Long-read sequencing of nascent RNA

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 CT 06520, 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

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. Because spliceosome assembly and catalysis occur co-transcriptionally, we hypothesized that introns are removed in the order of their transcription in genomes dominated by constitutive splicing. Remarkably little is known about splicing order and the regulatory potential of nascent transcript remodeling by splicing, due to the limitations of existing methods that focus on analysis of mature splicing products (mRNAs) rather than substrates and intermediates. Here, we overcome this obstacle through long-read RNA sequencing of nascent, multi-intron transcripts in the fission yeast *Schizosaccharomyces pombe*. Most multi-intron transcripts were fully spliced, consistent with rapid co-transcriptional splicing. However, an unexpectedly high proportion of transcripts were either fully spliced or fully unspliced, suggesting that splicing of any given intron is dependent on the splicing status of other introns in the transcript. Supporting this, mild inhibition of splicing by a temperature-sensitive mutation in Prp2, the homolog of vertebrate U2AF65, increased the frequency of fully unspliced transcripts. Importantly, fully unspliced transcripts displayed transcriptional read-through at the polyA site and were degraded co-transcriptionally by the nuclear exosome. Finally, we show that cellular mRNA levels were reduced in genes with a high number of unspliced nascent transcripts during caffeine treatment, showing regulatory significance of co-transcriptional splicing. Therefore, overall splicing of individual nascent transcripts, 3′ end formation, and mRNA half-life depend on the splicing status of neighboring introns, suggesting crosstalk among spliceosomes and the polyA cleavage machinery 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 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 influence 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; Wong et al. 2016). 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, intron definition 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 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 (Carrillo 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 due to a lack of methods that detect the order of intron removal.

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 *Schizosaccharomyces pombe*, where >1200 protein-coding genes have

multiple introns. The *S. pombe* splicing machinery is more similar to higher eukaryotes than to *S. cerevisiae* (Kaufer and Potashkin 2000), and alternative splicing occurs particularly in meiosis (Averbeck et al. 2005; Awan et al. 2013; Bitton et al. 2014; Bitton et al. 2015; Kilchert et al. 2015; Kuang et al. 2017). Splicing rates in *S. pombe* are similar to those reported in *S. cerevisiae* (Alexander et al. 2010; Barrass et al. 2015; Carrillo Oesterreich et al. 2016; Eser et al. 2016). *S. pombe* introns are very short with a median length of 56 nt, whereas exon lengths are similar to 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 *S. pombe* (Romfo et al. 2000; Shao et al. 2012; Fair and Pleiss 2017), suggesting that splicing of individual introns in multi-intron transcripts may occur independently of one other. To globally investigate the progression of splicing during transcription, nascent RNA (nRNA) prepared from proliferating *S. pombe* cells was analyzed by both short- and long-read sequencing (LRS). We report an unexpected coordination between splicing of neighboring introns and polyA site cleavage.

Results

Proteomic and transcriptomic characterization of *S. pombe* chromatin

To isolate nascent RNA, *S. pombe* chromatin was purified and analyzed for RNA and protein composition (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 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 rRNA in the chromatin fraction. Mass spectrometry 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 on chromatin (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 RNA and rRNA, using bead-based negative selection with oligo-dT and the RiboZero kit (Fig. 1A, S1). 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. To validate our splicing quantification, 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.

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 might be 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). Consistent with the earlier observation of lower first intron splicing, 176 first introns were spliced less (52% of 5′ less group) and only 73 were spliced more (27% of 5′ more group) than the second intron. Next, we compared intron features between the more and less frequently spliced introns in the pairs with significant splicing differences. The 42 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 S6, S9 for a complete overview of intron features). Fifteen features, e.g. cytoplasmic mRNA-seq intron coverage and 5'SS and BPS strength, were significantly associated with low or high splicing 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 less co-transcriptionally spliced than internal introns.

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, Fig. S4C-G, Table S4). 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 transcript sequence informs on the history of transcription and splicing of every RNA molecule. Nine examples from 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 to well-defined positions downstream of the existing annotation (e.g. SPAC16E8.15). 3′ ends correspond to the nucleotide position of elongating Pol II at the time of cell lysis. As seen previously for single intron nascent RNAs, introns were often spliced out when Pol II was positioned just downstream of the intron (Carrillo Oesterreich et al. 2016) (Fig. 3A, S4H). To evaluate the order of intron removal, it was 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 recapitulated 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 as '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 contained 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.

"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. 4A), 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. 4A).

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. For 99 of those genes the observed fraction of "all or none" splicing was greater than the predicted fraction (Fig. 4B). To estimate the degree of co-association between

individual splicing events, a co-association score was calculated, which we defined as the log2-fold change of observed and predicted fraction of "all or none" transcripts (Fig. 4C, S5E). Note that the maximal co-association score increases with increasing intron number (Fig. S5F). Most genes had co-association scores close to the maximal possible score (Fig. 4C, S5G). Hence, we conclude that partial transcript splicing is the exception in *S. pombe* and suggest that most introns are removed cooperatively. Interestingly, the number of introns in the transcript does not influence the gene-by-gene tendency for co-association (Fig. S5G). Here, the combined analysis of short- and long-read nascent RNA sequencing data enabled us to distinguish between completely spliced mRNAs, mRNAs with retained introns, or completely unspliced RNAs. Our data indicate that splicing in *S. pombe* occurs soon after introns are fully transcribed and may stimulate splicing of further downstream introns, resulting in "all or none" splicing of 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 neighboring introns within the same transcript. To test this hypothesis, we profiled nascent RNA splicing 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. 5A, S6A). After a two-hour shift to the non-permissive temperature, a ~50% reduction in pre-mRNA splicing was observed for five analyzed introns in *prp2-1* relative to WT in both cytoplasmic and chromatin fractions (Fig. 5A, S6A-B). LRS was performed on nRNA isolated from *prp2-1* and WT cells in three biological replicates. Similar to the RT-PCR experiment, splicing levels were reduced by half in *prp2-1* (Fig. 5B-C). Global differences in transcription of intron-containing genes were not detected

between the two stains (Fig. S6E). Classification of transcripts into "all unspliced", "all spliced" and "partially spliced" transcripts revealed that the proportion of "all unspliced" transcripts increased from 40 to 59% in *prp2-1* and had only a minor impact on the fraction of "partially spliced" transcripts (Fig. 5B, D). This is consistent with an inhibitory effect of individual unspliced introns on the splicing of the other introns in the same transcript. This 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). Finally, we asked whether the *prp2-1* mutation affected mRNA levels. 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. 5E). 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.

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. 6A).

