Long-read sequencing of nascent RNA in *S. pombe* reveals coupling among RNA processing events

Lydia Herzel[1], Korinna Straube, and Karla M. Neugebauer[#]

Department of Molecular Biophysics and Biochemistry, Yale University, New Haven USA

[#]corresponding author

karla.neugebauer@yale.edu

[1]present address: Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02142, USA.

running title: Long-read sequencing of nascent RNA in *S. pombe*

keywords: pre-mRNA splicing, RNA sequencing, polyadenylation cleavage, transcription

Abstract

Pre-mRNA splicing is accomplished by the spliceosome, a megadalton complex that assembles *de novo* on each intron. Spliceosome assembly and catalysis occur co-transcriptionally, suggesting that multi-intron transcripts are spliced in the order of their transcription. Here, we test this prediction by long-read RNA sequencing of nascent multi-intron transcripts in the fission yeast *Schizosaccharomyces pombe*. Most multi-intron transcripts were fully spliced, indicating that splicing is rapid and proceeds in the direction of transcription. Rare cases of intron retention in partially spliced nascent transcripts were associated with first and last intron position, among other features, and underwent export to the cytoplasm. We observed an unexpectedly high proportion of transcripts that were either fully spliced or fully unspliced, indicating that splicing of any given intron is dependent on the splicing status of the other introns in the transcript. Indeed, mild splicing inhibition in the temperature-sensitive *prp2-1* strain, carrying a mutation in the *S. pombe* homolog of vertebrate U2AF65, led to increased frequency of fully unspliced transcripts while the partially spliced transcript frequency was unchanged. Interestingly, fully unspliced transcripts were strongly correlated with reduced polyA site cleavage and transcriptional read-through. Those transcripts were degraded in the nucleus and not exported. Finally, we provide evidence that delaying splicing triggers degradation and alters cellular mRNA levels, showing regulatory significance of co-transcriptional splicing. We conclude that splicing and transcriptional read-through are dependent on the splicing status of neighboring introns, suggesting crosstalk among spliceosomes and polyA site cleavage machinery as they assemble during transcription elongation.

Introduction

Pre-mRNA splicing results in excision of non-coding introns and ligation of exons to produce mature mRNA. The two-step transesterification reaction involving the 5′ and 3′ splice sites (5′ and 3′ SSs) and the branch point sequence (BPS) is catalyzed by the spliceosome, which must assemble *de novo* on each intron (Wahl et al. 2009). Global short-read sequencing (RNA-seq) of nascent RNA from yeast to human has revealed that most introns are removed from the pre-mRNA during transcription by RNA polymerase II (Pol II) (Brugiolo et al. 2013). Therefore, spliceosome assembly and splicing occur on nascent RNA and are closely connected to other processes shaping mRNA expression, from the transcription process itself to 5′ end capping, 3′ end cleavage, and mRNA decay (Bentley 2014; Herzel et al. 2017). For example, 5′ end capping and recruitment of the nuclear cap binding complex promotes spliceosome assembly on the first intron (Gornemann et al. 2005), and histone modifications and variants can promote spliceosome assembly (Gunderson and Johnson 2009; Gunderson et al. 2011; Neves et al. 2017; Nissen et al. 2017). Moreover, splicing and transcription can influence one another to yield splicing-dependent transcriptional pausing and determine alternative splicing patterns (Schor et al. 2013; Dujardin et al. 2014; Fong et al. 2014; Saldi et al. 2016). The mechanisms underlying coordination between RNA processing and transcription are the subject of intense investigation.

Recent efforts have focused on determining the *in vivo* dynamics of gene expression in multiple species and biological contexts, providing insight into the coordination between splicing and transcription. For example, unspliced introns can lead to nuclear retention of 3′ end cleaved transcripts, suggesting that splicing rates can be regulated to promote particular gene expression programs (Bhatt et al. 2012; Boutz et al. 2015). Over the past ten years, numerous studies have employed a variety of methods – including live cell imaging, metabolic labeling and analysis of nascent RNA – to determine *in vivo* splicing rates; the data indicate a wide range of splicing kinetics that vary according to method and species (Alpert et al. 2017). The core spliceosome itself is

highly conserved (Fabrizio et al. 2009); yet, gene architecture is strongly species-dependent and related to splicing mechanisms and likely a key factor defining splicing kinetics. Specifically, the length of introns, their SS and BPS diversity, and intron number per gene increase with the complexity of the organism. Typical human genes contain eight introns (Sakharkar et al. 2005), increasing cellular demands for splicing machinery and generating vast potential for alternative splicing (Lee and Rio 2015). In higher metazoans, introns are ten times longer than exons and typical internal exons are 150 nt long (Zhang 1998), and spliceosomes are thought to assemble following an exon definition mechanism, whereby protein interactions across internal exons are required (Robberson et al. 1990). When introns are shorter than exons, a so-called intron definition mechanism triggers spliceosome assembly through bridging of 5′ and 3′ SSs across the intron by the U1 and U2 small nuclear ribonucleoproteins (snRNPs), respectively. The observed high frequency of co-transcriptional splicing despite the relative shortness of intron-containing genes in *Saccharomyces cerevisiae* suggests efficient spliceosome assembly through the intron definition mechanism (Alexander et al. 2010; Carrillo Oesterreich et al. 2010; Barrass et al. 2015; Wallace and Beggs 2017). Recently, our lab used two single molecule RNA sequencing methods to analyze the position of Pol II when exon ligation occurs (Oesterreich et al. 2016). We found that 50% of splicing events are complete when Pol II is only 45 nucleotides (nt) downstream of 3′SSs, indicating that the spliceosome is physically close to Pol II during catalysis. This raises the important question of how splicing is achieved in an organism that, unlike budding yeast, has multi-intron genes. If splicing occurs as rapidly in multi-intron genes as in single intron budding yeast genes, introns would be spliced in the order of their transcription. However, this has not yet been determined.

Here, we address co-transcriptional splicing efficiency and the possibility that rapid splicing imposes order on intron removal – from first to last intron as transcription proceeds – in the fission yeast *Schizosaccharomyces pombe.* Splicing rates in *S. pombe* are similar to those reported in *S. cerevisiae* (Alexander et al. 2010; Barrass et al. 2015; Eser et

al. 2016; Oesterreich et al. 2016). *S. pombe* introns are very short with a median length of 56 nt, whereas exons are similar in length to their counterparts in higher eukaryotes with an internal exon median of 137 nt (Kupfer et al. 2004; Herzel 2015). This gene architecture is consistent with an intron definition mechanism. In contrast to *S. cerevisiae*, ~50% of *S. pombe* protein-coding genes contain introns, of which >1200 have multiple introns. Further, the *S. pombe* pre-mRNA splicing machinery is more similar to higher eukaryotes than to *S. cerevisiae* (Kaufer and Potashkin 2000). Alternative splicing (predominantly intron retention) has been reported in *S. pombe* (Awan et al. 2013; Bitton et al. 2014; Bitton et al. 2015; Kuang et al. 2017) and is important in meiosis (Averbeck et al. 2005; Kilchert et al. 2015; Kuang et al. 2017). To observe the progression of splicing during transcription, *S. pombe* nascent RNA (nRNA) was purified from a chromatin fraction and subjected to both short- and long-read sequencing (LRS). LRS of full length nascent transcripts directly reports on the splicing status of individual introns in multi-intron genes relative to the position of Pol II. A key feature we uncovered is that the order of intron removal cannot be inferred from short-read RNA-seq data. Instead, LRS yields different populations of nascent RNA that contain disproportionate fractions of either fully spliced or fully unspliced introns, expanding the range of splicing outcome on a per transcript basis and revealing unexpected coordination between splicing of neighboring introns and polyA site cleavage.

Results

Proteomic and transcriptomic characterization of *S. pombe* chromatin

To isolate nascent RNAs, *S. pombe* chromatin was purified and the RNA and protein composition was determined (Fig. 1A). *S. pombe* cells were harvested in early exponential growth and fractionated into cytoplasm and nuclei (Fig. S1A). Lysed nuclei were further separated into an insoluble chromatin pellet and a soluble nucleoplasmic fraction. Western blot analysis of all fractions and electrophoresis of nucleic acids was used to evaluate the fractionation (Fig. 1B), showing enrichment of chromatin-associated proteins and genomic DNA and depletion of mature ribosomal RNA in the chromatin fraction. Mass spectrometry was used to further assess the biochemical fractionation and confirmed that nuclear, chromatin-associated protein complexes were present, whereas the cytoplasmic fraction contained mainly known cytoplasmic proteins (Table 1, S1B-C, Table S3). Interestingly, many splicing factors were detected in the chromatin fraction, likely reflecting the greater prevalence of intron-containing genes and their co-transcriptional splicing compared to *S. cerevisiae* (Table 1), where splicing factors were not detected in the chromatin fraction by mass spectrometry (Carrillo Oesterreich et al. 2010). Overall, the obtained RNA, DNA and protein signature upon *S. pombe* cell fractionation reflects purified chromatin.

To enrich for nascent protein-coding transcripts, RNA purified from the chromatin fraction was further depleted of polyadenylated RNAs and rRNA, using bead-based negative selection with oligo-dT and the RiboZero kit (Fig. 1A, S1). As polyadenylated transcripts can be retained on chromatin (Bhatt et al. 2012) and/or contaminate the chromatin fraction, the nascent RNA signal would possibly be masked in sequencing experiments without polyA+ RNA removal. Three biological replicates of chromatin-associated, non-polyadenylated RNA and mRNA from the respective cytoplasmic fraction were prepared for RNA sequencing on the Illumina platform, with an average of 14 Mio mapped reads per replicate. As expected, introns and regions downstream of genes showed higher signal coverage in the nascent RNA than in

mRNA, indicating the presence of unspliced pre-mRNA and uncleaved, non-terminated RNA (Fig. 1C).

