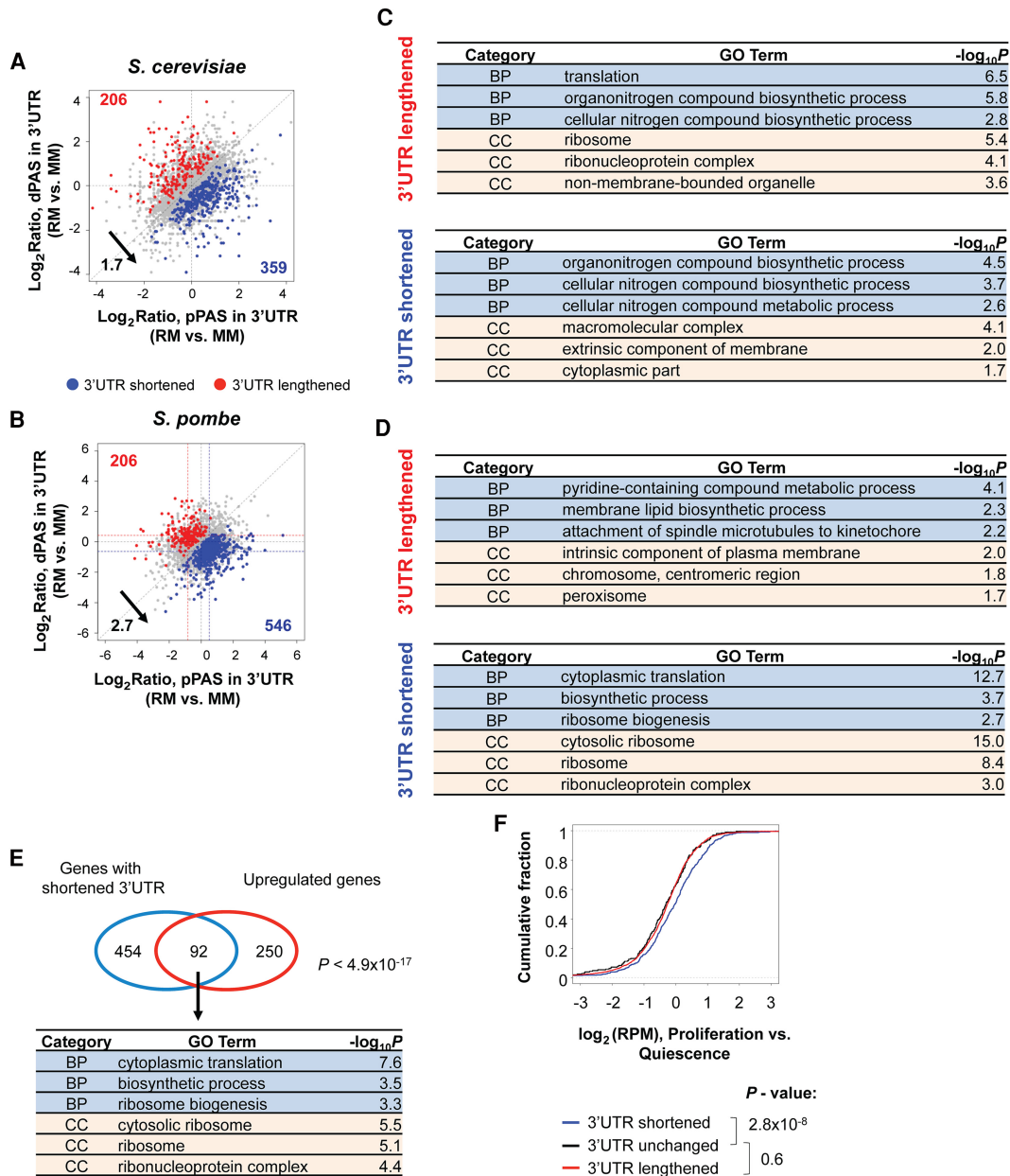
A *S. cerevisiae*

B *S. pombe*

C

3'UTR lengthened

Category	GO Term	-log$_{10}$P
BP	translation	6.5
BP	organonitrogen compound biosynthetic process	5.8
BP	cellular nitrogen compound biosynthetic process	2.8
CC	ribosome	5.4
CC	ribonucleoprotein complex	4.1
CC	non-membrane-bounded organelle	3.6

3'UTR shortened

Category	GO Term	-log$_{10}$P
BP	organonitrogen compound biosynthetic process	4.5
BP	cellular nitrogen compound biosynthetic process	3.7
BP	cellular nitrogen compound metabolic process	2.6
CC	macromolecular complex	4.1
CC	extrinsic component of membrane	2.0
CC	cytoplasmic part	1.7

