Transcription rate strongly affects splicing fidelity and cotranscriptionality in budding yeast

Vahid Aslanzadeh,1,3 Yuanhua Huang,2,4 Guido Sanguinetti,2 and Jean D. Beggs1

1Wellcome Centre for Cell Biology, University of Edinburgh, Edinburgh EH9 3BF, United Kingdom; 2School of Informatics, University of Edinburgh, Edinburgh EH8 9AB, United Kingdom

The functional consequences of alternative splicing on altering the transcription rate have been the subject of intensive study in mammalian cells but less is known about effects of splicing on changing the transcription rate in yeast. We present several lines of evidence showing that slow RNA polymerase II elongation increases both cotranscriptional splicing and splicing efficiency and that faster elongation reduces cotranscriptional splicing and splicing efficiency in budding yeast, suggesting that splicing is more efficient when cotranscriptional. Moreover, we demonstrate that altering the RNA polymerase II elongation rate in either direction compromises splicing fidelity, and we reveal that splicing fidelity depends largely on intron length together with secondary structure and splice site score. These effects are notably stronger for the highly expressed ribosomal protein coding transcripts. We propose that transcription by RNA polymerase II is tuned to optimize the efficiency and accuracy of ribosomal protein gene expression, while allowing flexibility in splice site choice with the non-ribosomal protein transcripts.

[Supplemental material is available for this article.]
Komblihtt et al. 2006; Das et al. 2007; Luco et al. 2010; Huang et al. 2012; Dujardin et al. 2014).

Although alternative splicing is a relatively rare event in budding yeast, it is important for the regulated expression of some genes. For example, intron retention in PTC7 transcripts and alternative splicing of an intron at the 3’ end of FES1 create mRNAs that code for different protein isoforms (Juneau et al. 2009; Gowda et al. 2016). Also, alternative 5’SS use in SRC1 transcripts results in different cellular localization of the Src1 protein isoforms, and expression of the APE2 gene is regulated upon heat shock by alternative 3’SS use (Meyer et al. 2011; Mishra et al. 2011). As alternative splicing requires flexibility in splice site use, not all splice sites have consensus sequences. Splicing quality control mechanisms are thought to monitor the configuration of substrate RNAs in splicing complexes to determine whether appropriate splice site sequences have been selected (Semlow and Staley 2012). As most splicing occurs cotranscriptionally, quality control mechanisms should also function cotranscriptionally, raising questions as to whether aspects of transcription could affect splicing quality control.

Here, we investigate the effect of different transcription elongation speeds on various splicing characteristics in S. cerevisiae. Following very brief metabolic labeling of RNA with 4-thiouracil (4U), we isolate newly synthesized RNA (Barras et al. 2015) and monitor transcription and splicing at high temporal resolution in yeast cells with fast and slow RNAPII mutants. We also assess cotranscriptional splicing by measuring coprecipitation of the spliced mRNA and excised intron that are associated with RNAPII. By using deep RNA sequencing, we determine for the first time how transcription elongation mutants affect splicing fidelity and alternative splicing in budding yeast.

Results

Altering transcription elongation rate changes the amount of nascent RNA that is spliced cotranscriptionally

To gain further insight into the coupling between splicing and transcription, we examined the splicing of newly synthesized transcripts produced by RNAPII elongation mutants that transcribe faster (rpb1-G1097D) or slower (rpb1-H1085Y) than the wild type (WT; RPB1) (Kaplan et al. 2012). The mutations reside in the trigger loop of the large subunit, a mobile domain that is responsible for substrate selection and catalysis. The rpb1-G1097D allele has been determined to result in approximately four times faster elongation in vitro, whereas the rpb1-H1085Y allele results in about eight times slower elongation in vitro, and both mutants reduce the growth rate compared to WT (Supplemental Fig. S1; Kaplan et al. 2012).

Splicing efficiency is generally determined by measuring levels of spliced and unspliced RNAs at steady state (Braberg et al. 2013). However, the level of these RNA species is determined by the rates of transcription and degradation as well by the efficiency of splicing. By briefly labeling RNA with 4U, we could affinity-select nascent, thiolated RNA using streptavidin beads. This method enables capture of the majority of short-lived RNA species, including unspliced pre-mRNAs before they are degraded (Barras et al. 2015). By analyzing the nascent RNA by reverse transcription-quantitative PCR (RT-qPCR), we followed the kinetics of splicing for three well-expressed intron-containing genes: RPL39, RPL28, and ACT1 (Fig. 1A). In the fast mutant, we observe an initial rapid increase in the pre-mRNA/mRNA ratio, indicating that pre-mRNA was synthesized faster than it was removed by splicing (Fig. 1B). After 2.5 min, this ratio declined gradually toward the steady-state level. Conversely, with the slow mutant, transcripts were spliced before they accumulated to a significant level such that the pre-mRNA/mRNA ratio was initially less than for the WT RNAPII even in the case of RPL28 transcripts that splice slowly (Barras et al. 2015).

