



Published in final edited form as:

Cell Rep. 2019 February 12; 26(7): 1919–1933.e5. doi:10.1016/j.celrep.2019.01.051.

## Ipa1 Is an RNA Polymerase II Elongation Factor that Facilitates Termination by Maintaining Levels of the Poly(A) Site Endonuclease Ysh1

Erika L. Pearson<sup>1,3,4</sup>, Joel H. Graber<sup>2,3</sup>, Susan D. Lee<sup>1</sup>, Kristoph S. Naggert<sup>2</sup>, Claire L. Moore<sup>1,5,\*</sup>

<sup>1</sup>Department of Development, Molecular, and Chemical Biology and Sackler School of Graduate Biomedical Science, Tufts University School of Medicine, Boston, MA 02111, USA

<sup>2</sup>Mount Desert Island Biological Laboratory, Bar Harbor, ME 04609, USA

<sup>3</sup>These authors contributed equally

<sup>4</sup>Present address: Sigilon Therapeutics, Cambridge, MA 02142, USA

<sup>5</sup>Lead Contact

### SUMMARY

The yeast protein Ipa1 was recently discovered to interact with the Ysh1 endonuclease of the prem-RNA cleavage and polyadenylation (C/P) machinery, and Ipa1 mutation impairs 3' end processing. We report that Ipa1 globally promotes proper transcription termination and poly(A) site selection, but with variable effects on genes depending upon the specific configurations of polyadenylation signals. Our findings suggest that the role of Ipa1 in termination is mediated through interaction with Ysh1, since Ipa1 mutation leads to decrease in Ysh1 and poor recruitment of the C/P complex to a transcribed gene. The Ipa1 association with transcriptionally active chromatin resembles that of elongation factors, and the mutant shows defective Pol II elongation kinetics *in vivo*. Ysh1 overexpression in the Ipa1 mutant rescues the termination defect, but not the mutant's sensitivity to 6-azauracil, an indicator of defective elongation. Our findings support a model in which an Ipa1/Ysh1 complex helps coordinate transcription elongation and 3' end processing.

### In Brief

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

\*Correspondence: [claire.moore@tufts.edu](mailto:claire.moore@tufts.edu).

#### AUTHOR CONTRIBUTIONS

Conceptualization, E.L.P. and C.L.M.; Methodology, E.L.P., C.L.M., and J.H.G.; Software, K.S.N. and J.H.G.; Formal Analysis, K.S.N. and J.H.G.; Investigation, E.L.P., S.D.L., and C.L.M.; Resources, J.H.G.; Data Curation, J.H.G.; Writing – Original Draft, E.L.P., C.L.M., and J.H.G.; Writing – Review & Editing, E.L.P., C.L.M., and J.H.G.; Project Administration, C.L.M.; Supervision, C.L.M.; Funding Acquisition, C.L.M.

#### DECLARATION OF INTERESTS

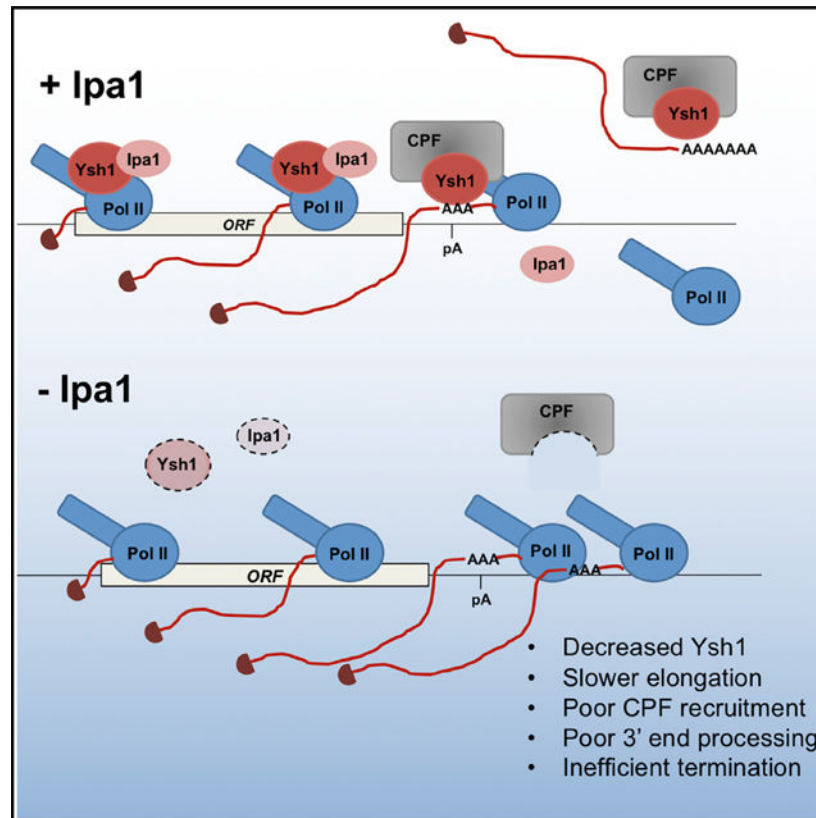
The authors declare no competing interests.

#### SUPPLEMENTAL INFORMATION

Supplemental Information includes three tables and can be found with this article online at <https://doi.org/10.1016/j.celrep.2019.01.051>.

The essential, uncharacterized Ipa1 protein was recently discovered to interact with the Ysh1 endonuclease of the pre-mRNA cleavage and polyadenylation machinery. Pearson et al. propose that the Ipa1/Ysh1 interaction provides the cell with a means to coordinate and regulate transcription elongation with 3' end processing in accordance with the cell's needs.

## Graphical Abstract



## INTRODUCTION

RNA polymerase II (Pol II) is responsible for synthesis of eukaryotic mRNA and several classes of non-coding RNAs. The Pol II transcription cycle on protein-coding genes consists of three discrete stages (initiation, elongation, and termination) that are distinguished from one another by the recruitment of specific sets of transcription co-factors, which differentially influence the behavior of Pol II and chromatin (Guo and Price, 2013; Porrua et al., 2016; Shandilya and Roberts, 2012). Upon productive initiation, the Pol II transcription complex enters the elongation stage with the recruitment of factors that assist in maintaining highly processive Pol II transcription throughout the body of a gene. Once the poly(A) site at the gene's end is transcribed, Pol II experiences a shift in processivity that is associated with the eviction of elongation factors and recruitment of termination and pre-mRNA 3' end-processing factors. The 3' end factors promote dismantling of the transcription complex as well as polyadenylation of mRNAs of almost all genes in the termination stage of the transcription cycle. Efficient removal of Pol II from chromatin requires cleavage of RNA at

the poly(A) site by the yeast Ysh1 protein or its mammalian homolog CPSF73 to create an entry site for the Rat1/Xrn2 exonuclease that co-transcriptionally degrades the polymerase-associated RNA (Eaton et al., 2018; Kim et al., 2004b).

Analysis of the association and dissociation of individual transcription factors at a genome-wide level has enhanced our understanding of transcription cycle control and has provided an intricate picture of rising and falling gradients of transcription factors that associate with Pol II in a tightly coordinated temporal-spatial manner (Kim et al., 2010; Mayer et al., 2010). Transcriptome-wide studies have revealed additional layers of regulation by illuminating the widespread use of alternative polyadenylation sites to produce mRNAs subject to different post-transcriptional fates (Elkon et al., 2013; Tian and Manley, 2017). Manipulating the levels of core components of the cleavage and polyadenylation (C/P) complex affects these poly(A) site choices in ways that resemble normal changes seen in different cell states. Our understanding about how sequences around poly(A) sites determines susceptibility to alterations in specific C/P subunits is growing but incomplete.

In addition, newly discovered factors are enriching our perspective of the transcriptional and co-transcriptional regulation paradigm. For example, mutation of the previously uncharacterized, essential yeast *IPA1* (Important for PolyAdenylation) gene causes defects in pre-mRNA 3' end processing in addition to a slow growth phenotype (Costanzo et al., 2016). Ipa1 physically interacts with only the Ysh1 and Mpe1 subunits of the 15-subunit cleavage and polyadenylation factor (CPF), and *IPA1* mutants share many genetic interactions in common with genes involved in mRNA 3' end processing (Casañal et al., 2017; Costanzo et al., 2016). Ipa1 shares sequence homology with HECT-like E3 ubiquitin-protein ligases (Lutz et al., 2018) and has orthologs in higher eukaryotes, indicating that it is evolutionarily conserved. UBE3D, the human ortholog, may be involved in cell-cycle regulation (Kobirumaki et al., 2005) and the immune response (Huang et al., 2015; Offenbacher et al., 2016). However, the mechanism by which Ipa1 affects mRNA synthesis has not been clarified.

In our current work, we show that Ipa1 participates in both 3' end formation and transcript elongation and importantly could facilitate coordination of these activities. Ipa1 promotes transcription termination and co-transcriptional 3' end processing by maintaining wild-type levels of Ysh1, the conserved subunit of CPF that cleaves pre-mRNA, and by contributing to proper CPF recruitment. Loss of Ipa1 function affects most mRNA genes to some extent but most strongly affects those with specific configurations of polyadenylation signals. Our findings also reveal that Ipa1 mutation results in deficient transcription elongation. Restoration of Ysh1 levels rescues some but not all of the defects of the *ipa1-1* mutant, supporting a larger role for Ipa1 in addition to maintenance of Ysh1 expression. We propose a model in which Ipa1 likely functions as an elongation factor while also serving as a molecular chaperone to deliver the Ysh1 endonuclease to the poly(A) site for 3' end processing and transcription termination.

## RESULTS

### Loss of Ipa1 Function Leads to Transcriptome-wide Reduction in Polyadenylation Activity and a Correlated Increased Average Length of mRNAs

To determine global changes in poly(A) processing caused by the *ipa1-1* mutation, we acquired the previously published genome-wide poly(A) site mapping data for *ipa1-1* and wild-type (WT) cells (Costanzo et al., 2016). This study showed a significant bias toward use of downstream poly(A) sites in the *ipa1-1* mutant, but features that determined an Ipa1-responsive site were not evaluated. Because such information could give in-sights into how use of alternative poly(A) sites is regulated, we re-analyzed these data as described below. All comparisons below were made based upon three *ipa1-1* samples and four WT samples. Previous genomic analyses of poly(A) sites in *S. cerevisiae* have shown that the majority map to the 3' UTR (Graber et al., 2013; Johnson et al., 2011; Liu et al., 2017; Ozsolak et al., 2010; Yoon and Brem, 2010), and we focused our analysis on this category. For statistical robustness, we restricted analysis of 3' UTR features to 4,377 genes that exceeded an arbitrary cutoff of at least 250 sequence tags summed across all seven samples.

We first characterized changes in the poly(A) site positions for each gene in order to derive the average 3' UTR length for that gene. After calculating a genotype-specific weighted average 3' UTR length for each gene (STAR Methods), we used t test (2-sided, unequal variance) to compare the average 3' UTR lengths of the *ipa1-1* samples to the WT samples on a gene-by-gene basis. More than half of the genes (2,399) passed a false discovery rate (FDR) threshold of <0.2. Of these, 2,367 showed increased 3' UTR length in *ipa1-1*, while only 32 were decreased (Figure 1A), indicating that the dominant effect of the *ipa1-1* mutation is a general extension in transcript length. This change is also evident when the transcriptome-wide distribution of 3' UTR lengths is plotted for mutant and WT (Figure 1B). The 3' UTR lengths extend from a WT median length of 124 nt to an *ipa1-1* median of 138 nt. Similarly, the average length increased from 148 nt in WT to 164 nt in *ipa1-1* samples. The measured WT values are consistent with previous studies (Graber et al., 2013; Liu et al., 2017). This analysis indicates that the *ipa1-1* mutation results in an extension of the 3' UTR length of over half of all genes on average.

We also used the poly(A) sequence data to calculate how site-specific polyadenylation processing probability changed with the *ipa1-1* mutation. In brief, our model presumes 5'-to-3' processing, such that each site's processing probability can be estimated from the ratio of tag counts at the site to sum of tag counts for all sites further downstream within the same gene. We restricted the analysis to 14,965 "strong" poly(A) sites, defined as sites that had an average processing probability of greater than 0.10 in WT samples. We then used a t test (2-sided, unequal variance) to identify sites with a significant difference in estimated processing probability in *ipa1-1* compared to WT. Under these stringent restrictions, 637 sites had a significant variation (FDR <0.2), with 627 sites (from 534 genes) showing suppression (a reduction in polyadenylation processing probability) and 10 sites (from 8 genes) showing enhancement in the *ipa1-1* mutant (Figure 1C). Thus, while the majority of sites could not pass a stringent test for variation, the subset that did pass were almost all suppressed rather than enhanced. Suppression of sites, as well as lengthening of the 3' UTR

described above, is consistent with the general reduction in the efficiency of polyadenylation in the *ipa1-1* mutant that we have reported previously (Costanzo et al., 2016).

### Specific Configurations of Poly(A) Signals Characterize Sites that Are Most Susceptible to the *ipa1-1* Mutant

The yeast poly(A) control sequence has been studied extensively (Graber et al., 2002; Shalem et al., 2015; Tian and Graber, 2012) and has been successfully modeled as a five-component sequence element, as depicted in Table 1. (In the description that follows, analysis was performed on genomic sequences, and we replace U with T in sequence descriptions.) These components are (1) the “efficiency element” with optimal sequence TATATA and a broad positioning distribution that peaks 35–40 nt upstream of the poly(A) site, (2) the “positioning element” with optimal sequence AATAAA and a location focused 10–30 nt upstream of the poly(A) site, (3) an upstream T-rich element with optimal sequence TTTTTT 5–15 nt upstream of the poly(A) site, (4) the cleavage site, optimally a pyrimidine followed by one or more A residues, and (5) the downstream T-rich element, with optimal sequence TTTTTT, located 1–20 nt downstream of the poly(A) site. Previous computational and mutagenesis studies (Graber et al., 2002; Guo and Sherman, 1996; Moqtaderi et al., 2013; Shalem et al., 2015) demonstrated that while an optimal configuration can be defined in terms of sequence and positioning, the sequences can be highly variable, with the total efficiency of a poly(A) signal being a complex mixture of all elements that remains poorly understood. Additionally, little is known in yeast about how specific variations in sequence correlate with altered poly(A) site use caused by mutations of processing factors or changes in growth conditions.

To determine which *cis* elements select a site for susceptibility to Ipa1 deficiency, we segregated the 4,377 genes described above based on two criteria. First, we classified genes into two categories (tight or spread) based on the distribution of their ensemble of poly(A) sites in WT samples. For this classification, we used the 5′-to-3′ cumulative polyadenylation distribution (CPD) (Graber et al., 2013) to measure the separation in nucleotides between the 10<sup>th</sup> and 90<sup>th</sup> percentiles. Examples of CPD plots are shown in Figure 1E. Genes with a tight WT configuration (*SSB1* and *RPS3*) have a single dominant poly(A) site (or cluster of closely spaced sites), whereas genes with a spread distribution (*RPL26A* and *ARF1*) have multiple sites spread over a larger distance. Based on manual inspection of the distribution of poly(A) site separation (Figure 1D), we selected a threshold value of 15 nt to separate genes with a “tight” configuration from those with a “spread” configuration of poly(A) sites.