To determine the fraction of spliced introns in nascent RNA and mRNA, we quantified splicing per intron (SPI) values using the fraction of spliced junction reads compared to all junction reads of a particular intron (Herzel and Neugebauer 2015), for the majority of introns (4,481) in the *S. pombe* genome (Figs. 1C-D and S2A; Table S6). An SPI of 1 reflects 100% splicing and an SPI of 0 no splicing of the respective intron. The SPI distribution median for nRNA was 0.59 and for mRNA 0.95. Hence, 50% of introns were spliced to ≥59% in our nRNA dataset. Accurate splicing quantification of first introns might be confounded by biases due to lower RNA-seq coverage in short first exons. To account for this, only introns with a distance to the transcription start site (TSS) greater than 76 nt were included in the analysis (Supplemental methods). Furthermore, 41 introns were randomly selected for semi-quantitative RT-PCR validation, out of which 33 could be evaluated using primers located in the exons flanking the intron of interest (Supplemental methods and Table S2). A strong positive correlation to our nascent RNA-seq data was observed (R=0.7, Fig. S2B), despite the smaller dynamic range of the quantification by RT-PCR. We conclude that *S. pombe* nRNA-seq accurately quantifies co-transcriptional splicing levels for individual introns in the *S. pombe* genome.

Table 1: Mass spectrometry of *S. pombe* chromatin with a comparison to *S. cerevisiae*. *S. cerevisiae* data from (Carrillo Oesterreich et al. 2010). 437 proteins were enriched in the chromatin fraction. The table lists the number of detected proteins relative to the number of annotated proteins for a certain function or complex (complete list in Table S3). A dash indicates complete absence of subunits of a protein complex. ISW1/2 are not present in the *S. pombe* genome (NA).

		S. cerevisiae	*S. pombe*
RNA polymerases	Pol I	12/14	10/14
	Pol II	10/12	8/12
	Pol III	12/15	7/18
Splicing factors	U1	-	3/9
	U2	-	6/17
	Prp19	-	9/14
	U5	-	2/6
	Sm-Ring	-	6/7
	SR-like	-	1/2
Nucleosomes	Histones	6/6	4/6
Chromatin remodelers	RSC	14/14	12/13
	yINO80	10/14	11/14
	SWI/SNF	8/12	7/12
	ISW1	4/4	NA
	ISW2	3/4	NA
	CHD-type	1/1	2/2
	FACT	1/2	2/2
Replication machinery	Pol ε	1/5	-
	MCM	6/6	-
	Factor C	5/6	5/6
	Factor A	2/3	1/3
Ribosome biogenesis		163/429	144/315
other		167	197

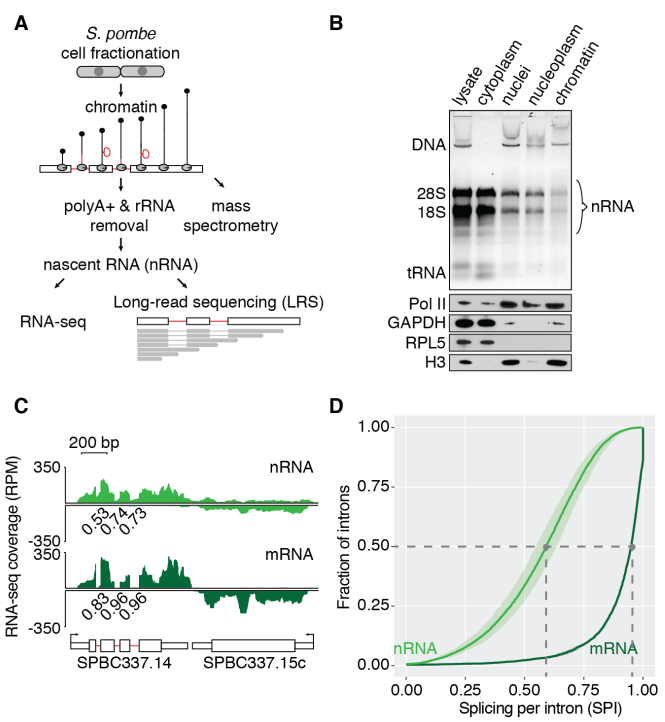

Figure 1: Transcriptome analysis of *S. pombe* chromatin reveals co-transcriptional splicing activity. (A) Schematic of nascent RNA (nRNA) purification from chromatin for short- (RNA-seq) and long-read RNA sequencing (LRS). (B) Enrichment of genomic DNA (DNA) and nRNA in the chromatin fraction and depletion of mature rRNA (18S & 28S) and tRNA revealed by gel electrophoresis and staining with Gelstar (Lonza). Western blot analysis with antibodies specific for chromatin-associated proteins Pol II and Histone 3 (H3) and cytoplasmic marker proteins GAPDH and RPL5. (C) Nascent and mRNA-seq read coverage (RPM) over a 3-intron gene and a convergent intronless gene. The pooled coverage from 3 biological replicates for each cellular fraction is shown. To assess splicing levels in nRNA and mRNA, splicing per intron (SPI) was calculated from intron junction reads. SPI values are shown for these representative introns underneath the RNA-seq coverage track. (D) Cumulative SPI distribution for *S. pombe* introns (nRNA, n=4,481 introns; mRNA, n=2,181). Mean values (line) with standard deviation (shading) are shown for 3 biological replicates. Grey dashed lines indicate the median splicing levels in the two populations (nRNA 0.59, mRNA 0.95).

A multitude of intron-specific features correlate with nascent RNA splicing levels

If pre-mRNA splicing of individual introns occurs in the direction of transcription, a decrease of global pre-mRNA splicing levels towards 3′ ends of genes is expected. Global nRNA splicing levels did not change in the direction of transcription, but rather with the relative distance to the transcript start and end, generally showing higher splicing for internal introns (Fig. 2A). The group of single, first and last introns were spliced to a lesser extent than internal introns, which showed similar SPIs independent of their position (median SPI 0.62). This trend was also detected in cytoplasmic mRNA (Fig. 2B), suggesting that incompletely spliced transcripts can be exported to the cytoplasm. To test if terminal introns show generally lower splicing in individual genes, the difference of SPIs of adjacent introns was calculated (SPI of 5′ intron – SPI of 3′ intron). 25% of those intron pairs showed significant differences in splicing levels (Fig. 2C). Intron pairs were grouped with respect to the SPI of the 5′ intron into introns that are spliced significantly less than the 3′ intron (5′ less) and spliced significantly more than the 3′ intron (5′ more). Consistent with the earlier observation of lower first intron splicing, 176 first introns were present in the 5′ less group and only 73 in the 5′ more group. Next, we compared intron features between the two groups. The 44 analyzed intron features included RNA-seq-derived measures, e.g. SPIs and coverage of exons and introns in RPKM, and gene architecture measures of exons and introns, e.g. splice site and BPS strength, length, conservation in the fission yeast clade and GC-content (see Supplemental methods and Table S5, S8 for a complete overview of intron features). 5′ less introns showed intron signatures reminiscent of first introns, such as short distance to the TSS and early intron position in the gene (from TSS). In total 21 features, e.g. cytoplasmic mRNA-seq intron coverage and 5′SS and BPS strength, were significantly associated with less or more splicing of the 5′ intron in an intron pair (Fig. 2D). Interestingly, mRNA expression also correlated with nascent RNA splicing, with the highest spliced introns belonging to mRNAs with the highest levels (Fig. S2C). We conclude that nRNA intron splicing does not decrease in the direction of transcription,

but first, last and single introns tend to be spliced to less extent co-transcriptionally than internal introns. Intron feature analysis supports that lower splicing results in intron retention detectable in the cytoplasm.

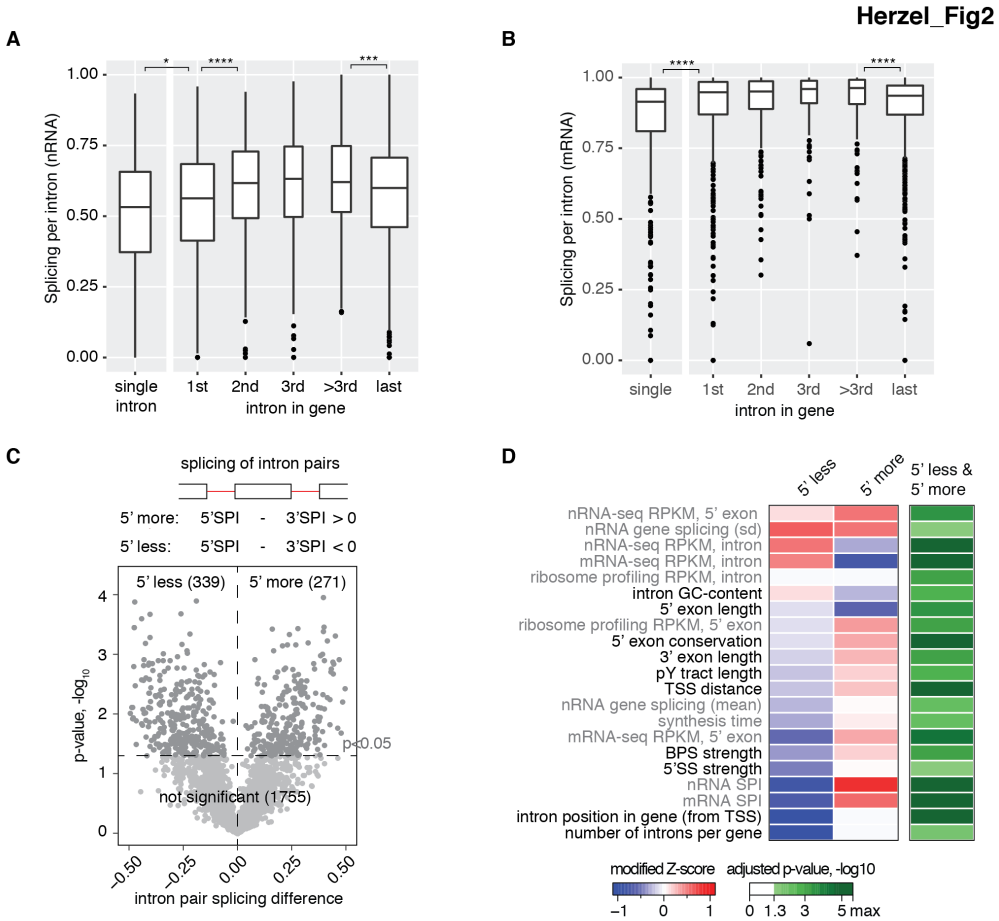

Figure 2: The relative position of introns in the gene and other intron features correlate with the extent of co-transcriptional intron splicing. (A) Co-transcriptional splicing levels differ between introns in different gene positions. The boxplot shows the distribution of nascent RNA SPIs for the group of single intron genes and first, internal (2nd, 3rd and other) or last introns in multi-intron genes. The boxwidth corresponds to the respective group size. (B) mRNA splicing levels differ between introns in different gene positions. The boxplot shows the distribution of mRNA SPIs for the group of single intron genes and first, internal (2nd, 3rd and other) or last introns in multi-intron genes. The boxwidth corresponds to the respective group size. (C) One quarter of introns are significantly less or more spliced than the next downstream (3') intron in nRNA. This is depicted as volcano plot, summarizing data from 3 biological replicates. (D) 21/44 analyzed gene architecture features correlate significantly with differentially spliced

