THE UNIVERSITY OF CHICAGO

THE LIFE AND DEATH OF LARIAT RNAS:
INSIGHTS INTO CO-TRANSCRIPTIONAL SPLICING AND MRNA EXPORT

A DISSERTATION SUBMITTED TO

THE FACULTY OF THE DIVISION OF THE BIOLOGICAL SCIENCES

AND THE PRITZKER SCHOOL OF MEDICINE

IN CANDIDACY FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

GRADUATE PROGRAM IN CELL AND MOLECULAR BIOLOGY

BY

YI ZENG

CHICAGO, ILLINOIS

DECEMBER 2020

To Xiaolan and Huilin

COPYRIGHT © 2020 BY YI ZENG

ALL RIGHTS RESERVED

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENT

First of all, I would like to thank my advisor Jon Staley for his patience, support, and guidance. I am grateful to Jon for showing me how to be a good scientist. I thank Jon for giving me the freedom and opportunities to explore new ideas. The curiosity toward science, the ability to make unexpected connections, and the intellectual rigor that Jon has are traits I hope to attain.

I thank the members of my thesis committee, Ben Glick, Doug Bishop, and Alex Ruthenburg, for their helpful discussions and their support on my projects, and for helping me see the big picture and improve my communication skills.

I thank past and present members of the Staley lab for many helpful discussions and making the lab such an awesome place to do science. I thank Daoming for his support inside and outside of the lab. I thank Aiswarya and Yichen for their support on the CoLa-seq project. I thank Klaus for being a big brother, who I can always count on. I will always cherish the conversations I had with Chris and Cody about science, basketball, and life. I am grateful to Ling for her support both inside and outside of the lab and for always making time to help me troubleshoot my experiments.

I thank Ben Fair and Yang Li, who have provided the technical support necessary to analyze the CoLa-seq data and have brought fresh ideas and new perspectives to the project. I thank John Hall in the Ruthenburg lab for getting me started with the CoLa-seq project. I thank Huilin Zeng for helping me build the computational models to interpret the CoLa-seq data.

I thank my friends and family. I am indebted to my mom, who has encouraged and supported me to pursue my interests for as long as I remember, and who has always made sure I'm okay. Lastly, I thank my fiancée Huilin, who has believed in me even when I doubted myself, and who I selfishly hope will continue to be my biggest supporter for many years to come.

ABSTRACT

Most eukaryotic genes are expressed as precursor messenger RNAs (pre-mRNAs) that are converted to mRNAs by splicing, an essential step of gene expression. Splicing is catalyzed by the spliceosome, a large and dynamic ribonucleoprotein machine, which removes non-coding intervening sequences (introns) from pre-mRNAs and ligates flanking coding sequences (exons) to produce mature mRNAs for protein synthesis in the cytoplasm. In humans, nearly all pre-mRNAs undergo alternative splicing to produce multiple mRNA isoforms with distinct functions, greatly expanding the complexity of the human transcriptome and proteome. Further, alternative splicing is regulated in time and space, playing critical roles in many biological processes. In fact, mutations and misregulation of splicing have been associated with a large number of human diseases. While *in vitro* splicing can occur in the absence of transcription, *in vivo* splicing is intimately coupled with transcription both physically and functionally. However, our understanding of co-transcriptional splicing remains incomplete in part due to limited information on the timing of splicing relative to transcription. Splicing is also a highly accurate process. To promote fidelity in splicing, the spliceosome discriminates and discards suboptimal splicing substrates that have engaged the spliceosome. Although nuclear quality control mechanisms have been proposed to retain immature mRNPs, discarded splicing substrates, including lariat intermediates, do export to the cytoplasm. However, the mechanism for exporting these species has remained unknown.

Here, we investigated the dynamics and regulations of co-transcriptional splicing in humans and characterized the nuclear export pathway of lariat intermediates in yeast. To study co-transcriptional splicing in humans, we developed a genome-wide approach, CoLa-seq, or co-transcriptional lariat sequencing. We show that CoLa-seq enables efficient mapping of branch

points, an essential reactant in the splicing reaction, in a cell type-specific manner. Additionally, for the first time *in vivo*, we show that adjacent introns can undergo concurrent splicing. Notably, we provide evidence that the timing of splicing can vary dramatically both across introns and even within the same intron. Further, we show that the splicing of a given intron can occur both through intron definition and exon definition pathway, which has implications for the regulation of alternative splicing. Importantly, we identified key *cis*-elements and *trans*-acting factors that correlate with the timing of co-transcriptional splicing.

Using a combination of ensemble and single molecule approaches, we demonstrate that the spliceosome-discarded lariat intermediates use the same nuclear export pathway as mature mRNAs do. Further, our findings suggest that during mRNA export, Tom1-mediated ubiquitylation of Yra1p undocks mRNA from the nuclear basket of the nuclear pore, allowing mRNA to transit through the nuclear pore for export.

CHAPTER 1

INTRODUCTION TO PRE-mRNA SPLICING AND mRNA EXPORT

Overview

Precursor messenger RNA (pre-mRNA) from essentially all human genes undergo a series of RNA processing events to form a mRNA, which is then exported into the cytoplasm for translation. These RNA processing events include 5' end capping, splicing, as well as 3' end cleavage and polyadenylation. Synthesis of pre-mRNA begins when RNA polymerase II (RNAP II) binds to a promoter to initiate transcription. After transcription initiation, RNAP II synthesizes the pre-mRNA as it elongates through the chromatin template and terminates often many kilobases (kbs) downstream (Selth et al., 2010). At the beginning of transcription, a 7-methyl guanosine cap is added to the 5' end of the pre-mRNA to protect it from endonucleases (Galloway and Cowling, 2019). During transcription, introns in the pre-mRNA are removed by splicing. At the end of the transcription, the transcribed pre-mRNA is cleaved at a polyadenylation (polyA) site and then polyadenylated to yield the mature 3' end (Sun et al., 2020), after which the mature mRNA is exported into the cytoplasm for protein synthesis.

In addition to being one of many essential processing steps during mRNA maturation, splicing is harnessed to produce multiple mRNAs from a given pre-mRNA transcript, thereby greatly increasing the complexity of transcriptome and proteome. In fact, more than 95% of human genes undergo alternative splicing (Pan et al., 2008; Wang et al., 2008). Further, alternative splicing is tightly regulated in a cell-, tissue-, development stage-, and signal-dependent manner. Importantly, mutations or misregulation of splicing are the cause of many human diseases or contribute to the severity of diseases (Manning and Cooper, 2017). Besides its diverse functions in different cellular processes, splicing is also important for efficient transcription (Furger et al., 2002), RNA stability (Moore and Proudfoot, 2009), RNA export (Valencia et al., 2008), as well as the biogenesis of non-coding RNAs, including telomerase RNA in *Schizosaccharomyces pombe*,

snoRNAs, and microRNAs (Kannan et al., 2013; Pawlicki and Steitz, 2010; Westholm and Lai, 2011).

Splicing is catalyzed by the spliceosome, an evolutionarily conserved yet extraordinarily dynamic ribonucleoprotein machine. Using powerful splicing assays developed in both mammalian and yeast cells, we have gained great understanding of splicing mechanisms and the inner workings of the spliceosome (Wahl et al., 2009). Facilitated by recent advances in cryo-electron microscopy, the formation and actions of the spliceosome have been illuminated at remarkable detail (Kastner et al., 2019; Plaschka et al., 2019; Yan et al., 2019; Zhang et al., 2019). While *in vitro* splicing can occur in the absence of other processes, *in vivo* splicing is predominantly co-transcriptional (Beyer and Osheim, 1988; Bhatt et al., 2012; Neugebauer, 2019; Wuarin and Schibler, 1994). Mounting evidence indicates that splicing is deeply coupled to transcription and chromatin. In this chapter, I will first review the assembly and dynamics of the spliceosome, briefly discuss splicing fidelity, and then summarize our understanding of co-transcription splicing. In the end, I will provide an overview of mRNA export.

Exon/intron architecture and chemistry of pre-mRNA splicing

Almost all human pre-mRNA transcripts contain multiple short exons separated by long introns, ten introns on average. Human exons (median size: 147bp) are much shorter than introns, which are often several kbs long. Although human introns are much longer than yeast introns, intron structures are well conserved (Wahl et al., 2009). The 5' and 3' ends of an intron are respectively marked by the 5' splice site (SS) and the 3' SS, two essential reactants during the splicing reaction. The third essential reactant, branch point (BP), is located 10 to 50 nts upstream of the 3' SS (**Fig. 1.1A**). In metazoans, the sequence between the BP and the 3' SS contains a

polypyrimidine tract (PPT), which helps define branch point during the spliceosome assembly and may play a role in the exon ligation (**Fig. 1.1A**).

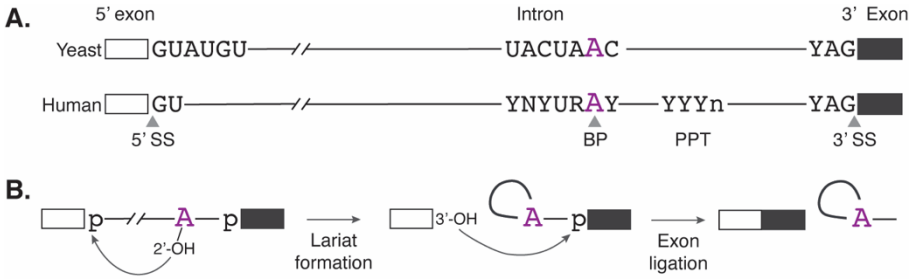


Figure 1.1. Schematic representations of yeast and human pre-mRNA substrates and the splicing reaction.
A. Intron/exon boundaries are defined by the 5' SS and 3' SS separately. Each intron contains three short conserved sequences at the 5' SS, the BP, and the 3' SS, which are conserved in yeast but more degenerate in humans. These sequences are recognized multiple times during the splicing cycle to maintain the fidelity of the splicing reactions. Purple indicates the BP adenosine. **B.** Introns are removed by two sequential transesterification reactions, lariat formation and exon ligation.

The spliceosome recognizes the conserved yet short sequences at the 5' SS, the BP, the PPT, and the 3' SS and carries out two sequential transesterification reactions (Wilkinson et al., 2020). In the first step of splicing, referred to as lariat formation, the 2' hydroxyl of a conserved BP adenosine acts as a nucleophile to attack the 5' SS phosphate, generating a branched lariat intermediate and a cleaved 5' exon with a free 3' hydroxyl (**Fig. 1.1B**). In the second step, referred to as exon ligation, the free 3' hydroxyl of the cleaved 5' exon attacks the 3' SS phosphate, ligating the exons and excising the branched lariat intron (**Fig. 1.1B**). The excised lariat intron is further debranched and degraded. These spliceosome catalyzed reactions are indistinguishable from those catalyzed by self-splicing group II introns, indicating that the spliceosome and group II introns have a common evolutionary origin (Fica et al., 2013).

Global annotation of branch points

The BP is one of three essential reactants participating in the splicing reaction and its usage directly impacts 3' SS selection. Mutations altering BP usage are associated with diseases (Darman et al., 2015). Despite its essentiality, BP annotation in the human transcriptome has proven to be challenging and lags far behind 5' and 3' SS annotations. Because 5’ and 3’ SS are present in the mature mRNAs, they can be precisely mapped by aligning reads from spliced RNAs to the genome (Wang and Burge, 2008). In contrast, BPs only reside in lariat RNA species, including lariat intermediates and excised lariat introns, which are transiently present at low levels, and surrounding sequences of BPs are highly degenerative, making BPs hard to capture and predict computationally. Recent large-scale approaches improved BP annotation by analyzing lariat RNA-derived inverted reads, in which BPs are invertedly located upstream of the 5' SSs (Mercer et al., 2015; Taggart et al., 2017; Pineda and Bradley, 2018). These inverted reads were incidentally captured during the cDNA synthesis step of RNA-seq library construction when reverse transcriptase traversed through the 2’-5’ bonds formed between the BP and the 5’ SS of the lariat RNA species. However, these reads were present at such low frequency that their analyses required either targeted enrichment or pooling of thousands of RNA-seq samples across diverse tissue and cell types (Mercer et al., 2015; Pineda and Bradley, 2018), either of which was still unable to capture most used branch points. Further, targeted enrichment limited BP mapping to only known introns, and pooling of different RNA-seq samples causes loss of cell/sample-type specificity. Importantly, both approaches did not directly capture the corresponding 3’ SS of individual BPs. A recent study in yeast approached BP annotation from a different perspective (W. Chen et al., 2018). Because the branched structure in lariat RNAs can effectively stop a reverse transcriptase, a cDNA synthesis starting downstream of the BP would stop at the nucleotide immediately

downstream of the BP, generating a cDNA fragment containing the precise location of the BP at its 3' end. In doing so, the authors identified more BPs from these RT-stopped reads than inverted reads. However, this method relies on cumbersome procedures involving immunoprecipitation of spliceosomes. Thus, this approach may not be readily feasible if the input sample size is small or a good-quality antibody is not available. In our recent work, we developed a robust method to map BPs in humans. Instead of immunoprecipitating the spliceosome, we used lariat RNA species enriched from chromatin-associated RNAs for BP mapping, which I will discuss in more detail in chapter 2. As a result, this method greatly expanded BP annotation in the human genome. Its efficient branch point mapping from a single sample further opens the door for comparing BP usages between different conditions. For instance, it can be used to profile changes in BP usage in cancer cells containing SF3B1 mutations, which were found to alter BP usage and consequently 3' SS selection, yielding aberrant mRNAs (Darman et al., 2015).

Spliceosome assembly and catalysis

The building blocks for the spliceosome are five small nuclear ribonucleoprotein particles (snRNPs): U1, U2, U4, U5, and U6. Each snRNP is composed of a small nuclear RNA (snRNA), whose 3' end is encircled by seven Sm proteins (or seven Lsm proteins in the case of U6 snRNA), and various snRNP-specific proteins. In cells, the U4, U6, and U5 snRNPs form a U4/U6.U5 tri-snRNP, in which the U4 and U6 snRNAs are base-paired.

The spliceosome assembles *de novo* on to each intron via sequential additions of snRNPs and non-snRNP protein factors. This initial assembly and downstream steps require extensive compositional and conformational rearrangements of the spliceosome that are driven by eight conserved DEXD/H-box RNA-dependent ATPases (Staley and Guthrie, 1998; Cordin and Beggs, 2013). After exon ligation, the spliceosome disassembles and recycles for a new round of splicing.

During the initial spliceosome assembly (**Fig. 1.2A**), U1 snRNP binds to the 5' SS via base pairing between the 5' end of U1 snRNA and the conserved sequences at the 5' SS. At the 3' end of the intron, splicing factor 1 (SF1) binds to the BP (Kastner et al., 2019). Binding of SF1 is stabilized by the U2 auxiliary factor (U2AF), a heterodimer made of U2AF1 and U2AF2 that binds to the terminal AG dinucleotide of the 3' SS and the PPT, respectively (Berglund et al., 1998; Kastner et al., 2019). Introns with strong PPTs do not require the 3' SS recognition by U2AF1 and can undergo lariat formation in the absence of U2AF1 (Guth et al., 1999; Wu et al., 1999). Indeed, structural studies have found that the strength of the PPT can influence the conformation of the U2AF heterodimer, thus impacting the recruitment of the U2 snRNP (Mackereth et al., 2011; Warnasooriya et al., 2020). Together, these initial interactions result in the formation of the E complex in an ATP-independent manner.

Subsequently, U2 snRNP replaces SF1 at the branch point to form the A complex (**Fig. 1.2A**), in which U2 snRNA is paired with the BP sequence with the BP adenosine bulged out. This interaction of the U2 snRNP with the BP branch requires the activity of two ATPases, Prp5 and UAP56 (Sub2 in yeast). Efficient splice site recognition is further promoted by cooperative interactions between the U1 snRNP and the U2 snRNP (Abovich and Rosbash, 1997; Dönmez et al., 2007; Michaud and Reed, 1993; Shao et al., 2012). While the establishment of such cross-intron interactions is required for subsequent steps, initial splice site recognition in multi-exonic human pre-mRNAs are thought to occur across exons in a process called exon definition, owing to the fact that introns are much longer than exons in humans and thus may antagonize efficient physical interaction between U1 and U2 snRNPs. In exon definition, the U1 snRNP and U2AF/U2 snRNP from adjacent introns are thought to interact first across an exon, which helps define the 3' SS of the upstream intron (**Fig. 1.2B**). This cross-exon interaction is often stabilized by SR proteins

that bind to exonic sequences via their RRM domain and interact through their RS domains with U2AF1 and U1-70K within U1 snRNP (Wu and Maniatis, 1993). Through a not well understood process (Schneider et al., 2010), this exon definition complex is eventually converted to the intron definition complex at this or later stage, establishing cross-intron interactions between the U1 snRNP and the U2 snRNP, which are bound to the 5' SS and the 3' SS of the upstream intron, respectively (**Fig. 1.2B**). It is proposed that exon definition is the dominant pathway of spliceosome assembly (Berget, 1995; Hollander et al., 2016), though recent genome-wide data indicate that the intron definition pathway is also utilized (Reimer et al., 2020). It would be important to understand if intron definition also serves as a major mode of spliceosome assembly in humans as intron definition allows flexibility in splice site choice at the earliest stages of spliceosome assembly. Indeed, these early stages of spliceosome assembly are reversible, allowing splice site sampling, and subject to extensive regulation (Chen and Manley, 2009; Fu and Ares, 2014; Hoskins et al., 2011). For example, regulation of AS is often achieved by promoting or repressing the binding of the U1 or U2 snRNP to the splice sites (Chen and Manley, 2009).

Following the formation of the A complex, the U4/U6.U5 tri-snRNP is recruited (Galej et al., 2013; Wilkinson et al., 2020). In the resulting pre-B complex (**Fig. 1.2A**), U6 snRNA is still in its inactive form, but its 3' end base pairs with the 5' end of the U2 snRNA to form a short duplex (U2/U6 helix II). To further integrate the tri-snRNP into the spliceosome, the DEAD-box helicase Prp28 releases U1 snRNP by disrupting the U1/5' SS helix so that the 5' SS can base pair with the ACAGA box within U6 snRNA (Staley and Guthrie, 1999) and the last few nucleotides of exons also base pairs with the loop I of U5 snRNA (Newman and Norman, 1992). In the resulting B complex (**Fig. 1.2A**), Brr2 is positioned to unwind the U4/U6 duplex (Charenton et al., 2019), leading to the displacement of U4 snRNPs and recruitment of multiple new proteins

including a large Prp19 complex, Prp19 related proteins, the pentameric intron-binding complex (IBC), and the retention and splicing (RES) complex (Kastner et al., 2019). Concomitantly, RNA and proteins undergo extensive conformational changes, resulting in the formation of the B^{act} complex. In the B^{act} complex (**Fig. 1.2A**), the freed U6 snRNA pairs with U2 snRNA to fold into a triple helix, which forms the active site of the spliceosome and positions two catalytic Mg^{2+} ions (Fica et al., 2014; Rauhut et al., 2016; Yan et al., 2019). The 5' SS is also docked into the active site and stabilized by proteins and the interaction of U5 snRNA with the last few nucleotides of the 5' exon. Although the 5' SS is docked into the active site, the BP adenosine is kept away from the active site as the U2/BP helix is encapsulated by SF3B1 of the SF3B complex (Haselbach et al., 2018). Therefore, the B^{act} complex requires additional remodeling to become catalytically competent.

The last step of spliceosome activation requires DEAH-box ATPase Prp2, which remodels the spliceosome to become the catalytically competent B* complex (**Fig. 1.2A**), which catalyzes the first step of splicing, leading to 5' exon cleavage and lariat formation. In the resulting C complex (**Fig. 1.2A**), the exon junction complex (EJC), a complex specific to higher eukaryotes, is loaded onto a region about 20 nucleotides upstream of the 5' SS (Le Hir et al., 2016; Schlautmann and Gehring, 2020). In order to catalyze the second step of splicing, another DEAH-box ATPase Prp16 remodels the C complex to form the C* complex (**Fig. 1.2A**), which catalyzes the second step of splicing, leading to exon ligation and excision of lariat intron (Schwer and Guthrie, 1992). In the resulting P complex (**Fig. 1.2A**), DEAH-box ATPase Prp22 releases mRNAs (Schwer, 2008), yielding the intron lariat spliceosome (ILS; **Fig. 1.2A**). Finally, the ILS is disassembled by DEAH-box ATPase Prp43. The released lariat intron is debranched and degraded, whereas the U2, U5, and U6 snRNPs are recycled for new rounds of splicing.

The four ATPases (Prp2, Prp16, Prp22, and Prp43) in the splicing cycle belong to the same DEAH family (Cordin and Beggs, 2013). Members of this family can bind to single stranded RNA, hydrolyze ATP in an RNA dependent manner, and translocate on the RNA in a 3'-5' direction (He et al., 2017). Despite their essential roles in splicing, it is not yet clear how they remodel the spliceosomes. However, based on recent biochemical studies and their peripheral locations on the spliceosome (Semlow et al., 2016; Wilkinson et al., 2020), it is plausible that all four ATPases acts as a pullase, which functions at a distance by pulling on the RNA, rather than a translocase, which translocates through the RNA to disrupt RNA-RNA or RNA-protein interactions (Semlow et al., 2016).

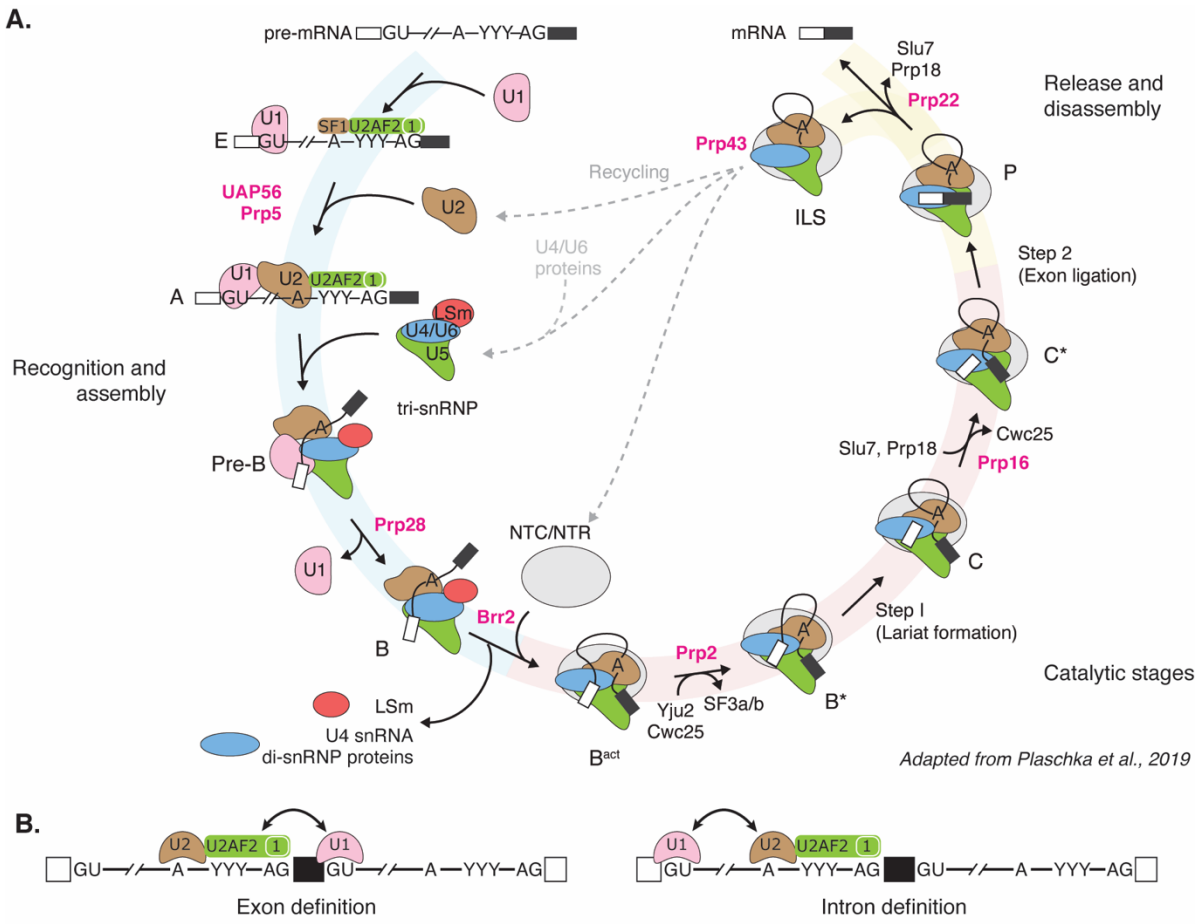


Figure 1.2. Schematic representation of the splicing cycle, the exon definition, and the intron definition.

A. The spliceosome is assembled in a stepwise manner (Recognition and assembly). Then, the spliceosome is activated to carry lariat formation and exon ligation reactions (Catalytic stages), after which the spliceosome releases mRNA and disassembles (Release and disassembly). During the splicing cycle, the spliceosome undergoes extensive compositional and conformational rearrangements, which are promoted by various DExD/H ATPases, including UAP56, Prp5, Prp28, Brr2, Prp2, Prp16, Prp22, and Prp43. ILS, intron-lariat spliceosome; NTC, Prp19-associated complex; NTR, Prp19-related complex. Schematic is adapted from Plaschka et al., 2019**. B.** In exon definition, the initial splice site recognition occurs across an exon, with U1 snRNP binding to the downstream 5' SS and U2AF/U2 snRNP binding to the 3' SS and the BP of the upstream intron. In the intron definition, U1 snRNP and U2AF/U2 snRNP recognition the 5' SS and the 3' SS of the same intron, respectively. The exon definition complex is eventually converted to the intron definition complex.

Alternative splicing

Alternative splicing (AS) allows a given pre-mRNA transcript to produce multiple mRNA isoforms by using combinations of different splice sites. AS is crucial for regulating gene expression and enhancing the complexity of transcriptome and proteome. In humans, more than 95% of multi-exon genes undergo AS (Pan et al., 2008; Wang et al., 2008), and the frequency of AS scales with cell type and species complexity (Barbosa-Morais et al., 2012; Merkin et al., 2012). At least 50% of alternatively spliced isoforms are differentially expressed among human tissues (Wang et al., 2008). Recent large studies have found that alternative exons in protein coding regions are often located in regions that are sites of post-translational modifications and protein-protein interactions, suggesting that a likely general function of AS is to control protein-protein interactions (Yang et al., 2016; Ule and Blencowe, 2019). Further, AS plays critical roles in multiple biological processes including cell cycle, cell growth, cell differentiation, cell death, development, and response to cell signaling (Braunschweig et al., 2014; Dominguez et al., 2016; Manning and Cooper, 2017; Ule and Blencowe, 2019). Its widespread nature and diverse functions underscore the importance of understanding the mechanisms of AS and the regulation of AS.

The common forms of AS in humans are cassette exon skipping, alternative 5' SS usage, alternative 3' SS usage, alternative intron retention, and mutually exclusive exons. Exon skipping counts for over 50% of AS events observed in humans (Hyung et al., 2018). The decisions as to which splice site to include or skip are often made during initial splice site recognition and early spliceosome assembly (Chen and Manley, 2009). However, recent studies have shown that this decisions can also be made after lariat formation (Lallena et al., 2002; Semlow et al., 2016). Choices of AS involve multiple parameters including *cis* RNA elements, *trans*-acting factors, transcriptional kinetics, nucleosome positioning, and epigenetic modifications.

Outside of the sequence information at the core splicing signals (the 5' SS, the 3' SS, and the BP), numerous cis-regulatory sequence elements are present to ensure accurate usage of splice sites (Wang and Burge, 2008). The vast majority of these regulatory elements are short and linear sequence motifs, although structured RNA elements have been discovered to function in splice site selection (McManus and Graveley, 2011). Depending on their positions and functions, these cis-regulatory elements are divided into four categories: exonic splicing enhancers (ESEs), exonic splicing silencers (ESSs), intronic splicing enhancers (ISEs), and intronic splicing silencers (ISSs). The concerted actions of multiple enhancers and silencers are often required to ensure the correct selection of splice sites (Barash et al., 2010). In particular, enhancers that support constitutive exons are often located in exons, whereas enhancers and silencers involved in alternative exon usage are often located in flanking introns, in agreement with the observation that intronic regions (>=150 bps) surrounding alternative exons are far more conserved than those surrounding constitutive exons (Sorek and Ast, 2003).

In general, these regulatory elements function by recruiting trans-acting factors that either activate or suppress splice site recognition or spliceosome assembly. These trans-acting factors generally belong to three classes: 1). the SR (Ser–Arg) proteins; 2). heterogeneous ribonucleoproteins (hnRNPs); and 3). several tissue-specific RNA-binding proteins, such as Nova, neuronal PTB, nSR100/SRRM4, the Rbfox family, and the muscleblind/CELF family (Fu and Ares, 2014). SR proteins consist of RNA recognition motifs (RRMs) and RS domains, which mediate protein-protein and protein-RNA interactions. SR proteins are generally considered as positive splicing regulators by binding to ESEs (Fu and Ares, 2014). Recent studies have found that SR proteins can bind to ISEs. For example, neuronal SR100 binds to ISEs upstream of 3' SS of neuronal exons and opposes the repressive function of PTBP1 (Gonatopoulos-Pournatzis et al.,

2018; Raj et al., 2014). However, several studies have shown that SR protein could suppress splicing when they are bound to intronic regions downstream of target exons (Erkelenz et al., 2013; Ibrahim et al., 2005). In contrast to SR proteins, hnRNPs proteins were first established as splicing repressors by binding to ESSs or ISSs (Fu and Ares, 2014). These proteins can function via a variety of mechanisms. For example, PTB (hnRNP I) can bind to PPTs to block the binding of U2AF2 (Saulière et al., 2006), whereas hnRNP A1 can bind to both sides of exon 7B in its own pre-mRNA and "loop out" the exon to promote exon skipping (Nasim et al., 2002).

While it is generally helpful to consider SR proteins as activators and hnRNPs as repressors, members of each class can either activate or repress splice site selection depending on their binding locations and surrounding sequence context (Fu and Ares, 2014). For instance, hnRNP H suppresses splicing when bound to exons but promotes nearby 5' SS usage when bound to intronic G-rich sequences (Chou et al., 1999; Mauger et al., 2008; Xiao et al., 2009). Similarly, neuron specific splicing factor Nova promotes exon inclusion when bound to the intronic sequence of the alternative exon but represses its inclusion when bound upstream of the alternative exon (Dredge et al., 2005; Ule et al., 2006).

Enhancers and silencers often function additively to increase either the binding affinity for trans-acting factors or the local concentrations of these factors. Different enhancers or silencers could also function cooperatively to enhance their functions. Indeed, many RNA binding proteins (RBPs) bind cooperatively by interacting with repeated motifs, bipartite motifs, or multivalent motifs. Interestingly, many RBPs contain intrinsically disordered regions (IDRs), which may allow these RBPs to form multivalent interactions when bound to RNAs (Ule and Blencowe, 2019). In fact, purified RBPs (i.e. PTBP1) can undergo liquid-liquid phase separation in the presence of RNA (Li et al., 2012). Such phase separation dramatically increases local protein concentrations,

likely resulting in more stable RNA binding and thus efficient splicing regulation. In support of this model, several recent studies have found that IDRs of several RBPs drive condensation and are required for regulating AS. For example, Rbfox forms a large RNP complex, called LASR (large assembly of splicing regulators), through its tyrosine-rich IDR. The integration of Rbfox into LASR is required for regulating splicing in a subset of alternative exons (Damianov et al., 2016; Ying et al., 2017). Another recent study found that exons that correspond to IDRs of many RBPs (i.e. hnRNP A and D families) can undergo AS, implying that differential inclusion of exons can be used to control multivalent interactions and in turn their functions (Gueroussov et al., 2017). Together, protein-RNA and protein-protein interactions form regulatory networks to control AS.

Co-transcriptional splicing

Much of our knowledge about the spliceosome assembly and catalysis have been acquired from *in vitro* studies. In powerful *in vitro* systems, splicing can occur in the absence of transcription and other RNA processing events. However, accumulating evidence indicates that *in vivo* splicing occurs during RNAP II-mediated transcription, and splicing is tightly coupled to transcription and chromatin (Neugebauer, 2019). Thus, a key to fully understand the splicing mechanisms and roles of other processes in splicing is understanding the timing of splicing relative to transcription.

Initial evidence of co-transcriptional splicing was from electron micrography analysis of embryonic transcript units using chromatin spreads from *Drosophila melanogaster* (Beyer et al., 1981; Beyer and Osheim, 1988). These electron micrographs indicated that spliceosome assembly, lariat formation, and exon ligation can occur while the transcript is tethered to the chromatin template. Analysis of the E74 gene in *Drosophila melanogaster* also revealed that nascent E74 transcripts were spliced before polyadenylation (LeMaire and Thummel, 1990). Similar observations were also implicated by observations obtained for the endogenous APRT gene in

Chinese hamster ovary cells (Kessler et al., 1993). Later studies have observed efficient splicing and both in-order and out-of-order splicing during transcription using transcribing RNAs from micro-dissected Balbiani Ring (BR) genes (Baurén and Wieslander, 1994; Wetterberg et al., 1996). Recent single module fluorescence *in situ* hybridization analysis of individual transcripts in mammals revealed that constitutive introns were spliced co-transcriptionally, whereas introns with strong secondary structures or introns that were alternatively spliced were spliced post-transcriptionally (Vargas et al., 2011). Consistent with these observations, by immunofluorescence or chromatin immunoprecipitation (ChIP), studies in both yeast and mammals revealed that spliceosomal snRNPs and splicing factors associated with intron-containing genes but not with intron-less genes during transcription (Huang and Spector, 1991, 1996; Lacadie and Rosbash, 2005; Lacadie, 2006; Görnemann et al., 2005; Listerman et al., 2006).

These targeted gene analyses have since been extended to the transcriptome level. By analyzing RNA-seq data produced from different RNA samples including total RNA, nuclear RNA, nucleoplasmic RNA, and chromatin-associated RNA, studies obtained different estimates of co-transcriptional splicing efficiency but reached a similar conclusion that most splicing is co-transcriptional (Ameur et al., 2011; Bhatt et al., 2012; Khodor et al., 2011; Tilgner et al., 2012). These studies indicate that post-transcriptionally spliced introns are enriched for alternatively spliced introns. Consistent with this observation, an antibody specific for the active, phosphorylated form of SF3B1 revealed 80% of the active spliceosome on chromatin and the remaining 20% in the nucleoplasm (Girard et al., 2012).

However, splicing being co-transcriptional does not guarantee that introns transcribed first are spliced first. In fact, frequent out-of-order splicing events, where later transcribed introns are spliced before earlier transcribed introns, are observed. Out-of-order splicing is associated with

specific sequence features (Kim et al., 2017; Drexler et al., 2020), implying that order of splicing may be involved in the regulation of AS. Indeed, a previous *in vitro* study found that out-of-order splicing brings close a regulatory element from downstream exon to promote splicing of the upstream intron, thereby leading to exon inclusion (Nasim et al., 1990), supporting the view that modulating the order of splicing could be used to influence the outcomes of AS.

The observation of transcription and splicing occurring at the same time does not necessarily indicate that these two processes are functionally coupled; however, mounting evidence indicates that splicing and transcription are tightly coupled both physically and functionally and supports two non-mutually exclusive mechanisms: recruitment coupling and kinetic coupling (Bentley, 2014; Saldi et al., 2016). Recruitment coupling involves the differential association of splicing factors with the transcription machinery; whereas kinetic coupling is achieved by modulation of the transcription rate.

In the recruitment coupling model, splicing factors are first recruited to the transcription elongation complex via the C-terminal domain (CTD) of RNAP II and then get deposited to specific sequences of the transcript as it emerges from RNAP II. The CTD consists of 52 heptad repeats with the consensus $Y_1S_2P_3T_4S_5P_6S_7$. Heptads in the N-terminal half of the CTD mostly conform to the consensus sequence, while heptads in the C-terminal half are more degenerate. The CTD is subject to extensive phosphorylation, which plays important and distinct functions in transcription and RNA processing. The initial support for the role of the CTD in splicing was from truncation studies, in which truncation of CTDs severely impaired 5' end capping, splicing, and 3' end processing of splicing reporters (Mccracken et al., 1998; McCracken et al., 1997; Misteli et al., 1997), as well as the outcomes of AS (de la Mata and Kornblihtt, 2006; Rosonina and Blencowe, 2004). *In vitro* studies also observed that phosphorylated CTD peptides stimulated splicing (Hirose

et al., 1999; Zeng and Berget, 2000), implying that the CTD might act as a platform to facilitate spliceosome assembly. Indeed, the phosphorylated CTD directly interacts with U1 snRNP-associated protein Prp40 in yeast, CA150, PSF, and U2AF (Hsin and Manley, 2012), suggesting that the CTD could help assemble the spliceosome on both ends of the intron during transcription. Consistent with this idea, the interaction of U2AF with the CTD recruits the PRP19 complex (David et al., 2011). Using RNAP II antibodies specific for serine 5 phosphorylated CTD (S5P-CTD) to immnunoprecipitate RNAP II from MNase-digested chromatin, a recent study identified by mass spectrometry that the S5P-CTD antibody co-precipitated SR proteins as well as spliceosomal components involved in the different steps of splicing cycle (Nojima et al., 2018). This result suggests that RNAP II with serine 5 phosphorylated CTD interacts with the active spliceosome during co-transcriptional splicing, providing direct support for early *in vivo* studies described above. Similar observations were also observed in yeast (Harlen et al., 2016). It is worth noting that some antibodies against the phosphorylated CTDs can cross-react with phosphorylated SR splicing factors (Doyle et al., 2002), leading to the capture of RNAs that are not directly associated with RNAP II, such as spliced mRNAs. Interestingly, a recent study found that components of the SAGA complex, a conserved transcription co-activator, directly interacted with Prp5 and modulates its proofreading function (Shao et al., 2020), implying that transcription might affect splicing fidelity.

The second way to achieve functional coupling is via kinetic coupling, in which the rate of transcription governs RNA processing. This model predicts that changes of the transcription rate would affect splicing efficiency and AS. Consistent with this idea, a slow RNAP II mutant enhanced constitutive splicing in yeast and flies (Khodor et al., 2011; Braberg et al., 2013; Aslanzadeh et al., 2018) and exon inclusions in the fibronectin and NACAM genes in humans (de

la Mata et al., 2003). Changes in the transcription rate can potentially impact the binding of both positive and negative splicing regulators, therefore slow transcription could also favor exon skipping. Indeed, in a slow RNAP II mutant, CTFR exon 9 was preferentially skipped as negative splicing factor ETR-3 prevented binding of U2AF2 to the PPT (Dujardin et al., 2014). However, a recent genome-wide study found that both slow and fast RNAP II mutants had the same impact on the outcomes of AS; both mutants increased or decreased inclusions of many alternative exons (Fong et al., 2014), suggesting that an optimal transcription rate is essential for accurate splicing during transcription. However, it is unclear how much of these effects are direct as changes in transcription dramatically alter the profile of gene expression.

Through kinetic coupling, factors can influence the outcomes of AS by modulating the transcription rate. These factors include promoter identity, transcriptional activators, DNA sequence, nucleosome occupancy, chromatin remodeling, and chromatin modifications (Batsché et al., 2006; Cramer et al., 1999, 1997; de Almeida et al., 2011; Dowhan et al., 2005; Guo et al., 2014; Iannone et al., 2015; Jimeno-González et al., 2015; Luco et al., 2010; Pradeepa et al., 2012). For example, the transcription barrier posed by nucleosomes is implicated in influencing AS. In progesterone-treated breast cancer cells, hormone-induced exon inclusions correlated with higher nucleosome occupancy and increased RNAP II density in the included exons, consistent with RNAP II pausing (Iannone et al., 2015).

As co-transcriptional splicing occurs in close proximity to chromatin, it is perhaps not surprising that chromatin features can shape the outcomes of AS. It has been proposed that specific histone marks recruit splicing factors or their adaptor proteins. H3K36me3, which is enriched over exons and added co-transcriptionally by SETD2 (de Almeida et al., 2011; Kizer et al., 2005; Yoh et al., 2008), was found to be important for splicing. The level of H3K36me3 in cassette exons

correlates with the inclusion level of these cassette exons (Kolasinska-Zwierz et al., 2009; Spies et al., 2009). Further, H3K36me3 could influence AS by recruiting PTB, SRSF1, and EFTUD2 through specific chromatin readers MRG15, Psip1, and BS69, respectively(Guo et al., 2014; Luco et al., 2010; Pradeepa et al., 2012). However, the affected splicing events displayed modest changes, suggesting that these factors may be only important for augmenting or stabilizing AS patterns. Further, it is unclear if some aspects of these studies were confounded by the specificity of H3K36me3 antibodies, as some antibodies against histone modifications were found to be unable to distinguish different methylation states (Rothbart et al., 2015; Shah et al., 2018).

Lastly, to achieve kinetic coupling, factors can directly coordinate transcription and splicing. One such factor is the DBC1-ZIRD (DBIRD) complex, which directly interacts with RNAP II and hnRNP A1. DBIRD depletion slowed down transcription elongation and reduced exon skipping for cassette exons in the AT-rich region (Close et al., 2012). Another factor is BRM, an ATPase subunit of the SWI/SNF complex, which was found to promote the inclusion of alternative exons in CD44 transcripts (Batsché et al., 2006). As BRM directly interacts with RNAP II and spliceosomal components, increased RNAP II density in these alternative exons suggested that BRM promotes exon inclusion by reducing elongation rate and facilitating local spliceosome assembly. Interestingly, the ATPase activity of BRM is not required in this process, suggesting its function in AS is independent of its chromatin remodeling activity.

It is clear that changes in splicing efficiency and AS patterns can be achieved by a variety of mechanisms that either modulate the transcription rate both globally and locally or mediate differential recruitment of splicing factors. However, it is likely that recruitment coupling and kinetic coupling are interwoven because factors that are recruited to RNAP II or chromatin could

affect the transcription rate, and in other cases, altered transcription rate could influence recruitment of splicing factors.