Next, we sought to determine the contribution of splicing inhibition to physiologically relevant cell signaling and shifting gene expression programs. To alter mRNA levels globally, exponentially growing *S. pombe* cells were treated with 10 mM caffeine for 15 min (Fig. 6B), which elicits a cellular response similar to nitrogen starvation (Rallis and Bahler 2013). Gene expression and splicing profiles were quantified by nRNA-seq and cytoplasmic mRNA-seq for three biological replicates. In agreement with a previous study, ~1,500 mRNAs changed levels significantly after treatment (Fig. 6C, 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. 6D), indicating that failure to splice co-transcriptionally leads to lower gene expression.

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

Strong links between splicing and polyA site cleavage (Herzel et al. 2017) led us to determine 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. 7A-C). The majority of intron-containing transcripts that had a 3′ end after the annotated polyA cleavage site were completely unspliced (Fig. 7B). Nine out of ten of these instances were validated by RT-PCR (Fig. S8A). Conversely, 95% of polyadenylated transcripts were spliced (Fig. S8D). Given our previous evidence

that unspliced transcripts are preferentially degraded (see above), we considered it likely that unspliced transcripts extending over the polyA site 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. 7D, 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. 7E). 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.

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 nascent 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; Herzel and Neugebauer 2015). It is generally assumed that remaining introns are post-transcriptionally spliced. Instead, our findings indicate that, at least in *S. pombe*, the majority of introns that are not spliced co-transcriptionally will be degraded. We demonstrate that degradation of unspliced RNA contributes significantly to shaping the transcriptome in response to

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 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 *S. cerevisiae* chromatin (Carrillo Oesterreich et al. 2010). This likely reflects the 10x higher number of intron-containing genes in the *S. pombe* genome compared to *S. cerevisiae* and is consistent with co-transcriptional splicing and/or the retention of post-splicing intron lariat spliceosomes on chromatin (Chen et al. 2014). Analysis of splicing frequency on a per intron basis from nascent RNA-seq showed that most *S. pombe* introns are spliced co-transcriptionally (median 0.59), which is lower than in *S. cerevisiae* (median 0.74), *Drosophila* or human cells (Brugiolo et al. 2013). Notably, 5% of mRNAs contain introns; recent studies suggest that translation of retained introns contributes to protein diversity in *S. pombe* (Duncan and Mata 2014; Kuang et al. 2017). Intron length is often associated with reduced co-transcriptional splicing (Carrillo Oesterreich et al. 2010; Brugiolo et al. 2013). However, we (see Figure 2A and S2C) and others (Eser et al. 2016) find that intron length is one of the few parameters that does *not* correlate with nascent RNA splicing frequency, while 5'SS and BPS strength do. Low co-transcriptional splicing of first introns in *S. pombe* and other species suggests that the first intron can persist while downstream introns are spliced (Kessler et al. 1993; Khodor et al. 2011; Khodor et al. 2012; Tilgner et al. 2012).

Full-length nascent transcript sequences reveal the life history of individual (pre-)mRNAs that cannot be inferred from conventional short-read sequencing. 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; Tilgner et al. 2018) 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 mRNA splicing products (Sharon et al. 2013; Kuang et al. 2017). Here, we used it to analyze splicing substrates: full-length nascent transcripts. The mapping of nascent RNA 3′ ends reports on Pol II position within genes. 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). Although higher coverage will be required to investigate Pol II dynamics, LRS of nascent RNA represents an additional powerful approach for investigating RNA processing and transcription regulation.

Full length sequencing of individual nascent RNAs enabled us to determine the fate of unspliced transcripts. Most often, unspliced nascent transcripts fail to undergo polyA site cleavage, leading to read-through transcription and RNA degradation by the nuclear exosome. Therefore, our data complement prior reports showing that unspliced transcripts are degraded in an *rrp6*-dependent manner (Bousquet-Antonelli et al. 2000; Bitton et al. 2015; Zhou et al. 2015; Kilchert et al. 2016). Our work does not exclude the action of additional degradation pathways for unspliced transcripts, like the nuclear 5′-3′ exonuclease Dhp1/Rat1/Xrn2 and cytoplasmic nonsense mediated decay (Vargas et al. 2011; Davidson et al. 2012; Girard et al. 2012; Kervestin and Jacobson 2012; Brogna et al. 2016; Kilchert et al. 2016). Physical interactions have been detected between spliceosomal components and the nuclear decay machinery to provide a layer of surveillance based on splicing fidelity (Nag and Steitz 2012; Zhou et al. 2015; Herzel et al. 2017). Our findings suggest that these interactions take place co-transcriptionally, explaining why poor co-transcriptional splicing is associated with lowered mRNA levels independent of transcriptional changes upon caffeine treatment (this study) and possibly why splicing levels within mRNA correlate with expression levels (Wilhelm et al. 2008). Cooperativity could drive the formation of fully spliced mRNAs over partially spliced transcripts to evade decay in the nucleus or cytoplasm. Also, increased mRNA half-life was observed *in vitro* when splicing occurred in a transcription-coupled system (Hicks et al. 2006). Conversely, fully unspliced transcripts might be degraded more efficiently. Perhaps

these transcripts arise during the process of shutting down gene expression. For example, read-through transcription and inhibition of splicing can occur during mitosis, cellular senescence, and/or under conditions of cellular stress (Leong et al. 2014; Shalgi et al. 2014; Vilborg et al. 2015; Muniz et al. 2017), though read-through transcription has not yet been linked to splicing inhibition in these cases. Taken together, the link between co-transcriptional splicing and mRNA half-life is an important determinant of gene expression.

Is cooperativity among splicing events widespread? We found that a single spliced or unspliced intron can mediate positive and negative effects on the splicing of other introns within the transcript. Our results *in vivo* are completely consistent with previous *in vitro* experiments showing that splicing of multi-intron transcripts is more efficient following an initial intron removal event (Crabb et al. 2010). In *S. cerevisiae*, constitutive splicing of the 2-intron *SUS1* gene has been found to occur cooperatively (Bonde et al. 2014). Co-association, the likelihood that distinct splicing events are correlated, has been detected for multiple alternative splicing events in human cells, *C. elegans* and *S. pombe* (Fededa et al. 2005; Glauser et al. 2011; Tilgner et al. 2015; Kuang et al. 2017; Tilgner et al. 2018). Importantly, a recent study concluded that splicing is regulated by adjacent processing events, based on deduced order of intron removal from human paired end RNA-seq (Kim et al. 2017). In the context of alternative splicing, intron removal can be delayed relative to an intron downstream of a cassette exon (Kim et al. 2017); interestingly, these delayed introns can already be committed to splicing (de la Mata et al. 2010). Thus, this under-appreciated phenomenon of splicing co-association has been observed in multiple species in total RNA. Here, it is extended to full-length nascent RNA, detecting intermediate states in RNA processing before and after splicing and enabling further mechanistic insights into how crosstalk is mediated.