D

3'UTR lengthened

Category	GO Term	-log$_{10}$P
BP	pyridine-containing compound metabolic process	4.1
BP	membrane lipid biosynthetic process	2.3
BP	attachment of spindle microtubules to kinetochore	2.2
CC	intrinsic component of plasma membrane	2.0
CC	chromosome, centromeric region	1.8
CC	peroxisome	1.7

3'UTR shortened

Category	GO Term	-log$_{10}$P
BP	cytoplasmic translation	12.7
BP	biosynthetic process	3.7
BP	ribosome biogenesis	2.7
CC	cytosolic ribosome	15.0
CC	ribosome	8.4
CC	ribonucleoprotein complex	3.0

E Genes with shortened 3'UTR — Upregulated genes

$P < 4.9 \times 10^{-17}$

Category	GO Term	-log$_{10}$P
BP	cytoplasmic translation	7.6
BP	biosynthetic process	3.5
BP	ribosome biogenesis	3.3
CC	cytosolic ribosome	5.5
CC	ribosome	5.1
CC	ribonucleoprotein complex	4.4

F

P - value:
— 3'UTR shortened ⎤ 2.8×10^{-8}
— 3'UTR unchanged ⎦
— 3'UTR lengthened   0.6

Figure 5. Alternative polyadenylation changes between cells grown in rich versus minimal media. (*A*) Scatter plot comparing expression changes of proximal PAS (pPAS) isoform (*x*-axis) between *S. cerevisiae* cells grown in rich media (RM) versus minimal media (MM) with that of distal PAS (dPAS) isoform (*y*-axis). Both pPAS and dPAS are in the 3' UTR. Genes with shortened 3' UTRs are highlighted in blue, and those with lengthened 3' UTRs are in red. Gene numbers for both types are shown, and their ratio (number of blue dots to number of red dots) is indicated. (*B*) As in *A*, except that data for *S. pombe* are shown. (*C,D*) Top Gene Ontology (GO) terms associated with *S. cerevisiae* (*C*) or *S. pombe* (*D*) genes with shortened 3' UTRs or lengthened 3' UTRs in RM versus MM. (BP) biological process; (CC) cellular component. *P*-value is based on the Fisher's exact test. (*E*) Venn diagram comparing the genes with shortened 3' UTRs and genes with up-regulated expression. The *P*-value (hypergeometric test) indicates the significance of overlap between two genes sets. The enriched GO terms are shown at the *bottom* with *P*-values indicated. (*F*) Distribution of gene expression changes in cell proliferation versus quiescence, as previously reported, for genes with different 3' UTR changes as analyzed in *B*. *P*-values (K-S test) indicating significance of difference between different gene sets are indicated.

pab2. However, we cannot rule out the possibility that this difference is caused by different growth conditions (rich media versus minimal media) for the two mutant strains.