intron pairs (sequence-based in black font and RNA-seq derived in grey font). Characteristics in TSS distance, 5′ exon length and intron position in gene (first intron – 1, second intron – 2 etc.) underline enrichment of first introns in "5′ less spliced" group. The median modified Z-score is shown for each feature with significant difference between the "5′ less" and "5′ more" groups and the respective negative log10 of the Bonferroni-corrected p-value is given. Asterisks indicate significance of direct neighbors according the Wilcoxon-rank sum test ($p < 0.05$ *, $p < 0.01$ **, $p < 0.001$ ***, $p < 0.0001$ **** after Bonferroni-correction) in A & B.

Full-length nascent RNA sequencing of multi-intron transcripts

To assess the order of intron removal directly, nascent RNA was converted into double-stranded cDNA for long-read sequencing (LRS) on the Pacific Biosciences platform. A DNA adaptor with five random nucleotides at the 5′ end was ligated to the 3′ end of all nascent RNAs. This diminishes possible ligation biases from specific 3′ nucleotides and preserves single molecule information (Zhuang et al. 2012; Herzel 2015; Mayer et al. 2015). Template-switching reverse transcription enabled the generation of full-length double-stranded cDNA by PCR, by attaching a universal sequence to the 5′ end. The inclusion of sample-specific barcodes allowed pooling of cDNA preparations, prior to Pacific Biosciences library preparation (Fig. S4). A total of 8 SMRT cells yielded 169,000 high-quality, non-polyadenylated, full-length transcripts from two biological replicates (Supplemental methods). 21% of those transcripts overlapped with intron-containing genes with a median of 7 transcripts per gene. Transcript counts per gene correlated well with expression data from nascent RNA-seq (Fig. S4F).

Full-length transcripts inform about the history of transcription and splicing of every molecule. Nine examples of multi-intron genes and full-length nascent transcripts detected by long-read sequencing are shown in Fig. 3A. Most transcript 5′ ends mapped close to the annotated TSS or at a well defined position downstream of the existing annotation (e.g. SPAC16E8.15). 3′ ends correspond to the nt positions occupied by elongating Pol II at the time of cell lysis. As expected, transcript 3′ ends mapped to varying positions mostly within the gene body, indicating that our approach yields full-

length information on nascent transcripts (Fig. 3A). As seen previously for single intron genes, introns can be already spliced in nascent RNAs when Pol II just transcribed the intron end (Oesterreich et al. 2016) (Fig. 3A, S4H). To evaluate the order of intron removal, it is necessary to identify partially spliced transcripts. Transcripts overlapping multiple introns were classified as "all spliced", if all introns were removed, as "all unspliced", if all introns were present, or as "partially spliced", if some but not all of the introns were spliced (see Fig. 3A). Partially spliced transcripts in the same gene were often associated with a specific intron being (un)spliced, e.g. unspliced intron 1 in SPBC428.01c or the unspliced intron 1 in SPBP8B7.11.

A comparison of the partially spliced transcripts identified by LRS to our nRNA-seq data showed a highly significant overlap between datasets (p-value=$4*10^{-08}$, Fisher's exact test). 5' introns classified as 'unspliced' or 'spliced' in partially spliced transcripts distributed according to the difference in nRNA-seq splicing values between adjacent intron pairs and thus recapitulate the differences in SPI by nRNA-seq (Fig. S5A). To analyze partial splicing patterns with respect to the direction of transcription, partially spliced transcripts were classified into 'in order' (5' introns spliced and one or more 3' introns unspliced), 'not in order' (at least one unspliced intron followed by at least one spliced intron) or transcripts with 'mixed' pattern that contain one or more unspliced intron(s) upstream of a spliced intron and downstream of a spliced intron (Fig. 3B, middle panel). Thus, 'mixed' transcripts combine both 'in order' and 'not in order' splicing. The fraction of 'in order' and 'not in order' transcripts was similar, suggesting that ordered intron removal is not strictly enforced. This is consistent with the nascent RNA-seq data, which did not show a splicing decrease from 5' to 3', but lower splicing of first and last introns. Intron feature analysis showed that 5'SS and BPS strength were reduced in unspliced introns compared to spliced introns of partially spliced transcripts, as well as the nRNA SPIs (Fig. S5B). Further, unspliced introns were associated with higher RNA-seq coverage in our cytoplasmic mRNA-seq data set. Comparison of our nascent LRS and published mRNA LRS data showed a decrease of partially spliced

transcripts to 43% from nascent RNA to mRNA (Fig. S5C), indicating that a substantial fraction of partially spliced transcripts can be polyadenylated and exported to the cytoplasm, reflecting examples of intron retention in *S. pombe*. In conclusion, LRS of nRNA provides a snapshot of splicing states of multi-intron transcripts. The subcategory of partially spliced transcripts recapitulates the splicing differences seen by nRNA-seq and carries signatures of gene-specific intron retention.

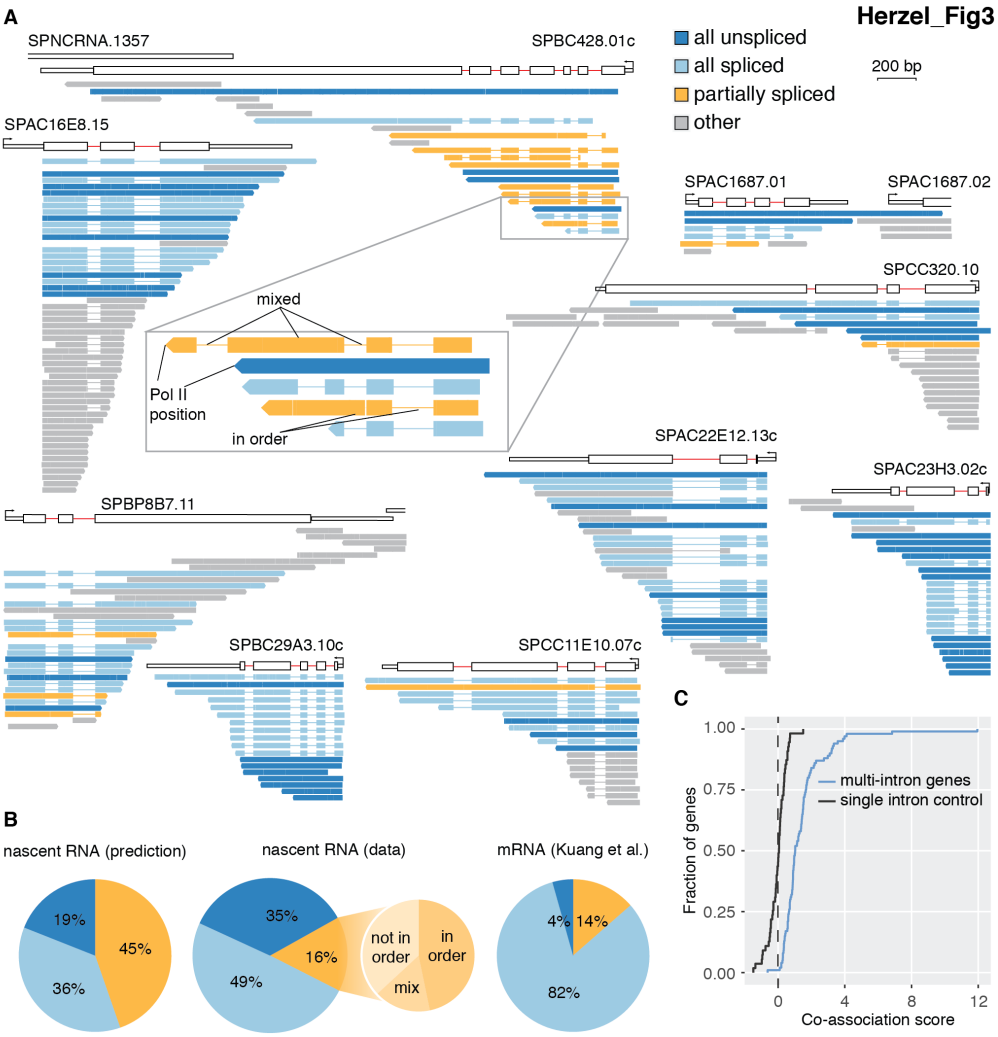

Figure 3: Single molecule long-read sequencing reveals predominantly "all or none" splicing of multi-intron transcripts. (A) Full-length transcripts mapping to nine genes with more than one intron are shown underneath each gene diagram. Transcripts are color-coded according to their splicing profile: "all unspliced" (dark blue), "all spliced" (blue) and "partially spliced" (orange). Transcripts overlapping with <2 introns were

grouped as "others" (grey), because they cannot be used to analyze the order of intron splicing. The inset shows five transcripts of gene SPBC428.01c, highlighting two subclasses of partially spliced transcripts and that transcript 3' ends refer to Pol II positions during transcription. (B) Expected and detected all unspliced, all spliced and partially spliced fraction in the transcriptome. Left panel, prediction of splicing categories from nRNA-seq intron splicing and gene expression data, assuming independence of splicing of adjacent introns. The prediction was done for the first two introns of genes with two and more introns for nRNA (see Supplemental methods for details). Middle panel, proportion of the three splicing categories from nRNA LRS. The small pie chart depicts the fraction of partially spliced transcripts, which show intron removal 'in order' (all introns are spliced upstream an unspliced intron in a particular transcript, compare zoom in of SPBC428.01c in A), 'not in order' (at least one intron is unspliced upstream of a spliced intron in a particular transcript) or a mixed pattern (compare inset of SPBC428.01c in A). Right panel, proportion of the three splicing categories from total mRNA LRS (data from (Kuang et al. 2017)). (C) Splicing of introns in individual genes is co-associated. For the 100 genes with 10 or more sequenced transcripts a co-association score was calculated as the log2-fold change of the observed to predicted fraction of "all or none" splicing. The predicted co-association score is derived from nRNA-seq splicing estimates and assumes independence (see Supplemental methods for details). As a negative control, co-association estimates were calculated for 55 pairs of single intron genes with matched read counts in the LRS data and quantified co-transcriptional splicing from nRNA-seq.