Timing of *in vivo* splicing

While a large body of work has established the functional coupling between splicing and transcription, our understanding of co-transcriptional splicing remains incomplete in part due to our limited information on the kinetics of *in vivo* splicing and the timing of *in vivo* splicing relative to transcription. In contrast to synthetic pre-mRNA substrates used for *in vitro* studies, sequence elements appear at different times during transcription. As a result, at any given moment, only a subset of splice sites and regulatory elements are available to the spliceosome and trans-acting factors. Such availability can be determined by the transcription rate. For instance, fast transcription would increase the number of available splice sites within a given time frame, whereas slow transcription would reduce the number of available splice sites. Transcription also influences DNA, RNA, and even chromatin structures, which in turn affect the binding of regulatory factors and the rate of splicing in the case of RNA secondary structure. Therefore, obtaining the positions of RNAP II during the course of splicing would reveal which sequence elements are necessary for the first or the second step of splicing.

A variety of approaches have been used to estimate the rate of *in vivo* splicing. Using live-cell imaging or qRT-PCR, the rates of *in vivo* splicing have been reported for a few endogenous genes. Although these studies provided rate estimates ranging from seconds to minutes, it is unclear if these estimates are truly representative given the high heterogeneity of the human transcriptome. Using RNA-seq data collected from total RNA, studies estimated the rates of *in vivo* splicing globally (Ameur et al., 2011; Zeisel et al., 2011). The apparent saw-tooth patterns of decreasing read density were observed across introns and interpreted as an indicator of rapid co-

transcriptional splicing. However, it is unclear if such estimates accurately reflect the kinetics of co-transcriptional splicing as total RNA contains not only information about splicing, but also information about RNA synthesis and degradation, which may confound accurate analysis of splicing rate. To overcome this limitation, studies have estimated the rates of *in vivo* splicing from chromatin-associated RNAs, concluding that *in vivo* splicing occurs on the order of minutes, ranging from 15 to 120 minutes (Panda-Jones, 2013). However, it is unclear if such estimates are bounded by experimental time resolution as their earliest time point was 15 minutes after stimulation.

Alternatively, metabolic RNA labeling with 4-thiouracil (4sU) followed by sequencing has been used to measure the rate of *in vivo* splicing. This approach relies on 4sU incorporation during transcription, resulting in the labeling of nascent RNA transcripts. However, this approach introduces bias by capturing both labeled and unlabeled regions of RNAs, the latter of which were synthesized and mostly spliced before the labeling, leading to potential bias in the rate estimate. To overcome this bias, an alternative approach (TT-seq) fragmented the labeled RNA before 4sU pull down to eliminate the unlabeled portion of RNA. Applying this approach to human K562 cells coupled with mathematical modeling, a recent study estimated that the *in vivo* rate for both lariat formation and exon ligation, concluding that splicing occurs on the order of minutes with a median of 7.2 minutes. The rate of *in vivo* splicing was also found to be influenced by both sequence features as well as the interaction of introns with the spliceosomal snRNAs. It is worth noting that fragmented RNAs in this study were mostly hundreds of nucleotides long, which led to bias against short RNA fragments and thus fast splicing events. It is also unclear whether 4sU labeling affects spliceosome assembly and catalysis due to altered base-pairing between U-rich RNA elements in introns and snRNAs (Testa et al., 1999).

Instead of measuring co-transcriptional splicing in terms of time, some studies have measured the timing of splicing relative to transcription progress, such as positions of RNAP II associated with each nascent RNA. In doing so, using single-molecule intron tracking (SMIT) and long-read sequencing in yeast, studies found that splicing is ~50% complete when RNAP II is 45 nts downstream of the 3' SS in yeast (Carrillo Oesterreich et al., 2016). Using a similar approach, a recent study concluded that splicing is also fast in mammals; 50% of splicing is complete when RNAP II is ~150 nts downstream of the 3' SS in murine cells (Reimer et al., 2020). Using direct RNA sequencing, a different study estimated that splicing in humans often completes when RNAP II is at least 4 kbs downstream of the 3' SS; whereas splicing in *Drosophila* completes when RNAP II is within 2 kbs downstream (Drexler et al., 2020). Both studies have identified sequence features influencing the timing of co-transcriptional splicing and observed out-of-order splicing.

Instead of relying on un-spliced and spliced nascent RNA transcripts, our work focused on nascent lariat intermediates to define the timing of co-transcriptional splicing. Compared to linear RNA counterparts, nascent lariat intermediates offer several advantages. First, every lariat intermediate is a splicing intermediate. Therefore, the lariat intermediate not only reports the position of RNAP II at its 3' end but also indicates that splicing is in progress; by contrast, linear RNA transcripts only report whether splicing has taken place or not. Second, the lariat intermediate reports on both the lariat formation and exon ligation steps of splicing, in contrast to mRNA, which reports on the combination of these two steps; specifically, the appearance of lariat intermediate reflects the lariat formation step and the disappearance of lariat intermediate reflects the exon ligation step. Since most splicing regulation occurs at the early stages of spliceosome assembly, the timing of lariat formation provides a more representative assay of splicing regulation than the timing of overall mRNA formation does, as the latter cannot accurately differentiate the

contribution of the timing of lariat formation and exon ligation on the timing of mRNA formation. Third, the lariat intermediate, unlike linear RNA transcripts, report BP usage, which has direct implications for 3' SS selection. Fourth, since mRNAs are abundant in cells, it is challenging to independently assay a relatively small fraction of nascent RNA transcripts. Contaminating mRNAs can complicate co-transcriptional analyses. In contrast, owing to their transient and intermediate nature, probing lariat intermediates allows examining nascent lariat intermediates independently from post-transcriptional lariat intermediates. Lastly, in part because over 90% of the BPs are within 50 nts upstream of the 3' SS, nascent lariat intermediates permit the interrogation of splicing events over a wide range of timeframes by short-read sequencing, thereby circumventing the current limitations on depth imposed by long-read sequencing. Based on these advantages, we developed co-transcriptional lariat-sequencing (CoLa-seq) to investigate the timing of co-transcriptional splicing. Using CoLa-seq, we observed not only in-order and out-of-order splicing but also concurrent splicing for the first time *in vivo.* Notably, we found that the timing at which splicing occurs can vary dramatically both across introns and even for the same intron, implying that the splicing of the same intron can occur both through an intron definition and exon definition pathway. By studying introns with varying timing of splicing, we identified key cis-sequence and trans-acting factors that determine early and late lariat formation. Remarkably, we found evidence that AG/U2AF1-independent introns can undergo ultra-fast lariat formation.

Collectively, these findings strengthen the proposal that understanding the timing of splicing relative to transcription will be essential before we can accurately describe the mechanisms of co-transcriptional splicing and its regulation.

Fidelity in splicing

Considering that splice sites contain minimal sequence information and are located at ends of introns that are often several kbs apart in humans, the spliceosome faces a formidable challenge in identifying and juxtaposing the appropriate splice sites to ensure accurate splicing. At the same time, the spliceosome must maintain significant flexibility to accommodate the extensive usage of alternative splice sites. Ultra-deep sequencing of a mini-gene reporter estimated that the overall rate of splicing error was maintained at a very low level, on the order of 10^{-5} (Reynolds and Hertel, 2019). Another analysis of RNA-seq data estimated the splicing error rate to be about 0.7% (Pickrell et al., 2010). Such high fidelity of splicing underscores its importance in gene expression. Indeed, it is estimated that up to 35% of disease-causing point mutations affect splicing (Manning and Cooper, 2017).

To ensure accurate splicing, there are several safeguards in play at both the splice site recognition step and later steps of the splicing cycle. The high-fidelity process of splice site recognition is ensured by multiple layers of parameters, such as *cis*-regulatory elements, *trans*-acting factors, transcriptional kinetics, nucleosome positioning, and epigenetic modifications (Braunschweig et al., 2013), some of which are discussed above. At later steps, core splicing factors appear to play major roles by actively discriminating and discarding suboptimal RNA substrates. In this section, I will focus on the roles of core splicing factors, such as DExD/H-box ATPases, in controlling splicing fidelity.

Multiple DExD/H-box ATPases act as proofreading factors by not only promoting splicing of optimal substrates but also antagonizing suboptimal substrates (Koodathingal and Staley, 2013). Prp16, the archetype of these ATPases, acts not only after the lariat formation step to remodel the spliceosome to adopt an exon ligation confirmation but also before the lariat formation step to

suppress splicing substrates with suboptimal BPs (Burgess and Guthrie, 1993; Koodathingal et al., 2010; Schwer and Guthrie, 1992). The dual function of Prp16 has led to a proposal that Prp16 proofreads the BPs through a kinetic proofreading mechanism. In kinetic proofreading, the productive pathway and the discard pathway compete with each other to enhance splicing fidelity (Semlow and Staley, 2012). This kinetic proofreading mechanism is also observed for Prp28, Prp5, Prp22, and Prp28, which proofread the 5' SS, the BP, and the 3' SS, respectively (Mayas et al., 2006; Xu and Query, 2007; Yang et al., 2013). Interestingly, a recent study found that Prp16 and Prp22 not only reject suboptimal substrates but also promote the selection of alternative splice sites, implying a relationship between proofreading and AS (Semlow et al., 2016).

Importantly, another ATPase Prp43 is required to dissociate the rejected suboptimal splicing substrates from the spliceosome by disassembling the spliceosome (Koodathingal et al., 2010; Mayas et al., 2010). Although nuclear quality control mechanisms have been proposed to retain immature mRNAs, these discarded suboptimal splicing substrates, including pre-mRNAs and splicing intermediates, are exported to the cytoplasm for degradation, as indicated by their translation and/or degradation by cytoplasmic nucleases, (Hilleren and Parker, 2003; Mayas et al., 2010). These observations raise questions about whether a robust quality control mechanism exists to distinguish against not only pre-mRNA that failed to engage the spliceosome but also pre-mRNA that did engage the spliceosome but subsequently suffered discard either at the pre-mRNA or intermediate stage. Further, the mechanism for exporting these species has remained unknown (see Chapter 3).

Overview of mRNA export

The hallmark of eukaryotes is the appearance of the nuclear envelope, which encloses the nuclear genome inside the nucleus and results in the compartmentalization of different cellular

activities. During gene expression, pre-mRNAs are transcribed and processed into mature mRNAs in the nucleus. The mature mRNAs are then exported through the nuclear pore complexes (NPCs) into the cytoplasm for translation. This nuclear mRNA export process is highly conserved and essential for gene expression and cellular functions (Köhler and Hurt, 2007). Further, the assembly of export competent mRNA ribonucleoprotein (mRNP) occurs co-transcriptionally; the general export receptor Mex67-Mtr2 complex (TAP-P15 complex in humans) is recruited to the transcribing RNA by export adaptor proteins (Ben-Yishay and Shav-Tal, 2019; Ashkenazy-Titelman et al., 2020).

During transcription, RNAP II recruits the evolutionarily conserved TREX complex, which consists of the THO complex (Hpr1, Mft1, Tph2, and Tho2), ATPase Sub2 (UAP56 in humans), and the mRNA export adaptor Yra1 (Aly/REF in humans; Chávez et al., 2000; Sträßer et al., 2002). The TREX complex has been implicated in transcription elongation, mRNA export, genome stability, and 3' end processing (guez-Navarro and Hurt, 2011; Tutucci and Stutz, 2011). Specifically, recruitment of Yra1 occurs via a multi-step process (Tutucci and Stutz, 2011). Yra1 is first recruited by Pcf11, which is a component of the 3' end processing complex CF1A and binds to the S2P-CTD of RNAP II (Johnson et al., 2009). Next, Pcf11 hands over Yra1 to Sub2. Subsequently, Sub2 transfers Yra1 to the transcribing RNA, after which Yra1 recruits Mex67-Mtr2. In addition, Dbp2 and the Prp19 complex were also found to help recruit Yra1 to the RNA (Chanarat et al., 2012; Ma et al., 2013). In addition to Yra1, Hpr1 mediates the co-transcriptional binding of Mex67 (Gwizdek et al., 2006). This interaction requires ubiquitylation of Hpr1 by E3 ubiquitin ligase Rsp5 and the C-terminal ubiquitin-associated (UBA) domain of Mex67 (Gwizdek et al., 2006, 2005).

Another export adaptor is Nab2, a polyA tail binding protein. Nab2 was found to directly interact with Mex67 and forms a stable ternary complex with Mex67 and Yra1(Iglesias et al., 2010). Further, Nab2 directly interacts with Mlp1, a component of the nuclear basket of the NPC (Deanna M. Green et al., 2003; Grant et al., 2008). Such interaction is thought to allow the exporting mRNP to dock to the nuclear basket once the mRNP reaches the nuclear envelope (see below).

The third general export adaptor Npl3, an SR-like protein, is preferentially recruited to the 5' end portion of the RNA likely through its direct interaction with serine 2 phosphorylated CTD (S2P-CTD) of RNAP II (Gilbert and Guthrie, 2004; Meinel et al., 2013; Baejen et al., 2014). As Npl3 is imported into the nucleus in a phosphorylated state, mRNA bound Npl3 is dephosphorylated by the nuclear phosphatase Glc7 so as to recruit the general receptor complex Mex67-Mtr2 (Gilbert and Guthrie, 2004). After the mRNP is exported into the cytoplasm, Npl3 is phosphorylated by Kinase Sky1, which dissociates Npl3 from the mRNA and Mex67-Mtr2, resulting in nuclear import of Npl3 for subsequent mRNA export (Gilbert et al., 2001). This successive phosphorylation and dephosphorylation of Npl3 are proposed to help enforce the directionality of the mRNA export pathway (Huang and Steitz, 2005).

In addition to these general export adaptors, Gbp2 and Hrb1 act as export adaptors for a subset of genes (Häcker and Krebber, 2004; Hackmann et al., 2014). Specialized recruitment of Mex67-Mtr2 was also found under stressed conditions. Interestingly, general export adaptors are not responsible for recruiting Mex67/Mtr2 complex under heat shock conditions. Mex67-Mtr2 is recruited to heat-shock mRNAs via a heat-shock transcription factor Hsf1 (Zander et al., 2016). Although there are only a few adaptor proteins discovered so far in yeast, there are likely more export adaptor proteins given the large number of RNA binding proteins in cells. In fact, SR

proteins in humans can function as export factors in addition to their essential functions in splicing (Ben-Yishay and Shav-Tal, 2019).

Once the mRNP reaches the nuclear envelope, the mRNP first encounters the nuclear basket of the NPC. Single particle tracking experiments implied that mRNP docks to the nuclear basket through the interaction betweenNab2, an export adaptor, and Mlp1 and Mlp2, components of the nuclear basket (Grant et al., 2008; Deanna M Green et al., 2003; Grünwald and Singer, 2010; Saroufim et al., 2015). The mRNP is thought to undergo further rearrangement to undock from the nuclear basket and enter the NPC channel for export. However, mechanisms for mRNP undocking remain poorly characterized. Previous studies have suggested that mRNP undocking requires the E3 ubiquitin ligase Tom1, which ubiquitylates Yra1, causing Yra1 to dissociate from nuclear mRNPs (Iglesias et al., 2010). Such Yra1 ubiquitylation is thought to take place on the nuclear basket, as deletion of the nuclear basket component *NUP60* led to the accumulation of Yra1 on mRNPs and deletion of another nuclear basket component *MLP2* suppressed the temperature sensitivity of *TOM1* (Iglesias et al., 2010; Lund and Guthrie, 2005). These observations support a mRNP undocking model, in which Tom1-mediated ubiquitylation of Yra1 causes Yra1 removal from the mRNP, triggering the mRNP to undock from the nuclear basket and then enter the NPC for export. In Chapter 3, we provide direct evidence for the support of this model.

Once the mRNP enters the NPC, the mRNP translocates across the NPC through the interactions of Mex67/Mtr2 with FG repeats that fill the central channel of the NPC (Ashkenazy-Titelman et al., 2020). Given the channel size of the NPC, the mRNP is expected to undergo extensive structural rearrangements while translocating through the pore. Recent structural studies revealed that cytoplasmic filaments formed by the Nup82 complex are positioned toward the central channel of the NPC (Fernandez-Martinez et al., 2016). Such positioning is proposed to help

capture exporting mRNP once it reaches the cytoplasmic face the NPC. Once captured, the DEAD-box ATPase Dbp5, in coordination with Gle1, remodels the mRNP (Tieg and Krebber, 2013). During remodeling, Mex67-Mtr2, Nab2, and other export factors are removed, thereby preventing the mRNA from traveling back to the nucleus. ATP hydrolysis by Dbp5 finally leads to the release of the mRNA into the cytoplasm (Tran et al., 2007). Notably, in yeast, the small molecule inositol hexakisphosphate (InsP6) is involved in activation of Dbp5 (Alcázar-Román et al., 2006; Weirich et al., 2006). In human cells, however, such a role for InsP6 is still controversial (Adams et al., 2017; Lin et al., 2018).

Previous studies have proposed that several mechanisms for the quality control of mRNA export, including the selective deposition of export factors and the nuclear retention of immature species mechanisms for quality control, the latter of which involves Mlp1, a component of the nuclear basket of the NPC. Reporter assays imply that Mlp1 acts as the general quality control factor to prevent faulty transcripts, such as pre-mRNAs, from export (Galy et al., 2004; Vinciguerra et al., 2005). However, through examining endogenous genes in different organisms, other studies found that Mlp1 instead promotes mRNA export (Aksenova et al., 2019; Bae et al., 2009; Deanna M. Green et al., 2003; Xu et al., 2007). Further, pre-mRNAs and splicing intermediates, such as free 5' exons and lariat intermediates, have been observed in the cytoplasm (Carvalho et al., 2017; Hilleren and Parker, 2003; Mayas et al., 2010; Sayani and Chanfreau, 2012). Disabling nonsense-mediated decay also stabilizes pre-mRNAs in the cytoplasm, indicating that unprocessed RNA transcripts can indeed export to the cytoplasm (Carvalho et al., 2017; Egecioglu and Chanfreau, 2011). These observations raise questions about a strict requirement for quality control in mRNA export and, specifically, the nuclear retention role of Mlp1. Further, the cytoplasmic localization of these incompletely processed RNA species raises questions concerning

the pathways utilized for export and their relation to mRNA export, specifically because a cleaved 5' exon lacks a poly(A) tail and a lariat intermediate lacks a cap, both features of mRNA implicated in promoting export. In chapter 3, we provide direct evidence that the export of lariat intermediates requires the mRNA export pathway, establishing that both mRNAs and lariat intermediates utilize the same export machinery. To our surprise, we found evidence that Mlp1 does not retain but instead promotes the export of lariat intermediates, consistent with its positive role in mRNA export observed in different organisms. Importantly, the efficient export of discarded suboptimal splicing substrates ensures that the spliceosome-discarded substrates do not re-enter the splicing pathway, thereby reinforcing the splicing fidelity.

CHAPTER 2

PROFILING NASCENT LARIAT INTERMEDIATES REVEALS THE GENOMIC BASIS OF SPLICING TIMING*

* This chapter is in preparation for publication and adapted from the unpublished manuscript "Profiling nascent lariat intermediates reveals the genomic basis of splicing timing", with authors Yi Zeng, Aiswarya Krishnamohan, Benjamin Fair, Huilin Zeng, Yichen Hou, Alexander J. Ruthenburg, Yang I. Li, and Jonathan P. Staley. The experiments described herein were performed by Y. Z. except those described in Fig. 2.2A, 2.3A, and 2.15, which were performed by Y.Z. and A.K.; those described in Fig. 2.2C-F, which was performed by Y.Z. and B. F.; those described in Fig. 2.3B, which were performed by Y. Z., A.K., and Y.H.; and those described in Fig. 2.7A, 2.8A, 2.8B, 2.12A, and 2.13A, which were performed by Y.Z. and H. Z..

Abstract

Pre-mRNA splicing is an essential step during eukaryotic gene expression and is intimately coupled with transcription both physically and functionally. To study the mechanisms behind this coupling in humans, we developed a high-throughput genomic approach called CoLa-seq, or co-transcriptional lariat sequencing, that reports the timing of lariat formation relative to transcription through analysis of nascent lariat intermediates. The capture of nascent lariat intermediates followed by short-read high-throughput sequencing allowed us to dramatically expand our annotation of human branch points in a cell type-specific manner (**Fig. 2.2F**). CoLa-seq data further allowed us to identify early splicing at an unprecedented resolution and observe *in vivo* concurrent splicing for the first time. Notably, we found that the timing at which splicing occurs can vary dramatically across introns and even for the same intron (**Fig. 2.5**). We also provide evidence that the splicing of a given intron can occur both through an intron definition and an exon definition pathway (**Fig. 2.10**), implying a potential mechanism for the regulation of alternative splicing. By analyzing introns with varying timing of splicing, we identified key *cis* element and *trans*-acting factors, such as intron size, regional GC content, and U2AF binding, that predict early and late splicing, explaining 30-40% of the variance in splicing timing (**Fig. 2.7; Fig. 2.12**). Remarkably, we found evidence that AG/U2AF1-independent introns can undergo ultra-fast lariat formation (**Fig. 2.14**). Together, these findings provide critical insights into human co-transcriptional splicing and establish a key framework for defining the mechanisms of splicing regulation.

Introduction

Pre-mRNA splicing is an essential step of gene expression, which converts pre-mRNAs to mRNAs by removing introns and ligating flanking exons. In humans, almost all pre-mRNAs

undergo alternative splicing (AS) to produce multiple mRNA isoforms, which greatly expand the human transcriptome and proteome. Further, AS is regulated in time and space, contributing a critical layer of regulatory control of gene expression beyond transcription. In fact, mutations or misregulation of splicing are associated with a large number of human diseases (Manning and Cooper, 2017). Importantly, *in vivo* splicing is tightly coupled to transcription. Therefore, it is of great importance to understand the mechanism of pre-mRNA splicing in the context of transcription.

Splicing is catalyzed by the spliceosome, a large and dynamic machine in which five small ribonucleoprotein particles (snRNPs) and numerous non-snRNP factors act together. The spliceosome recognizes conserved sequences at the 5’ splice site (SS), the branch point (BP), and the 3’ SS and catalyzes splicing in two sequential transesterification reactions (Kastner et al., 2019). The spliceosome assembly begins with the U1 snRNP binding to the 5' SS at the 5’ end of the intron, followed by SF1 binding to the BP near the 3’ end of the intron. Binding of SF1 is stabilized by U2 snRNP auxiliary factor (U2AF), a heterodimer of U2AF1 and U2AF2 that binds to the terminal AG dinucleotide at the 3' SS and its upstream poly-pyrimidine tract (PPT), respectively. Subsequently, the U2 snRNP replaces SF1 at the BP in an ATP-dependent manner. Efficient splice site recognition is further promoted by cooperative interactions between the U1 snRNP and U2AF/U2 snRNP (Chen and Manley, 2009). While the establishment of such interactions spanning a single intron is essential for subsequent steps, initial splice site recognition in multi-exonic human pre-mRNAs may also occur across exons in a process called exon definition. Since introns are much longer than exons in humans, exon definition is thought to be the predominant mode of splicing initiation (Berget, 1995; Hollander et al., 2016). In exon definition, the U1 snRNP and U2AF/U2 snRNP from adjacent introns are thought to interact across an intervening exon, which

helps define the 3’ SS of the upstream intron. This cross-exon interaction is often stabilized by SR proteins that bind to exonic sequences (Wu and Maniatis, 1993). The exon definition complex is eventually converted to the intron definition complex, establishing interactions between the U1 snRNP and the U2 snRNP across the upstream intron. Although it is proposed that the exon definition is the dominant pathway in humans, recent genomic studies have observed splicing via the intron definition pathway (Drexler et al., 2020; Reimer et al., 2020). However, it is unclear how prevalent the intron definition pathway is in humans.

Subsequently, the spliceosome is joined by the U4/U5.U6 tri-snRNP. Through extensive compositional and structural rearrangements promoted by multiple RNA helicases, the spliceosome is activated to catalyze two sequential transesterification reactions (Cordin and Beggs, 2013; Kastner et al., 2019). In the first step, referred to as lariat formation, the 2' hydroxyl of a conserved BP adenosine attacks the 5' SS phosphate, forming a branched lariat intermediate and free 5' exon. In the second step, referred to as exon ligation, the 3' hydroxyl of the cleaved 5' exon attacks the 3' SS phosphate, ligating the exons and excising the lariat intron.

Much of our knowledge about spliceosome assembly and catalysis have been acquired from *in vitro* studies. In the *in vitro* systems, splicing can occur in the absence of transcription. However, accumulating evidence indicates that *in vivo* splicing occurs predominantly during RNAP II-mediated transcription (Beyer and Osheim, 1988; Kessler et al., 1993; Tilgner et al., 2012). Importantly, splicing is tightly coupled to transcription and chromatin—transcription rate and chromatin influence not only splicing efficiency but also the outcomes of AS; likewise, splicing impacts RNAP II elongation and pausing (Braunschweig et al., 2013; Herzel et al., 2017). However, our understanding of co-transcriptional splicing remains incomplete in part due to limited information on the timing of *in vivo* splicing relative to transcription.

In contrast to synthetic pre-mRNA substrates used for *in vitro* studies, sequence elements appear at different times *in vivo* due to ongoing transcription. As a result, at any given moment, only a subset of splice sites and regulatory elements are available to the spliceosome and trans-acting splicing factors. Further, introns transcribed first are not always spliced first *in vivo*. In fact, frequent out-of-order splicing events have been observed and associated with specific sequence features (Kessler et al., 1993; Kim et al., 2017). Further, out-of-order splicing was found to bring regulatory elements from the downstream exon close to the upstream exon to promote its inclusion (Nasim et al., 1990). Thus, a key to fully understand splicing mechanisms and the impact of other processes on splicing is understanding the timing of splicing of a given intron relative to transcription and to adjacent introns.

Recent studies have come to different estimates of timing of *in vivo* splicing (Wachutka et al., 2019; Drexler et al., 2020; Reimer et al., 2020), underscoring the complexity of co-transcriptional splicing kinetics. It is also unclear if these estimates are free of biases introduced by preferential enrichment of U-rich unspliced transcripts or longer RNA fragments, limited depth of long-read sequencing, and/or potential contamination of mRNAs. Further, many unknowns remain, including the pathway of mRNA biogenesis, such as BP usage, the timing of lariat formation, and the timing of the transition between two steps of splicing.

In this study, we present a genome-wide approach, called CoLa-seq (co-transcriptional lariat sequencing), to study the dynamics and regulation of co-transcriptional splicing with single nucleotide resolution. This method exploits nascent lariat intermediates, which in particular encode BP usage and the position of RNAP II during transcription at the time of splicing. Using CoLa-seq, we observed not only in-order and out-of-order splicing but also *in vivo* concurrent splicing for the first time. Interestingly, AS associates with later timing of splicing, either concurrent or

out-of-order splicing, implying the potential impact of the order of splicing on alternative splice site choices. Strikingly, we found that the timing at which splicing occurs can vary drastically across introns and even for the same intron. Intriguingly, we found evidence that the splicing of a given intron can occur both through an intron definition and exon definition pathway, which can potentially regulate AS. Importantly, we identified key *cis* elements and *trans*-acting factors that correlate with the timing of lariat formation. Remarkably, we found evidence that AG/U2AF1-independent introns can undergo ultra-fast lariat formation. Together, these findings broaden our understanding of co-transcriptional splicing and how it impacts splicing outcomes.

Results

CoLa-seq captures co-transcriptional lariat intermediates and excised lariat introns

Previous investigations of the timing of co-transcriptional splicing have relied on assays of nascent pre-mRNA and mRNA; these assays have both advantages and disadvantages. Here, we have instead exploited the nascent lariat intermediate (NLI; **Fig. 2.1A**) to define the timing of co-transcriptional splicing for the following reasons. First, every lariat intermediate reports on the timing of splicing, because its 3' end reports on the position of RNAP II, as in previous assays, but the branch point (BP) of the same lariat intermediate inherently indicates that splicing is in progress; by contrast, unspliced and spliced nascent transcripts are generally not actively splicing, thereby only indicating that splicing has not yet occurred or already occurred, respectively. Second, the lariat intermediate reports on both the lariat formation and exon ligation steps of splicing, in contrast to mRNA, which reports on the combination of these two steps; specifically, the appearance of a lariat intermediate reflects the lariat formation step and the disappearance of lariat intermediate reflects the exon ligation step. Notably, the regulation of splicing is thought to focus generally on the earliest steps of splicing, so the lariat formation step of splicing offers a more

representative assay of regulatory steps than does overall mRNA formation. Third, lariat RNAs, unlike linear mRNA, report on the BP used during splicing catalysis, which has direct implications for 3' SS usage. Fourth, the abundance of mRNA in a cell presents a challenge to assay independently the small fraction of nascent mRNA, so contaminating mRNA can complicate co-transcriptional analyses (Drexler et al., 2020), and solutions involving metabolically labeling are potentially problematic (Testa et al., 1999); by contrast, the transient nature and low-abundance of the lariat intermediate minimizes any challenge of assaying nascent lariat intermediates independently from post-transcriptional lariat intermediates. Fifth, the unique chemical nature of the BP allows for enrichment of the lariat intermediate and consequently for molecules specifically reporting on the timing of splicing. This advantage circumvents the requirement for gene-specific amplification, which would limit the scope of a co-transcriptional splicing analysis in humans where there are >375,000 introns (Piovesan et al., 2016). Lastly and importantly, because splicing can generally occur only when the entire intron is transcribed and ~90% of BPs reside within the last 50 nts of introns, nascent lariat intermediates permit the interrogation of splicing events over a wide range of timeframes via short-read sequencing (see below), thereby circumventing the current limitations on depth imposed by long-read sequencing.

To investigate the timing of pre-mRNA splicing genome-wide in humans, as well as branch point usage, we developed a novel technique, CoLa-seq, to capture and sequence nascent lariat intermediates (**Fig. 2.1B**). CoLa-seq begins with enrichment for nascent RNAs through their tight association with chromatin (**Fig. 2.2A**). To enrich further for nascent lariat intermediates, we degrade linear RNAs by treating nascent RNA with the decapping enzyme, RppH, and the 5'-to-3' exoribonuclease, XRN-1. Next, to capture RNAP II positions for nascent lariat intermediates, we ligate an adaptor to the 3' ends of our isolated RNAs, an adapter with a unique molecular identifier

(UMI) to correct for PCR duplication and RT mis-priming. Then, to capture the branch point of lariat intermediates, the signature of this species, we reverse transcribe nascent RNA under standard conditions, which stops reverse transcriptase at the 2'-5' linkage of the BP. Finally, we construct a sequencing library comprising PCR products of 150-750 bps for 100 bp paired-end sequencing. As a result, a CoLa-seq read for a given lariat intermediate captures both the BP location and the position of RNAP II at a point in time of transcription after branching but before excision of the intron (**Fig. 2.1A, B**). In addition to nascent lariat intermediates, CoLa-seq captures fully excised lariat introns, which, like nascent lariat intermediates, have a BP and a free 3' hydroxyl at the 3' ends and remain associated with chromatin at least until the spliceosome releases nascent mRNA (**Fig. 2.1A, B**).

After applying CoLa-seq to K562 cells, we deduplicated uniquely mapped CoLa-seq reads and sought to verify that they were enriched for reads derived from lariat RNA species. To this end, we analyzed the distribution of RNA 5' ends from all reads around annotated BPs (Pineda and Bradley, 2018). We observed that the 5' ends peaked 1 nt downstream of known BPs, confirming that CoLa-seq specifically enriched for lariat RNA species (**Fig. 2.1C**). Consequently, we developed a computational pipeline to systematically identify reads starting at BPs and to classify reads as deriving from nascent lariat intermediates (NLIs) or excised lariat introns (ELIs). First, we filtered for and classified reads as (i) putative NLI reads if they traversed 3' SSs and ended downstream of 3' SSs and (ii) putative ELI reads if they ended at the 3' SS, the last nucleotide of the intron. Remarkably, through this simple classification, 5' ends of putative ELI reads already resembled known BPs in terms of sequence motifs (i.e., resemblance to YUNAN consensus; Gao et al., 2008) and positions relative to 3' SSs (i.e., with a median position at 31 nts upstream of 3' SS; **Fig. 2.2B, bottom panels**), suggesting that putative ELI reads derive largely

from excised lariat introns. Similarly, 5' ends of putative NLI reads resembled known BPs (**Fig. 2.2B; top panels**), although to a lesser degree, implying that many but not all putative NLI reads derived from authentic nascent lariat intermediates. Thus, as the second step, we built a classifier to annotate authentic BPs (see Methods), which we further used to filter for lariat RNA-derived NLI and ELI reads. This classifier integrates motif and positional constraints derived from published BPs (Pineda and Bradley, 2018) with CoLa-seq data, which independently reports on BP usage (**Fig. 2.2C, D**). In total, our classifier identified 154,187 high-quality BPs (false positive rate, 1%; **Fig. 2.2E**), with expected BP characteristics (**Fig. 2.1D**). Remarkably, the classifier identified more BPs from CoLa-seq libraries from a single cell-type than the most recent large-scale study (**Fig. 2.2F**), which used 1000 times as many reads pooled across 17,164 of RNA-seq libraries from 65 studies of diverse tissues and cell-types (Pineda and Bradley, 2018); the CoLa-seq BPs overlap with previously experimentally identified BPs to a similar extent as this recent study, implying that the CoLa-seq BP annotations are of similar quality as previous studies. Of the CoLa-seq BPs, 66% (101,792) were novel in comparison to previously identified BPs combined by large scale studies (Briese et al., 2019; Mercer et al., 2015; Pineda and Bradley, 2018; Taggart et al., 2017). These novel sites show a consensus motif and distribution relative to the 3' SS that is similar to annotated sites (**Fig. 2.2G**); indeed, we verified 14,060 BPs, including 5,733 novel ones, by lariat-sequencing, a direct but less efficient method for identifying BPs through inverted reads that cross the branched structures in lariat RNA (Mayerle et al., 2017). Overall, this analysis establishes the utility of CoLa-seq data in defining BPs, and these BP annotations allowed us to extensively identify NLI and ELI reads.

We detected NLI and ELI reads primarily from protein coding genes, as anticipated, and with strong reproducibility between biological replicates (Pearson's $r > 0.96$; **Fig. 2.1E, top**

panels). Further, we confirmed that NLI and ELI reads depend on spliceosome assembly by showing that the corresponding 5' and 3' read are drastically diminished from cells that were treated with splicing inhibitor Pladienolide B (**Fig. 2.3A, B**). Lastly, we demonstrated that NLI and ELI reads are both substantially reduced from cells treated with the transcription elongation inhibitor flavopiridol, establishing the nascent nature of the RNAs yielding NLI and ELI reads (**Fig. 2.3B**). Importantly, of NLI-associated BPs, 69% overlapped with ELI-associated BPs (**Fig. 2.2I**). Despite expectations that not all NLI reads and ELI reads associated with the same BPs should correlate (**Fig. 2.2H, J**), we see a modest correlation of read counts (Pearson's *r* 0.32), consistent with the functional conversion of NLI species to ELI species.

NLI and ELI reads derived from a significant fraction of human introns with many introns yielding abundant reads (**Fig. 2.1E, bottom panels**). In total, we obtained 2,201,036 NLI reads associated with 73,562 3' SSs (>= 3 NLI reads), and we obtained 2,574,753 ELI reads associated with 57,309 3' SSs (>= 3 ELI reads). In a subset of genes, we detected NLI reads and ELI reads from every intron (2,941 and 1,432 gene major isoforms, respectively). The depth and breadth of these datasets allowed us to study BP selection and co-transcriptional splicing with unprecedented resolution.

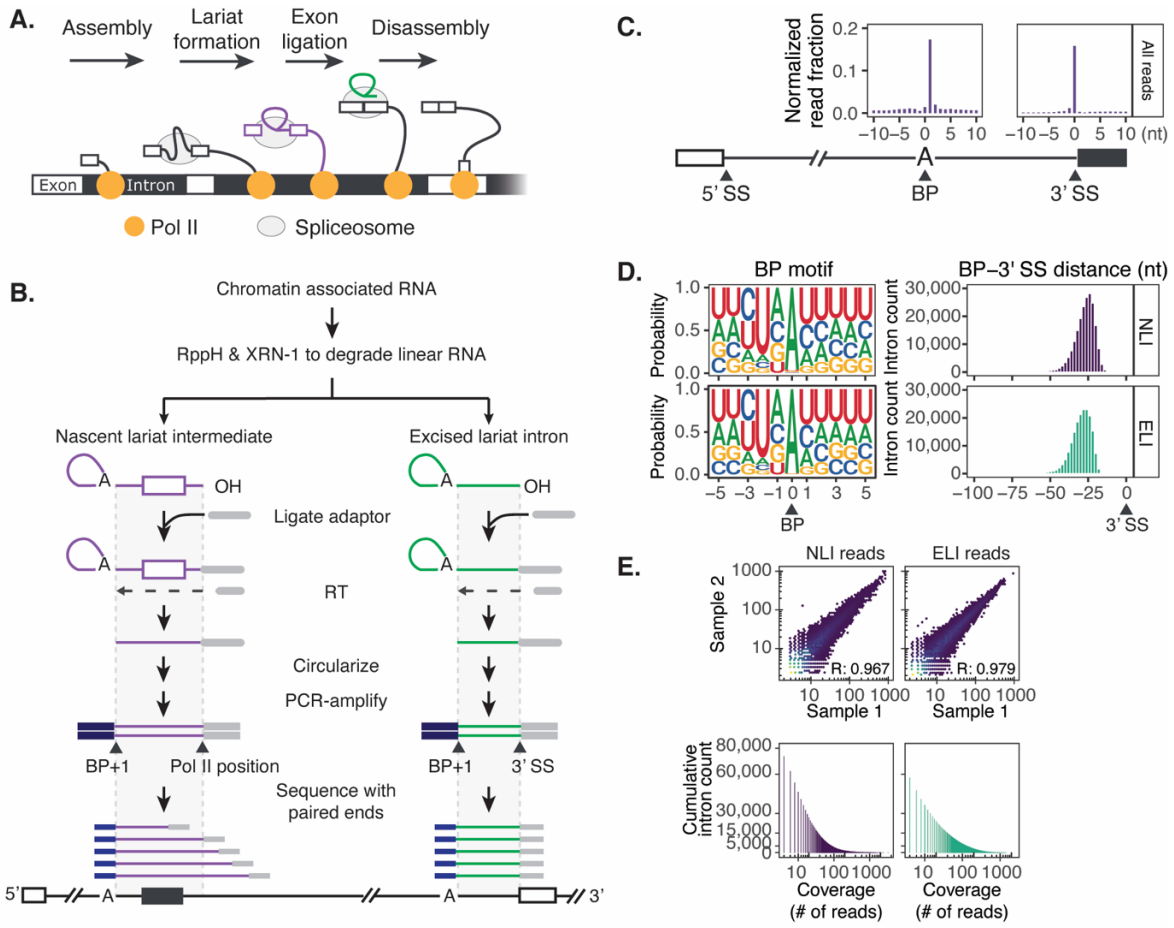


Figure 2.1. CoLa-seq captures co-transcriptional lariat intermediates and excised lariat introns.

A. Schematic view of the co-transcriptional splicing process. 5' exon and lariat intermediate are colored in purple; excised lariat intron is colored in green. **B.** Schematic overview of the CoLa-seq procedure. Chromatin-associated RNA is treated with RppH and XRN-1 to degrade linear RNAs and enrich for lariat RNAs: nascent lariat intermediates (NLIs) and excised lariat introns (ELIs). Next, the 3' ends of these lariat RNAs are ligated to an adaptor to capture pol II positions or 3' SSs. Branch point (BP) positions are captured during cDNA synthesis because the branched structures in the lariat RNAs terminate reverse transcription. cDNA circularization and PCR amplification generate the final library for paired-end sequencing. The resulting reads are classified as NLI reads or ELI reads for downstream analysis. **C.** Unfiltered CoLa-seq reads are enriched for lariat RNA species. Bar plots illustrate that the 5' ends of uniquely mapped reads peaked 1nt downstream of annotated BPs and that their 3' ends peaked at the last nucleotides of introns. Only constitutive 3' SSs were used. +/- 10 nts were shown for each region. **D.** NLI reads and ELI reads exhibit expected BP characteristics. Seqlogo plots (left panels) to illustrate BP motifs from NLI reads and ELI reads, respectively. Histograms (right panels) plot BP-3' SS distance distribution of NLI reads and ELI reads, respectively. **E.** CoLa-seq is highly reproducible (top panels) and yielded good coverage across many introns (bottom panels) for both NLI and ELI reads. Scatter plots (top panels) compare number of reads at individual 3' splice sites (>= 3 reads) from two biological replicates. Pearson's correlation coefficient (R) is shown. Cumulative bar plots (bottom panels) to show cumulative intron counts with increasing coverage thresholds.