Cooperativity among splicing events and with polyA cleavage likely occurs at the level of assembling spliceosomes. 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 at non-saturating levels or in the context of competition with RNA decay (Crabb et al. 2010; Munding et al. 2013). The cellular concentration of polyadenylation factors has also been linked to coupled alternative polyadenylation and splicing of the terminal intron (Movassat et al. 2016). Additional features contributing to cooperativity may include the local chromatin environment and/or changes in Pol II elongation and modifications (Gunderson and Johnson 2009; Gunderson et al. 2011; Patrick et al. 2015; Milligan et al. 2017; Neves et al. 2017; Nissen et al. 2017). An EM tomography study of the Balbiani ring 3 gene containing 38 introns suggested that only one spliceosome may associate with each nascent RNA at any given time (Wetterberg et al. 2001). While this remains to be tested in other systems, spliceosomal components can remain associated with multi-intron pre-mRNAs during splicing (Crabb et al. 2010; de la Mata et al. 2010). This would be consistent with the possibility that the presence of pre-spliceosomal components, assembling spliceosomes, or remaining spliceosomal components after splicing may enhance or antagonize cooperativity among introns (Huang et al. 2002; Crabb et al. 2010; Chen et al. 2014). In mammalian cells, U2AF interacts directly with the polyA cleavage machinery (Vagner et al. 2000; Millevoi et al. 2006), while U1 snRNP interactions with nascent RNA – particularly in first introns – suppress premature polyA cleavage, prevent polyadenylation, and enhance transcription elongation through the gene's entirety (Gunderson et al. 1998; Berg et al. 2012; Park and Hannenhalli 2015; Oh et al. 2017; Chiu et al. 2018). Perhaps U1 snRNP lingering on unspliced introns can promote elongation and inhibit cleavage at correct polyA sites as well, producing the observed read-through transcription. The overrepresentation of unspliced first introns in all organisms tested leads us to speculate that splicing of the first intron represents a tipping point in a transcript-specific decision of whether to proceed with transcription, splicing, and 3′ end formation.

Methods

S. pombe strains, genome version and annotation, growth conditions

For strain information please refer to Table S1. The *S. pombe* genome version EF2 (http://goo.gl/PuaBni) and gene annotation ASM294v2.31 were used (Eser et al. 2016) (Table S5). *S. pombe* cell cultures were grown in complete liquid media (YES) at 30°C and 250 rpm. The *Δrrp6* strain was grown under G418 selection. Cells were harvested by centrifugation in exponential growth at OD (595nm) ~0.5. Cell pellets were washed with ice-cold PBS, quick-frozen in liquid nitrogen in 6 aliquots and kept at -80°C. For *prp2-1* splicing analysis YES cultures originating from single *prp2-1* and 972h cell colonies were grown at 26°C to an OD of 0.4, then shifted to 37°C for 2 hours before harvesting.

Protein analysis

For western blotting, ~3 µg of protein were loaded per lane. The following antibodies were used: Rpb1 (8WG16), Histone H3 (ab1791, Abcam), GAPDH (Novus Biologicals, NB300-221) and ribosomal protein L5 (Santa Cruz, sc-103865). Coomassie gel staining and mass spectrometry were performed as described (Shevchenko et al. 2006) at the mass spectrometry facility of MPI-CBG Dresden (Supplemental methods).

RNA preparation

Nascent RNA was prepared as described (Carrillo Oesterreich et al. 2016) and used for nascent RNA sequencing and long-read sequencing (Supplemental methods). PolyA-RNA was obtained using oligo-dT coated magnetic beads binding to polyA+ RNA (Dynabeads mRNA DIRECT Micro Purification Kit, Life technologies). rRNA was removed from ~5 µg chromatin-associated polyA- RNA using the Ribo-Zero Gold rRNA Removal Kit (Yeast) from Epicentre/Illumina.

Qualitative and quantitative analysis of nucleic acids

RNA and DNA samples were analyzed by agarose (1-1.5%) or TBE-Urea polyacrylamide (10 or 15%, Invitrogen) gel electrophoresis. DNA and RNA concentrations were determined by UV/Vis spectroscopy with the Nanodrop2000 (ThermoScientific) or fluorometric measurements with the Qubit dsDNA BR Assay or the RNA BR Assay (Life technologies). Reverse transcription of RNA was done using SuperScript III reverse transcriptase (Invitrogen) with random hexamers, oligo(dT) or gene-specific primers. For details on RT, qPCR and PCR protocols see the Supplemental methods.

Short-read RNA sequencing

For RNA-seq of different cellular fractions RNA samples were submitted to the Yale Center for Genome Analysis (YCGA). PolyA+ RNA depleted, rRNA depleted nascent RNA and cytoplasmic polyA+ RNA were analyzed by single-end RNA-seq with 76bp read length. Random hexamer primed Illumina libraries were prepared with standard protocols.

Mapping short-read RNA-seq data to the *S. pombe* genome

Data were mapped to the genome with Tophat2 (version 2.0.12) (Kim et al. 2013) using the following settings: tophat2 -p 5 -i 30 -I 900 –g 1 -N 2 -G <Spombe_EF2> –segment-length 25 –library-type fr-firststrand –minanchor-length 8 –splice-mismatches 0 –min-coverage-intron 30 –max-coverageintron 900 –min-segment-intron 30 –max-segment-intron 900. RNA-seq coverage was adjusted for library size, excluding reads mapped to rDNA regions, weighted using the R-package edgeR (calcNormFactors without additional parameters, further details in Supplemental methods). To determine gene expression values between replicates and different samples cufflinks version 2.2.1 and cuffdiff were used with the following settings: cufflinks -p 20 -G <Spombe_EF2> -b <Spombe_EF2.fasta>; cuffdiff –frag-bias-correct <Spombe_EF2.fasta> –num-threads 20 –library-type fr-firststrand –library-normmethod geometric <Spombe_EF2.gtf>.

Intron splicing quantification

For intron splicing calculation junction reads originating from spliced (split, cigar contains "N" and unspliced (unsplit, cigar contains "M", but no "N") transcripts were extracted from all mapped reads. Overlaps with 8 nucleotide windows around annotated 5′ SS and 3′ SS junctions were searched for split and unsplit reads separately with bedtools intersect. An overlap of at least 4 nucleotides on each side of the junction was required. The sum of all split reads per 5′SS (identical with 3′SS split read count) was divided by the sum of the 5′SS split read count and the average of unsplit reads over 5′ and 3′SS per intron. This resulted in a splicing score (splicing per intron, SPI) ranging from 0 to 1 with 1 being 100% spliced (Herzel and Neugebauer 2015). A cutoff of at least 10 reads per junction to report an SPI was applied.