Consistent with global up-regulation of distal PASs in PAB2- and Pcf11-deficient cells, we found that the percentage of PAS reads mapped to intergenic regions substantially increased in *pab2*Δ and P_{nmt1}-*pcf11* strains (10.6% and 14.3%, respectively), as compared to the wild-type strain (5 .3%) (Fig. 6D). We thus examined all PASs across the genome in P_{nmt1}-*pcf11* and *pab2*Δ strains. Significant increases of antisense PAS usages in the upstream regions of the TSS and start codon (Fig. 6E) could be discerned in both cells. Consistent with 3' UTR lengthening, inhibition of *pcf11* expression substantially reduced the height of the PAS peak around the TES or near the stop codon and caused up-
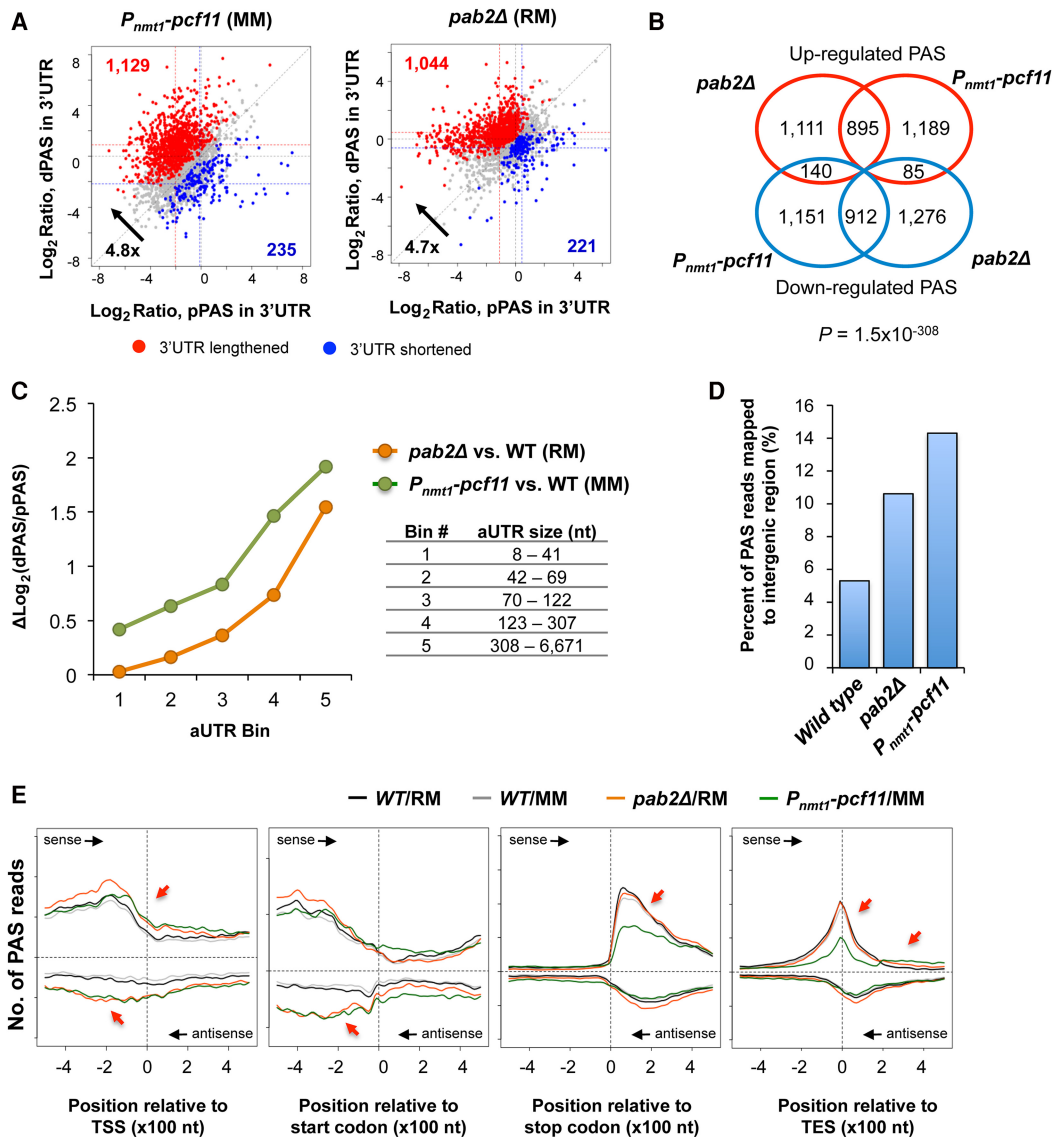
Figure 6. APA regulation by PA factors. (*A*) 3′ UTR APA analysis of Pcf11-deficient (P_{nmt1}-*pcf11*; *left*) and *pab2*-null (*pab2Δ*; *right*) strains of *S. pombe.* Each scatter plot compares expression change of proximal PAS isoform (pPAS, *x*-axis) between an analyzed strain and control, as well as that of distal PAS isoform (dPAS, *y*-axis) between the strains. Both pPAS and dPAS are in the 3′ UTR. Genes with shortened 3′ UTRs are highlighted in blue, and those with lengthened 3′ UTRs in red. Gene numbers for both types are shown, and their ratio (number of blue dots to number of red dots) is indicated. Note that P_{nmt1}-*pcf11* was grown in minimal media (MM) and *pab2Δ* in rich media (RM). (*B*) Venn diagram comparing significantly regulated PASs in P_{nmt1}-*pcf11* with those in *pab2Δ* cells. *P*-value (Fisher's exact test) is based on analysis of commonly regulated PASs. (*C*) Relationship between aUTR size (distance between proximal and distal PASs) and regulation of APA. The two most abundant 3′ UTR isoforms based on PAS reads were selected from each gene. APA regulation of each gene was based on the difference in log₂(dPAS/pPAS) between an analyzed strain and control. A higher value indicates greater up-regulation of the distal PAS isoform. Genes were divided into five groups based on aUTR size. The average value of each group was plotted. (*D*) Percentage of PAS reads mapped to intergenic regions in wild type, *pab2Δ*, and P_{nmt1}-*pcf11* strains. The data on wild type were based on both cells grown in RM and in MM. (*E*) Metagene analysis of PASs in different strains of *S. pombe.* Significantly altered PAS peaks are highlighted with red arrows.

regulation of PAS reads further downstream. This trend, however, was not observed in the *pab2Δ* mutant, further indicating that 3′ UTR lengthening elicited by ablation of *pab2* is much milder than by suppression of *pcf11* expression.