"All or none" splicing of single transcripts predominates

The number of partially spliced transcripts (697) for the above analysis of splicing order was surprisingly small. This caused us to consider how many partially spliced transcripts would be expected from our dataset. Given that each spliceosome assembles anew on every intron in each transcript and that *S. pombe* introns are mainly defined by intron definition, we expected that pre-mRNA splicing events of neighboring introns would be independent of one another. This is illustrated in a prediction for all expressed *S. pombe* intron-containing genes with more than one intron (Fig. 3B), utilizing the intron splicing frequencies from our nascent RNA-seq data. In this scenario, 81% of transcripts would be either partially or completely spliced and only 19% completely unspliced. Surprisingly, this prediction was not recapitulated by our LRS data, in which 84% of transcripts were either completely spliced or unspliced. Only 16% of transcripts were

partially spliced, 2.8 times less than expected if splicing of adjacent introns were independent of each other (Fig. 3B).

To challenge the robustness of this finding, a similar analysis was carried out on a gene-by-gene basis for 100 genes with 10 or more transcripts spanning multiple introns (Fig. 3C). A co-association score was calculated, which we defined as the log2-fold change of observed and predicted fraction of "all or none" transcripts. The median co-association score for the analyzed genes was 1, reflecting a two-fold deviation from independence between adjacent introns. Hence, we conclude that partial transcript splicing is the exception in *S. pombe* on a per transcript basis, and suggest that most introns are removed in a cooperative way. The combined analysis of short- and long-read nascent RNA sequencing data enabled us to distinguish between transcript splicing, resulting in either completely spliced mRNAs, mRNAs with retained introns, or completely unspliced RNAs. Our data support that intron splicing in *S. pombe* not only occurs soon after introns are fully transcribed, but also stimulates splicing of further downstream introns, which manifests itself in an "all or none" fashion in transcripts with multiple introns.

Mild global splicing inhibition increases "all unspliced" transcripts

The above findings suggest that the presence of an unspliced intron might negatively impact the splicing of the neighboring intron in the same transcript. To test this hypothesis, we profiled splicing states of nascent transcripts with multiple introns in the temperature-sensitive mutant *prp2-1* (homolog of mammalian U2AF65, involved in 3'SS recognition), which is known to reduce mRNA splicing levels upon shifting to the non-permissive temperature (Sridharan et al. 2011; Lipp et al. 2015). By semi-quantitative RT-PCR, no splicing differences between wild-type (WT) and *prp2-1* cells were detected at the permissive temperature (Fig. 4A, S6A). After a two-hour shift to the non-permissive temperature, a reduction in pre-mRNA splicing was observed for five analyzed introns in *prp2-1* relative to WT in both cytoplasmic and chromatin fractions (Fig. 4A, S6A-B).

LRS was performed on nRNA isolated from *prp2-1* and WT cells grown for two hours at the non-permissive temperature in three biological replicates. Similar to the RT-PCR experiment, splicing levels were reduced by half in *prp2-1* compared to WT cells (Fig. 4B-C). Classification of transcripts into "all unspliced", "all spliced" and "partially spliced" transcripts revealed that splicing inhibition by *prp2-1* increased the class of "all unspliced" transcripts from 40 to 59% and had only a minor impact on the fraction of "partially spliced" transcripts (Fig. 4B,D). This is consistent with an inhibitory effect of individual unspliced introns on the splicing of the other introns in the same transcript. The latter is supported by adjusting the prediction from nRNA-seq to match the degree of global splicing inhibition seen in the *prp2-1* LRS data (Fig. S6C). Are unspliced transcripts post-transcriptionally spliced or degraded? If unspliced transcripts were predominantly degraded, the relatively mild inhibition of co-transcriptional splicing by *prp2-1* should specifically reduce mRNA levels for intron-containing genes. Indeed, reanalysis of *prp2-1* and WT mRNA expression data measured by microarrays in (Lipp et al. 2015) showed stronger downregulation of mRNA levels for intron-containing genes than intronless genes at the non-permissive temperature for *prp2-1* (Fig. 4E). Changes in gene expression of intronless genes may be indirect effects of altered expression of the large number of intron-containing genes. We conclude from these data that the *prp2-1* mutation increases the levels of "all unspliced" nascent transcripts, consistent with a pervasive inhibitory effect on the splicing of the other introns in the same transcript, and ultimately leads to downregulation of mRNA levels.

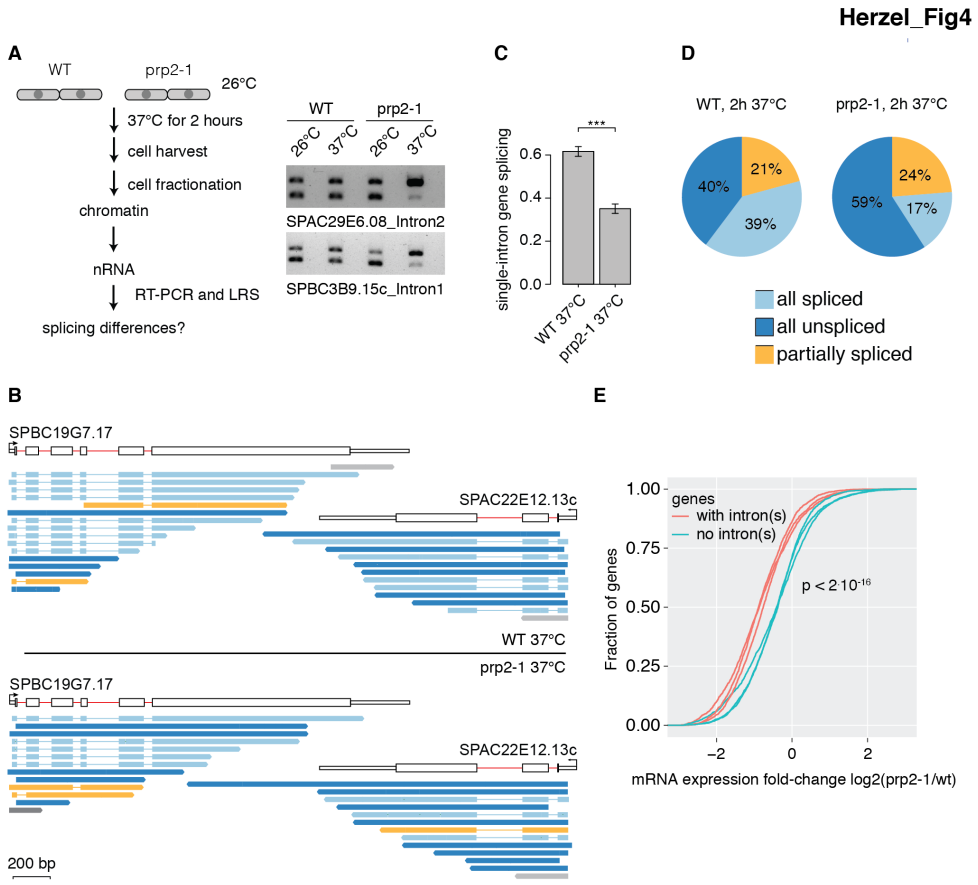

Figure 4: Inhibition of pre-mRNA splicing increases "all unspliced" transcript levels and reduces mRNA levels. (A) Left panel, schematic of the workflow to generate splicing profiles in the temperature-sensitive *prp2-1* mutant and WT strains at non-permissive temperatures. Right panel, RT-PCR shows increased levels of unspliced nascent RNA in the *prp2-1* mutant after 2 hours of growth at 37°C for 2 introns compared to the WT *972h-* strain. –RT control was loaded in lanes (empty) adjacent to the +RT samples. (B) Nascent RNA WT and *prp2-1* LRS read coverage over 2 genes. (C) Barplot indicating significant reduction in splicing frequency for single intron genes in *prp2-1* relative to WT at 37°C calculated from LRS data. The standard deviation from 3 biological replicates is given and asterisks indicate significance according the Student's t-test ($p < 0.001$ ***). (D) Comparison of the "all unspliced", "all spliced" and "partially spliced" fractions in the transcriptome in WT and *prp2-1*. (E) Reduced mRNA levels for intron-containing genes compared to intronless genes in *prp2-1* at 37°C. Cumulative distribution of expression changes between *prp2-1* mutant and WT for intron-containing and intronless genes (data from 3 replicates from (Lipp et al. 2015), p-value from Kolmogorov-Smirnoff test between intronless and intron-containing group).

Failure to splice co-transcriptionally reduces cytoplasmic mRNA levels

Downregulation of mRNA levels caused by degradation of unspliced pre-mRNAs suggests a possible mechanism to fine tune gene expression through co-transcriptional splicing. To determine whether this is a global phenomenon in cells, genes were grouped into two classes following nRNA- and mRNA-seq expression quantification using cufflinks: 1) genes with mRNA levels lower than nRNA levels and 2) genes with mRNA levels higher than or similar to nRNA levels. Higher mRNA levels than nRNA levels can arise in cases where the mRNA half-life is long (Fig. S6F). Identical mRNA and nRNA levels suggest that transcription levels define mRNA expression. If mRNA levels are lower than nRNA levels, RNA decay of pre-mRNA reduces the fraction of mRNA. For group 1 genes, the average pre-mRNA splicing levels were significantly lower than for group 2 genes, underlining the contribution of unspliced RNAs to establish mRNA expression levels (Fig. 5A).