3′ end ligation and long-read sequencing library preparation

3′ end ligation and long-read sequencing library preparation of polyA+ and rRNA-depleted, chromatin-associated RNA was done as described previously (Carrillo Oesterreich et al. 2016). Sample-specific barcodes were included in the PCR oligonucleotides (Table S2) for the 3 *prp2-1* cDNA libraries and the corresponding 972h- cDNA libraries. Double-stranded cDNA (>1 μg) was submitted to the Yale Center for Genomic Analysis (YCGA) for Pacific Biosciences library preparation and sequencing with standard protocols (SMRTbell Template Prep Kit 1.0). Sequencing was done with either diffusion- or magbead-loading (further details in Supplemental methods).

Long-read sequencing data processing and mapping

3′ end linker sequences, Clontech adaptor sequences (SMARTer cDNA synthesis kit, Clontech) and the 5 nt random 3′ barcode were removed with cutadapt (Martin 2011) and the FASTX toolkit (http://hannonlab.cshl.edu/fastx_toolkit/index.html). Processed reads were mapped to the respective genome using gmap (Wu and Watanabe 2005). To remove potential mRNA contaminants reads ending within +/-100 nt of an annotated

polyA site and short polyA tails (> 4 nt) were removed from the dataset.

Classification of multi-intron transcripts

Reads overlapping completely with multiple introns were considered multi-intron transcripts. If the number of exons in a read was identical to the number of overlapping introns +1, a read was considered to be "all spliced". If the block count was 1 irrespective of the number of overlapping introns, a read was considered to be "all unspliced". Reads, which had a block count greater 1, but fewer block counts than the number of overlapping introns +1, were classified as "partially spliced" (further details in Supplemental methods).

Prediction of transcript splicing status from nRNA-seq

To determine the expectation of the abundance of partially spliced transcripts, in the case that individual intron splicing is independent of each other, we assume that the SPI, calculated from short read RNA-seq, reflects the probability of splicing of a particular intron in the chromatin fraction at steady state. Further, we assume that the gene expression values from nRNA-seq reflect an estimate of relative transcript number per gene in the chromatin fraction at steady state. For each intron 'spliced' or 'unspliced' sampling with replacement was performed according to its SPI and the average transcript number per gene. Simulated data were combined into a matrix with columns representing each intron and rows representing each transcript. If a row contained both 'spliced' and 'unspliced' column entries, it was classified as 'partially spliced'. If a row contained only 'spliced' or only 'unspliced' column entries, it was classified as 'all spliced' or 'all unspliced', respectively. In the gene-specific analysis a numerical solution using SPIs as splicing probability was calculated (Supplemental methods). From the gene-specific observed and predicted fraction of "all or none" splicing we calculated a co-association score, which we define as the log2-ratio between observed and predicted fraction of "all or none" splicing.

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 from Harvard Dataverse doi:10.7910/DVN/PW1KEG (Herzel 2017).

Acknowledgements

We thank Gene-Wei Li and members of the Neugebauer and Li labs, Hanspeter Herzel, Michael Weber, Fernando Carrillo Oesterreich and Charles Query for discussions and comments on the manuscript. We are grateful to Iva Tolic, Tamas Fischer and Charles Query for gifts of strains, Jeremy Schofield for help with preliminary experiments, and Guilin Wang and the Yale Center for Genome Analysis for technical assistance and advice. The 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**.

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.

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, 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.

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.

Chiu AC, Suzuki HI, Wu X, Mahat DB, Kriz AJ, Sharp PA. 2018. Transcriptional Pause Sites Delineate Stable Nucleosome-Associated Premature Polyadenylation Suppressed by U1 snRNP. *Mol Cell* **69**: 648-663 e647.

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.

Duncan CD, Mata J. 2014. The translational landscape of fission-yeast meiosis and sporulation. *Nat Struct Mol Biol* **21**: 641-647.

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.

Gunderson SI, Polycarpou-Schwarz M, Mattaj IW. 1998. U1 snRNP inhibits pre-mRNA polyadenylation through a direct interaction between U1 70K and poly(A) polymerase. *Mol Cell* **1**: 255-264.

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

Herzel L. 2017. Long-read nascent RNA sequencing from S. pombe wild-type and prp2-1 strains. doi:10.7910/DVN/PW1KEG.

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 D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. 2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* **14**: R36.

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.

Leong HS, Dawson K, Wirth C, Li Y, Connolly Y, Smith DL, Wilkinson CR, Miller CJ. 2014. A global non-coding RNA system modulates fission yeast protein levels in response to stress. *Nat Commun* **5**: 3947.

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.

Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *2011* **17**.

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.

Millevoi S, Loulergue C, Dettwiler S, Karaa SZ, Keller W, Antoniou M, Vagner S. 2006. An interaction between U2AF 65 and CF I(m) links the splicing and 3' end processing machineries. *EMBO J* **25**: 4854-4864.

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**.

Movassat M, Crabb TL, Busch A, Yao C, Reynolds DJ, Shi Y, Hertel KJ. 2016. Coupling between alternative polyadenylation and alternative splicing is limited to terminal introns. *RNA Biol* **13**: 646-655.

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.

Muniz L, Deb MK, Aguirrebengoa M, Lazorthes S, Trouche D, Nicolas E. 2017. Control of Gene Expression in Senescence through Transcriptional Read-Through of Convergent Protein-Coding Genes. *Cell Rep* **21**: 2433-2446.

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.

Oh JM, Di C, Venters CC, Guo J, Arai C, So BR, Pinto AM, Zhang Z, Wan L, Younis I et al. 2017. U1 snRNP telescripting regulates a size-function-stratified human genome. *Nat Struct Mol Biol* **24**: 993-999.

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.

Shalgi R, Hurt JA, Lindquist S, Burge CB. 2014. Widespread inhibition of posttranscriptional splicing shapes the cellular transcriptome following heat shock. *Cell Rep* **7**: 1362-1370.

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.

Shevchenko A, Tomas H, Havlis J, Olsen JV, Mann M. 2006. In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat Protoc* **1**: 2856-2860.

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, Jahanbani F, Gupta I, Collier P, Wei E, Rasmussen M, Snyder M. 2018. Microfluidic isoform sequencing shows widespread splicing coordination in the human transcriptome. *Genome Res* **28**: 231-242.

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.

Vagner S, Vagner C, Mattaj IW. 2000. The carboxyl terminus of vertebrate poly(A) polymerase interacts with U2AF 65 to couple 3'-end processing and splicing. *Genes Dev* **14**: 403-413.

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.

Vilborg A, Passarelli MC, Yario TA, Tycowski KT, Steitz JA. 2015. Widespread Inducible Transcription Downstream of Human Genes. *Mol Cell* **59**: 449-461.

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.

Wetterberg I, Zhao J, Masich S, Wieslander L, Skoglund U. 2001. In situ transcription and splicing in the Balbiani ring 3 gene. *EMBO J* **20**: 2564-2574.

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.

Wong JJ, Au AY, Ritchie W, Rasko JE. 2016. Intron retention in mRNA: No longer nonsense: Known and putative roles of intron retention in normal and disease biology. *Bioessays* **38**: 41-49.