We also measured the pre-mRNA/mRNA ratio by RT-qPCR of several intron-containing transcripts in total (steady-state) RNA, again finding a higher ratio than WT (indicating intron retention) with the fast mutant and lower for the slow mutant (Fig. 1C). Next, we performed RNA sequencing and measured steady-state splicing efficiency transcriptome-wide. We observe that the effect of the fast mutant to reduce splicing efficiency (more unspliced pre-mRNA) affects mainly ribosomal protein (RP) transcripts, whereas the slow mutant improves splicing efficiency for both RP and non-RP transcripts (Fig. 1D). By reanalyzing published data (Braberg et al. 2013) from a splicing-specific microarray analysis of steady-state RNA from the same fast RNAPII mutant as used here and a different slow mutant, we found the same effect (Supplemental Fig. S2A). Moreover, based on our RNA-seq data, the levels of pre-mRNAs and mRNAs in the RNAPII mutants reveal a negative correlation for splicing, with the fast mutant reducing mRNA levels and the slow mutant increasing mRNA levels (Supplemental Fig. S2B,C).

The highly stable association of nascent transcripts with elongating RNAPII permits cotranscriptional splicing to be measured by analysis of transcripts that copurify with RNAPII (Churchman and Weissman 2011). For this purpose, the TAP-tagged Rpb3 subunit of RNAPII was pulled down, and copurified RNAs were analyzed by RT-qPCR to quantify the mRNA and lariat-intron-containing products of splicing. Our results reveal increased association of the mRNA and lariat-intron with slow RNAPII, indicating enhanced cotranscriptional splicing, and less association with the fast RNAPII compared with WT (Fig. 1E). This is compatible with the 4U analysis (Fig. 1B).

Altering RNAPII elongation rate affects splicing fidelity

To examine whether RNAPII mutants that alter elongation rate affect splicing fidelity, we carried out strand-specific deep sequencing of total RNA from strains in which UPF1 was deleted in order to protect misspliced transcripts from NMD. After aligning the reads to the genome, we obtained, on average, 5,804,817 uniquely mapped split reads per sample, supporting about 3358 potential splicing events. Events that were unlikely to be products of spliceosome activity, such as with mitochondrial transcripts, and events with fewer than five supporting reads (in at least one of the samples) were excluded from further analysis. The remaining exon-exon splice junctions were divided into two groups, known and novel. The known junctions are those that are annotated in the GTF file (Ensembl, version R64-1-1.75) and the novel junctions result from splicing events that are not annotated in the GTF file. Merging the list of novel splicing events found here with those reported in recent transcriptome-wide studies creates a database of alternative and cryptic splicing events for S. cerevisiae (Supplemental Table S1). This includes 21 alternative splicing events that have been reported repeatedly (Kawashima et al. 2014; Schreiber et al. 2015; Gould et al. 2016) and hundreds of new splicing events that were not in those data sets (Fig. 2A). In order to ensure robustness of the splicing fidelity analysis, we performed further filtering, selecting only events that were supported by at least five reads in both replicates of one or more strains. Finally, 244 novel
alternative splicing events were retained, and the splicing error frequency (SEF) was calculated for each event (Fig. 2B; Supplemental Table S1). First, by analyzing novel events occurring in the strain with WT RNAPII, we note a marked difference between RP and non-RP transcripts (Fig. 2C). Despite the greater abundance of RP transcripts, their SEFs (ranging from one in 100,000 up to one in 250) are orders of magnitude lower than those of non-RP transcripts (SEFs ranging from one in 100 to seven in 100). Some novel events were detected with much higher frequency; for example, IWR1 and SPT14 could potentially produce similar amounts of their two alternative isoforms (Fig. 2C). The SEF and the mRNA abundance are anti-correlated for both RNAPII mutants and WT (Fig. 2D), indicating that the detection of alternative splicing events in high-depth RNA-seq data is not due to the higher abundance of mRNAs.
number of reads for highly expressed transcripts. The position distribution of the novel alternative splicing events in the WT strain shows that most of the 166 alternative 3′SSs occur a very short distance from the annotated 3′SS. Although there are only half as many alternative 3′SSs (78 events) (Supplemental Fig. S3), alternative 3′SSs and 5′SSs have similar SEF distributions, and these values are significantly lower for RPs than for non-RPs (boxplot in Supplemental Fig. S3). The majority of the novel alternative splicing events introduced premature translation termination codons in the coding region of the spliced RNAs, which would normally be recognized by the NMD system, leading to transcript degradation (Supplemental Table S1). By using the Fisher’s exact test (P < 0.01, false-discovery rate [FDR] < 0.03) to determine which novel splicing events occur more or less frequently in the fast mutant relative to WT, we identified 57 in RP transcripts, with an approximately equal numbers with reduced and enhanced fidelity. With the slow mutant, we identified 48 novel events in RP transcripts, of which reduced fidelity occurred approximately 1.6 times more often than increased fidelity (Fig. 2E). Similar to the fast mutant, we detected equal numbers of events with reduced and enhanced fidelity with the slow mutant in non-RP transcripts. Collectively, these results demonstrate that fidelity of splicing RP intron-containing transcripts is more sensitive to changes in transcription elongation rate than for non-RP transcripts.