We also segregated genes according to the mutation-induced change in average 3′ UTR length as either “elongated,” “unchanged,” “truncated,” or “indeterminate,” with “indeterminate” used for genes that have too much within-genotype variability to pass a statistical test for a change in 3′ UTR length, but that also are too divergent to classify as “unchanged.” We focused on the following sets of genes: 2,010 elongated genes (with 176 in a tight configuration and 1,834 in a spread configuration), and 1,319 unchanged genes (with 400 tight and 919 spread). We did not pursue analysis of the 1,021 genes in the

indeterminate category because of their variability and did not examine genes in the truncated category (32 total) due to the low number of such genes.

To further focus our sequence analysis, we reduced each gene to a single poly(A) site, using the specific poly(A) site within each gene that had the highest calculated average poly(A) processing probability. For *ipa1-1* elongated genes in spread configuration, we added an additional constraint, choosing only to work with sites that had both the highest poly(A) probability in WT and the largest change in poly(A) probability in *ipa1-1* compared to WT. This restriction reduced the size of the elongated and spread dataset from 1,834 to 772 (and total number of elongated genes to 948) but increased the rigor of our analysis. Genes with a tight configuration are more highly expressed on average than those in the spread group, for both unchanged and elongated transcripts (Table S1).

Motif analysis of yeast poly(A) signals with generalized pattern recognition tools is difficult since the composite signal is a complex of AT-rich signals on the AT-rich background of 3' UTR sequences. Because we know the optimal elements from previous studies, we focused our analysis to search sequences flanking poly(A) sites for the fraction of sequences that match the known poly(A) control elements described above. For ease of presentation, we discuss the results of a search for the optimal variants (Table 1). However, searches for more divergent matches showed consistent results (Table S2).

A prominent overall feature of tight poly(A) sites is the high percentage of sites with an exact match to a TATATA efficiency element. This characteristic is seen regardless of whether the poly(A) site is suppressed or unchanged by the *ipa1-1* mutation (45% and 49%, respectively, versus 31.5% of all sites, Table 1). In contrast, spread poly(A) sites overall are less likely to have this element (24% for suppressed sites and 28% for unchanged sites). Differences in other components of the poly(A) signal are found when inspecting the elongated and unchanged gene sets. Suppressed spread sites, which give rise to elongated transcripts, are characterized by a decreased likelihood to have a downstream T-rich element compared to unchanged spread sites (6.7% and 12.3%, respectively). The suppressed tight sites are most notable for the increased presence of the AATAAA positioning element compared to unchanged tight sites (11.4% and 5.4%, respectively).

In summary, our analysis indicates that the primary difference between genes with a spread or tight configuration is the presence of a strong efficiency element (TATATA) in the tight group, suggesting that this motif contributes to a strong overall poly(A) site and therefore a tight gene distribution. The defining features of spread poly(A) sites that are suppressed by the *ipa1-1* mutation are a significantly weaker downstream T-rich element and a somewhat weaker efficiency element but a normal positioning and upstream T-rich elements. Interestingly, tight suppressed sites have a stronger positioning element. Implications of these poly(A) sequence differences are further explored in the Discussion.

### **The *ipa1-1* Mutation Causes Pol II Enrichment Downstream of Most Poly(A) Sites**

Given that the *ipa1-1* mutant causes defects in cleavage and polyadenylation, we suspected that Ipa1 would participate in proper Pol II transcription termination, as termination and processing are intricately coordinated. To determine whether *ipa1-1* exhibited termination

defects, we conducted a genome-wide survey of Pol II occupancy by performing chromatin immunoprecipitation sequencing (ChIP-seq) experiments. To analyze the data, we normalized the Pol II coverage in WT and *ipa1-1* mutant backgrounds to their respective inputs and generated log<sub>2</sub> ratio profiles.

Our analysis for four representative genes (*RPS13*, *PMA1*, *ADE5,7*, and *GPM1*) is shown in Figure 2 and integrates the previously generated genome-wide poly(A) site mapping data (Costanzo et al., 2016) with our ChIP-seq analysis. For each gene, the top panel shows the 5'-to-3' CPD across each gene, with positions of the major poly(A) sites evident from a sharp increase in polyadenylation probability downstream of the stop codon. The middle panel gives the Pol II occupancy across the gene for WT and *ipa1-1*. As expected, each gene shows a decline in Pol II beyond the poly(A) site, indicative of transcription termination. The *ipa1-1* mutation causes gene-specific changes in average Pol II occupancy, decreasing for some such as *PMA1*, increasing for others such as *ADE5,7* and *GPM1*, or remaining the same, as with *RPS13*. Overall, there is a modest transcriptome-wide trend toward decreased Pol II occupancy, with 58.6% of the genes showing a loss and 41.4% showing an increase due to the *ipa1-1* mutation. On the whole, however, it is a small change, in that 84.4% of the genes show less than 10% difference in Pol II enrichment (*ipa1-1* versus WT) and 95% show less than 20% difference. If we examine those with “larger change” in Pol II occupancy, it is still modestly biased toward a loss in Pol II enrichment, with 2.7% (157 out of 5,837 genes) showing a decrease of more than 20% but only 2.3% (132 out of 5,837 genes) showing an increase of more than 20%. Reduction in Pol II levels within the gene body has been previously observed with mRNA 3' end-processing mutants (Eaton et al., 2018; Kuehner et al., 2017; Luna et al., 2005; Mapendano et al., 2010), and the decrease in some genes in *ipa1-1* may be related to its processing defect.

To focus our analysis on changes in Pol II processivity rather than occupancy levels, we calculated the average Pol II enrichment across each gene's coding sequence for both WT and *ipa1-1* and then used this ratio to scale the *ipa1-1* plot, effectively normalizing to equal Pol II occupancy. Examination of this normalized difference in Pol II occupancy reveals that the zone of termination expands downstream in *ipa1-1* for each gene (Figures 2A–2D, bottom panels). To globally assess termination defects, we generated anchor plots aligning the *ipa1-1*:WT difference in Pol II occupancy to the poly(A) site position for genes with unchanged sites and for those with suppressed sites (Figure 2E). Accumulation of Pol II downstream of poly(A) sites is evident in the mutant for both sets of sites and extends until ~400 bp. It has been reported that in WT yeast, termination occurs within ~200 bp from the poly(A) site (Baejen et al., 2017; Schaughency et al., 2014). Our analysis shows that, in *ipa1-1*, termination is delayed on most mRNA genes, as might be expected if release of Pol II is delayed because 3' end cleavage is less efficient.

Mutations in proteins needed for mRNA 3' end processing can also cause defects in termination at genes encoding small nucleolar (sno) RNAs (Garas et al., 2008; Mischo and Proudfoot, 2013). These genes are transcribed by Pol II but their 3' ends are generated by termination or by RNase III-mediated cleavage, followed by exonuclease-mediated trimming of the 3' end, and not by the cleavage machinery that acts on the yeast protein-coding transcripts (Peart et al., 2013). We analyzed Pol II distribution on the 76 yeast snoRNA

genes and found that, in the *ipa1-1* mutant, the occupancy of Pol II increased downstream of mapped snoRNA ends (Figure 2F). Thus, Ipa1 is important for termination of both mRNA and snoRNA genes.

### Ipa1 Promotes Pol II Transcription Termination on a Naked DNA Template

To assess the mechanisms by which Pol II termination is altered in the *ipa1-1* mutant, we first performed a multi-round *in vitro* transcription termination assay (Mariconti et al., 2010) using extracts prepared from the WT and *ipa1-1* strains. This assay uses two transcription templates constructed by Mariconti et al. (2010), which contain five tandem G-less cassettes of varying lengths. On one of the templates, the first two cassettes are separated from the last three by a functional *CYCI* poly(A) sequence element known to terminate transcription *in vivo* and *in vitro* (Figure 3A). Body radio-labeled RNAs transcribed from this template in extracts were digested with RNase T1, which cleaves only 3' of guanosines, and the resulting RNase T1-resistant G-less fragments were resolved by denaturing gel electrophoresis (Figure 3B). To measure transcriptional readthrough, the radioactive signals of the bands corresponding to the cassettes downstream of the *CYCI* poly(A) element were normalized to that of the band corresponding to the 100 nt upstream cassette (Figure 3C). As shown previously (Mariconti et al., 2010; Pearson and Moore, 2014), the *CYCI* poly(A) element directs transcription termination, with only 5% of the transcripts extending past the poly(A) site in the WT extract. However, roughly 25% of the transcripts are extended in the *ipa1-1* extract, indicating that Pol II termination *in vitro* is much less efficient in the mutant background and is defective in the absence of chromatin. Examination of a template lacking a *CYCI* poly(A) site showed that WT and *ipa1-1* extracts have similar levels of Pol II processivity on a non-chromatin template (Figures 3B and 3D), with no difference in signal even at the 145 nt cassette, which is located approximately 1.3 kb from the transcription start site.

### The *ipa1-1* Mutant Causes Changes in Phosphorylation of Pol II CTD and Diminishes Recruitment of the CPF 3' End-Processing Factor

We next performed ChIP experiments in the WT and *ipa1-1* strains to identify changes in the transcription complex that might explain the *ipa1-1* termination defect. We examined the *RPS13* gene, which shows a clear accumulation of Pol II downstream of its poly(A) site in our ChIP-seq analysis but identical levels of Pol II across the gene body (Figure 2A). Chromatin immunoprecipitated by Pol II antibody was analyzed by qPCR using primer pairs across *RPS13* (Figure 4A, top panel) to generate a snapshot of Pol II distribution along this gene. Consistent with the ChIP-seq pattern for *RPS13*, there was a 2- to 3-fold increase in Pol II occupancy 150 and 440 bp past the poly(A) site in *ipa1-1* (primer pairs 1220 and 1507, Figure 4A).

We also used ChIP to analyze the phosphorylation patterns of Ser2 and Ser5 of the heptad repeat of the Pol II C-terminal domain (CTD). This phosphorylation is coupled to transitions in transcription elongation to transcription termination (Heidemann et al., 2013; Hsin and Manley, 2012). The level of Ser5 phosphorylation, which does not affect C/P factor recruitment, is similar on *RPS13* in WT and mutant backgrounds (Figure 4C). However, we found that Ser2 phosphorylation levels, which are coupled to C/P factor recruitment, are



reduced in the mutant throughout the *RPS13* open reading frame (ORF) and especially in the region downstream of the poly(A) site (Figure 4D).

We next determined how well the C/P factors were recruited to *RPS13* by examining the occupancy of Pta1, a subunit of CPF, and of Rna15, a subunit of CF IA, another factor needed for 3' end processing (Figures 4E and 4F). The Pta1 and Rna15 signals were normalized to Pol II in Figure 4A. Consistent with previous reports (Kim et al., 2004a; Mayer et al., 2012; Nedeá et al., 2003), in WT cells, Pta1 and Rna15 are found at low levels in the gene body and spike to a much higher level at the gene's 3' end. Interestingly, Pta1 recruitment to Pol II is strongly reduced in the *ipa1-1* mutant downstream of the poly(A) site of *RPS13* (Figure 4E). However, Rna15 is recruited to WT levels (Figure 4F) in spite of the decrease in Ser2 phosphorylation. In summary, the termination defect at *RPS13* is correlated with a severe reduction in Ser2P and in CPF, but not in CF IA recruitment.

### Overexpression of the Ysh1 Endonuclease Rescues the *ipa1-1* Termination Defect

We next confirmed by western blot with antibodies against Pta1 that the change in recruitment of Pta1 to the *RPS13* poly(A) site was not due to a change in its relative abundance in *ipa1-1* cells (Figure 4B, lanes 1 and 2). The Rna15 level is also unchanged in *ipa1-1*. Recent work has demonstrated a physical interaction between Ipa1 and Ysh1 (Casañal et al., 2017; Costanzo et al., 2016). Western blotting reveals that Ysh1 abundance is severely reduced in *ipa1-1* cells (Figure 4B, lanes 1 and 2). By poly(A) tag counts, the relative amount of *YSH1* mRNA, however, is not decreased in the *ipa1-1* mutant (Costanzo et al., 2016). In work to be described elsewhere (S.D.L. and C.L.M., unpublished data), we find that Ysh1 is the only subunit of the processing complex that is decreased in the *ipa1-1* mutant, and it is likely that the Ipa1-Ysh1 interaction has a stabilizing effect on the Ysh1 protein. Introduction of additional copies of Myc-tagged *YSH1* (Myc-Ysh1) on a low-copy plasmid into WT and *ipa1-1* cells increases the level of the Ysh1 protein (Figure 4B). When *YSH1* is overexpressed in this way, the amount of Pol II downstream of the *RPS13* poly(A) site, as measured by ChIP-qPCR, is reduced almost to WT levels (Figures 5A and 5B). This result indicates that additional copies of Ysh1 can restore transcription termination in the absence of functional Ipa1.

To further dissect the connection between the Ipa1-Ysh1 interaction and transcription termination, we examined the Myc-Ysh1 and Pta1 ChIP-qPCR profiles in WT and *ipa1-1*. Normalized to Pol II occupancy, Myc-Ysh1 is enriched in WT and mutant backgrounds downstream of the poly(A) site of *RPS13* (Figure 5C), suggesting that Myc-Ysh1 is recruited to the 3' end in the absence of functional Ipa1. Overexpression of *YSH1* increases the recruitment of Pta1 to the 3' end of *RPS13* in the *ipa1-1* mutant (Figure 5D), with the amount of Pta1 in this region now exceeding that seen in WT cells. This result is not due to an overall increase in the steady-state levels of Pta1 in the cell (Figure 4B). These findings indicate that while Pta1 occupancy and appropriate transcription termination is recovered upon introduction of exogenous Myc-Ysh1 in *ipa1-1* cells, functional Ipa1 is necessary for balanced Pta1 recruitment to Pol II.

## Ipa1 Facilitates Proper Transcription Elongation Kinetics

Along with examining recruitment of processing factors, we determined whether Ipa1 was recruited to actively transcribed chromatin in a pattern similar to that of the processing factors, as might be expected from the physical interaction between Ipa1 and Ysh1 (Costanzo et al., 2016). We tagged a chromosomal copy of *IPA1* with Myc and performed ChIP-qPCR. Using the *PMA1* gene as an example, we found that Ipa1 localizes to the coding sequence and 3' UTR of the *PMA1* locus (Figure 6B) and is thus associated with transcriptionally active chromatin. Subunits of the C/P complex typically show some ChIP signal in the body of actively transcribed genes but spike in occupancy at the 3' end (Kim et al., 2004a; Nedeá et al., 2003). We observed this pattern for Pta1 and Rna15 on *PMA1*, with the spike occurring at position 3347 (Figure 6B). However, unlike Pta1 and Rna15, Ipa1 occupancy begins to decline at position 3347, a point where there is also a large decrease in Pol II occupancy (Figure 6C).