Wu TD, Watanabe CK. 2005. GMAP: a genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics* **21**: 1859-1875.

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.

Table and figure legends

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).

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 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 (Herzel and Neugebauer 2015). 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). The inset shows schematizes how the SPI is calculated from the number of 'spliced' and 'unspliced' junction reads spanning a particular intron.

Figure 2: Pre-mRNA features that 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) 15/42 analyzed gene architecture features correlate significantly with differentially spliced intron pairs (sequence-based in black font and RNA-seq derived in grey font). The smaller intron position for "low" spliced introns in a pair (first intron – 1, second intron – 2 etc.) is consistent with enrichment of first introns in the "5' less spliced" group. The median modified Z-score is shown for each feature with significant difference between the "low" and "high" groups and the respective negative log10 of the Bonferroni-corrected p-value is given. For two features, no change in the median modified Z-score is visible. The respective feature distribution is thus presented as boxplot below. 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 & D.

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). Thin lines in transcripts indicate that the intron sequence is absent (due to splicing). Transcripts overlapping with <2 introns ("others", grey) 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) All unspliced, all spliced and partially spliced fraction in the transcriptome. The large pie chart shows the 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).

Figure 4: High degree of splicing co-association in multi-intron transcripts. (A) Predicted and observed "all unspliced", "all spliced" and "partially spliced" fractions in the transcriptome. Left panel, prediction of splicing categories from nRNA-seq, assuming independence of splicing of adjacent introns. Predicted values were calculated 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 as in 3B. Right panel, proportion of the three splicing categories from total mRNA LRS (data from (Kuang et al. 2017)). (B) Predominant "all or none" splicing in multi-intron genes. The predicted fraction of "all or none" splicing (fraction of all spliced or unspliced transcripts calculated from nRNA-seq, assuming splicing independence) is plotted against the observed fraction of "all or none" splicing for 100 genes with 10 or more sequenced transcripts by LRS. (C) Co-association of splicing is close to the maximal possible co-association value. A co-association score was calculated as the log2-fold change of the observed (LRS) to predicted (RNA-seq) fraction of "all or none" splicing. The maximal possible co-association score for 2, 3 and 4-intron genes was calculated (log2(1/gene splicing)) and plotted versus the mean gene splicing value (solid line). Co-association scores for individual genes fall below or on top of this line, suggesting maximal co-association for most genes (median 86%, calculated as ratio of observed co-association score over maximal co-association score at particular gene

splicing). Outliers are indicated as lighter points and can be explained by inconsistencies in gene annotation.

Figure 5: 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 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).

Figure 6: Co-transcriptional splicing correlates with higher mRNA levels. (A) 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) 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).

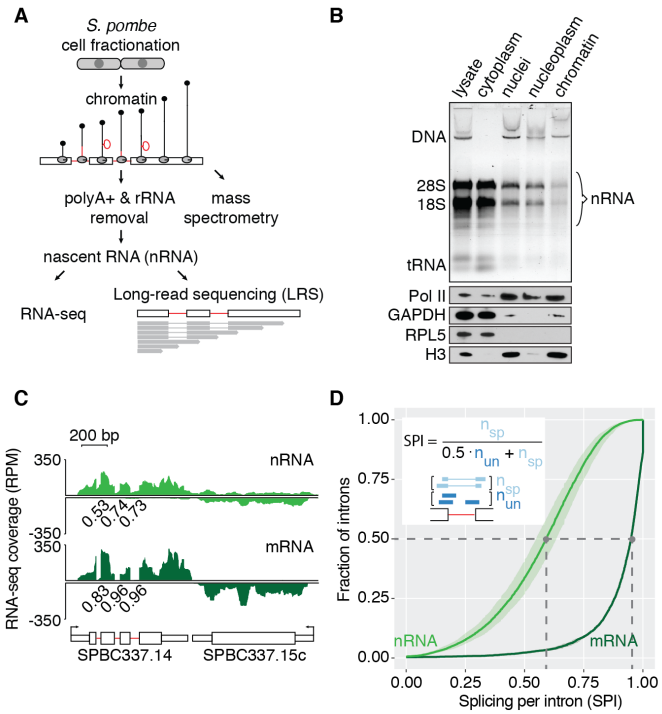
Figure 7: Coupling between co-transcriptional splicing, polyA site cleavage, and mRNA stability. (A) Examples of intron-containing genes with unspliced transcripts extending over the polyA site. Black triangles mark the polyA cleavage sites. Full representation of all sequenced transcripts in Fig. S8B. (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 ($p < 0.05$ *, $p < 0.01$ **, $p < 0.001$ ***).