With WT RNAPII, novel upstream 3′SS use is mainly restricted to within a window of ~50 nucleotides (nt) upstream of the annotated 3′SSs (Supplemental Fig. S3). Notably, position distribution of the alternative splice sites whose SEF is significantly different in the fast (Fig. 3A–D) and slow (Fig. 3E–H) mutants relative to WT shows increased use of alternative 3′SSs that occur upstream of the annotated branch point (BP) (note the change in slope of the plot of upstream novel 3′SSs in Fig. 3D,H compared with Supplemental Fig. S3). These events likely make use of more 5′ suboptimal BPs for splicing. For example, two novel 3′SS events in RPL13B are located 330 and 322 nt upstream of the annotated 3′SS, and lariat sequencing (Gould et al. 2016) has detected a cryptic BP (AACTAAT) located 21 and 13 nt upstream of these cryptic 3′SSs (Supplemental Fig. S4). The SEFs for these novel events in RPL13B, and similar events in other transcripts, were increased by both faster and slower transcription, most likely due to increased utilization of more 5′ cryptic BPs (Fig. 3D,H). Notably, fast and slow transcription do not necessarily reduce or increase a specific splicing event in opposite directions. For instance, an alternative downstream 5′SS in RPS27B increases with faster transcription. However, both the fast and slow mutants increase use of an alternative upstream 3′SS in RPL26B (Supplemental Fig. S4). The elongation mutants do not affect occurrence of some
alternative splicing events; for example, alternative 3′SSs in SEC14 and UBC12 are used with the same frequency in all three strains (P > 0.05) (Supplemental Fig. S4). We validated several of these alternative and novel splicing events by RT-PCR (Supplemental Fig. S5).

Features associated with splicing fidelity

In order to identify features that might be responsible for modulating splicing fidelity, we investigated the correlation between splicing error rate in all three strains and various transcript features, including intron length, predicted intron secondary structure (free folding energy, normalized by the intron length), 5′SS, BP and 3′SS scores, and frequency of 1- to 3-mers in the intron (Supplemental Fig. S6A). Transcripts with introns predicted to have less stable structure (higher free folding energy) have higher SEF (P < 1 × 10^-10) (Fig. 4A). There is also a significant negative correlation between the intron length and SEF, indicating that shorter introns are more susceptible to splicing errors (P < 1 × 10^-23) (Fig. 4B). This leads to the prediction that within pairs of RP paralogs that have introns of different lengths (but very similar or identical exons), the one with a shorter intron will have higher SEF than the other. Indeed, this is the case for seven RP paralogs whose intron lengths differ by >100 nt (Fig. 4C; Supplemental Fig. S7A). A similar comparison of seven RP paralogs whose introns differ by <34 nt found no significant difference in their SEF (Supplemental Fig. S7B). As might be predicted, 3′SS score anti-correlates with splicing error rate (P < 1 × 10^-13) (Fig. 4D). Notably, the effect on splicing error rate of having a poor 3′SS score is also significantly greater within RP genes (P < 0.05) (Fig. 4D). There is also a weak but significant negative correlation (P < 1 × 10^-6) between BP score and mean SEF, indicating a small contribution of the BP in maintaining splicing fidelity (Fig. 4E). However, delta G of the BP-3′SS region and splicing fidelity show no correlation (P > 0.05) (Fig. 4F). Collectively, greater SEF (lower fidelity) correlates with shorter introns together with less stable intron secondary structure and poor 3′SS score. We also generated sequence logos for 5′SSs and 3′SSs of novel introns, finding that the novel introns predominantly have suboptimal splice sites (Fig. 4G). The polypyrimidine tract preceding the novel 3′SSs also

Figure 3. Distribution of the novel splice sites in RP (purple) and non-RP (orange) intron-containing transcripts whose SEF is significantly different in the fast and slow mutants relative to WT (P < 0.01). (A,B) Alternative 5′SS and 3′SS splicing events whose fidelity increased with the fast mutant. (C,D) Alternative 5′SS and 3′SS splicing events whose fidelity reduced with the fast mutant. (E,F) Alternative 5′SS and 3′SS splicing events whose fidelity increased with the slow mutant. (G,H) Alternative 5′SS and 3′SS splicing events whose fidelity reduced with the slow mutant. Green region upstream of 3′SS is the mean distance between BP and annotated 3′SS in budding yeast (~37 nt).
shows a relaxation from the consensus sequence, highlighting the significance of the polypyrimidine tract sequence in correct recognition of the 3′SS, even in budding yeast (Plass et al. 2012). Additionally, our analysis shows that novel RP introns whose use increases with fast or slow mutants have significantly higher predicted free energy (Supplemental Fig. S7C,D). In contrast, novel events that are less frequent (increased fidelity) with the RNAPII mutants do not have different free energy compared with annotated introns. This suggests that selection of novel introns with less stable secondary structure increases with changing transcription elongation rate. Using these and other features listed in Supplemental Figure S6A to predict the splicing error rate by a random forest regression model reveals a good correlation (R = 0.804) between observed and predicted splicing error rate (Fig. 4H).

**Figure 4.** Features associated with SEF. Orange points represent novel events in non-RP transcripts, and purple points denote RP transcripts. (A) Positive correlation between delta G of intron (see Methods) and mean SEF of the transcripts. Delta G of intron was divided by the intron length to achieve delta G per nucleotide. More negative delta G values represent more structured introns. (B) Negative correlation between intron length and mean SEF. (C) Mean SEF of seven pairs of paralogs. Paralogs with shorter intron length (IL) have higher SEF. (D) Negative correlation between 3′SS score (see Methods) and mean SEF. (E) Weak negative correlation between BP score and mean SEF. (F) Absence of correlation between delta G of BP-3′SS region and mean SEF. Delta G was divided by the distance between BP and 3′SS to achieve delta G per nucleotide (G) Sequence logos were generated for 5′SSs and 3′SSs of all novel alternative splicing events. For annotated introns, sequence logos were generated only from splice sites of transcripts that had novel alternative splicing events. (H) Correlation between observed and predicted SEF (as described in Methods).