The similarity between the chromatin occupancy patterns of Ipa1 and known elongation factors (Kim et al., 2004a; Mayer et al., 2010) suggests that Ipa1 might influence transcription elongation in addition to termination. To address this question, we spotted serial dilutions of WT and *ipa1-1* cells on a medium containing 6-azauracil (6-AU), a chemical that depletes intracellular GTP and UTP pools and can exaggerate transcription elongation defects (Gaillard et al., 2009; Powell and Reines, 1996; Riles et al., 2004). The *ipa1-1* strain is very sensitive to 6-AU, suggesting that transcription elongation is affected on a global scale in this mutant (Figure 6D). To further delineate the defect, we examined two mRNA 3' end-processing mutants (*cft2-1* and *pcf11-2*) that have the same strain background as our *ipa1-1* mutant. All three mutants are thermosensitive for growth at 37°C (Figure 6D, right-hand panel), and we have previously demonstrated that these mutants are all defective for cleavage and polyadenylation *in vitro*, with *cft2-1* being much more impaired compared to *ipa1-1* and *pcf11-2* (Costanzo et al., 2016). We found that 6-AU sensitivity does not correlate with the 3' end-processing defect, as *cft2-1* shows no growth inhibition, *pcf11-2* shows intermediate inhibition, and *ipa1-1* shows severe inhibition (Figure 6D, middle panel).

To confirm an effect on elongation, we employed a ChIP-based *in vivo* transcription assay, which measures Pol II kinetics (Mason and Struhl, 2005). This assay relies on the *GAL1* promoter fused to a naturally occurring, long ORF in yeast, *YLR454*, as a means to activate transcription via galactose induction and to shut off transcription via addition of glucose (Figure 6F). Using primer pairs at 2 kb intervals, the last wave of transcribing Pol II molecules along *YLR454* can be observed upon glucose shutoff. In the presence of galactose, Pol II occupancy is observed at relatively even levels along the ORF in the WT and *ipa1-1* backgrounds (Figure 6G). This result suggests that Pol II processivity through chromatin is similar between the two strains and corroborates the observation that Pol II has similar processivity on a naked DNA template in both WT and mutant extracts (Figure 3). Four minutes after glucose addition, Pol II occupancy in WT is reduced 5- to 6-fold over the entire length of the gene, when compared to that in galactose, as transcription is shut off (Figure 6H). In *ipa1-1*, Pol II occupancy is reduced at the 5' end of the ORF compared to growth in galactose but increases toward the 3' end of the ORF (Figure 6H). This striking

Pol II occupancy pattern represents the last wave of Pol II molecules transcribing to the end of the ORF once transcription has been shut off. This observation is in agreement with the observed 6-AU sensitivity and indicates that Ipa1 participates in maintaining proper Pol II transcription elongation *in vivo*.

To determine whether restoration of Ysh1 expression could also rescue the *ipa1-1*-mediated elongation defect, we tested the growth of cells expressing plasmid-borne *YSH1* on a 6-AU-containing medium. Extra copies of Ysh1 produced from a high-copy plasmid could not relieve the 6-AU sensitivity of *ipa1-1* (Figure 6E), suggesting that, while the transcription termination activity is dependent upon Ysh1 (and can be restored without functional Ipa1), functionally intact Ipa1 is critical for transcription elongation.

## DISCUSSION

In this report, we describe an unexpected interaction by which the cell uses the Ipa1 protein to coordinate and balance transcription and pre-mRNA processing, thus insuring proper gene expression. Ipa1 was originally identified as important for mRNA polyadenylation (Costanzo et al., 2016), but the mechanism by which it exerted this effect was not known. Here, we show that inactivation of Ipa1 causes a severe reduction in Ysh1, the endonuclease that cleaves the pre-mRNA precursor at the poly(A) site. This loss of Ysh1 leads to diminished recruitment of CPF to the 3' ends of genes and to termination defects. Importantly, we find that the role of Ipa1 extends beyond acting at the 3' end of genes, with Ipa1 promoting the elongation phase of the Pol II transcription cycle. Restoring expression of Ysh1 to *ipa1-1* mutant cells permits accumulation of CPF in the 3' UTR and concomitant rescue of the defective termination phenotype. Despite the recovered termination activity, the restoration of Ysh1 alone is insufficient to rescue the 6-AU sensitivity of *ipa1-1*, which may instead reflect a function of Ipa1 at other steps in gene expression.

### **Ipa1 Associates with Chromatin in the Manner of an Elongation Factor and Loss of Ipa1 Function Impairs Elongation**

We have shown that Ipa1 is recruited to chromatin over the entire length of a gene's ORF, with reduction in occupancy beyond the poly(A) site. We interpret this result to mean that Ipa1 associates with actively transcribed chromatin during early elongation and dissociates during pre-mRNA 3' end processing and termination, a pattern that resembles that of known elongation factors (Kim et al., 2004a; Mayer et al., 2010). The defect in elongation kinetics that we observe in *ipa1-1* is also consistent with Ipa1 functioning at the elongation step. We do not observe an elongation defect in transcription assays using cell extract, indicating that loss of Ipa1 function impedes Pol II progression through chromatin but not on a naked DNA template. In further support of the role of Ipa1 in elongation, loss of the Cdc73 subunit of the Paf1 complex (Paf1C), a crucial elongation factor, displays a synthetic lethal genetic interaction with the *ipa1-1* mutation (van Pel et al., 2013), suggesting that Ipa1 and Paf1C operate in overlapping or parallel pathways. The *ipa1-1* mutant also exhibits a negative genetic interaction with the capping enzyme subunit Ceg1 (Costanzo et al., 2016). Ceg1 recruits a second subunit, Cet1, which in turn promotes the transition to elongation (Sen et al., 2017).

By serving as an elongation factor, Ipa1 might facilitate the coupling of transcription and mRNA 3' end maturation. Insight into how this could happen comes from recent structural analysis of CPF showing that the nuclease, poly(A) polymerase, and phosphatase activities of CPF are organized into three modules (Casañal et al., 2017). An earlier study has shown that Cft1, a key scaffolding protein of the polymerase module, physically interacts with Paf1C (Nordick et al., 2008). As an elongation factor (Costa and Arndt, 2000; Squazzo et al., 2002; Tomson and Arndt, 2013; Tous et al., 2011), Paf1C closely associates with the transcription complex shortly after promoter escape and dissociates near the poly(A) site (Kim et al., 2004a; Mayer et al., 2010). Similar to Ipa1, Paf1C also influences 3' end activities and is needed for proper levels of CTD Ser2 phosphorylation (Chen et al., 2015; Mueller et al., 2004; Nordick et al., 2008; Penheiter et al., 2005; Yu et al., 2015). Nordick et al. proposed that Paf1C recruits Cft1 early on to the elongation complex, travels with the transcriptional apparatus in a complex with Cft1, and then dissociates, leaving Cft1 behind with the transcription complex once the poly(A) site has been transcribed. Similarly, the CF IA factor may be assembled only after the poly(A) site is reached. A study of the interaction of the CF IA subunit Pcf11 with the export factor Yra1 suggests that Pcf11 hands off Yra1 to the mRNP assembly apparatus before joining Clp1, Rna14, and Rna15 (the remaining CF IA subunits) to function in processing at the poly(A) site (Johnson et al., 2011). An association with the Spt5 elongation factor helps bring Rna14, Rna15, and Clp1 to transcribed genes to promote termination (Baejen et al., 2017; Mayer et al., 2012).

Ipa1 associates with the Ysh1 and Mpe1 subunits of the CPF nuclease module, but not with Cft2, the remaining component of this module, or with other proteins in the C/P complex (Casañal et al., 2017; Costanzo et al., 2016). These studies suggest that Ipa1 is not a stable component of CPF but exists in a complex with Ysh1, Mpe1, and possibly other not-yet-identified proteins. In a fashion analogous to those described above, the Ipa1 elongation factor may travel with the transcription machinery in a complex with Ysh1 and Mpe1 and subsequently deliver these proteins to the poly(A) site once it is exposed. Timely delivery would allow tightly coordinated assembly of the rest of the CPF into a fully functional processing apparatus.

A consequence of the *ipa1-1* mutation is increased retention of Pol II beyond poly(A) sites, and overexpression of *YSH1* can rescue this defect on the *RPS13* gene. This finding supports the conclusions drawn from other studies that mutation or loss of Ysh1 or its CPSF73 mammalian homolog causes termination defects on mRNA-encoding genes (Baejen et al., 2017; Eaton et al., 2018; Garas et al., 2008; Nojima et al., 2013; Schaughency et al., 2014). Ysh1/CPSF-73 is also important for termination of snoRNA genes in budding yeast (Garas et al., 2008) and in fission yeast (Larochelle et al., 2018). Therefore, the Ysh1 depletion caused by the Ipa1 mutation may explain the delayed termination of Pol II at both mRNA and snoRNA genes.

### **Sequence Dependency of Poly(A) Processing in Response to Ipa1 Mutation Suggests a Mechanism for Site-Specific Tuning**

Efficient mRNA 3' end processing requires several contacts between the processing complex and specific RNA sequences surrounding the poly(A) site, as depicted in Table 1

for yeast. A composite of these elements will determine the strength of a particular site, and variations likely affect how well the processing complex assembles around the poly(A) site and how Ysh1 is positioned to carry out its function in cleavage. The UAUUA efficiency element has been shown to most strongly correlate with the amount of protein expression (Shalem et al., 2015), and our analysis indicates that it is also critical in determining whether cleavage of the mRNA 3' end is tightly focused or instead spread over tens or hundreds of nucleotides. We also found that sites most likely to be resistant to the *ipa1-1* mutation, regardless of whether they have a tight or spread configuration, have a good match to the downstream U-rich motif. As evident from the tightly focused *ipa1-1*-suppressed sites (Table 1), even a combination of strong UAUUA and A-rich elements cannot compensate for a poor downstream element. In yeast, mRNA polyadenylation is performed by a complex of Hrp1, which binds to the UAUUA motif, and two multi-subunit factors, CPF and CF IA. If the *ipa1-1* mutation causes a scarcity of intact CPF, those sites that can most stably recruit CPF are likely to be processed more efficiently and accurately in the mutant. Our analysis indicates that the downstream U-rich motif is critical for this recruitment, possibly through interaction the Cft2 subunit of CPF and Rna15 of CF IA, which crosslink to this element *in vivo* (Baejen et al., 2014).

Several mammalian studies have shown that alternative polyadenylation can be regulated by the amount of core C/P subunits, and that the poly(A) sites most affected by loss of subunits such as hFip1 and CFI<sub>M</sub> 68 are enriched in the binding sites for these factors, and therefore more dependent on these proteins for 3' end processing (Lackford et al., 2014; Li et al., 2015; Tian and Manley, 2017). Alternatively, affected poly(A) sites may have poorer matches to the preferred binding sequence, as has been seen for CstF64/CstF64 $\tau$  depletion (Yao et al., 2013). In agreement with these studies, our analysis indicates that the 3' end processing of a gene's transcript can be exquisitely tunable depending on the nature of the polyadenylation signals that specify each poly(A) site.

Historically, studies and modeling of multi-partite regulatory sequences have necessarily treated variation from the optimal sequence as random noise. However, our results and those described above suggest that such variations are part of regulatory mechanisms that facilitate changes in gene expression. In our experimental system, Ipa1 and Ysh1 expression was lost through mutation, but their expression might also be modulated naturally by the cell in response to environmental change. Our findings suggest that these changes would target a specific subset of all poly(A) sites as part of the cellular response. It will be interesting in the future to determine whether mutations in other RNA-binding subunits of the yeast C/P complex affect specific subsets of genes, and whether rubrics developed from such analyses will allow prediction of which specific C/P proteins are likely to be regulated when the cell state changes.

In summary, we propose that the Ipa1/Ysh1 interaction provides the cell with a means to coordinate transcription elongation with pre-mRNA 3' end processing and perhaps simultaneously regulate both of these steps in mRNA synthesis according to the cell's needs. In our current study, we have found that inactivation of Ipa1 impairs elongation. Thus, if the cell needs to slow mRNA synthesis, a decrease of Ipa1 and the subsequent decrease of Ysh1 would correspondingly slow both elongation and processing. We have recently reported that

a mutation in Ysh1 that is defective for 3' end processing also causes slower elongation (McGinty et al., 2017). This finding, together with earlier studies showing that mutations in CF IA also cause elongation defects (Luna et al., 2005; Tous et al., 2011), suggests that a poorly functioning processing complex can also feedback to slow elongation.

Ipa1 is conserved in higher eukaryotes, including humans, and the human ortholog of Ipa1, UBE3D, was found to physically interact in quantitative proteomics screens (Hein et al., 2015; Huttlin et al., 2017) with CPSF73, the ortholog of Ysh1. This evidence points to the potential of a highly conserved mechanism of transcriptional and 3' end RNA processing control imparted by the interaction between Ipa1/UBE3D and Ysh1/CPSF73. Future investigations may reveal a widespread “molecular chaperone” mechanism in which critical subunits of co-transcriptional complexes are accompanied by transcription elongation factors and are subsequently delivered to their specific sites of action in a spatially and temporally coordinated manner.

## STAR★METHODS

### CONTACT FOR REAGENT AND RESOURCE SHARING

Requests for further information and reagents may be directed to the Lead Contact, Dr. Claire Moore, at Tufts University (Claire.moore@tufts.edu).

### EXPERIMENTAL MODEL AND SUBJECT DETAILS

**Yeast strains**—Yeast strains BY4741 (Wild-type), TSA1248 (*ipa1-1*), TS801 (*cft2-1*), and TSA685 (*pcf11-2*) were obtained from Charles Boone, University of Toronto (Costanzo et al., 2016). Yeast were grown in YPAD (YPD supplemented with adenine) rich medium or in Complete Media minus uracil or leucine at 30°C and as indicated, shifted to the non-permissive temperature of 37°C for 1 hour. For spot growth assays, 5 or 10-fold dilutions were prepared in a 96 well plate prior to using a replica pin plater to spot cultures onto agar plates. For the *YSH1* overexpression studies, yeast were transformed with the indicated plasmids and transformants selected on selective medium.