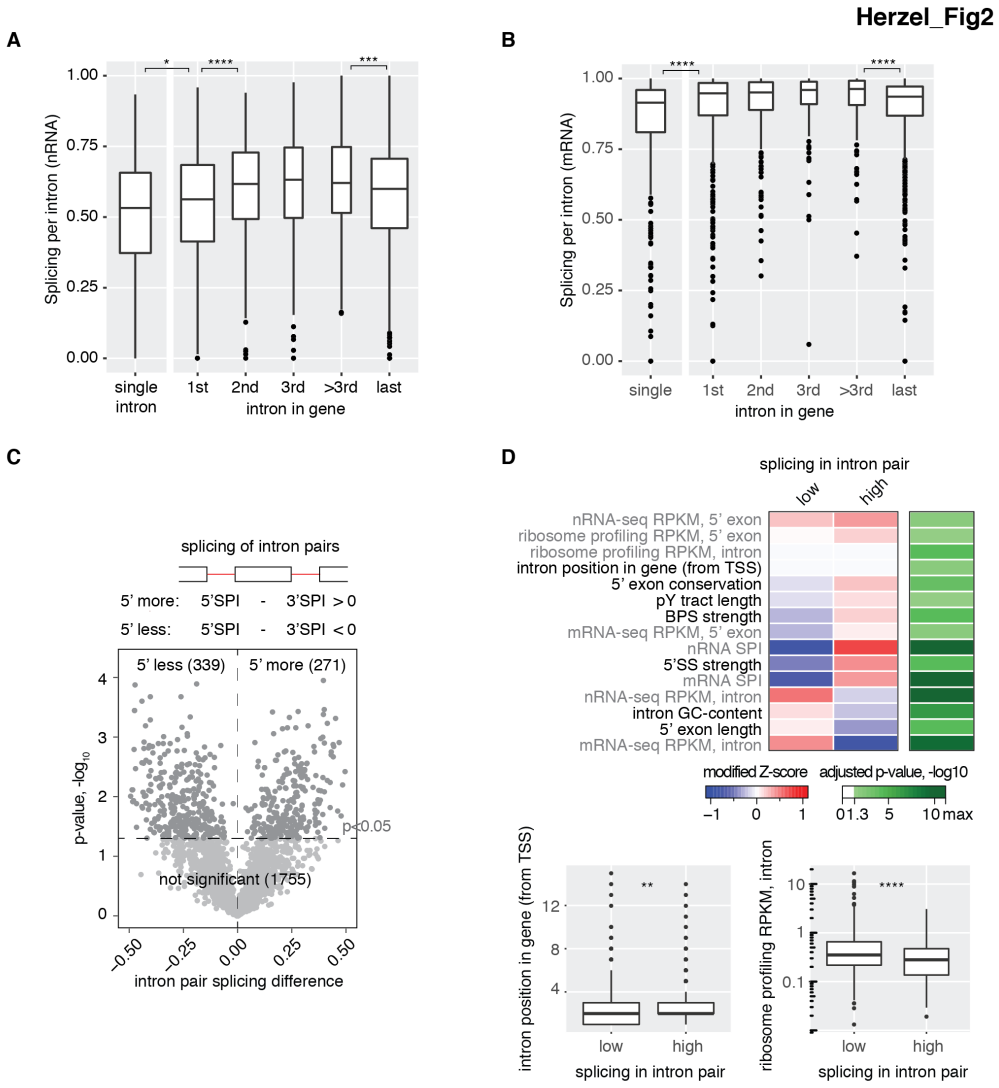
Table and figures

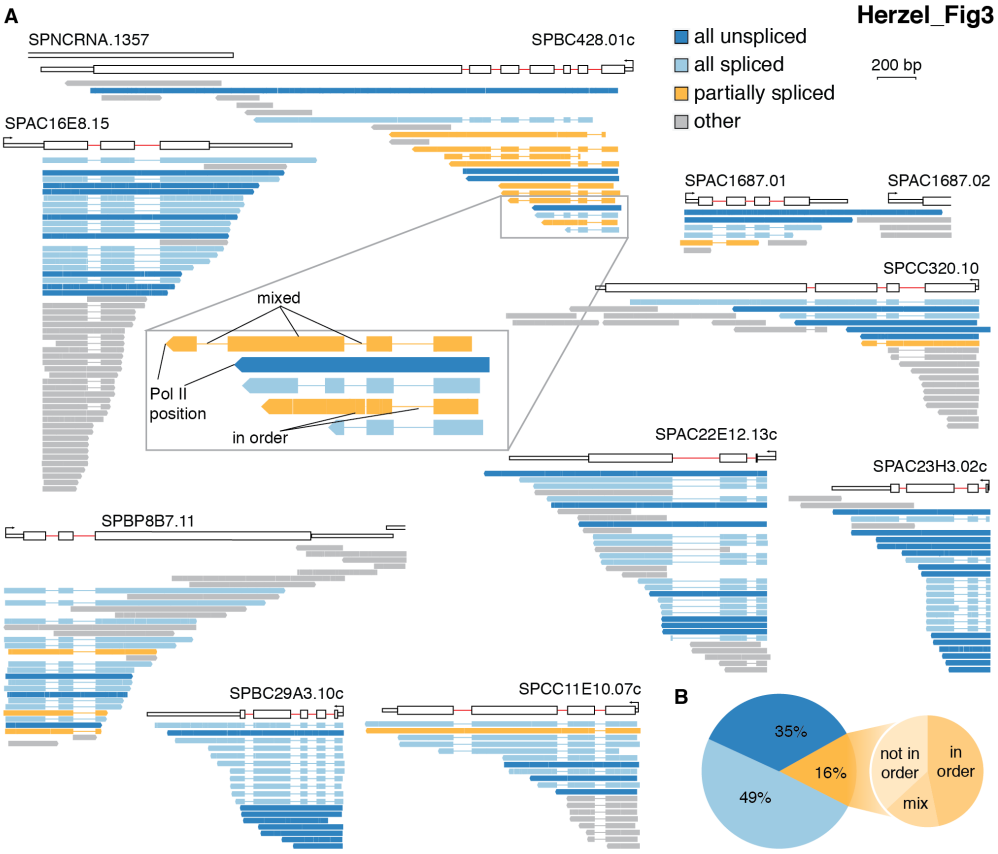
Table 1

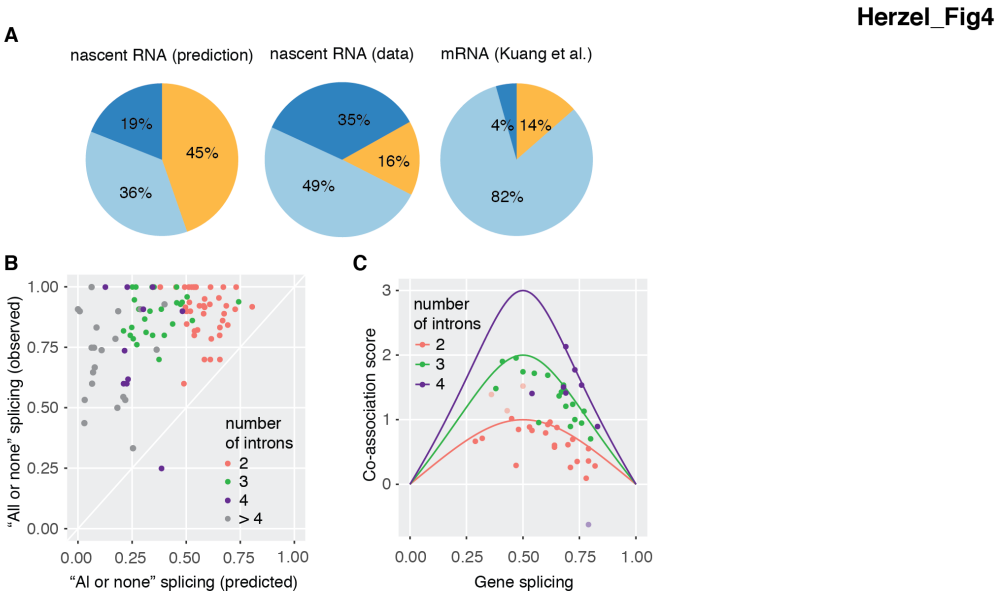
		S. cerevisiae	*S. pombe*
RNA polymerases	Pol I	12/14	10/14
	Pol II	10/12	8/12
	Pol III	12/15	7/18
Splicing factors	U1	-	3/9
	U2	-	6/17
	Prp19	-	9/14
	U5	-	2/6
	Sm-Ring	-	6/7
	SR-like	-	1/2
Nucleosomes	Histones	6/6	4/6
Chromatin remodelers	RSC	14/14	12/13
	yINO80	10/14	11/14
	SWI/SNF	8/12	7/12
	ISW1	4/4	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