Next, we sought to determine the contribution of splicing inhibition to physiologically relevant cell signaling and shifting gene expression programs. We employed caffeine treatment, which elicits a cellular response similar to nitrogen starvation, to alter mRNA expression levels globally (Rallis and Bahler 2013). Exponentially growing *S. pombe* cells were treated with 10 mM caffeine for 15 min (Fig. 5B). Gene expression and splicing profiles were quantified by nRNA-seq and cytoplasmic mRNA-seq for three biological replicates. Similar to a previous study, ~1,500 genes changed their expression significantly during this short period of time (Fig. 5C, S7A). No significant global splicing changes were detected, when grouping genes according their respective expression profiles in nRNA and mRNA (Fig. S7B). However, that analysis did not account for potential changes in transcription. To determine if co-transcriptional splicing levels alone correlate with mRNA levels, we identified genes among the up- and down-regulated mRNAs whose nascent RNA levels did not change with treatment. Interestingly, significantly lower co-transcriptional splicing frequencies

were observed for the down-regulated mRNAs (Fig. 5D), indicating that failure to splice co-transcriptionally leads to lower gene expression.

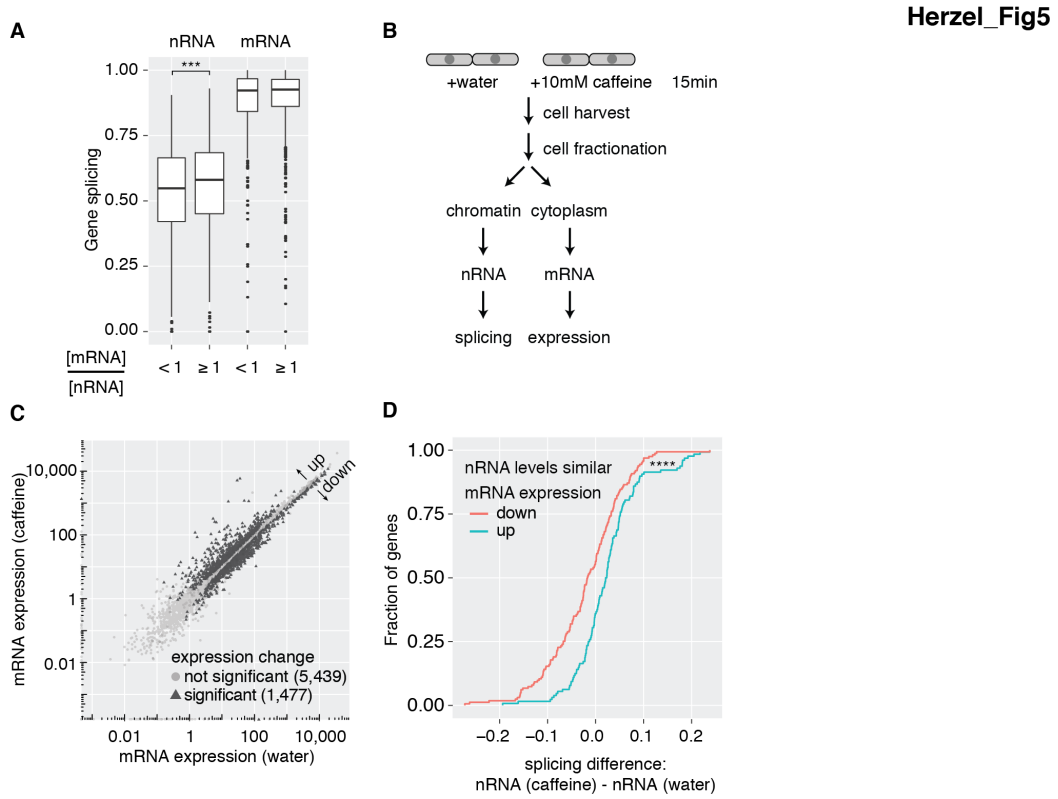

Figure 5: Co-transcriptional splicing correlates with higher mRNA levels. (A) Co-transcriptional gene splicing differences are associated with differences between nRNA and mRNA levels. Boxplot of nRNA and mRNA gene splicing levels after grouping according to cytoplasmic mRNA levels relative to nRNA levels (p-value $p < 0.001$ ***, Wilcoxon-rank sum test). (B) Experimental outline to induce changes in gene expression upon caffeine treatment in *S. pombe* cells. (C) Correlation of mRNA expression values between the two conditions identifies 1,477 differentially expressed genes (FDR-adjusted p-value < 0.05, FDR <= 0.05; 566 of those are intron-containing). (D) Reduced co-transcriptional splicing in genes with reduced mRNA expression upon caffeine treatment, but similar transcription levels. Cumulative distribution of nRNA gene splicing differences between caffeine treatment and control. Only genes without significant changes in nRNA levels, but significant differences in mRNA expression were considered ($p < 0.0001$ ****, Kolmogorov-Smirnoff test between 'mRNA expression up' (n=52) and 'mRNA expression down' (n=65) gene group).

Coordination of splicing status, polyA site cleavage, and nuclear degradation

What is the fate of unspliced transcripts? Strong links between splicing and polyA site cleavage (Herzel et al. 2017) led us to wonder whether the persistence of unspliced introns in nascent RNA might impact downstream RNA processing events such as polyA site cleavage. Indeed, 574 genes in our WT datasets contained unspliced introns and had 3′ ends located downstream of annotated polyA cleavage sites (Fig. 6A-C). The majority of intron-containing transcripts that had a 3′ end after the annotated polyA cleavage site were completely unspliced (Fig. 6B). Nine out of ten of these instances were validated by RT-PCR (Fig. S8A). Given our previous evidence that unspliced transcripts are preferentially degraded (see above), we considered it likely that these transcripts might be degraded rather than further processed. Indeed, total RNA-seq data from a nuclear exosome subunit deletion strain (*Δrrp6*) showed a general increase in the amount of unspliced RNA and higher sequencing read coverage downstream of polyA sites (Fig. 6D, Fig. S8B) (Zhou et al. 2015). For the high coverage gene SPAC1805.11c, the signal downstream of the polyA site was substantially higher in the *Δrrp6* strain than in WT, measured by RT-qPCR (Fig. 6E). This finding substantiates the notion that splicing cooperativity among introns extends to coordination with polyA site cleavage, highlighting a pathway to mRNA degradation based on reduced co-transcriptional splicing efficiency.

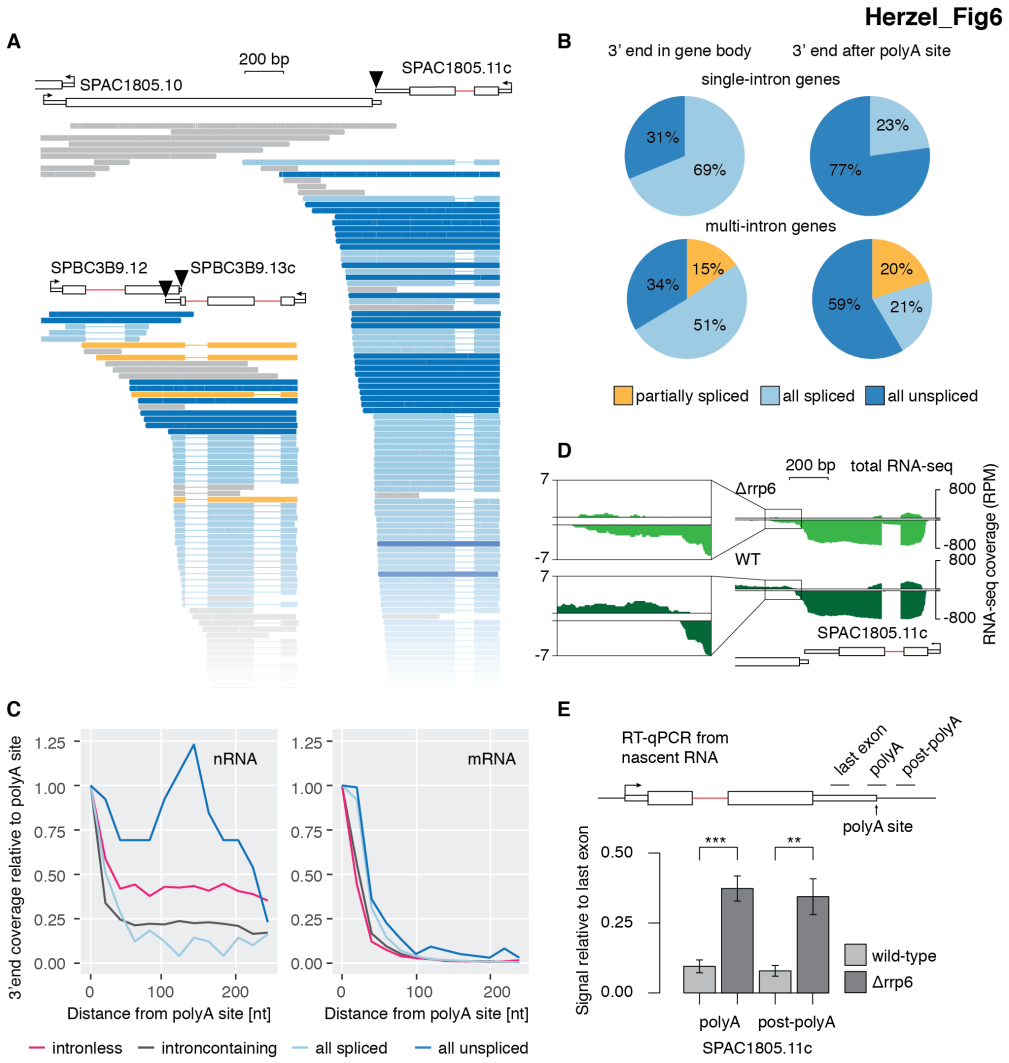

Figure 6: Coupling between co-transcriptional splicing, polyA site cleavage, and mRNA stability. (A) 2/574 examples of intron-containing genes with unspliced transcripts extending over the polyA site. Black triangles mark the polyA cleavage sites. Full representation of all sequenced transcripts in Fig. S8B. (B) Pie charts reflecting the fraction of spliced and unspliced transcripts in single or multi-intron genes with 3′ ends either within the gene body or downstream of the polyA site cleavage site. (C) 3′ end profiles downstream of annotated polyA cleavage sites for different transcript classes. Data are binned in 20 nt intervals and normalized to first bin (-20 nt – 0 nt from polyA cleavage site). (D) WT and Δ*rrp6* total RNA-seq read coverage over the same example gene as in A is shown (log-scale). Counts per nucleotide were normalized to library size. Data reanalyzed from (Zhou et al. 2015). The inset zooms into the region downstream of the annotated polyA site. (E) Nascent transcript levels with 3′ ends extending over the polyA site are increased in the exosome mutant Δ*rrp6*. RT-qPCR from Δ*rrp6* and WT