**Bacterial strains**—DH5 $\alpha$  cells were grown in LB medium at 37°C and used to propagate plasmids.

### METHOD DETAILS

***In vitro* transcription assay**—To generate extract for *in vitro* transcription, yeast (1 l) were grown to an OD<sub>600</sub> of 2.0–5.0. The cells were harvested and resuspended in one volume of AGK buffer [20 mM HEPES-KOH, pH 7.9; 200 mM KCl; 1.5 mM Mg<sub>2</sub>Cl; 10% glycerol; 0.5 mM Dithiothreitol (DTT)] supplemented with EDTA-free protease inhibitor cocktail (Roche). Cells were frozen in droplets in liquid nitrogen and lysed with cryo-grinding. The thawed lysate was cleared with ultracentrifugation first at 31,000 rpm in the TLA 100.3 rotor for 30 minutes and then at 65,000 rpm in the same rotor for 1 hour. Proteins in the cleared lysate were precipitated with 0.24 mg/ml finely ground ammonium sulfate (40% saturation) with stirring on ice for 30 minutes. The ammonium sulfate pellet was collected with ultracentrifugation at 31,000 rpm in the TLA 100.3 rotor for 20 minutes and

was carefully resuspended in 40  $\mu$ l of D-alternative buffer (20 mM HEPES, pH 7.9; 75 mM potassium acetate; 1.5 mM magnesium acetate; 20% glycerol; 1 mM DTT) per ml of sample prior to ammonium sulfate precipitation. The resuspension was dialyzed three times against 600 mL of D-alternative buffer, for one hour each time, and cleared with centrifugation for 2 minutes at 15,000 rpm in a tabletop microcentrifuge. The extracts were flash-frozen in liquid nitrogen and stored at  $-80^{\circ}\text{C}$ . Addition of ammonium sulfate (final concentration of 0.5 M) to the freshly lysed cells and incubation at  $4^{\circ}\text{C}$ , with rocking, prior to centrifugation resulted in transcriptionally inactive extract. Transcription reactions were performed as described previously (Mariconti et al., 2010), except that 100 mg extract and 0.5 mg plasmid DNA were used. The transcription-template plasmids pKS708 and pKS710 were kind gifts of Bernhard Dichtl (Universität Zürich). RNA products were digested with T1 RNase, and the fragments resolved on a 6% polyacrylamide/7 M urea gel. After Phosphorimager detection, the radioactive intensities of each band were measured using ImageQuant software and were normalized to the 100 nt G-less cassette band to calculate the termination efficiency at each downstream G-less cassette. Averages were generated from three independent experiments.

**Yeast extract preparation and western blotting**—For determination of total protein levels, cell extracts were prepared as described by Zhao et al. (Zhao et al., 1999) from cell cultures grown to mid-log phase at  $30^{\circ}\text{C}$ . Fifty micrograms of each extract was resolved onto a 10% polyacrylamide/Bis-Tris-MOPS gel ([https://openwetware.org/wiki/Sauer:bis-Tris\\_SDS-PAGE,\\_the\\_very\\_best#Running](https://openwetware.org/wiki/Sauer:bis-Tris_SDS-PAGE,_the_very_best#Running)). The electrophoresed proteins were transferred to a PVDF membrane and blots were probed with the indicated antibodies against Ysh1, Myc, Pta1, Rna15 and actin.

**Chromatin Immunoprecipitation (ChIP), quantitative PCR (qPCR), and ChIP-seq**—Yeast cells (50 ml) were grown to  $\text{OD}_{600}$  of 0.5, shifted to  $37^{\circ}\text{C}$  for 1 hour, fixed with 1.035% formaldehyde for 15 minutes and neutral-ized with 0.135 M glycine for 5 minutes. Washed cells were lysed in FA-lysis buffer (50 mM HEPES-KOH, pH 7.9; 150 mM NaCl; 1% Triton X-100; 1 mM EDTA; 0.1% sodium deoxycholate; 0.1% SDS) with grinding in liquid nitrogen. Crosslinked chromatin was sheared in a final volume of 500  $\mu$ L FA-lysis buffer in a 2 mL microcentrifuge tube using a Branson water bath sonicator at  $4^{\circ}\text{C}$  for 8 minutes. Two hundred ml of pre-cleared sheared chromatin were immunoprecipitated with 15  $\mu$ L protein A beads pre-equilibrated with the respective antibody, as indicated in the figure legends. For the IgM H5 anti-Ser2P antibody, Anti-mouse IgM – Agarose beads were used. Antibodies used in the ChIP analysis include the anti-pan CTD mouse monoclonal antibody, 4H8 (Santa Cruz); the anti-Rna15 rabbit polyclonal antibody and the anti-Pta1 rabbit polyclonal antibody (generous gifts of Horst Domdey); the anti-Myc clone E10 mouse monoclonal antibody (in house); the 3E8 anti-Ser5P antibody (Active Motif); and the H5 anti-Ser2P antibody (Covance). Five microliters of each Ab was used per IP. SamSamples were rotated 4–5 hours at  $4^{\circ}\text{C}$ . The beads were washed once with each of the following buffers: FA-lysis buffer + 275 mM NaCl; FA-lysis buffer + 500 mM NaCl; LiCl buffer (10 mM Tris-Cl, 1 mM EDTA, 0.25 M LiCl, 0.5% NP-40, 0.5% sodium deoxycholate, pH 8.0); and 10 mM Tris-Cl, 1 mM EDTA. Upon washing, chromatin was eluted with 250  $\mu$ L TE + 1% SDS with incubation at  $65^{\circ}\text{C}$  for 20 minutes followed by a rinse in 250  $\mu$ L TE. Samples were treated with Proteinase K for 1 hour at  $42^{\circ}\text{C}$  and de-crosslinked at  $65^{\circ}\text{C}$  overnight.

LiCl (to 0.4 M) and 20 mg glycogen were added prior to DNA purification. DNA was purified with phenol-chloroform extraction and ethanol precipitation and was resuspended in 200  $\mu$ L qPCR-grade water.

Quantitative PCR (qPCR) was conducted in 20  $\mu$ L reaction volumes consisting of SYBR Green PCR master mix (BioRad), 0.5 mM primers (Table S4) and 2  $\mu$ L of IP or input samples. Up to 40 cycles were used for each experiment. The relative occupancy was calculated as a percentage of input using the equation:  $Ct = 2^{-(IP_{Ct} - input_{Ct})}$ . Average relative occupancy values are presented, and the error bars represent the standard deviation from these average values generated from two to four independent experiments.

For ChIP-seq experiments, yeast cultures were scaled up to 400 mL and were fixed and lysed as described above. The resulting 4 mL chromatin were incubated with 40  $\mu$ L 4H8 antibody and 200  $\mu$ L pre-washed protein G beads for 5 hours at 4°C. The beads were washed as described above and were then eluted in two steps: 1) in 150  $\mu$ L TE + 1% SDS at 65°C for 15 minutes, and 2) in 150  $\mu$ L TE + 0.67% SDS at 65°C for 10 minutes. Samples were de-crosslinked overnight at 65°C and then treated with Proteinase K for 2–4 hours at 42°C. LiCl (to 0.4 M), 20 mg glycogen and 1 mL 100% ethanol were added to precipitate the DNA. The DNA pellets were washed in 70% ethanol and were further purified using MinElute columns (QIAGEN). The eluted DNA was submitted to the Tufts Genomics Core Facility for TruSeq ChIP library preparation and for 50 nt single-end sequencing on the Illumina HiSeq 2500 system.

For the *in vivo* RNA Pol II elongation assay, experiments were performed essentially as described previously (Mason and Struhl, 2005). To assay elongation on the *YLR454* gene, two strains were created from the WT and *ipa1-1* strains by single-step integration of a *TRPI* plasmid containing the *GALI* promoter fused to the 5'-most 300 bp of the *YLR454w* open-reading frame into the *YLR454w* locus. These strains were grown to early mid-log in raffinose-containing minimal medium, induced with 2% galactose for 2.5 hours, shifted to the non-permissive temperature for 1 hour and spiked with 2% glucose for 4 minutes before fixation with 1% formaldehyde. ChIP was performed with the anti-pan CTD mouse monoclonal antibody.

## QUANTIFICATION AND STATISTICAL ANALYSIS

### Computational analysis of poly(A) site usage

**FASTQ file pre-processing:** For all analyses, we used the poly(A) site mapping datasets obtained previously for *IPA1* and *ipa1-1* cells (Costanzo et al., 2016). All comparisons below were made based upon three *ipa1-1* samples (labeled TS1248 in the original data) and four BY4741 wild-type (WT) (labeled BY in the original data) samples. These samples showed a minor batch effect (data not shown), but all were used in our analysis. In brief, the sequence tags were preprocessed to reduce them to a non-redundant set, aligned to the yeast genome (sacCer3), and then post-processed to generate sample-specific maps of the poly(A) sites for each yeast protein-coding gene. For statistical robustness, we restricted analysis of 3'-UTR features to 4377 genes that exceeded an arbitrary cutoff of at least 250 sequence tags summed across all seven samples.



Because poly(A)-site sequences have a very different distribution than standard RNaseq data, specifically in that the poly(A) site sequence data have much higher redundancy, we used the following procedure for our analysis. Each sequence fragment is putatively a reverse-complement read with the first base corresponding to the last base before addition of the poly(A) tail with subsequent reads progressing upstream of the poly(A) site. Sequences were first trimmed of any leading T bases because they are ambiguous as to whether they are of genomic origin, and then trimmed to a common length of 30 nt, a length chosen as a tradeoff between uniqueness and ease of computational manipulation. Each sequence set was then condensed to only its unique sequence “tags” while retaining the exact count of how many times the sequence occurred in the set. This removal of sequence redundancy necessarily means that quality scores were discarded, but we operated from the presumption that the statistics of the occurrence of each tag and its near matches (representing putative errors) would adequately compensate for the absence of quality data. Sequence tags were sorted in decreasing order of occurrence and relabeled according to the pattern “seq\_N\_C” where N was the rank in terms of abundance, and C was the count. The resulting dataset was stored as a fasta sequence file.

**Alignment to the yeast genome:** The reduced sequence tag set was aligned with the sensitive alignment program blat (Kent, 2002) using custom parameters “-t=dna -q=dna -tileSize=10 -stepSize=3 -minIdentity=85 -minScore=24” which were manually optimized to align the maximal number of short tags. The target for alignment was a composite file consisting of the *Saccharomyces cerevisiae* genome, version 3 (saccer3), combined with the sequence of the yeast 2 micron plasmid.

**Alignment post-processing to count tags at putative poly(A) sites:** Custom perl and c++ programs were created to post process the blat-produced psl files through the following steps. (1) count and record the number of times each tag aligned to the genome, (2) count and record the number of tags that aligned at each position in the genome, (3) merge the results of the first two steps with the count of each tag in the dataset to finally generate a file that scored each putative poly(A) site in the genome by the total number of tags that aligned there, the total number of distinct sequences represented within those tags, and the average number of times this set of tags aligned across the genome. For statistical robustness, we restricted analysis of 3'-UTR features to 4377 genes that exceeded an arbitrary cutoff of at least 250 sequence tags summed across all seven samples.

**Assignment of putative sites to genes and other genomic features:** The genomic locations of putative poly(A) sites were assigned in two distinct manners. First the closest properly oriented mRNA gene was chosen, based up on the distance to the stop codon. Second the closest genomic feature of any type was also identified. Annotations were taken from the table SGD\_features.tab, downloaded from the [yeastgenome.org](http://yeastgenome.org) website in January 2013, and merged with the set of SUT and CUT genes as reported by Xu et al. (Xu et al., 2009). In this first assignment, no distance restrictions were imposed.

We examined various means of assigning aligned tags to neighboring genes, specifically comparing using only protein-coding genes versus using protein-coding genes combined with known CUT and SUT targets and concluded that use of protein-coding genes was more

likely to be an accurate reflection of the molecular changes in *ipa1-1* mutants. This conclusion was based on manual examination of the CUT and SUT transcripts that were identified as significantly changed between samples in a preliminary analysis performed with DESeq2 (Love et al., 2014). In nearly all cases, the SUT or CUT transcript (a) was increased in apparent expression in *ipa1-1* compared to WT, (b) showed very low expression in the WT, and (c) was situated on the genome in a configuration downstream on the same strand and relatively close (typically tens of nucleotides) to a relatively highly expressed protein-coding gene. These findings, taken in the context of our broader finding that *ipa1-1* mutation leads to a general lengthening of transcripts from the majority of yeast genes, led us to conclude that the sequence tags in question were more likely to have been generated from extended transcripts of the coding gene than increased initiation at the SUT or CUT genes. Accordingly, we subsequently carried out all subsequent analysis with assignment of poly(A) sequence tags to the nearest properly oriented protein-coding gene.

**Extraction of flanking sites and filtering of putative false priming events and restriction by distance to genes:** For final analysis of expression levels and poly(A) site usage, putative processing sites were limited to those which had 8 or fewer A or G residues in the next 10 nucleotides downstream. In addition, sites were limited to those that occurred between 80 nt upstream of the start codon and 1000 nt downstream of the stop codon. While these limitations might eliminate some true sites, previous studies (Graber et al., 1999; van Helden et al., 2000) suggest that the fraction lost will be well below 10%. Finally, multi-alignment tags were not eliminated, but were instead scaled through multiplication by the inverse of the average number of genomic alignments for tags at the site.

**Calculation of average UTR length for each gene:** For each individual gene within each sample dataset, the average 3'-UTR length was computed as a weighted average of the 3'-UTR length of all transcripts associated with the gene, restricting the analysis to only tags in the 3'-UTR, with Equation 1, where  $\langle U \rangle$  is the average 3'-UTR length, the summations are all over all 3'-UTR polyA sites, and  $U_i$  and  $n_i$  are respectively the 3'-UTR length and number of sequence tags associated with polyA site  $i$ .

$$\langle U \rangle = \frac{\sum_{i = \text{sites}} U_i n_i}{\sum_{i = \text{sites}} n_i} \quad \text{equation 1}$$

**Calculation of site-specific polyadenylation probability:** For each putative poly(A) site within each individual gene within each sample dataset, a poly(A) probability was calculated based on the rationale that transcripts are processed from 5' to 3' and that at each site, the probability represents the choice between 3' end processing or extension of the transcript further in the downstream (3') direction. In this model, the processing probability is independent of upstream (5') sites, and is estimated by the ratio of the count of tags at the current site to the sum of all tags counted from the current site to the 3'-most site associated with the gene. For numerical robustness, a Bayesian prior was incorporated, using the same counts (at the site and downstream) summed for the same gene across all samples in the experiment and down-weighted by a factor of 0.01, as shown in Equation 2, where  $n_i$  is the

tag count at the current site ( $i$ ) in the sample of interest,  $n_{is}$  and  $n_{js}$  are the tag counts at site  $i$  (or  $j$ ) in sample  $s$ , the summation  $j$  is over all sites from the current site to the 3'-most site assigned to this gene, and the summation  $s$  is over all samples in the experiment.

$$p_i = \frac{n_i + 0.01 \sum_{s = \text{samples}} n_{is}}{\sum_{j = \text{sites}} n_j + 0.01 \sum_{j = \text{sites}, s = \text{samples}} n_{js}} \quad \text{equation 2}$$

**Calculation of expression levels and transcript-truncation probabilities based on poly(A) tags:** expression level estimates of each gene's expression within each sample were obtained using the summation of all counts classified as within the 3'-UTR (corresponding to transcripts with a complete coding sequence). In addition, each gene was scored for the fraction of transcripts that were either CDS-truncated (a properly oriented tag occurring upstream of the stop codon) or promoter-proximal (a properly oriented tag occurring within the 5'-UTR or less than 100 nt (arbitrarily chosen) downstream of the start codon). The normalizing denominator in each case was the total count of tags assigned to the gene (5'-UTR, CDS, or 3'-UTR) for a given sample.

**Sequence analysis of poly(A)-site flanking sequences:** to investigate putative poly(A) control sequence elements, we extracted sequences spanning 100 nt upstream to 100nt downstream of the putative sites, using custom C++ programs. Sequence analysis was restricted to only one representative site per gene in order to reduce the likelihood of biased results due to highly similar sequences. Occurrence of common regulatory element hexamers was measured with command-line scripts.