**Altering transcription speed increases splicing of cryptic introns and enhances exon skipping**

The results described above refer to errors in splicing transcripts from the ∼5% of genes that are known to contain introns. Due to the depth of the RNA sequencing, splicing of cryptic introns was detected for 140 normally unspliced transcripts, as evidenced by at least five unique splice junction reads in the WT sequence data (Supplemental Table S1). Perturbing RNAPII elongation rate in either direction increases splicing of these cryptic introns compared with WT (Fig. 5A). This is presumably a consequence of fast and slow transcription reducing splicing fidelity. Interestingly, splicing of the intron at the 3′ end of FES1 (Gowda et al. 2016) is affected differentially with elongation rate.
With the fast mutant, a twofold increase in the FES1 shorter isoform that results from using the promoter-proximal polyadenylation site and a slight increase in splicing of the intron (generating longer isoform) with the slow mutant suggest cotranscriptional regulation of FES1 splicing (Fig. 5B; Supplemental Fig. S7E). This result does not conform with a simple kinetic model that would predict that faster elongation might increase 3' SS recognition and enhance splicing. However, as we have shown that fast transcription decreases cotranscriptional splicing, this result may be explained if cotranscriptional cleavage and polyadenylation is less affected by faster elongation. This is consistent with preferential induction of the shorter isoform due to increased transcription upon heat shock (Gowda et al. 2016). However, splicing of some cryptic introns found with WT RNAPII appear insensitive to alteration in transcription speed, for example, PRPS and MRM2 that have 167- and 63-nt introns, respectively, with canonical splice sites (Miura et al. 2006; Yassour et al. 2009). Previously, only RPL30 was reported to have a novel intron in the second exon (Schreiber et al. 2015). Here we report five more genes like this (Supplemental Table S1). As with other splicing events that disrupt translational open reading frames, such nonproductive splicing may provide a mechanism to reduce expression of these genes by promoting the turnover of the spliced RNA, referred to as spliceosome-mediated decay (Volanakis et al. 2013). Recently, Gould et al. (2016) reported finding an unusual (for yeast) AT-AC intron nested inside the annotated intron of RPL30, but neither we nor two other studies (Kawashima et al. 2014; Schreiber et al. 2015) found evidence for this.

Only 10 genes in S. cerevisiae contain two introns, and RT-PCR evidence of exon skipping was reported for few of these (Hossain et al. 2011; Egecioglu et al. 2012). Searching for evidence of exon skipping in our RNA-seq data, we found a threefold increase in skipping of the TAD3 second exon in the fast mutant (Fig. 5C). Exon skipping is detected with low frequency in other two-intron transcripts, including DYN2 (Howe et al. 2003) but with insufficient read density to draw a clear conclusion (Supplemental Fig. S8). There is no sign of exon skipping for the highly expressed RPL7A and RPL7B, the only RP transcripts with three exons, which is consistent with higher splicing fidelity in RP transcripts (Fig. 2C).

Discussion

Our analysis of the kinetics of splicing nascent 4U-labeled RNA detects production of the pre-mRNA before it is spliced (Fig. 1). Greater accumulation of the pre-mRNA with the fast RNAPII compared with the WT strain indicates that splicing does not keep up with a faster elongating RNAPII. This is in accord with results from
show that splicing fidelity of RP transcripts is reduced by changes in transcription speed, especially by faster transcription, whereas non-RP introns are less sensitive to changes in elongation rate, with fidelity being moderately increased or reduced with the same likelihood (Fig. 2E). We conclude that, especially for RP transcripts, speed of transcription is critical for both the efficiency and the fidelity of splicing.

With WT RNAPII, more novel 3'SSs are used than novel 5'SSs, and it is the frequency with which these novel splice sites are used, rather than the number of different types of splicing events, that is lower for RP compared with non-RP transcripts. Meyer et al. (2011) reported that real (canonical) 3'SSs are more accessible than are cryptic 3'SSs due to effects of intron structure, which might explain why cryptic splice sites are less frequently used in RP transcripts that, in general, have more highly structured introns. Non-RP introns, on average, have lower 3'SS scores than RP introns (Fig. 4D), which may contribute to their higher error frequency. Moreover, the productive alternative splicing events identified to date in budding yeast are exclusively in non-RP transcripts, reflecting the need for greater flexibility in splice site choice.

The higher SEF for RP transcripts with the RNAPII mutants is due mainly to an increased error frequency for the same novel splice sites found with WT RNAPII, rather than use of new splice sites in the mutants (Fig. 3). Also, more errors occur in 3'SS selection with RP transcripts (Fig. 3); therefore although 3'SS score correlates with fidelity in yeast, this does not seem to protect against the effects of changing elongation rate, at least for the RP transcripts. This is unlike the situation with human fibronectin exon domain I (ED7), where responsiveness of exon skipping to elongation rate was inversely proportional to 3'SS strength (Noguès et al. 2003).

It has been proposed that spliceosomes can use any 3'SS located within an optimal distance (10–45 nt) downstream from the BP, but RNA structures can make optimal 3'SSs more accessible while masking suboptimal 3'SSs (Meyer et al. 2011). For example, there is a cryptic 3'SS between BP and annotated 3'SS of RPS23B, but it is masked by secondary structures (Meyer et al. 2011). Interestingly, the fast mutant enhances use of this cryptic 3'SS, suggesting that faster transcription weakens the structure of pre-mRNA in this region, making the suboptimal 3'SS more accessible (Supplemental Fig. S4G). In contrast, the elongation mutants do not increase SEF of an alternative upstream 3'SS event in APE2 (Supplemental Fig. S4H), which was shown to be spliced with higher frequency upon heat shock, probably because of unstable RNA structure in the BP-3'SS region at elevated temperature (Yassour et al. 2009; Meyer et al. 2011). These different effects of transcription elongation on cryptic 3'SS use in RPS23B and APE2 might suggest that changes in elongation rate preferentially affect stable RNA structure of the RP transcripts, which is consistent with the observed correlation between intron structure and splicing fidelity (fig. 4A). However, we did not observe a genome-wide correlation between SEF and delta G of BP-3'SS regions (Fig. 4F). Conceivably, BP-3'SS structure affects use of canonical 3'SSs in a gene-specific manner, and a more detailed analysis is required to investigate these effects.