strains confirmed higher levels of nascent RNA uncleaved at the polyA cleavage site, using qPCR primers to generate amplicons (black line above gene diagram) bridging (polyA) or downstream (post-polyA) of the polyA cleavage site (RT with random hexamers). SDs from 4 biological replicates are given. Asterisks indicate significance according to the Student's t-test ($p < 0.05$ *, $p < 0.01$ **, $p < 0.001$ ***).

Discussion

In this study, we sought to determine whether rapid co-transcriptional splicing in *S. pombe* enforces the removal of introns in the order of their transcription. In depth analysis of nascent RNA by short- and long-read sequencing revealed that distinct populations of fully spliced, fully unspliced and partially spliced nascent RNA are uniquely detectable by LRS. The preponderance of fully spliced nascent transcripts indicates that splicing proceeds rapidly and efficiently in the order of intron synthesis with few exceptions. Moreover, we show that fully unspliced transcripts, which account for the majority of unspliced introns, fail to terminate properly and are degraded in the nucleus. This evidence extends previous co-transcriptional splicing studies from multiple species. Specifically, bulk determinations by short-read RNA-seq yield co-transcriptional splicing efficiencies on a per intron basis, ranging from 45-84% depending on the species and platform (Brugiolo et al. 2013). It is generally assumed that remaining introns are post-transcriptionally spliced. Our findings indicate that, at least in *S. pombe*, the majority of introns that are not spliced co-transcriptionally will be degraded. We present evidence that this mechanism of degrading unspliced RNA contributes significantly to shaping the transcriptome in response to a cellular stimuli changing gene expression. An apparent cooperativity among neighboring introns may lead to "all or none" splicing of individual transcripts, suggesting a specific mechanism(s) for crosstalk among spliceosomes and the polyA site cleavage machinery as they assemble during Pol II elongation.

Our analysis highlights intensive chromatin-associated splicing activity in *S. pombe*. Mass spectrometry of the isolated chromatin fraction identified numerous

splicing factors that were not detected in the *S. cerevisiae* analysis (Carrillo Oesterreich et al. 2010). This likely reflects the 10x higher number of intron-containing genes in the *S. pombe* genome compared to the *S. cerevisiae* genome and is consistent with co-transcriptional splicing and/or the retention on chromatin of the post-splicing intron lariat spliceosome (Chen et al. 2014). Through nascent RNA-seq, we determined splicing frequency on a per intron basis, finding that most *S. pombe* introns are spliced co-transcriptionally (median 0.59). This splicing frequency is lower than in *S. cerevisiae* (median 0.74), *Drosophila* or human cells (Brugiolo et al. 2013). Intron length is often associated with co-transcriptional splicing efficiency (Carrillo Oesterreich et al. 2010; Brugiolo et al. 2013). However, our nRNA-seq analysis (see Figure 2A and S2C) and the analysis of splicing times by others (Eser et al. 2016) show that intron length is one of the few parameters that does *not* correlate with splicing frequency in nascent RNA, while 5'SS and BPS strength do. Low splicing efficiency for first introns suggests that in some cases splicing does not occur in the order of intron synthesis. In other words, the first intron can persist while downstream introns are spliced. In the context of alternative splicing, intron removal can be delayed relative to an intron downstream of a cassette exon; interestingly, these delayed introns can already be committed to splicing (de la Mata et al. 2010). Globally, lower first than internal intron splicing has been detected in numerous other species and datasets (Kessler et al. 1993; Khodor et al. 2011; Khodor et al. 2012; Tilgner et al. 2012). In most species first introns are longer than internal introns and undergo different types of regulation, such as premature transcription termination that can be suppressed by U1 snRNP binding (Berg et al. 2012; Park and Hannenhalli 2015).

Full-length nascent transcript sequences provide information about the life history of the (pre-)mRNA. Data on full-length transcripts can be obtained by RT-(q)PCR (Kessler et al. 1993; Pandya-Jones and Black 2009; Glauser et al. 2011; Bonde et al. 2014), synthetic long-read sequencing (Tilgner et al. 2015) or LRS on the Oxford Nanopore and Pacific Biosciences platforms (Byrne et al. 2017). The latter platform has been previously

used to characterize full-length mature transcripts (Sharon et al. 2013; Kuang et al. 2017). Here, we used it to analyze full-length nascent transcripts genome-wide. Average splicing values for single intron genes in LRS and RT-PCR confirmed results from nRNA-seq. In the body of each transcript, the sequence identifies whether introns are removed or not and how exons are connected to one another. The mapping of 3′ ends of nascent transcripts reports on progression of Pol II through the transcription unit. In terms of transcription analysis, higher transcript coverage per gene is required for information on Pol II dynamics. Promoter-proximal Pol II pausing was identified in *S. pombe* (Booth et al. 2016) and NET-Seq was recently used to analyze transcription elongation (Shetty et al. 2017). LRS of nascent RNA represents an additional promising approach for the analysis of splicing and transcription regulation.

What is the fate of unspliced transcripts? Our data indicate that unspliced transcripts will not terminate properly due to a failure of polyA site cleavage. These transcripts are then degraded by the nuclear exosome. Therefore, our data agree with prior reports based on total RNA-seq, showing that unspliced transcripts are subjected to degradation in an *rrp6*-dependent manner and that splicing competes with nuclear decay (Bousquet-Antonelli et al. 2000; Bitton et al. 2015; Zhou et al. 2015; Kilchert et al. 2016). Previous work identified additional degradation pathways for unspliced transcripts, including the nuclear 5′-3′ exonuclease Dhp1/Rat1/Xrn2 and nonsense mediated decay in the cytoplasm (Vargas et al. 2011; Davidson et al. 2012; Girard et al. 2012; Kervestin and Jacobson 2012; Brogna et al. 2016; Kilchert et al. 2016). We cannot exclude these additional mechanisms. Furthermore, multiple physical interactions have been detected between spliceosomal components and the nuclear decay machinery to directly link splicing fidelity with RNA degradation (Nag and Steitz 2012; Zhou et al. 2015; Herzel et al. 2017). Based on our findings, we suggest that these interactions take place co-transcriptionally. This would explain our observation that poor co-transcriptional splicing is associated with down-regulation of mRNA levels upon caffeine treatment, which we showed are independent of transcriptional changes. This

extends the finding of a previous analysis of mRNA, that pre-mRNA splicing correlates with mRNA expression in *S. pombe* (Wilhelm et al. 2008). Interestingly, a prior study also identified increased mRNA half-life when splicing occurred *in vitro* in a transcription-coupled system (Hicks et al. 2006). Due to the clear link between co-transcriptional intron retention and mRNA decay, we conclude that co-transcriptional splicing is an important determinant of gene expression.

Our observation of unexpectedly high levels of fully spliced and fully unspliced transcripts suggests crosstalk among introns, perhaps at the level of assembling spliceosomes. Considering previous results that pre-mRNA splicing occurs immediately after intron synthesis in *S. cerevisiae* and *S. pombe* (Oesterreich et al. 2016), it was conceivable that introns could be removed in the direction of transcription. Splice site recognition occurs via intron definition in *S. pombe* (Romfo et al. 2000; Shao et al. 2012; Fair and Pleiss 2017), implying that splicing of individual introns in multi-intron transcripts would occur independently of each other. However, we show that the opposite is the case, raising the question of why cells splice multi-intron transcripts cooperatively. Median spliceosomal protein copy number per cell is almost identical to the number of intron-containing genes in *S. pombe* (Marguerat et al. 2012). Hence, high local concentrations of splicing factors at sites of active transcription might promote splicing when the splicing machinery is non-saturating levels or in the context of competition with RNA decay (Munding et al. 2013). Cooperativity could drive the formation of fully spliced mRNAs over partially spliced transcripts to evade decay in the nucleus or cytoplasm. On the other hand, fully unspliced transcripts could arise at genes or in cells in the process of shutting down gene expression, e.g. at entry to mitosis. In addition, completely unspliced transcripts might be degraded more efficiently.