**Analysis of RNA Polymerase II ChIP-seq datasets:** Fastq sequence files were imported to the public Galaxy server (Afgan et al., 2016) (<https://usegalaxy.org/>) and analyzed in the following sequence.

1. Sequences were analyzed for quality control with the program fastqc, revealing no issues.
2. All sequences were then aligned to the *S. cerevisiae* genome (sacCer3 as provided in Galaxy), using BWA (ID <https://toolshed.g2.bx.psu.edu/repos/devteam/bwa/bwa/0.7.15.1>) using default parameters (Li and Durbin, 2009, 2010).
3. The resulting BAM output files were converted to bigwig coverage using the DeepTools Bam Coverage program (ID [toolshed.g2.bx.psu.edu/repos/bgruening/deeptools\\_bam\\_coverage/deeptools\\_bam\\_coverage/2.5.0.0](https://toolshed.g2.bx.psu.edu/repos/bgruening/deeptools_bam_coverage/deeptools_bam_coverage/2.5.0.0)), using bin-size = 1 (every base independent) (Ramírez et al., 2016) but with averaging over a window 21 nt wide centered on the current base, but with all other parameters left at default values. Normalization was set to “constant coverage on the genome” such that the total coverage across all samples is forced to be equal.
4. The bigwig files for each pair of PolII and matched input dataset were compared with the Deep Tools bigwigCompare program (ID <https://toolshed.g2.bx.psu.edu/>

[repos/bgruening/deeptools\\_bigwig\\_compare/deeptools\\_bigwig\\_compare/2.5.0.0](https://github.com/bgruening/deeptools_bigwig_compare/deeptools_bigwig_compare/2.5.0.0)) (Ramírez et al., 2016), producing a log<sub>2</sub>ratio of the Pol II to input coverage, again using window size = 1

The four resulting Pol II-enrichment files (representing two replicates for the wild-type (WT) and two replicates for the *ipa1-1* mutant) were downloaded and processed with command-line tools to generate a joined file that included calculation of the average value at each genomic position (chromosome-nucleotide-position), as well as the difference between them, represented as *ipa1-1* average minus WT average.

Since we are interested in the relative distribution of Pol II along each gene (because *ipa1-1* has been associated with polyadenylation), we decided to normalize each gene's local neighborhood independently for meta-gene analysis. Conventional wisdom holds that polII enrichment correlates well with expression level, so we calculated the average value of the difference in enrichment within the coding region of each gene, using the gene CDS boundaries from [yeastgenome.org](http://yeastgenome.org) to define the region of interest, and used the ratio of the WT and *ipa1-1* averages to scale *ipa1-1* enrichments before comparing with WT values.

**Pol II enrichment anchor plot generation:** Plots anchored at the poly(A) site were generated for “locally normalized” data (as described above), by aligning related sites that were oriented 5'-to-3' in terms of transcription direction of the gene. Mean and standard error of the mean values at each position were calculated and plotted using Microsoft Excel.

**Composite polyA and ChIp-Seq plot generation:** The composite cumulative polyadenylation distribution (CPD) and Pol II enrichment plots were generated from the output text tables described above. Bash shell scripts were written to generate a display script to be interpreted and displayed by the plotting program Gnuplot, version 5.2.

**Statistical analysis:** Gene-specific features (such as average 3'-UTR length) and polyadenylation site-specific features (such as polyadenylation processing probability) were all computed separately for each of the four WT and 3 *ipa1-1* samples and then compared with a two-sided t test, assuming unequal variances. Calculations were made on tables stored in Microsoft Excel. Multiple hypothesis testing was accomplished by controlling the false discovery rate (FDR) (Benjamini and Hochberg, 1995).

## DATA AND SOFTWARE AVAILABILITY

Pol II ChIp-seq fastq datasets have been submitted to the Sequence Read Archive, and the accession number for ChIP-seq data reported in this paper is (Database): accession GEO: GSE117402. All software used in this study is listed in the Key Resources Table. All locally generated software used herein are available without restriction on request from the authors.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## ACKNOWLEDGMENTS

We thank members of the Moore, Bohm, and Gaglia labs for helpful discussions, Jason Kuehner and Katarzyna Kaczmarek Michaels for feedback on the manuscript, and Kevin Harlen for help with initial analysis of the ChIP-seq data. This research was supported by grant MCB-1714603 from the United States National Science Foundation (NSF) to C.L.M. J.H.G. was partially supported by an Institutional Development Award (IDeA) from the National Institute of General Medical Sciences of the National Institutes of Health under grant numbers P20GM103423 and P20GM104318.

## REFERENCES

- Afgan E, Baker D, van den Beek M, Blankenberg D, Bouvier D, Dieck M, Chilton J, Clements D, Coraor N, Eberhard C, et al. (2016). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. *Nucleic Acids Res.* 44 (W1), W3–W10. [PubMed: 27137889]
- Baejen C, Torkler P, Gressel S, Essig K, Söding J, and Cramer P (2014). Transcriptome maps of mRNP biogenesis factors define pre-mRNA recognition. *Mol. Cell* 55, 745–757. [PubMed: 25192364]
- Baejen C, Andreani J, Torkler P, Battaglia S, Schwalb B, Lidschreiber M, Maier KC, Boltendahl A, Rus P, Esslinger S, et al. (2017). Genome-wide analysis of RNA polymerase II termination at protein-coding genes. *Mol. Cell* 66, 38–49. [PubMed: 28318822]
- Benjamini Y, and Hochberg Y (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* 57, 289–300.
- Casañal A, Kumar A, Hill CH, Easter AD, Emsley P, Degliesposti G, Gordiyenko Y, Santhanam B, Wolf J, Wiederhold K, et al. (2017). Architecture of eukaryotic mRNA 3′-end processing machinery. *Science* 358, 1056–1059. [PubMed: 29074584]
- Chen FX, Woodfin AR, Gardini A, Rickels RA, Marshall SA, Smith ER, Shiekhhattar R, and Shilatifard A (2015). PAF1, a molecular regulator of promoter-proximal pausing by RNA Polymerase II. *Cell* 162, 1003–1015. [PubMed: 26279188]
- Costa PJ, and Arndt KM (2000). Synthetic lethal interactions suggest a role for the *Saccharomyces cerevisiae* Rtf1 protein in transcription elongation. *Genetics* 156, 535–547. [PubMed: 11014804]
- Costanzo M, VanderSluis B, Koch EN, Baryshnikova A, Pons C, Tan G, Wang W, Usaj M, Hanchard J, Lee SD, et al. (2016). A global genetic interaction network maps a wiring diagram of cellular function. *Science* 353, Published online September 23, 2016. 10.1126/science.aaf1420.
- Eaton JD, Davidson L, Bauer DLV, Natsume T, Kanemaki MT, and West S (2018). Xrn2 accelerates termination by RNA polymerase II, which is underpinned by CPSF73 activity. *Genes Dev.* 32, 127–139. [PubMed: 29432121]
- Elkon R, Ugalde AP, and Agami R (2013). Alternative cleavage and polyadenylation: Extent, regulation and function. *Nat. Rev. Genet.* 14, 496–506. [PubMed: 23774734]
- Gaillard H, Tous C, Botet J, González-Aguilera C, Quintero MJ, Viladevall L, García-Rubio ML, Rodríguez-Gil A, Marín A, Ariño J, et al. (2009). Genome-wide analysis of factors affecting transcription elongation and DNA repair: A new role for PAF and Ccr4-not in transcription-coupled repair. *PLoS Genet.* 5, e1000364. [PubMed: 19197357]
- Garas M, Dichtl B, and Keller W (2008). The role of the putative 3′ end processing endonuclease Ysh1p in mRNA and snoRNA synthesis. *RNA* 14, 2671–2684. [PubMed: 18971324]
- Graber JH, Cantor CR, Mohr SC, and Smith TF (1999). Genomic detection of new yeast pre-mRNA 3′-end-processing signals. *Nucleic Acids Res.* 27, 888–894. [PubMed: 9889288]
- Graber JH, McAllister GD, and Smith TF (2002). Probabilistic prediction of *Saccharomyces cerevisiae* mRNA 3′-processing sites. *Nucleic Acids Res.* 30, 1851–1858. [PubMed: 11937640]
- Graber JH, Nazeer FI, Yeh PC, Kuehner JN, Borikar S, Hoskinson D, and Moore CL (2013). DNA damage induces targeted, genome-wide variation of poly(A) sites in budding yeast. *Genome Res.* 23, 1690–1703. [PubMed: 23788651]
- Guo J, and Price DH (2013). RNA polymerase II transcription elongation control. *Chem. Rev.* 113, 8583–8603. [PubMed: 23919563]

- Guo Z, and Sherman F (1996). 3'-end-forming signals of yeast mRNA. *Trends Biochem. Sci.* 21, 477–481. [PubMed: 9009831]
- Heidemann M, Hintermair C, Voß K, and Eick D (2013). Dynamic phosphorylation patterns of RNA polymerase II CTD during transcription. *Biochim. Biophys. Acta* 1829, 55–62. [PubMed: 22982363]
- Hein MY, Hubner NC, Poser I, Cox J, Nagaraj N, Toyoda Y, Gak IA, Weisswange I, Mansfeld J, Buchholz F, et al. (2015). A human interactome in three quantitative dimensions organized by stoichiometries and abundances. *Cell* 163, 712–723. [PubMed: 26496610]
- Hsin JP, and Manley JL (2012). The RNA polymerase II CTD coordinates transcription and RNA processing. *Genes Dev.* 26, 2119–2137. [PubMed: 23028141]
- Huang LZ, Li YJ, Xie XF, Zhang JJ, Cheng CY, Yamashiro K, Chen LJ, Ma XY, Cheung CM, Wang YS, et al. (2015). Whole-exome sequencing implicates UBE3D in age-related macular degeneration in East Asian populations. *Nat. Commun.* 6, 6687. [PubMed: 25872646]
- Huttlin EL, Bruckner RJ, Paulo JA, Cannon JR, Ting L, Baltier K, Colby G, Gebreab F, Gygi MP, Parzen H, et al. (2017). Architecture of the human interactome defines protein communities and disease networks. *Nature* 545, 505–509. [PubMed: 28514442]
- Johnson SA, Kim H, Erickson B, and Bentley DL (2011). The export factor Yra1 modulates mRNA 3' end processing. *Nat. Struct. Mol. Biol.* 18, 1164–1171. [PubMed: 21947206]
- Kent WJ (2002). BLAT—the BLAST-like alignment tool. *Genome Res.* 12, 656–664. [PubMed: 11932250]
- Kim M, Ahn SH, Krogan NJ, Greenblatt JF, and Buratowski S (2004a). Transitions in RNA polymerase II elongation complexes at the 3' ends of genes. *EMBO J.* 23, 354–364. [PubMed: 14739930]
- Kim M, Krogan NJ, Vasiljeva L, Rando OJ, Nedeá E, Greenblatt JF, and Buratowski S (2004b). The yeast Rat1 exonuclease promotes transcription termination by RNA polymerase II. *Nature* 432, 517–522. [PubMed: 15565157]
- Kim H, Erickson B, Luo W, Seward D, Graber JH, Pollock DD, Megee PC, and Bentley DL (2010). Gene-specific RNA polymerase II phosphorylation and the CTD code. *Nat. Struct. Mol. Biol.* 17, 1279–1286. [PubMed: 20835241]
- Kobirumaki F, Miyauchi Y, Fukami K, and Tanaka H (2005). A novel UbcH10-binding protein facilitates the ubiquitinylation of cyclin B in vitro. *J. Biochem.* 137, 133–139. [PubMed: 15749827]
- Kuehner JN, Kaufman JW, and Moore C (2017). Stimulation of RNA polymerase II ubiquitination and degradation by yeast mRNA 3'-end processing factors is a conserved DNA damage response in eukaryotes. *DNA Repair (Amst.)* 57, 151–160. [PubMed: 28783563]
- Lackford B, Yao C, Charles GM, Weng L, Zheng X, Choi EA, Xie X, Wan J, Xing Y, Freudenberg JM, et al. (2014). Fip1 regulates mRNA alternative polyadenylation to promote stem cell self-renewal. *EMBO J.* 33, 878–889. [PubMed: 24596251]
- Larochelle M, Robert MA, Hébert JN, Liu X, Matteau D, Rodrigue S, Tian B, Jacques PE, and Bachand F (2018). Common mechanism of transcription termination at coding and noncoding RNA genes in fission yeast. *Nat. Commun.* 9, 4364. [PubMed: 30341288]
- Li H, and Durbin R (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760. [PubMed: 19451168]
- Li H, and Durbin R (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 26, 589–595. [PubMed: 20080505]
- Li W, You B, Hoque M, Zheng D, Luo W, Ji Z, Park JY, Gunderson SI, Kalsotra A, Manley JL, and Tian B (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. [PubMed: 25906188]
- Liu X, Hoque M, Larochelle M, Lemay JF, Yurko N, Manley JL, Bach- and F, and Tian B (2017). Comparative analysis of alternative polyadenylation in *S. cerevisiae* and *S. pombe*. *Genome Res.* 27, 1685–1695. [PubMed: 28916539]
- Love MI, Huber W, and Anders S (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15, 550. [PubMed: 25516281]