Downstream novel 3'SSs mainly occur very close to the annotated 3'SSs (Supplemental Fig. S3), presumably constrained by the 10- to 45-nt optimal distance from the BP (Meyer et al. 2011) rather than any lack of downstream “AG” dinucleotides (potential 3'SSs) (Supplemental Fig. S6B). Those few novel 3'SSs that occur further downstream could use cryptic BPs. For example, lariat sequencing detected two additional BPs in TDA5 downstream from the
annotated BP (Gould et al. 2016) that could explain our finding of a suboptimal 3′SS 127 nt downstream from the annotated 3′SS in TDA5 (Supplemental Fig. S4A).

Howe et al. (2003) showed that mutations that create a suboptimal BP sequence in the first intron of DY2 caused second exon skipping and that the defect was partially alleviated by slower transcription. This suggests that slow transcription provides more time for the assembling spliceosome to select the suboptimal BP (within the first intron) before transcription of the competing downstream BP (within the second intron). However, our results show that both faster and slower transcription reduces BP fidelity, with increased utilization of suboptimal BP's upstream of the annotated BP, leading to splicing of proximal cryptic 3′SSs (Fig. 3D,H; Supplemental Fig. S4).

How does transcription elongation affect splicing fidelity? The effects on splicing fidelity of changing transcription rate could reflect less efficient elimination of splicing errors as a result of perturbing cotranscriptional quality control mechanisms, such as kinetic proofreading of splicing (Hopfield 1974; Semlow and Staley 2012). Additionally, altering transcription elongation rate was reported to change multiple properties of nascent RNA and the chromatin environment around sites of active transcription, affecting alternative splicing decisions (Buratti and Baralle 2004; Dujardin et al. 2013; Saldi et al. 2016). The CTD was recently shown to form multivalent interactions with low-complexity domains in other proteins, including certain splicing factors, suggesting another mechanism by which RNAPII might affect cotranscriptional splicing, by compartmentalizing these two processes (Harlen and Churchman 2017). Therefore, a finely tuned coordination in time and space between transcription and splicing may be influenced by many factors.

The relationship that we observe between transcription elongation rate and the use of alternative/cryptic splice sites is not readily explained by a simple "window of opportunity" model, as both faster and slower transcription reduce splicing fidelity. This is more in keeping with the "Goldilocks" model in which an optimal elongation rate is required for a normal processing outcome (Fong et al., 2014). We propose that RNAPII speed may be tuned to optimize splicing efficiency and to balance fidelity with flexibility for making alternative splice site choice in yeast, as in mammals. Studies in budding yeast revealed that RP transcripts tend to be spliced faster and more cotranscriptionally than non-RP transcripts (Barrass et al. 2015; Wallace and Beggs 2017). Our data now show that the efficiency and fidelity of splicing RP transcripts are affected by altered transcription rates, especially by faster elongation, significantly more than for non-RP transcripts. Together, these observations suggest that splicing of RP transcripts is more functionally coupled to transcription than is splicing of non-RP transcripts, and may indicate that it is beneficial to splice cotranscriptionally. RNAPII speed is an abundant and vital component of the translation machinery, and RP gene expression is coordinated to match cell growth rate according to the requirement for more or fewer ribosomes in rapidly or slowly dividing cells respectively (Warner 1999; Pleiss et al. 2007). Overall, the transcription and splicing of RP transcripts seem to be tuned to meet the demands of high expression without compromising quantity and quality of the spliced transcripts. It was previously proposed that there is an optimal transcription rate for splicing different introns in mammalian cells (Fong et al. 2014). In view of the greater sensitivity of RP splicing to transcription speed, we propose that, in budding yeast, the RNAPII elongation rate has evolved in tune with RP transcript splicing to optimize the expression of these highly important genes while allowing greater flexibility in splice site choice for non-RP genes. Important goals for the future will be to determine how the transcription and splicing machineries communicate and how these interactions are regulated to ensure the appropriate outcome.

Methods

Yeast strains and plasmids
RPB1 (CKY690), rpb1-G1097D (CKY772), and rpb1-H1085Y (CKY691) strains were generously provided by Craig Kaplan (Kaplan et al. 2012). UPF1 was deleted from these strains as previously described (Longtine et al. 1998), selecting upf1::HPMX6 transformants on hygromycin B plates (Supplemental Table S2). For 4U labeling, strains contained pRS426-FU1 that encodes the uracil permease Fu1 to boost 4U uptake by cells.