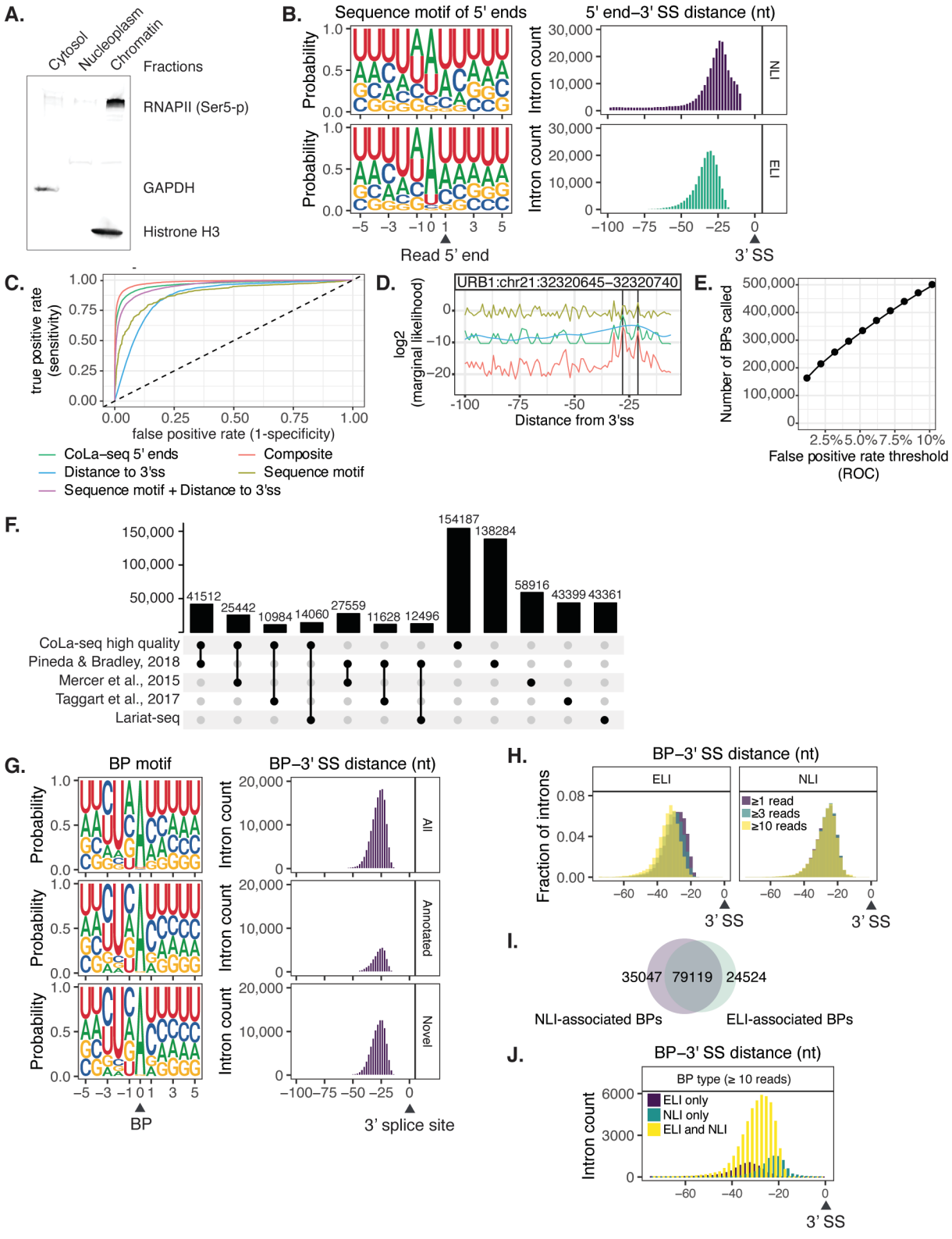


Figure 2.2. CoLa-seq maps BPs in human genome to an unprecedented depth.

A. Western blot confirmed successful subcellular fractionation. GAPDH was probed as a cytoplasmic marker. Serine 5 phosphorylated RNAP II and histone H3 are probed as chromatin

Figure 2.2. (continued) markers. **B.** Seqlogo and histograms illustrate that 5' ends of putative NLI and ELI reads resemble known BPs. **C.** ROC (receiver operating characteristic) plot illustrates that the contribution of different features on calling BPs. CoLa-seq data alone outperformed positional and motif information at identifying known BPs. **D.** A specific example to illustrate the contribution of individual features on calling BPs. Features are the same in C. **E.** Point plot depicts the number of BPs identified at different false positive rates. **F.** Upset plot depicts the overlap of the BPs identified in CoLa-seq and those identified in previous studies. **G.** Seqlog and histograms illustrate BP motif and BP-3' SS distance distributions for all CoLa-seq identified (all), previously annotated (annotated), and unique CoLa-seq identified (novel) BPs, respectively. **H.** Histograms illustrate BP-3' SS distance distributions for BPs containing different numbers of ELI or NLI reads, respectively. **I.** Venn diagram illustrates the overlap between NLI-associated BPs and ELI-associated BPs. **J.** Histograms illustrate BP-3' SS distance distributions for the BPs that were observed in both NLI and ELI reads (ELI and NLI), only in ELI reads (ELI only), and only in NLI reads, respectively.

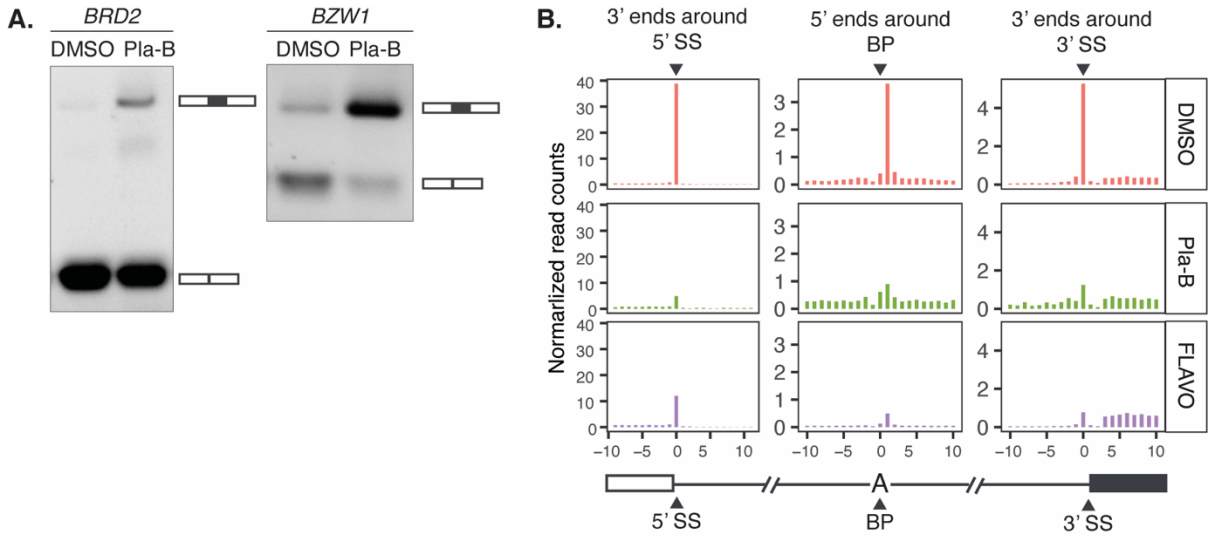


Figure 2.3. NLI and ELI reads depend on spliceosome assembly and transcription.
A. RT-PCR confirmed that Pladienolide B treatment inhibited splicing. **B**. Lariat reads captured by CoLa-seq depend on spliceosome assembly and transcription. CoLa-seq was performed on cells treated with DMSO, Pladienolide B (Pla-B), or Flavopiridol (FLAVO). Bar plots illustrate that the 3' ends of CoLa-seq reads in regions around the 5' and 3' SS (+/- 10 nts), representing cleaved 5' exons and ELIs, respectively, and their 5' ends in regions around the BPs (+/- 10 nts), representing NLIs and ELIs.

Excised lariat introns reveal coupling between BP and 3' SS usage

The direct linkage between BP and 3' SSs in ELI reads not only facilitates BP annotation, relative to 3' SSs, but also reports on the functional coupling, or lack thereof, between BP and 3' SS selection. We will report more deeply on the insights revealed by ELI reads elsewhere, but we highlight two findings here. First, ELI reads revealed cases of strict coupling between BP choice and 3' SS choice. For example, for two alternative 3' SSs in the 12th intron of the gene *RNF111*, the upstream 3' SS always coupled to an upstream BP, whereas the downstream 3' SS always coupled to a downstream BP (**Fig. 2.4A**); in this case, the upstream BP is likely too far from the downstream 3' SS to allow their coupling, and the downstream BP is too close to the upstream 3' SS to allow their coupling. Although recognition of the 3' SS by U2AF can impact BP selection prior to lariat formation, here BP choice must ultimately drive 3' SS selection, which follows lariat formation. Second, ELI reads revealed that nearby alternative 3' SSs are chosen independent of the BP. Specifically, when two competing 3' SSs were within 12 nts, 85% of the 3' SS pairs used the same set of BPs (**Fig. 2.4B, C**), indicating that 3' SS choice occurred likely during exon ligation. Thus, ELI reads distinguish between alternative 3' SSs that are chosen prior to the lariat formation step from those that are chosen at the exon ligation step.

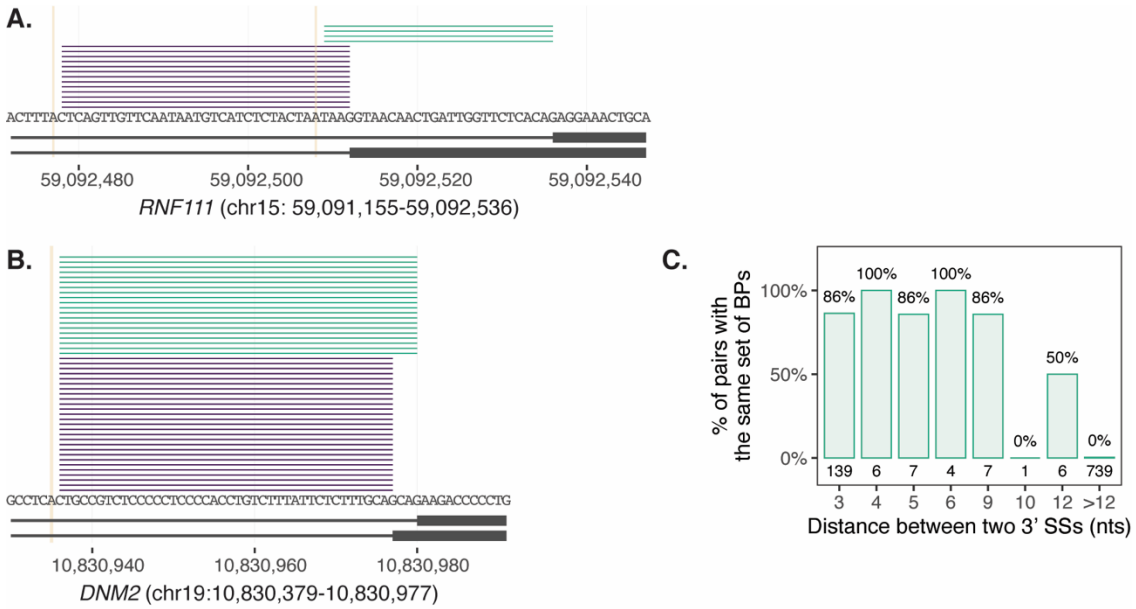


Figure 2.4. Excised lariat introns reveal coupling between BP and 3' SS usage.
A. BP choice enforces 3' SS selection. ELI reads are visualized for the 12th intron of *RNF111* gene. **B. C.** Nearby alternative 3' SSs are chosen independent of the BP. In panel **B**, ELI reads are visualized for the 20th intron of *DNM2*gene. In panel **C**, bar plot illustrates the percentages (labeled on top of each bar) of two alternative 3' SSs using the same set of BPs with increasing distances between the two 3' SSs. The number of 3' SS pairs at each distance is shown at the bottom of each bar.

CoLa-seq reveals in-order, out-of-order, and concurrent splicing

Because previous studies have shown that introns are not necessarily spliced in a 5' to 3' direction (Kessler et al., 1993; Kim et al., 2017), we first interrogated the timing of splicing of adjacent introns to investigate the timing of co-transcriptional splicing. Specifically, for NLI reads indicating an intermediate state of splicing for an upstream intron, we interrogated the splicing status of the downstream intron. We observed three classes of NLI reads, reflecting three different states of splicing order: (i) in-order splicing, (ii) out-of-order splicing, and, unexpectedly, (iii) concurrent splicing, each indicating different timings of splicing, for a given upstream intron (**Fig. 2.5A, B; Fig. 2.6A-D**). In-order splicing, the class representing the earliest splicing timing for the upstream intron, is reflected by NLI reads that indicate an unspliced downstream intron. These NLI reads could contain either a 3' end upstream of the downstream 3' SS, indicating that RNAP II has not yet completed the transcription of the downstream intron, or a transcribed downstream intron that remains unspliced (**Fig. 2.5A, B**). We observed in-order splicing in 93% of detected introns, and across all detected introns we observed a median value of 43% in-order splicing (**Fig. 2.5C; Fig. 2.6B**). Out-of-order splicing, the class representing the latest splicing timing for the upstream intron, is reflected by NLI reads that indicate a spliced downstream intron, signifying transcription and excision of the downstream intron and ligation of the flanking exons (**Fig. 2.5A, B**). While CoLa-seq is biased toward early splicing events, we observed out-of-order splicing in 46% of detected introns, and across all detected introns we observed a median value of 12% out-order splicing (**Fig. 2.5C; Fig. 2.6B**).

Strikingly, we have also observed concurrent splicing, when two adjacent introns splice at the same time. To our knowledge, concurrent splicing of adjacent introns has only been observed for a synthetic substrate (Christofori et al., 1987), but has not only been detected directly *in vivo*.

Concurrent splicing, the class representing a late timing of splicing for the upstream intron, is reflected by NLI reads that end at the last nucleotide of the intervening exon, indicating that the downstream intron is transcribed, undergoing splicing, and at the intermediate stage (**Fig. 2.5A, B**). Thus, concurrent splicing represents a special scenario when both the upstream and downstream introns are at the lariat intermediate stage. Indeed, a meta-analysis of all NLIs aligned to the downstream 5' SS reveals a prominent peak of 3' NLI ends at the last nucleotide of the exon (**Fig. 2.5D; Fig. 2.6E**). Remarkably, we observed concurrent splicing in 90% of detected introns, and across all detected introns we observed a median value of 53% concurrent splicing (**Fig. 2.5C; Fig. 2.6B**).

Importantly, both out-of-order and concurrent splicing allows us to investigate late splicing events, in which the distances between the 3' SSs and positions of RNAP II are longer than the size limitations imposed by short-read sequencing. Concurrent splicing is particularly informative because cleavage of the 5' SS of the downstream intron both (i) indicates that RNAP II has transcribed the downstream intron and must be positioned somewhere downstream of this intron and (ii) shortens the NLI to well within the range of short-read sequencing, given the median size for exons is 133 nts in humans (Piovesan et al., 2016). Notably, we observed concurrent splicing for adjacent intron pairs in which the downstream intron sizes span four orders of magnitude – from 56 nts to 178,737 nts, with a median of 1,095 nts (**Fig. 2.5E**), indicating that the timing of lariat formation varies significantly across introns. Assuming that the average transcription rate is 2 kb/min (Veloso et al., 2014) and the median exon size is 133 nts, our data indicate that the splicing could occur as fast as 5 seconds or as long as 9 minutes, with a median at 37 seconds.

As the timing of splicing has been found to influence AS (Fong et al., 2014; Aslanzadeh et al., 2018), we next examined the order of splicing of adjacent introns flanking cassette exons that

are preferentially included in K562 cells (Yee et al., 2019). Compared to introns upstream of constitutive exons, introns upstream of cassette exons had a lower level of in-order splicing but a higher level of concurrent splicing (**Fig. 2.5F**), indicating that these introns splice later relative to their transcription. Similarly, compared to constitutive 3' SSs, alternative 3' SSs are associated with a comparable level of in-order splicing, a lower level of concurrent splicing, but a higher level of out-of-order splicing (**Fig. 2.5G**), indicating that alternative 3' SSs also splice later. Interestingly, the 3' SSs in introns upstream of cassette exons and alternative 3' SSs are weaker than constitutive 3' SSs (**Fig. 2.6F**), consistent with previous findings that AS is usually associated with weak splice signals (Clark and Thanaraj, 2002). Because of weak splice signals, splicing of the upstream intron is slowed down so that all possible splice sites are transcribed and available for selection at the time of spliceosome assembly and splicing catalysis. In the case of cassette exons, delayed splicing of the upstream intron ensures that the 5' SSs from both upstream and downstream introns are available for selection by the time the 3' SS of the downstream intron has been transcribed, enabling skipping of an exon. However, since concurrent splicing leads to exon inclusion, changing the level of concurrent splicing can influence the level of exon skipping. In the case of alternative 3' SSs, it is plausible that the out-of-order splicing brings regulatory elements in the downstream exons close to the upstream intron and thus influence alternative 3' SS selection.

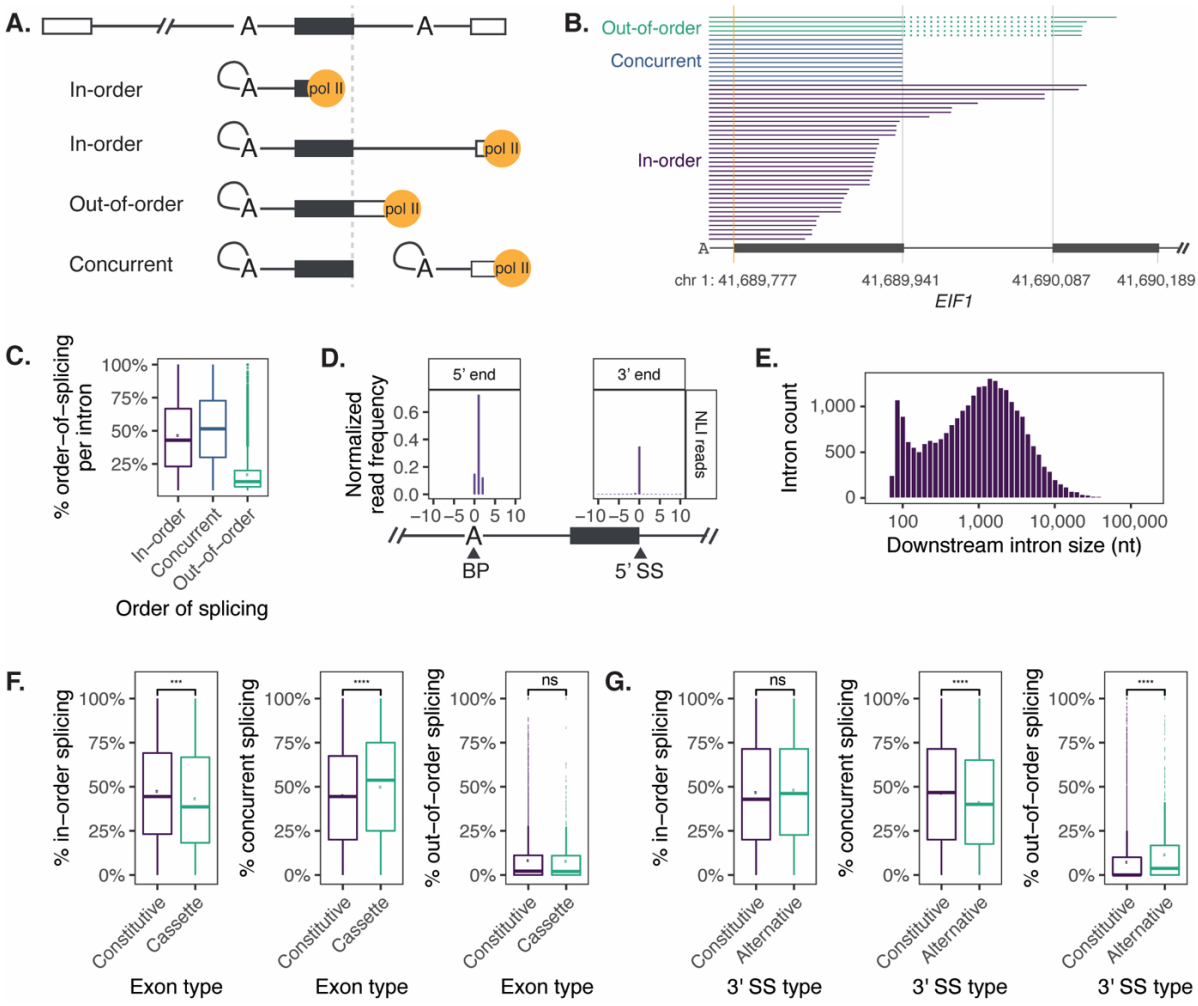
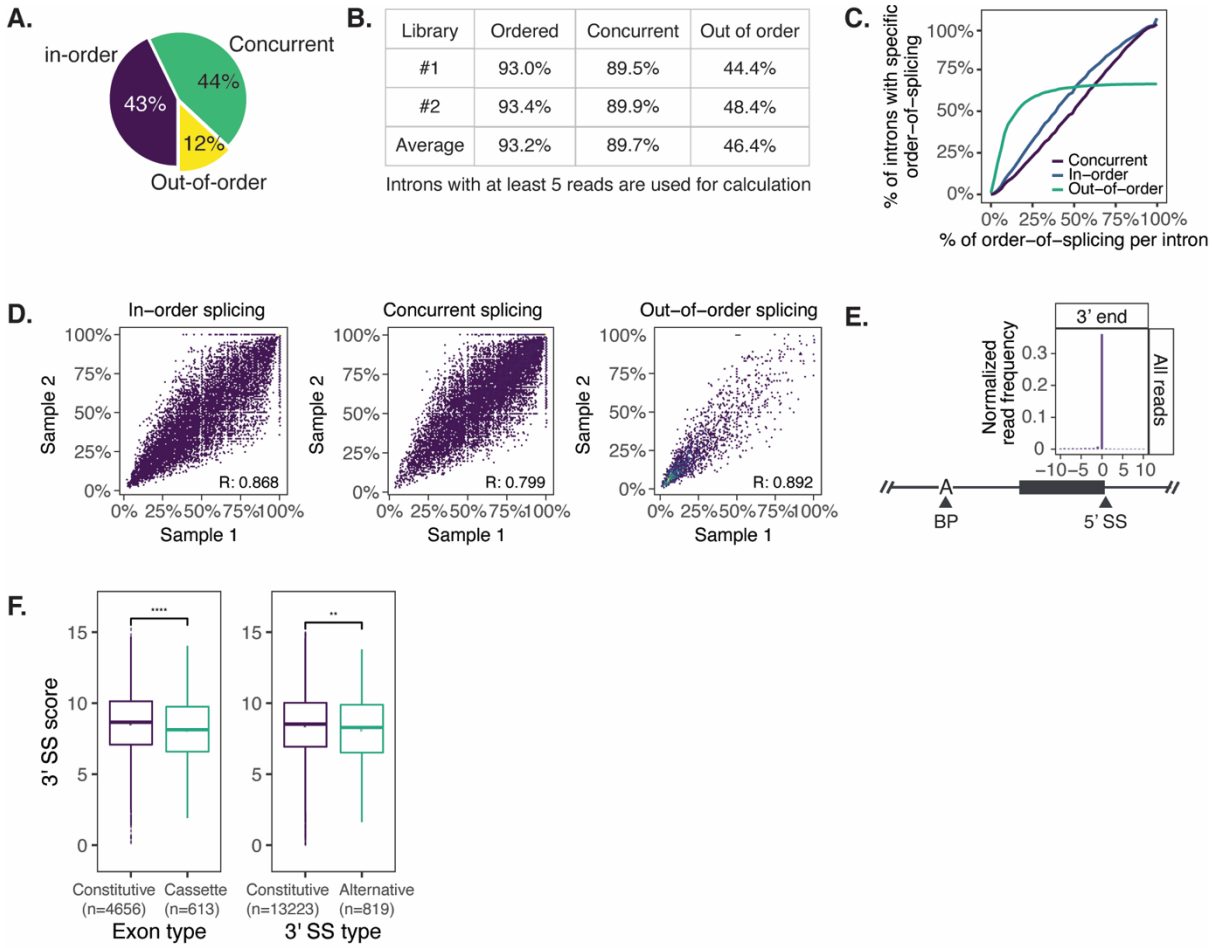


Figure 2.5. CoLa-seq reveals in-order, out-of-order, and concurrent splicing.
A. Schematic view of the relative order of splicing for adjacent introns: in-order splicing, out-of-order splicing, and concurrent splicing. **B.** NLI reads corresponding to different classes of splicing order are visualized for intron 1 of *EIF1* gene. **C.** Introns have a mixed pattern of splicing order and employ each of the splicing order with a different frequency. Box plot illustrates usage distributions of in-order splicing, concurrent splicing, and out-of-order splicing per intron. **D.** A meta-analysis of all NLIs aligned to the downstream 5' SS reveals a prominent peak of 3' NLI ends at the last nucleotide of exons. Bar plots illustrate that the 5' ends of NLI reads around annotated BPs and their 3' ends around the downstream 5' SS. (+/- 10 nts were shown for each region). **E**. Histogram illustrates the wide-ranging distribution of downstream introns in the concurrent splicing intron pairs. **F.** Box plots illustrate that, compared to introns upstream of constitutive exons (n=4,692), introns upstream of cassette exons (n=621) are associated with a lower level of in-order splicing, a higher level of concurrent splicing, and a similar level of out-of-order splicing. **G.** Box plots illustrate that, compared to constitutive 3' SSs (n=13,300), alternative 3' SSs (n=831) are associated with a comparable level of in-order splicing, a lower level of concurrent splicing, and a higher level of out-of-order splicing. The *p*-values were calculated by Mann-Whitney test; ns (not significant), $p > 0.05$; *, $p \leq 0.05$; **, $p \leq 0.01$; ***, $p \leq 0.001$; ****, $p \leq 0.0001$.



Library	Ordered	Concurrent	Out of order
#1	93.0%	89.5%	44.4%
#2	93.4%	89.9%	48.4%
Average	93.2%	89.7%	46.4%

Figure 2.6. CoLa-seq reveals different classes of the order of splicing.
A. Pie chart illustrates the percentage of NLI reads corresponding to each class of splicing order. **B.** Percentage of introns within each class of splicing order in two biological replicates. **C.** Cumulative plot illustrates the distribution of each class of splicing order per intron. **D.** Scatter plot illustrates that the distribution of each class of splicing order is highly reproducible between two biological replicates. **E.** Bar plot illustrates 3' ends of all reads around the downstream 5' SS (+/- 10 nts). **F.** Box plot illustrates that in the CoLa-seq dataset 3' SSs associated with AS are weaker than constitutive 3' SSs, as expected. The *p*-values were calculated by Mann-Whitney test.

Intronic elements, gene architecture and genomic context predict the order of co-transcriptional splicing

To search for genetic attributes associated with a particular order of splicing, we applied a gradient boosting based algorithm, XGBoost (Chen and Guestrin, 2016). Briefly, XGBoost uses decision tree ensembles to identify key features among a large set of features that associate with a given target variable. Using 22 pre-selected features, we separately trained models for in-order splicing, concurrent splicing, and out-of-order splicing; to focus on differences between these classes, we normalized in-order splicing relative to all classes of splicing and compared concurrent splicing relative to in-order splicing and out-of-order splicing relative to in-order splicing. Overall, the resulting models accounted for 39.8%, 38.9%, and 30.7% of the variance for in-order splicing, concurrent splicing, and out-of-order splicing, respectively. These XGboost-based models outperformed linear regression-based models by 44-241% (**Fig. 2.8A**).

Within each trained model, we quantified the contribution of each feature using SHAP (SHapley Additive exPlanation), which assigns each feature an importance value for a particular prediction (Lundberg and Lee, 2017). Three general classes of elements emerged as predictive for the order of splicing: (i) canonical intronic splicing elements, especially the 3' splice site, (ii) the sizes of introns and exons, and (iii) GC content (**Fig. 2.7A; Fig. 2.8B**). These features appear to function by promoting or antagonizing splicing of either the upstream or the downstream intron.

As expected, the strength of canonical intronic splicing elements helps predict the distribution of NLIs between in-order, concurrent, and out-of-order splicing. These elements include the 5' splice site, BP-3' SS distance, poly-pyrimidine tract (U content), and 3' SS (**Fig. 2.7A; Fig. 2.8B, C**). Further, the mean free energy (MFE) of pairing between the 5' SS and U1 or U6 as well as between the BP and U2 help predict NLI distribution between classes (**Fig. 2.7A;**

Fig. 2.8B, C). Specifically, a strong upstream 5' SS score correlates with in-order splicing, relative to out-of-order splicing, whereas a strong downstream 5' SS score correlates with out-of-order splicing, relative to in-order splicing. In parallel, the MFE of U1/5' SS base pairing for the downstream intron correlates with concurrent splicing, relative to in-order splicing. Additionally, a strong upstream 3' SS score correlates with in-order splicing, relative to all NLI, whereas a strong downstream 3' SS score correlates with out-of-order splicing, relative to in-order splicing. Consistent with this observation, the strongest predictor of the order of splicing is the %U in the BS-3'ss region; a high percentage of U in this region correlates with in-order splicing, relative to all NLIs, presumably reflecting the importance of PPT tracts. In general, the 3' SS score provides more predictive power than the 5' SS score, perhaps because the 3' SS is transcribed long after the 5' SS, imposing greater constraints on 3' SS recognition for early splicing (see below). The predictive value of these elemental features validates the model, that the relative order of splicing reflects differences in splicing timing between the upstream and downstream introns. Nevertheless, these elemental intronic elements are not the strongest predictors of splicing order.

One of the features that most strongly predicts the order of splicing is the size of the downstream intron (average impact size: 6.4%), with longer introns favoring in-order splicing, relative to concurrent and out-of-order splicing (**Fig. 2.7B; Fig. 2.9A**). We infer that this predictive power indicates that splicing timing of the downstream intron, relative to the upstream intron, is generally delayed by the time required to transcribe the downstream intron, such that longer downstream introns require more time to be transcribed. This would reduce the possibility of concurrent and out-of-order splicing. This inference validates that CoLa-seq reports on splicing timing relative to transcription.

The upstream intron size also predicts the order of splicing, although to a lesser extent (0.8%), with shorter introns favoring in-order splicing, relative to concurrent splicing and out-of-order splicing (**Fig. 2.7B, C; Fig. 2.8C**). The correlation between upstream intron size and in-order splicing is particularly strong for introns less than 250 nts (**Fig. 2.7C**). Interestingly, this trend reverses at 100 nts (**Fig. 2.7C**), which corresponds to intron sizes recently reported to require dedicated factors, Smu1/Red1, for efficient splicing (Keiper et al., 2019); indeed, we find that Smu1/Red1-dependent introns are more likely found in this reversed phase of the plot (**Fig. 2.9B**). The correlation between the downstream intron and concurrent splicing demonstrated a similar trend including both the sharp rise in SHAP values as introns shorten below 250 nt as well as the reversal at the shortest intron sizes (<100 nts; **Fig. 2.9C**). These observations validate that our model not only has the predictive power for a population of observed introns but also can resolve the contribution of features on the timing of splicing at the level of individual introns.

Surprisingly, exon size emerged as the strongest predictor of in-order splicing, relative to concurrent and out-of-order splicing, accounting for 6.5% of the model's predictive power. Although the median exon size is 10x shorter than the median intron size in humans and varies by only 1.4 fold across quintiles of in-order splicing (**Fig. 2.7B**), increasing exon size could correlate with in-order splicing by delaying downstream intron splicing just as increasing downstream intron size likely does – by increasing the time delay between transcription of the upstream and downstream introns. Indeed, exons have been implicated in slowing transcription (Wissink et al., 2019) to an extent that would yield a time delay on par with the time delay imposed by transcription of a median-sized intron. Curiously, the correlation of in-order splicing with exon size indicates that introns in the highest quintile have a median exon size on par with the average exon in the human genome, while those in lower quintiles have less-than-average sizes (**Fig. 2.7B**), implying

that average exons are optimized at least in terms of size for in-order splicing and further that shorter exon sizes may poise associated elements within them to regulate splicing outcomes. The greater apparent importance of transcription over intronic elements highlights transcription as a mechanism for AS regulation, as has been proposed (Saldi et al., 2016).

Among the features that anti-correlate with in-order splicing, and correlate with concurrent splicing, GC content is a top contributor (**Fig. 2.7A; Fig. 2.8B**). In-order splicing anti-correlated with higher GC content in the upstream intron, exon, as well as the downstream intron (**Fig. 2.8C**), suggesting that higher GC content of introns showing low in-order splicing extended to the downstream exon and intron. Indeed, the GC content of these three regions within the same gene is highly correlated (**Fig. 2.9D**). A number of studies have found that GC content bias can extend several kilobase pairs, in regions termed isochores (Costantini et al., 2006; Costantini and Musto, 2017). Further, one study showed that exons and flanking introns can be separated into two general classes – those with higher GC content that is relatively uniform across exons and introns and those with lower GC content that is even lower in the flanking introns than the exons (Amit et al., 2012); our analysis indicates that an intron of the lower GC class correlates with high levels of in-order splicing, whereas the higher GC class correlates with lower levels of in-order splicing (**Fig. 2.7D**). Additionally, a recent study identified GC-rich cassette exons associated with high GC isochores and AT-rich cassette exons associated with low GC isochores; each class of exons associates with distinct splicing regulators (Lemaire et al., 2019). Further, we found that introns upstream of AT-rich exons have higher than average in-order splicing and that introns upstream of GC-rich cassette exons have lower than average in-order splicing (**Fig. 2.7E**). Together, these results imply that the broader genomic context of an intron dictates splicing timing, which has direct implications for splicing regulation.

The higher GC content of low in-order splicing introns has the potential to impact splicing timing in several ways that each impact splice sites. First, low in-order splicing is associated with weaker intron consensus signals in the upstream as well as downstream introns, as noted above, and the higher GC content of these introns rationalizes the weaker splice sites, which show increased G and C in the consensus sites (data not shown). Second, GC content 50 nts around splice sites in our dataset is strongly correlated with the mean free energy of potential secondary structure (Pearson's *r*, 0.75 and 0.60 for the 5' and 3' SSs, respectively). Strong RNA secondary structure around splice sites correlates inversely with in-order splicing, presumably reflecting sequestration of the splice sites (**Fig. 2.9E**). Indeed, the GC content of the first 25 nts of downstream exon anti-correlates with out-of-order splicing but not concurrent splicing, presumably because secondary structure at the downstream 3' splice site antagonizes conversion of the downstream lariat intermediate to mRNA, prohibiting out-of-order splicing but not concurrent splicing. Consistent with this model, GC-rich cassette exons are particularly sensitive to the levels of the U1 snRNP as well as the two DEAD RNA helicases DDX5 and DDX17 known to promote U1 snRNP binding (Lemaire et al., 2019). It is notable that although GC content is known to slow transcription, slowing transcription would likely yield the opposite effect – more in-order splicing. Therefore, the impact of GC content, which extends across both the upstream and downstream introns, must be strong enough to reduce the significance of the transcriptional delay of the downstream intron and thereby increasing the probability of concurrent splicing.

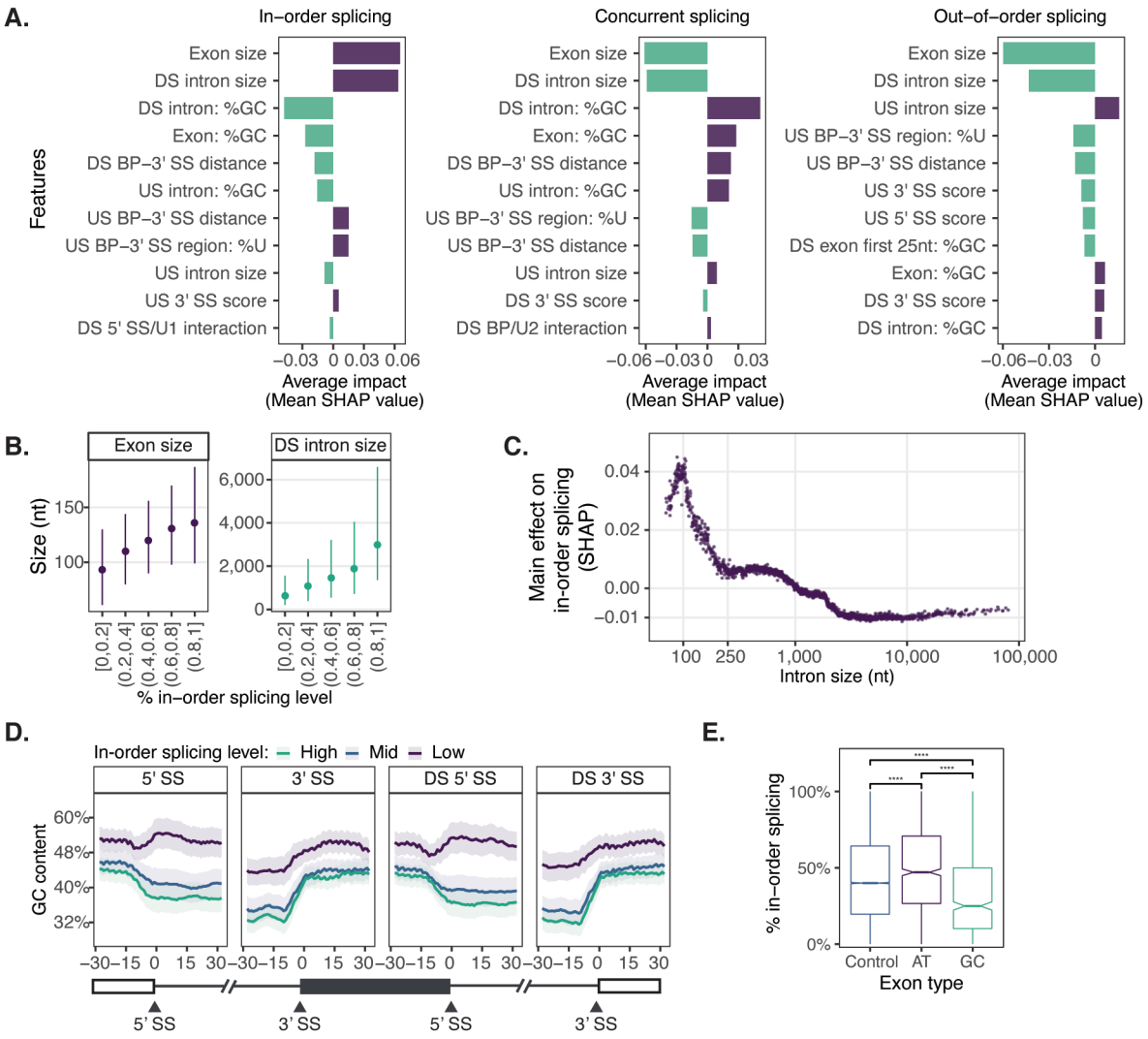


Figure 2.7. Intronic elements, gene architecture and genomic context predict the order of co-transcriptional splicing.

A. The average impact (average of absolute SHAP values) of top features is visualized for in-order splicing, concurrent splicing, and out-of-order splicing, respectively. Correlation is calculated with Spearman's correlation. **B**. Exon size and downstream intron size are positively correlated with in-order splicing. Point range plots illustrate the size distribution at each quintile of in-order splicing. The solid point represents the median value and the vertical line indicates the range of data between the 25th percentile and the 75th percentile. **C.** The effect of intron size on in-order splicing shows regional differences. Scatter plot to illustrate the main effect of intron-size on in-order splicing (5,000 3' SSs are sampled for this analysis). **D**. Introns with different levels of in-order splicing (low, mid, high) exhibit different patterns of GC content across splice sites of the upstream and downstream introns (+/- 30 nts). **E**. Cassette exons with different GC content are associated with different levels of in-order splicing. Box plots to illustrate the percentage distribution of in-order splicing in different types of exons (Control (n = 27,655), AT-rich cassette exons (n = 1,020), GC-rich cassette exons (n = 536)). The *p*-values were calculated by Mann-Whitney test.

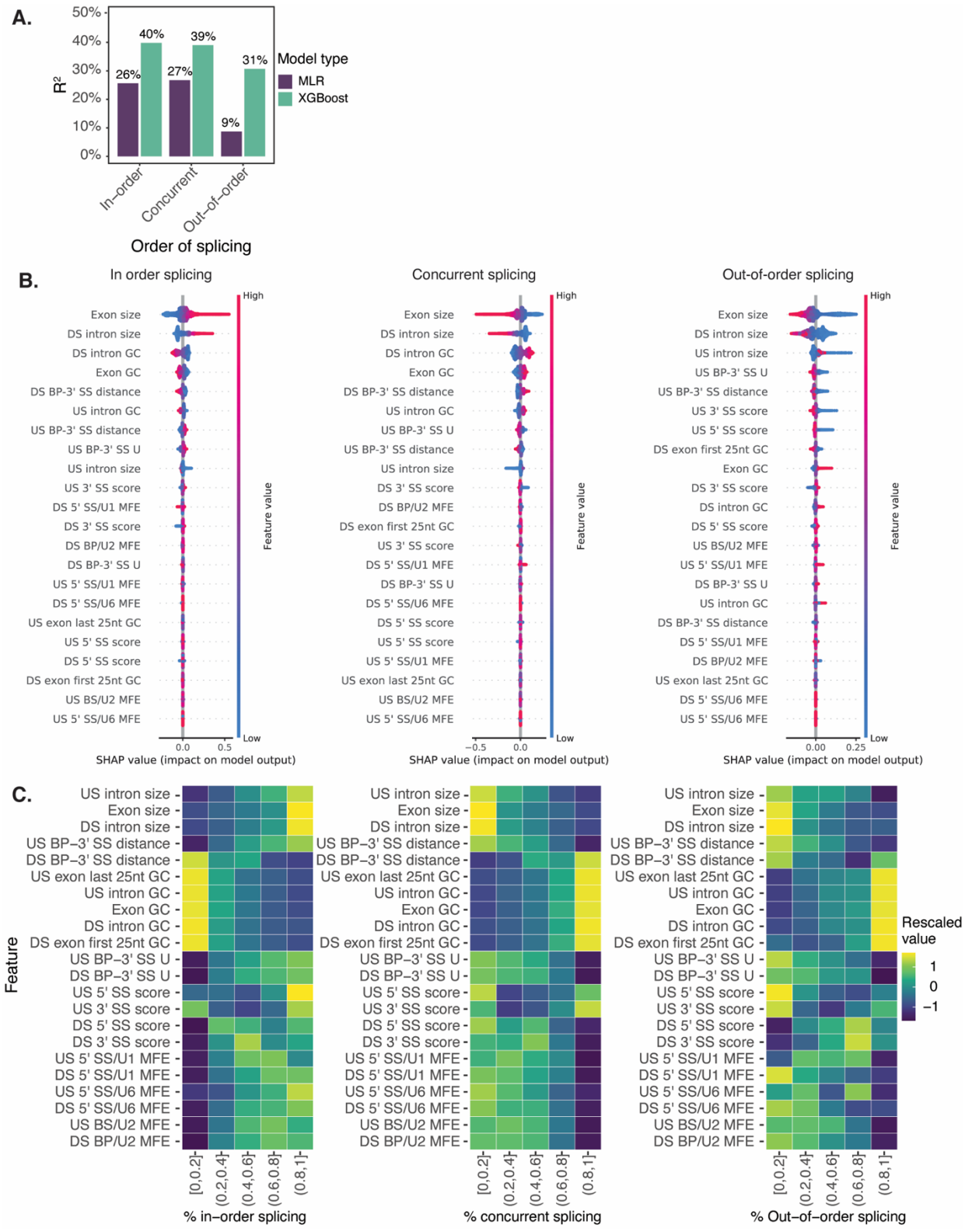


Figure 2.8. Characteristics of features associated with the order of splicing.
A. XGboost-based models outperformed multilinear regression (MLR)-based models. Model performance is evaluated by the coefficient of determination (R^2). **B.** SHAP summary plot to

Figure 2.8. (continued) illustrate the distribution of the impacts each feature has on in-order splicing, concurrent splicing, and out-of-order splicing, respectively. Features are ordered from high to low by the sum of SHAP values of each feature. **C.** Heatmap to illustrate the relationship between each feature and in-order splicing, concurrent splicing, and out-of-order splicing, respectively. Feature values are normalized within each feature.