- Luna R, Jimeno S, Marín M, Huertas P, García-Rubio M, and Aguilera A (2005). Interdependence between transcription and mRNP processing and export, and its impact on genetic stability. *Mol. Cell* 18, 711–722. [PubMed: 15949445]
- Lutz AP, Schladebeck S, Renicke C, Spadaccini R, Mösch HU, and Taxis C (2018). Proteasome activity is influenced by the HECT<sub>2</sub> Protein Ipa1 in budding yeast. *Genetics* 209, 157–171. [PubMed: 29519818]
- Mapendano CK, Lykke-Andersen S, Kjems J, Bertrand E, and Jensen TH (2010). Crosstalk between mRNA 3' end processing and transcription initiation. *Mol. Cell* 40, 410–422. [PubMed: 21070967]
- Mariconti L, Loll B, Schlinkmann K, Wengi A, Meinhart A, and Dichtl B (2010). Coupled RNA polymerase II transcription and 3' end formation with yeast whole-cell extracts. *RNA* 16, 2205–2217. [PubMed: 20810619]
- Mason PB, and Struhl K (2005). Distinction and relationship between elongation rate and processivity of RNA polymerase II in vivo. *Mol. Cell* 17, 831–840. [PubMed: 15780939]
- Mayer A, Lidschreiber M, Siebert M, Leike K, Söding J, and Cramer P (2010). Uniform transitions of the general RNA polymerase II transcription complex. *Nat. Struct. Mol. Biol.* 17, 1272–1278. [PubMed: 20818391]
- Mayer A, Schreieck A, Lidschreiber M, Leike K, Martin DE, and Cramer P (2012). The spt5 C-terminal region recruits yeast 3' RNA cleavage factor I. *Mol. Cell. Biol.* 32, 1321–1331. [PubMed: 22290438]
- McGinty RJ, Puleo F, Aksenova AY, Hisey JA, Shishkin AA, Pearson EL, Wang ET, Housman DE, Moore C, and Mirkin SM (2017). A defective mRNA cleavage and polyadenylation complex facilitates expansions of transcribed (GAA)<sub>n</sub> repeats associated with Friedreich's Ataxia. *Cell Rep.* 20, 2490–2500. [PubMed: 28877480]
- Mischo HE, and Proudfoot NJ (2013). Disengaging polymerase: Terminating RNA polymerase II transcription in budding yeast. *Biochim. Biophys. Acta* 1829, 174–185. [PubMed: 23085255]
- Moqtaderi Z, Geisberg JV, Jin Y, Fan X, and Struhl K (2013). Species-specific factors mediate extensive heterogeneity of mRNA 3' ends in yeasts. *Proc. Natl. Acad. Sci. USA* 110, 11073–11078. [PubMed: 23776204]
- Mueller CL, Porter SE, Hoffman MG, and Jaehning JA (2004). The Paf1 complex has functions independent of actively transcribing RNA polymerase II. *Mol. Cell* 14, 447–456. [PubMed: 15149594]
- Nedea E, He X, Kim M, Pootoolal J, Zhong G, Canadien V, Hughes T, Buratowski S, Moore CL, and Greenblatt J (2003). Organization and function of APT, a subcomplex of the yeast cleavage and polyadenylation factor involved in the formation of mRNA and small nucleolar RNA 3'-ends. *J. Biol. Chem.* 278, 33000–33010. [PubMed: 12819204]
- Nojima T, Dienstbier M, Murphy S, Proudfoot NJ, and Dye MJ (2013). Definition of RNA polymerase II CoTC terminator elements in the human genome. *Cell Rep.* 3, 1080–1092. [PubMed: 23562152]
- Nordick K, Hoffman MG, Betz JL, and Jaehning JA (2008). Direct interactions between the Paf1 complex and a cleavage and polyadenylation factor are revealed by dissociation of Paf1 from RNA polymerase II. *Eukaryot. Cell* 7, 1158–1167. [PubMed: 18469135]
- Offenbacher S, Divaris K, Barros SP, Moss KL, Marchesan JT, Morelli T, Zhang S, Kim S, Sun L, Beck JD, et al. (2016). Genome-wide association study of biologically informed periodontal complex traits offers novel in-sights into the genetic basis of periodontal disease. *Hum. Mol. Genet.* 25, 2113–2129. [PubMed: 26962152]
- Ozsolak F, Kapranov P, Foissac S, Kim SW, Fishilevich E, Monaghan AP, John B, and Milos PM (2010). Comprehensive polyadenylation site maps in yeast and human reveal pervasive alternative polyadenylation. *Cell* 143, 1018–1029. [PubMed: 21145465]
- Pearson E, and Moore C (2014). The evolutionarily conserved Pol II flap loop contributes to proper transcription termination on short yeast genes. *Cell Rep.* 9, 821–828. [PubMed: 25437538]
- Peart N, Sataluri A, Baillat D, and Wagner EJ (2013). Non-mRNA 3' end formation: How the other half lives. *Wiley Interdiscip. Rev. RNA* 4, 491–506. [PubMed: 23754627]

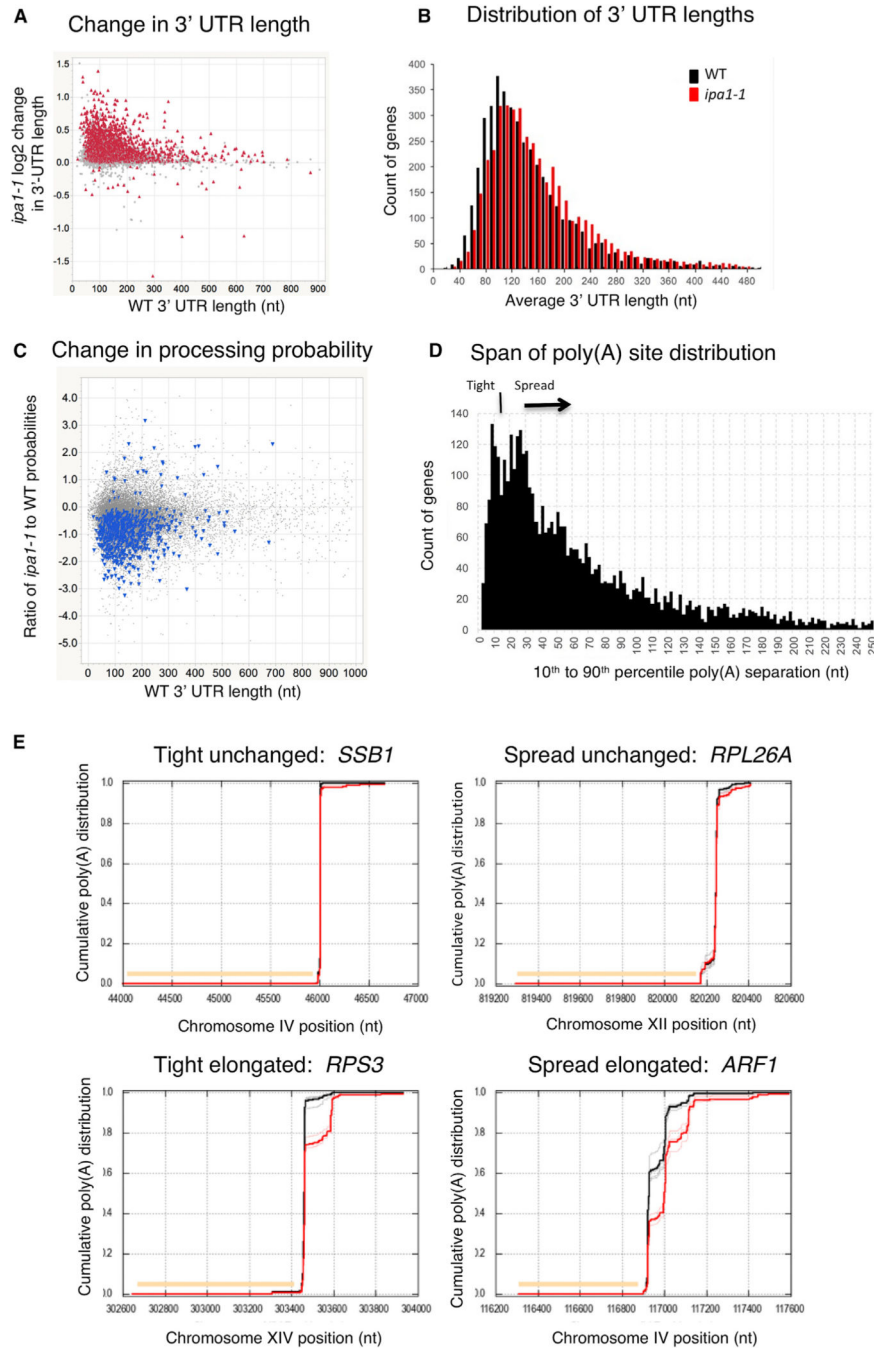
- Penheiter KL, Washburn TM, Porter SE, Hoffman MG, and Jaehning JA (2005). A posttranscriptional role for the yeast Paf1-RNA polymerase II complex is revealed by identification of primary targets. *Mol. Cell* 20, 213–223. [PubMed: 16246724]
- Porrua O, Boudvillain M, and Libri D (2016). Transcription termination: Variations on common themes. *Trends Genet.* 32, 508–522. [PubMed: 27371117]
- Powell W, and Reines D (1996). Mutations in the second largest subunit of RNA polymerase II cause 6-azauracil sensitivity in yeast and increased transcriptional arrest in vitro. *J. Biol. Chem.* 271, 6866–6873. [PubMed: 8636112]
- Ramírez F, Ryan DP, Grüning B, Bhardwaj V, Kilpert F, Richter AS, Heyne S, Dündar F, and Manke T (2016). deepTools2: A next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* 44, 160–165.
- Riles L, Shaw RJ, Johnston M, and Reines D (2004). Large-scale screening of yeast mutants for sensitivity to the IMP dehydrogenase inhibitor 6-azauracil. *Yeast* 21, 241–248. [PubMed: 14968429]
- Schaughency P, Merran J, and Corden JL (2014). Genome-wide mapping of yeast RNA polymerase II termination. *PLoS Genet.* 10, e1004632. [PubMed: 25299594]
- Sen R, Kaja A, Ferdoush J, Lahudkar S, Barman P, and Bhaumik SR (2017). An mRNA capping enzyme targets FACT to the active gene to enhance the engagement of RNA Polymerase II into transcriptional elongation. *Mol. Cell. Biol.* 37, Published online June 15, 2017. 10.1128/ MCB.00029-17.
- Shalem O, Sharon E, Lubliner S, Regev I, Lotan-Pompan M, Yakhini Z, and Segal E (2015). Systematic dissection of the sequence determinants of gene 3' end mediated expression control. *PLoS Genet.* 11, e1005147. [PubMed: 25875337]
- Shandilya J, and Roberts SG (2012). The transcription cycle in eukaryotes: From productive initiation to RNA polymerase II recycling. *Biochim. Biophys. Acta* 1819, 391–400. [PubMed: 22306664]
- Squazzo SL, Costa PJ, Lindstrom DL, Kumer KE, Simic R, Jennings JL, Link AJ, Arndt KM, and Hartzog GA (2002). The Paf1 complex physically and functionally associates with transcription elongation factors in vivo. *EMBO J.* 21, 1764–1774. [PubMed: 11927560]
- Tian B, and Graber JH (2012). Signals for pre-mRNA cleavage and polyadenylation. *Wiley Interdiscip. Rev. RNA* 3, 385–396. [PubMed: 22012871]
- Tian B, and Manley JL (2017). Alternative polyadenylation of mRNA precursors. *Nat. Rev. Mol. Cell Biol.* 18, 18–30. [PubMed: 27677860]
- Tomson BN, and Arndt KM (2013). The many roles of the conserved eukaryotic Paf1 complex in regulating transcription, histone modifications, and disease states. *Biochim. Biophys. Acta* 1829, 116–126. [PubMed: 22982193]
- Tous C, Rondón AG, García-Rubio M, González-Aguilera C, Luna R, and Aguilera A (2011). A novel assay identifies transcript elongation roles for the Nup84 complex and RNA processing factors. *EMBO J.* 30, 1953–1964. [PubMed: 21478823]
- van Helden J, del Olmo M, and Pérez-Ortín JE (2000). Statistical analysis of yeast genomic downstream sequences reveals putative polyadenylation signals. *Nucleic Acids Res.* 28, 1000–1010. [PubMed: 10648794]
- van Pel DM, Stirling PC, Minaker SW, Sipahimalani P, and Hieter P (2013). *Saccharomyces cerevisiae* genetics predicts candidate therapeutic genetic interactions at the mammalian replication fork. *G3 (Bethesda)* 3, 273–282. [PubMed: 23390603]
- Xu Z, Wei W, Gagneur J, Perocchi F, Clauder-Münster S, Camblong J, Guffanti E, Stutz F, Huber W, and Steinmetz LM (2009). Bidirectional promoters generate pervasive transcription in yeast. *Nature* 457, 1033–1037. [PubMed: 19169243]
- Yao C, Choi EA, Weng L, Xie X, Wan J, Xing Y, Moresco JJ, Tu PG, Yates JR 3rd, and Shi Y (2013). Overlapping and distinct functions of CstF64 and CstF64 $\tau$  in mammalian mRNA 3' processing. *RNA* 19, 1781–1790. [PubMed: 24149845]
- Yoon OK, and Brem RB (2010). Noncanonical transcript forms in yeast and their regulation during environmental stress. *RNA* 16, 1256–1267. [PubMed: 20421314]



- Yu M, Yang W, Ni T, Tang Z, Nakadai T, Zhu J, and Roeder RG (2015). RNA polymerase II-associated factor 1 regulates the release and phosphorylation of paused RNA polymerase II. *Science* 350, 1383–1386. [PubMed: 26659056]
- Zhao J, Kessler M, Helmling S, O'Connor JP, and Moore C (1999). Pta1, a component of yeast CF II, is required for both cleavage and poly(A) addition of mRNA precursor. *Mol. Cell. Biol.* 19, 7733–7740. [PubMed: 10523662]

### Highlights

- Ipa1 mutation delays transcription termination and decreases elongation efficiency
- Specific configurations of poly(A) signals define an Ipa1-responsive site
- Ipa1 stabilizes Ysh1 levels and facilitates CPF recruitment to a transcribed gene
- Ysh1 overexpression rescues the *ipa1-1* termination, but not the elongation, defect



**Figure 1. Loss of Ipa1 Function Leads to Transcriptome-wide Reduction in Polyadenylation Activity and a Correlated Increased Average Length of mRNAs**

(A) Plot of change in the average 3' UTR length for each gene. Each gene is represented by a single point, with the change in 3' UTR in the *ipa1-1* mutant on the y reaxis and the WT average on the x axis. A t test on the average 3' UTR length was performed, followed by an FDR correction. Genes that pass a threshold of FDR <0.2 are highlighted in red. Those without significant changes are indicated in gray.