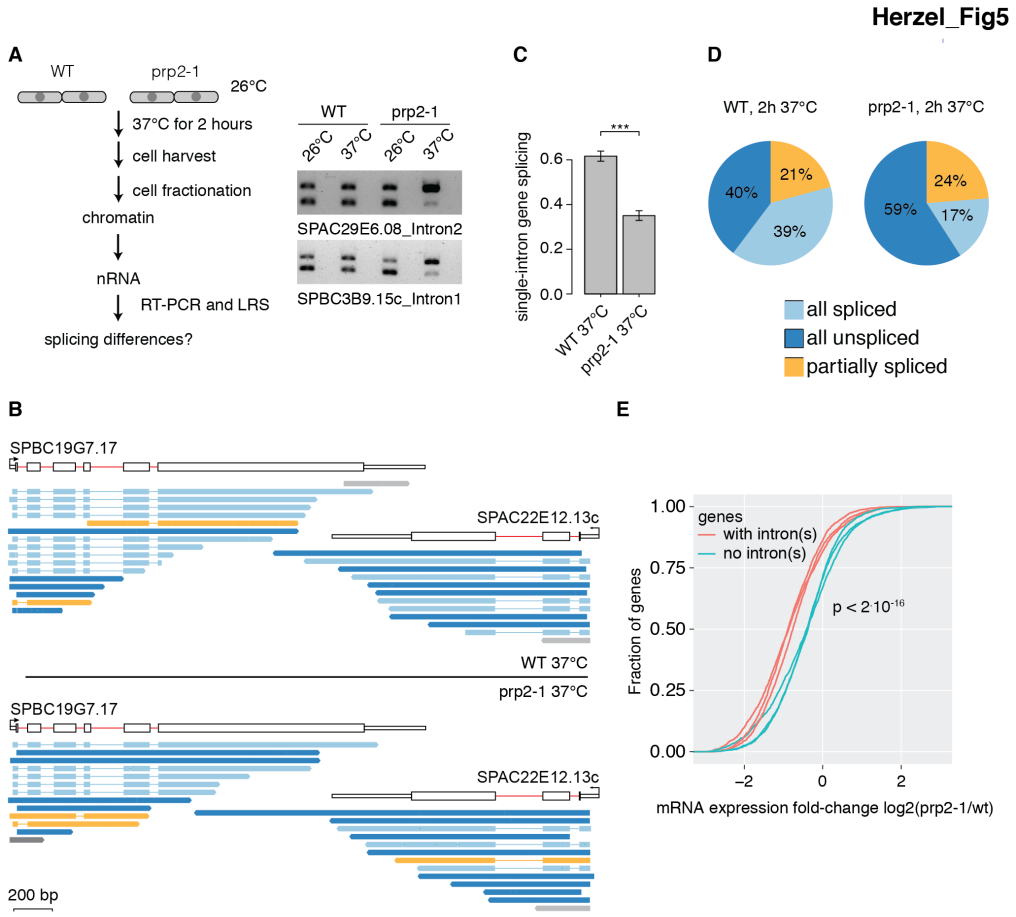
Herzel_Fig1


A
S. pombe
cell fractionation
chromatin
polyA+ & rRNA
removal
mass
spectrometry
nascent RNA (nRNA)
RNA-seq
Long-read sequencing (LRS)
B
lysate
cytoplasm
nuclei
nucleoplasm
chromatin
DNA
28S
18S
nRNA
tRNA
Pol II
GAPDH
RPL5
H3
C
200 bp
RNA-seq coverage (RPM)
350
-350
nRNA
0.53
0.74
0.73
mRNA
0.83
0.96
0.96
SPBC337.14
SPBC337.15c
D
Fraction of introns
1.00
0.75
0.50
0.25
0.00
SPI = n_sp / (0.5 · n_un + n_sp)
n_sp
n_un
nRNA
mRNA
0.00
0.25
0.50
0.75
1.00
Splicing per intron (SPI)

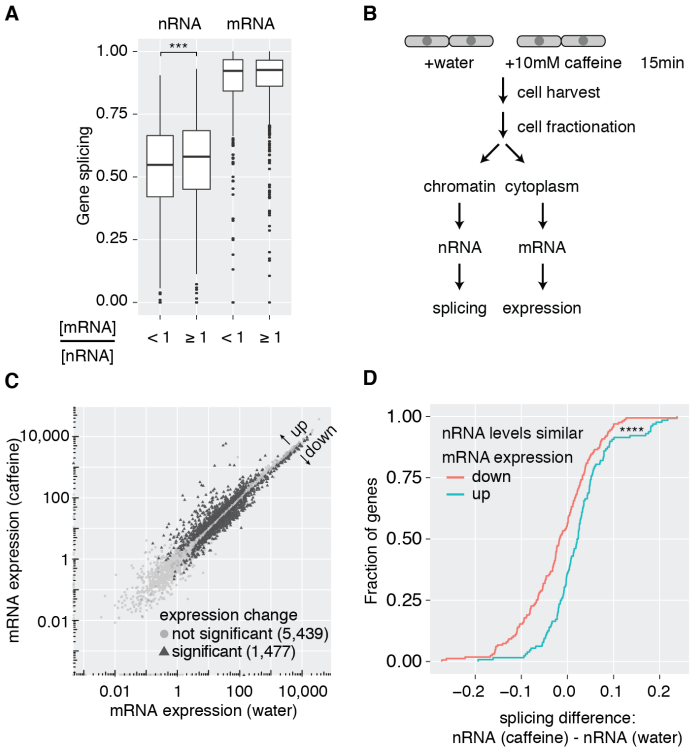

Herzel_Fig2
A
Splicing per intron (nRNA)
1.00
0.75
0.50
0.25
0.00
*
****
***
single intron
1st
2nd
3rd
>3rd
last
intron in gene
B
Splicing per intron (mRNA)
****
****
single
1st
2nd
3rd
>3rd
last
intron in gene
C
splicing of intron pairs
5' more: 5'SPI - 3'SPI > 0
5' less: 5'SPI - 3'SPI < 0
5' less (339)
5' more (271)
p-value, -log10
p<0.05
not significant (1755)
-0.50
-0.25
0.00
0.25
0.50
intron pair splicing difference
D
splicing in intron pair
low
high
nRNA-seq RPKM, 5' exon
ribosome profiling RPKM, 5' exon
ribosome profiling RPKM, intron
intron position in gene (from TSS)
5' exon conservation
pY tract length
BPS strength
mRNA-seq RPKM, 5' exon
nRNA SPI
5'SS strength
mRNA SPI
nRNA-seq RPKM, intron
intron GC-content
5' exon length
mRNA-seq RPKM, intron
modified Z-score
adjusted p-value, -log10
–1
0
1
01.3
5
10 max
intron position in gene (from TSS)
**
low
high
splicing in intron pair
ribosome profiling RPKM, intron
****
low
high
splicing in intron pair