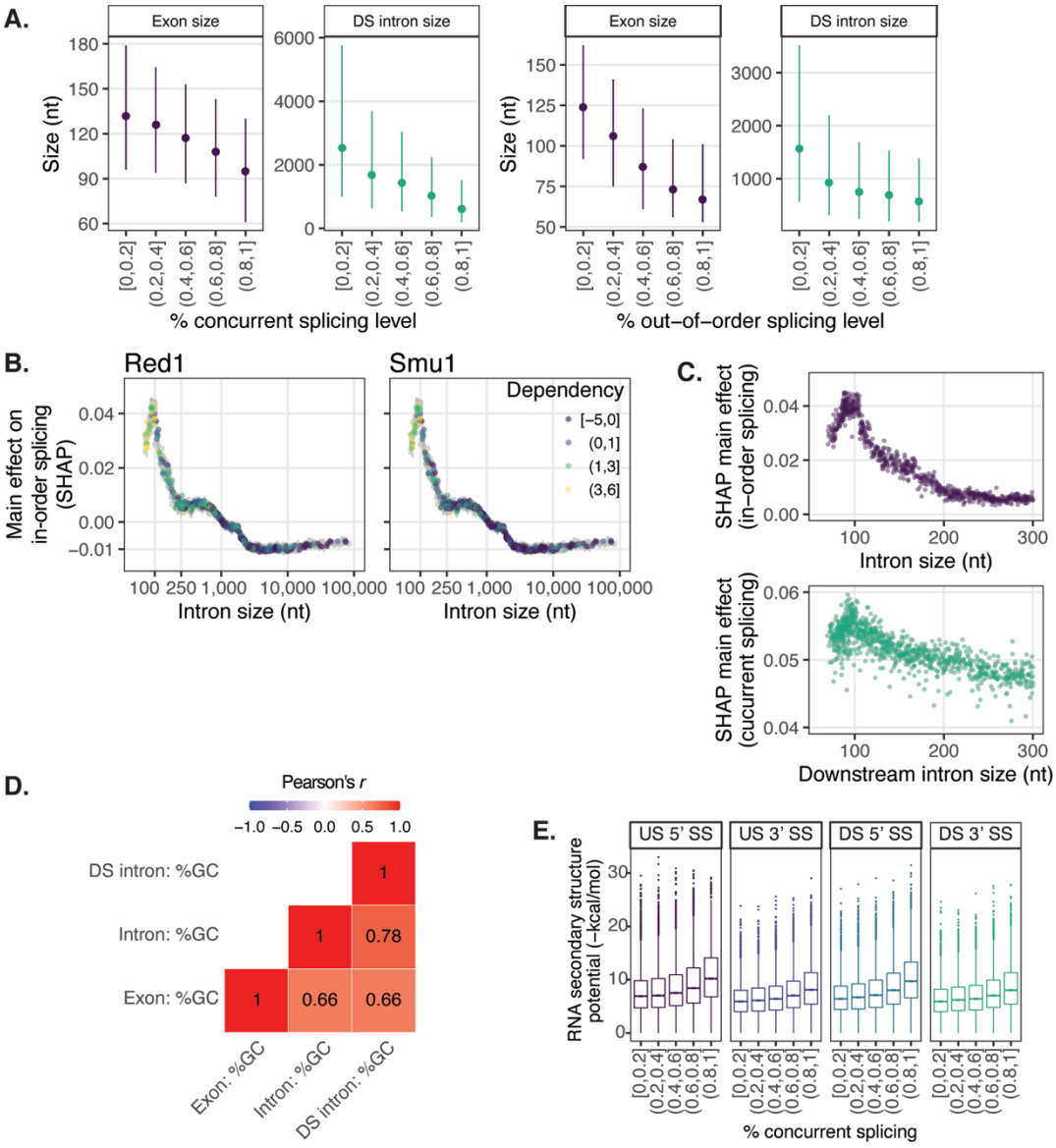


Figure 2.9. Features associated with the order of splicing.

A. Exon size and downstream intron size are negatively correlated with either concurrent or out-of-order splicing. Point range plots to illustrate the size distribution at each quintile of concurrent and out-of-order splicing, respectively. **B.** For Smu1/Red1-dependent introns, intron sizes are more likely to be positively correlated with the level of in-order splicing. Fold changes (log2) of intron retention after Red or Smu1 knockdown (Keiper et al., 2019) are overlayed on the scatter plot illustrating the main effect of intron size on in-order splicing. **C.** Scatter plot to illustrate that for introns smaller than 300 nts, the main effect of intron size on in-order splicing parallels the main effect of downstream intron size on concurrent splicing. **D.** Heatmap to illustrate that the GC content of intron, exon, and downstream intron in a single gene are highly correlated. Pearson correlation is calculated. **E.** Box plots illustrate that higher concurrent splicing for adjacent intron is associated with stronger RNA secondary structure potential (-kcal/mol) around splice sites.

Early co-transcriptional lariat formation indicates widespread usage of intron definition

In-order splicing not only indicates fast lariat formation relative to concurrent and out-of-order splicing but also reports on the precise timing of splicing, relative to transcription, through its tail length, with nucleotide precision. To assess the timing of lariat formation for in-order splicing, we analyzed the size distribution of in-order spliced NLI reads, as an indicator of the NLI tail length (**Fig. 2.11A**). We found that 10% of NLI 3' ends (RNAP IIs) are within 52 nts downstream of BPs, 50% are within 144 nts, and 90% are within 410 nts (**Fig. 2.10A; Fig. 2.11B**), implying that many lariat intermediates form when RNAP II is not far past the 3' SS. Moreover, the wide-ranging size distribution indicates that the timing of lariat formation varies among introns. Indeed, we found evidence, for example, that intron 16 of *EIF4G2* first forms lariat intermediate when RNAP II is 69 nts downstream of the 3' SS, whereas intron 21 of *OPLAH* first forms lariat intermediate when RNAP II is 157 nts downstream of the 3' SS (**Fig. 2.10B**). More surprisingly, intron 16 of *HNRNPK* first forms lariat intermediate when RNAP II is only 3 nts downstream of the 3' SS (**Fig. 2.10B**). Given that the exit channel of RNAP II is about 18-21 nts (Bernecky et al., 2017; Chen et al., 2009), this observation implies that when intron 16 of *HNRNPK* first undergoes lariat intermediate formation, the 3' SS is still buried within RNAP II,. This observation further implies the independence of lariat formation on this 3' SS (see below).

Unexpectedly, 58% of NLI 3' ends, the inferred positions of RNAP IIs, are upstream of the downstream 5' SS (**Fig. 2.10 B, C, D**); in total, 90% of introns in our dataset yielded NLIs with 3' ends upstream of the downstream 5' SS for at least a population of splicing events (**Fig. 2.11C**). These observations are surprising because they imply that NLI, which ended upstream of the downstream 5' SS, formed via the intron definition pathway. The intron definition pathway does not require transcription of the downstream 5' SS to initiate lariat formation, whereas the proposed

dominant pathway, the exon definition pathway, does (Berget, 1995; Hollander et al., 2016). Remarkably, introns with various sizes have NLI reads indicative of intron definition events (**Fig. 2.10E**), implying the intron definition pathway is not just a rare occurrence, but plays a substantial role in initiating splicing of introns genome-wide. Future work will be required to reconcile these observations with the prevailing view of exon definition as the operable mode of splicing and determine how introns are recognized in humans.

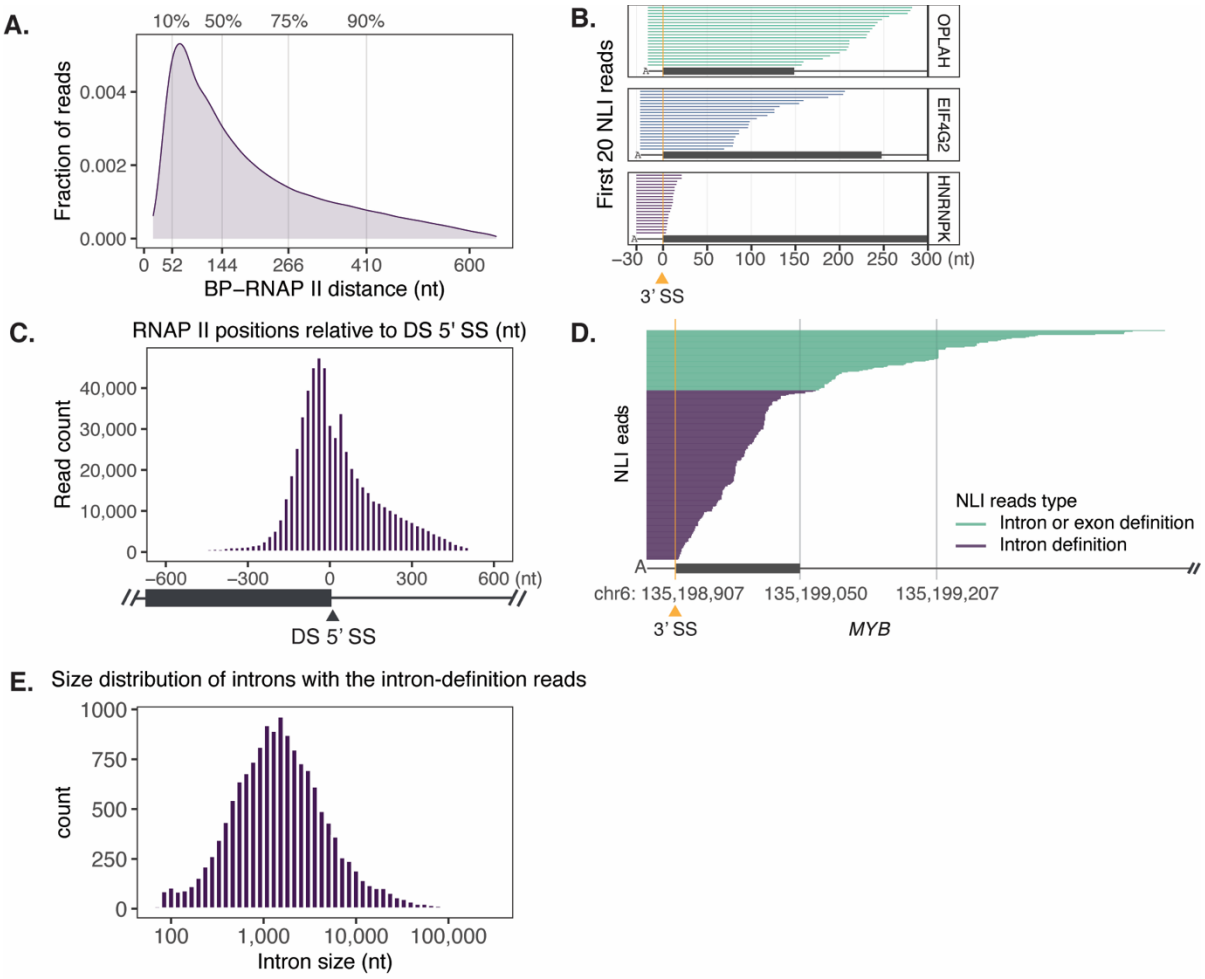


Figure 2.10. Early co-transcriptional lariat formation indicates widespread usage of intron definition.

A. For in-order spliced nascent lariat intermediates, RNAP IIs are located at various distances from the BPs. Density plot illustrates the BP-RNAP II distance distribution of in-order NLI reads. **B.** Initial lariat formation varies among introns. First 20 in-order NLI reads ranked by length are visualized for the 20th intron of *OPLAH* gene, the 16th intron of *EIFG2* gene, the 16th intron of *HNRPNK* gene, respectively. **C.** Histogram illustrates RNAP II positions relative to the downstream 5' SSs for in-order spliced nascent lariat intermediates. **D.** In-order NLI reads derived from the 8th intron of *MYB* gene are visualized for lariat formation events that occurred when RNAPIIs were in the downstream exon (lines labeled in purple) or in the downstream introns (lines labeled in green). **E.** Histogram illustrates the wide-ranging size distribution of introns that contain NLI reads derived from the intron definition events.

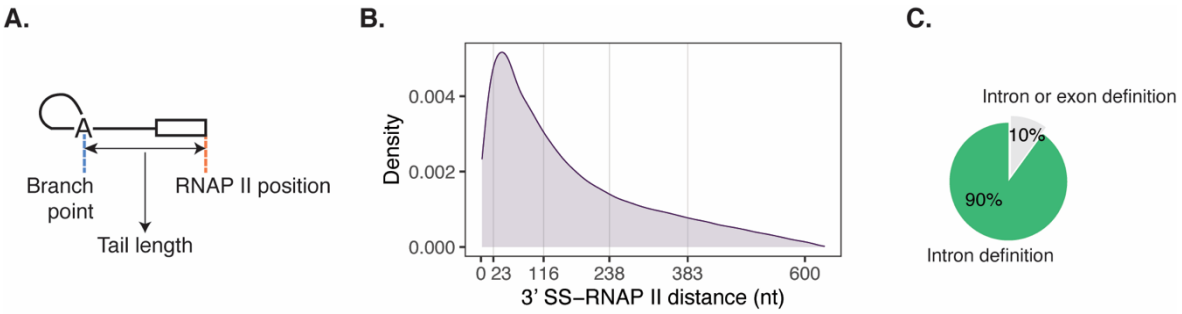


Figure 2.11. Early co-transcriptional lariat formation indicates widespread usage of intron definition.

A. Schematic view of nascent lariat intermediate tail, which is the region between BP and RNAP II position. **B.** Density plot to illustrate the 3' SS-RNAP II distance distribution of in-order NLI reads. **C.** Pie chart to illustrate the percentage of introns containing NLI reads derived from the intron definition events.

Gene architecture and nucleotide composition at the 3' end of an intron predict the timing of in-order splicing

Toward understanding the basis for the timing of lariat formation for in-order splicing, we focused on the earliest timing of lariat formation for reach intron, because such timing likely reflects the minimal requirements for each intron to undergo lariat formation. Specifically, we applied XGboost to identify key sequence features that correlate with the tail length of the shortest NLI read in each intron.

The resulting model accounted for 35% of the variance in the timing of initial lariat formation. We then quantified the contribution of each feature using SHAP values (**Fig. 2.12A; Fig. 2.13A**). Of all the features, features around the BP and the 3' SS are top contributors. Specifically, long BP-3' SS distance, high U content or low C content in between the BP and the 3' SS as well as around the 3' SS and high U around the BP, and weak RNA secondary structure around the 3' SS are associated with early lariat formation (**Fig. 2.12A; Fig. 2.13A-D**). Another top feature is intron size, which contributes positively to the timing of initial lariat formation. This observation is surprising, as given the exon definition model, one would expect longer introns to disfavor early splicing; one rationale may be that longer introns allow for more time for spliceosome assembly before transcription of the intron is complete (**Fig. 2.12A; Fig. 2.13E**).

Given that U content between the BP and the 3' SS is strongly and positively associated with early splicing, we asked whether high U content would indicate strong binding of U2AF1/U2AF2 heterodimer, in which U2AF1 binds to the terminal AG at the 3' SS and U2AF2 binds to U-rich PPT. Consistent with U2AF2 binding, hexamer enrichment analysis revealed that the BP-3' SS region of early introns is enriched for U-rich hexamers in early introns (**Fig. 2.13F**); *de novo* motif analysis confirmed that these early splicing introns are enriched for U2AF2 binding

sites (**Fig. 2.12B**). To test for U2AF binding, we compared U2AF1/U2AF2 crosslinking in early and late splicing introns by analyzing eCLIP data (Van Nostrand et al., 2020). Significantly, early splicing introns showed increased interactions of the U2AF1/U2AF2 heterodimer than late splicing introns do (**Fig. 2.12C**). Together, these results suggest that strong U2AF2 binding promotes early lariat formation.

Considering that the U2AF1/U2AF2 heterodimer showed different interactions with the 3' ends of introns between early splicing and late splicing introns, we examined whether such differential interactions also exist in other *trans*-acting factors. We expanded our analysis to other core splicing factors examined by eCLIP-seq in K562 cells (Van Nostrand et al., 2020). Interestingly, we observed in early splicing introns enrichment of GPKOW, a factor involved in the spliceosome activation (**Fig. 2.12C**), implying faster formation of the catalytically active B^{act} spliceosome complex. In contrast, EFTUD2 and PRPF8, two essential splicing factors that remain associated with the spliceosome during the entire process of splicing, are enriched in late splicing introns (**Fig. 2.12B**). Such enrichment may not necessarily indicate strong factor binding but reflect long residence time on nascent transcripts, since late lariat formation events likely result from slow splicing. Together, these results establish key roles for local sequence features and *trans*-acting factors in determining the timing of initial lariat formation.

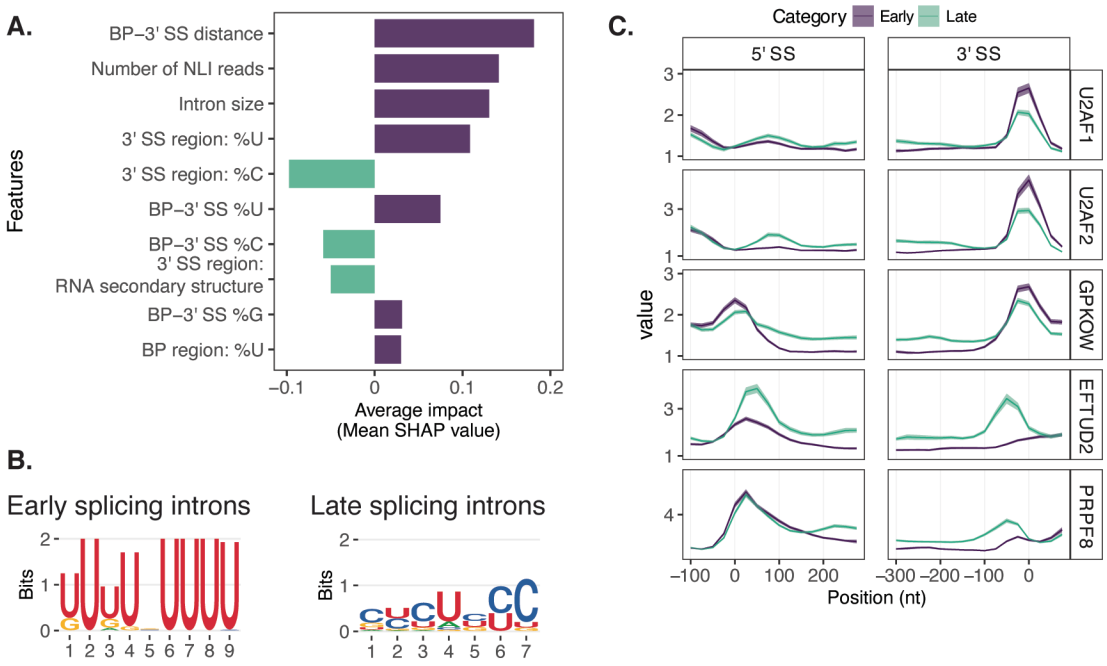


Figure 2.12. Gene architecture and nucleotide composition at the 3' end of an intron predict the timing of in-order splicing.

A. The average impact (average of absolute SHAP values) of top features is visualized for initial lariat formation. Correlation is calculated using Spearman's rank correlation coefficient. **B.** Seqlogo illustrates that U2AF2 binding motif is enriched in early splicing introns. **C.** Splicing factors exhibited different eCLIP signals in early and late splicing introns. Averaged eCLIP signals of a given splicing factor were visualized for the 5' SS region and the 3' SS region, respectively. 95% confidence interval was shown by the shaded area. Early splicing introns have at least 70% of NLI reads end within 50 nts downstream of the 3' SSs; late splicing introns have at least 70% of NLI reads end at least 200 nts downstream of the 3' SSs.

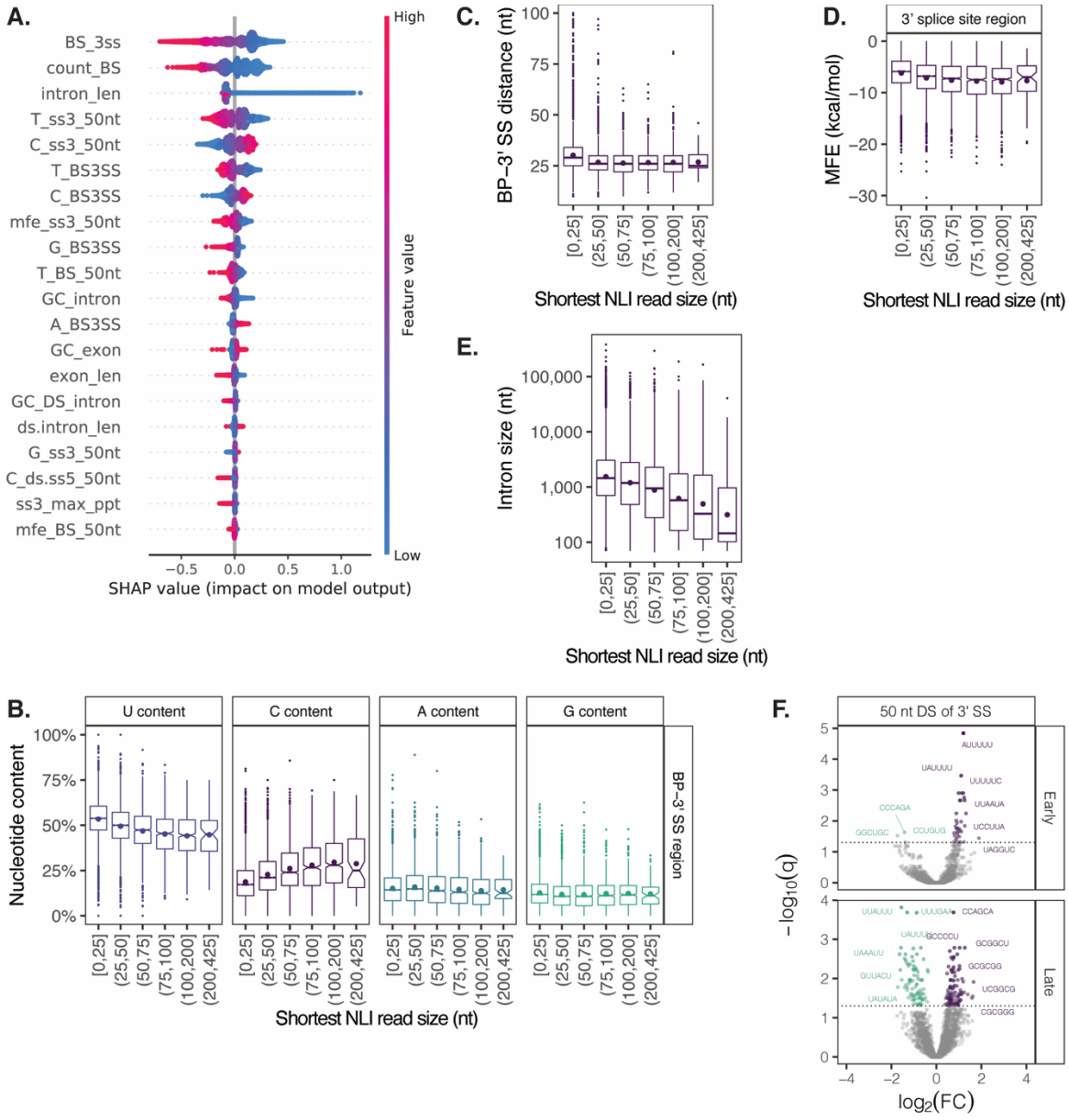


Figure 2.13. Local sequence features affect the timing of initial lariat formation.
A. SHAP summary plot illustrates the distribution of the impacts each feature has on initial lariat formation. **B.** Box plots illustrate the relationship between individual nucleotide content (U, C, A, G) in the BP-3' SS region and initial lariat formation. **C.** Box plots illustrate the relationship between the BP-3' SS distance and initial lariat formation. **D.** Box plots illustrate the relationship between the potential of RNA secondary structure formation (MFE, kcal/mol) around 3' SS and initial lariat formation. **E.** Box plots illustrate the relationship between the intron size and initial lariat formation. **F.** Volcano plots illustrate that U-rich hexamers are enriched in early splicing introns, whereas U-rich hexamers are depleted in late splicing introns. Early and late splicing introns are classified as in Fig. 2.12.

AG/U2AF1-independent human introns can undergo ultra-fast lariat formation

For introns that undergo early lariat formation, some of their nascent lariat intermediates formed so early that the 3' SSs could not have emerged from the exit channel of RNAP II (e.g. LI reads in intron 16 of *HNRPNK*; **Fig. 2.10B**). This ultra-fast splicing implies that lariat formation occurred before U2AF1 binds to the AG dinucleotide at the 3' SS. Consistently, previous findings indicate that a subpopulation of human introns can undergo lariat formation in an AG/U2AF1-independent manner (Guth et al., 1999; Shao et al., 2014; Wu et al., 1999). Biochemical studies found that AG/U2AF1-independent introns have strong PPTs (Guth et al., 1999; Wu et al., 1999), suggesting that U2AF2 can function independent of U2AF1 to stabilize SF1 and later recruit the U2 snRNP. Together with the observations that a long BP-3' SS distance and U2AF2 binding correlate with early splicing introns (**Fig. 2.12A, B**), we hypothesized that ultra-fast lariat formation events captured by CoLa-seq are AG/U2AF1 independent.

To test this hypothesis, we knocked down U2AF1 in K562 cells and performed CoLa-seq to assay for changes in the timing of lariat formation. To minimize indirect effects, we knocked down U2AF1 mildly (**Fig. 2.15A**), and, by qRT-PCR and RNA-seq, confirmed that the KD caused expected changes in AS based on previous studies (**Fig. 2.15B, C**; Shao et al., 2014). For late splicing introns, the U2AF1 KD shifted the tails of nascent lariat intermediates to longer lengths, implying that the timing of co-transcriptional splicing was delayed and that these introns were sensitive to the levels of U2AF1 (**Fig. 2.14A**). By contrast, for ultra-fast splicing introns, the U2AF1 KD did not impact the tail length distribution, implying that the timing of co-transcriptional splicing was not delayed and that these introns are insensitive to the levels of U2AF1 (**Fig. 2.14A**). Broadly, these data verify that NLI 3' ends reflect the activity of splicing factors and confirm that CoLa-seq is sensitive to determine changes in splicing due to different

expression of a splicing factor. More specifically, these results support our hypothesis that these introns have ultra-fast splicing because they are AG/U2AF1-independent. Indeed, AG/U2AF1-sensitive introns have low U content in the BP-3' SS region, a signature of weak PPTs, indicating that these introns rely on U2AF1 to stabilize the interaction between the PPT and U2AF2 to initiate splicing (**Fig. 2.14B**). Consistently, these U2AF1-sensitive introns have reduced in-order splicing but increased concurrent and out-of-order splicing (**Fig. 2.14C**), further confirming our hypothesis that strong binding of U2AF2 is necessary for fast splicing. Together, these results suggest that ultra-fast splicing introns can undergo AG/U2AF1-independent splicing; whereas late splicing introns undergo AG/U2AF1-dependent splicing (**Fig. 2.14D**).

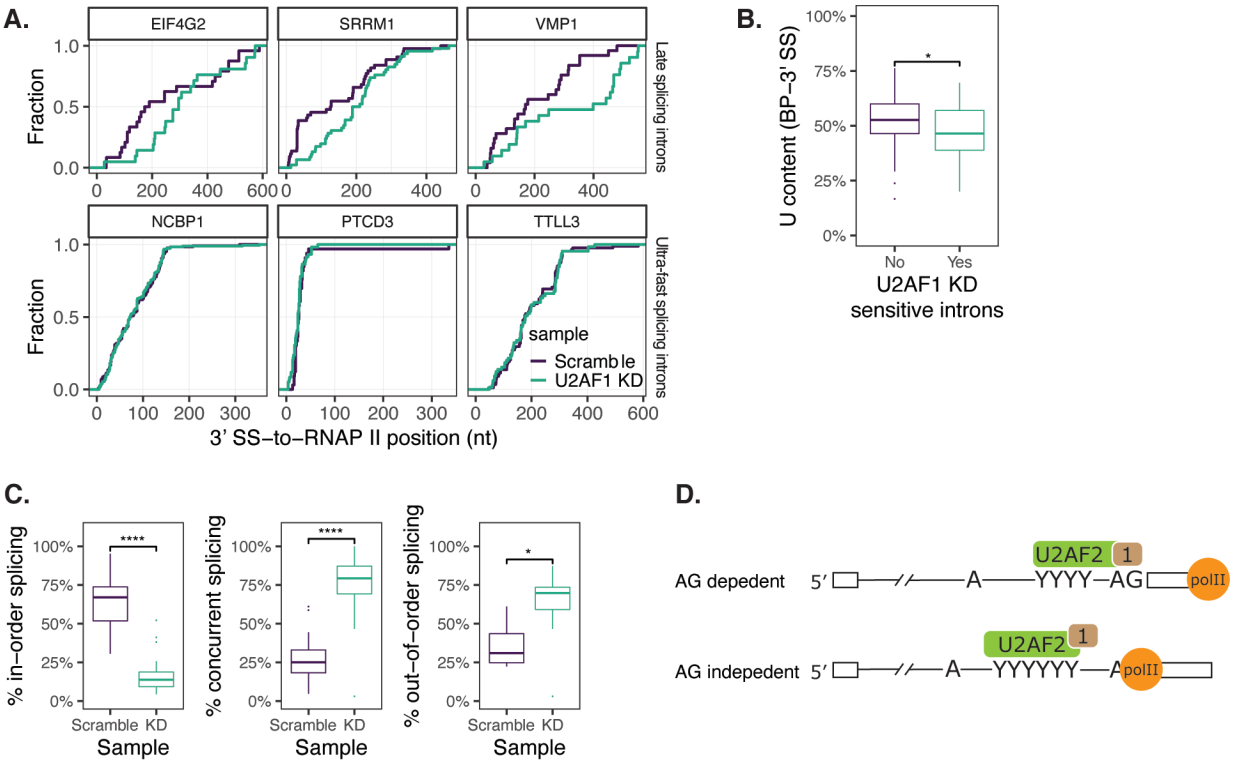


Figure 2.14. AG/U2AF1-independent human introns can undergo ultra-fast lariat formation. **A.** U2AF1 KD did not affect lariat formation in ultra-fast splicing introns but delayed it in late splicing introns. Empirical cumulative distribution plots to illustrate the distributions of 3' SS-RNAP II distance of lariat intermediates derived from specific introns of genes. **B.** Boxplots to illustrate that U2AF1 KD-sensitive introns have lower U content between BP and 3' SS, indicative of weaker PPTs. **C.** Boxplots to illustrate that U2AF1 KD-sensitive introns have a lower percentage of in-order splicing, but a higher percentage of concurrent and out-of-order splicing. **D.** Model of lariat formation with respect to AG/U2AF1 dependence. AG/U2AF1-dependent introns require binding of the U2AF heterodimer: U2AF2 binding to the PPT and U2AF1 binding to the AG at the 3' SS, thereby undergoing late lariat formation; whereas AG/U2AF1-independent introns only require binding of U2AF2 to the PPT, thereby undergoing ultra-fast lariat formation.

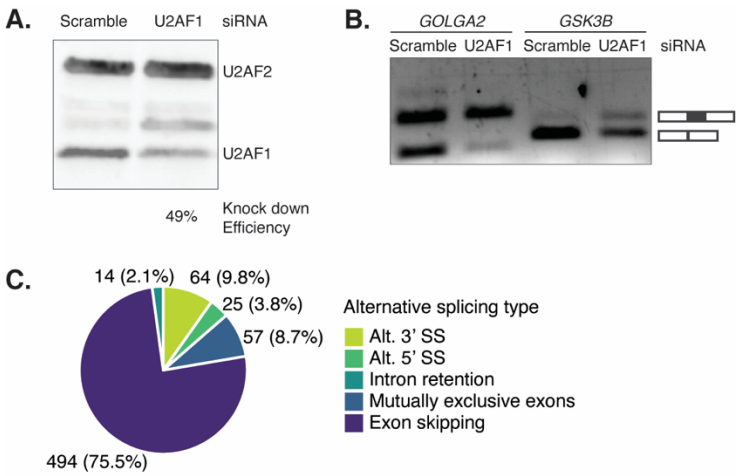


Figure 2.15. U2AF1 KD led to expected changes in AS.
A. Western blot confirmed mild U2AF1 KD (49%) by siRNA. **B.** RT-PCR confirmed that U2AF1 KD led to exon skipping, as were observed previously (Shao et al., 2014). **C.** RNA-seq confirmed that U2AF1 KD led to expected changes in AS.

Discussion

Co-transcriptional splicing is an essential step during gene expression. However, our understanding of splicing timing relative to transcription is limited. In this work, we have developed a novel approach, CoLa-seq, to study the dynamics and regulations of co-transcriptional splicing on a genome-wide scale in humans. By enriching for lariat RNA species, CoLa-seq greatly expanded BP annotations in the human genome. By detecting nascent lariat intermediates, CoLa-seq captured co-transcriptional splicing in action, revealing concurrent splicing for the first time *in vivo*. Notably, we observed that the timing of lariat formation varies dramatically across introns and even within the same intron. Further, we identified key *cis*-regulatory features and *trans*-acting factors that predict the timing of splicing. Intriguingly, our data also indicate that intron definition is used prevalently in humans. In fact, AG/U2AF1-independent introns can undergo lariat formation even before fully transcribed introns have emerged from RNAP II. Through nascent lariat intermediates, CoLa-seq has expanded our understanding of co-transcriptional splicing in humans.

The BP is one of three essential reactants participating in the splicing reaction and its usage directly impacts 3' SS selection. Mutations altering BP usage are associated with human diseases (Agrawal et al., 2018). Despite its essentiality, BP annotation in the human transcriptome has proven to be challenging and lags far behind 5' and 3' SS annotations, largely due to the degenerative sequence compositions surrounding BPs and the transient nature of lariat RNA species that BPs reside in. Recent large-scale approaches improved BP annotation by analyzing inverted reads in which BPs are located upstream of the 5’ SSs (Mercer et al., 2015; Taggart et al., 2017; Pineda and Bradley, 2018). These inverted reads were incidentally captured during the cDNA synthesis step of RNA-seq library construction when reverse transcriptase traversed

through the 2'-5' bonds formed between the BP and the 5' SS of the lariat RNA species. However, these reads were present at such low frequency that their analyses required either targeted enrichment or pooling of thousands of RNA-seq samples across many tissue and cell types (Mercer et al., 2015; Pineda and Bradley, 2018), both of which were still unable to capture most BPs. Further, targeted enrichment limited BP mapping to known introns, and pooling of different RNA-seq samples loses sample or cell-type specificity. To overcome these challenges, CoLa-seq enriched for lariat RNA species in an unbiased manner and efficiently captured the BPs at ends of individual reads as the branched structure in BPs stops RT proceeding further in the 3' to 5' direction. As a result, CoLa-seq greatly expanded BP annotations in the human genome (**Fig. 2.2F**). Further, its efficient BP mapping from a single sample opens the door for profiling BP usage under different conditions, such as cancer cells containing splicing factor mutations that are known to impact BP usage (Darman et al., 2015).

In addition to in-order and out-of-order splicing, CoLa-seq revealed that adjacent introns can undergo concurrent splicing (**Fig. 2.5B**), revealing a new splicing pathway for adjacent introns. Notably, splicing of specific intron pairs follows a preferred order, which is largely defined by the underlying sequence elements and gene architectures (**Fig. 2.7A**). As expected, we found that the strength of canonical splicing elements contributes to the order of splicing (**Fig. 2.7A; Fig. 2.8**). For example, when upstream introns contain strong PPTs, measured by U content, introns tend to have a high level of in-order splicing. This is likely because a strong PPT efficiently recruits the U2AF heterodimer, which in turn promotes efficient spliceosome assembly and lariat formation. Consistent with this idea, we detected strong U2AF eCLIP signals near 3' SSs of introns with a high level of in-order splicing. Notably, introns with high in-order splicing are also associated with long exons and long downstream introns, both of which are likely to increase the time required for

RNAP II to transcribe the downstream introns, thereby reducing the possibilities of concurrent and out-of-order splicing. These observations indicate that, analogous to its role in regulating exon inclusion and exon skipping, transcription is a key target for regulating in-order and concurrent splicing. Although GC content can delay transcription, which would increase in-order splicing, we observed the opposite effect – GC content decreases in-order splicing, indicating an impact outside of transcription (**Fig. 2.7**). In particular, we found that regional nucleotide composition strongly impacts the order of splicing; adjacent introns from low GC content regions preferentially undergo in-order splicing, whereas adjacent introns from high GC content regions preferentially undergo concurrent and out-of-order splicing. The impact of such regional bias also manifests at the gene level, as adjacent introns within the same genes share more similar splicing profiles than adjacent introns from different genes. Intriguingly, recent long-read sequencing studies observed that the splicing timings across multiple introns in a gene are often correlated (Drexler et al., 2020; Reimer et al., 2020). Our analyses indicate that such correlation is likely due to GC content bias, which extends beyond an intron up to several kbs, thereby creating a regional bias that imposes similar and significant impacts on splicing timing across multiple introns. High GC content likely impacts splicing timing in multiple ways, such as weakening splice site strength, disrupting interactions with snRNA or trans-acting factors, or stabilizing local RNA secondary structures. Therefore, high GC content slows down splicing in both introns and minimizes the impact of transcription delay caused by the downstream intron, thereby increasing the possibility of concurrent splicing. Together, these findings suggest that the order of splicing might be inherently encoded in the human genome through sequence features, gene architectures, and large genomic contexts to help coordinate splicing across multiple introns within the same gene.

Importantly, our results support both early and late splicing estimates in humans, as we found evidence that the timing of lariat formation varies significantly across introns and even within the same intron. In particular, 50% of RNAP IIs associated with in-order splicing were less than 120 nts downstream of the 3' SS, whereas 50% of RNAP IIs associated with concurrent splicing were located at least 1.2 kb downstream of the 3' SS. These results indicate that some introns undergo lariat formation as soon as they are transcribed, similar to previous findings in budding yeast, fission yeast, flies, and murine cells (Carrillo Oesterreich et al., 2016; Herzel et al., 2018; Drexler et al., 2020; Reimer et al., 2020); whereas other introns undergo lariat formation when RNAP II is much farther downstream of the 3' SSs, consistent with recent estimates in humans (Drexler et al., 2020; Wachutka et al., 2019).

Using computational modeling, we found that sequence features related to the BP, the PPT, and the 3' SS are associated with the timing of in-order splicing. Introns with strong PPT and U2AF binding undergo early lariat formation. Since previous studies have found that RNAP II recruits U2AF2 and Prp19 complex to pre-mRNAs and promote splicing activation (David et al., 2011), our findings suggest that U2AF2 and Prp19 complex are deposited to introns with strong PPTs right after they are transcribed, thereby leading to early lariat formation. Compared to features at the 3' end of introns, features related to the 5' SS have minimal effects. This is likely because 5' SS recognition is completed by the time RNAP II transcribes the 3' SS. Further, since the regulation of AS is often achieved by promoting or repressing the binding of the U1 snRNP, U2AF, or the U2 snRNP to the splice sites, the timing of lariat formation might provide a more accurate measurement for the impact of *in vivo* splicing regulation than the timing of mRNA formation, which is composed of both lariat formation and exon ligation, does.

As most human introns are much longer than exons and mutations at either the 5' or 3' SS often lead to exon skipping, the exon definition pathway is considered as the dominant pathway in humans (Berget, 1995; Hollander et al., 2016). Thus, it is surprising that CoLa-seq captured many short NLI reads, indicating that lariat formation occurred through the intron definition pathway (**Fig. 2.10**). It is formally possible that these NLI reads derived from intron-definition events that were incompetent for exon ligation. However, this is unlikely to be the case since most BPs and 3' SSs observed in these NLI reads were also observed in ELI reads from the same sample (**Fig. 2.2I**), indicating that these splicing signals are fully functional for both steps of splicing. Recent genomic studies using different approaches also found evidence for intron-definition mediated mRNA formation (Drexler et al., 2020; Reimer et al., 2020). Further, these intron-definition events could occur with the help of factors that bind to exonic splicing enhancers to facilitate 3' SS recognition (Wu and Maniatis, 1993; Ule and Blencowe, 2019). Lastly, these intron-definition events are associated with features that are known to promote lariat formation. In fact, CoLa-seq observed that introns with strong PPT for U2AF2 binding underwent ultra-fast lariat formation in an AG/U2AF1-independent manner (**Fig. 2.14A**), consistent with previous findings that a subpopulation of human introns does not require U2AF1 for lariat formation (Guth et al., 1999; Wu et al., 1999; Shao et al., 2014).

Collectively, CoLa-seq provides a powerful method to map BPs and study co-transcriptional splicing in humans. CoLa-seq can be readily applied to examine changes in the regulation of co-transcriptional splicing under different conditions as well as to interrogate the impact of different splicing factors and genetic variations on co-transcriptional splicing.