Discussion

In this study, we systematically identified PASs and their associated sequence motifs in *S. cerevisiae* and *S. pombe*, and compared several aspects of APA in these two species. PASs in both yeasts are closely

spaced to one another, forming an "end zone" as previously proposed by Struhl and colleagues (Moqtaderi et al. 2013). Notably, our PAS identification method is based on a robust approach aimed to find the optimal site-to-site distance to distinguish heterogeneous cleavage sites from PASs. The 7-nt cutoff identified in this study is largely consistent with those used in earlier studies (Mata 2013; Schlackow et al. 2013). Nevertheless, due to the diffusive nature of PAS motifs, it is possible that some heterogeneous cleavage sites were separated into different PASs and thus treated as APA sites. This potential technical issue, however, can be

mitigated by using the first and last PASs for analysis, as we did in this study.

Consistent with previous reports (Graber et al. 2002; Mata 2013; Schlackow et al. 2013), the PASs in the budding yeast are surrounded with sequence motifs distinct from those in the fission yeast. The motifs appear to be more loosely defined around the PASs in *S. cerevisiae* than in *S. pombe*. The PAS motifs of *S. pombe* resemble those of mammals (Hu et al. 2005) to a much greater extent than those of *S. cerevisiae*. Some differences, however, are still notable. The upstream UGUA motif and downstream UGUG motif in mammals were absent from *S. pombe*. In addition, AUUAAA is a very prominent motif in mammals but not in *S. pombe*. Instead, AAUGAA is highly enriched in the fission yeast, suggesting distinct sequence specificity of the PA complex. Interestingly, the PAS regions in both species contain motifs that resemble those of NNS complex binding sites, albeit with different sequences, i.e., CU-rich motifs in the budding yeast and the GUA motif in the fission yeast. This result implicates a general role of the NNS pathway in 3′-end processing of mRNA genes, probably as a fail-safe mechanism for termination of transcription (Rondon et al. 2009; Lemay and Bachand 2015).

We found that, thanks to the differential PAS placement on the antisense strand, *S. cerevisiae* and *S. pombe* differ considerably in their antisense RNA landscapes. It is possible that due to differences in handling double-stranded RNAs in *S. cerevisiae* versus *S. pombe*, e.g., the RNA inference mechanism exists in *S. pombe* but not in *S. cerevisiae*, the budding yeast can tolerate antisense transcripts better than the fission yeast. It remains to be seen, however, how double-stranded RNA structures formed by sense and antisense transcripts at the 3′ end may affect mRNA metabolisms, such as stability and translation.

We revealed a number of key APA features that are different between budding and fission yeasts, despite that both species have a high frequency of APA events among genes. *S. pombe* has a longer alternative 3′ UTR sequence than *S. cerevisiae*, indicating expansion of 3′ UTRs in the fission yeast, which may allow greater gene regulation through 3′ UTR motifs. Importantly, distal PASs in *S. pombe* tend to be associated with stronger PAS motifs than proximal ones, consistent with the higher abundances of long 3′ UTR isoforms than short ones. This feature is opposite in *S. cerevisiae* but is similar in higher species (Tian et al. 2005). Consistently, the sequences upstream of distal PASs in *S. pombe* are more evolutionarily conserved relative to the corresponding sequences in *S. cerevisiae*. Therefore, the APA configuration involving strong distal PASs and weak proximal PASs, which is common in higher species and can presumably ensure proper termination and efficient regulation of APA, is also established in *S. pombe*.

In both budding and fission yeasts, short 3′ UTR isoforms are up-regulated relative to long isoforms when cells are growing in nutrient-rich media. However, different gene groups are affected by this mechanism, highlighting the distinct functional consequences of APA in these two species. Significantly, 3′ UTR shortening in *S. pombe* coordinates with up-regulation of expression for genes involved in translation during cell proliferation, a key part of the regulatory program for cell growth (Marguerat et al. 2012). Therefore, different APA regulatory mechanisms and consequent post-transcriptional controls can substantially contribute to metabolic differences between budding and fission yeasts.

Our data indicate that regulation of core PA factors in yeast can significantly impact APA events genome-wide, similar to mammalian cells (Li et al. 2015). However, although the role of Pcf11 is consistent between yeast and higher species, Pab2 in *S.*

pombe has an opposite effect on APA as compared to its mammalian homolog PABPN1, whose role in APA has also been implicated in the etiology of oculopharyngeal muscular dystrophy (OPMD), an autosomal-dominant disease caused by a triplet repeat expansion in the *PABPN1* gene (de Klerk et al. 2012; Jenal et al. 2012; Banerjee et al. 2013). This finding appears consistent with a recent report indicating that *S. pombe* Pab2 and human PABPN1 have different biochemical properties in stimulation of poly(A) polymerase for polyadenylation. Future work is needed to unravel the stark difference between Pab2 and PABPN1 in PAS choice.