Is splicing cooperativity widespread and how is it achieved? Multiple alternative splicing events are coassociated in human cells, *C. elegans* and *S. pombe* (Fededa et al. 2005; Glauser et al. 2011; Tilgner et al. 2015; Kuang et al. 2017). In *S. cerevisiae*, constitutive splicing of the 2-intron *SUS1* gene has been found to occur cooperatively

(Bonde et al. 2014), and a recent study in human cells characterized individual splicing events that are coupled to adjacent introns (Kim et al. 2017). Thus, this under-appreciated phenomenon has been observed in multiple species. Splicing regulators bind *cis*-regulatory elements in introns and exons (ISEs and ESEs) and could mediate interactions between adjacent spliceosomes. Srp1 and Srp2, homologs of mammalian SR-proteins, could take up this role in *S. pombe* (Webb et al. 2005; Fair and Pleiss 2017). Emphasizing the importance of ESEs for multi-intron splicing, *S. cerevisiae* genome evolution did not only result in mainly single-intron genes, but also several factors implicated in ESE function were lost (Aravind et al. 2000; Kaufer and Potashkin 2000). In addition to splicing regulators, pre-formed spliceosome subcomplexes or remaining association of spliceosomal components after intron removal could also enhance cooperativity (Huang et al. 2002; Crabb et al. 2010; Chen et al. 2014). The presence of pre-spliceosomal components (e.g. U1 snRNP) on remaining upstream introns might also influence spliceosome assembly on downstream introns either positively or negatively. Finally, the local chromatin environment and changes in Pol II may be involved in mediating crosstalk (Gunderson and Johnson 2009; Gunderson et al. 2011; Patrick et al. 2015; Milligan et al. 2017; Neves et al. 2017; Nissen et al. 2017). Localization of promoter-proximal Pol II, general transcription factors, and active chromatin marks depend on the length of first exons in humans (Bieberstein et al. 2012), suggesting a special role for first introns in transcriptional output. Perhaps the overrepresentation of unspliced first introns in RNA-seq datasets from all organisms tested (see above) reflects a regulatory step that might trigger or repress additional downstream splicing events as transcription elongation proceeds.

Methods

In depth information on strains, experimental protocols and data analysis can be found in the Supplemental methods section.

Data access

The accession number for the data reported in this paper is GEO: GSE104681. Multiplexed WT and *prp2-1* CCS reads can be downloaded under doi:10.7910/DVN/PW1KEG.

Acknowledgements

Members of the laboratory of Karla Neugebauer and Gene-Wei Li, Hanspeter Herzel, Michael Weber, Fernando Carrillo Oesterreich and Charles Query provided crucial discussions and critical comments on the manuscript. Jeremy Schofield contributed to some of the preliminary experiments. We are grateful to Iva Tolic, Tamas Fischer and Charles Query for the gift of strains as well as Guilin Wang and the Yale Center for Genome Analysis for technical assistance and advice. The presented work was supported by funding by NIH R01GM112766 from the NIGMS. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the NIH. LH is a postdoctoral fellow of the Helen Hay Whitney foundation.

Author contributions

LH and KMN conceived the study and designed the experiments. Semi-quantitative RT-PCR experiments represented in Figure S2B were carried out by KS. All other experiments and data analyses were performed by LH. LH and KMN wrote the manuscript. All authors read and approved the manuscript.

Disclosure declaration

The authors declare no conflict of interest.

References

Alexander RD, Barrass JD, Dichtl B, Kos M, Obtulowicz T, Robert MC, Koper M, Karkusiewicz I, Mariconti L, Tollervey D et al. 2010. 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.

Alpert T, Herzel L, Neugebauer KM. 2017. Perfect timing: splicing and transcription rates in living cells. *Wiley Interdiscip Rev RNA* **8**.

Aravind L, Watanabe H, Lipman DJ, Koonin EV. 2000. Lineage-specific loss and divergence of functionally linked genes in eukaryotes. *Proc Natl Acad Sci U S A* **97**: 11319-11324.

Averbeck N, Sunder S, Sample N, Wise JA, Leatherwood J. 2005. Negative control contributes to an extensive program of meiotic splicing in fission yeast. *Mol Cell* **18**: 491-498.

Awan AR, Manfredo A, Pleiss JA. 2013. Lariat sequencing in a unicellular yeast identifies regulated alternative splicing of exons that are evolutionarily conserved with humans. *Proc Natl Acad Sci U S A* **110**: 12762-12767.

Barrass JD, Reid JE, 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.

Bentley DL. 2014. Coupling mRNA processing with transcription in time and space. *Nat Rev Genet* **15**: 163-175.

Berg MG, Singh LN, Younis I, Liu Q, Pinto AM, Kaida D, Zhang Z, Cho S, Sherrill-Mix S, Wan L et al. 2012. U1 snRNP determines mRNA length and regulates isoform expression. *Cell* **150**: 53-64.

Bhatt DM, Pandya-Jones A, Tong AJ, Barozzi I, Lissner MM, Natoli G, Black DL, Smale ST. 2012. Transcript dynamics of proinflammatory genes revealed by sequence analysis of subcellular RNA fractions. *Cell* **150**: 279-290.

Bieberstein NI, Carrillo Oesterreich F, Straube K, Neugebauer KM. 2012. First exon length controls active chromatin signatures and transcription. *Cell Rep* **2**: 62-68.

Bitton DA, Atkinson SR, Rallis C, Smith GC, Ellis DA, Chen YY, Malecki M, Codlin S, Lemay JF, Cotobal C et al. 2015. Widespread exon skipping triggers degradation by nuclear RNA surveillance in fission yeast. *Genome Res* **25**: 884-896.

Bitton DA, Rallis C, Jeffares DC, Smith GC, Chen YY, Codlin S, Marguerat S, Bahler J. 2014. LaSSO, a strategy for genome-wide mapping of intronic lariats and branch points using RNA-seq. *Genome Res* **24**: 1169-1179.

Bonde MM, Voegeli S, Baudrimont A, Seraphin B, Becskei A. 2014. Quantification of pre-mRNA escape rate and synergy in splicing. *Nucleic Acids Res* **42**: 12847-12860.

Booth GT, Wang IX, Cheung VG, Lis JT. 2016. Divergence of a conserved elongation factor and transcription regulation in budding and fission yeast. *Genome Res* **26**: 799-811.

Bousquet-Antonelli C, Presutti C, Tollervey D. 2000. Identification of a regulated pathway for nuclear pre-mRNA turnover. *Cell* **102**: 765-775.

Boutz PL, Bhutkar A, Sharp PA. 2015. Detained introns are a novel, widespread class of post-transcriptionally spliced introns. *Genes Dev* **29**: 63-80.

Brogna S, McLeod T, Petric M. 2016. The Meaning of NMD: Translate or Perish. *Trends Genet* **32**: 395-407.

Brugiolo M, Herzel L, Neugebauer KM. 2013. Counting on co-transcriptional splicing. *F1000Prime Rep* **5**: 9.

Byrne A, Beaudin AE, Olsen HE, Jain M, Cole C, Palmer T, DuBois RM, Forsberg EC, Akeson M, Vollmers C. 2017. Nanopore long-read RNAseq reveals widespread transcriptional variation among the surface receptors of individual B cells. *Nat Commun* **8**: 16027.

Carrillo Oesterreich F, Preibisch S, Neugebauer KM. 2010. Global analysis of nascent RNA reveals transcriptional pausing in terminal exons. *Mol Cell* **40**: 571-581.

Chen W, Shulha HP, Ashar-Patel A, Yan J, Green KM, Query CC, Rhind N, Weng Z, Moore MJ. 2014. Endogenous U2.U5.U6 snRNA complexes in S. pombe are intron lariat spliceosomes. *RNA* **20**: 308-320.

Crabb TL, Lam BJ, Hertel KJ. 2010. Retention of spliceosomal components along ligated exons ensures efficient removal of multiple introns. *RNA* **16**: 1786-1796.

Davidson L, Kerr A, West S. 2012. Co-transcriptional degradation of aberrant pre-mRNA by Xrn2. *EMBO J* **31**: 2566-2578.

de la Mata M, Lafaille C, Kornblihtt AR. 2010. First come, first served revisited: factors affecting the same alternative splicing event have different effects on the relative rates of intron removal. *RNA* **16**: 904-912.

Dujardin G, Lafaille C, de la Mata M, Marasco LE, Munoz MJ, Le Jossic-Corcos C, Corcos L, Kornblihtt AR. 2014. How slow RNA polymerase II elongation favors alternative exon skipping. *Mol Cell* **54**: 683-690.

Eser P, Wachutka L, Maier KC, Demel C, Boroni M, Iyer S, Cramer P, Gagneur J. 2016. Determinants of RNA metabolism in the Schizosaccharomyces pombe genome. *Mol Syst Biol* **12**: 857.

Fabrizio P, Dannenberg J, Dube P, Kastner B, Stark H, Urlaub H, Luhrmann R. 2009. The evolutionarily conserved core design of the catalytic activation step of the yeast spliceosome. *Mol Cell* **36**: 593-608.

Fair BJ, Pleiss JA. 2017. The power of fission: yeast as a tool for understanding complex splicing. *Curr Genet* **63**: 375-380.

Fededa JP, Petrillo E, Gelfand MS, Neverov AD, Kadener S, Nogues G, Pelisch F, Baralle FE, Muro AF, Kornblihtt AR. 2005. A polar mechanism coordinates different regions of alternative splicing within a single gene. *Mol Cell* **19**: 393-404.

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

Girard C, Will CL, Peng J, Makarov EM, Kastner B, Lemm I, Urlaub H, Hartmuth K, Luhrmann R. 2012. Post-transcriptional spliceosomes are retained in nuclear speckles until splicing completion. *Nat Commun* **3**: 994.

Glauser DA, Johnson BE, Aldrich RW, Goodman MB. 2011. Intragenic alternative splicing coordination is essential for Caenorhabditis elegans slo-1 gene function. *Proc Natl Acad Sci U S A* **108**: 20790-20795.

Gornemann J, Kotovic KM, Hujer K, Neugebauer KM. 2005. Cotranscriptional spliceosome assembly occurs in a stepwise fashion and requires the cap binding complex. *Mol Cell* **19**: 53-63.

Gunderson FQ, Johnson TL. 2009. Acetylation by the transcriptional coactivator Gcn5 plays a novel role in co-transcriptional spliceosome assembly. *PLoS Genet* **5**: e1000682.

Gunderson FQ, Merkhofer EC, Johnson TL. 2011. Dynamic histone acetylation is critical for cotranscriptional spliceosome assembly and spliceosomal rearrangements. *Proc Natl Acad Sci U S A* **108**: 2004-2009.

Herzel L. 2015. Co-transcriptional splicing in two yeasts. In *Biology*, Vol 1.0. Technical University Dresden.