Materials and Methods

Chromatin associated RNA isolation

Chromatin RNA was isolated from K562 cells similar to previously described methods (Pandya-Jones and Black, 2009; Wuarin and Schibler, 1994; Werner and Ruthenburg, 2015) with modifications noted below. All of the steps were performed with pre-chilled, freshly prepared RNase-free buffers using RNase-free equipment on ice or at 4˚C. 5% of subcellular fractions were saved for quality control. 50 million K562 cells were spun down and washed twice with 15 mL of 1X PBS and once with 1mL of ice-cold 1X PBS. Washed cells were then gently resuspended in 1mL of 0.1% (v/v) Triton-X-100 containing buffer A (10 mM HEPES (pH 7.5), 10 mM KCl, 10% Glycerol, 0.116 g/mL sucrose, 1 mM DTT, 1x protease inhibitor, 2 mM EDT) and lysed on ice for 9 min. After incubation, cells were centrifuged at 1200g for 5 minutes at 4˚C. The supernatant corresponding to the cytoplasmic fraction was carefully removed. The remaining nuclei pellet was gently resuspended in 250µL of NRB buffer [20 mM HEPES (pH 7.5), 75 mM NaCl, 50% glycerol, 0.5 mM EDTA, 1 mM DTT, 1x protease inhibitors]. 250 µL of NUN buffer [50 mM HEPES (pH 7.5), 300 mM NaCl, 0.2 mM EDTA, 1 M urea, 1 mM DTT, 1% NP-40] was slowly added and mixed by inversion. The resulting nuclei suspension was incubated on ice for 5 minutes and centrifuged at 1200g for 5 minutes at 4˚C. The supernatant corresponding to the nucleoplasmic fraction was carefully removed. The chromatin pellet was washed once with 500 µL of premixed NRB/NUN buffer (1:1) and centrifuged at 1200g for 5 minutes at 4˚C. The supernatant was carefully removed. The washed chromatin pellet was gently resuspended in 200 µL of NRB buffer. RNA from whole cells, cytoplasmic fraction, nucleoplasmic fraction, and chromatin pellet were extracted by Trizol reagent (Invitrogen) followed by a second extraction with phenol:chloroform:isoamyl alcohol (25:24:1) solution. The isolated RNA was furthered treated

with Turbo DNase (Invitrogen) to remove residual genomic DNA, followed by extraction with phenol:chloroform:isoamyl alcohol (25:24:1) solution. For each CoLa-seq library, chromatin RNA from 200 million cells were used.

Confirmation of fractionation

The efficiency of fractionation was tested by western blotting of the cytosolic fraction, the nucleoplasmic fraction, and the chromatin fraction with appropriate antibodies. The antibodies used were GAPDH (Santa Cruz Biotechnology), RNA Pol II (Ser-5 phosphorylated, MBL international), Histone H3 (Abcam)

Enrichment of lariat RNA species

In order to enrich for lariat RNA species (nascent lariat intermediates and excised lariat introns) in the chromatin-associated RNA, a size-selection procedure was performed using SPRI paramagnetic beads (Beckman coulter) to remove RNA shorter than 200 nucleotides. Next, ribosomal RNA was depleted using oligo-based Ribo-Zero Magnetic Gold Kit (Illumina) according to the manufacturer's protocol. During the optimization of this protocol, we noticed abundant ncRNA reads in our libraries, similar to previous observations in mNET-seq (Mayer et al., 2015). To overcome this issue, biotinylated anti-sense oligos were designed to specifically target and deplete ncRNAs that were found to be abundant in our initial libraries (**Table 2.1**). For subsequent libraries, ncRNA depletion was performed immediately following rRNA depletion using Dynabeads™ MyOne™ Streptavidin C1 beads (Invitrogen). Excess anti-sense oligos were removed by treatment with Turbo DNase (Invitrogen) and the RNA was purified using the RNA clean and concentrate kit (Zymo Research). Finally, purified RNA was treated with RppH (NEB) to decap linear RNA and the decapped linear RNA was degraded using Terminator 5'-3' exonuclease (Lucigen). The enzymatic treatments were performed according to the manufacturer's

protocols and the lariat enriched RNA was purified with the RNA clean and concentrate kit (Zymo research).

CoLa-seq library prep

Once the chromatin RNA was enriched for lariat species, we ligated a pre-adenylated 3' adapter, which contains 6 nt UMI at the 5' end (**Table 2.2**), using T4 RNA Ligase 2 (truncated K227Q, NEB) for 3 hours at 22˚C. Excess adapters after the reaction were removed by size selection using SPRI beads. cDNA was synthesized from the adapter-ligated RNA using Superscript III first strand synthesis system (Invitrogen) at 50 ˚C followed by cDNA size selection to remove the excessive primer. The resulting cDNA was circularized with Circligase (Lucigen). Final library amplification was performed on the circularized cDNA using Phusion polymerase (ThermoFischer Scientific) with primers to add the specific sequencing indices. The amplified product was run on a 6% TBE gel and the DNA fragments between 175 to 750 bp were gel purified and sequenced in a 100bp paired-end mode on an Illumina Hiseq 4000.

Nascent RNA sequencing

Chromatin-associated RNAs were depleted of ribosomal RNA depletion using Ribo-Zero Magnetic Gold Kit (Human/Mouse/Rat). Abundant ncRNAs were depleted as described above. The resulting RNAs were used as input for RNA-seq library prep. RNA-seq library was prepared using NEBNext Ultra™ Directional RNA Library Prep Kit for Illumina. The final library was sequenced in a 100bp paired-end mode on an Illumina Hiseq 4000. Nascent RNA-seq data were pre-processed using a custom snakemake (Köster and Rahmann, 2012) pipeline as followed: 1). Adaptors were trimmed using Cutadapt (Martin, 2011). Reads shorter than 15 nts were discarded; 2). The remaining reads were aligned to the human reference genome (hg38) with GENCODE transcriptome annotation (v26) using the STAR software (Dobin et al., 2013); 4). The aligned

reads were further filtered for unique mapping and properly aligned reads using samtools (Li et al., 2009). The resulting reads were used for downstream analyses; 5). Separately, after the adapter removal step, reads were used for gene expression quantitation using Kallisto (Bray et al., 2016).

Lariat sequencing

Lariat-seq was performed similarly to previously described methods (Mayerle et al., 2017; Mercer et al., 2015) with modifications. Briefly, chromatin-associated RNA was first isolated to capture both lariat intermediates and excised lariat introns. Then, ribosomal RNAs were depleted using Ribo-Zero Magnetic Gold Kit (Human/Mouse/Rat). To further enrich lariat RNA species, RNase R digestion was performed to degrade linear RNA and trim tails of lariat RNA species. Afterwards, RNA-seq library was prepared using NEBNext Ultra™ Directional RNA Library Prep Kit for Illumina. The final library was sequenced in a 100bp paired-end mode on an Illumina Hiseq 4000. The resulting reads were used to identify branch points with custom scripts. Briefly, chimeric reads were identified by mapping reads to the human genome (hg38) using STAR. Chimeric reads were further filtered for reads that consist of two segments: the 3' (right) portion of the segment is mapped to 5' splice site containing a portion of the intron; the 5' (left) segment is mapped to a region in the last 100 nucleotides of the intron. Lastly, the last nucleotide of the 5' segment is considered as the branch point, especially if it is a mismatch as RT often introduces a mismatch when it traverses through the 2'-5' linkage in the lariat RNAs.

Splicing inhibition and transcription inhibition

K562 cells at 90% confluency were treated for 2 hours with one of the following reagents (i) Pladienolide B (Santa Cruz Biotechnology) to a final concentration of 10 μM or (ii) Flavopiridol (Sigma) to a final concentration of 1 μM. A control treatment was also performed with an equal volume of DMSO. Successful splicing inhibition was confirmed by RT-PCR to detect previously

observed splicing changes (Nojima et al., 2015). Chromatin fractionation and CoLa-Seq libraries were prepared from the DMSO or drug treated cells as described above. Primers used are listed in Table 2.3.

U2AF1 KD

K562 cells at 70% confluency were transfected with 30 pmols of either U2AF1 siRNA or scrambled siRNA control using Lipofectamine RNAiMAX (ThermoFischer Scientific), according to manufacturer's protocol. The cells were harvested 48 hours post transfection. Successful knockdown was confirmed by western blot and qRT-PCR to observe previously described splicing changes (Shao et al., 2014). CoLa-Seq libraries were prepared and sequenced as described above. Primers used are listed in **Table 2.3**.

Data preprocessing

CoLa-seq data were reprocessed as followed: 1). 6 nt UMIs were trimmed from reads and attached to read names to maintain their association with reads using fastp (S. Chen et al., 2018); 2). Adaptors were trimmed using Cutadapt. Reads shorter than 15 nts were discarded; 3). The remaining reads were aligned to the human reference genome (hg38) with GENCODE transcriptome annotation (v26) using the STAR software; 4). The aligned reads were further filtered for unique mapping and properly aligned reads using samtools; 5). To ensure 3' ends of reads represent positions of RNAP IIs, RT mis-priming produced reads were identified if their UMI portion was mapped to the genome without any mismatch or if they contained soft clips at ends corresponding to RNA 3' ends. These RT mis-primed reads were identified and removed using samtools and custom python scripts; 6). PCR duplicates were removed using UMI-tools (Smith et al., 2017); 7). Reads overlapping with ncRNAs (snRNAs, snoRNAs, miRNAs, rRNAs)

were removed using GENCODE reference. The remaining reads were used for downstream analyses.

Enrichment analysis for reads derived from lariat RNA species

Using previously annotated branch points (Pineda and Bradley, 2018), 21 nt branch point regions were extracted. These branch point regions were then intersected with 5' ends of all unique-mapping and deduplicated reads from CoLa-seq. Read signals across each branch point region was normalized to range between 0 and 1 to remove expression differences so that each branch point region contributed equally. The branch point regions were aligned with the branch point in the center and the mean normalized signal was calculated at each position.

Sequence logo of branch point motif

Using Bioconductor tools and ggseqlogo package in R (R Core Team, 2013; Huber et al., 2015; Wagih, 2017), the sequence logo was generated as followed: 1). For each branch point, a 11 nt long sequence (branch point in the middle) was extracted. 2). The resulting sequence set was used to calculate the position weight matrix and then plot sequence logo for the branch point motif.

Branch point calling algorithm

ELI fragments are defined as sequenced fragments for which their 3’ end maps to an annotated 3’ splice site (Gencode v29). NLI fragments are identified by their 5’ end mapping to known branch points and extending without gaps past an annotated 3’ splice site. However, published databases of known branch points (Pineda and Bradley, 2018) are relatively incomplete. To expand the number of NLI reads available for downstream analysis, we built a probabilistic classifier to expand the set of branch point positions from which NLI reads are anchored. In detail: each position in the 5 to 100 bp region upstream of annotated 3’ splice sites is assigned a probability that it is a branch point based on various features, based on naive Bayes assumption of

independence. The probability distribution of features over the non-branch points are assumed uniform and therefore a constant that can be ignored at all positions. Therefore, at each position, X

$$Pr(C_{bp}) \propto \prod_{i=1}^{n} P\,r(x_i|C_{bp})$$

where is a classification of branch point C_{bp} or non-branch point C_n at position X, and x_i is the value of feature i at position X. The features considered are pentamer motif (-3 to +1 relative to position X), distance to annotated 3'ss, and relative coverage of CoLa-seq 5' read ends. The probability distribution of pentamers in C_{bp} was trained on branch points empirically identified in Pineda and Bradley (Pineda and Bradley, 2018). More specifically, all high-quality branch points were used to generate a position weight matrix (weighted by the read count supporting each branch point in Pineda et al) to assign branch point probabilities for each hexamer. Similarly, the empirical distance to the annotated 3'ss across (Pineda et al) branch points (weighted by read count) was used to create a smoothed probability distribution using the `density` function in R with `bw=3` bandwidth adjustment. To determine a branch point probability from relative coverage of CoLa-seq 5' ends ($\Pr\left(x_i\middle|C_{bp}\right)$ where $x_i$is the relative coverage of CoLa-seq 5' ends at known branch points C_{bp}), we did not consider the empirical probability of relative coverage from known branch points but rather, we devised an alternate probability function as follows: CoLa-seq reads from all experiments were combined and filtered for fragments that end at (ELI) or cross (putative NLI) an annotated 3'ss. For each 3'ss region with at least 10 ELI or putative NLI fragments, the coverage (normalized to the total coverage in the region, and after adding a 0.1 pseudocount to all positions) of the 5' position of fragments was used to create a smoothed probability distribution with the `density` function with bandwidth `bw=0.5`. Smoothing was introduced as a way to emulate the

natural imprecision of reverse transcription termination, as we have observed that reverse transcriptase sometimes terminates at the branch point and sometimes at the base(s) just upstream. Following the assumption that ELI and LI CoLa-seq reads most often terminate at the base downstream of the branch point, the probability distribution was shifted by one base such that the center of smoothed peaks corresponds to more likely branch points. Finally, to classify bases as belonging to a branch point region, we defined a threshold

$$\hat{y} = C_{bp} \text{ if } \prod_{i=1}^{n} Pr(x_i|C_{bp}) > \text{threshold};\ C_n \text{ otherwise}$$

Or equivalently:

$$\hat{y} = C_{bp} \text{ if } \sum_{i=1}^{n} log(Pr(x_i|C_{bp})) > \text{threshold};\ C_n \text{ otherwise}$$

The threshold was chosen based on ROC analysis using a validation set of branch point positions identified in both Pineda et al as well as lariat-seq in a matched cell-type as the true response. We chose a threshold that yields a false positive rate of 0.01 in the ROC analysis. Finally, after applying this classifier to all 114,540 3'ss regions with >10 ELI/LI reads, we merged branch points with one or zero bases in between each other into a single most likely branch point. The feature training and classification process was repeated separately for introns annotated as U2 or U12 introns (Olthof et al., 2019). In total we identified 153471 amongst 95686 U2 introns, and 716 branch points amongst 536 U12 introns.

Identification of nascent lariat intermediates and excised lariat introns

Exons from GENCODE transcriptome annotation (v26) and identified in nascent RNA-seq data were used for identifying nascent lariat intermediates and excised lariat introns. Exons whose

ends can be used as alternative 3' splice sites were removed to avoid ambiguity. To identify reads derived from nascent lariat intermediates (NLIs), the following filtering steps were performed using a custom snakemake pipeline wrapped around bedtools (Quinlan and Hall, 2010), samtools. and Picard tools ("Picard toolkit," 2019): 1). Candidate NLIs reads were identified if they overlapped with intron/exon boundaries and their 5' ends are within 100 nts upstream of 3' splice sites; 2). using branch points annotated above, NLI reads were identified if their 5' ends correspond to the BP+1 position. To identify reads derived from excised lariat introns, the following filtering steps were taken using performed using a custom snakemake pipeline wrapped around bedtools, samtools. and Picard tools: 1). Candidate ELI reads were identified if their 3' ends are at the last nt of introns and their 5' ends are within 100 nts upstream of 3' splice sites; 2) Using branch points annotated above, ELI reads were identified if their 5' ends correspond to the BP+1 position.

Identification of constitutive and alternative splice sites

Using Regtools (Feng et al., 2018) and leafcutter (Li et al., 2018), 5' splice sites and 3' splice sites were extracted from nascent RNA-seq data of this study, nuclear RNA-seq data from ENCODE, as well as RNA-seq data from ENCODE. Splice sites were considered as constitutive sites when their usage percentages were higher than 95%. Splice sites were considered as major alternative sites when their usage percentages were between 50-95%. Splice sites were considered minor alternative sites when their usage percentages were between 10-50%.

Quantification of branch point usage

Branch point usage was quantified as described in Mercer et al. 2015 (Mercer et al., 2015). Briefly, ELI read (or NLI read) counts of a branch point were divided by ELI read (or NLI read) counts of all branch points associated with the same 3' splice site.

Quantification of usage coupling of branch point and 3' splice site

Alternative 3' splice sites using the same 5' splice sites were paired together, their nucleotide distances were then calculated and used to sort these pairs to different groups. If both 3' splice sites within a pair use the same set of branch points, the coupling index of this pair was set to 1, whereas if they used different ones, the coupling index was set to 0. The coupling fraction of each group was then calculated by taking the average of the coupling index of that group.

Classification of in-order, concurrent, and out-of-order splicing

To further classify NLI reads into classes of in-order, concurrent, and out-of-order splicing, NLI reads that ended at the last nt of an annotated exon were classified as concurrent splicing. NLI reads that were split reads (N in cigar) were considered as out-of-order splicing. NLI reads that did not belong to the other two classes and were shorter than 600 nts (Upper size limit of the library) were considered as in-order splicing.

Sequence feature generation

For exons and their neighboring introns, sequence features such as size and nucleotide compositions (%A, %G, %C, %U, and %GC) were extracted and calculated using Bioconductor tools within R. For 5' splice site and 3' splice site, their scores were calculated using MaxEntscan algorithm (Yeo and Burge, 2004), their regional nucleotide compositions were extracted, RNA secondary structure potentials were calculated using viennaRNA (Lorenz et al., 2011). For 5' splice sites, the binding energy of U1 snRNA and U6 snRNAs were also calculated using viennaRNA. For branch points annotated by CoLa-seq, the following features were generated: branch point identity, branch point usage percentage, BP-3' SS distance, and nucleotide composition and RNA secondary structure around branch points and in regions between BP and 3' SS.

XGBoost models to identify key features that contribute to the order of splicing, and the timing of in-order splicing

Intron pairs containing at least 10 CoLa-seq reads were used for modeling in-order splicing (29,211), concurrent splicing (27,419), out-of-order splicing (15,494), and the timing of in-order splicing (13,975). The target variables were %in-order splicing, %concurrent splicing, %out-of-order splicing, and log-transformed NLI read sizes, respectively. For each target variable, introns were randomly split into a training set and a test set in a 7:3 ratio. In the training set, when two features were more than 80% correlated (Pearson correlation), one of them was randomly removed. For each target variable, a gradient boosting based regression was then performed using XGBoost with hyperparameter tuned to generalize the model performance. As a result, the hyper-parameters used were max_depth=5, learning_rate=0.01, n_estimatros=550, subsample=0.6, colsample_bynode=0.6, reg_alpha=0.5, reg_lambda=0.8. The test set was used for model performance evaluation (metric: R-squared). Furthermore, 5-fold cross-validation was performed on the training set to get the final model performance (metric: R-squared). To evaluate the contribution of each feature within each model, SHAP values were computed and visualized using the SHAP package in python.

GC content calculation across splice sites for introns with different levels of in-order splicing

Intron pairs were divided into quintiles according to the percentages of in-order splicing observed in upstream introns. The first, the third, and the fifth quintile were labeled as the low, the mid, and the high level of in-order splicing group, respectively, and used for further analysis. For each intron pair, 50 nts flanking each splice site were extracted. Sequences corresponding to splice site signals were removed. For 5' splice sites, the last 3 nts in the exons and first 6 nts in the introns were removed. For 3' splice sites, the last 3 nts in the introns and first 2 nts in the exons were

removed. GC content was then calculated at each position with a 10 nt sliding window. Within each group, the average GC content and 95% confidence interval were calculated at each position and plotted for an adjusted 60nt size window across splice sites of both upstream and downstream introns.

Classification of early splicing and late splicing introns

Introns that have at least 70% of NLI reads that end within 50 nts downstream of 3' splice sites are classified as early splicing introns; whereas introns that have at least 70% of NLI reads that end more than 200 nts downstream of 3' splice sites are classified as late splicing introns

Analysis of eCLIP-seq data in K562 cells from ENCODE

Bam files of input and IP for each RBP were downloaded from the ENCODE project website (https://www.encodeproject.org/; Van Nostrand et al., 2016). For each RBP, IP over input ratio was calculated using bamCompare from Deeptools (Ramírez et al., 2014). The resulting IP/input ratios were calculated using computeMatrix from Deeptools for 5' splice site regions (-100 to +300 nts) and 3' splice site regions (-300 to +100 nts) of fast splicing introns and slow splicing introns, respectively. RBPs with differential signals between fast splicing introns and slow splicing introns were identified and processed for visualization in R.

De novo motif identification and motif enrichment analysis

To identify *de novo* motifs in specific regions of early splicing and late splicing introns, meme from MEME SUITE (Bailey et al., 2009; Tanaka et al., 2014) were used with a custom markov model built using the same regions from all introns containing at 10 NLI reads. The identified motifs were further processed in R using universalmotif (Tremblay, 2019) and ggseqlogo for making sequence logos. The identified motifs were also fed into tomtom from

MEME SUITE to search for enriched known motifs of RNA binding protein against the human CISBP-RNA database (Ray et al., 2013).

Hexamer analysis

To identify enriched and depleted hexamers in specific regions of early splicing and late splicing introns, the k-mer analysis was performed using transite package in R (Krismer et al., 2020).

Quantification, statistical analysis, and plot generation

All quantification and statistical analyses were done in R and python. Analysis details can be found in figure legends and result sections. All plots were prepared using data.table (Dowle and Srinivasan, 2019), ggpubr (Kassambara, 2020), cowplot (Wilke, 2019), patchwork (Pedersen, 2019), gridExtra (Auguie, 2017), ggrepel (Slowikowski, 2020), and tidyverse tools (Wickham et al., 2019) in R, except for SHAP summary plots, which were made using SHAP and matplotlib (Hunter, 2007) in python.

Supplementary materials

Table 2.1 ncRNA depletion oligos

Target	Sequence	3' end modification
U1	CCACAAATTATGCAGTCGAGTTTCCCACATTTG	/3BioTEG/
U1	CCACAAATTATGCAGTCGAGTTTCCCACATTTG	/3BioTEG/
U4	CAGTCTCCGTAGAGACTGTCAAAAATTGCCAATGC	/3BioTEG/
U3	GAGAGAAGAACGATCATCAATGGCTGACGGCAGTT	/3BioTEG/
U3	TCAGGAGAAAACGCTACCTCTCTTCCTCG	/3BioTEG/
U3	AACGATCATCAATGGCTGACGGCAGTTGCA	/3BioTEG/
U3	GAGAGAAGAACGATCATCAATGGCTGACGGCAGTT	/3BioTEG/
U3	TCAGGAGAAAACGCTACCTCTCTTCCTCG	/3BioTEG/
snora61	CAAGACCAGTGTTCAGATCCGATGGGAAAGGGATC	/3BioTEG/
snora61	AACATTTAGGCCAGCTTCACTATTACTTTT	/3BioTEG/
snora48	AAGCTGGGATGAACAAGGAGTTGCTTTGTC	/3BioTEG/
snora73A/B	ACGAGGCCCAGCTTTATTTCCAACGTTGTG	/3BioTEG/
snora73A/B	GAAAGGGACTGTACATCATGGGGCAGAGCC	/3BioTEG/
snora71B	ATTGATTCCTCTCCCTGCACTATCAATGACCAGGG	/3BioTEG/
snora71B	GTTTGGAAGGGATAGGAGTGGACCCTTCAAACACG	/3BioTEG/
snord17	TAGAACAGGAACTGAGGCTTGGAAGAAGGTCAGTG	/3BioTEG/
U2	ACGTATCAGATATTAAACTGATAAGAACAGATACTACACTTG	/3BioTEG/
U2	ATACCAGGTCGATGCGTGGAGTGGA	/3BioTEG/
U5	CTCTCCACGGAAATCTTTAGTAAAAGGCGAAAGA	/3BioTEG/
scaRNA2	CCAGGCCGCTCTCCCTCCCTAAAAC	/3BioTEG/
RMRP	GGGAGGAACAGAGTCCTCAGTGTGTAG	/3BioTEG/
7SK	TCGTATACCCTTGACCGAAGACCGGTCCTC	/3BioTEG/
28S rRNA	TTTCCAGCCGCGCCCCGTTTCCCAGGACGA	/3BioTEG/
28S rRNA	TTCCCCACGAACGTGCGGTGCGTGACGGGC	/3BioTEG/
45S rRNA	GAGGAAGACGAACGGAAGGACGGACGGCGCCGGAC	/3BioTEG/
45S rRNA	GTGAACGGGGAGGAGGCGGGAACCGAAGAAGCGGG	/3BioTEG/
5.8S rRNA	TCGACGCACGAGCCGAGTGATCCACCGCTAAGAGTC	/3BioTEG/
5.8S rRNA	AGCGACGCTCAGACAGGCGTAGCCCCGGGAGG	/3BioTEG/

Table 2.2 Primer sequences used in library preparation

Oligo Name	Sequence
oYZ_adapter_V2	/5Phos/NNNNNNAGATCGGAAGAGCACACGTCTGAA/3ddC/
oYZ_V3_RT	/5Phos/AGATCGGAAGAGCGTCGTGTA/iSp18/TTCAGACGTGTGCTC
oYZ_V3_PCR1_F	CTTTCCCTACACGACGCTCTTCC
oYZ_V3_PCR1_R	GACTGGAGTTCAGACGTGTGCTCTTC
oYZ_V2_PCR_F1	aatgatacggcgaccaccgaGATCTACACTCTTTCCCTACACGACGCTC
oYZ_V2_PCR_index01_R	caagcagaagacggcatacgAGATcgtgatGTGACTGGAGTTCAGACGTGT*G
oYZ_V2_PCR_index02_R	caagcagaagacggcatacgAGATACATCGGTGACTGGAGTTCAGACGTGT*G
oYZ_V2_PCR_index05_R	caagcagaagacggcatacgAGATCACTGTGTGACTGGAGTTCAGACGTGT*G
oYZ_V2_PCR_index06_R	caagcagaagacggcatacgAGATATTGGCGTGACTGGAGTTCAGACGTGT*G
oYZ_V2_PCR_index07_R	caagcagaagacggcatacgAGATGATCTGGTGACTGGAGTTCAGACGTGT*G
oYZ_V2_PCR_index08_R	caagcagaagacggcatacgAGATTCAAGTGTGACTGGAGTTCAGACGTGT*G
oYZ_V2_PCR_index09_R	caagcagaagacggcatacgAGATCTGATCGTGACTGGAGTTCAGACGTGT*G
oYZ_V2_PCR_index11_R	caagcagaagacggcatacgAGATGTAGCCGTGACTGGAGTTCAGACGTGT*G
oYZ_V2_PCR_index12_R	caagcagaagacggcatacgAGATtacaagGTGACTGGAGTTCAGACGTGT*G
oYZ_V2_PCR_index15_R	caagcagaagacggcatacgAGATtgacatGTGACTGGAGTTCAGACGTGT*G
oYZ_V2_PCR_index18_R	caagcagaagacggcatacgAGATgcggacGTGACTGGAGTTCAGACGTGT*G
oYZ_V2_PCR_index19_R	caagcagaagacggcatacgAGATTTTCACGTGACTGGAGTTCAGACGTGT*G
oYZ_V2_PCR_index22_R	caagcagaagacggcatacgAGATcgtacgGTGACTGGAGTTCAGACGTGT*G
oYZ_V2_PCR_index25_R	caagcagaagacggcatacgAGATatcagtGTGACTGGAGTTCAGACGTGT*G
*: indicates phosphorothioate bond	

Table 2.3 Primer sequences used for validation experiments

Name	Sequence
BRD2_ex4_FW	CAAAATTATAAAACAGCCTATGGACATG
BRD2_ex5_RV	TTTTCCAGCGTTTGTGCCATTAGGA
BZW1_ex3_FW	TACCGTCGATATGCAGAAACA
BZW1_ex4_RV	GAGCAAATGCTTGCATGGTCT
GOLGA2_FW	CTGCATCGTCTGCTAACCTG
GOLGA2_RV	CGAGGATCCCTATGGTCTGA
GSK3B_FW	GAACTCCAACAAGGGAGCAA
GSK3B_RV	GTGGTGTTAGTCGGGCAGTT
U2AF1 KD siRNA	GCAAGTACGGGCTTGTCAAdTdT

CHAPTER 3

EXPORT OF DISCARDED, SPLICING INTERMEDIATES PROVIDES INSIGHT INTO mRNA EXPORT*

*This chapter is in preparation for publication and adapted from the unpublished manuscript "Export of discarded, splicing intermediates provides insight into mRNA export", with authors Yi Zeng, and Jonathan P. Staley. The experiments described herein were performed by Y. Z..

Abstract

To promote fidelity in nuclear pre-mRNA splicing, the spliceosome rejects and discards suboptimal splicing substrates after they have engaged the spliceosome. Although nuclear quality control mechanisms have been proposed to retain immature mRNPs, discarded splicing substrates, including lariat intermediates, do export to the cytoplasm, as indicated by their translation and/or degradation by cytoplasmic nucleases. However, the mechanism for exporting these species has remained unknown. By single molecule (sm)RNA-FISH and crosslinking in budding yeast, we have observed directly the nuclear export of lariat intermediates and have found that the export of lariat intermediates requires the general mRNA export receptor Mex67p and three of its mRNA export adapter proteins, Nab2p, Yra1p, and Nlp3, establishing that both mRNAs and lariat intermediates utilize the same export machinery. Unexpectedly, by export reporters and smRNA-FISH, the efficient export of lariat intermediates, but not mRNA, requires an interaction between Nab2p and Mlp1p, a nuclear basket component implicated in retaining immature mRNPs, including unspliced pre-mRNA, in the nucleus of budding yeast. Finally, the efficient export of lariat intermediates also relies on the E3 ubiquitin ligase Tom1p and its target sites in Yra1p. Thus, our data indicate that the nuclear basket can promote rather than antagonize the export of an immature mRNP. Further, our data imply that the export of discarded lariat intermediates requires both Mlp1p-dependent docking onto the nuclear basket and subsequent Tom1p-mediated undocking, a mechanism our data suggest generalizes to the export of mRNA but in a manner obscured by redundant pathways.

Introduction

In eukaryotes, the nuclear membrane uncouples translation from mRNA transcription and processing, thereby necessitating mRNA export through the nuclear pore complex (NPC), after

transcription and RNA processing (Niño et al., 2013; Xie and Ren, 2019). Because the NPC establishes a permeability barrier, the export of such substrates requires the binding of export receptors to enable passage through the pore. The export of most mRNAs requires the heterodimeric export receptor – Mex67p-Mtr2p in yeast or TAP-p15 in humans (Grüter et al., 1998; Katahira et al., 1999; Segref et al., 1997). Binding of Mex67p-Mtr2p to mRNA is mediated by adapters, which bind mRNA in a manner that is generally coupled to transcription and RNA processing (Ben-Yishay and Shav-Tal, 2019; Stewart, 2019). Such coupling imparts specificity to Mex67p-Mtr2p binding. In a manner complementary to such coupling, nuclear retention, in budding yeast and humans, has been invoked to enforce quality control mechanisms that restrict the export of immature messages, including unspliced pre-mRNA (Bonnet and Palancade, 2015; Soheilypour and Mofrad, 2018), but an absolute requirement for such quality control mechanisms is challenged by a number of observations that raise questions concerning the actual role of factors implicated in nuclear retention (Aksenova et al., 2019; Mayas et al., 2010; Palazzo and Lee, 2018; Sayani and Chanfreau, 2012). We have gained insight into these issues through an investigation of the proofreading of pre-mRNA splicing.

Pre-mRNA splicing is catalyzed by the spliceosome, a ribonucleoprotein machine that comprises over 80 conserved proteins and five small nuclear RNAs (snRNAs) (Wilkinson et al., 2020). The snRNA components play key roles in recognizing the substrate and catalyzing intron excision through two transesterification reactions; in the first reaction, branching, the 2’ hydroxyl of the branch site attacks the 5’ splice site, yielding a lariat intermediate and a free 5’ exon, which in the second reaction, exon ligation, attacks the 3’ splice site to yield mRNA and excised intron. Proofreading in splicing functions to enhance the specificity of splice site choice (Semlow and Staley, 2012). Grossly suboptimal splice sites never bind the spliceosome, but nearly optimal but

suboptimal splice sites do bind, necessitating downstream mechanisms to distinguish and reject such splice sites. Indeed, proofreading mechanisms have been implicated throughout the splicing cycle, including the phases of spliceosome assembly, activation, and catalysis. ATPases of the DEAD and DExH families of the helicase II superfamily play a central role in proofreading (Semlow and Staley, 2012). In addition to driving the splicing cycle forward in the case of optimal splice sites, these ATPases also reject and discard splicing substrates with suboptimal splice sites. For example, the DExH ATPase Prp22p, in the case of an optimal substrate, promotes the release of mRNA from the spliceosome after exon ligation, but in the case of a suboptimal 3' splice site, Prp22p rejects the splice site by undocking the site from the catalytic core before exon ligation (Mayas et al., 2006). If an alternative, optimal 3' splice site does not engage the spliceosome, then a second DExH ATPase Prp43p, which acts as a general terminator of splicing, disassembles the spliceosome, discarding the suboptimal substrate at the intermediate stage (Mayas et al., 2010). Interestingly, in a few fungal species including fission yeast, this proofreading pathway is repurposed for the biogenesis of telomerase RNA, which corresponds to a cleaved 5' exon in these species (Kannan et al., 2013, 2015). Thus, any mechanism proposed to enforce robust quality control of mRNA export would not only need to distinguish against pre-mRNA that failed to engage the spliceosome but also pre-mRNA that did engage the spliceosome but subsequently suffered discard either at the pre-mRNA or intermediate stage. A robust quality control mechanism to discriminate against the export of such species has been thought to be essential to preclude the translation of truncated protein products (but see below).

Evidence suggests several mechanisms for establishing the quality control of mRNA export, including the selective deposition of export factors and the nuclear retention of immature species (Tutucci and Stutz, 2011; Schmid and Jensen, 2018). For example, in budding yeast, the TREX

complex, which functions in transcription elongation and mRNA export, favors the export of spliced RNA (Tutucci and Stutz, 2011). Further, reporter assays have implicated Mlp1, a subunit of the nuclear basket, as the general quality control factor for retaining faulty RNA transcripts, especially unspliced pre-mRNAs, in the nucleus (Galy et al., 2004; Vinciguerra et al., 2005). However, a number of observations question the requirement for strict quality control mechanisms in mRNA export. First, pre-mRNA and splicing intermediates have been observed directly in the cytoplasm in both budding yeast and humans (Mayas et al., 2010; Harigaya and Parker, 2012; Sayani and Chanfreau, 2012; Carvalho et al., 2017; Talhouarne and Gall, 2018; Hilleren and Parker, 2003; Legrain and Rosbash, 1989). Second, disabling the nonsense-mediated decay machinery also stabilizes pre-mRNAs in the cytoplasm (Sayani and Chanfreau, 2012). Lastly, several studies found that Mlp1p does not act as a general quality control factor, but instead promotes mRNA export (Bangs et al., 1998; Aksenova et al., 2019; Deanna M. Green et al., 2003; Bae et al., 2009; Xu et al., 2007). These observations raise questions about a strict requirement for quality control in mRNA export and the role of factors implicated in nuclear retention, in particular the nuclear basket. Further, the cytoplasmic localization of these species raises questions concerning the pathways utilized for export and their relation to mRNA export; for example, because, for example, a cleaved 5' exon lacks a poly(A) tail and a lariat intermediate lacks a cap, both features of mRNA implicated in promoting export.

The export of mRNA is a multi-step process requiring the orchestration of numerous factors (Ashkenazy-Titelman et al., 2020; Stewart, 2019). First, adapters (see below) are recruited to the RNA during transcription. Then, the adapters recruit the export receptor Mex67p-Mtr2p. After transcription and 3' end formation, the export competent mRNP leaves the site of transcription and transits to the NPC, at which the mRNP docks to the nuclear basket of the NPC,

undocks, and then translocates across the NPC through interactions of Mex67p-Mtr2p with FG repeats that fill the central channel of the NPC (Oeffinger and Zenklusen, 2012). Once the mRNP reaches to the cytoplasmic side of the NPC, the mRNP is captured by cytoplasmic filaments, on which DEAD-box helicase-like ATPase Dbp5p remodels the mRNP to dissociate export factors and release the remodeled mRNP into the cytoplasm for translation (Tran et al., 2007). Live cell imaging of the mRNP export indicated that Dbp5p-mediated remodeling is the slowest step (Grünwald and Singer, 2010).

In yeast, several mRNA export adapters have been identified, including the SR-like protein Npl3p, Nab2p, the TREX-2 complex, and the TREX complex, featuring Yra1p (Tutucci and Stutz, 2011). In addition to its Mex67p-interaction domain, Npl3p includes an RNA-recognition motif (RRM) and an RS-like domain and is recruited during transcription (Tutucci and Stutz, 2011). Nab2p includes zinc-finger domains that interact directly with the polyA tail, a Q-rich motif, and an N-terminal domain that interacts directly with the C-terminal domain of the nuclear basket factor Mlp1p (Grant et al., 2008), providing a mechanism for docking (see below). In addition to binding Mex67p, Yra1p has been implicated in binding to the nuclear basket factors Mlp1p and Mlp2p in an RNA dependent manner (Vinciguerra et al., 2005). Interestingly, the E3 ubiquitin ligase Tom1p dissociates Yra1p from mRNP in the nucleus, thereby promoting mRNA export (Iglesias et al., 2010). Tom1p is required for the export of poly(A) RNA in budding yeast at high temperature (Duncan et al., 2000), but its role in export is unclear. Studies have implicated functions of Tom1p in promoting mRNP docking to the NPC, mRNP undocking from the nuclear basket, or mRNP quality control, ensuring that only mature mRNPs are exported, in part due to genetic evidence that in the absence of Tom1p function, mRNPs are retained by nuclear basket factors Mlp1p and Mlp2p (Iglesias et al., 2010).

Mlp1p and Mlp2p, as well as the mammalian ortholog Tpr, compose the elongated fiber-like structures of the nuclear basket, forming coiled-coils that extend roughly 300 nM into the nucleoplasm (Strambio-de-Castillia et al., 1999; Niepel et al., 2013; Kim et al., 2018). The N-termini of these proteins are anchored to the nuclear membrane via Nup1p and Nup60p and their C-termini extend into the nuclear interior (Bangs et al., 1998; Kim et al., 2018). In addition to the purported roles in retaining immature RNAs in the nucleus, Mlp1p and Tpr have been implicated in mRNA export in *S. pombe*, *Arabidopsis*, and mammalian cells as perturbations of their functions result in the accumulation of poly(A)+ RNA in the nucleus (Bae et al., 2009; Shibata et al., 2002; Xu et al., 2007). Consistent with such a role, evidence indicates that Mlp1 and Mlp2p interact with various mRNA export adapters, including Nab2p, Npl3p, Yra1p, and TREX-2 (Ashkenazy-Titelman et al., 2020; Fasken et al., 2008; Deanna M. Green et al., 2003; Vinciguerra et al., 2005). Further, single particle tracking implies a role of Mlp1p and Mlp2p in docking mRNPs to the NPC, via an interaction with Nab2p, to promote export (Saroufim et al., 2015), a role that also implies a requirement for mRNP undocking, consistent with the residency time of mRNA at the nuclear face of the NPC (Grünwald and Singer, 2010; Saroufim et al., 2015), but mechanisms for undocking remain poorly characterized. As in mammals, in budding yeast overexpression of Mlp1p and Mlp2p results in the accumulation of poly(A)+ RNA, but a knockout of both *MLP1* and *MLP2* is viable and shows no apparent accumulation of poly(A)+ RNA (Bangs et al., 1998; Deanna M. Green et al., 2003; Kosova et al., 2000), suggesting Mlp1p/Mlp2p/Tpr does not play a general role in export. Indeed, recent data from mammals implicates a role for Tpr in the export of only a subset of the transcripts that rely on Nxf1 for export, and another study implicates Tpr specifically in the export of short transcripts (Aksenova et al., 2019; Lee et al., 2019). The lack of a general role for Mlp1p and Mlp2p in mRNA export has suggested a broader role for Mlp1p and its orthologs in

nuclear retention of immature substrates, but the role of Mlp1p in both export and retention remains under investigation.

We have gained a unique perspective into the role of Mlp1p in mRNA export and quality control through an investigation of the export pathway for discarded lariat intermediates. We found that, like the export of mRNA, the export of discarded, lariat intermediates required the export receptor Mex67p and three of its adapters, Yra1p, Nab2p, and Npl3p. Unlike mRNA export, lariat intermediate export required Mlp1p and an interaction between Mlp1p and Nab2p. Further, unlike bulk mRNA export, lariat intermediate export required Tom1p-mediated ubiquitylation of Yra1p, a post-translational modification that releases Yra1p from export cargo in the nucleus. Unexpectedly, our data argue against a general role for Mlp1p in quality control through nuclear retention of immature RNAs and instead highlights a role for Mlp1p in exporting immature RNAs. Specifically, our data suggest a model in which the export of lariat intermediates first requires docking onto the nuclear basket and then undocking to permit transit through the nuclear pore, a pathway that likely also operates in the case of mRNA export but in a manner that is normally masked by redundant pathways.

Results

Export of lariat intermediates requires mRNA export factor Mex67p

To investigate the mechanism of lariat intermediate export in *Saccharomyces cerevisiae*, we tested whether the canonical mRNA export factors are required, given that the assembly of these factors initiates on a transcript co-transcriptionally. To observe the subcellular localization of lariat intermediates, we utilized a *lacZ*-expressing *ACT1* export reporter in which a branch site (br) A-to-G mutation accumulates lariat intermediates that localize to the cytoplasm, based on indirect assays (**Fig. 3.1A, B**; Mayas et al., 2010). To define the subcellular localization of reporters

directly, we targeted the *lacZ* gene in the 3' exon of the derived reporter, named brG, and the wild-type control, named brA, for smRNA-FISH. To resolve individual foci and thereby allow accurate, subcellular RNA counting, we modified the reporter by replacing its strong promoter with the weak *pSTE5* promoter. Consistent with previous indirect assays (Mayas et al., 2010), in a wild-type strain at 30 °C the vast majority of the brG reporter localized to the cytoplasm, as for the brA reporter, with only a minor fraction in the nucleus (16.7%±1.1 (SEM) and 17.2%±1.1, respectively; **Fig. 3.1C**). By primer extension analysis, the brG mutation accumulates lariat intermediate by 7-fold (**Fig. 3.1B**); although the mutation also accumulates pre-mRNA (by 2-fold), the pre-mRNA represents 33% of brG species (**Fig. 3.1B**), which is insufficient to account for the magnitude of the cytoplasmic signal from the brG reporter (83.3%), and a different reporter that accumulates lariat intermediate exclusively also localizes to the cytoplasm by smRNA-FISH (see below). These data provide direct evidence for the nuclear export of lariat intermediates.