Methods

S. cerevisiae samples

The haploid budding yeast strain W303a was used to generate RNA from rich and minimal media cultures. For rich medium, YPD supplemented with 120 mg/mL adenine and 200 mg/mL tryptophan was used (XY). For minimal medium, synthetic complete (SC) medium was used. Overnight cultures of W303a, in both XY and SC media, were used to inoculate 50 mL of the same media to an OD_{600} of 0.2. Cells were then grown to an OD_{600} of 0.6–0.8, and RNA extraction was performed according to Amberg et al. (2006). Briefly, cells were harvested and chilled in AE buffer (50 mM sodium acetate, 10 mM EDTA pH. 8.0), and then frozen and thawed three times in AE with 1% SDS and phenol using a dry ice/ethanol bath. The aqueous layer was precipitated, washed, and then resuspended in water for DNase I treatment (1 h) after absorbance was measured.

S. pombe samples

Fission yeast cells were grown at 30°C in rich YES media (3% glucose, 0.5% yeast extract, supplemented with adenine, histidine, leucine, and uracil) in the case of wild-type (WT) and *pab2* deletion (*pab2Δ*) strains (Perreault et al. 2007) or in Edinburgh minimal medium (EMM) supplemented with adenine, histidine, leucine, and uracil in the case of conditional *pcf11* ablation (P$_{nmt1}$-*pcf11*) strain (Larochelle et al. 2017). Conditional strain, in which the genomic copy of the essential gene *pcf11* is expressed from the thiamine-repressive (P$_{nmt1}$) promoter, was repressed by the addition of 60 μM thiamine in the EMM media for 12–15 h. In parallel, WT cells were also grown in EMM media treated with thiamine as a control. Cells were collected at OD_{600} of 0.5–0.7, and RNA extractions were performed by the hot-acid phenol method.

3′READS

3′READS was carried out as previously described (Hoque et al. 2014). Briefly, total RNA was subjected to one round of poly(A) selection using oligo(dT) beads (NEB), followed by fragmentation on-bead with RNase III (NEB). Poly(A)-containing RNA fragments were isolated using the MyOne streptavidin C1 beads (Invitrogen) coated with a 5′ biotinylated chimeric T_{45} U_5 oligo (Sigma), followed by washing, and then elution by removal of poly(A) with RNase H. Eluted RNA fragments were purified by Phenol:Chloroform extraction and ethanol precipitation, followed by sequential ligations to a 5′-adenylated 3′ adapter (5′-rApp/NNNNGATCG TCGGACTGTAGAACTCTGAAC/3ddC) with the truncated T4 RNA ligase II (Bioo Scientific) and to a 5′ adapter (5′-GUUCAGAG UUCUACAGUCCGACGAUC) with T4 RNA ligase I (NEB). The resultant RNA was reverse transcribed to cDNA with SuperScript III (Invitrogen), followed by 12 cycles of PCR amplification with Phusion high-fidelity polymerase (NEB). cDNA libraries were sequenced on an Illumina HiSeq (1×50 bases).

Data analysis

3′READ data were analyzed as previously described (Hoque et al. 2014). Briefly, 3′READS reads were mapped to yeast genomes (*S. cerevisiae*: EF4, Ensembl v16; *S. pombe*: ASM294v2, Ensembl v17) using Bowtie 2 (local mode) (Langmead and Salzberg 2012). Gene models were obtained from the Ensembl databases (Aken et al. 2017). The transcription start sites (TSSs) of genes in *S. cerevisiae* were previously annotated by Xu et al. (2009). To obtain more accurate annotations of the 3′-end region, the annotated 3′ ends of genes were further extended by strand-specific RNA-seq data from previous studies (GSE60181 from NCBI GEO, and E-MTAB-154 and E-MTAB-1824 from ArrayExpress) (Marguerat et al. 2012; Schlackow et al. 2013; Rosonina et al. 2014). Extension was based on continuous RNA-seq read coverage and did not go beyond the annotated TSS of the downstream gene. Uniquely mapped reads (with MAPQ score > 10) that had at least two additional 5′ Ts after genome alignment were considered as PAS reads. For APA analysis, we additionally required that the abundance level of each PAS isoform was ≥5% of all isoforms in a gene. APA analysis comparing proximal and distal PASs was based on the Fisher's exact test. Gene expression level of a gene was calculated as the total number of PAS reads assigned to the gene and was measured by reads per million (RPM) uniquely mapped PAS reads. Significance of gene regulation was based on *P*-value ≤ 0.001 (Fisher's exact test).

Separating APA from heterogeneous cleavage

We hypothesized that the distance between adjacent cleavage sites fell into two types: within a PAS cluster or between two PAS clusters. We thus used the expectation maximization (EM) algorithm in the mixtools package of R programming (Benaglia et al. 2009) to identify two distribution modes in the data. This was carried out first by generating a distribution of all distances (log$_2$-based) between adjacent cleavage sites (CSs), and then the mixtools program was used to find two modes in the distribution using default parameters. This method resulted in a cross point of ~7 nt between the two distribution modes for both yeast species. Because PAS clusters in the mouse genome are well separated (24 nt), and two modes can be easily identified (Tian et al. 2005), we used mouse PASs we previously identified (Hoque et al. 2013) as a control. Using the same parameters, the program found 24 nt as the cutoff for mouse PAS clusters, validating the method and parameters used for yeast PASs. We thus used this number to separate adjacent CSs, i.e., those separated by >7 nt were separated into two PASs, whereas those by ≤7 were clustered together into one PAS. The CS with the highest number of reads in a PAS cluster was chosen as the representative site for the cluster.