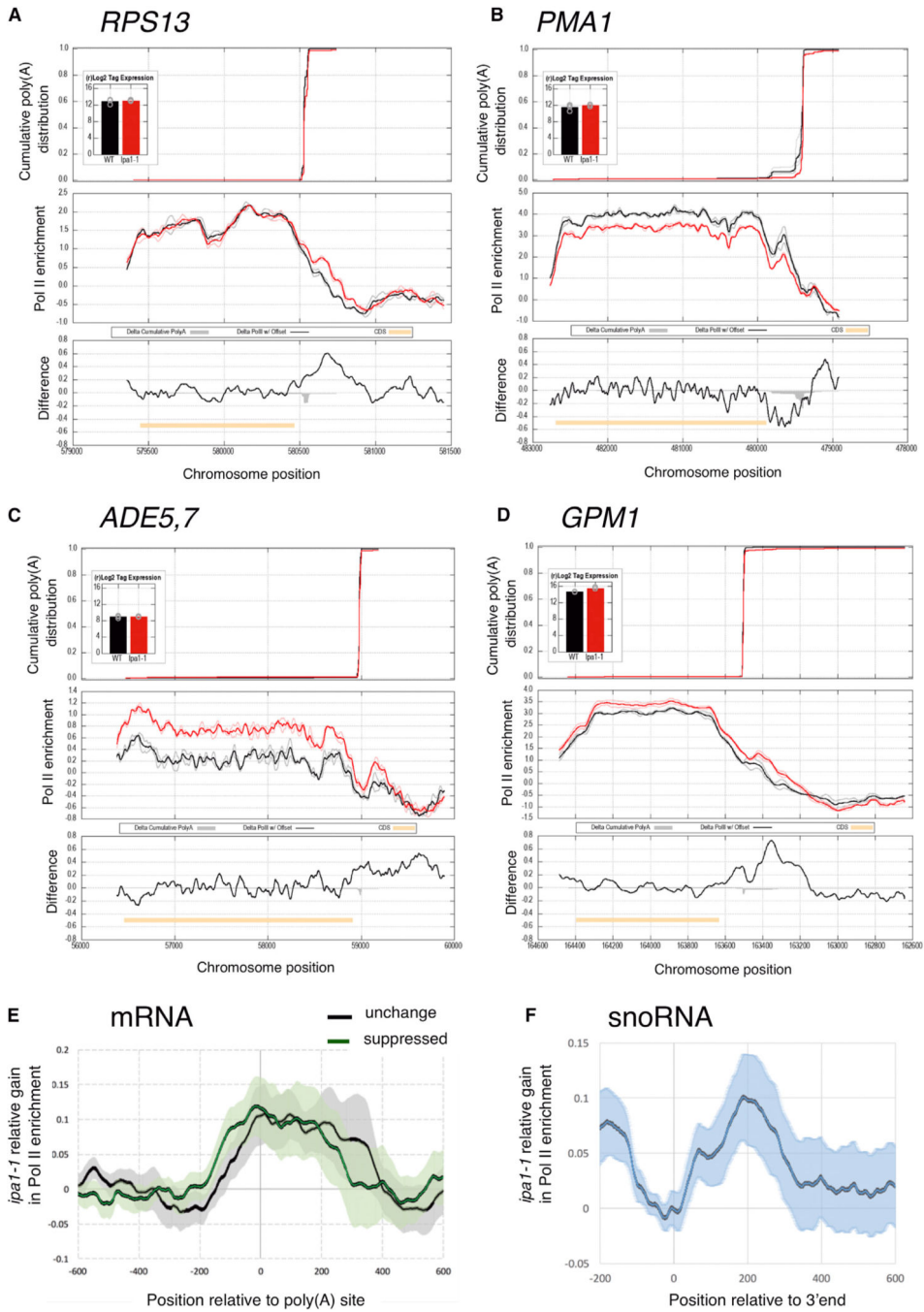
(B) The transcriptome-wide distribution of 3' UTR lengths. The number of genes in each 10 nt bin of 3' UTR length is plotted for WT (black) and *ipa1-1* (red). For the two plots in

Figure 1B, a Kolmogorov-Smirnov test on the average 3' UTR lengths gives a D-statistic = 0.091, which for matched sample sizes of 4,377, gives a significance level for rejection of the null hypothesis (that the two length distributions are equal) of approximately  $1.0e-16$ , indicating a significant difference in the WT and *ipa1-1* datasets.

(C) Site-specific changes in polyadenylation processing probability (represented on the y axis as base-2 logarithm of the ratio of *ipa1-1* to WT probabilities) plotted against the WT 3' UTR length. Each point represents a single poly(A) site. Significantly altered sites were identified based on a t test of calculated probabilities for four WT replicates versus three *ipa1-1* samples, followed by an FDR adjustment. Sites with an FDR  $<0.2$  are highlighted as blue triangles. Those without significant changes are indicated in gray.

(D) The span of poly(A) site distribution within each gene for WT cells is plotted, using the 5'-to-3' CPD, as described in (E), to measure the nucleotide separation between the 10<sup>th</sup> and 90<sup>th</sup> percentiles.

(E) Examples of the CPD of genes from each operational classes—tight unchanged, spread unchanged, tight elongated, and spread elongated. The average CPD illustrates how poly(A) site usage shifts in the *ipa1-1* mutant compared to WT. The CPD profiles were generated from genome-wide poly(A) site mapping of mRNA from WT and *ipa1-1* cells (Costanzo et al., 2016). TPa1 Promotes Pol II Transcription Terminationhe CPD plot for each gene reads from 5' to 3' through the gene and represents the empirical likelihood of polyadenylation at or before each position in the transcript. The four WT replicates are traced in gray and the three *ipa1-1* replicates in light red, and the average is shown in black or red. The yellow bar represents the CDS.



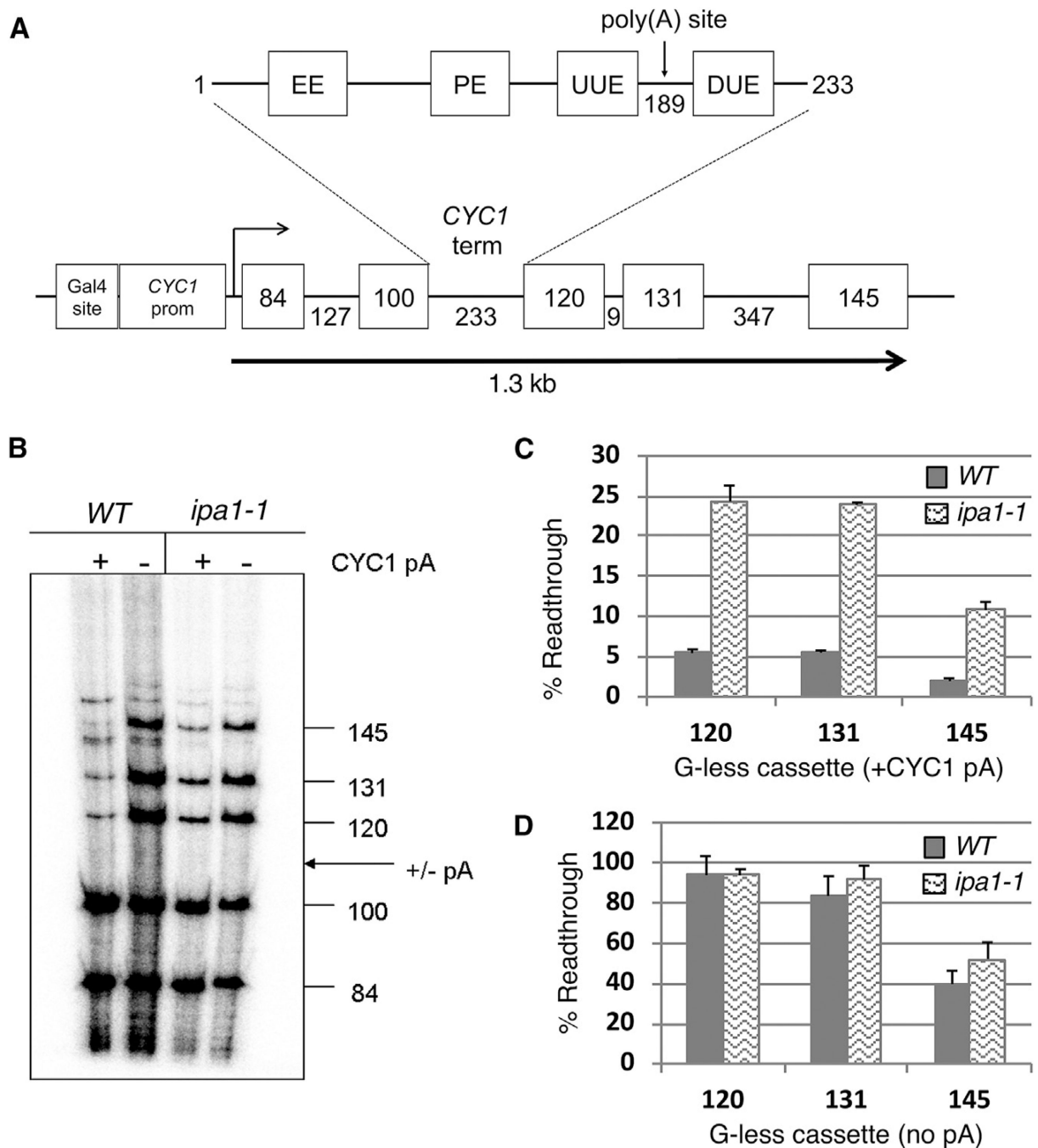
**Figure 2. The *ipa1-1* Mutation Causes Pol II Enrichment Downstream of Most Poly(A) Sites** (A–D) Analysis of poly(A) site distribution and Pol II occupancy of the *RPS13* (A), *PMA1* (B), *ADE5,7*(C), and *GPM1* (D) genes using RNA sequencing data and ChIP-seq analysis. Top panel: the average CPD illustrates poly(A) site usage in the *ipa1-1* mutant compared to WT. The CPD profiles were generated as described for Figure 1E using RNA sequencing tag counts from the genome-wide poly(A) site mapping data of Costanzo et al. (2016). The four WT replicates are traced in gray and the three *ipa1-1* replicates in light red, and the average is shown in black or red. The expression levels in WT and mutant determined from RNA

sequencing tag counts of full-length mRNAs are shown in the inset and in Table S3. Middle panel: Pol II enrichment determined by ChIP-seq, with the two WT replicates traced in gray and the two *ipa1-1* replicates in light red, and the average shown in black or red.

Bottom panel: the difference in Pol II occupancy between *ipa1-1* and WT after the *ipa1-1* value has been scaled to match the WT average value in the CDS (indicated in yellow). The gray area represents the region with changes in poly(A) site usage.

(E) Metagene analysis of locally normalized Pol II enrichment change anchored at the poly(A) site in WT and *ipa1-1* cells. The Pol II profile on genes with unchanged sites is traced in black and that of genes with *ipa1-1* suppressed sites in green. Plots are shown as the average across all genes, with the lightly shaded areas representing the error bars shown as SEM.

(F) Metagene analysis of differential Pol II occupancy at snoRNA genes in WT and *ipa1-1* cells.



**Figure 3. Pol II Termination *In Vitro* Is Less Efficient in the *ipa1-1* Mutant**

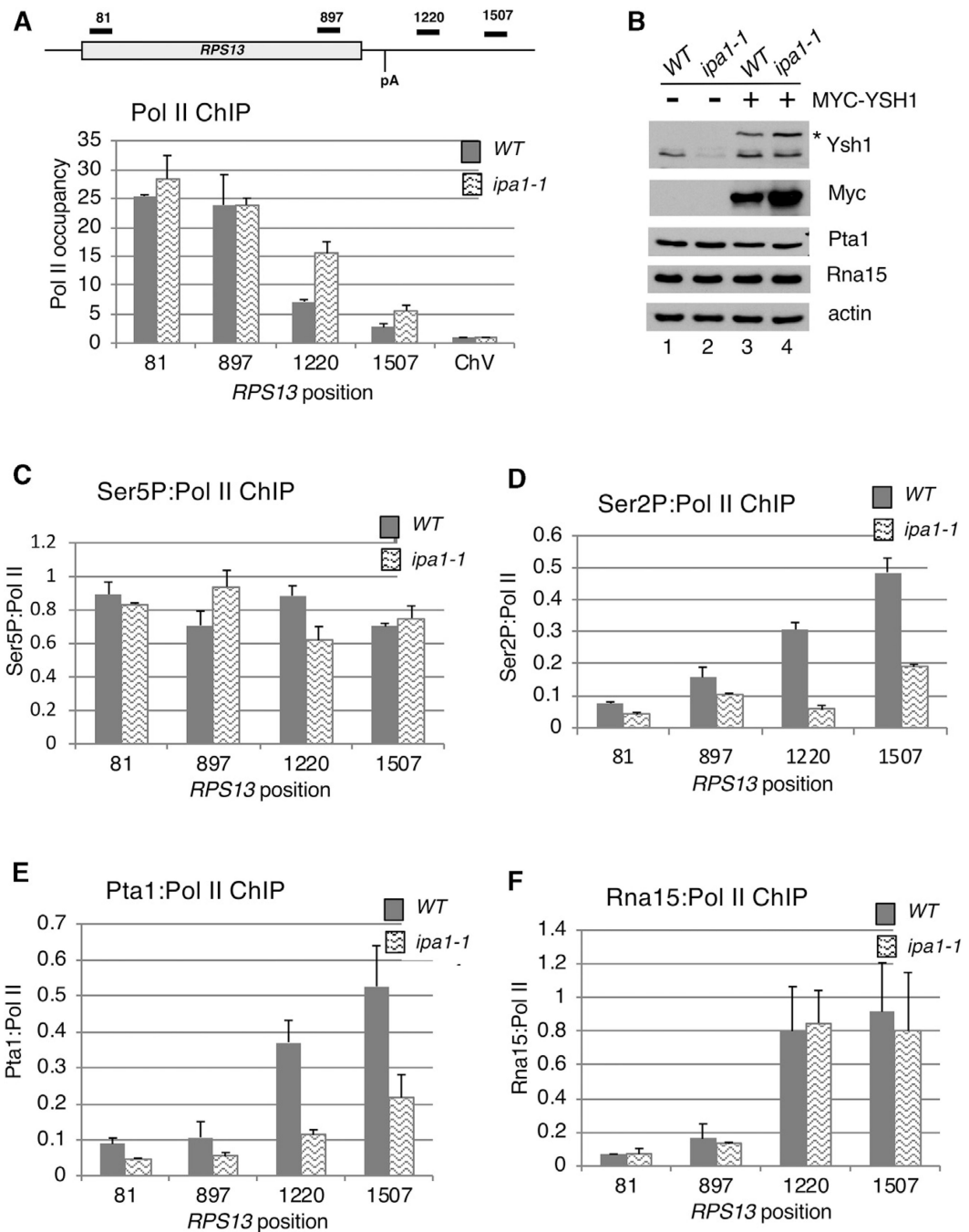
(A) Tandem G-less cassette transcription template. The transcription start site, the position, and lengths of the G-less cassettes, the position of the inserted *CYC1* terminator and location of the poly(A) site, and the distance in kilobases (kb) from the transcription start site to the end of the last cassette are indicated. EE, efficiency element; PE, positioning element; UUE, upstream U-rich element; DUE, downstream U-rich element.

(B) Radio-labeled G-less cassette transcription fragments synthesized in WT and *ipa1* extracts were resolved on a 6% polyacrylamide/7M urea gel. The two transcription

templates contain the *CYCI* poly(A) signal (*CYCI*) or no poly(A) elements (no pA). Lengths, in bases, of the G-less cassettes produced upon T1 RNase digestion of transcript are indicated.

(C and D) Quantification of transcription products in (B). The signals from the 120, 131, and 145 nt G-less cassettes in WT (solid gray) and *ipa1-1* (wavy lines) extracts are normalized to that of the 100 nt G-less cassette for the transcription template with the *CYCI* poly(A) signal (C) or with no poly(A) element (D). Error bars represent the SD from the average values of three independent experiments.





**Figure 4. The *ipa1-1* Mutant Causes Changes in Phosphorylation of Pol II CTD and Recruitment of the CPF 3' End-Processing Factor**

(A) Pol II occupancy in WT (solid gray) and *ipa1-1* (wavy lines) strains at indicated *RPS13* positions. The top panel shows positions of primer pairs used in the ChIP analysis in base pairs downstream of the start codon. Pol II signals were obtained with the 4H8 antibody, which recognizes both phosphorylated and unphosphorylated forms of the CTD. The y axis indicates fold enrichment over the non-transcribed background signal at the intergenic region on Chromosome V (ChrV), and error bars show SE calculated from two or three independent biological replicates, each with two technical replicates.

(B) Endogenous Ysh1 is depleted in the absence of functional Ipa1. Western blots of extracts prepared from WT and *ipa1-1* strains harboring either pRS315 or pRS315-Myc-YSH1 plasmids show the abundance of endogenous Ysh1, exogenous Myc-Ysh1, Pta1, and Rna15. Actin is included as a loading control. The Myc-Ysh1 band detected with the Ysh1 antibody is marked with an asterisk.