By smRNA-FISH as well as RNA-FISH, we examined whether the export of the lariat intermediates requires the general mRNA export receptor Mex67p. Indeed, in a temperature-sensitive *mex67-5* mutant that blocks mRNA export at the non-permissive temperature of 37 ˚C (Segref et al., 1997), both brG and brA reporters accumulated in the nucleus after a temperature shift to 37 ˚C. Specifically, the localization of the brG reporter in the nuclei of *mex67-5* cells increased from 19.1%±1.2 to 74.4%±1.9 (**Fig. 3.1C; Fig. 3.2B**), paralleling an increase in brA-derived mRNAs from 17.9%±0.9 to 70.8%±2.7 and an increase in poly(A)+ RNAs (**Fig. 3.1C; Fig. 3.2B**). In contrast, a shift of wild-type, *MEX67* cells did not significantly increase localization of the brG reporter in the nucleus (15.5%±1.2 at 37 ˚C compared to 16.7%±1.1 at 30 ˚C), as for brA-derived mRNA and poly(A)+ RNA (**Fig. 3.1C, D; Fig. 3.2B**). The tRNA export mutant *los1*Δ did not affect the cytoplasmic localization of the brG reporter, indicating that the export of lariat

intermediates is not promiscuous and requires a specific export pathway (**Fig. 3.2C**). Together, our results indicate that the export of lariat intermediates requires the general mRNA export factor Mex67p.

To test whether Mex67p interacts with lariat intermediates directly, we assayed for the interaction of Mex67p with lariat intermediates *in vivo*. We performed RNA co-immunoprecipitation (co-IP) from extracts of *MEX67-GFP* cells expressing the brG reporter and found that Mex67p-GFP co-immunoprecipitated lariat intermediates (**Fig. 3.2D**). To rule out that the interaction formed *in vitro*, we pre-mixed untagged *MEX67* cells expressing the brG reporter and TAP-tagged *MEX67* cells having a vector control and then performed RNA co-IP from lysates of the cell mixture; in this case, an interaction between Mex67p-TAP and lariat intermediates was not observed, in contrast with an interaction between Mex67p-TAP and *RPL21a* mRNA observed in both experiments (**Fig. 3.1E**), verifying that Mex67p interacts with lariat intermediates *in vivo* – not after cell lysis. Thus, these data support a direct role for Mex67p in lariat intermediate export.

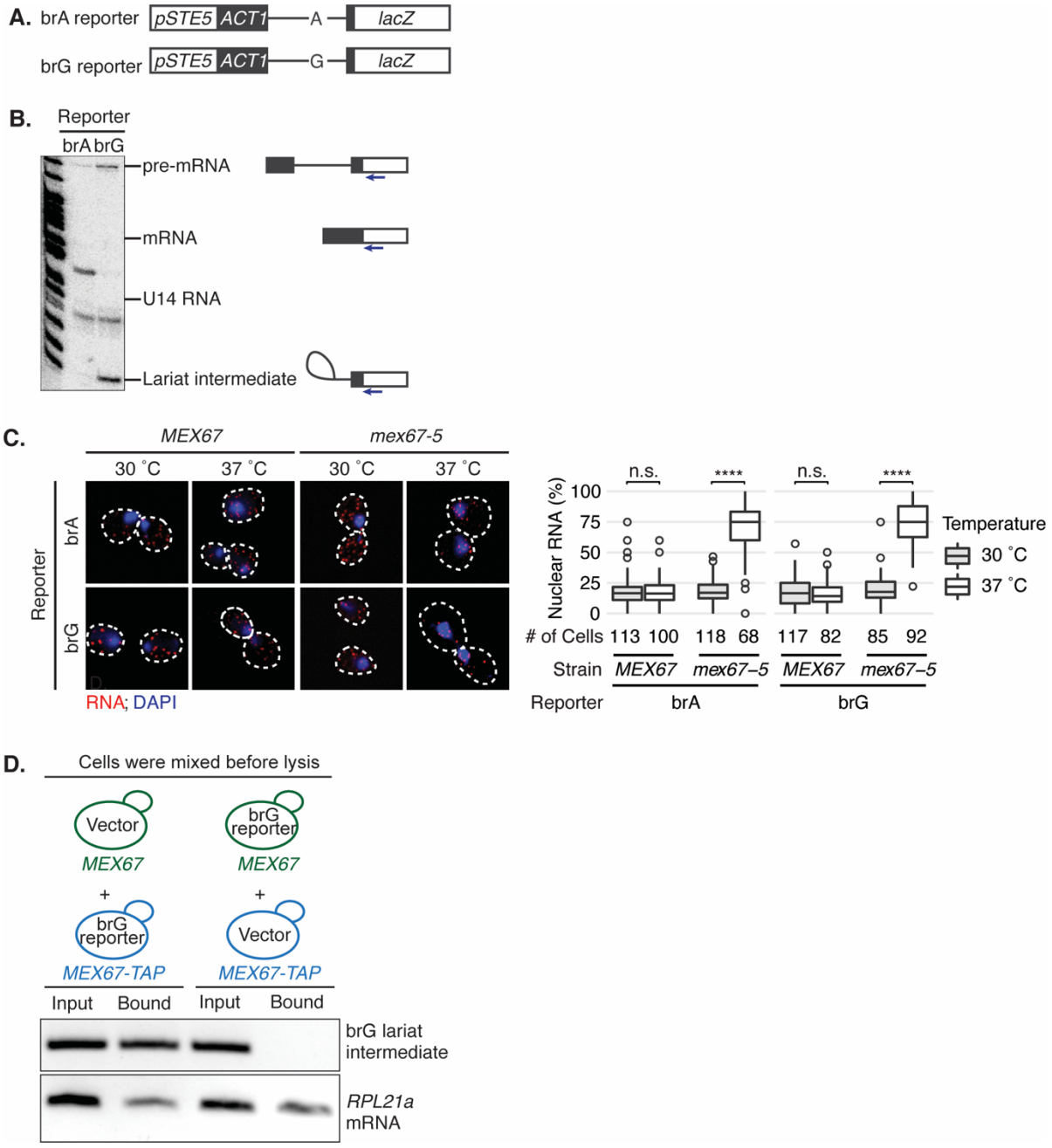


Figure 3.1. Export of a lariat intermediate requires the mRNA export factor Mex67p.
A. Schematic representation of the brA and brG export reporters. **B.** The brG reporter accumulates lariat intermediate, as well as some pre-mRNA, as shown by extension of a primer that anneals to the downstream exon. The migration of the indicated splicing species is shown. **C, D.** By smRNA-FISH, the export of the brG reporter is impeded by the *mex67-5* mutant at the non-permissive temperature. Wild-type or *mex67-5* mutant cells transformed with the indicated reporters were shifted from 30 ˚C to 37 ˚C for 30 minutes. In panel **D**, the fraction of nuclear RNA for each cell was quantitated and displayed as a box plot. The number of cells used for quantitation is indicated beneath each box plot. A *p*-value of 0.05 or less is indicated with asterisks (see the end of legend). The *lacZ* region of the reporter transcript was targeted by Cy3-labeled probes; DNA was probed

Figure 3.1 (continued) by DAPI. Representative cells are shown; dashed lines mark the cell boundary. **E.** By RNA co-IP, Mex67p interacts with lariat intermediates *in vivo*. Top, the indicated cells were mixed before lysis and co-IP. Bottom, the brG lariat intermediate, or *RPL21a* mRNA were detected by RT-PCR before (Input) and after (Bound) IP of Mex67p-TAP, expressed from the indicated strains (*MEX67-TAP*). The *p*-values were calculated by Mann-Whitney test; n.s. (not significant), $p > 0.05$; *, $p \leq 0.05$; **, $p \leq 0.01$; ***, $p \leq 0.001$; ****, $p \leq 0.0001$.

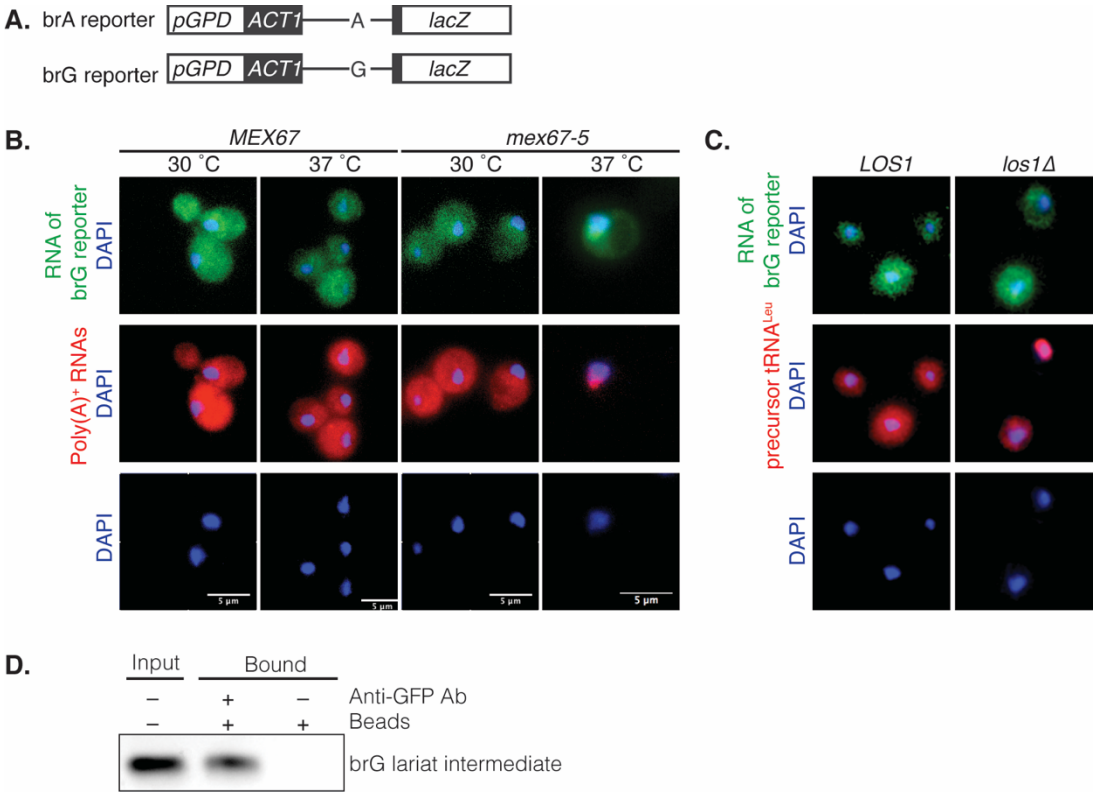


Figure 3.2. Export of a lariat intermediate requires the mRNA export factor Mex67p but not the tRNA export factor Los1.

A. Schematic representation of the brA and brG export reporters driven by *pGPD* promoter. **B.** By RNA-FISH**,** the export of the brG reporter, like poly(A)+ RNA, is impeded by the *mex67-5* mutant at the non-permissive temperature. Cells were shifted from 30 ˚C to 37 ˚C for 30 minutes. The intronic region of the brG reporter transcript was detected by Alexa Fluor 488-labelled probes; poly(A)+ RNA was detected in the same cells by a Cy3-labelled poly(dT)$_{50}$ probe; DNA was probed by DAPI, which is shown separately and overlayed. **C.** By RNA-FISH**,** the export of the brG reporter, unlike pre-tRNA, is not impeded by the *los1*Δ mutant. The intronic region of the brG reporter was detected as in panel A; the intronic region of the tRNA^{Leu} precursor was detected in the same cells by a Cy3-labelled probe (note that pre-tRNA splicing does not involve the spliceosome and occurs in the cytoplasm in budding yeast); DNA was probed by DAPI, which is shown separately and overlayed. **D**. By native RNA co-IP, Mex67-GFP interacts with brG lariat intermediates.

Export of lariat intermediates requires Mex67p export adaptors

In budding yeast, Mex67p is recruited to mRNA transcripts by three different export adaptors, Yra1p, Nab2p, and/or Npl3p (Iglesias et al., 2010). Therefore, we tested whether any of these adaptors is required for the export of lariat intermediates. First, to test for the requirement of Yra1p in the export of lariat intermediates, we tested whether the temperature-sensitive *GFP-yra1-8* mutant, which blocks mRNA export at 37 ˚C (Zenklusen et al., 2002), disrupts the export of lariat intermediates at 37 ˚C. Specifically, we shifted *GFP-yra1-8* cells to 37 ˚C and assessed the cellular localization of transformed reporters by smRNA-FISH. As for the nuclear localization of the brA reporter, the nuclear localization of the brG reporter did not increase in the wild-type *YRA1* cells but did increase in the *GFP-yra1-8* cells and from 27.6%±1.4 to 72.4%±3.6 (**Fig. 3.3A, B**), establishing evidence that the export of lariat intermediates requires Yra1p. To test whether lariat intermediates interact directly with Yra1p, we expressed the brG reporter in *HA-YRA1* cells, formaldehyde crosslinked the cells, performed denaturing RNA co-IP from extracts, and then assayed for RNA by RT-qPCR. Indeed, HA-Yra1p not only co-immunoprecipitated endogenous *RPL21a* mRNA, as expected, but also brG-derived lariat intermediates (**Fig. 3.3C**). These data support a direct role for Yra1p in the export of lariat intermediates.

Next, to test for a requirement of Nab2p, which interacts with polyA tails, in the export of lariat intermediates, we assessed by smRNA-FISH whether the export of the brG reporter was compromised in a cold-sensitive *nab2*-Δ*N* mutant, which accumulates poly(A)+ RNA in the nucleus at 16 ˚C (Marfatia et al., 2003). Indeed, when *nab2*-Δ*N* cells transformed with the reporters were shifted to 16 ˚C, the nuclear localization of the brG reporter increased from 35.1%±4.0 to 52.4%±3.2, similar to the increase of 31.6%±2.1 to 47.3%±2.4 for the brA reporter (**Fig. 3.3D, E**), implying that Nab2p is also required for the export of lariat intermediates. To test whether Nab2p

interacts directly with lariat intermediates, we expressed the brG reporter in HTB-tagged *NAB2-HTB* cells, UV-crosslinked the cells, performed denaturing RNA pull-down from extracts, and then assayed for associated RNAs by RT-qPCR. Indeed, Nab2p-HTB not only pulled down *RPL21a* mRNA in a UV-dependent manner, as expected, but also brG-derived lariat intermediate (**Fig. 3.3F**). These data support a direct role for Nab2p in the export of lariat intermediates.

Finally, to examine whether Npl3p plays a role in the export of lariat intermediates, we tested whether a temperature-sensitive *npl3-1* mutant, which displays a strong defect in mRNA export after a shift to 37 ˚C (Lee et al., 1996), blocks the export of lariat intermediates. After a shift from 25 ˚C to 37 ˚C for 2 hours, the nuclear localization of the brG reporter, as well as the brA reporter, increased in the *npl3-1* cells, from 24.3%±3.1 to 63.1%±3.9 and 32.6%±2.4 to 59.9%±2.8, respectively (**Fig. 3.3G, H**). An Npl3-TAP IP was unsuccessful. Thus, like Yra1p and Nab2p, Npl3p contributes to the export of lariat intermediates. Taken together, these results establish that the export of lariat intermediates requires Mex67p and its adaptors and therefore the canonical mRNA export pathway.

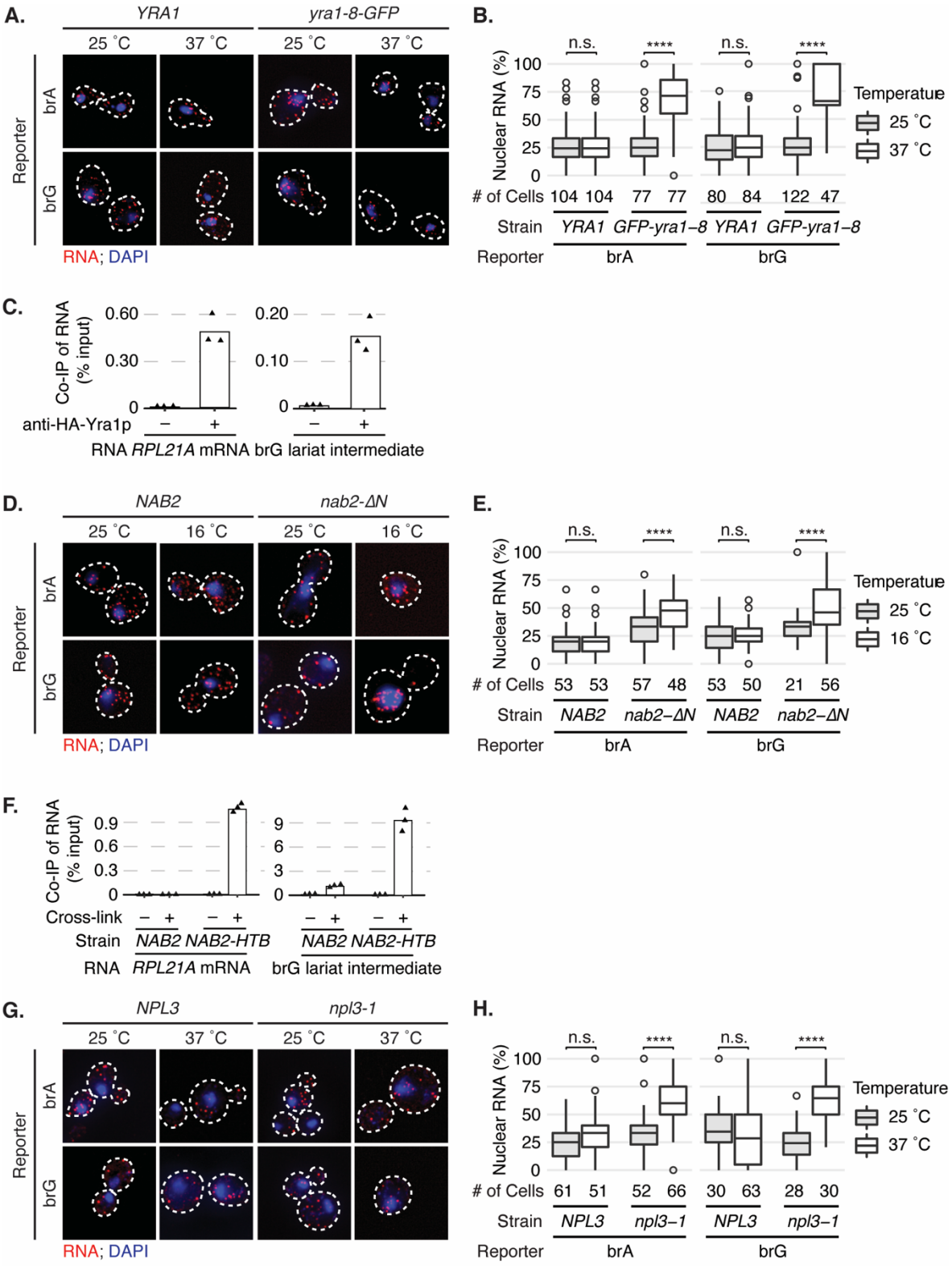


Figure 3.3. Export of a lariat intermediate requires the mRNA export adapters Yra1p, Nab2p, and Npl3p.

A, B. By smRNA-FISH**,** export of the brG reporter requires *YRA1*. In panel **A**, wild-type or

Figure 3.3. (continued) *yra1-8-GFP* mutant cells were shifted from 25 ˚C to 37 ˚C for 2 hours. Panel **B** shows the quantitation of the nuclear fraction of reporter RNA. **C.** By RNA co-IP, Yra1p interacts directly with lariat intermediates. After formaldehyde crosslinking, HA-Yra1p was IP'd from cell extracts under denaturing conditions with or without antibodies, and then associated brG lariat intermediate or *RPL21a* mRNA were detected by RT-qPCR. The levels of IP'd RNA are shown as a percentage of input for three technical replicates; the height of the bar indicates the mean. **D, E.** By smRNA-FISH, the export of the brG reporter requires *NAB2*. Wild-type or *nab2-*Δ*N* mutant cells were shifted from 30 ˚C to 16 ˚C for 2 hours. Panel **D** shows representative images; panel **E** shows quantitation of the nuclear fraction of reporter RNA. **F.** By RNA co-IP, Nab2p interacts directly with lariat intermediates. HTB-tagged *NAB2-HTB* cells and untagged control cells were treated with and without UV, Nab2p-HTB was pulled down from cell extracts under denaturing conditions using Ni-NTA agarose, and then associated brG lariat intermediate or *RPL21a* mRNA was assayed by qRT-PCR. The levels of pulled-down RNA were quantitated and illustrated as in panel **C**. **G, H.** By smRNA-FISH**,** the export of the brG reporter requires *NPL3*. Wild-type or *npl3-1* mutant cells were shifted from 25 ˚C to 37 ˚C for 2 hours. Panel **G** shows representative images; panel **H** shows quantitation of the nuclear fraction of reporter RNA. Throughout, for smRNA-FISH cells were probed and demarked as in Fig. 3.1C (left panel); the nuclear fraction was calculated and displayed as in Fig. 3.1C (right panel).

Efficient export of lariat intermediates requires the nuclear basket component Mlp1p

Because previous studies have determined that Nab2p physically interacts with the nuclear basket factor Mlp1p and that the *nab2*-Δ*N* mutant lacks the Mlp1p-interacting domain (Marfatia et al., 2003; Grant et al., 2008), we hypothesized that Nab2p mediates lariat intermediate export through its interaction with Mlp1p, even though Mlp1p has been implicated as a quality-control factor that retains immature RNPs in the nucleus (Bonnet and Palancade, 2015; Galy et al., 2004; Vinciguerra et al., 2005). As a first test of this hypothesis, we assayed whether lariat intermediate export requires Mlp1p. To assay export efficiency, we used our previously described *lacZ*-based export reporter having an internal ribosomal entry site (IRES) in the second exon just upstream of *lacZ*, which enables the translation of exported species, whether pre-mRNA, lariat intermediate, or mRNA (**Fig. 3.4A**; Mayas et al., 2010). Indeed, deletion of *MLP1* significantly reduced the β-galactosidase activity of the brG-IRES reporter by 50%; by contrast deletion of *MLP1* did not significantly affect the β-galactosidase activity of the brA-IRES reporter (**Fig. 3.4B**), consistent with previous observations (Mayas et al., 2010). This reduced β-galactosidase activity in *mlp1*Δ cells cannot be accounted for by changes in RNA levels, which were not perturbed (**Fig. 3.5B, C**). Similarly, in *mlp1*Δ cells a G1a reporter that accumulates lariat intermediates had reduced β-galactosidase activity (**Fig. 3.5B, C, D**). These data are consistent with a requirement for Mlp1p in the efficient export of lariat intermediates.

To determine directly whether *MLP1* promotes the export of lariat intermediates, we used smRNA-FISH to examine the cellular localization of the brG reporter in *mlp1*Δ cells. Consistent with our findings by the β-galactosidase assay, the *mlp1*Δ mutation decreased the cytoplasmic signal and concomitantly increased the nuclear signal of the brG reporter; the nuclear fraction increased by 1.8-fold – from 26.9%±1.2 in wild-type *MLP1* cells to 49.1%±1.5 in *mlp1*Δ cells (**Fig.**

3.4C). By contrast, the *mlp1*Δ mutation did not impact the subcellular localization of the brA reporter (**Fig. 3.4C**). These data support a role for Mlp1p in the export of lariat intermediates, specifically.

Because the brG reporter accumulates some pre-mRNA (38%), in addition to lariat intermediate (58%; **Fig. 3.5B, C**) and the *mlp1*Δ mutant only partially compromised export of the brG reporter (by 50%; **Fig. 3.4B**), we tested whether the dependence of the brG reporter on Mlp1p might reflect a dependence of the export pre-mRNA export, in addition to lariat intermediate, on Mlp1p. To test this possibility, we assessed the impact of the *mlp1*Δ deletion on the export of an isogenic reporter having a G1c mutation at the 5' splice site, a mutation that accumulates similar levels of pre-mRNA (2.4-fold, relative to wild type; **Fig. 3.5B, C**) but no lariat intermediate; in fact, the levels of lariat intermediate are even lower than for a wild-type reporter (**Fig. 3.5B**). Significantly, the *mlp1*Δ deletion did not reduce the levels of β-galactosidase expressed from the G1c reporter, relative to the wild-type strain (**Fig. 3.5E)**. Similarly, in *mlp1*Δ cells, a brC reporter that similarly accumulates pre-mRNA did not reduce β-galactosidase activity (**Fig. 3.5B, C, E**). These data provide strong evidence that the export of brG pre-mRNA is not dependent on Mlp1p; these data, therefore, imply that the efficient export of brG lariat intermediate is dependent on Mlp1p.

To test for a requirement for Mlp1p in the export of lariat intermediates explicitly, we assessed the impact of the *mlp1*Δ mutation on the export of an isogenic reporter having a UAc mutation at the 3' splice site. Like the brG mutation, the UAc mutation compromises exon ligation; unlike the brG mutation, the UAc mutation does not also substantially compromise the conversion of pre-mRNA to lariat intermediate, so UAc pre-mRNA levels are insignificant (e.g. **Fig. 3.6A**). However, whereas the brG lariat intermediate, having a non-consensus branch linkage, is resistant

to debranching by Dbr1 and turnover; the UAc lariat intermediate, with a consensus branch linkage, is subject to rapid debranching and turnover. Nevertheless, in a *dbr1*Δ strain the UAc lariat intermediate is stabilized (e.g. **Fig. 3.6A, B**), and we have shown previously that this stabilized lariat intermediate is efficiently translated in the cytoplasm (**Fig. 3.4D**; Mayas et al., 2010). Thus, we tested in a *dbr1*Δ background whether the export of the UAc lariat intermediate requires *MLP1* for export. Indeed, the *mlp1*Δ mutation compromised export; specifically, whereas the *mlp1*Δ mutation did not significantly reduce the β-galactosidase activity of the brA-IRES, the *mlp1*Δ mutation reduced the β-galactosidase activity of the UAc-IRES reporter by roughly 40% (**Fig. 3.4D**), similar to the brG reporter (**Fig. 3.4B**). These data provide compelling evidence that *MLP1* is required explicitly and specifically for the efficient export of lariat intermediates.

To test directly for a requirement for *MLP1* in the export of the UAc lariat intermediate, we assayed for localization of the UAc reporter in the *dbr1*Δ background by smRNA-FISH. Whereas the *mlp1*Δ mutation did not impact the subcellular localization of the brA reporter, the *mlp1*Δ mutation did shift the localization of the UAc reporter from the cytoplasm to the nucleus; specifically, we observed a shift in the nuclear fraction from 33.1%±1.5 to 45.4%%±2.7 (**Fig. 3.4E**), similar to the shift of the brG reporter in a *DBR1* background (**Fig. 3.4C**); by contrast, the *mlp1*Δ mutation did not shift the localization of the brA reporter from the cytoplasm to the nucleus (**Fig. 4E**). These data provide direct evidence for a requirement for *MLP1* in efficient lariat intermediate export.

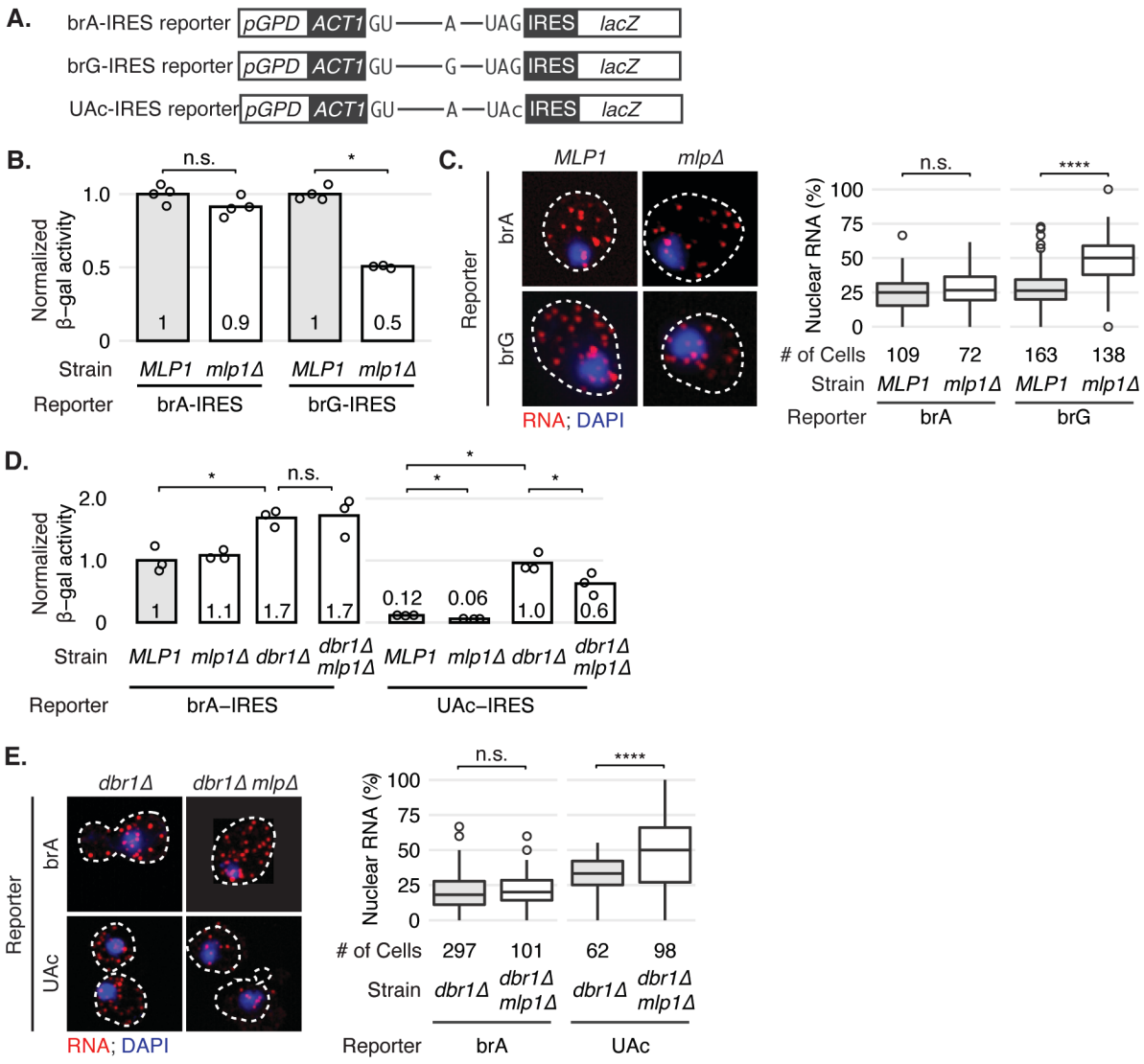


Figure 3.4. The export of lariat intermediates requires Mlp1p.

A. Schematic representation of the *ACT1-IRES-lacZ* export reporters. **B, C.** Efficient export of the brG but not the brA reporter requires *MLP1*. In panel **B**, *MLP1* or *mlp1*Δ cells were grown at 30 °C, and the cytoplasmic localization of the reporters was assayed by measuring the β-galactosidase activity of cell extracts. The normalized β-galactosidase activities of four technical replicates are shown; the height of the bar and the number indicate the normalized mean activity. Values for each reporter were normalized to the mean activity of the same reporter observed in the *MLP1* strain; a *p*-value of 0.05 or less is indicated with asterisks (see the end of legend). In panel **C**, *MLP1* or *mlp1*Δ cells were grown at 30 °C, and the cytoplasmic localization of the reporters was assayed directly by smRNA-FISH (left); the nuclear fraction of reporter RNA was quantitated (right). **D, E**. Export of the UAc lariat intermediate is comprised by the *mlp1*Δ mutation. In panel **D**, *MLP1*, *mlp1*Δ, *dbr1*Δ, or *mlp1*Δ *dbr1*Δ cells were grown at 30 °C, and the cytoplasmic localization of the brA- and UAc-IRES reporters was assayed by measuring β-galactosidase activity of cell extracts. The β-galactosidase activity was quantitated as in B; values were normalized to the brA-IRES reporter in the wild-type strain (*MLP1*). In panel **E**, mutant *dbr1*Δ or double mutant *mlp1*Δ *dbr1*Δ cells were grown at 30 °C and then assayed for reporter localization

Figure 3.4 (continued) by smRNA-FISH (left); the nuclear fraction of reporter RNA was quantitated (right). Throughout, for smRNA-FISH cells were probed and demarked as in Fig. 3.1C; the nuclear fraction was calculated and displayed as in Fig. 3.1D. The *p*-values were calculated by Mann-Whitney test.

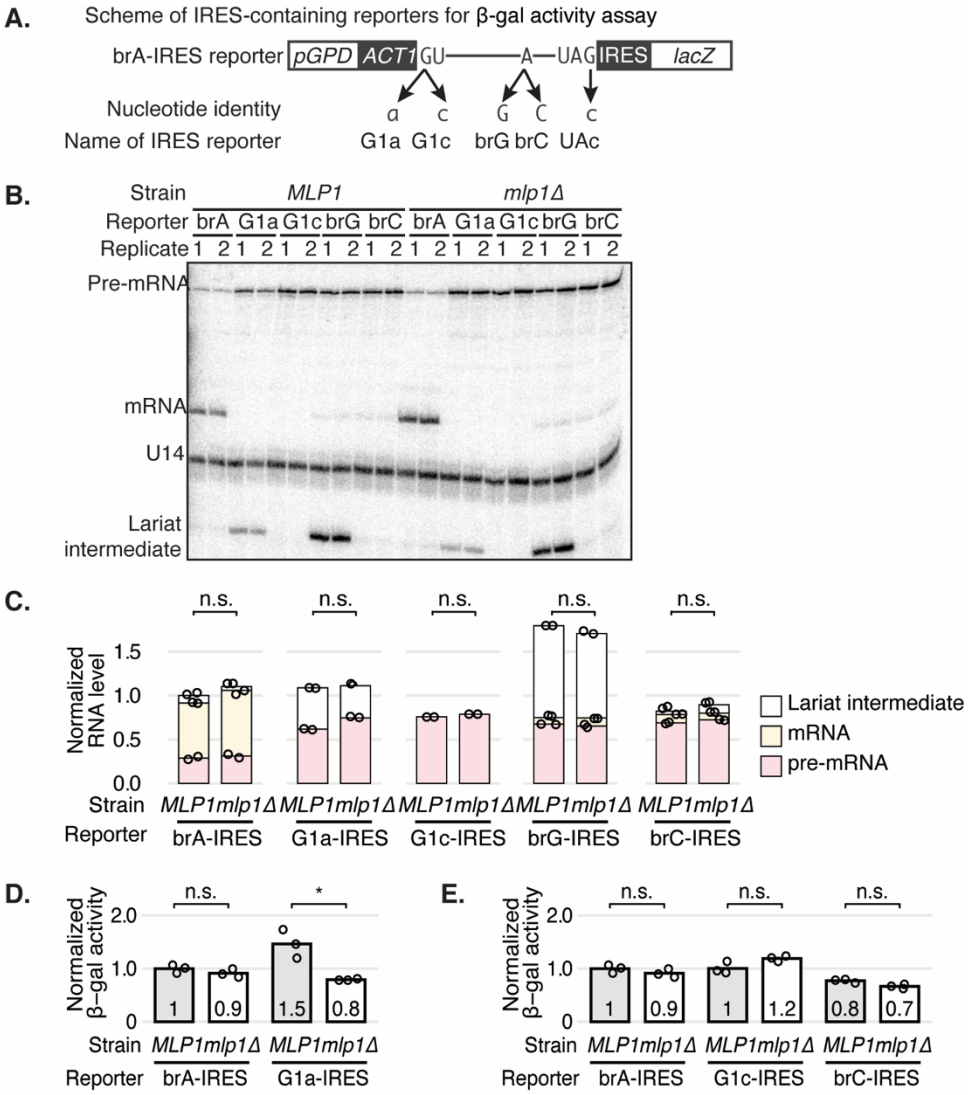


Figure 3.5. Whereas the export of lariat intermediates requires MLP1, the export of pre-mRNA does not.

A. Schematic representation of different *ACT1-IRES-lacZ* export reporters driven by the *pGPD* promoter. **B, C**. The *mlp1*Δ mutation does not significantly impact the levels of splicing species for IRES-containing reporters containing either the brA, brG, or other mutated splice sites, thereby supporting a direct impact of the *mlp1*Δ mutation on the export of brG lariat intermediates. In panel **B**, splicing of each reporter was analyzed in *MLP1* and *mlp1*Δ cells, grown at 30 °C. The levels of splicing precursor, lariat intermediate, and mRNA, were detected by extension of a primer that annealed to the downstream exon; the levels of the snoRNA U14 were detected similarly and served as an internal control; two pairs are shown. Panel **C** shows normalized RNA levels of different splicing species from two technical replicates in B; the height of bar indicates the mean RNA level. Levels of pre-mRNA (salmon), lariat intermediate (yellow), and mRNA (white) were stacked and normalized to RNA levels of the *brA-IRES-lacZ* reporter in the *MLP1* strain to reflect total 3′ exon levels. **D**. Export of the G1a lariat intermediate is comprised by the *mlp1*Δ mutation. *MLP1* or *mlp1*Δ cells were grown at 30 °C, and the cytoplasmic localization of the brA- and *G1a-IRES-lacZ* reporters was assayed by measuring β-galactosidase activity of cell extracts. Activity

Figure 3.5 (continued) was quantitated as in Fig. 3.3B; values were normalized to the *brA-IRES-lacZ* reporter in the *MLP1* strain. **E**. Export of pre-mRNA is not comprised by the *mlp1*Δ mutation. *MLP1* or *mlp1*Δ cells were grown at 30 °C, and the cytoplasmic localization of the brA-, *G1c*- and *brC-IRES-lacZ* reporters was assayed by measuring β-galactosidase activity of cell extracts. Activity was quantitated as in Fig. 3.4B; values were normalized to the brA-IRES reporter in the *MLP1* strain. The activities for brA reporter were reproduced from **D**.

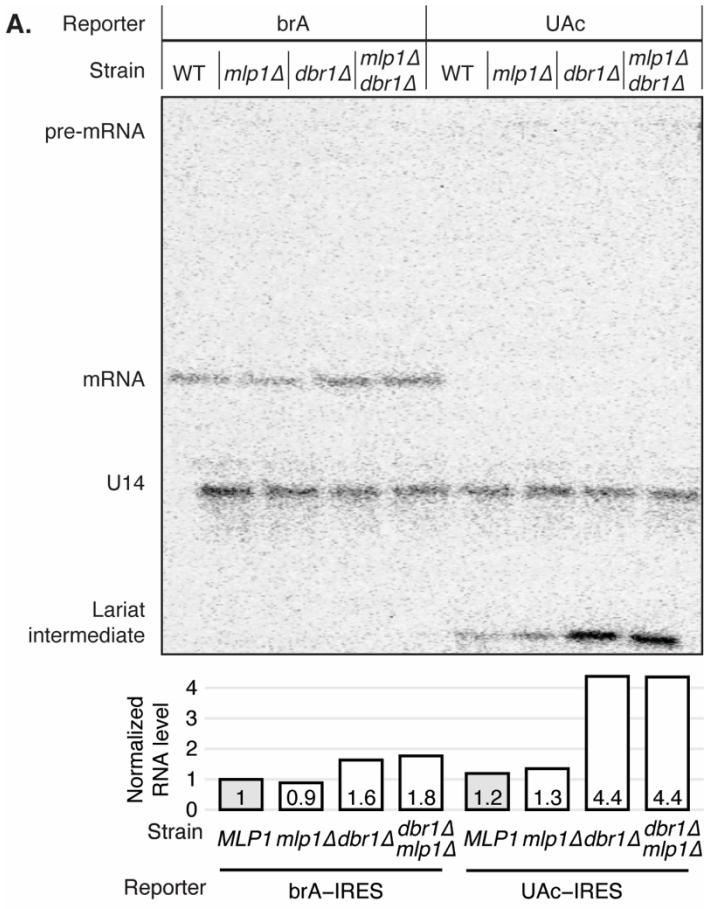


Figure 3.6. The mlp1Δ mutation does not significantly impact the levels of splicing species for either the brA- or UAc-IRES reporters in the presence or absence of Dbr1p.
In panel **A**, *MLP1*, *mlp1*Δ, *dbr1*Δ, or *mlp1*Δ *dbr1*Δ cells were grown at 30 °C and splicing of the reporters was analyzed in *MLP1* and *mlp1*Δ cells by primer extension, as described in panel Fig. 3.S2B. Panel **B** shows the quantitation of RNA levels. Values were normalized to the brA-IRES reporter in the *MLP1* strain. Note that the lariat intermediate dominates the splicing species in the UAc-IRES reporter, and the *dbr1*Δ stabilizes this species, specifically.

Export of lariat intermediates requires an interaction between Nab2p and Mlp1p

As a second test of our hypothesis that Nab2p mediates lariat intermediate export through its interaction with Mlp1p, we probed more deeply for a requirement for the Nab2p-Mlp1p interface in the export of lariat intermediates. First, we tested for a requirement for the Mlp1p component of the interface utilizing the *mlp1-Δ1586-1768* mutation, which lacks the Nab2p-interacting domain. This mutation compromised the β-galactosidase activity of the brG-IRES reporter; specifically, expression of wild-type *MLP1* in *mlp1*Δ cells fully restored β-galactosidase activity of the brG-IRES reporter; whereas expression of the *mlp1-Δ1586-1768* mutated gene failed to do so (**Fig. 3.7A**). By fluorescence microscopy, we confirmed that the mutated Mlp1p-Δ1586-1768 protein is localized properly to the nuclear periphery, as observed for Mlp1p (Fig. 3.S4A; Galy et al., 2004; Niepel et al., 2013). These results are consistent with the hypothesis that efficient lariat intermediate export requires the interaction between Nab2p and Mlp1p.