Cis element analysis

We used the PROBE method to identify significantly enriched sequence motifs around PASs, which we previously used for human PAS analysis (Hu et al. 2005). Briefly, for each *k*-mer (6-mer or 4-mer), we enumerated its frequency (observed value) in a defined region of PAS and compared it with the frequency (expected value) using randomized sequences of the same region based on the first order Markov chain model. The enrichment score was calculated as a *Z*-score based on the difference between the observed and expected frequencies. Three regions around the PAS were analyzed, including −60 to −31 nt, −30 to −1 nt, and +1 to +30 nt, with the PAS set at position 0. For comparison between different PAS sets, the Fisher's exact test was used to examine whether a *k*-mer was enriched for or depleted from each set.

PhastCons score analysis

The PhastCons scores for sequence conservation of *S. cerevisiae* and of *M. musculus* were obtained from UCSC Genome Browser (*S. cerevisiae*: phastCons7way; *M. musculus*: phastCons30way). The PhastCons scores of *S. pombe* were obtained from the *Schizosaccharomyces* Comparative Genome Project (Rhind et al. 2011). PhastCons scores around PASs were plotted to show the general conservation profiles. Stop codon and intergenic regions were used as references. The median PhastCons scores for each position were used for plotting. The lowest function in R with default setting was used to smooth lines.

Gene Ontology analysis

Gene Ontology (GO) annotation of genes was obtained from the Ensembl database (Aken et al. 2017). GO entries were tested for significance of association with regulated APA genes using the hypergeometric test. GO terms associated with more than 2000 genes were considered too generic and were discarded. To remove redundancy, each reported GO term was required to have at least 30% of genes that had not been associated with another term with a more significant *P*-value.

Data access

All sequencing data from this study have been submitted to the NCBI Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo/) under accession number GSE95139.

References

Aken BL, Achuthan P, Akanni W, Amode MR, Bernsdorff F, Bhai J, Billis K, Carvalho-Silva D, Cummins C, Clapham P, et al. 2017. Ensembl 2017. *Nucleic Acids Res* **45:** D635–D642.
Amberg DC, Burke DJ, Strathern JN. 2006. Yeast RNA isolation: small-scale. *CSH Protoc* doi: 10.1101/pdb.prot4155.
Arndt KM, Reines D. 2015. Termination of transcription of short noncoding RNAs by RNA polymerase II. *Annu Rev Biochem* **84:** 381–404.
Banerjee A, Apponi LH, Pavlath GK, Corbett AH. 2013. PABPN1: molecular function and muscle disease. *FEBS J* **280:** 4230–4250.
Benaglia T, Chauveau D, Hunter DR, Young DS. 2009. mixtools: an R package for analyzing mixture models. *J Stat Softw* **32:** 29.
Cheng Y, Miura RM, Tian B. 2006. Prediction of mRNA polyadenylation sites by support vector machine. *Bioinformatics* **22:** 2320–2325.
de Klerk E, Venema A, Anvar SY, Goeman JJ, Hu O, Trollet C, Dickson G, den Dunnen JT, van der Maarel SM, Raz V, et al. 2012. Poly(A) binding protein nuclear 1 levels affect alternative polyadenylation. *Nucleic Acids Res* **40:** 9089–9101.
Dujon B. 2010. Yeast evolutionary genomics. *Nat Rev Genet* **11:** 512–524.
Elkon R, Ugalde AP, Agami R. 2013. Alternative cleavage and polyadenylation: extent, regulation and function. *Nat Rev Genet* **14:** 496–506.
Flavell SW, Kim TK, Gray JM, Harmin DA, Hemberg M, Hong EJ, Markenscoff-Papadimitriou E, Bear DM, Greenberg ME. 2008. Genome-wide analysis of MEF2 transcriptional program reveals synaptic target genes and neuronal activity-dependent polyadenylation site selection. *Neuron* **60:** 1022–1038.
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.

Graber JH, Cantor CR, Mohr SC, Smith TF. 1999. *In silico* detection of control signals: mRNA 3′-end-processing sequences in diverse species. *Proc Natl Acad Sci* **96:** 14055–14060.

Graber JH, McAllister GD, Smith TF. 2002. Probabilistic prediction of *Saccharomyces cerevisiae* mRNA 3′-processing sites. *Nucleic Acids Res* **30:** 1851–1858.