RNA extraction
For total RNA, log phase cells were pelleted and lysed in a bead-beater (BioSci Products) with zirconia beads (Thistle Scientific) and phenol, and then extracted with phenol:chloroform and precipitated with ethanol. For nascent RNA, 4U was added (final concentration 100 µM) to log phase culture at OD600 of 0.8. After 1, 1.5, 2.5, 5, and 10 min, samples were snap frozen by pouring into chilled methanol sitting on dry ice, and then 4U-labeled RNA was purified as previously described (Barrass et al. 2015). The yield of 4U-labeled RNA was similar for all three strains (Supplemental Fig. S9), as estimated by physical measurements (NanoDrop and Agilent Bioanalyzer). For RT-qPCR analysis, equal amounts of nascent RNA were used based on incorporated 4U. To further correct for amount of input RNA, values for pre-mRNA and mRNA were separately normalized to an intron-less transcript (ALG9) in the same sample.

Native elongating transcript purification
Immunoprecipitation of RNAPII and purification of nascent RNA was performed essentially as previously described (Churchman and Weissman 2001). RNAPII was bound via TAP-tagged Rpb3 to IgG-Sepharose beads, eluted by incubation with ProTEV protease (Promega) for 90 min at 18°C, and then collected by Promega spin column. After proteinase K digestion, RNA was precipitated in ethanol.

RT-qPCR
RT-qPCR was carried out as described previously (Alexander et al. 2010a) using oligonucleotides listed in Supplemental Table S3.

RNA sequencing
Total RNA was extracted from log phase culture grown in YPD and quality controlled by Agilent 2100 bioanalyzer RNA NanoChip. BGI (formerly Beijing Genomics Institute) performed the library preparation and strand-specific, 150 base paired-end deep sequencing on an Illumina HiSeq 4000 for two independent biological replicates per strain.

Processing of the raw sequencing data
RNA-seq reads were mapped to S. cerevisiae strain S288C reference genome (version R64) with STAR, a splice-aware aligner (Dobin et al. 2013). To quantify pre-mRNA fractions, the dice-count function from the DICEseq package (Huang and Sanguinetti 2016) was used to obtain the counts for specific read classes; reads that only
belong to pre-mRNA are boundary and intron reads, while reads that only belong to mature mRNA are junction reads. We define the pre-mRNA fraction as pre-mRNA reads/(pre-mRNA reads + mRNA reads).

**Estimation of SEF for novel splicing events**

Alternative or novel splicing events are defined as splicing events that are within intron-containing transcripts and are not annotated in the *S. cerevisiae* annotation file from Ensembl (version R64-1-1.75). We used default settings of STAR aligner for detecting novel splice junctions. By default, STAR filters all annotated splice junctions with overhang (anchor sequence) length less than three on both sides and requires overhang sequence longer than 12 on both sides for reporting unannotated canonical splice junctions. Events supported by less than five unique reads were filtered out. The frequency of each alternative splicing event is defined as

\[
\text{SEF} = \frac{\text{number of reads that align to the novel splice junction}}{\text{number of reads that align to the annotated splice junction}}.
\]

**Estimation of SEF for cryptic introns in normally intron-less transcripts**

Cryptic introns are those spliced from normally intron-less transcripts and are not annotated in the *Saccharomyces* annotation file. Events supported by less than five unique reads were filtered out. The frequency of splicing cryptic introns is defined as

\[
\text{SEF} = \frac{\text{number of reads that align to the cryptic splice junction of the transcript}}{\text{RPKM of the transcript}}.
\]

**Sequence features and prediction of SEF**

To predict splicing error from sequence, we considered the following features as predictors: intron length, Delta G for intron and for BP-3′SS region, 5′SS score, 3′SS score, BP score, and frequency of 84 short sequences (1– to 3-mers) from both annotated introns and novel introns. For the splice site score, we examined the frequency of seeing a certain sequence \(s_g\) ranging from zero to 100 (Crooks et al. 2004). The Delta G is calculated by the empirical frequency of each nucleotide at a certain position, we could define the probability of seeing a certain sequence \(s_g\) = \(s_{g,1}, \ldots, s_{g,k}\) on a gene g as

\[
P(s_g) = \prod_{i=1}^{k} p(s_{g,i}).
\]

By first applying a log_{2} transformation of the probability \(X_g = \log_{2}(P(s_g))\) and then using a linear re-scale \(Y_g = (X_g - \min(X)) / (\max(X) - \min(X)) \times 100\), we have the splice score \(Y_g\) for each gene g, ranging from zero to 100 (Crooks et al. 2004). The Delta G is a free energy score for RNA secondary structure, which is predicted by mfold v3.6 (Zuker 2003). All these features were fetched with pysealib (https://github.com/huangyh99/PySealib). In total, 176 features were used to predict the SEF scores for novel splicing events. Based on these features, a random forest regression model was trained to predict the SEF score at log_{2} scale, and threefold cross-validation was used to evaluate the prediction performance.

**Data access**

All sequencing data from this study have been submitted to the NCBI Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo/) under accession number GSE99161.

**Acknowledgments**

We thank Craig Kaplan for providing the mutant RNAPII yeast strains, members of the Beggs laboratory for advice, and especially Edward Wallace and Susana De Lucas for helpful suggestions regarding the manuscript. We thank Sander Granneman for guidance on the RNA-seq data analysis. This work was supported by Wellcome funding to J.D.B. [104648] and a European Research Council grant MLC3 306999 to G.S. V.A. was supported by a Darwin Trust of Edinburgh studentship. Y.H. was supported by a University of Edinburgh Principal Career Development scholarship. Work in the Wellcome Centre for Cell Biology is supported by Wellcome core funding [092076].