We showed above that the cold-sensitive *nab2-ΔN* mutant, which lacks the Mlp1p-interacting domain, compromises export of the brG reporter at low temperature (**Fig. 3.3D, E**), but this mutant also compromises the export of the brA reporter (**Fig. 3.3D, E**), in addition to poly(A)+ RNA, phenotypes that do not parallel the phenotypes of *mlp1* mutants, so we next assayed whether lariat intermediate export was sensitive to a subtler *NAB2* double mutant, *nab2-F72A/F73A*, that also disrupts the Nab2p-Mlp1p interaction (Fasken et al., 2008; Grant et al., 2008) but displays no growth defect at any temperature (**Fig. 3.8B**). Indeed, although the *nab2-F72A/F73A* mutation did not significantly affect the β-galactosidase activity of the brA-IRES reporter, the *nab2-F72A/F73A* mutation did reduce the β-galactosidase activity of the brG-IRES reporter by 60% (**Fig. 3.7B**), quantitatively similar to the reduction of β-galactosidase activity observed for *mlp1*Δ or *mlp1-Δ1586-1768* mutant cells (**Fig. 3.4B; Fig. 3.7A**). The diminished β-galactosidase activity of the

brG *lacZ* reporter in *nab2-F72A/F73A* cells did not result from changes in RNA levels, because the *nab2-F72A/F73A* mutant did not alter RNA levels of either reporter (**Fig. 3.8C, D**). These data are also consistent with the hypothesis that efficient lariat intermediate export requires the interaction between Nab2p and Mlp1p.

To determine directly whether the Nab2p-Mlp1p interaction promotes the export of lariat intermediates, we used smRNA-FISH to examine the cellular localization of the brG reporter in *nab2-F72A/F73A* mutant cells. As in *mlp1*Δ cells (**Fig. 3.4C**), in *nab2-F72A/F73A* cells the nuclear fraction of the brG reporter increased by 2.1-fold, from 22.7%±2.5 in wild-type *NAB2* cells to 47.6%±4.3 in *nab2-F72A/F73A* cells, whereas the nuclear fraction of the brA reporter did not change/increase significantly (**Fig. 3.7C**). Together, these results establish that the efficient export of lariat intermediates in particular requires the interaction between Nab2p and Mlp1p.

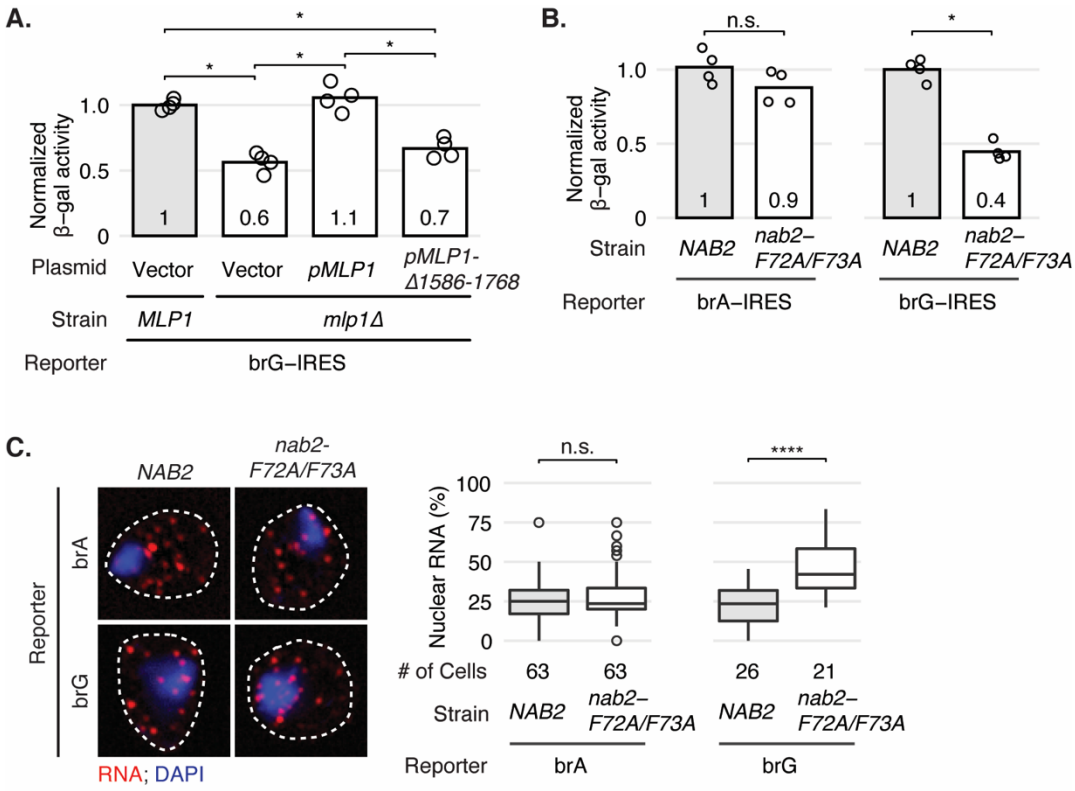


Figure 3.7. The export of lariat intermediates requires an interaction between Nab2p and Mlp1p.

A. Deletion in the Nab2p-interacting domain of Mlp1p impedes the efficient export of the brG reporter. Mutant *mlp1*Δ cells were transformed with a vector control, *MLP1*, or *mlp1*Δ*1586-1768*, grown at 30 °C, and then lysed to assay β-galactosidase activity of the brG reporter. The activity was quantitated as in Fig. 3.3B; values were normalized to the brG reporter in the *MLP1* strain transformed with a vector control. Panels **B** and **C** show that a double mutation in the Mlp1p-interacting domain of Nab2p impedes the efficient export of the brG reporter, specifically. In panel **B**, *NAB2* or *nab2-F72AF73A* cells were grown at 30 °C and then lysed to assay β-galactosidase activity. The activity was quantitated as in Fig. 3.3B; values for each reporter were normalized to the mean activity of the same reporter observed in the *NAB2* strain. In panel **C,** *NAB2* or *nab2-F72AF73A* cells were grown at 30 °C and then assayed for reporter localization by smRNA-FISH (left); cells were probed and demarked as in Fig. 3.1C. The nuclear fraction of reporter RNA was quantitated and displayed, on the right, as in Fig. 3.1C.

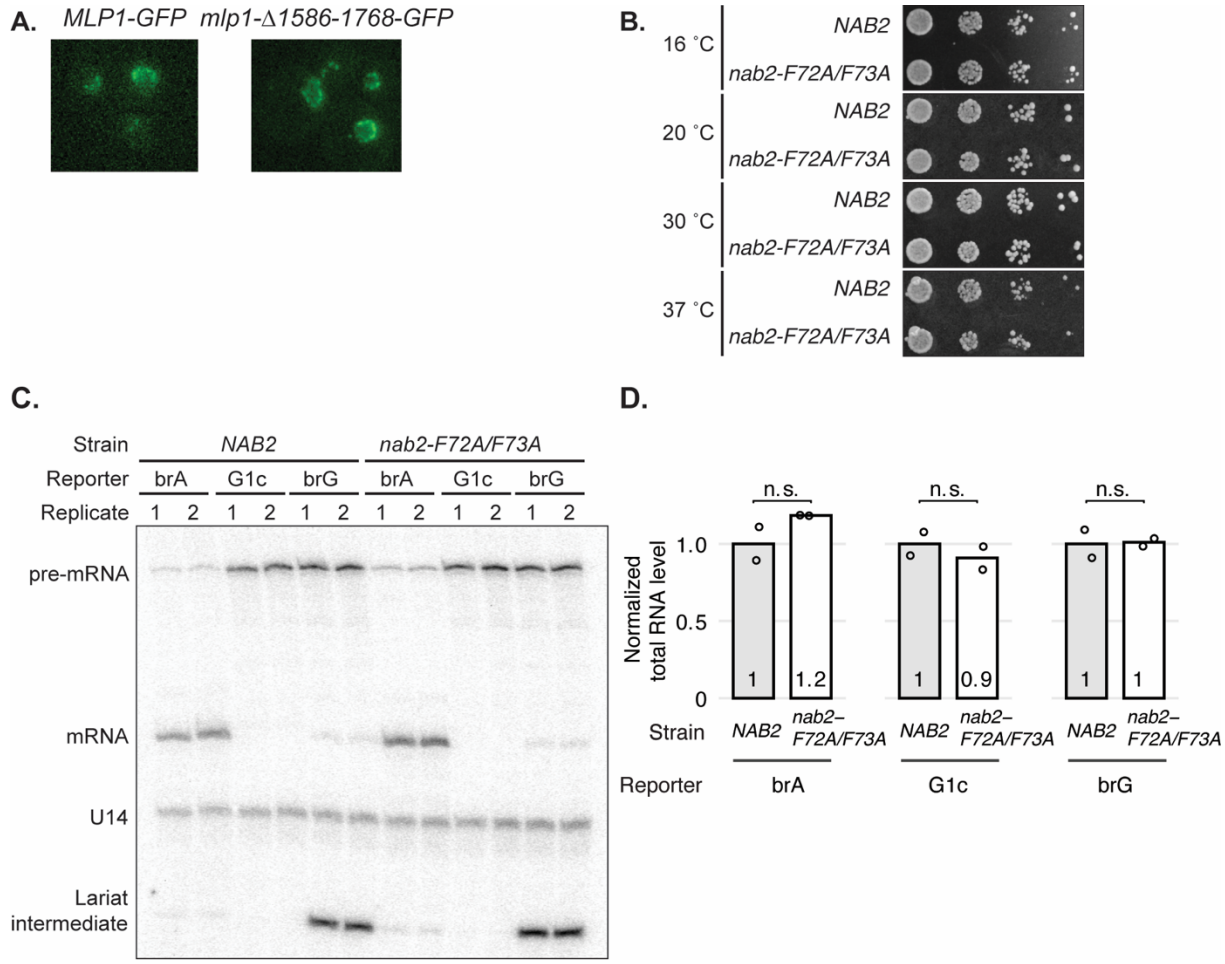


Figure 3.8. Deletion of Nab2-interacting domain Mlp1p does not compromise Mlp1p localization and nab2-F72A/F73A does not compromise either growth or splicing.
A. Mutated mlp1p-Δ1586-1768-GFP localizes to the nuclear periphery, just as wild-type Mlp1p-GFP does. GFP was probed (in live cells) by epi-fluorescence microscopy. Note that the crescent shape is characteristic of Mlp1p localization, which reflects exclusion from the nuclear periphery where the nucleolus resides. Also note that whereas mlp1p-Δ1586-1768-GFP lacks the Nab2p-interacting domain, the mutated protein does include its nuclear localization signal. **B.** The subtle *nab2-F72A/F73A* mutant displays no growth defect over a range of temperatures. Growth of the indicated strains was assayed at the indicated temperatures by spotting serial dilutions of cultures onto rich, solid media (YPDA plates) and incubating for 5 days at 16 °C, 4 days at 20 °C, 2 days at 30 °C, and 1 day at 37 °C. **C, D.** The *nab2-F72A/F73A* double mutation does not impact the splicing of the brA, G1c, or brG reporters. In panel **C**, splicing was analyzed by primer extension as in Fig. 3.S2B. Panel **D** shows the quantitation of total RNA level. Total RNA of each reporter was normalized to its RNA level in the *NAB2* strain.

Export of lariat intermediates requires Tom1p-mediated ubiquitylation of Yra1p

Yra1p is ubiquitylated by the E3 ligase Tom1p, and this ubiquitylation is required for efficient poly(A)+ mRNA export at a higher temperature (37 ˚C; Iglesias et al., 2010). Additionally, based on genetic interactions between Tom1p or Yra1p and Mlp1p or Mlp2p, Tom1p has been proposed to surveil mRNAs for maturity at the nuclear basket (Iglesias et al., 2010). Thus, we investigated the impact of Tom1p on lariat intermediate export. First, we assayed for the consequences of deleting *TOM1* on the export of the brA- and brG-IRES reporters. At the permissive temperature (30 °C), the *tom1*Δ mutation did not significantly reduce the β-galactosidase activity of the brA-IRES reporter, but the deletion mutation did significantly reduce the activity of the brG-IRES reporter (**Fig. 3.9A**), suggesting unexpectedly that *TOM1, like MLP1,* promotes, rather than antagonizes, the export of incompletely processed mRNPs; indeed, our data imply that Tom1p, like Mlp1p and its interaction with Nab2p (**Fig. 3.4B; Fig. 3.7C**), is more important for the export of incompletely processed mRNPs than mature mRNPs.

Given this evidence that Tom1p is required specifically for the export of the brG-IRES reporter, we next tested whether ubiquitylation of the Tom1p target Yra1p is also required, exploiting the mutant *yra1-KR-all*, which mutates all potential ubiquitylation sites and precludes Tom1p-mediated ubiquitylation. At the semi-permissive temperature of 25 ˚C, the *yra1-KR-all* mutation reduced the β-galactosidase activity of the brA reporter but by a modest degree (20%; **Fig. 3.9B**), in contrast with the *nab2*, *mlp1*, and *tom1* mutants; this reduction may reflect the growth defect of this strain at 25 ˚C and/or the mutation of all lysines, which are not all functional targets for Tom1p (**Fig. 3.10**; Iglesias et al., 2010). The *yra1-KR-all* mutation reduced the β-galactosidase activity of the brG-IRES reporter by a greater degree (52%; **Fig. 3.9B**). Notably, the *mlp1*Δ, *nab2-F72A/F73A, tom1*Δ, and *yra1-KR-all* mutants all reduced the β-galactosidase activity

of the brG-IRES reporter by a similar degree, roughly 50% (**Fig. 3.4B; Fig. 3.7B; Fig. 3.9A, B**), suggesting that Nab2p, Mlp1p, Tom1p, and Yra1p act in the same pathway.

Because Tom1p-mediated ubiquitylation of Yra1p promotes dissociation of Yra1p from its bound mRNPs before nuclear exit (Iglesias et al., 2010), we examined whether Yra1p ubiquitylation modulated the association of Yra1p with lariat intermediates, as documented in Fig. 3.2C, at 25 ˚C. By formaldehyde cross-linking followed by RNA co-IP using HA-tagged Yra1p, the *yra1-KR-all* mutation increased, as expected, binding to a control mRNA (*RPL21A*) by 1.5-fold; significantly, the *yra1-KR-all* mutation increased binding to the brG lariat intermediate by 3.2-fold (**Fig. 3.9C**). These data indicate a role for Tom1p-mediated ubiquitylation of Yra1p in promoting the export of lariat intermediates through dissociation of Yra1p.

Although Tom1p-mediated ubiquitylation of Yra1p has been found to promote its dissociation from mRNP before nuclear exit, it is still not clear whether Tom1p functions before or after mRNPs dock onto the nuclear pores. Genetic evidence has suggested that Tom1p acts after mRNPs dock to the nuclear pores (Iglesias et al., 2010). However, live tracking of newly born mRNAs from an induced reporter implied that Tom1p promotes mRNP docking, that Tom1p acts before mRNP docking (Saroufim et al., 2015). To distinguish these two models, we first examined directly the localization of lariat intermediates and mRNAs. By smRNA-FISH, after a temperature shift of *tom1*Δ to 37 °C, the brG reporter, as well as the brA reporter, localized primarily to the nucleus (78.1%±1.6 and 82.2%±1.8, respectively), relative to the wild-type strain (35.2%±1.4 and 29.6%±1.1, respectively; **Fig. 3.9D**); similarly, after a temperature shift of the *yra1-KR-all* strain to 37 °C, the brG reporter, as well as the brA reporter, localized primarily to the nucleus (68.4%±3.0 and 74.2%±3.3, respectively), relative to the wild-type strain (32.2%±2.1 and 33.8%±3.3, respectively; **Fig. 3.9E**). These data indicate that at a non-permissive temperature of

37 °C, Tom1p-mediated ubiquitylation of Yra1p is required for the efficient export of not just mature poly(A)+ mRNA but also an immature RNA.

Together with the observation that *MLP1* and *MLP2* mutations suppress *TOM1* and *YRA1* mutants (Iglesias et al., 2010), our smRNA-FISH results suggest that mRNPs in *tom1*Δ and *yra1-KR-all* mutant cells are stalled at the nuclear basket (**Fig. 3.9D, E**). To test this idea, we examined whether the disruption of the nuclear basket in the *tom1*Δ strain would rescue the export defect. Consistent with previous genetic data showing that deletion of *MLP2* rescues the temperature sensitivity of the *tom1*Δ strain (Iglesias et al., 2010), deletion of *MLP2* in the *tom1*Δ strain substantially rescued the export defect of the brA reporter, releasing stalled mRNPs from the nuclear periphery to the cytoplasm and reducing nuclear localization from 82.2%±1.8 to 54.9%±1.8 and increasing cytoplasmic localization by 2.5-fold (p-value=2.2×10^{-16}**; Fig. 3.9F**); similarly, deletion of *MLP1* in the *tom1*Δ strain modestly but significantly reduced nuclear localization from 82.2%±1.8 to 74.4%±1.7 and increased cytoplasmic localization by 1.4-fold (p-value=8.682×10^{-4}; **Fig. 3.9F**). Interestingly, deletion of *MLP2* but not *MLP1* in the *tom1*Δ strain also slightly but significantly rescued the export defect of the brG reporter, reducing nuclear localization from 78.1%+1.57 to 67.5%+3.1 and increased cytoplasmic localization by 1.5-fold (p-value=0.003; **Fig. 3.9F**). The smaller increase of the brG reporter into the cytoplasm is consistent with a requirement for Mlp1p, if not also Mlp2p, in exporting lariat intermediates, relative to mRNAs. These results indicate that Mlp1p and Mlp2p do not retain incompletely processed lariat intermediates in the nucleus but instead promote the export of such species; indeed, together these data support a model in which Mlp1p and Mlp2p promote the export of lariat intermediates by facilitating docking at the nuclear basket and Tom1p-mediated ubiquitylation of Yra1p promotes undocking, rather than docking, and thereby nuclear pore entry.

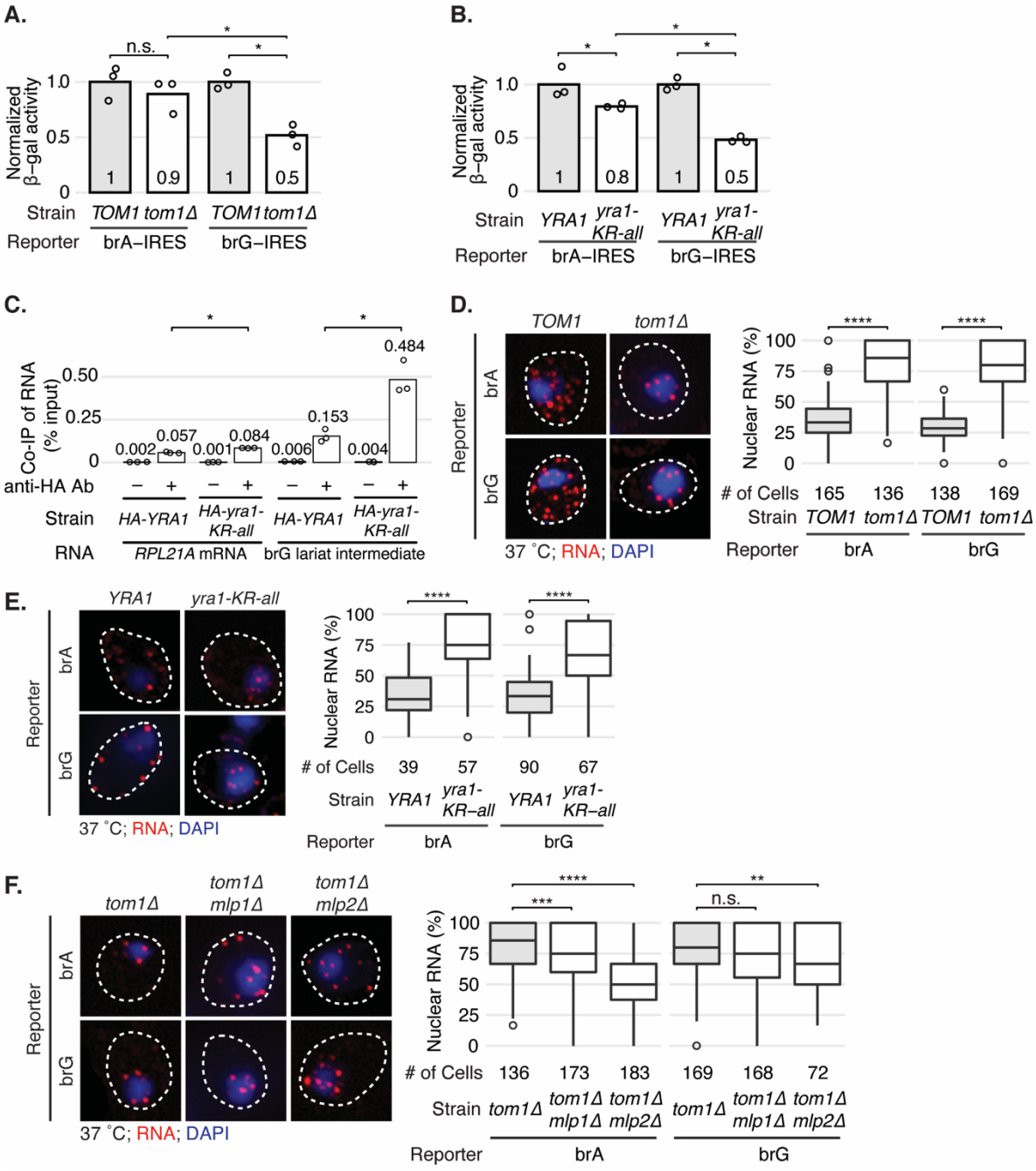


Figure 3.9. The export of lariat intermediate requires Tom1p-mediated ubiquitination of Yra1p.

A, B. The efficient export of lariat intermediates requires *TOM1* (**A**) and sites of *TOM1*-dependent ubiquitylation in *YRA1* (**B**). *TOM1* and *tom1*Δ cells (**A**) were grown at a permissive temperature of 30 ˚C, and *YRA1* and *yra1-KR-all* cells (**B**) were grown at semi-permissive temperature of 25 °C. The cytoplasmic localization of the reporters was assayed by measuring β-galactosidase activity of cell extracts. The activity of three technical replicates was quantitated and illustrated as in Fig. 3.3B; values for each reporter were normalized to the mean activity of the same reporter in the *TOM1* strain (**A**) or *YRA1* stain (**B**). **C.** The release of Yra1p from lariat intermediate, as for mRNA, requires sites of *TOM1*-dependent ubiquitylation. Crosslinking and RNA co-IP were

Figure 3.9 (continued) executed as in Fig. 3.2C, except that both *YRA1-HA* and *yra1-KR-all-HA* cells were shifted to 37 ˚C for 2 hours before crosslinking, the levels of IP'd RNA were quantitated as in Fig. 3.2C. **D, E.** Export of lariat intermediate, as well as mRNA, requires Tom1p (**D**) and ubiquitylation of Yra1p (**E**), and compromising these activities accumulates lariat intermediate and mRNA in the nucleus. Cells were shifted to the non-permissive temperature of 37 ˚C for 2 hours and then assayed for reporter localization by smRNA-FISH (left); the nuclear fraction of reporter RNA was quantitated (right). **F**. The nuclear basket factors *MLP1* and *MLP2* are required for the nuclear peripheral localization of lariat intermediate, in addition to mRNA, indicating a role for the nuclear basket in promoting the export of lariat intermediate by recruiting the lariat intermediate to the pore. Cells were shifted to the non-permissive temperature of 37 ˚C for 2 hours and then assayed for reporter localization by smRNA-FISH (left); the nuclear fraction of reporter RNA was quantitated (right). Throughout, for smRNA-FISH cells were probed and demarked as in Fig. 3.1C; the nuclear fraction was calculated and displayed as in Fig. 3.1C.

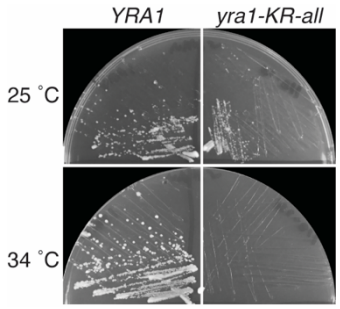


Figure 3.10. The yra1-KR-all mutant displayed a mild growth defect at 25 ˚C.
YRA1 and *yra1-KR-all* cells were streaked onto YPDA, solid media and incubated for 2 days at 25 ˚C or 34 ˚C.

Discussion

In this work, we undertook a combination of ensemble and single molecule approaches to demonstrate that the spliceosome-discarded lariat intermediates use the same nuclear export pathway as mature mRNAs do, requiring the general export receptor Mex67p (**Fig. 3.1**) and three of its adaptors, Yra1p, Nab2p, and Npl3p (**Fig. 3.3**), establishing both mRNAs and lariat intermediates utilize the same export machinery. Unexpectedly, we found that the purported quality control factor Mlp1p did not retain lariat intermediates in the nucleus but instead promoted the export of lariat intermediates through its interaction with the export adaptor Nab2p (**Fig. 3.4; Fig. 3.7**). Further, the efficient export of lariat intermediates also relies on the E3 ubiquitin ligase Tom1p and its target sites in Yra1p (**Fig. 3.9**). Importantly, our findings imply that Tom1-mediated ubiquitylation of Yra1p undocks mRNPs from the nuclear basket of the NPC, allowing not only spliceosome-discarded mRNPs but also the mRNPs to transit through the NPC.

In this work, we found that the spliceosome-discarded lariat intermediates use the same nuclear export pathway as mature mRNAs do. Since lariat intermediates do not have a canonical 5' cap and thus lack the cap binding complex (CBC), our data indicate that the 5' cap and CBC are not necessary for nuclear mRNA export, but instead facilitate nuclear mRNA export. As these intermediates are degraded by the cytoplasmic exosome, our data further establish the mRNA export pathway as the key step for turning over RNA substrates discarded by the splicing proofreading mechanism. Nuclear export ensures that these discarded substrates do not get inspected by the spliceosome again. In fact, as both splicing and the assembly of mRNA export machinery are co-transcriptional, it is likely that these intermediates are already export competent when they are rejected by the spliceosome and thereby nuclear export by default. Indeed, a range of suboptimal pre-mRNAs and splicing intermediates are exported in the cytoplasm for

degradation (Harigaya and Parker, 2012; Hilleren and Parker, 2003; Legrain and Rosbash, 1989; Mayas et al., 2010; Sayani and Chanfreau, 2012; Talhouarne and Gall, 2018), implying that pre-mRNA becomes export competent during the early steps of the spliceosome assembly. Indeed, Sub2p, which recruits Yra1p, is essential for pre-spliceosome assembly (Cordin and Beggs, 2013; Tutucci and Stutz, 2011), further supporting the view that pre-mRNAs become export competent during the early steps of the splicing cycle. Nevertheless, future work is needed to determine at which step of the splicing pathway a pre-mRNA transcript becomes export competent.

Unexpectedly, we found that the purported quality control factor Mlp1p did not retain lariat intermediates in the nucleus but instead promoted the export of lariat intermediates through its interaction with the export adaptor Nab2p (**Fig. 3.4; Fig. 3.7**). Our results raise questions concerning the actual role of Mlp1p in retaining immature mRNPs in the nucleus. As the original study, which identified Mlp1p as a quality control factor, focused on only one artificial substrate with poor splicing efficiency (Galy et al., 2004), it is likely that its role in quality control is substrate specific. Consistently, a genome-wide study on endogenous genes did not find evidence that Mlp1 acts as a general retention factor (Sayani and Chanfreau, 2012). Future work is needed to determine what are the endogenous substrates of Mlp1p.

While we did not find evidence that Mlp1p is required for mRNA export, we reason that, in *mlp1*Δ cells, other functionally redundant factors compensate for the loss of Mlp1p to ensure mRNA export. In fact, multiple known interactions exist between the mRNP and the NPC; therefore, loss of one interaction, such as deletion of *MLP1*, would not grossly affect mRNA export in budding yeast. By contrast, since the lariat intermediate is already an export deficient substrate since it lacks a 5' cap and a portion of RNA, its export dependents on Mlp1p. Therefore, our observation that Mlp1p promotes the export of lariat intermediates is consistent with the positive

role of Mlp1p and its orthologs in mRNA export observed in a number of organisms (Aksenova et al., 2019; Bae et al., 2009; Lee et al., 2019; Marfatia et al., 2003; Xu et al., 2007). To test this functional redundancy mechanism, future work is needed to test if by artificially tethering cap binding proteins to the lariat intermediate would rescue its export defect in *mlp1*Δ cells.

In addition to the lariat intermediate, the cleaved 5' exon, the other first step intermediate, is also exported into the cytoplasm (Mayas et al., 2010 ;data not shown). However, it is unclear how it is exported. Our results imply that the 5' exon is likely exported through the same export pathway as mRNA and the lariat intermediate do. If so, since both the 5' exon and the lariat intermediate are discarded at the same time, it would be interesting to test if they are exported together or separately. Since both the 5' exon and the lariat intermediate do not contain the complete set of export factors, they can be used proxies to examine the functional redundancy mechanism during mRNA export.

Further, lariat intermediate export required Tom1p-mediated ubiquitylation of Yra1p (**Fig. 3.9**), which is proposed to take place at the nuclear basket. Importantly, deletion of *MLP2* rescued the export defect of both mRNAs and lariat intermediates in *tom1*Δ cells (**Fig. 3.9**), indicating that Yra1p-containing mRNPs are stalled on the nuclear basket. Collectively, our data suggest a model, in which the export of lariat intermediates first requires docking onto the nuclear basket through the interactions between Nab2p and Mlp1p and between Yra1 and Mlp2p. Tom1-mediated ubiquitylation of Yra1 undocks mRNPs, allowing mRNPs to transit through the NPC, a pathway that likely also operates in the case of mRNA export but in a manner that is normally masked by redundant pathways.

Materials and Methods

Yeast strain and plasmid construction

Yeast strains and plasmids used in this study are listed in table 3.1 and 3.2. Yeast strains were constructed with exogenous copy present on a shuffle plasmid. Plasmids were made either using topo cloning and/or site-directed mutagenesis. *pRS313-MLP1* plasmid was generated by cloning *MLP1* locus (ORF with 500 bp flanking sequences) from the *BY4741* strain into *pRS313* empty vector. *pRS313-mlp1-Δ1586-1768* plasmid was generated by site directed mutagenesis of *pRS313-MLP1*. The NLS in *pRS313-mlp1-Δ1586-1768* is intact. GFP tagged *MLP1* plasmids (*pRS313-MLP1-GFP* and *pRS313-mlp1-Δ1586-1768-GFP*) were generated by inserting the enhanced monomeric GFP sequence immediately downstream of the penultimate amino acid in frame. *pRS315-nab2-F72AF73A* plasmid was generated by site-directed mutagenesis of *pRS315-NAB2* plasmid. *pRS316-MEX67* plasmid was generated by cloning *MEX67* locus (ORF with 500 bp flanking sequences) from the *BY4741* strain into *pRS316* empty vector. *pRS313-MEX67* was generated by subcloning *MEX67* from *pRS316-MEX67* to *pRS313*. *mex67-5* mutant was generated by site directed mutagenesis of *pRS313-MEX67*. Reporters used in FISH experiments were generated as followed. The *lacZ* reporters containing different mutations were subcloned from *pRS426* to *pRS316*. Then, the strong *GPD* promoter was replaced by a weak *STE5* promoter. *MEX67* strains were generated as followed. *MEX67* shuffle strain was generated by replacing the endogenous *MEX67* with *HPHMX4* in the *BY4741* strain expressing *pRS316-MEX67*. *MEX67* shuffle strain was conferred by PCR and 5-FOA selection. Afterward, *pRS316-MEX67* in the *MEX67* shuffle strain was replaced by either *pRS313-MEX67* or *pRS313-mex67-5*, generating final *MEX67* strain or *mex67-5* strain. Similar strategy was applied to generate *mlp1*Δ *dbr1*Δ strain from *dbr1*Δ strain, *mlp1*Δ *tom1*Δ strain from *tom1*Δ, and *mlp2*Δ *tom2*Δ strain from *tom1*Δ strain.

RNA-FISH

RNA-FISH was performed as described in Raj and Tyagi, 2010 (Raj and Tyagi, 2010). Cells were imaged using Olympus IX81 inverted widefield microscope equipped with Hamamatsu Orca Flash 4.0 camera with 4 megapixels and 100x 1.45NA oil objective lens. Single RNA molecule counting was conducted using custom macros in imageJ and statistical analysis was conducted in R. 10% formamide was used for smRNA-FISH experiments and 40% formamide was used for RNA-FISH experiments. For smRNA-FISH, probes were designed by Biosearch targeting the *lacZ* portion of the reporters. For RNA-FISH, a single Cy3-labeled oligo-dT(50) probe was used to target polyA+ RNAs and a single Alexa488-labeled probe was used to target the brG reporter.

Immunofluorescence

Cells expressing *MLP1-GFP* or *MLP1-*Δ*1586-1768-GFP* were imaged in an anti-fade buffer using the epifluorescence microscopy with 100X magnification.

RNA co-immunoprecipitation

For Mex67-GFP RNA co-IP, cell lysates were prepared from *MEX67-GFP* cells expressing the brG reporter (pJPS1488) and incubated with beads pre-conjugated with either anti-GFP antibody or IgG. After a 2-hour incubation at 4 ˚C, beads were washed five times and put through RNA extractions using Phenol:Chloroform:Isoamyl Alcohol (25:24:1, v/v). To ensure Mex67 binds lariat intermediates *in vivo*, tagged *MEX67-TAP* cells expressing the empty vector were mixed with untagged *MEX67* cells expressing the brG reporter. In parallel, tagged *MEX67-TAP* cells expressing the brG reporter were mixed with untagged *MEX67* cells expressing the empty vector. Lysates were then prepared from these mixed cells and incubated with IgG beads at 4 ˚C for 2 hours. After incubation, beads were washed five times. Bound *MEX67* containing mRNPs

were released by TEV cleavage and put through RNA extractions using Phenol:Chloroform:Isoamyl Alcohol (25:24:1, v/v). For Nab2-HTB denaturing RNA co-IP and HA-Yra1 denaturing RNA co-IP, cells were grown to OD600 of 0.6-0.8, cross-linked for 10 minutes by 3.7% formaldehyde, and quenched with glycine. Then, cell lysates were prepared using glass beads. Cell lysates were incubated with either Ni-NTA beads in the case of Nab2-HTB or anti-HA beads in the case HA-Yra1 at 4 ˚C for 2 hours, After incubation, beads were washed five times and put through RNA extractions using Phenol:Chloroform:Isoamyl Alcohol (25:24:1, v/v). Lariat intermediates were assayed using lariat specific RT-PCR. mRNA of endogenous *RPL21* was assay by RT-PCR as an internal control.

In vivo RNA analysis

For analysis of RNA in vivo, cells transformed with ACT1-IRES-lacZ splicing reporters were cultured in selective media at 30 °C to an OD600 of 0.6–0.8, lysed, and then assayed for RNA. RNA was analyzed by primer extension using 32P-radiolabeled primers and AMV-reverse transcriptase, followed by separation of products on a 6% denaturing polyacrylamide gel. Two primmer were used, one binding the *lacZ* portion of the reporters and the other binding U14 snoRNA for an internal control. Data were visualized using a phosphorimager (Molecular Dynamics) and quantitated using ImageQuant (GE healthcare).

β-galactosidase assays

Liquid assays were performed as described in Mayas et al., 2010. Specifically, 1.5 mL of liquid cultures in selective media were harvested at OD 600 of 0.6–0.8, and washed with Z buffer (60 mM Na_2HPO_4, 40 mM NaH_2PO_4, 10 mM KCl, 1 mM $MgSO_4$; pH 7.0). Cells were resuspended in 100 µL of Z buffer and lysed by 6 cycles of freeze-thawing (30 seconds each in liquid nitrogen and in a 42 °C water bath). Lysed cells were incubated with 700 µL of prewarmed (30 °C) Z buffer

that included 1 mg mL *ortho*-nitrophenyl-β−galactopyranoside and 50 mM β-mercaptoethanol. Reactions were incubated at 30 °C for 30 minutes to 2 hours and stopped by the addition of 0.5 mL 1M Na_2CO_3. After cell debris was pelleted, the OD_{420nm} of the supernatant was measured. Activity in Miller units was calculated as (OD4200nm × 1000)/(OD600nm × (minutes elapsed) × 1.5 mL).

Supplementary materials

Table 3.1 Yeast strains used in this study

Name	Genotype	Reference
BY4741	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0*	Open Biosystems
MEX67	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 mex67Δ::hphMX4 pRS313-MEX67*	This study
mex67-5	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 mex67Δ::hphMX4 pRS313-mex67-5*	This study
MEX67-TAP	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 MEX67-TAP::HIS3MX6*	Open Biosystems
HA-YRA1	*MATa ade2 his3 leu2 trp1 ura3 yra1Δ::HIS3 pTRP1-HA-YRA1 cDNA*	Iglesias et al., 2010
HA-yra1-KR-all	*MATa ade2 his3 leu2 trp1 ura3 yra1Δ::HIS3 pTRP1-HA-YRA1 KR1-8*	Iglesias et al., 2010
HA-yra1-8-GFP	*MATa ade2 his3 leu2 trp1 ura3 yra1Δ::HIS3 pTRP1-HA-GFP-yra1-8*	Zenklusen et al., 2002
NAB2	*MATa leu2Δ ura3Δ his3Δ nab2Δ::HIS3 pRS315-NAB2*	Grant et al., 2008
nab2-ΔN	*MATa leu2Δ ura3Δ his3Δ nab2Δ::HIS3 pRS315NAB2-ΔN*	Grant et al., 2008
nab2-F72A/F73A	*MATa leu2Δ ura3Δ his3Δ nab2Δ::HIS3 pRS315-NAB2-F72AF73A*	This study
NAB2-HTB	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 NAB2-HTB::hphMX4*	This study
NPL3	*MATa ura3-52 his3-11,15 leu2-3,112 trp1-1 (W303)*	Kress et al., 2008
npl3-1	*MATa ura3-1 leu2-3,112 trp1-1 ade2-1 lys2 his3 his 4 (W303 background)*	Kress et al., 2008
mlp1Δ	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 mlp1Δ::kanMX4*	Open Biosystems
MLP1-GFP	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 mlp1Δ::kanMX4 pRS313-MLP1-GFP*	This study
mlp1-Δ1586-1768-GFP	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 mlp1Δ::kanMX4 pRS313-mlp1-Δ1586-1768-GFP*	This study
dbr1Δ	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 dbr1Δ::kanMX4*	Open Biosystems
mlp1Δ dbr1Δ	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 dbr1Δ::kanMX4 mlp1Δ::hphMX4*	This study
tom1Δ	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 tom1Δ::kanMX4*	Open Biosystems
tom1Δ mlp2Δ	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 tom1Δ::kanMX4 mlp1Δ::hphMX4*	This study
tom1Δ mlp1Δ	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 tom1Δ::kanMX4 mlp2Δ::hphMX4*	This study
los1Δ	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 los1Δ::kanMX4*	Open Biosystems

Table 3.2 Plasmids used in this study

Plasmid ID	Name	Reference
pYZ1	*pRS313-MLP1*	This study
pYZ4	*pRS313-mlp1-Δ1586-1768*	This study
pYZ9	*pRS313-MLP1-GFP*	This study
pYZ10	*pRS313-mlp1-Δ1586-1768-GFP*	This study
pAC717	*pRS315-NAB2 (pAC717)*	Green et al., 2002
pYZ25	*pRS315-nab2-F72AF73A (derived from pAC717)*	This study
pYZ17	*pRS315-MEX67*	This study
pYZ18	*pRS315-mex67-5*	This study
pYZ19	*pRS316-MEX67*	This study
	pUN100–LEU2–MEX67–GFP	Segref et al., 1997
pJPS1481	*pRS426-GPDpromoter-ACT1(WT)-IRES-lacZ*	Mayas et al., 2010
pJPS1483	*pRS426-GPDpromoter-ACT1(G303c)-IRES-lacZ*	Mayas et al., 2010
pJPS1485	*pRS426-GPDpromoter-ACT1(G1a)-IRES-lacZ*	Mayas et al., 2010
pJPS1486	*pRS426-GPDpromoter-ACT1(G1c)-IRES-lacZ*	Mayas et al., 2010
pJPS1488	*pRS426-GPDpromoter-ACT1(brG)-IRES-lacZ*	Mayas et al., 2010
pJPS1514	*pRS426-GPDpromoter-ACT1(brC)-IRES-lacZ*	Mayas et al., 2010
	pRS316-STE5promoter-ACT1(WT)-IRES-lacZ	This study
	pRS316-STE5promoter-ACT1(brG)-IRES-lacZ	This study
	pRS316-STE5promoter-ACT1(G303c)-IRES-lacZ	This study
	pRS316-STE5promoter-ACT1(WT)-lacZ	This study
	pRS316-STE5promoter-ACT1(brG)-lacZ	This study

CHAPTER 4

CONCLUSIONS AND PERSPECTIVES

In this work, we investigated the dynamics and regulations of co-transcriptional splicing and characterized the nuclear export pathway of lariat intermediates with implications in general mRNA export. To study co-transcriptional splicing in humans, we developed a genome-wide approach, CoLa-seq, which allowed us to efficiently map BPs in the human genome, uncover a new splicing pathway for adjacent introns, concurrent splicing, observe co-transcriptional lariat formation events with diverse kinetics, and, surprisingly, obtain evidence for the prevalent usage of the intron definition pathway in humans. Using computational modeling, we identified key *cis*-regulatory sequence elements and *trans*-acting factors, such as intron size, regional GC content, and U2AF binding, that correlate with the timing of co-transcriptional splicing. To study the nuclear export of lariat intermediates in yeast, we undertook a combination of ensemble and single molecule approaches to demonstrate that the spliceosome-discarded lariat intermediates use the same nuclear export pathway as mature mRNAs do, requiring the general export receptor Mex67p and three of its adaptors, Yra1p, Nab2p, and Npl3p. We further found that the purported quality control factor Mlp1p did not retain lariat intermediates in the nucleus but instead promoted the export of lariat intermediates through its interaction with the export adaptor Nab2p. Lastly, our findings indicate that Tom1-mediated ubiquitylation of Yra1p undocks mRNPs from the nuclear basket of the NPC, allowing mRNPs to transit through the NPC. Together, our work deepened our understandings of two key steps during gene expression: splicing and mRNA export.