Herzel L, Neugebauer KM. 2015. Quantification of co-transcriptional splicing from RNA-Seq data. *Methods* **85**: 36-43.

Herzel L, Ottoz DSM, Alpert T, Neugebauer KM. 2017. Splicing and transcription touch base: co-transcriptional spliceosome assembly and function. *Nat Rev Mol Cell Biol* **18**: 637-650.

Hicks MJ, Yang CR, Kotlajich MV, Hertel KJ. 2006. Linking splicing to Pol II transcription stabilizes pre-mRNAs and influences splicing patterns. *PLoS Biol* **4**: e147.

Huang T, Vilardell J, Query CC. 2002. Pre-spliceosome formation in S.pombe requires a stable complex of SF1-U2AF(59)-U2AF(23). *EMBO J* **21**: 5516-5526.

Kaufer NF, Potashkin J. 2000. Analysis of the splicing machinery in fission yeast: a comparison with budding yeast and mammals. *Nucleic Acids Res* **28**: 3003-3010.

Kervestin S, Jacobson A. 2012. NMD: a multifaceted response to premature translational termination. *Nat Rev Mol Cell Biol* **13**: 700-712.

Kessler O, Jiang Y, Chasin LA. 1993. Order of intron removal during splicing of endogenous adenine phosphoribosyltransferase and dihydrofolate reductase pre-mRNA. *Mol Cell Biol* **13**: 6211-6222.

Khodor YL, Menet JS, Tolan M, Rosbash M. 2012. Cotranscriptional splicing efficiency differs dramatically between Drosophila and mouse. *RNA* **18**: 2174-2186.

Khodor YL, Rodriguez J, Abruzzi KC, Tang CH, Marr MT, 2nd, Rosbash M. 2011. Nascent-seq indicates widespread cotranscriptional pre-mRNA splicing in Drosophila. *Genes Dev* **25**: 2502-2512.

Kilchert C, Wittmann S, Passoni M, Shah S, Granneman S, Vasiljeva L. 2015. Regulation of mRNA Levels by Decay-Promoting Introns that Recruit the Exosome Specificity Factor Mmi1. *Cell Rep* **13**: 2504-2515.

Kilchert C, Wittmann S, Vasiljeva L. 2016. The regulation and functions of the nuclear RNA exosome complex. *Nat Rev Mol Cell Biol* **17**: 227-239.

Kim SW, Taggart AJ, Heintzelman C, Cygan KJ, Hull CG, Wang J, Shrestha B, Fairbrother WG. 2017. Widespread intra-dependencies in the removal of introns from human transcripts. *Nucleic Acids Res* **45**: 9503-9513.

Kuang Z, Boeke JD, Canzar S. 2017. The dynamic landscape of fission yeast meiosis alternative-splice isoforms. *Genome Res* **27**: 145-156.

Kupfer DM, Drabenstot SD, Buchanan KL, Lai H, Zhu H, Dyer DW, Roe BA, Murphy JW. 2004. Introns and splicing elements of five diverse fungi. *Eukaryot Cell* **3**: 1088-1100.

Lee Y, Rio DC. 2015. Mechanisms and Regulation of Alternative Pre-mRNA Splicing. *Annu Rev Biochem* **84**: 291-323.

Lipp JJ, Marvin MC, Shokat KM, Guthrie C. 2015. SR protein kinases promote splicing of nonconsensus introns. *Nat Struct Mol Biol* **22**: 611-617.

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

Mayer A, di Iulio J, Maleri S, Eser U, Vierstra J, Reynolds A, Sandstrom R, Stamatoyannopoulos JA, Churchman LS. 2015. Native elongating transcript sequencing reveals human transcriptional activity at nucleotide resolution. *Cell* **161**: 541-554.

Milligan L, Sayou C, Tuck A, Auchynnikava T, Reid JEA, Alexander R, Alves FL, Allshire R, Spanos C, Rappsilber J et al. 2017. RNA polymerase II stalling at pre-mRNA splice sites is enforced by ubiquitination of the catalytic subunit. *Elife* **6**.

Munding EM, Shiue L, Katzman S, Donohue JP, Ares M, Jr. 2013. Competition between pre-mRNAs for the splicing machinery drives global regulation of splicing. *Mol Cell* **51**: 338-348.

Nag A, Steitz JA. 2012. Tri-snRNP-associated proteins interact with subunits of the TRAMP and nuclear exosome complexes, linking RNA decay and pre-mRNA splicing. *RNA Biol* **9**: 334-342.

Neves LT, Douglass S, Spreafico R, Venkataramanan S, Kress TL, Johnson TL. 2017. The histone variant H2A.Z promotes efficient cotranscriptional splicing in S. cerevisiae. *Genes Dev* **31**: 702-717.

Nissen KE, Homer CM, Ryan CJ, Shales M, Krogan NJ, Patrick KL, Guthrie C. 2017. The histone variant H2A.Z promotes splicing of weak introns. *Genes Dev* **31**: 688-701.

Oesterreich FC, Herzel L, Straube K, Hujer K, Howard J, Neugebauer KM. 2016. Splicing of Nascent RNA Coincides with Intron Exit from RNA Polymerase II. *Cell* **165**: 372-381.

Pandya-Jones A, Black DL. 2009. Co-transcriptional splicing of constitutive and alternative exons. *RNA* **15**: 1896-1908.

Park SG, Hannenhalli S. 2015. First intron length in mammals is associated with 5′ exon skipping rate. *bioRxiv* doi:10.1101/024463.

Patrick KL, Ryan CJ, Xu J, Lipp JJ, Nissen KE, Roguev A, Shales M, Krogan NJ, Guthrie C. 2015. Genetic interaction mapping reveals a role for the SWI/SNF nucleosome remodeler in spliceosome activation in fission yeast. *PLoS Genet* **11**: e1005074.

Rallis C, Bahler J. 2013. Inhibition of TORC1 signaling and increased lifespan: gained in translation? *Aging (Albany NY)* **5**: 335-336.

Robberson BL, Cote GJ, Berget SM. 1990. Exon definition may facilitate splice site selection in RNAs with multiple exons. *Mol Cell Biol* **10**: 84-94.

Romfo CM, Alvarez CJ, van Heeckeren WJ, Webb CJ, Wise JA. 2000. Evidence for splice site pairing via intron definition in Schizosaccharomyces pombe. *Mol Cell Biol* **20**: 7955-7970.

Sakharkar MK, Perumal BS, Sakharkar KR, Kangueane P. 2005. An analysis on gene architecture in human and mouse genomes. *In Silico Biol* **5**: 347-365.

Saldi T, Cortazar MA, Sheridan RM, Bentley DL. 2016. Coupling of RNA Polymerase II Transcription Elongation with Pre-mRNA Splicing. *J Mol Biol* **428**: 2623-2635.

Schor IE, Gomez Acuna LI, Kornblihtt AR. 2013. Coupling between transcription and alternative splicing. *Cancer Treat Res* **158**: 1-24.

Shao W, Kim HS, Cao Y, Xu YZ, Query CC. 2012. A U1-U2 snRNP interaction network during intron definition. *Mol Cell Biol* **32**: 470-478.

Sharon D, Tilgner H, Grubert F, Snyder M. 2013. A single-molecule long-read survey of the human transcriptome. *Nat Biotechnol* **31**: 1009-1014.

Shetty A, Kallgren SP, Demel C, Maier KC, Spatt D, Alver BH, Cramer P, Park PJ, Winston F. 2017. Spt5 Plays Vital Roles in the Control of Sense and Antisense Transcription Elongation. *Mol Cell* **66**: 77-88 e75.

Sridharan V, Heimiller J, Singh R. 2011. Genomic mRNA profiling reveals compensatory mechanisms for the requirement of the essential splicing factor U2AF. *Mol Cell Biol* **31**: 652-661.

Tilgner H, Jahanbani F, Blauwkamp T, Moshrefi A, Jaeger E, Chen F, Harel I, Bustamante CD, Rasmussen M, Snyder MP. 2015. Comprehensive transcriptome analysis using synthetic long-read sequencing reveals molecular co-association of distant splicing events. *Nat Biotechnol* **33**: 736-742.

Tilgner H, Knowles DG, Johnson R, Davis CA, Chakrabortty S, Djebali S, Curado J, Snyder M, Gingeras TR, Guigo R. 2012. Deep sequencing of subcellular RNA fractions shows splicing to be predominantly co-transcriptional in the human genome but inefficient for lncRNAs. *Genome Res* **22**: 1616-1625.

Vargas DY, Shah K, Batish M, Levandoski M, Sinha S, Marras SA, Schedl P, Tyagi S. 2011. Single-molecule imaging of transcriptionally coupled and uncoupled splicing. *Cell* **147**: 1054-1065.

Wahl MC, Will CL, Luhrmann R. 2009. The spliceosome: design principles of a dynamic RNP machine. *Cell* **136**: 701-718.

Wallace EWJ, Beggs JD. 2017. Extremely fast and incredibly close: cotranscriptional splicing in budding yeast. *RNA* **23**: 601-610.

Webb CJ, Romfo CM, van Heeckeren WJ, Wise JA. 2005. Exonic splicing enhancers in fission yeast: functional conservation demonstrates an early evolutionary origin. *Genes Dev* **19**: 242-254.

Wilhelm BT, Marguerat S, Watt S, Schubert F, Wood V, Goodhead I, Penkett CJ, Rogers J, Bahler J. 2008. Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution. *Nature* **453**: 1239-1243.

Zhang MQ. 1998. Statistical features of human exons and their flanking regions. *Hum Mol Genet* **7**: 919-932.

Zhou Y, Zhu J, Schermann G, Ohle C, Bendrin K, Sugioka-Sugiyama R, Sugiyama T, Fischer T. 2015. The fission yeast MTREC complex targets CUTs and unspliced pre-mRNAs to the nuclear exosome. *Nat Commun* **6**: 7050.

Zhuang F, Fuchs RT, Sun Z, Zheng Y, Robb GB. 2012. Structural bias in T4 RNA ligase-mediated 3'-adapter ligation. *Nucleic Acids Res* **40**: e54.