**References**

Alexander RD, Barrass JD, Dietchl B, Kos M, Ohtulowicz T, Robert M-C, Koper M, Karkusiewicz I, Mariconta I, Tollervy D, et al. 2010a. RiboSybs, a high-resolution, quantitative approach to measure the in vivo kinetics of pre-mRNA splicing and 3′-end processing in *Saccharomyces cerevisiae*. RNA 16: 2570–2580.

Alexander RD, Innocente SA, Barrass JD, Beggs JD. 2010b. Splicing-dependent RNA polymerase pausing in yeast. Mol Cell 40: 582–593.

Ameer A, Zaghloul A, Halvarsson J, Wetterbom A, Gyllensten U, Cavelier L, Feuk L. 2011. Total RNA sequencing reveals nascent transcription and widespread co-transcriptional splicing in the human brain. Nat Struct Mol Biol 18: 1435–1440.

Ares M, Grate L, Pauling MH. 1999. A handful of intron-containing genes produces the lion’s share of yeast mRNA. RNA 5: 1138–1139.

Barrass JD, Reid JEA, Huang Y, Hector RD, Sanguinetti G, Beggs JD, Granneman S. 2015. Transcriptome-wide RNA processing kinetics revealed using extremely short 4TU labeling. Genome Biol 16: 282.

Braberg H, Jin H, Moehle EA, Chan YA, Wang S, Shales M, Benschop JJ, Morris JH, Qiu C, Hu F, et al. 2013. From structure to systems: high-resolution, quantitative genetic analysis of RNA polymerase II. Cell 154: 755–789.

Brinster RL, Allen JM, Behringer RR, Gelines RE, Palmer MD. 1988. Intron increase transcriptional efficiency in transgenic mice. Proc Natl Acad Sci USA 85: 836–840.

Buturi E, Baralle FE. 2004. Influence of RNA secondary structure on the pre-mRNA splicing process. Mol Cell Biol 24: 10505–10514.

Carrillo Oesterreich F, Herzel L, Straube K, Hujer K, Howard J, Neugebauer K. 2010a. RiboSys, a high-resolution, quantitative approach to measure the in vivo kinetics of pre-mRNA splicing and 3′-end processing in *Saccharomyces cerevisiae*. RNA 16: 2570–2580.

Churchman LS, Weissman JS. 2001. Native elongating transcript sequencing (NET-seq). Curr Protoc Mol Biol doi: 10.1002/0471142727. mb0414.98.

Churchman LS, Weissman JS. 2011. Nascent transcript sequencing visualizes transcription at nucleotide resolution. Nature 469: 368–373.

Crooks GE, Hon G, Chandonia J-M, Brenner SE. 2004. Weblogo: a sequence logo generator. Genome Res 14: 1188–1190.

Das R, Yu J, Zhang Z, Gygi MP, Kraner AR, Gygi SP, Reed R. 2007. SR protein function in coupling RNP II transcription to pre-mRNA splicing. Mol Cell 26: 867–881.

de la Mata M, Kornblitht AR. 2006. RNA polymerase II C-terminal domain mediates regulation of alternative splicing by Srp52. Nat Struct Mol Biol 13: 973–980.

de la Mata M, Alonso CR, Kardener S, Fededa JP, Blaustein M, Pelisch F, Aslanzadeh E, de la Mata M, Kornblitht AR. 2003. A slow RNA polymerase II affects alternative splicing in vivo. Mol Cell 12: 529–532.

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013.STAR: ultrafast universal RNA-seq alignment. Bioinformatics 29: 15–21.

Dujardin G, Lafaille C, Petrillo E, Baggio V, Gomez Acuna LI, Fiszbein A, Godoy Herz MA, Nieto Moreno N, Muñoz MJ, Alló M, et al. 2013.
Transcriptional elongation and alternative splicing. *Biochim Biophys Acta* **1829**: 134–140.

Dujardin J, Laflaie C, de la Mata M, Marasco LE, Muñoz MJ, Le Jossic-Corcos C, Corcos L, Kombilhtrt AR. 2014. How slow RNA polymerase II elongation favors alternative exon skipping. *Mol Cell* **85**: 685–690.

Egecioglu DE, Kawashima TR, Chanfreau GF. 2012. Quality control of MATα1 splicing and exon skipping by nuclear RNA degradation. *Nucleic Acids Res* **40**: 1787–1796.

Eperon LP, Graham JR, Griffiths AD, Eperon IC. 1988. Effects of RNA secondary structure on alternative splicing of pre-mRNA: Is folding limited to a region behind the transcribing RNA polymerase? *Cell* **54**: 393–401.

Fong YW, Zhou Q. 2001. Stimulatory effect of splicing factors on transcriptional elongation. *Nature* **414**: 929–933.

Fong N, Kim H, Zhou Y, Ji X, Qiu J, Saltdi T, Diener K, Jones K, Fu X-D, Bentley DL. 2014. Pre-mRNA splicing is facilitated by an optimal RNA polymerase II elongation rate. *Genes Dev* **28**: 2663–2676.

Furger A, O'Sullivan JM, Binnie A, Lee BA, Proudfout NJ. 2002. Promoter proximal splice sites enhance transcription. *Genes Dev* **16**: 2792–2799.

Gould GM, Paggi JM, Guo Y, Phizicky DV, Zinshetyn B, Wang ET, Gilbert WV, Gifford DK, Burge CB. 2016. Identification of new branch points and unconventional introns in *Saccharomyces cerevisiae*. *RNA* **22**: 1522–1534.