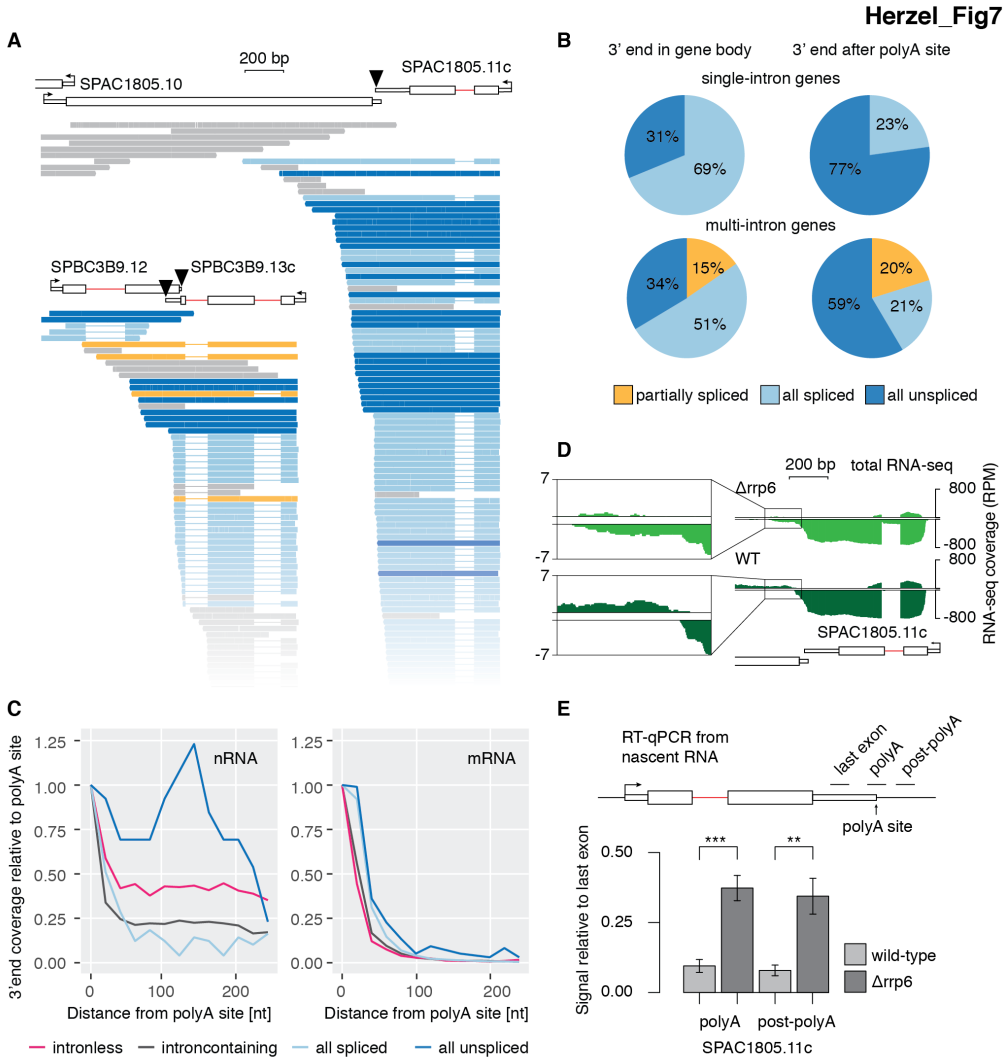

Herzel_Fig3
A
SPNCRNA.1357
SPBC428.01c
all unspliced
all spliced
partially spliced
other
200 bp
SPAC16E8.15
SPAC1687.01
SPAC1687.02
SPCC320.10
mixed
Pol II
position
in order
SPAC22E12.13c
SPAC23H3.02c
SPBP8B7.11
SPBC29A3.10c
SPCC11E10.07c
B
35%
16%
49%
not in
order
in
order
mix


Herzel_Fig4
A
nascent RNA (prediction)
nascent RNA (data)
mRNA (Kuang et al.)
19%
45%
36%
35%
16%
49%
4%
14%
82%
B
"All or none" splicing (observed)
"Al or none" splicing (predicted)
number
of introns
2
3
4
> 4
C
Co-association score
Gene splicing
number
of introns
2
3
4


Herzel_Fig5
A
WT
prp2-1
26°C
37°C for 2 hours
cell harvest
cell fractionation
chromatin
nRNA
RT-PCR and LRS
splicing differences?
WT
prp2-1
26°C
37°C
26°C
37°C
SPAC29E6.08_Intron2
SPBC3B9.15c_Intron1
C
single-intron gene splicing
0.6
0.4
0.2
0.0
***
WT 37°C
prp2-1 37°C
D
WT, 2h 37°C
prp2-1, 2h 37°C
40%
21%
39%
59%
24%
17%
all spliced
all unspliced
partially spliced
B
SPBC19G7.17
SPAC22E12.13c
WT 37°C
prp2-1 37°C
SPBC19G7.17
SPAC22E12.13c
200 bp
E
genes
with intron(s)
no intron(s)
p < 2·10-16
Fraction of genes
1.00
0.75
0.50
0.25
0.00
−2
0
2
mRNA expression fold-change log2(prp2-1/wt)

Herzel_Fig6


A
nRNA
mRNA
***
Gene splicing
1.00
0.75
0.50
0.25
0.00
[mRNA]
[nRNA]
< 1
≥ 1
< 1
≥ 1
B
+water
+10mM caffeine
15min
cell harvest
cell fractionation
chromatin
cytoplasm
nRNA
mRNA
splicing
expression
C
mRNA expression (caffeine)
10,000
100
1
0.01
up
down
expression change
not significant (5,439)
significant (1,477)
0.01
1
100
10,000
mRNA expression (water)
D
nRNA levels similar
mRNA expression
down
up
****
Fraction of genes
1.00
0.75
0.50
0.25
0.00
−0.2
−0.1
0.0
0.1
0.2
splicing difference:
nRNA (caffeine) - nRNA (water)


Herzel_Fig7
A
200 bp
SPAC1805.10
SPAC1805.11c
SPBC3B9.12
SPBC3B9.13c
B
3' end in gene body
3' end after polyA site
single-intron genes
31%
69%
23%
77%
multi-intron genes
34%
15%
51%
20%
59%
21%
partially spliced
all spliced
all unspliced
D
200 bp
total RNA-seq
Δrrp6
WT
SPAC1805.11c
RNA-seq coverage (RPM)
C
3'end coverage relative to polyA site
nRNA
mRNA
Distance from polyA site [nt]
intronless
introncontaining
all spliced
all unspliced
E
RT-qPCR from nascent RNA
last exon
polyA
post-polyA
polyA site
Signal relative to last exon
wild-type
Δrrp6
polyA
post-polyA
SPAC1805.11c