Graber JH, Nazeer FI, Yeh PC, Kuehner JN, Borikar S, Hoskinson D, Moore CL. 2013. DNA damage induces targeted, genome-wide variation of poly(A) sites in budding yeast. *Genome Res* **23:** 1690–1703.

Hoffman CS, Wood V, Fantes PA. 2015. An ancient yeast for young geneticists: a primer on the *Schizosaccharomyces pombe* model system. *Genetics* **201:** 403–423.

Hollerer I, Curk T, Haase B, Benes V, Hauer C, Neu-Yilik G, Bhuvanagiri M, Hentze MW, Kulozik AE. 2016. The differential expression of alternatively polyadenylated transcripts is a common stress-induced response mechanism that modulates mammalian mRNA expression in a quantitative and qualitative fashion. *RNA* **22:** 1441–1453.

Hoque M, Ji Z, Zheng D, Luo W, Li W, You B, Park JY, Yehia G, Tian B. 2013. Analysis of alternative cleavage and polyadenylation by 3′ region extraction and deep sequencing. *Nat Methods* **10:** 133–139.

Hoque M, Li W, Tian B. 2014. Accurate mapping of cleavage and polyadenylation sites by 3′ region extraction and deep sequencing. *Methods Mol Biol* **1125:** 119–129.

Hu J, Lutz CS, Wilusz J, Tian B. 2005. Bioinformatic identification of candidate *cis*-regulatory elements involved in human mRNA polyadenylation. *RNA* **11:** 1485–1493.

Jenal M, Elkon R, Loayza-Puch F, van Haaften G, Kühn U, Menzies FM, Oude Vrielink JA, Bos AJ, Drost J, Rooijers K, et al. 2012. The poly(A)-binding protein nuclear 1 suppresses alternative cleavage and polyadenylation sites. *Cell* **149:** 538–553.

Ji Z, Tian B. 2009. Reprogramming of 3′ untranslated regions of mRNAs by alternative polyadenylation in generation of pluripotent stem cells from different cell types. *PLoS One* **4:** e8419.

Ji Z, Lee JY, Pan Z, Jiang B, Tian B. 2009. Progressive lengthening of 3′ untranslated regions of mRNAs by alternative polyadenylation during mouse embryonic development. *Proc Natl Acad Sci* **106:** 7028–7033.

Kim Guisbert KS, Li H, Guthrie C. 2007. Alternative 3′ pre-mRNA processing in *Saccharomyces cerevisiae* is modulated by Nab4/Hrp1 in vivo. *PLoS Biol* **5:** e6.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9:** 357–359.

Larochelle M, Hunyadkürti J, Bachand F. 2017. Polyadenylation site selection: linking transcription and RNA processing via a conserved carboxy-terminal domain (CTD)-interacting protein. *Curr Genet* **63:** 195–199.

Legendre M, Gautheret D. 2003. Sequence determinants in human polyadenylation site selection. *BMC Genomics* **4:** 7.

Lemay JF, Bachand F. 2015. Fail-safe transcription termination: because one is never enough. *RNA Biol* **12:** 927–932.

Lemay JF, Marguerat S, Larochelle M, Liu X, van Nues R, Hunyadkürti J, Hoque M, Tian B, Granneman S, Bähler J, et al. 2016. The Nrd1-like protein Seb1 coordinates cotranscriptional 3′ end processing and polyadenylation site selection. *Genes Dev* **30:** 1558–1572.

Li W, You B, Hoque M, Zheng D, Luo W, Ji Z, Park JY, Gunderson SI, Kalsotra A, Manley JL, et al. 2015. Systematic profiling of poly(A)+ transcripts modulated by core 3′ end processing and splicing factors reveals regulatory rules of alternative cleavage and polyadenylation. *PLoS Genet* **11:** e1005166.

Luo W, Johnson AW, Bentley DL. 2006. The role of Rat1 in coupling mRNA 3′-end processing to transcription termination: implications for a unified allosteric–torpedo model. *Genes Dev* **20:** 954–965.

Mandel CR, Bai Y, Tong L. 2008. Protein factors in pre-mRNA 3′-end processing. *Cell Mol Life Sci* **65:** 1099–1122.

Marguerat S, Schmidt A, Codlin S, Chen W, Aebersold R, Bähler J. 2012. Quantitative analysis of fission yeast transcriptomes and proteomes in proliferating and quiescent cells. *Cell* **151:** 671–683.

Mata J. 2013. Genome-wide mapping of polyadenylation sites in fission yeast reveals widespread alternative polyadenylation. *RNA Biol* **10:** 1407–1414.

Mayr C. 2016. Evolution and biological roles of alternative 3′UTRs. *Trends Cell Biol* **26:** 227–237.

Meinhart A, Cramer P. 2004. Recognition of RNA polymerase II carboxy-terminal domain by 3′-RNA-processing factors. *Nature* **430:** 223–226.