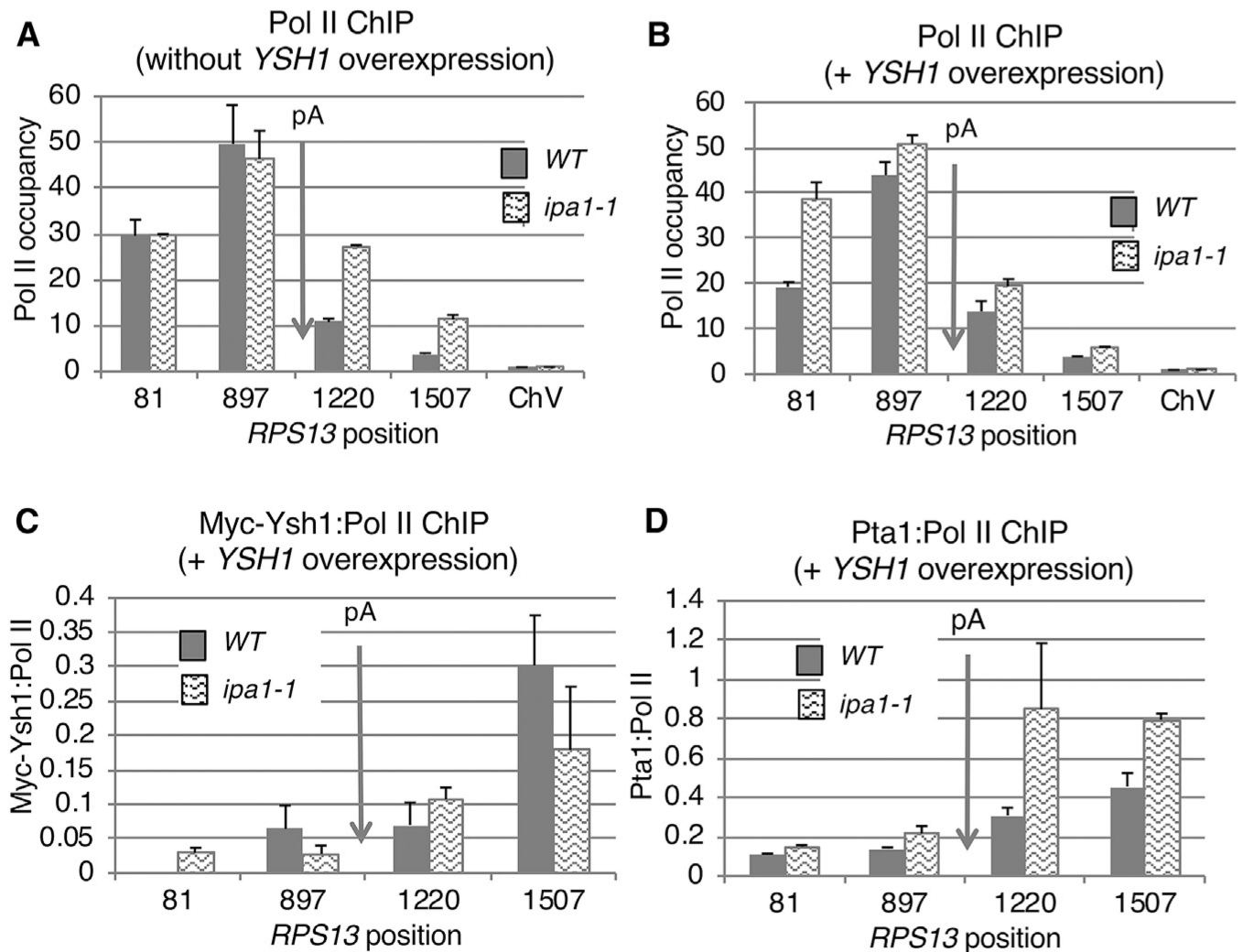
(C) Ser5P:Pol II occupancy in WT and *ipa1-1* strains at *RPS13* positions.

(D) Ser2P:Pol II occupancy in WT and *ipa1-1* strains at *RPS13* positions.

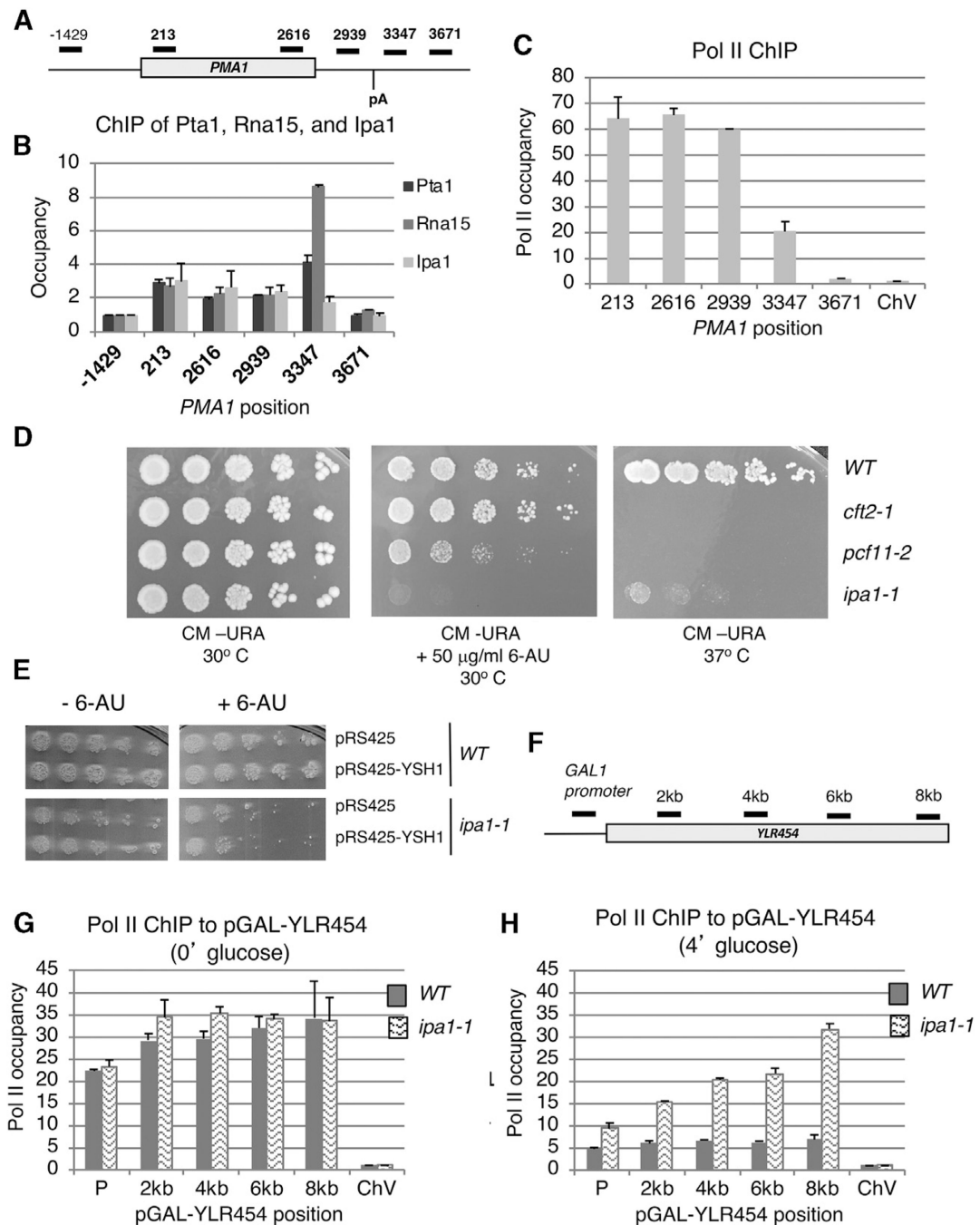
(E) Pta1:Pol II occupancy in WT and *ipa1-1* strains at *RPS13* positions.

(F) Rna15:Pol II occupancy in WT and *ipa1-1* strains at *RPS13* positions.

For (C–F), ChIP was conducted with antibodies against Pta1, Rna15, or Pol II CTD, and qPCR signals were normalized to that of Pol II. For these and ChIP analyses presented in Figures 5 and 6, error bars show SE from two to four independent experiments.



**Figure 5. Overexpression of the Ysh1 Endonuclease Rescues the *ipa1-1* Termination Defect** (A and B) Pol II occupancy in WT and *ipa1-1* strains harboring either pRS315 (A) or pRS315\_Myc-YSH1 (B), respectively, at indicated *RPS13* positions. (C and D) Myc-Ysh1:Pol II occupancy (C) and Pta1:Pol II occupancy (D), respectively, in WT and *ipa1-1* strains harboring pRS315-Myc\_YSH1. Error bars show standard error from two to four independent experiments.



### Figure 6. *IPAI* Facilitates Proper Transcription Elongation Kinetics

(A) Positions of primer pairs used in ChIP analysis in base pairs relative to the start codon of *PMA1*.

(B) Pta1, Rna15, and Ipa1-Myc occupancy across the *PMA1* gene in WT cells.

(C) Pol II occupancy across the *PMA1* gene in WT cells.

(D) Serial dilution spot assay of WT and mutant strains on media in the absence or presence of 6-AU at the indicated temperatures.

(E) Overexpression of exogenous Ysh1 cannot rescue the 6-azauracil sensitivity in the absence of Ipa1. A serial dilution spot assay was performed using *IPA1* and *ipa1-1* strains harboring the 2 mm, high-copy pRS425 or pRS425-YSH1 plasmids on media in the absence and presence of 6-AU.

(F) Schematic of the galactose-inducible YLR454 locus.

(G and H) Pol II occupancy in WT and *ipa1-1* at the indicated *YLR454* positions in (G) galactose (0' glucose) and or (H) 4 min after glucose addition (4' glucose).

Error bars show standard error from two to four independent experiments.

Table 1.

Percentage of Sequences that Match Optimal Poly(A) Signal Elements

		Poly(A) site					
		Efficiency	Positioning	Upstream	Downstream	Cleavage Y-A	Downstream +1 to +20
		5'	-----TATATA-----AATAAA-----TTTTTT--YA <sub>n</sub> --TTTTTT-----	3'	TTTTTT	TTTTTT	TTTTTT
		-60 to -20	-30 to -10	-15 to -5	+1 to +20		
<i>ipa1-1</i> Variation	WT Poly(A) Config	n	Efficiency -60 to -20 TATATA	Positioning -30 to -10 AATAAA	Upstream -15 to -5 TTTTTT	Cleavage Y-A	Downstream +1 to +20 TTTTTT
Suppressed site (elongated transcript)	spread	772	23.7 (L) <sup>a,b</sup>	7.1	6.5	51.6 (L) <sup>a</sup>	6.7 (L) <sup>a,b</sup>
	tight	176	44.9 (H) <sup>a,b</sup>	11.4 (H) <sup>a,b</sup>	6.8 (H) <sup>a</sup>	50.0 (L) <sup>a</sup>	9.1
Unchanged site (unchanged transcript)	spread	917	27.9 (L) <sup>a,b</sup>	7.2	4.8	55.8	12.3 (H) <sup>a,b</sup>
	tight	391	49.1 (H) <sup>a,b</sup>	5.4 (L) <sup>a</sup>	5.1	58.3 (H) <sup>a</sup>	12.5 (H) <sup>a</sup>
Union of all sites		2,256	31.5	7.2	5.6	54.3	10.2
95% confidence interval			29.6–33.4	6.1–8.2	4.6–6.5	52.3–56.4	8.9–11.4

The first row shows the optimal yeast poly(A) control elements and spacing.

<sup>a</sup> Values that are higher (H) or lower (L), respectively, than the 95% confidence interval calculated for the union of all sites. A simple test on equal proportions, with the null hypothesis that each of the four subsets are random draws from the union of all four, was also performed.

<sup>b</sup> Probability associated with the resulting Z score is below 0.05. The cases that are outside the 95% confidence interval, but not  $p < 0.05$ , are all in the range  $0.1 < p < 0.2$ , with the exception of the upstream signal in suppressed-tight, where the relatively smaller number of sites (176) is likely the cause.

## KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER Add
Antibodies		
RNA Pol II pan-CTD	Santa Cruz	4H8, Cat# sc-47701; RRID:AB_677353
Rna15	Horst Domdey	Rabbit polyclonal
Pta1	Craig Peebles	Mouse monoclonal
Ysh1	Horst Domdey	Rabbit polyclonal
Myc	E10	Tufts Antibody and Cell Culture Facility
RNA Pol II Ser5P CTD	Active Motif	3E8, Cat# 61085; RRID:AB_2687451
RNA Pol II Ser2P CTD	Covance	H5, Cat# MMS-129R-200; RRID:AB_10143905
beta Actin	Abcam	Cat# ab8224; RRID:AB_449644
Rabbit IgG-HRP	Fisher	OB 4050-05
Mouse IgG-HRP	BioRad	1705047
Bacterial and Virus Strains		
DH5a	Lab stock	N/A
Chemicals, Peptides, and Recombinant Proteins		
Anti-mouse IgM - Agarose	Abcam	ab65867
EDTA-free protease inhibitor cocktail	Fisher	50-720-3178
RNase T1	Ambion	AM2282
Protein A beads	Santa Cruz	SC2001
Protein G beads	Santa Cruz	SC2002
Proteinase K	US Biological	P9100
qPCR grade water	Fisher	10-977-015
Critical Commercial Assays		
SYBR Green PCR master mix	BioRad	1708885
MinElute columns	QIAGEN	28004
TruSeq ChIP Sample Prep Kit	Illumina	IP-202-1012
SuperSignal West Pico PLUS Chemiluminescent Substrate	lhermoFisher	3458Q
Deposited Data		
ChIP-seq datasets	This paper	GEO: GSE117402
Experimental Models: Organisms/Strains		
Yeast strain BY4741 (Wild-type)	Charles Boone, University of Toronto	(Costanzo et al., 2016)
Yeast strain TSA1248 (BY4741 with the <i>ipa1-1</i> mutation)	Charles Boone, University of Toronto	(Costanzo et al., 2016)
Yeast strain TS801 (BY4741 with the <i>cfi2-1</i> mutation)	Charles Boone, University of Toronto	(Costanzo et al., 2016)
Yeast strain TSA685 (BY4741 with the <i>pcf11-2</i> mutation)	Charles Boone, University of Toronto	(Costanzo et al., 2016)
BY4741 strain with GAL:YLR454	This paper	ELP1

<b>REAGENT or RESOURCE</b>	<b>SOURCE</b>	<b>IDENTIFIER Add</b>
TSA1248 with GAL:YLR454	This paper	ELP2
Oligonucleotides		
PCR primers	Integrated DNA Technologies	See Table S4
Recombinant DNA		
pKS708	Bernhard Dichtl Universität Zürich	(Mariconti et al., 2010)
pKS710	Bernhard Dichtl Universität Zürich	(Mariconti et al., 2010)
pRS315	Addgene	N/A
pRS315-MYC-YSH1	This paper	N/A
pRS425	Addgene	N/A
pRS425-YSH1	This paper	N/A

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript