Transcription termination promotes splicing efficiency and fidelity in a compact genome

Keaton Barr[1,2], Kevin L He[1,2], Andreas J. Krumbein[1] and Guillaume F. Chanfreau[1,3]

1 Department of Chemistry and Biochemistry and the Molecular Biology Institute, University of California Los Angeles, Los Angeles CA 90095-1569.

2. Equal contribution.

3. Correspondence: guillom@ucla.edu

Abstract

Splicing of terminal introns is coupled to 3′-end processing by cleavage and polyadenylation (CPA) of mRNAs in mammalian genes. Whether this functional coupling is universally conserved across eukaryotes is unknown. Here we show using long read RNA sequencing in *S.cerevisiae* that splicing inactivation does not result in widespread CPA impairment and that inactivation of CPA does not lead to global splicing defects. However, 5′-extensions due to termination defects from upstream genes lead to splicing inhibition in a length-dependent manner. Additionally, for some extended RNAs resulting from failed termination, we observed decreased splicing fidelity resulting in novel intergenic and long-range intragenic splicing events. These results argue against a broad coupling of splicing to CPA in *S.cerevisiae* but show that efficient CPA-mediated transcription termination is critical for splicing fidelity and efficiency in a compact genome.

Introduction

Synthesis of most eukaryotic mRNAs requires three main processing steps: capping of the 5′-end of the mRNA by addition of a 7-methyl guanosine(7-meG), removal of intronic sequences by splicing, and 3′-end formation by cleavage and polyadenylation (CPA). In addition to generating mRNA 3′-ends, CPA also promotes transcription termination by RNA polymerase II. The three major mRNA processing reactions occur co-transcriptionally, as factors involved are known to associate with RNA polymerase II and/or the nascent transcripts (reviewed in [1]). These three main processing reactions also influence each other. At the 5′-end of the mRNA, capping stimulates splicing of the first, cap-proximal intron ([2–4]; reviewed in [5]). This stimulation is conferred by the cap-binding complex (CBC), which binds the 7-meG cap and promotes recruitment of the U1 snRNP, resulting in enhanced recognition of the first, cap-proximal exon 5′ splice sites (SS) by splicing factors [2,3,6.]. At the 3′-end of the mRNA, splicing of the terminal intron and recognition of the CPA site are intimately connected (reviewed in [1,7]). Interestingly, this coupling of splicing and 3′-end processing can occur independently from transcription[8,9], and several interactions have been described between 3′-end processing factors and splicing factors. At the genomic scale, perturbation of 3′-end processing has been shown to result in inhibition of splicing of terminal introns[10].

While there is broad evidence for reciprocal stimulation of splicing of terminal introns and 3′-end processing by CPA in mammalian genes, it is unclear whether this mechanism is universally conserved across eukaryotes. In support of this conservation, inactivation of the *S.pombe* homologue of splicing factor U2AF65 triggers 3′-end formation defects and transcriptional read-through[11]. In the yeast *S.cerevisiae*, most intron-containing genes contain only a single intron, which defines most introns as *de facto* ′terminal′ and close to the cleavage and polyadenylation sites, especially for introns of ribosomal protein genes which constitute the quantitative majority of splicing substrates [12]. Some indirect evidence suggests that there might exist functional coupling between splicing and 3′-end formation in *S.cerevisiae*. For instance, the Ysh1 endonuclease involved in the cleavage reaction of CPA was first identified in a genetic screen for splicing mutants[13,14] . In addition, physical and genetic interactions have been identified between splicing factors and the CPA machinery (reviewed in [15]). However no study has yet directly addressed the impact of inactivating CPA factors on splicing, or of inactivating splicing factors on 3′-end formation. In this study we address this question, using long-read RNA sequencing technology. Using rapid inactivation of splicing or 3′-end processing factors, we do not find evidence for broad reciprocal coupling between splicing and 3′-end processing. However, loss of termination promoting readthrough into intron-containing genes results in decreased splicing efficiency, likely because of the increased distance of the intron from the cap structure. In addition, loss of termination revealed dormant splice sites and resulted in the production of intergenic spliced mRNAs. These results underscore the importance of proper CPA-mediated termination to limit aberrant splicing events and enhance splicing efficiency.

Inactivation of the U1 snRNP component Snp1 does not lead to prevalent 3′-end formation or termination defects.

To analyze the reciprocal impact of splicing or 3′-end processing factors on splicing and polyadenylation, we assessed the impact of rapidly inactivating these factors on the transcriptome. Most splicing or 3′-end processing factors are essential in *S.cerevisiae*. To rapidly inactivate these RNA processing factors, we used the anchor away (AA) technique[16], which promotes rapid export of a nuclear protein of choice out of the nucleus, resulting in their

functional inactivation. This technique had previously been used to rapidly inactivate splicing[17] or 3′-end processing [18,19] factors in *S.cerevisiae*. We used AA strains of the 3′-end processing factors Ysh1 and Rna15, as well as the U1 snRNP component Snp1p, the yeast orthologue of the U1-70K protein. We chose to inactivate this early splicing factor to prevent engagement of transcripts in the splicing pathway and binding of proteins that may recognize splicing signals following E complex formation. Following treatment with Rapamycin to induce the export of tagged proteins out of the nucleus, RNA processing efficiency was analyzed *in vivo* using long-read RNA sequencing. An untagged strain from the same genetic background (which alleviates any toxic effects of rapamycin treatment and downstream effects of gene expression) was treated for 2hrs with Rapamycin and used as a negative control (mentioned as WT in the remainder of this study). RNA samples were obtained for three replicates for each strain and sequenced by direct RNA sequencing using the Oxford Nanopore Technology (ONT) platform. We obtained an average of ca. 3 million reads per replicate (Table S1) and excellent reproducibility of RNA detection between replicates (Fig.S1). We then analyzed the impact of inactivation of these factors on 3′-end formation by measuring 3′ UTR length. Analysis of global 3′-UTR lengths did not reveal any global impact of Snp1 inactivation on 3′-UTR length (Fig.1A).

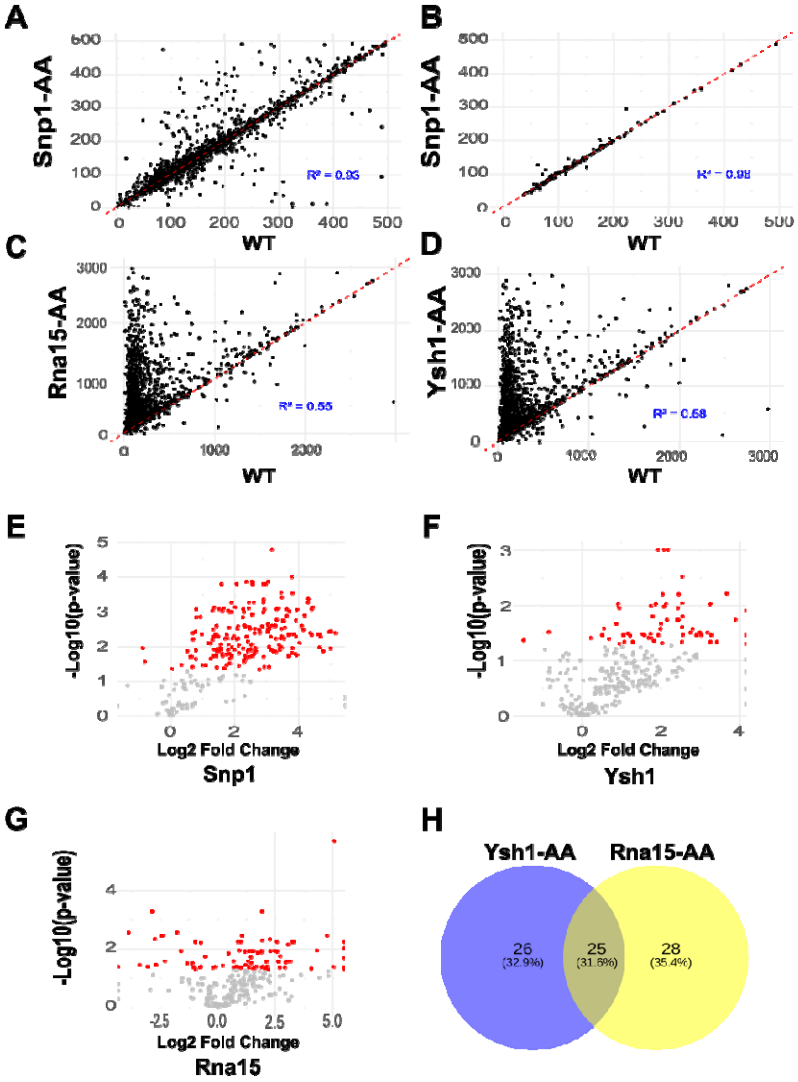


Figure 1. Global analysis of 3′-end formation and splicing upon CPA or splicing inactivation.

A. Comparison of 3′UTR length between the Snp1-AA strain and the WT control for all *S.cerevisiae* genes detected by Nanopore sequencing. Numbers indicate the 3′UTR lengths in nucleosides.

B. Comparison of 3′UTR length between the Snp1-AA strain and the WT control for all *S.cerevisiae* intron-containing genes detected by Nanopore sequencing.

C. Comparison of 3′UTR length between the Rna15-AA and the WT control for all *S.cerevisiae* genes detected by Nanopore sequencing.

D. Same as (C) for the Ysh1-AA strain.

E. Volcano plot showing the ICGs exhibiting splicing defects in the Snp1-AA strain compared to WT. Red dots correspond to ICGs showing significant splicing changes.
F. Same as (E) for the Ysh1-AA strain.
G. Same as (E) for the Rna15-AA strain.
H. Venn Diagram showing the overlap of ICGs showing significant splicing defects in the Ysh1-AA and Rna15-AA strains.

This lack of effect could be due to the fact that intron-containing genes (ICG) represent only ~5% of genes in *S.cerevisiae*. We therefore analyzed 3′-UTR length only for ICG, and did not detect any significant changes in UTR length upon Snp1 inactivation (Fig.1B). This is in contrast to what was observed for Ysh1 or Rna15 inactivation, which resulted in a global increase in 3′-UTR size (Figure 1C), as expected since loss of these factors is known to induce termination defects.
Since Snp1 inactivation resulted in general splicing defects (see below), we conclude that splicing inactivation at an early stage of the splicing pathway does not result in major perturbations of 3′-end processing *in vivo*.

Inactivation of the cleavage and polyadenylation factors Ysh1 and Rna15 does not induce widespread splicing defects

We next focused on analyzing the impact of inactivating Snp1, Ysh1 or Rna15 on pre-mRNA splicing. To this end, we quantitated unspliced reads in strains in which these factors were anchored away. Analysis of splicing efficiency in the Snp1-AA strain showed widespread increase of unspliced reads for many ICGs, as expected (Fig.1D). Some genes did not show significant accumulation of unspliced reads: either because their splicing is robust enough to resist to short-term inactivation of the U1 snRNP, or because unspliced pre-mRNAs accumulation is limited by rapid degradation [20,21]. By contrast, inactivation of Rna15 or Ysh1 had a more modest effect on splicing (Fig1E, 1F). Some splicing defects overlap was observed upon inactivation of Rna15 or Ysh1(Fig.1G). However inactivation of Ysh1 and Rna15 also resulted in unique set of genes for which splicing was affected (Fig.1G). This suggests that the impact of Ysh1 or Rna15 inactivation on splicing is not due to a general effect of CPA inactivation on splicing, but that each of these factors impact expression of the affected genes in a unique way which perturbs splicing indirectly, rather than through a general coupling mechanism between splicing and CPA.

Splicing defects are not prevalent in 3′-extended RNAs and unprocessed RNAs are subject to nuclear degradation

To understand the mechanistic basis for the perturbation of splicing of genes impacted by Ysh1 or Rna15 inactivation, we first analyzed the splicing signals of ICGs showing a significant splicing defect in the Ysh1-AA or Rna15-AA samples. Sequence logos obtained for the 5′SS or branchpoint sequences of the RNAs with splicing defects in the Ysh1 or Rna15-AA strain were similar to those of all ICG (Fig.S2), suggesting that deficient splicing for these mRNAs was not due to intrinsically weaker splicing signals. To better understand the mechanism responsible for these splicing defects, we inspected ONT reads of ICGs to search for the presence of unspliced RNAs in transcripts with 3′-readthroughs in the Ysh1-AA or Rna15-AA samples. In general, intron-containing transcripts with long 3′-extensions following Ysh1 or Rna15 inactivation did not show increased accumulation of unspliced species. One example is shown in Fig.2A for the *RPL26A* gene, which exhibits frequent readthrough transcripts onto the neighbor gene *YLR345W*,

none of which were unspliced. We calculated the splicing efficiency of RNA with 3′-extensions in the Ysh1-AA or Rna15-AA or control strain genome-wide, based on the percentage of unspliced reads, and compared this metric to the splicing efficiency of RNAs with no extensions. RNAs with 3′-extensions found in the Ysh1-AA or Rna15-AA datasets did not show a large increase of unspliced reads compared to unextended transcripts found in the control wild-type sample or unextended transcripts detected in the Ysh1-AA or Rna15-AA samples (Fig.2B). Some exceptions include *RPL28*, for which intron retention was twice as abundant in 3′-extended transcripts compared to unextended transcripts in the Ysh1-AA or Rna15-AA (Fig.2C; Fig.S3). Overall, these results argue that loss of recognition of the primary CPA or termination site does not impact the efficiency of splicing of upstream introns genome-wide.

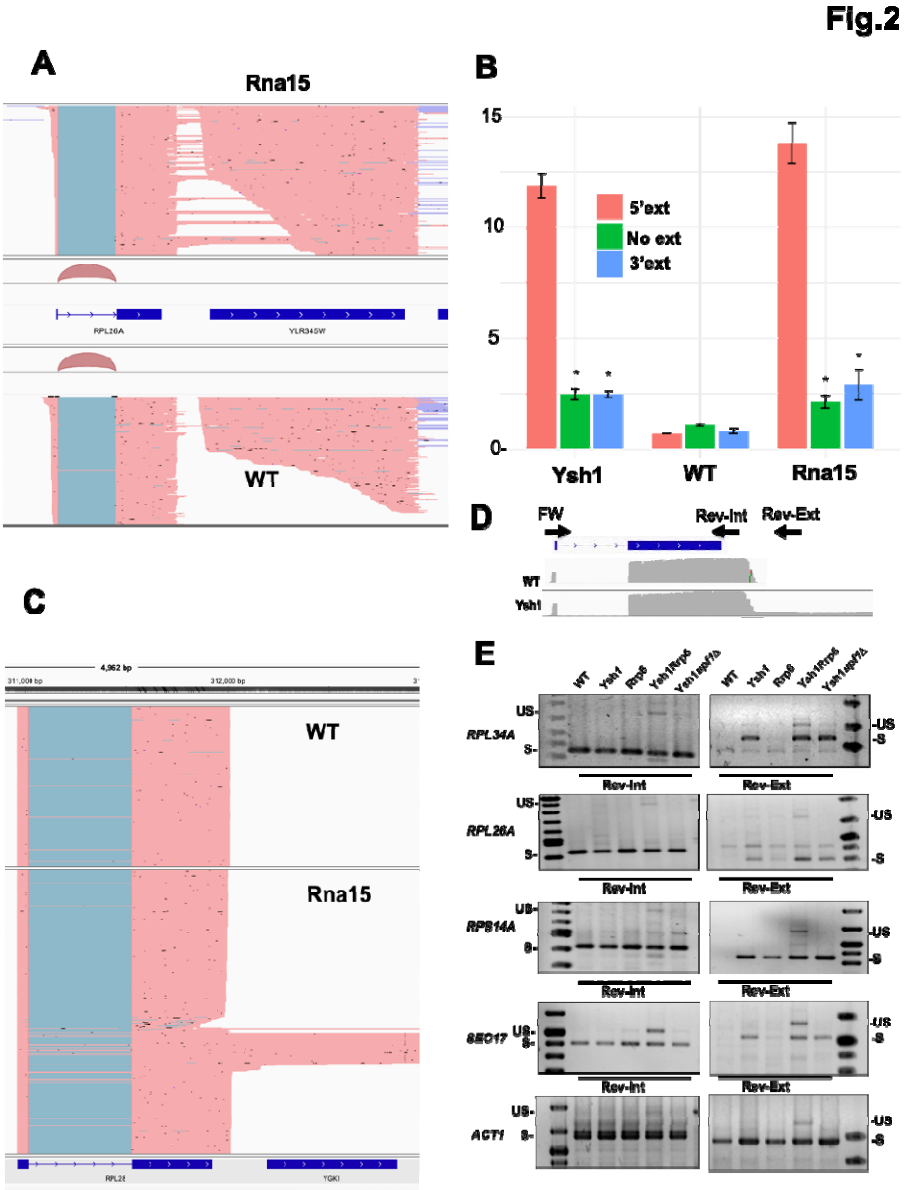


Figure 2. Impact of CPA inactivation on splicing efficiency.

A. Selection of ONT reads from WT or Rna15AA in the *MRPL24-RPL26A-YLR345W* region showing the accumulation of 3′ extended reads of *RPL26A* in the Rna15AA strain. Reads are colored in salmon; splicing event connecting reads are colored in blue. Sashimi plots showing the splicing of *RPL26A* are also included.

B. Bar graph showing the percentage of unspliced reads genome-wide as a function of 5′ extension size in the Ysh1-AA, WT or Rna15-AA strains for RNAs with no extensions. The p-values obtained for the comparisons are: Rna15vs WT: No Ext: 0.005848); 3' extended: 0.003757; Ysh1 vs WT: No Ext: 0.007634; 3' extended: 0.01342.

C. Selection of ONT reads from WT or Rna15-AA showing the accumulation of unspliced and 3′ extended reads of *RPL28* in the Rna15AA strain.

D. RT-PCR scheme used to analyze the splicing status of normal and 3′-extended mRNAs. Shown is the Oxford Nanopore sequencing read coverage obtained for the WT and Ysh1-AA samples for one example gene (*RPS14A*) gene and approximate location of the primers used for the RT-PCR analysis. The Rev-Ext primer only anneal to 3′-extended forms detected upon Ysh1 or Rna15 inactivation. Similar schemes are used for the other genes.
E. RT-PCR analysis of splicing of normal and 3′-extended mRNAs. panels show RT-PCR products obtained for each of the genes indicated in the indicated strains. S=spliced; US = unspliced.

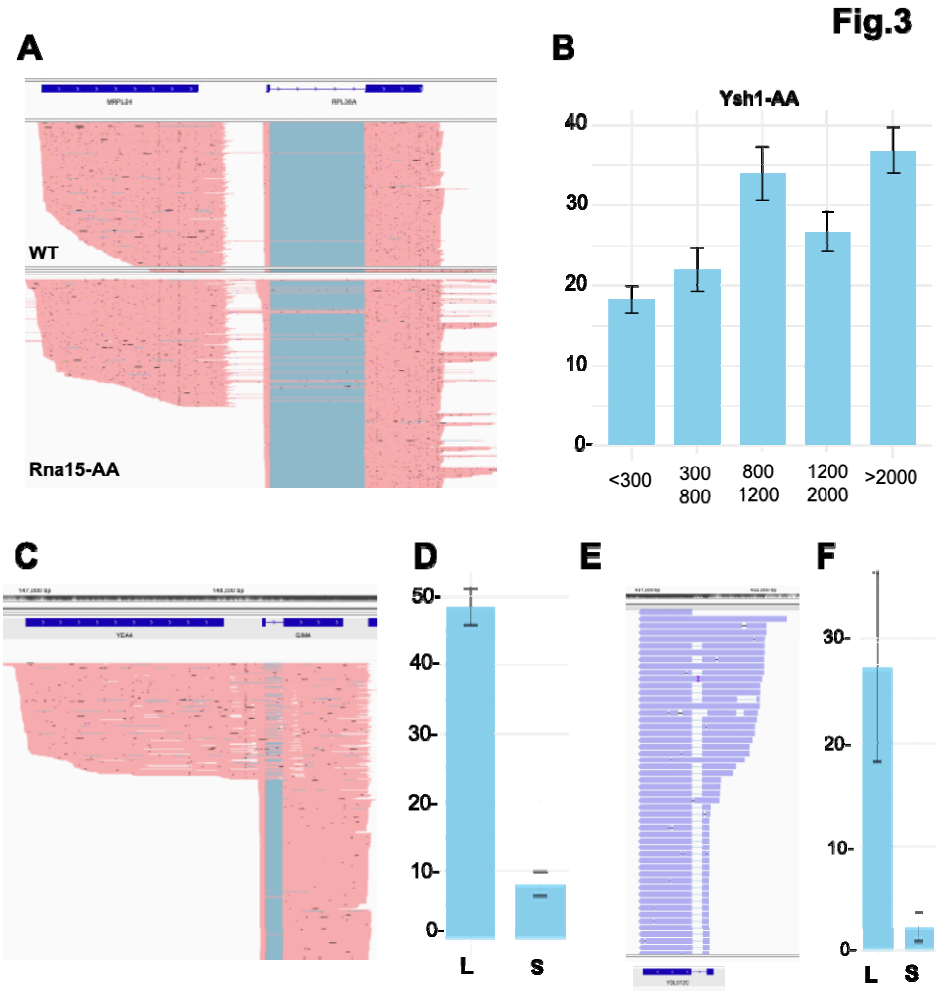
The ONT RNA sequencing platform relies on the presence of poly(A) tails for sequencing. For this reason, mRNAs that are 3′ extended but unpolyadenylated were undetectable in the sequencing experiments described above. Therefore, we could not rule out that unpolyadenylated, 3′-extended readthrough transcripts might be predominantly unspliced when the CPA machinery is inactivated. In addition, we could not rule out that RNA degradation pathway such as the nuclear exosome, or nonsense-mediated decay might degrade unspliced transcripts, limiting their detection by Nanopore sequencing. To investigate the splicing status of unpolyadenylated, 3′-extended readthrough transcripts, we analyzed selected transcripts that did not show prevalent splicing defects in 3′-extended Nanopore reads by RT-PCR using total RNAs, which included unpolyadenylated RNAs. We used two different reverse primers: an internal reverse primer (Rev-Int) detects normal mRNAs, as well as 3′-extended ones; or a Rev-Ext primer that anneals only to 3′-extended RNAs and allowed the analysis of splicing for 3′extended forms (Fig.2D). Concomittant inactivation of the nuclear exosome component Rp6p by anchor away, or deletion of the gene encoding the NMD factor Upf1p was used to assess if unprocessed transcripts are subject to nuclear or cytoplasmic surveillance.
As show in Fig.2E, RT-PCR analysis using reverse primers that anneal only to extended RNAs showed that 3′-extended transcripts that were fully spliced were readily detected but absent in WT (or present at much lower level for ACT1). This analysis did not reveal any major accumulation of unspliced 3′-extended transcripts in the Ysh1-AA strain. For all genes analyzed, a minor accumulation of unspliced 3′-extended transcripts could be detected upon inactivation of both Ysh1 and Rrp6,. This result shows that these 3′-extended and unspliced RNAs are efficiently degraded by the nuclear exosome. By contrast, inactivation of NMD did not result in an increase in unspliced RNAs detection, which showed that these transcripts are likely retained in the nucleus, perhaps due their lack of poly(A) tails. A similar minor accumulation of unspliced transcripts was observed with the Rev-Int primer in the Ysh1/Rrp6 double anchor away strain. We conclude that a fraction of unprocessed, 3′-extended RNAs fail to undergo splicing but they do not accumulate to detectable levels unless the nuclear exosome is inactivated. We note that concurrent inactivation of CPA factors and of Rrp6p has been shown to induce relocation of mRNAs to discrete nuclear compartments[22], which may complicate the interpretation of these data.

Upstream termination defects result in length-dependent splicing defects for downstream intron-containing genes

Inspection of the genes showing significant splicing defects in the Ysh1-AA or Rna15-AA samples revealed that unspliced reads were mainly found for RNAs with 5′-extensions resulting from readthrough transcription from upstream genes. These 5′-extensions are due to deficient termination from upstream transcription units, resulting in transcripts containing the upstream mRNAs, the intergenic region and the downstream ICG mRNA. One example is shown on Fig.3A where deficient termination of *MRPL24* transcription generated readthrough transcripts

into the downstream ICG, leading to accumulation of unspliced *RPL36A* reads (Fig.3A). This effect was detected genome-wide, as shown by a large increase in the fraction of unspliced reads for reads containing 5′-extensions found in the Rna15-AA or Ysh1-AA strain relative to unextended transcripts found in any of the strains (Fig.2B). This effect is reminiscent to what was observed upon nuclear depletion of the nuclear poly(A) binding protein Nab2[23], which results in readthrough transcripts that are frequently unspliced.

Figure 3. Impact of upstream readthrough transcripts and 5′ extensions on splicing efficiency.



A. Selection of ONT reads from WT or Rna15-AA in the *MRPL24-RP36A* region showing the accumulation of unspliced 5′ extended reads of *RPL36A* in the Rna15-AA strain.

B. Bargraph showing the percentage of unspliced reads genome-wide as a function of 5′ extension size in the Ysh1-AA strain. The p-values obtained for the comparisons are: <300 vs 800 -1200: 0.01486; <300 vs 1200-2000: 0.03149; <300 vs >2000: 0.009079.

C. Selected Nanopore reads from WT control cells in the *YEA4-GIM4* region. Reads at the top of the picture are readthroughs transcripts from *YEA4* gene into *GIM4* which show a high fraction of unspliced reads.

D. Quantification of splicing defects for normal *GIM4* mRNAs (S), or readthrough mRNAs from the *YEA4* gene (L); p-value L vs S: 0.0003384. Numbers indicate the percentage of unspliced reads..

E. Selection of Nanopore reads from WT control cells in the *YDL012C* region.

F. Quantification of splicing defects for *YDL012C* mRNAs with short 5′exons (S), or with 5′ exons longer than the annotated 5′ exon (L); p-value L vs S: 0.01879. Numbers indicate the percentage of unspliced reads.

The results described above suggest that an increased distance from the 5′-end of the transcript to the intron is detrimental to splicing. To provide support for this hypothesis, we compared the length of 5′ extensions found in transcripts showing significant splicing defect or no splicing defects. In the Ysh1-AA strain, the median 5′ extension size for ICGs which showed a significant

splicing defect was 1320 nt, while the median 5′ extension size for ICGs which did not show a significant splicing defect was only 768 nt. In the Rna15-AA, the corresponding numbers were very similar, at 1426 nt and 833 nt, respectively. Binning reads obtained from the Ysh1-AA strain according to 5′ extension lengths showed a threshold pattern, where transcripts with 5′ extensions smaller than 800 nt showed a low level of intron retention, while transcripts with longer 5′ extensions showed significantly stronger splicing defects (Fig.3D). A similar effect was observed for the Rna15-AA strain (Fig.S4). Overall, these data demonstrate that long 5′ extensions that are caused by readthrough transcription upon CPA inactivation decrease splicing efficiency of downstream ICGs.

mRNAs with naturally long 5′-extensions exhibit low splicing efficiency.

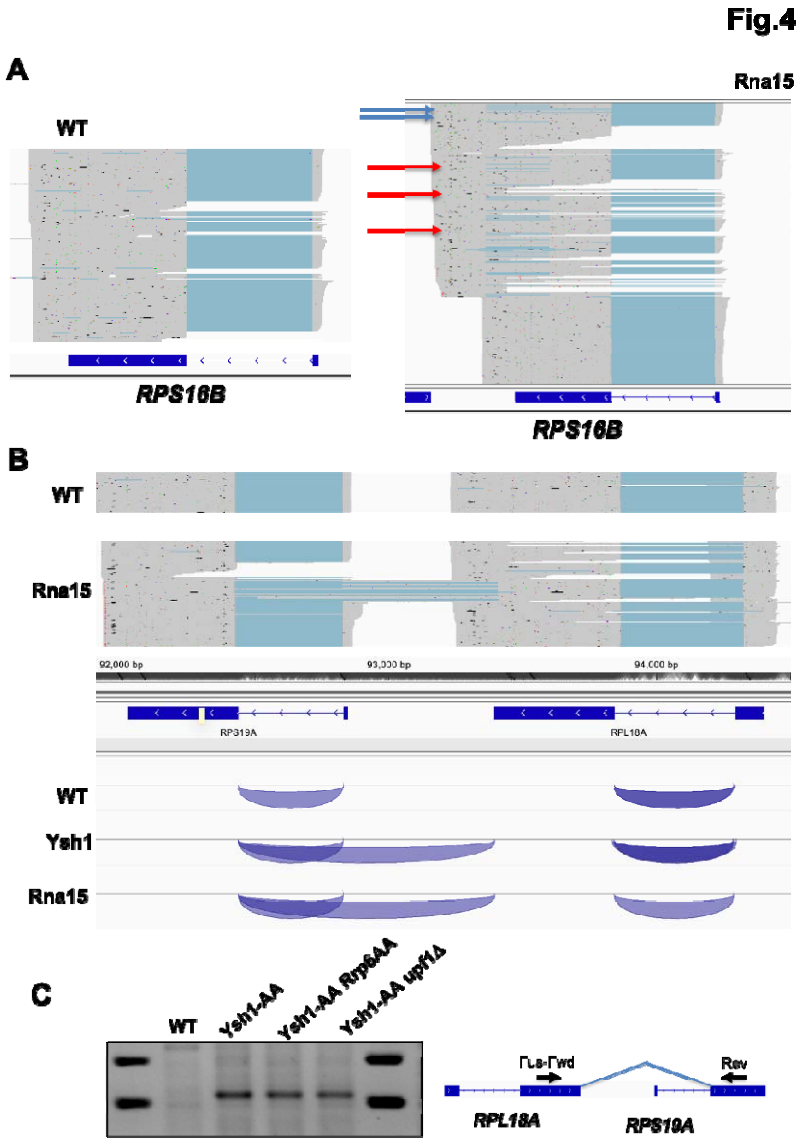
The previous data showed that loss of termination due to nuclear depletion of CPA factors led to inefficiently spliced transcripts with 5′-extensions. To determine if this effect was solely due to the length of the 5′-exons/extensions or to indirect effects from CPA factors inactivation, we analyzed ONT data from WT cells to search for ICGs exhibiting reads with heterogenous 5′-ends and analyzed their splicing efficiency as a function of 5′-exon length. The *GIM4* ICG is located just downstream of the *YEA4* gene, and we detected a high number of transcripts originating from *YEA4* and reading through *GIM4* in WT samples (Fig.3C). Strikingly, these readthrough transcripts showed a much lower splicing efficiency than the mRNAs originating from the bona fide *GIM4* promoter (Fig.3C,3D) as the level of unspliced reads for readthrough mRNA reached close to 50% (Fig.3D). For *YDL012C*, we detected some transcripts with heterogeneity in 5′-exon length, possibly due to usage of alternative transcription start sites (Fig.3E). As described above for *GIM4*, *YDL012C* mRNAs with naturally long 5′-exons were less efficiently spliced than those with short 5′-exons (Fig.3F), with an increase of ~5-fold in the amount of unspliced reads. We note that the numbers shown on this figure might be an underrepresentation of the actual splicing defects for these genes, as unspliced mRNAs are known to be targeted by various RNA surveillance systems[21]. Overall, these data demonstrate that long 5′-exons interfere with splicing efficiency. The most direct explanation for this effect is that long 5′-exon length increase the distance from the 5′-cap to the intronic sequences, decreasing the activation of splicing by the cap-binding complex [5].

Deficient transcription termination reveals dormant splice sites and promotes new splicing events

We next analyzed the impact of inactivating Ysh1 or Rna15 on splicing fidelity, by searching for novel splicing events induced by inactivation of these factors. We hypothesized that RNA extensions caused by loss of termination upon CPA inactivation would expose sequences matching splice sites, which are otherwise not used in the context of normal termination. Inspection of the Ysh1-AA and Rna15-AA strain datasets identified several such cases. The 5′ splice site of *RPS16B* was found to be spliced to a new 3′-splice site found within the *RPS16B* second exon (blue arrows, Fig 4A), in the Ysh1-AA or Rna15-AA strains, but not in WT. In addition, several reads showed frequent usage of a novel 5′ splice in the middle of *RPS16B* exon2, spliced to the aforementioned 3′ SS (red arrows, Fig.4A; Fig.S5). It is possible that the 3′SS found used in the Rna15 or Ysh1-AA strains is naturally used in a wild-type context. However, this would result in a very short transcript considering the length of the exons used, which is likely to be undetectable by Nanopore sequencing given the size limit for detection. In another case, we detected frequent usage of a novel 5′SS found at the end of exon2 of *RPL18A* spliced to the 3′SS of the following *RPS19A* gene upon Ysh1 or Rna15 inactivation (Fig.4B).

This novel 5′SS showed a GUAAGU sequence which overlaps with the UAA stop codon of *RPL18A*.(Fig.S6). This splicing event resulted in the production of chimeric *RPL18A-RPS19A* transcripts, which corresponds to about 10% of the splicing events for extended species of *RPL18A* in Rna15 or Ysh1-AA. (Fig.4B).

Figure 4. Activation of novel Splice sites promoted by deficient transcription termination.
A. Selection of Nanopore sequencing reads for RPS16B in the WT control or Rna15AA strain. Blue arrows point to reads using the alternative 3′SS detected in the 3′UTR of *RPS16B*. Red arrows point to the isoform using the alternative 5′SS and 3′SS.
B. Selection of Nanopore sequencing reads for the RPL18A and RPS19A in the WT control or Rna15AA strain and Sashimi plots for WT, Rna15AA, and Ysh1AA showing the intergenic *RPL18A-RPS19A* splicing using the new 5′SS found at the end of *RPL18A*.
C. RT-PCR detection of the intergenic *RPL18A-RPS19A* splicing. RNAs extracted from WT control or Ysh1AA ,Ysh1Rrp6AA, and Ysh1upf1 were analyzed by RT-PCR using the forward (Fwd) and reverse (Rev) primers indicated.



We validated the accumulation of this spliced product by RT-PCR in the Ysh1-AA (Fig.4C). Interestingly, this spliced isoform did not appear to be targeted by the nuclear exosome (Ysh1-Rrr6AA, Fig.4C), nor by nonsense-mediated decay (NMD) (Ysh1AA-*upf1*Δ, Fig.4C). The sequence of this chimeric mRNA would encode the Rpl18 protein in its entirety, fused to the protein sequence encoded by the second exon of *RPS19A*. The absence of premature stop codon explains the lack of effect of inactivating NMD. In the case of the duplicated *RPL18B* gene paralogue, the corresponding sequence is GUAAUC and we did not detect splicing between *RPL18B* and *RPS19B* at that site. The mismatches at positions 5 and 6 compared to canonical 5′SS sequences explains why such intergenic splicing events were not detected for the paralogue genes.

Deficient transcription termination promotes long-distance intergenic splicing events, including splicing-dormant genes.

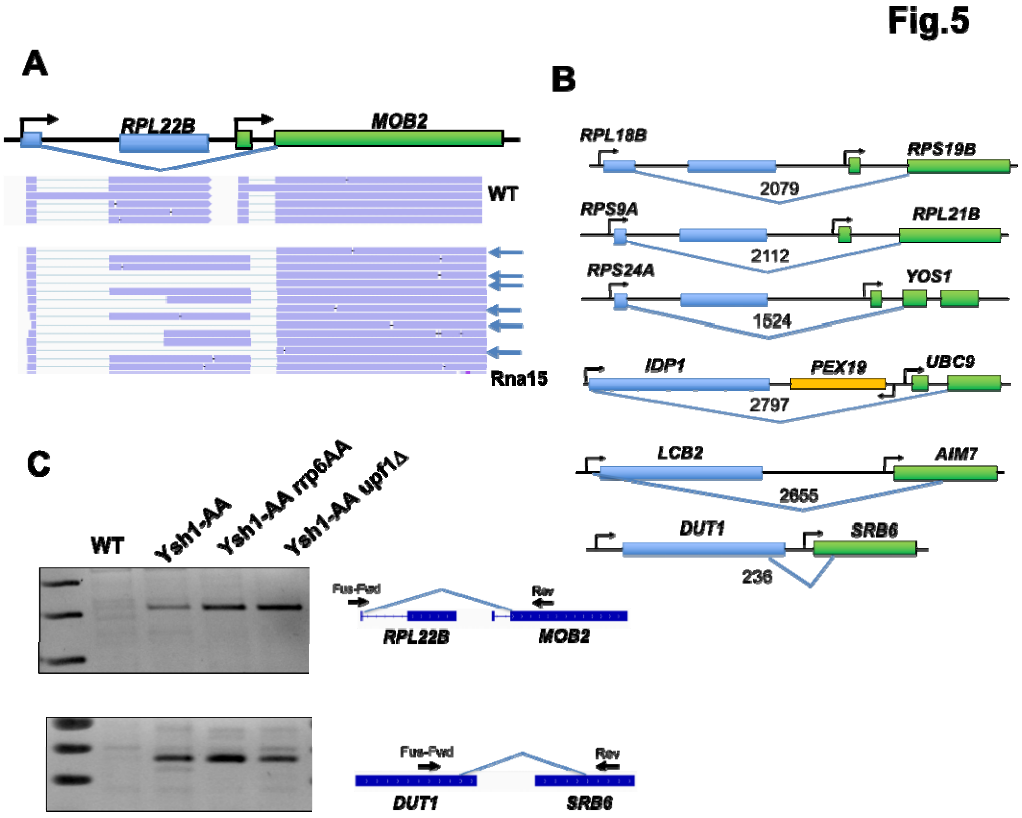
The previous example showed that loss of termination could promote the production of chimeric *RPL18A-RPS19A* transcripts. In *S.cerevisiae*, several intron containing genes (ICG) are located in tandem on the same strand (Table S2). Deficient termination due to inactivation of Ysh1 or Rna15 may therefore result in the production of transcripts covering multiple introns containing several splicing signals competing for usage, providing the opportunity for splicing events between tandem genes. We searched for such intergenic splicing events by identifying reads that showed splicing junctions between sequences belonging to different genomic features. After curating these junctions for overlapping transcription units, and for read alignment artefacts, we identified several cases of intergenic splicing triggered by loss of termination in the Ysh1-AA and/or Rna15-AA strains. The most frequent event detected was the splicing of the 5′ SS of *RPL22B* to the 3′ SS of the downstream *MOB2* gene (Fig.5A).

Figure 5. Intergenic splicing events detected upon inactivation of CPA factors.

A. *RPL22B-MOB2* genomic region and examples of Oxford Nanopore reads obtained from WT or RNA15-anchor away strains. Reads highlighted with blue arrows indicate intergenic *RPL22B-MOB2* splicing events.

B. Other examples of intergenic splicing events detected in the Ysh1-AA and/or Rna15-AA strains. Numbers indicate the distance between the 5′ and 3′ SS used.

C. RT-PCR analysis of intergenic splicing in WT, Ysh1-AA and derivative strains. The approximate locations of the forward (Fwd) and Reverse (Rev) primers used is indicated.



It was shown previously that an alternative intronic 5′ SS is present in *RPL22B* [24]. This alternative 5′ SS was also found to be spliced to the downstream 3′ SS of *MOB2*. Reads showing splicing of exon1 of one intron-containing gene to the exon2 of a downstream ICG were also found for several other gene pairs: *RPL18B-RPS19B*, *RPS9A-RPL21B* and *RPS24A-YOS1* (Fig.5B). These events were found to occur upon inactivation of either Rna14 or Ysh1, showing that deficient transcription termination can promote intergenic splicing.

In addition to these events where natural splice sites were used, we also discovered cases where new splice sites are revealed. In the case of *IDP1*, a cryptic 5′ SS is present at the beginning of the ORF, which becomes spliced to the 3′ SS of the downstream UBC9 gene (Fig.5B; Fig.S7). Strikingly, in the case of the *LCB2-AIM7* or *DUT1-SRB6* gene pairs, none of these genes are naturally spliced. However, Ysh1 or Rna15 inactivation generates readthrough species which results in the intergenic splicing between a 5′ SS present in the 5′UTR or ORF of the upstream gene (*LCB2* or *DUT1*) and a 3′ SS in the ORF of the downstream gene (*AIM7* or *SRB6*, respectively; Fig.5B; Fig.S8). We used RT-PCR to validate two of these splicing events (Fig.5C), and test whether these intergenic spliced species are subject to nuclear surveillance or degradation by NMD. In agreement with the Nanopore sequencing data, intergenic splicing between *RPL22B* and *MOB2* or between *DUT1* and *SRB6* was detected only upon inactivation of Ysh1, but not in the WT control (Fig.5C). There was no major effect of inactivating Rrp6 or the NMD factor Upf1 on the accumulation of these intergenic spliced products, suggesting that these species are not a major target of these RNA degradation systems.

Discussion

In this work we have studied the impact of rapid inactivation of splicing or 3′-end processing factors on splicing and 3′-end formation using long read ONT sequencing. The use of this platform allowed us to directly investigate the splicing efficiency and 3′-end of transcripts genome-wide. Strikingly, we did not find evidence for global reciprocal coupling between splicing and usage of downstream 3′-end processing signals. One of the limitations of this conclusion is that ONT sequencing relies on polyadenylated RNAs, and RNAs showing 3′-extensions in our sequencing data are still polyadenylated albeit at more distant sites. RT-PCR analysis of selected transcripts showed that a minor population of unspliced and 3′- extended transcripts can be detected in total RNAs (which include both polyadenylated and unpolyadenylated RNAs) if the nuclear exosome component Rrp6 is inactivated concomitantly with CPA factors (Fig.2E). While this may suggest some level of coupling of splicing to 3′-end processing, we note that the majority of these 3′-extended transcripts are still efficiently spliced (Fig.2E). In addition, simultaneous inactivation of CPA factors and of Rrp6p has been shown to induce relocation of mRNAs to discrete nuclear compartments[22], which may complicate the interpretation of data involving these double mutants. The involvement of the RNA exosome in degrading unprocessed RNAs is in agreement with previous studies showing that readthrough transcripts generated upon genetic inactivation of some CPA factors can be degraded by the exosome[25]. Overall our genome-wide analysis of the Rna15-AA or Ysh1-AA showed a global increase in 3′-UTR length (Fig.1); therefore we think that our genome-wide analysis still revealed the full extent processing defects in these strains, albeit at possibly lower levels than if all RNA surveillance pathways had been inactivated. Based on these experiments we conclude that there is no global coupling of splicing to 3′-end formation for most *S.cerevisiae* ICGs, although coupling might exist for a few transcripts such as *RPL28* (Fig.2). One reason for the loss of this coupling mechanism might be due to the fact that most transcription units in *S.cerevisiae* do not contain intronic sequences, unlike in mammalian genomes. Therefore the evolutionary pressure to maintain functional coupling between these two processes may have been lost if the signals that dictate the efficiency of each process are strong enough.

One of the major conclusions from our work is that loss of termination promoted by inactivation of CPA factors result in intron-containing transcripts with long 5′-extensions, which strongly

decreased splicing efficiency. Strikingly, the size of the 5′-extensions negatively correlates with splicing efficiency. A negative impact on splicing for readthrough transcripts was reported previously upon nuclear depletion of the nuclear poly(A) binding protein Nab2[23]. While Nab2 is not a *bona fide* component of the CPA machinery, it is possible that the global impact of nuclear depletion of Nab2 on nuclear RNA decay[26] and/or on mRNA export [27] indirectly impairs transcription termination and/or CPA. Interestingly, we found that the deleterious impact of long 5′extensions can be detected not only upon CPA inactivation, but also in a normal cells for transcripts exhibiting readthrough transcription from an upstream gene such as *GIM4* (Fig.3), or for mRNAs with heterogeneous transcription start sites such as *YDL012C*. These results shows that the splicing defects induced by long extensions are not intrinsically due to CPA deficiency but suggest that an increase distance of the splicing signals to the 5′-cap structure is responsible for the impairment of splicing, as it would decrease the positive effect of the cap-binding complex on splicing. Alternatively, we cannot rule out that long 5′-extensions might promote secondary structures that impair the recognition of downstream splicing signals. However, given the general widespread effect of 5′ extension length on splicing, we believe this hypothesis to be unlikely. Overall these data underscore the importance of proper termination to limit readthrough transcription onto downstream genes, which would largely inhibit splicing.

One of the most striking effect of inactivating CPA is the production of intergenic transcripts with multiple splicing signals, resulting in the detection of novel splicing events (Fig.4, 5). These results provide further evidence that inhibition of CPA does not impair splicing, as the extended species that are produced upon CPA inactivation are now spliced with novel patterns. Intergenic splicing events can be highly abundant in the case of *RPL22B-MOB2* or more rare, as they are sometimes detected with only a few reads in the Ysh1-AA or Rna15-AA samples. These spliced species might actually be underrepresented in our ONT sequencing datasets for two major reasons. First some of these species might be targeted by RNA surveillance pathways, even if we did not detect a major effect upon inactivating the nuclear exosome or NMD (Fig.4 and 5). For instance, exon skipped products (which functionally correspond to the intergenic spliced products we describe here) have been shown to be targeted by the exosome and Rat1 in *S.cerevisiae* [28]. In addition, some 5′-exons are very short, and while Nanopore sequencing offers good coverage of mRNAs, some 3′-end bias can still be observed in yeast[29], which results in underrepresentation of transcript with short 5′-exons. We note that the distance involving some of these splicing events exceeds the size of the largest natural intron in *S.cerevisiae*, including several detected splicing events involving splice sites that separated than more than 2kb, including 2.6 and 2.8kb (Fig.5B). These results indicate that the yeast splicing machinery can operate at longer distances than previously thought. On the other hand, readthrough between closely positioned transcription units can promote short distance splicing events such as the one detected between *DUT1* and *SRB6*. It is possible that some of these intergenic splicing events may occur in natural conditions and produce functional products if natural levels of transcriptional readthrough occurs, such as in stress conditions as observed in mammalian cells[30]. The compactness of the yeast genome and the close spatial positioning of transcription units make this more likely, and could potentially offer a way to produce proteins resulting from splicing of adjacent genes.

Acknowledgements

We thank Douglas Black and members of the GFC lab for discussion and Alina Tong for comments on the manuscript. KH was supported by a Dissertation Year Fellowship from the UCLA Division of Graduate Education. Supported by NIGMS grant GM130370 to GFC.

Materials and Methods

Yeast Strains and Growth.

All yeast growth was performed in YPD media (1% w/v yeast extract, 2% w/v peptone, and 2% w/v dextrose). In all experiments cells were back diluted from an overnight solution to an O.D.600nm between 0.1 and 0.13 in 50mL per time-point and grown for approximately two doublings at 30°C, shaken at 200rpm. 50mL of cell culture would be harvested via centrifugation and flash frozen in liquid nitrogen for downstream use. Anchor away experiments were performed by replacing cell culture media of log-phase yeast with YPD+ 1µg/mL of rapamycin as described originally in [16]. Cells were then grown for an additional 2 hours in YPD+rapamycin before being harvested by centrifugation and flash frozen for downstream use. Strains used in this work can be found in Table S3.

RNA Extraction and Direct RNA sequencing.

Total RNAs were extracted in triplicate via the standard phenol-chloroform technique [31] from anchor-away yeast strains. RNAs for the wild-type, ysh1-AA, and rna15-AA samples were then further processed according to the protocol described in the manufacturer′s manual for ONT product (SQK-RNA004) for sequencing on the PromethION system (PRO-SEQ002) using RNA Flow Cells (FLO-PRO004RA). The only exception was the use of Maxima-H Minus as the RT enzyme as this was the recommended enzyme at the time of these experiments. RNAs for the snp1-AA samples were further processed per the manufacturer′s manual for ONT product (SQK-RNA004) for sequencing on the MinION system (MIN-101B) using RNA Flow Cells (FLO-MIN004RA) using the current Induro RT enzyme protocol. RNA Flow Cells were washed according to the manufacturer′s protocol for the Flow Cell Wash Kit (EXP-WSH00) in between sample runs.

RT-PCR Analysis

RNAs for RT-PCR were generated from yeast cells as described above. RNA samples were then DNase I treated according to the manufacturer′s instructions (Thermo Fisher EN0521). For cDNA synthesis from total RNAs the standard Maxima H Minus RT protocol (Thermo Fisher EP0751) was followed with gene specific primers and 100 units of enzyme per reaction. 1uL of each RT reaction was then used as template for a 20uL HiFi-PCR reaction using the corresponding gene specific primers (Supplemental Table S1). 5uL PCR products were then mixed with a 6x Bromophenol Blue and Xylene Cyanol loading dye and analyzed by agarose gel using 1.2-1.8% w/v agarose in 1xTAE (40mM Tris Base, 20mM acetic acid, 1mM EDTA, pH 8.2-8.4) depending on expected product size. Gels were then visualized by Ethidium Bromide staining on an Azure 300 Imaging System.

Bioinformatics and Statistical Analysis

Reads were basecalled in real-time using the default algorithm for direct RNA sequencing on MinKNOW (version 6.2.6) with Dorado (version 7.6.7). Reads were then mapped to the

Saccharomyces cerevisiae genome (S288C_reference_sequence_R64-5-1) acquired from SGD using Minimap 2 (Version 2.17-r941). The following parameters were used during alignment "minimap2 -ax splice -G 3000" which set a maximum intron length of 3000 bp. Resulting bam files were coordinate sorted and converted to bed files using SAMtools[32] and Bedtools (https://bedtools.readthedocs.io/en/latest/) for downstream analyses. Reads were visualized using IGV (Version 2.12.3).

Reads were annotated with overlapping genes using bedtools intersect (v2.30.0). For read counts, we calculated the number of reads mapping to each gene using featureCounts as part of the Rsubread package (version 2.12.3) allowing for multiple overlaps. DESeq2 [33] was used to quantify changes in gene expression and for normalizing counts to library size.

Counts were generated separately for whole genes and introns specifically. For a read to be counted as unspliced it needed to overlap an annotated intron by at least 6 bases. Ratios of unspliced were generated by dividing intron counts by the gene counts. For extension analysis, reads were classified into 4 groups. "3 prime extended" are reads which overlap the gene of interest but terminate more than 150 bp downstream of the annotated PAS. "5 prime extended" overlap the gene of interest, but begin more than 100 bp upstream of the TSS. If a read met both of these conditions, it was assigned as an "intrusive transcript". All other reads fall into the "no extension" category. Experimentally derived UTR annotations[34] were used for the TSS and PAS locations.

All statistical tests and p-values reported were based on a welch's t-test. For the splicing volcano plots the p-value was adjusted further using the Benjamini-Hochberg method to control for False Discovery Rate.

Data availability

Custom Nanopore processing code is available on GitHub: https://github.com/kevinh97/Yeast-Transcription-termination. The long-read RNA sequencing data reported in this paper are accessible in the NIH SRA database under the accession number PRJNA1229592.

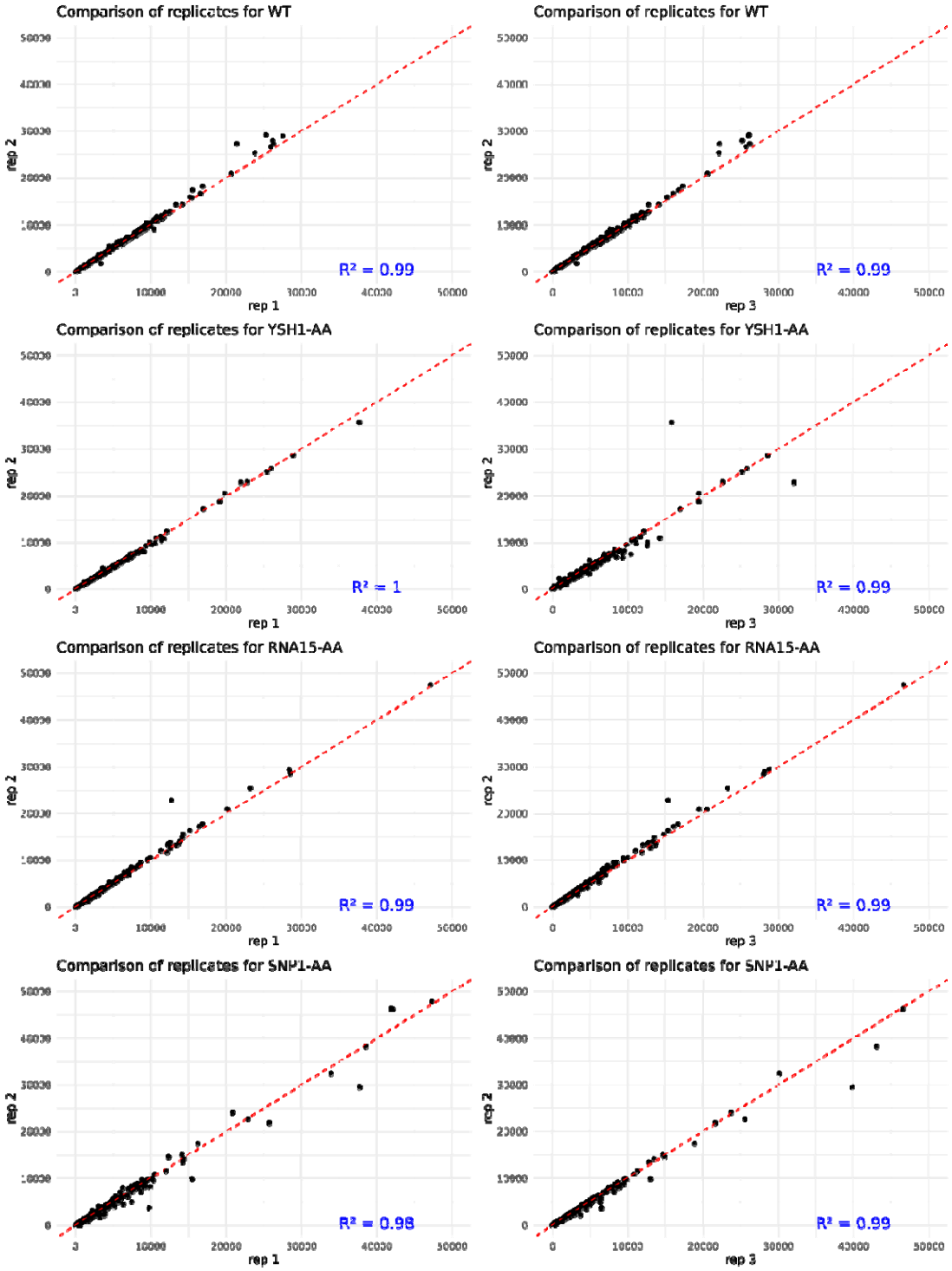
References

1. Shenasa, H. & Bentley, D. L. Pre-mRNA splicing and its cotranscriptional connections. *Trends in Genetics* **39**, 672–685 (2023).
2. Colot, H. V., Stutz, F. & Rosbash, M. The yeast splicing factor Mud13p is a commitment complex component and corresponds to CBP20, the small subunit of the nuclear cap-binding complex. *Genes Dev* **10**, 1699–1708 (1996).
3. Lewis, J. D., Izaurralde, E., Jarmolowski, A., McGuigan, C. & Mattaj, L. W. A nuclear cap-binding complex facilitates association of U1 snRNP with the cap-proximal 5' splice site. *Genes Dev* **10**, 1683–1698 (1996).
4. Izaurralde, E. *et al.* A nuclear cap binding protein complex involved in pre-mRNA splicing. *Cell* **78**, 657–668 (1994).
5. Rambout, X. & Maquat, L. E. The nuclear cap-binding complex as choreographer of gene transcription and pre-mRNA processing. *Genes Dev* **34**, 1113–1127 (2020).
6. Görnemann, J., Kotovic, K. M., Hujer, K. & Neugebauer, K. M. Cotranscriptional Spliceosome Assembly Occurs in a Stepwise Fashion and Requires the Cap Binding Complex. *Mol Cell* **19**, 53–63 (2005).
7. Martinson, H. G. An active role for splicing in 3′-end formation. *Wiley Interdiscip Rev RNA* **2**, 459–470 (2011).
8. Niwa, M. & Berget, S. M. Mutation of the AAUAAA polyadenylation signal depresses in vitro splicing of proximal but not distal introns. *Genes Dev* **5**, 2086–2095 (1991).
9. Niwa, M., Rose, S. D. & Berget, S. M. In vitro polyadenylation is stimulated by the presence of an upstream intron. *Genes Dev* **4**, 1552–1559 (1990).
10. Movassat, M. *et al.* Coupling between alternative polyadenylation and alternative splicing is limited to terminal introns. *RNA Biol* **13**, 646–655 (2016).
11. Herzel, L., Straube, K. & Neugebauer, K. M. Long-read sequencing of nascent RNA reveals coupling among RNA processing events. *Genome Res* **28**, 1008–1019 (2018).
12. Ares Jr., M., Grate, L. & Pauling, M. H. A handful of intron-containing genes produces the lion's share of yeast mRNA. *RNA* **5**, 1138–1139 (1999).
13. Noble, S. M. & Guthrie, C. Identification of novel genes required for yeast pre-mRNA splicing by means of cold-sensitive mutations. *Genetics* **143**, 67–80 (1996).
14. Chanfreau, G., Noble, S. M. & Guthrie, C. Essential yeast protein with unexpected similarity to subunits of mammalian cleavage and polyadenylation specificity factor (CPSF). *Science (1979)* **274**, 1511–1514 (1996).
15. Herzel, L., Ottoz, D. S. M., Alpert, T. & Neugebauer, K. M. Splicing and transcription touch base: co-transcriptional spliceosome assembly and function. *Nat Rev Mol Cell Biol* **18**, 637–650 (2017).
16. Haruki, H., Nishikawa, J. & Laemmli, U. K. The Anchor-Away Technique: Rapid, Conditional Establishment of Yeast Mutant Phenotypes. *Mol Cell* **31**, 925–932 (2008).
17. Liu, Y. *et al.* Splicing inactivation generates hybrid mRNA-snoRNA transcripts targeted by cytoplasmic RNA decay. *Proc Natl Acad Sci U S A* **119**, (2022).
18. Wang, C. *et al.* Stress-induced inhibition of mRNA export triggers RNase III-mediated decay of the BDF2 mRNA. *RNA* **27**, 1545–1556 (2021).
19. Han, Z. *et al.* DNA-directed termination of RNA polymerase II transcription. *Mol Cell* **83**, 3253-3267.e7 (2023).

20. Sayani, S., Janis, M., Lee, C. Y., Toesca, I. & Chanfreau, G. F. Widespread Impact of Nonsense-Mediated mRNA Decay on the Yeast Intronome. *Mol Cell* **31**, 360–370 (2008).
21. Sayani, S. & Chanfreau, G. F. Sequential RNA degradation pathways provide a fail-safe mechanism to limit the accumulation of unspliced transcripts in Saccharomyces cerevisiae. *RNA* **18**, 1563–72 (2012).
22. Carneiro, T. *et al.* Inactivation of cleavage factor I components Rna14p and Rna15p induces sequestration of small nucleolar ribonucleoproteins at discrete sites in the nucleus. *Mol Biol Cell* **19**, 1499–1508 (2008).
23. Alpert, T., Straube, K., Carrillo Oesterreich, F. & Neugebauer, K. M. Widespread Transcriptional Readthrough Caused by Nab2 Depletion Leads to Chimeric Transcripts with Retained Introns. *Cell Rep* **33**, (2020).
24. Gabunilas, J. & Chanfreau, G. Splicing-Mediated Autoregulation Modulates Rpl22p Expression in Saccharomyces cerevisiae. *PLoS Genet* **12**, e1005999 (2016).
25. Torchet, C. *et al.* Processing of 3'-extended read-through transcripts by the exosome can generate functional mRNAs. *Mol Cell* **9**, 1285–1296 (2002).
26. Schmid, M. *et al.* The Nuclear PolyA-Binding Protein Nab2p Is Essential for mRNA Production. *Cell Rep* **12**, 128–139 (2015).
27. Green, D. M. *et al.* Nab2p is required for poly(A) RNA export in Saccharomyces cerevisiae and is regulated by arginine methylation via Hmt1p. *J Biol Chem* **277**, 7752–7760 (2002).
28. Egecioglu, D. E., Kawashima, T. R. & Chanfreau, G. F. Quality control of MATa1 splicing and exon skipping by nuclear RNA degradation. *Nucleic Acids Res* **40**, 1787–1796 (2012).
29. DeMario, S., Xu, K., He, K. & Chanfreau, G. NanoBlot: A Simple Tool for Visualization of RNA Isoform Usage From Third Generation RNA-sequencing Data. *bioRxiv* 2022.10.26.513894 (2022) doi:10.1101/2022.10.26.513894.
30. Vilborg, A., Passarelli, M. C., Yario, T. A., Tycowski, K. T. & Steitz, J. A. Widespread Inducible Transcription Downstream of Human Genes. *Mol Cell* **59**, 449–461 (2015).
31. Wang, C. *et al.* Rrp6 Moonlights in an RNA Exosome-Independent Manner to Promote Cell Survival and Gene Expression during Stress. *Cell Rep* **31**, (2020).
32. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
33. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 1–21 (2014).
34. Nagalakshmi, U. *et al.* The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science (1979)* **320**, 1344–1349 (2008).

Supplementary Figures and Tables – Barr, He et al.

Figure S1. Reproducibility of Oxford Nanopore Long Read RNA Sequencing data.
Plots showing normalized gene counts for each replicate among the 4 different strains sequenced (WT, YSH1-AA, RNA15-AA, and SNP1-AA). Counts were generated using FeatureCounts as part of the Rsubread package, DESeq2 was used for normalizing counts to library size.



Figure

S2.

Figure S3. Unspliced reads quantifications for *RPL28* in WT, Rna15-AA or Ysh1-AA for reads with no extensions or 3'extensions.

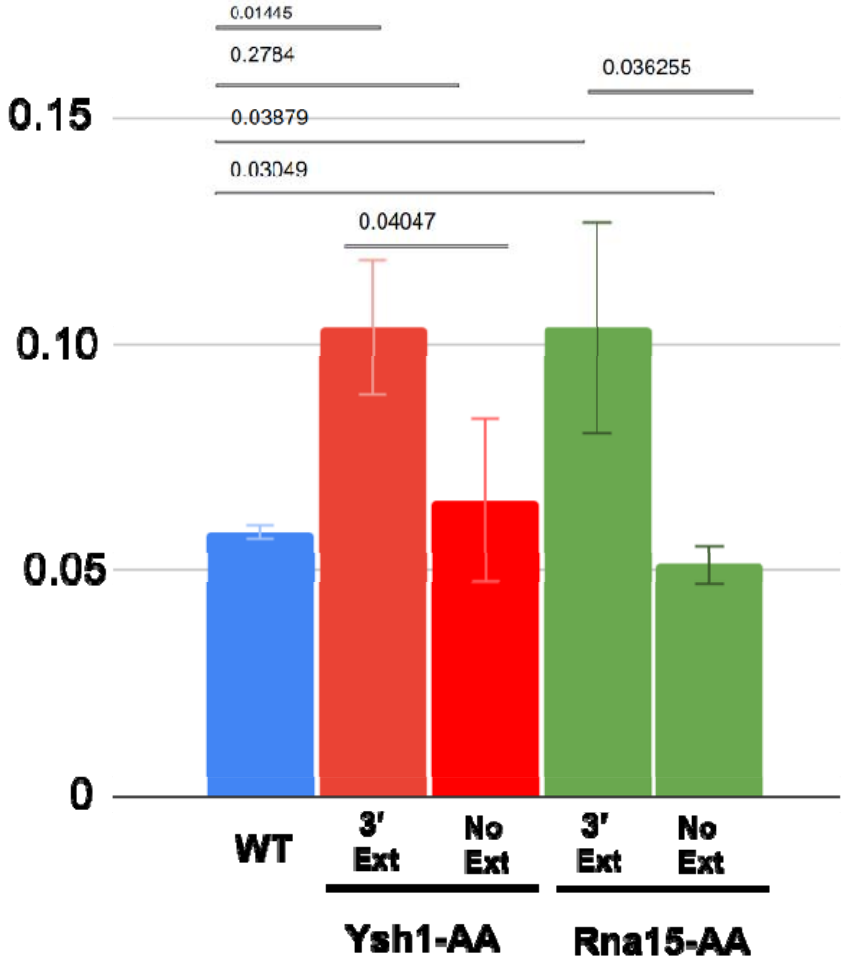


Fig.S4. Percentage of unspliced reads genome-wide as a function of 5' extension size in the Ysh1-AA and Rna15-AA strains.

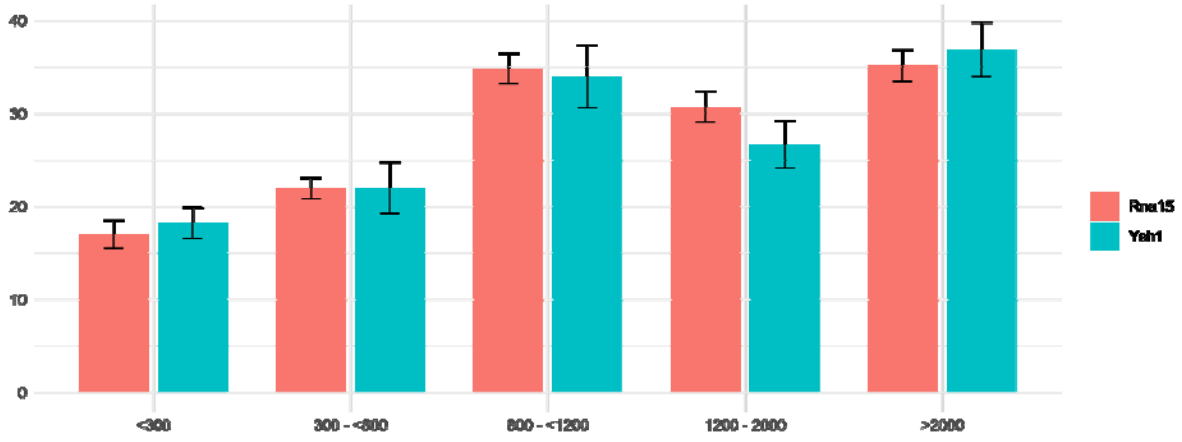


Figure S5.
A new intron is activated in *RPS16B* in the Ysh1-AA or Rna15AA strains.
Shown are sashimi plot of splicing events detected in the WT or Rna15AA strain as well as reads showing usage of the cryptic intron in the *RPS16B* 2nd exon. The 5′ SS and 3′ SS used are highlighted in red.