Gowda NKC, Kaimal JM, Masser AE, Kang W, Friedländer MR, Andréasson C. 2016. Cytosolic splice isoform of Hsp70 nucleotide exchange factor Fes1 is required for the degradation of misfolded proteins in yeast. *Mol Biol Cell* **27**: 1210–1219.

Harlen KM, Churchman LS. 2017. The code and beyond: transcription regulation by the RNA polymerase II carboxy-terminal domain. *Nat Rev Mol Cell Biol* **18**: 263–273.

Hopfield JJ. 1974. Kineticproofreading: a new mechanism for reducing errors in biosynthetic processes requiring high specificity. *Proc Natl Acad Sci U S A* **71**: 4135–4139.

Hossain MA, Rodríguez CM, Johnson TL. 2011. Key features of the two-intron *Saccharomyces cerevisiae* gene SUS1 contribute to its alternative splicing. *Nucleic Acids Res* **39**: 8612–8627.

Howe kj, Kane CM, Ares M. 2003. Perturbation of transcription elongation influences the fidelity of internal exon inclusion in *Saccharomyces cerevisiae*. *RNA* **9**: 993–1006.

Huang Y, Sanguinetti G. 2016. Statistical modeling of isoform splicing dynamics from RNA-seq time series data. *Bioinformatics* **32**: 2963–2972.

Luo S, Khrebtukova I, Gnirke A, et al. 2009. Ab initio construction of a large-scale full-length cDNA analysis to explore the budding yeast transcriptome. *Genome Res* **19**: 235–241.

Sayani S, Janis M, Lee CY, Toesca I, Chanfreau GF. 2008. Widespread impact of nonsense-mediated mRNA decay on the yeast intronome. *Mol Cell* **31**: 360–370.

Schreiber K, Csaba G, Hasbeck M, Zimmer R. 2015. Alternative splicing in the next generation sequencing data of *Saccharomyces cerevisiae*. *PLoS One* **10**: e0140487.

Seminow DR, Staley JP. 2012. Staying on message: ensuring fidelity in pre-mRNA splicing. *Trends Biochem Sci* **37**: 263–273.

Volanakis A, Passoni M, Hector RD, Shah S, Kilchert C, Granneman S, Vasiljeva L. 2013. Spliceosome-mediated decay (SMD) regulates expression by RNA Pol II elongation: Faster is not always better. *JMBB* **68**: 235–241.

Monsalve M, Wu Z, Adelmann G, Puigserver F, Fan M, Spiegelman BM. 2000. Direct coupling of transcription and mRNA processing through the thermogenic coactivator PGC-1. *Mol Cell* **6**: 307–316.

Nafteberg S, Schor IE, Ast G, Kombilhtrt AR. 2015. Regulation of alternative splicing through coupling with transcription and chromatin structure. *Annu Rev Biochem* **84**: 165–198.

Nogués G, Kadener S, Cramer P, de la Mata M, Fededa JP, Blaustein M, Srebrow A, Kombilhtrt AR. 2003. Control of alternative pre-mRNA splicing by RNA Pol II elongation: Faster is not always better. *JMBB* **68**: 235–241.

Plass M, Codony-Servat C, Ferreiraga PG, Vilardell J, Eyres E. 2012. RNA secondary structure mediates alternative 3’s selection in *Saccharomyces cerevisiae*. *RNA* **18**: 1103–1115.

Pleiss JA, Whitworth GB, Bergkessel M, Guthrie C. 2007. Rapid, transcript-specific changes in splicing in response to environmental stress. *Mol Cell* **27**: 928–937.

Roberts GC, Gooding C, Mak HY, Proudfout NJ, Smith CW. 1998. Co-transcriptional commitment to alternative splice site selection. *Nucleic Acids Res* **26**: 5568–5572.

Saldi T, Cortazar MA, Sheridan RM, Bentley DL. 2016. Coupling of RNA polymerase II transcription elongation with pre-mRNA splicing. *J Mol Biol* **426**: 2625–2635.

Lewis BP, Green RE, Brenner SE. 2003. Evidence for the widespread coupling of alternative splicing and nonsense-mediated mRNA decay in humans. *Proc Natl Acad Sci U S A* **100**: 189–192.

Longtine MS, McKenzie A, Dematini DJ, Shah NG, Wach A, Brachat A, Philippens P, Pringle JR. 1998. Additional modules for versatile and economical PCR-based gene deletion and modification in *Saccharomyces cerevisiae*. *Yeast Chichester Engl* **14**: 953–961.

Luco RF, Pan Q, Tominaga K, Blencowe BJ, Pereira-Smith OM, Misteli T. 2010. Regulation of alternative splicing by histone modifications. *Science* **327**: 996–1000.

Meyer M, Plass M, Pérez-Valle J, Eyras E, Vilardeil J. 2011. Deciphering 3's selection in the yeast genome reveals an mRNA thermosensor that mediates alternative splicing. *Mol Cell* **45**: 1033–1039.

Mishra SK, Ammon T, Popowicz GM, Krajewski M, Nagel RJ, Ares M, Holak TA, Jentsch S. 2011. Role of the ubiquitin-like protein Hub1 in splice-site usage and alternative splicing. *Nature* **474**: 173–178.

Murata F, Kawaguchi N, Sese J, Toyoda A, Hattori M, Morishita S, Ito T. 2006. A large-scale full-length cDNA analysis to explore the budding yeast transcriptome. *Proc Natl Acad Sci U S A* **103**: 17846–17851.

Received May 25, 2017; accepted in revised form December 14, 2017.