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* **110:** 11073–11078.

Nam DK, Lee S, Zhou G, Cao X, Wang C, Clark T, Chen J, Rowley JD, Wang SM. 2002. Oligo(dT) primer generates a high frequency of truncated cDNAs through internal poly(A) priming during reverse transcription. *Proc Natl Acad Sci* **99:** 6152–6156.

Perreault A, Lemieux C, Bachand F. 2007. Regulation of the nuclear poly(A)-binding protein by arginine methylation in fission yeast. *J Biol Chem* **282:** 7552–7562.

Porrua O, Libri D. 2015. Transcription termination and the control of the transcriptome: why, where and how to stop. *Nat Rev Mol Cell Biol* **16:** 190–202.

Proudfoot NJ. 2016. Transcriptional termination in mammals: Stopping the RNA polymerase II juggernaut. *Science* **352:** aad9926.

Rhind N, Chen Z, Yassour M, Thompson DA, Haas BJ, Habib N, Wapinski I, Roy S, Lin MF, Heiman DI, et al. 2011. Comparative functional genomics of the fission yeasts. *Science* **332:** 930–936.

Rondon AG, Mischo HE, Kawauchi J, Proudfoot NJ. 2009. Fail-safe transcriptional termination for protein-coding genes in *S. cerevisiae*. *Mol Cell* **36:** 88–98.

Rosonina E, Yurko N, Li W, Hoque M, Tian B, Manley JL. 2014. Threonine-4 of the budding yeast RNAP II CTD couples transcription with Htz1-mediated chromatin remodeling. *Proc Natl Acad Sci* **111:** 11924–11931.

Sandberg R, Neilson JR, Sarma A, Sharp PA, Burge CB. 2008. Proliferating cells express mRNAs with shortened 3′ untranslated regions and fewer microRNA target sites. *Science* **320:** 1643–1647.

Schlackow M, Marguerat S, Proudfoot NJ, Bähler J, Erban R, Gullerova M. 2013. Genome-wide analysis of poly(A) site selection in *Schizosaccharomyces pombe*. *RNA* **19:** 1617–1631.

Shepard PJ, Choi EA, Lu J, Flanagan LA, Hertel KJ, Shi Y. 2011. Complex and dynamic landscape of RNA polyadenylation revealed by PAS-Seq. *RNA* **17:** 761–772.

Shi Y. 2012. Alternative polyadenylation: new insights from global analyses. *RNA* **18:** 2105–2117.

Shi Y, Manley JL. 2015. The end of the message: multiple protein-RNA interactions define the mRNA polyadenylation site. *Genes Dev* **29:** 889–897.

Shi Y, Di Giammartino DC, Taylor D, Sarkeshik A, Rice WJ, Yates JR III, Frank J, Manley JL. 2009. Molecular architecture of the human pre-mRNA 3′ processing complex. *Mol Cell* **33:** 365–376.

Tian B, Graber JH. 2012. Signals for pre-mRNA cleavage and polyadenylation. *Wiley Interdiscip Rev RNA* **3:** 385–396.

Tian B, Manley JL. 2017. Alternative polyadenylation of mRNA precursors. *Nat Rev Mol Cell Biol* **18:** 18–30.

Tian B, Hu J, Zhang H, Lutz CS. 2005. A large-scale analysis of mRNA polyadenylation of human and mouse genes. *Nucleic Acids Res* **33:** 201–212.

Weng L, Li Y, Xie X, Shi Y. 2016. Poly(A) code analyses reveal key determinants for tissue-specific mRNA alternative polyadenylation. *RNA* **22:** 813–821.

Wilkening S, Pelechano V, Järvelin AI, Tekkedil MM, Anders S, Benes V, Steinmetz LM. 2013. An efficient method for genome-wide polyadenylation site mapping and RNA quantification. *Nucleic Acids Res* **41:** e65.

Xu Z, Wei W, Gagneur J, Perocchi F, Clauder-Münster S, Camblong J, Guffanti E, Stutz F, Huber W, Steinmetz LM. 2009. Bidirectional promoters generate pervasive transcription in yeast. *Nature* **457:** 1033–1037.

Xu Z, Wei W, Gagneur J, Clauder-Münster S, Smolik M, Huber W, Steinmetz LM. 2011. Antisense expression increases gene expression variability and locus interdependency. *Mol Syst Biol* **7:** 468.

Yoon OK, Brem RB. 2010. Noncanonical transcript forms in yeast and their regulation during environmental stress. *RNA* **16:** 1256–1267.

Zhang Z, Gilmour DS. 2006. Pcf11 is a termination factor in *Drosophila* that dismantles the elongation complex by bridging the CTD of RNA polymerase II to the nascent transcript. *Mol Cell* **21:** 65–74.