Figure S6. A 5′ SS used in the Ysh1-AA or Rna15-AA overlaps with the stop codon of *RPL18A*.

Shown are sequencing reads from the Rna15-AA strain highlighting usage of the GUAAGU 5′ SS highlighted in red overlapping with the UAA stop codon of RPL18A. Reads polarity are from right to left since the gene is localized on the C (-) strand.

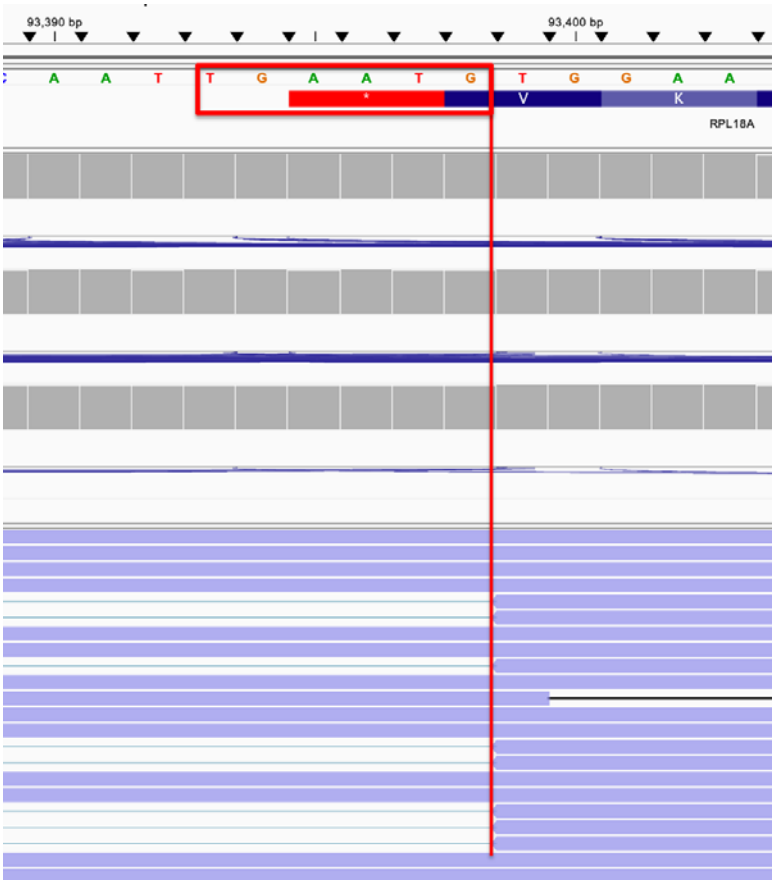


Figure S7.
Shown are reads produced by the *IDP1-UBC9* intergenic splicing events detected in the Rna15 or Ysh1AA strain. The cryptic 5′ SS sequence at the beginning of the *IDP1* ORF is highlighted in red.

Figure S8.
Shown are reads produced by the LCB2-AIM7 intergenic splicing events detected in the Ysh1AA strain. The cryptic 5′ SS sequence in the LCB2 5′UTR and the putative branchpoint and 3′ SS used in AIM7 are is highlighted in red.

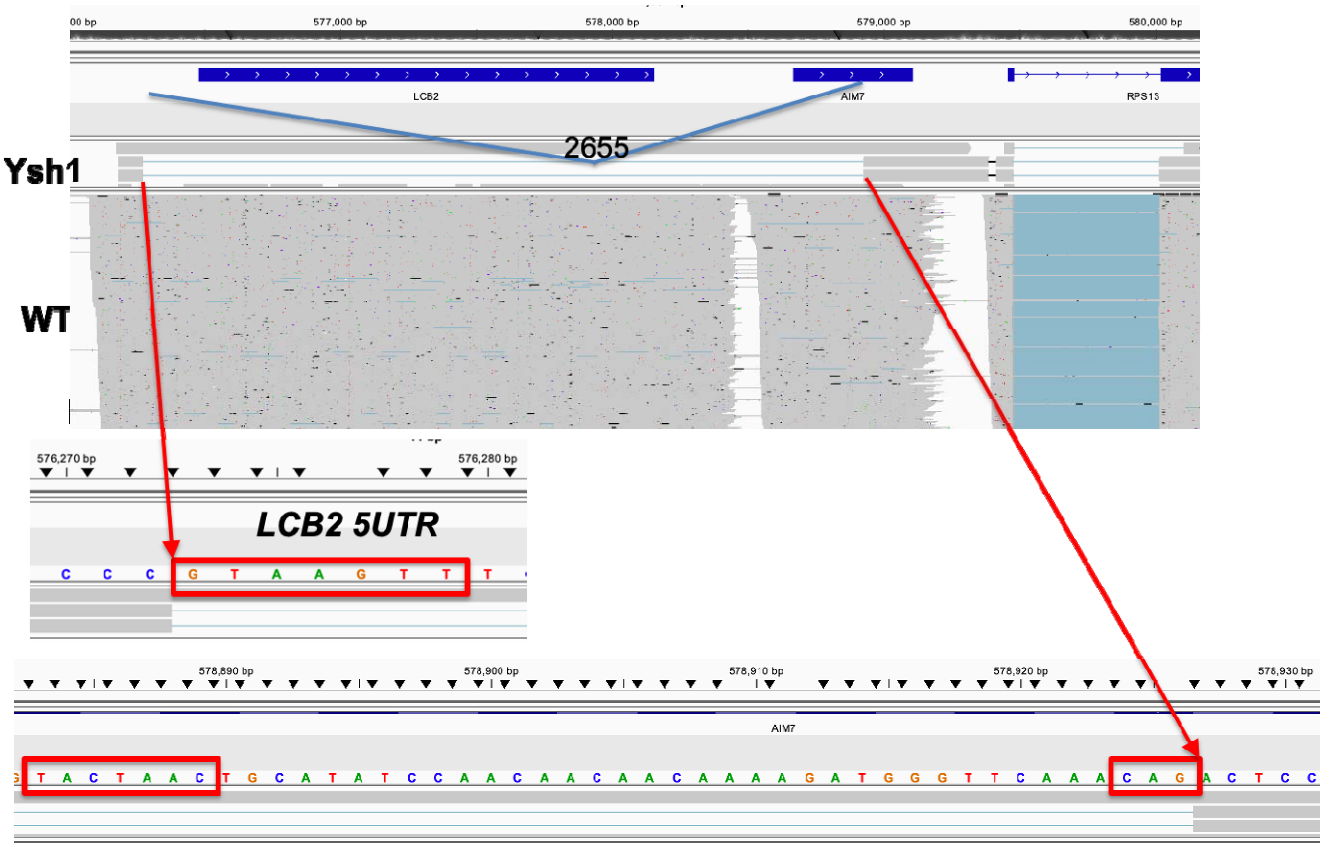


Table S1
Statistics of the number of reads obtained by Oxford Nanopore Sequencing for each strain and replicate samples. Also shown is the number of reads that aligned to the S.cerevisiae genome and the percentage of reads aligned.

Strain	Total reads	Aligned reads	Percent Aligned
WT rep1	4185196	3809153	91.014%
WT rep2	4152656	3715956	89.483%
WT rep3	2745690	2497176	90.948%
RNA15 rep1	1052882	908599	86.296%
RNA15 rep2	1835357	1593438	86.818%
RNA15 rep3	915672	787530	86.005%
YSH1 rep1	2780820	2358478	84.812%
YSH1 rep2	4253350	3603440	84.720%
YSH1 rep3	1787771	1556736	87.076%
SNP1 rep1	3298600	2838622	86.055%
SNP1 rep2	6436917	5471358	84.999%
SNP1 rep3	2681055	2289439	85.393%

Table S2. List of adjacent Intron-Containing genes localized in tandem in S.cerevisiae

Upstream Gene	Downstream Gene
ARF2	RPL35B
RPS9A	RPL21B
RPS9B	RPL21A
RPL19B	LSM2
RPS24A	YOS1
RPL22B	MOB2
APE2	RPS27A
RPL18B	RPS19B
RPL18A	RPS19A
YML6	RPS18B.

Table S3. *S.cerevisiae* strains used in this study

Strain Name in Paper	Abbreviated Genotype	Full Genotype
Wild-Type Anchor Away	HHY168 *fpr1Δ::hphMX4*	ade 2-1 trp1-1 can1-100 leu2-3,112 his3-11,15 ura3 GAL psi+ MATalpha tor1-1 fpr1::HYG RPL13A-2×FKBP12::TRP1 (derived in house)
YSH1 Anchor Away	HHY168 fpr1Δ::hphMX4 YSH1-FRB::His	derived from HHY168 fpr1Δ::hphMX4
RNA15 Anchor Away	HHY168 *fpr1Δ::hphMX4 RNA15-FRB::His*	derived from HHY168 fpr1Δ::hphMX4
RRP6 Anchor Away	HHY168 *fpr1Δ::hphMX4 RRP6-FRB::His*	derived from HHY168 fpr1Δ::hphMX4
YSH1+RRP6 Anchor Away	HHy168 fpr1Δ::hphMX4 YSH1-FRB::KanMX6 RRP6-FRB::His	derived from HHY168 fpr1Δ::hphMX4 RRP6-FRB::His
SNP1 Anchor Away	HHY168 *fpr1Δ::hphMX4 SNP1-FRB::kanMX6*	derived from HHY168 fpr1Δ::hphMX4
Wild-Type	BY4741	MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0