CoLa-seq enables efficient BP annotation in humans

Although the BP is one of three essential reactants of splicing reactions, its annotation is far from complete. Mutations altering BP usage are associated with many human diseases (Darman et al., 2015). By enriching for lariat RNA species, we identified more BPs from a single cell type than the most recent large-scale study (**Fig. 2.2F**), which used 1000 times as many reads pooled

across thousands of RNA-seq libraries from diverse tissues and cell-types (Pineda and Bradley, 2018), establishing CoLa-seq as an efficient approach for global BP annotation in humans. Notably, 66% of BPs identified in our work are novel, therefore, our expanded BP annotations will be a valuable resource for future work on BP usage. Additionally, through excised lariat introns, CoLa-seq enables direct examination of BP-3' SS coupling for the first time on a genome-wide scale. We obtained direct evidence that the BP selection directly impacts 3' SS selection. Previous studies also found that suboptimal BPs compromise exon ligation efficiency (Query et al., 1996). In fact, biochemical studies from the Staley lab have found that Prp16 not only proofreads the BP but also promotes alternative BP selection (Semlow et al., 2016). Therefore, in the future, it will be interesting to determine if Prp16 contributes to BP choice and thus 3' SS selection in a regulated manner.

Moreover, CoLa-seq enables profiling BP usage as well as BP-3' SS coupling under different conditions. One attractive application is studying BP usage in cancer cells containing SF3B1 mutations. SF3B1 is one of the most frequently mutated genes in various cancer types including myelodysplastic syndromes, chronic lymphocytic leukemia, uveal melanoma, and pancreatic cancer; however, the molecular and functional consequences of SF3B1 mutations are not fully understood (Teng et al., 2017). Targeted, anecdotal gene analyses have suggested that cancer-associated SF3B1 mutations impact the selection of BPs, resulting in changes in the 3' SS usage (Darman et al., 2015), but a systematic analysis has not been performed due to the lack of a practical genome-wide strategy to map BPs. Therefore, CoLa-seq provides an effective strategy to investigate the impact of SF3B1 mutations on the usage of BPs during splicing in patient samples and cancer cell lines carrying cancer-associated SF3B1 mutations. Further, CoLa-seq is suitable for studying how candidate drugs targeting SF3B1 impact BP usage in these cancer cells.

Adjacent introns undergo not only in-order and out-of-order splicing but concurrent splicing

As splicing and transcription are coupled, it was initially thought that the order in which introns are removed follows the order of transcription (Aebi and Weissman, 1987). However, accumulating evidence indicates that introns within a given pre-mRNA do not always complete splicing according to the order of transcription (Kessler et al., 1993; Kim et al., 2017; Drexler et al., 2020). Our results indicate that adjacent introns can undergo not only in-order and out-of-order splicing but also concurrent splicing, providing a potential way for coordinating splicing of adjacent introns and thus splicing regulation. Notably, splicing of specific intron pairs follows a preferred order. Using computational modeling, we found that specific sequence features, gene architecture, and genomic contexts contribute to the order of splicing (**Fig. 2.7**), implying that the order of splicing might be inherently encoded in the human genome to help coordinate splicing across multiple introns within the same gene.

As one of the top features, GC content is strongly associated with the potential of RNA secondary structure formation, implying that RNA structures can influence the order of splicing. In fact, GC rich exons are sensitive to the levels of U1 snRNP components and helicases, DDX5 and DDX17, implying that RNA secondary structures are limiting splicing for these exons (Lemaire et al., 2019). To test this hypothesis, *in vivo* RNA structural probing can be used (Flynn et al., 2016; Zubradt et al., 2017). Since RNA folds co-transcriptionally and RNA can adopt different structures depending on the surrounding sequences (Herzel et al., 2017), it would be important to determine nascent RNA structures as opposed to steady-state RNA structures (Liu et al., 2019; Sun et al., 2019). The nucleotide composition also manifests on the gene level; therefore, it would be interesting to examine how larger genomic contexts, such as nucleosome positioning, chromatin modifications, and even higher-order chromosome structures, influence the order of

splicing. For example, regions with high GC content would lead to high nucleosome occupancy, which in turn might influence the order of splicing.

Although concurrent splicing is prevalent across introns, it is unknown if concurrent splicing rises from two independent splicing reactions or two coordinated splicing reactions. Nevertheless, concurrent splicing might enable adjacent introns to coordinate their splicing. Consistent with this idea, results from a recent long-read sequencing study suggested splicing coordination across multiple introns within a given pre-mRNA (Drexler et al., 2020). However, it is unknown whether such coordination is achieved via concurrent splicing. Future experiments are required to explicitly test the coordination in concurrent splicing. The key prediction from coordinated splicing is that disrupting splicing of one intron would block splicing of the other. To test this prediction, one could block splicing of the downstream intron using an anti-sense oligo targeting the downstream BP and examine whether splicing of the upstream intron is affected or not. If splicing of the upstream intron is blocked, it would indicate that concurrent splicing is coordinated. Otherwise, concurrent splicing is not coordinated.

Intriguingly, specific patterns of the order of splicing are also associated with alternatively spliced introns. For instance, compared to introns flanking constitutive introns, introns flanking included cassette exons are associated with a lower level of in-order splicing but a higher level of concurrent splicing (**Fig. 2.5**). Similarly, alternative 3' SSs are associated with a similar level of in-order splicing, a lower level of concurrent splicing, but a higher level of out-order splicing (**Fig. 2.5**). However, it is unclear whether these specific patterns contribute to the outcomes of AS. It is plausible that certain orders of splicing influences local RNA structures or leads to the juxtaposition of regulatory elements, which in turn affects the outcome of AS. In the future, it will

be interesting to test if the order of splicing impacts AS patterns using CRISPR-based assays as well as mini-gene based reporters.

Timing of co-transcriptional splicing varies across introns

With CoLa-seq, we found that splicing timing can vary significantly across introns and even within the same intron. Our findings contrast with some aspects of previous studies and complement others (Carrillo Oesterreich et al., 2016; Drexler et al., 2020; Herzel et al., 2018; Pai et al., 2017; Reimer et al., 2020; Wachutka et al., 2019). Using metabolic labeling followed by either short sequencing or direct RNA long-read sequencing, two recent studies found that splicing occurs in the range of minutes, indicating that splicing occurs when RNAP II is not in close proximity to the 3' SSs (Wachutka et al., 2019; Drexler et al., 2020). However, other studies, done in yeasts and murine cells, concluded that splicing occurs when RNAP II is still in close proximity to the 3' SSs (Carrillo Oesterreich et al., 2016; Herzel et al., 2018; Reimer et al., 2020). Notably, our results support both early and late splicing estimates in humans. It is likely that experimental designs in these studies led to the preferential capture of one class of splicing events over the other. For instance, in the case of metabolic studies using 4sU RNA labeling, given that U content is much higher in un-spliced transcripts than in spliced transcripts, un-spliced transcripts were preferentially enriched over spliced transcripts during the sample preparation, leading to an underestimate of splicing efficiency. Further, it is unclear whether 4sU RNA labeling affects spliceosome assembly and catalysis between U-rich RNA elements in introns and snRNAs, as 4sU increases the stability of G-U base pairing by 10-fold. (Testa et al., 1999). In the case of TT-seq (Wachutka et al., 2019), the authors also noted that their data missed fast splicing events likely because RNA fragments were hundreds of nucleotides long. In the case of Nano-cop (Drexler et al., 2020), because of the limited read depth of the long-read sequencing, the authors had to

calculate the splicing efficiency by pooling reads across different introns. Given the complex nature of the human transcriptome, it is unclear if such an averaged estimate could accurately represent the dynamics of splicing timing. For studies concluding that splicing is fast, they relied on chromatin-associated RNAs. While nascent lariat intermediates captured by CoLa-seq are present at a very low level and associated with chromatin because of co-transcriptional splicing, linear RNA species, including mature mRNAs and degradation intermediates, are highly abundant and thus can be associated with chromatin independent of transcription. Thus, it is challenging to obtain purified chromatin-associated RNAs that are free of mature mRNAs and degradation intermediates, leading to a likely overestimation of splicing efficiency. Considering the complex nature of human transcription and extensive regulation of AS, it is reasonable to assume that optimal splicing and gene regulation requires dynamic co-transcriptional splicing.

Notably, we found that introns with early lariat formation have strong PPT and U2AF binding (**Fig. 2.12**). Since *in vitro* studies suggested that RNAP II recruits U2AF2 and Prp19 complex to pre-mRNAs through its phosphorylated CTD, thereby promoting splicing activation (David et al., 2011), we propose the following model: for introns with strong PPT, RNAP II-associated U2AF2 and Prp19 complex become associated soon after the 3' SSs are transcribed, leading to rapid spliceosome assembly, activation, and lariat formation; whereas, for introns with weak PPT, RNAP II-associated U2AF2 and Prp19 complex may require help from additional factors, such as SR proteins (Wu and Maniatis, 1993), to become associated, thereby leading to late lariat formation. Our model also explains why AG/U2AF1-independent introns can undergo ultra-fast lariat formation (**Fig. 2.14**). Since AG/U2AF1-independent introns have strong PPT, RNAP-II U2AF2 and the Prp19 complex become associated with introns soon after the PPT emerges from RNAP II, thereby leading to ultra-fast lariat formation. Compared to features at the

3’ ends of the intron, features related to the 5’ SS have minimal effect on the timing of lariat formation. As U1 snRNP-associated protein Prp40 interacts with RNAP II CTD in yeast (Morris and Greenleaf, 2000), it is likely that RNAPII-associated U1 snRNP begins 5’ SS recognition when the 5’ SS is transcribed and completes 5’ SS recognition by the time RNAP II transcribes the 3’ SS. Consistently, we found that longer introns promote early lariat formation (**Fig. 2.12**), as it takes longer time for RNAP II to transcribe longer introns, allowing more time for U1 snRNP to complete 5’ SS recognition.

Since the regulation of AS is often achieved by promoting or repressing the binding of the U1 snRNP, U2AF, or the U2 snRNP to the splice sites (Chen and Manley, 2009), our measurement of lariat formation timing might provide a more accurate and sensitive assessment for *in vivo* splicing regulation than the final splicing timing does, as the final splicing timing comprises timing of lariat formation and timing of exon ligation. As we only looked at a single condition in our current work, we essentially examined the differences and similarities between different introns. Therefore, in the future, it will be interesting to apply CoLa-seq to different cell types and cellular conditions to study the functions of different regulatory factors and examine the impact of genetic variations on co-transcriptional splicing.

Further development of CoLa-seq to study co-transcriptional splicing

The great depth afforded by short-read sequencing allowed us to capture many nascent lariat intermediates across many introns (**Fig. 2.1**). However, as most human introns are several kbs long, the size limitation imposed by the short-read sequencing has prevented us to directly capture late splicing events, in which RNAP IIs are located more than 1 kb downstream of the 3’ SS. Although, fortuitously, concurrent splicing reads captured by CoLa-seq allowed us to independently infer the timing of late lariat formation, these reads do not report the precise

positions of RNAP IIs. Based on our data, the timing of lariat formation varies dramatically across introns, therefore, in order to obtain the full picture of co-transcriptional splicing dynamics, it is critical for us to capture both early and late lariat formation events. To overcome such size limitation imposed by the short-read sequencing, the important next step for CoLa-seq is to integrate the long-read sequencing into the CoLa-seq pipeline, which will allow us to capture positions of RNAP IIs associated with late lariat formation. It is worth noting that the limited depth of the long-read sequencing will likely restrict our analyses to fewer introns. Although it is not ideal, we should be able to capture representative pictures of co-transcriptional splicing by performing deep and targeted analysis of selected genes. Recent developments in the long-read sequencing have also shown promising signs for obtaining datasets with deep coverage (Amarasinghe et al., 2020; Logsdon et al., 2020).

To capture the full dynamics of co-transcriptional splicing, we need to capture not only nascent lariat intermediates but also nascent un-spliced and spliced transcripts so that we can calculate splicing efficiency at single nucleotide position. Although recent studies have applied the long-read sequencing to examine the splicing timing using un-spliced and spliced transcripts (Drexler et al., 2020; Reimer et al., 2020), their approaches are likely confounded by technical biases (see above). To overcome these biases, one approach is to directly assay RNAP II-associated nascent RNA transcripts. This approach has been successfully used in recent mNET-seq studies (Nojima et al., 2015). Importantly, recent work done in the Staley lab has identified working conditions to integrate RNAP II IP to the CoLa-seq pipeline (Y. Hou and J. Staley, unpublished data).

While CoLa-seq greatly enriched for lariat RNA species, reads derived from lariat RNA species are still a small portion in our current dataset due to contaminations of highly abundant

small RNAs, such as snRNAs and snoRNAs, and degradation intermediates from the linear RNA degradation step. To further improve the enrichment of lariat RNA species, the following strategies should be considered. The first strategy involves using a more stringent condition, which should reduce the level of small RNAs that are not directly associated with the transcription elongation complex. Studies from Proudfoot lab have identified detergent empigen as a potential candidate since it can disrupt the interactions between the spliceosome and RNAP II (Nojima et al., 2018; Schlackow et al., 2017). The second strategy involves a more efficient method to degrade linear RNAs. In the current design, we use RppH and XRN-1 to degrade linear RNA species. Studies on the processivity of XRN-1 have found that certain RNA structures can restrict the processivity of XRN-1 (Chapman et al., 2014). Therefore, one could use a thermo-stable XRN-1 so that the reaction can be performed at a higher temperature with less RNA secondary structures. Alternatively, one could use cap-specific antibodies to deplete capped transcripts. However, it is unknown how much of chromatin-associated RNA transcripts are capped. Recent work in the Staley lab showed that RNAP II IP in the presence of empigen shows the greatest promise (Y. Hou and J. Staley, unpublished data).

Overall, CoLa-seq developed in this work has the potential to greatly facilitate future studies on co-transcriptional splicing and its regulation. For instance, CoLa-seq can be used to examine how modulating the transcription rate affects the timing of co-transcriptional splicing. CoLa-seq can also be used to study how different splicing factors influence the outcomes of AS in the context of transcription. With further development, CoLa-seq will be able to capture the full picture of co-transcriptional splicing dynamics.

Nuclear export of spliceosome-discarded lariat intermediates sheds light on mRNA export

Using smRNA-FISH and *lacZ*-based export assays, we found spliceosome-discarded pre-mRNAs and splicing intermediates are exported into the cytoplasm for degradation. However, studies also found that many defective transcripts can be degraded in the nucleus. For instance, human nuclear exosome targeting (NEXT) complex targets early and unprocessed transcripts (i.e. promoter upstream transcripts) for nuclear degradation, and ZFC3H1 targets extensively polyadenylated transcripts for nuclear degradation (Lubas et al., 2011; Meola et al., 2016). Therefore, it is not entirely clear why some defective RNAs are exported for degradation. As the export machinery is assembled during transcription, it is plausible that the nuclear export of defective transcripts merely indicates that these defective transcripts are already export competent when they are discarded from either transcription or other RNA processing steps. Consistently, recent studies found that blockage of mRNA export would cause nuclear degradation of even properly processed mature mRNAs, suggesting that there is a competition between nuclear degradation and mRNA export (Silla et al., 2018; Tudek et al., 2019).

Our results indicate that Tom1p-mediated ubiquitylation of Yra1p occurs on the nuclear basket and Yra1 ubiquitylation undocks mRNPs, allowing mRNPs to transit through the NPC. To further test this model, *in vitro* ubiquitylation system (Sung et al., 2016) can be used to test if Tom1-mediated ubiquitylation can release mRNPs from recombinant Mlp1p. Further, reconstituted nuclear pores (Hülsmann et al., 2012) can be used to test if Yra1 ubiquitylation is sufficient to allow the mRNP to transit through the NPC.

Together, our findings provide a framework for future work on studying the nuclear mRNP docking/undocking at the nuclear basket of the NPC.

REFERENCES

Abovich, N., Rosbash, M., 1997. Cross-Intron Bridging Interactions in the Yeast Commitment Complex Are Conserved in Mammals. Cell 89, 403–412. https://doi.org/10.1016/S0092-8674(00)80221-4

Adams, R.L., Mason, A.C., Glass, L., Aditi, null, Wente, S.R., 2017. Nup42 and IP6 coordinate Gle1 stimulation of Dbp5/DDX19B for mRNA export in yeast and human cells. Traffic Cph. Den. 18, 776–790. https://doi.org/10.1111/tra.12526

Aebi, M., Weissman, C., 1987. Precision and orderliness in splicing. Trends Genet. 3, 102–107. https://doi.org/10.1016/0168-9525(87)90193-4

Agrawal, A.A., Yu, L., Smith, P.G., Buonamici, S., 2018. Targeting splicing abnormalities in cancer. Curr. Opin. Genet. Dev., Cancer genomics 48, 67–74. https://doi.org/10.1016/j.gde.2017.10.010

Aksenova, V., Lee, H.N., Smith, A., Chen, S., Bhat, P., Iben, J., Echeverria, C., Fontoura, B., Arnaoutov, A., Dasso, M., 2019. Distinct Basket Nucleoporins roles in Nuclear Pore Function and Gene Expression: Tpr is an integral component of the TREX-2 mRNA export pathway. bioRxiv 685263. https://doi.org/10.1101/685263

Alcázar-Román, A.R., Tran, E.J., Guo, S., Wente, S.R., 2006. Inositol hexakisphosphate and Gle1 activate the DEAD-box protein Dbp5 for nuclear mRNA export. Nat. Cell Biol. 8, 711–716. https://doi.org/10.1038/ncb1427

Amarasinghe, S.L., Su, S., Dong, X., Zappia, L., Ritchie, M.E., Gouil, Q., 2020. Opportunities and challenges in long-read sequencing data analysis. Genome Biol. 21, 30. https://doi.org/10.1186/s13059-020-1935-5

Ameur, A., Zaghlool, A., Halvardson, J., Wetterbom, A., Gyllensten, U., Cavelier, L., Feuk, L., 2011. Total RNA sequencing reveals nascent transcription and widespread co-transcriptional splicing in the human brain. Nat. Struct. Mol. Biol. 18, 1435–1440. https://doi.org/10.1038/nsmb.2143

Amit, M., Donyo, M., Hollander, D., Goren, A., Kim, E., Gelfman, S., Lev-Maor, G., Burstein, D., Schwartz, S., Postolsky, B., Pupko, T., Ast, G., 2012. Differential GC Content between Exons and Introns Establishes Distinct Strategies of Splice-Site Recognition. Cell Rep. 1, 543–556. https://doi.org/10.1016/j.celrep.2012.03.013

Ashkenazy-Titelman, A., Shav-Tal, Y., Kehlenbach, R.H., 2020. Into the basket and beyond: the journey of mRNA through the nuclear pore complex. Biochem. J. 477, 23–44. https://doi.org/10.1042/BCJ20190132

Aslanzadeh, V., Huang, Y., Sanguinetti, G., Beggs, J.D., 2018. Transcription rate strongly affects splicing fidelity and cotranscriptionality in budding yeast. Genome Res. 28, 203–213. https://doi.org/10.1101/gr.225615.117

Auguie, B., 2017. gridExtra: Miscellaneous Functions for “Grid” Graphics.

Bae, J.-A., Moon, D., Yoon, J.H., 2009. Nup211, the fission yeast homolog of Mlp1/Tpr, is involved in mRNA export. J. Microbiol. Seoul Korea 47, 337–343. https://doi.org/10.1007/s12275-009-0125-7

Baejen, C., Torkler, P., Gressel, S., Essig, K., Soding, J., Cramer, P., 2014. Transcriptome Maps of mRNP Biogenesis Factors Define Pre-mRNA Recognition. Mol. Cell 55, 745–757.

Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L., Ren, J., Li, W.W., Noble, W.S., 2009. MEME Suite: tools for motif discovery and searching. Nucleic Acids Res. 37, W202–W208. https://doi.org/10.1093/nar/gkp335

Bangs, P., Burke, B., Powers, C., Craig, R., Purohit, A., Doxsey, S., 1998. Functional analysis of Tpr: identification of nuclear pore complex association and nuclear localization domains and a role in mRNA export. J. Cell Biol. 143, 1801–1812.

Barash, Y., Calarco, J.A., Gao, W., Pan, Q., Wang, X., Shai, O., Blencowe, B.J., Frey, B.J., 2010. Deciphering the splicing code. Nature 465, 53–59. https://doi.org/10.1038/nature09000

Barbosa-Morais, N.L., Irimia, M., Pan, Q., Xiong, H.Y., Gueroussov, S., Lee, L.J., Slobodeniuc, V., Kutter, C., Watt, S., Colak, R., Kim, T., Misquitta-Ali, C.M., Wilson, M.D., Kim, P.M., Odom, D.T., Frey, B.J., Blencowe, B.J., 2012. The Evolutionary Landscape of Alternative Splicing in Vertebrate Species. Science 338, 1587–1593. https://doi.org/10.1126/science.1230612

Batsché, E., Yaniv, M., Muchardt, C., 2006. The human SWI/SNF subunit Brm is a regulator of alternative splicing. Nat. Struct. Mol. Biol. 13, 22–29. https://doi.org/10.1038/nsmb1030

Baurén, G., Wieslander, L., 1994. Splicing of Balbiani ring 1 gene pre-mRNA occurs simultaneously with transcription. Cell 76, 183–192. https://doi.org/10.1016/0092-8674(94)90182-1

Bentley, D.L., 2014. Coupling mRNA processing with transcription in time and space. Nat. Rev. Genet. https://doi.org/10.1038/nrg3662

Ben-Yishay, R., Shav-Tal, Y., 2019. The dynamic lifecycle of mRNA in the nucleus. Curr. Opin. Cell Biol., Cell Nucleus 58, 69–75. https://doi.org/10.1016/j.ceb.2019.02.007

Berget, S.M., 1995. Exon Recognition in Vertebrate Splicing. J. Biol. Chem. 270, 2411–2414. https://doi.org/10.1074/jbc.270.6.2411

Berglund, J.A., Fleming, M.L., Rosbash, M., 1998. The KH domain of the branchpoint sequence binding protein determines specificity for the pre-mRNA branchpoint sequence. RNA N. Y. N 4, 998–1006.

Bernecky, C., Plitzko, J.M., Cramer, P., 2017. Structure of a transcribing RNA polymerase II–DSIF complex reveals a multidentate DNA–RNA clamp. Nat. Struct. Mol. Biol. 24, 809–815. https://doi.org/10.1038/nsmb.3465

Beyer, A.L., Bouton, A.H., Miller, O.L., 1981. Correlation of hnRNP structure and nascent transcript cleavage. Cell 26, 155–165.

Beyer, A.L., Osheim, Y.N., 1988. Splice site selection, rate of splicing, and alternative splicing on nascent transcripts. Genes Dev. 2, 754–765.

Bhatt, D.M., Pandya-Jones, A., Tong, A.J., Cell, I.B., Black, D.L., 2012. Transcript dynamics of proinflammatory genes revealed by sequence analysis of subcellular RNA fractions. Cell.

Bonnet, A., Palancade, B., 2015. Intron or no intron: a matter for nuclear pore complexes. Nucleus 6, 455–461. https://doi.org/10.1080/19491034.2015.1116660

Braberg, H., Jin, H., Moehle, E.A., Chan, Y.A., Wang, S., Shales, M., Benschop, J.J., Morris, J.H., Qiu, C., Hu, F., Tang, L.K., Fraser, J.S., Holstege, F.C.P., Hieter, P., Guthrie, C., Kaplan, C.D., Krogan, N.J., 2013. From Structure to Systems: High-Resolution, Quantitative Genetic Analysis of RNA Polymerase II. Cell 1–14. https://doi.org/10.1016/j.cell.2013.07.033

Braunschweig, U., Barbosa-Morais, N.L., Pan, Q., Nachman, E.N., Alipanahi, B., Gonatopoulos-Pournatzis, T., Frey, B., Irimia, M., Blencowe, B.J., 2014. Widespread intron retention in mammals functionally tunes transcriptomes. Genome Res. 24, 1774–1786. https://doi.org/10.1101/gr.177790.114

Braunschweig, U., Gueroussov, S., Plocik, A.M., Graveley, B.R., Blencowe, B.J., 2013. Dynamic Integration of Splicing within Gene Regulatory Pathways. Cell 152, 1252–1269. https://doi.org/10.1016/j.cell.2013.02.034

Bray, N.L., Pimentel, H., Melsted, P., Pachter, L., 2016. Near-optimal probabilistic RNA-seq quantification. Nat. Biotechnol. 34, 525–527. https://doi.org/10.1038/nbt.3519

Briese, M., Haberman, N., Sibley, C.R., Faraway, R., Elser, A.S., Chakrabarti, A.M., Wang, Z., König, J., Perera, D., Wickramasinghe, V.O., Venkitaraman, A.R., Luscombe, N.M., Saieva, L., Pellizzoni, L., Smith, C.W.J., Curk, T., Ule, J., 2019. A systems view of spliceosomal assembly and branchpoints with iCLIP. Nat. Struct. Mol. Biol. 26, 930–940. https://doi.org/10.1038/s41594-019-0300-4

Burgess, S.M., Guthrie, C., 1993. A mechanism to enhance mRNA splicing fidelity: the RNA-dependent ATPase Prp16 governs usage of a discard pathway for aberrant lariat intermediates. Cell 73, 1377–1391.

Carrillo Oesterreich, F., Herzel, L., Straube, K., Hujer, K., Howard, J., Neugebauer, K.M., 2016. Splicing of Nascent RNA Coincides with Intron Exit from RNA Polymerase II. Cell 165, 372–381. https://doi.org/10.1016/j.cell.2016.02.045

Carvalho, T., Martins, S., Rino, J., Marinho, S., Carmo-Fonseca, M., 2017. Pharmacological inhibition of the spliceosome subunit SF3b triggers exon junction complex-independent nonsense-mediated decay. J. Cell Sci. 130, 1519–1531. https://doi.org/10.1242/jcs.202200

Chanarat, S., Burkert-Kautzsch, C., Meinel, D.M., Sträßer, K., 2012. Prp19C and TREX: Interacting to promote transcription elongation mRNA export. Transcription 3, 8–12. https://doi.org/10.4161/trns.3.1.19078

Chapman, E.G., Moon, S.L., Wilusz, J., Kieft, J.S., 2014. RNA structures that resist degradation by Xrn1 produce a pathogenic Dengue virus RNA. eLife 3, e01892. https://doi.org/10.7554/eLife.01892

Charenton, C., Wilkinson, M.E., Nagai, K., 2019. Mechanism of 5′ splice site transfer for human spliceosome activation. Science 364, 362–367. https://doi.org/10.1126/science.aax3289

Chávez, S., Beilharz, T., Rondón, A.G., Erdjument-Bromage, H., Tempst, P., Svejstrup, J.Q., Lithgow, T., Aguilera, A., 2000. A protein complex containing Tho2, Hpr1, Mft1 and a novel protein, Thp2, connects transcription elongation with mitotic recombination in Saccharomyces cerevisiae. EMBO J. 19, 5824–5834. https://doi.org/10.1093/emboj/19.21.5824

Chen, C.-Y., Chang, C.-C., Yen, C.-F., Chiu, M.T.-K., Chang, W.-H., 2009. Mapping RNA exit channel on transcribing RNA polymerase II by FRET analysis. Proc. Natl. Acad. Sci. 106, 127–132. https://doi.org/10.1073/pnas.0811689106

Chen, M., Manley, J.L., 2009. Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. Nat. Rev. Mol. Cell Biol. 10, 741–754. https://doi.org/10.1038/nrm2777

Chen, S., Zhou, Y., Chen, Y., Gu, J., 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 34, i884–i890. https://doi.org/10.1093/bioinformatics/bty560

Chen, T., Guestrin, C., 2016. XGBoost: A Scalable Tree Boosting System. Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Min. 785–794. https://doi.org/10.1145/2939672.2939785

Chen, W., Moore, J., Ozadam, H., Shulha, H.P., Rhind, N., Weng, Z., Moore, M.J., 2018. Transcriptome-wide Interrogation of the Functional Intronome by Spliceosome Profiling. Cell 173, 1031-1044.e13. https://doi.org/10.1016/j.cell.2018.03.062

Chou, M.Y., Rooke, N., Turck, C.W., Black, D.L., 1999. hnRNP H is a component of a splicing enhancer complex that activates a c-src alternative exon in neuronal cells. Mol. Cell. Biol. 19, 69–77. https://doi.org/10.1128/mcb.19.1.69

Christofori, G., Frendewey, D., Keller, W., 1987. Two spliceosomes can form simultaneously and independently on synthetic double-intron messenger RNA precursors. EMBO J. 6, 1747–1755.

Clark, F., Thanaraj, T.A., 2002. Categorization and characterization of transcript-confirmed constitutively and alternatively spliced introns and exons from human. Hum. Mol. Genet. 11, 451–464. https://doi.org/10.1093/hmg/11.4.451

Close, P., East, P., Dirac-Svejstrup, A.B., Hartmann, H., Heron, M., Maslen, S., Chariot, A., Söding, J., Skehel, M., Svejstrup, J.Q., 2012. DBIRD complex integrates alternative mRNA splicing with RNA polymerase II transcript elongation. Nature 484, 386–389. https://doi.org/10.1038/nature10925

Cordin, O., Beggs, J.D., 2013. RNA helicases in splicing. RNA Biol. 10, 83–95. https://doi.org/10.4161/rna.22547

Costantini, M., Clay, O., Auletta, F., Bernardi, G., 2006. An isochore map of human chromosomes. Genome Res. 16, 536–541. https://doi.org/10.1101/gr.4910606

Costantini, M., Musto, H., 2017. The Isochores as a Fundamental Level of Genome Structure and Organization: A General Overview. J. Mol. Evol. 84, 93–103. https://doi.org/10.1007/s00239-017-9785-9

Cramer, P., Cáceres, J.F., Cazalla, D., Kadener, S., Muro, A.F., Baralle, F.E., Kornblihtt, A.R., 1999. Coupling of transcription with alternative splicing: RNA pol II promoters modulate SF2/ASF and 9G8 effects on an exonic splicing enhancer. Mol. Cell 4, 251–258. https://doi.org/10.1016/s1097-2765(00)80372-x

Cramer, P., Pesce, C.G., Baralle, F.E., Kornblihtt, A.R., 1997. Functional association between promoter structure and transcript alternative splicing. Proc. Natl. Acad. Sci. U. S. A. 94, 11456–11460. https://doi.org/10.1073/pnas.94.21.11456

Damianov, A., Ying, Y., Lin, C.-H., Lee, J.-A., Tran, D., Vashisht, A.A., Bahrami-Samani, E., Xing, Y., Martin, K.C., Wohlschlegel, J.A., Black, D.L., 2016. Rbfox Proteins Regulate Splicing as Part of a Large Multiprotein Complex LASR. Cell 165, 606–619. https://doi.org/10.1016/j.cell.2016.03.040

Darman, R.B., Seiler, M., Agrawal, A.A., Lim, K.H., Peng, S., Aird, D., Bailey, S.L., Bhavsar, E.B., Chan, B., Colla, S., Corson, L., Feala, J., Fekkes, P., Ichikawa, K., Keaney, G.F., Lee, L., Kumar, P., Kunii, K., MacKenzie, C., Matijevic, M., Mizui, Y., Myint, K., Park, E.S., Puyang, X., Selvaraj, A., Thomas, M.P., Tsai, J., Wang, J.Y., Warmuth, M., Yang, H., Zhu, P., Garcia-Manero, G., Furman, R.R., Yu, L., Smith, P.G., Buonamici, S., 2015. Cancer-Associated SF3B1 Hotspot Mutations Induce Cryptic 3′ Splice Site Selection through Use of a Different Branch Point. Cell Rep. 13, 1033–1045. https://doi.org/10.1016/j.celrep.2015.09.053

David, C.J., Boyne, A.R., Millhouse, S.R., Manley, J.L., 2011. The RNA polymerase II C-terminal domain promotes splicing activation through recruitment of a U2AF65-Prp19 complex. Genes Dev. 25, 972–983. https://doi.org/10.1101/gad.2038011

de Almeida, S.F., Grosso, A.R., Koch, F., Fenouil, R., Carvalho, S., Andrade, J., Levezinho, H., Gut, M., Eick, D., Gut, I., Andrau, J.-C., Ferrier, P., Carmo-Fonseca, M., 2011. Splicing

enhances recruitment of methyltransferase HYPB/Setd2 and methylation of histone H3 Lys36. Nat. Struct. Mol. Biol. 18, 977–983. https://doi.org/10.1038/nsmb.2123

de la Mata, M., Alonso, C.R., Kadener, S., Fededa, J.P., Blaustein, M., Pelisch, F., Cramer, P., Bentley, D., Kornblihtt, A.R., 2003. A Slow RNA Polymerase II Affects Alternative Splicing In Vivo. Mol. Cell 12, 525–532. https://doi.org/10.1016/j.molcel.2003.08.001

de la Mata, M., Kornblihtt, A.R., 2006. RNA polymerase II C-terminal domain mediates regulation of alternative splicing by SRp20. Nat. Struct. Mol. Biol. 13, 973–980. https://doi.org/10.1038/nsmb1155

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., Gingeras, T.R., 2013. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29, 15–21. https://doi.org/10.1093/bioinformatics/bts635

Dominguez, D., Tsai, Y.-H., Weatheritt, R., Wang, Y., Blencowe, B.J., Wang, Z., 2016. An extensive program of periodic alternative splicing linked to cell cycle progression. eLife 5, 185. https://doi.org/10.7554/eLife.10288

Dönmez, G., Hartmuth, K., Kastner, B., Will, C.L., Lührmann, R., 2007. The 5′ End of U2 snRNA Is in Close Proximity to U1 and Functional Sites of the Pre-mRNA in Early Spliceosomal Complexes. Mol. Cell 25, 399–411. https://doi.org/10.1016/j.molcel.2006.12.019

Dowhan, D.H., Hong, E.P., Auboeuf, D., Dennis, A.P., Wilson, M.M., Berget, S.M., O'Malley, B.W., 2005. Steroid hormone receptor coactivation and alternative RNA splicing by U2AF65-related proteins CAPERalpha and CAPERbeta. Mol. Cell 17, 429–439. https://doi.org/10.1016/j.molcel.2004.12.025

Dowle, M., Srinivasan, A., 2019. data.table: Extension of `data.frame`.

Doyle, O., Corden, J.L., Murphy, C., Gall, J.G., 2002. The distribution of RNA polymerase II largest subunit (RPB1) in the Xenopus germinal vesicle. J. Struct. Biol. 140, 154–166. https://doi.org/10.1016/S1047-8477(02)00547-6

Dredge, B.K., Stefani, G., Engelhard, C.C., Darnell, R.B., 2005. Nova autoregulation reveals dual functions in neuronal splicing. EMBO J. 24, 1608–1620. https://doi.org/10.1038/sj.emboj.7600630

Drexler, H.L., Choquet, K., Churchman, L.S., 2020. Splicing Kinetics and Coordination Revealed by Direct Nascent RNA Sequencing through Nanopores. Mol. Cell 77, 985-998.e8. https://doi.org/10.1016/j.molcel.2019.11.017

Dujardin, G., Lafaille, C., de la Mata, M., Marasco, L.E., Muñoz, M.J., Le Jossic-Corcos, C., Corcos, L., Kornblihtt, A.R., 2014. How Slow RNA Polymerase II Elongation Favors Alternative Exon Skipping. Mol. Cell 1–8. https://doi.org/10.1016/j.molcel.2014.03.044

Duncan, K., Umen, J.G., Guthrie, C., 2000. A putative ubiquitin ligase required for efficient mRNA export differentially affects hnRNP transport. Curr. Biol. CB 10, 687–696. https://doi.org/10.1016/S0960-9822(00)00527-3

Egecioglu, D.E., Chanfreau, G., 2011. Proofreading and spellchecking: A two-tier strategy for pre-mRNA splicing quality control. RNA N. Y. N 17, 383–389. https://doi.org/10.1261/rna.2454711

Erkelenz, S., Mueller, W.F., Evans, M.S., Busch, A., Schöneweis, K., Hertel, K.J., Schaal, H., 2013. Position-dependent splicing activation and repression by SR and hnRNP proteins rely on common mechanisms. RNA N. Y. N 19, 96–102. https://doi.org/10.1261/rna.037044.112

Fasken, M.B., Stewart, M., Corbett, A.H., 2008. Functional significance of the interaction between the mRNA-binding protein, Nab2, and the nuclear pore-associated protein, Mlp1, in mRNA export. J. Biol. Chem. 283, 27130–27143. https://doi.org/10.1074/jbc.M803649200

Feng, Y.-Y., Ramu, A., Cotto, K.C., Skidmore, Z.L., Kunisaki, J., Conrad, D.F., Lin, Y., Chapman, W.C., Uppaluri, R., Govindan, R., Griffith, O.L., Griffith, M., 2018. RegTools: Integrated analysis of genomic and transcriptomic data for discovery of splicing variants in cancer. bioRxiv 436634. https://doi.org/10.1101/436634

Fernandez-Martinez, J., Kim, S.J., Shi, Y., Upla, P., Pellarin, R., Gagnon, M., Chemmama, I.E., Wang, J., Nudelman, I., Zhang, W., Williams, R., Rice, W.J., Stokes, D.L., Zenklusen, D., Chait, B.T., Sali, A., Rout, M.P., 2016. Structure and Function of the Nuclear Pore Complex Cytoplasmic mRNA Export Platform. Cell 167, 1215-1228.e25. https://doi.org/10.1016/j.cell.2016.10.028

Fica, S.M., Mefford, M.A., Piccirilli, J.A., Staley, J.P., 2014. Evidence for a group II intron–like catalytic triplex in the spliceosome. Nat. Struct. Mol. Biol. 21, 464–471. https://doi.org/10.1038/nsmb.2815

Fica, S.M., Tuttle, N., Novak, T., Li, N.-S., Lu, J., Koodathingal, P., Dai, Q., Staley, J.P., Piccirilli, J.A., 2013. RNA catalyses nuclear pre-mRNA splicing. Nature 503, 229–234. https://doi.org/10.1038/nature12734

Flynn, R.A., Zhang, Q.C., Spitale, R.C., Lee, B., Mumbach, M.R., Chang, H.Y., 2016. Transcriptome-wide interrogation of RNA secondary structure in living cells with icSHAPE. Nat. Protoc. 11, 273–290. https://doi.org/10.1038/nprot.2016.011

Fong, N., Kim, H., Zhou, Y., Ji, X., Qiu, J., Saldi, T., Diener, K., Jones, K., Fu, X.-D., Bentley, D.L., 2014. Pre-mRNA splicing is facilitated by an optimal RNA polymerase II elongation rate. Genes Dev. 28, 2663–2676. https://doi.org/10.1101/gad.252106.114

Fu, X.-D., Ares, M., 2014. Context-dependent control of alternative splicing by RNA-binding proteins. Nat. Rev. Genet. 15, 689–701. https://doi.org/10.1038/nrg3778

Furger, A., O'Sullivan, J.M., Binnie, A., Lee, B.A., Proudfoot, N.J., 2002. Promoter proximal splice sites enhance transcription. Genes Dev. 16, 2792–2799. https://doi.org/10.1101/gad.983602

Galej, W.P., Oubridge, C., Newman, A.J., Nagai, K., 2013. Crystal structure of Prp8 reveals active site cavity of the spliceosome. Nature 493, 638–643. https://doi.org/10.1038/nature11843

Galloway, A., Cowling, V.H., 2019. mRNA cap regulation in mammalian cell function and fate. Biochim. Biophys. Acta BBA - Gene Regul. Mech., mRNA modifications in gene expression control 1862, 270–279. https://doi.org/10.1016/j.bbagrm.2018.09.011

Galy, V., Gadal, O., Fromont-Racine, M., Romano, A., Jacquier, A., Nehrbass, U., 2004. Nuclear retention of unspliced mRNAs in yeast is mediated by perinuclear Mlp1. Cell 116, 63–73.

Gao, K., Masuda, A., Matsuura, T., Ohno, K., 2008. Human branch point consensus sequence is yUnAy. Nucleic Acids Res. 36, 2257–2267. https://doi.org/10.1093/nar/gkn073

Gilbert, W., Guthrie, C., 2004. The Glc7p nuclear phosphatase promotes mRNA export by facilitating association of Mex67p with mRNA. Mol. Cell 13, 201–212.

Gilbert, W., Siebel, C.W., Guthrie, C., 2001. Phosphorylation by Sky1p promotes Npl3p shuttling and mRNA dissociation. RNA N. Y. N 7, 302–313.

Girard, C., Will, C.L., Peng, J., Makarov, E.M., Kastner, B., Lemm, I., Urlaub, H., Hartmuth, K., Lührmann, R., 2012. Post-transcriptional spliceosomes are retained in nuclear speckles until splicing completion. Nat. Commun. 3, 994. https://doi.org/10.1038/ncomms1998

Gonatopoulos-Pournatzis, T., Wu, M., Braunschweig, U., Roth, J., Han, H., Best, A.J., Raj, B., Aregger, M., O'Hanlon, D., Ellis, J.D., Calarco, J.A., Moffat, J., Gingras, A.-C., Blencowe, B.J., 2018. Genome-wide CRISPR-Cas9 Interrogation of Splicing Networks Reveals a Mechanism for Recognition of Autism-Misregulated Neuronal Microexons. Mol. Cell 72, 510-524.e12. https://doi.org/10.1016/j.molcel.2018.10.008

Görnemann, J., Kotovic, K.M., Hujer, K., Neugebauer, K.M., 2005. Cotranscriptional Spliceosome Assembly Occurs in a Stepwise Fashion and Requires the Cap Binding Complex. Mol. Cell 19, 53–63. https://doi.org/10.1016/j.molcel.2005.05.007

Grant, R.P., Marshall, N.J., Yang, J.-C., Fasken, M.B., Kelly, S.M., Harreman, M.T., Neuhaus, D., Corbett, A.H., Stewart, M., 2008. Structure of the N-terminal Mlp1-binding domain of the Saccharomyces cerevisiae mRNA-binding protein, Nab2. J. Mol. Biol. 376, 1048–1059. https://doi.org/10.1016/j.jmb.2007.11.087

Green, Deanna M., Johnson, C.P., Hagan, H., Corbett, A.H., 2003. The C-Terminal Domain of Myosin-like Protein 1 (Mlp1p) Is a Docking Site for Heterogeneous Nuclear Ribonucleoproteins That Are Required for mRNA Export. Proc. Natl. Acad. Sci. U. S. A. 100, 1010–1015. https://doi.org/10.2307/3138276?ref=search-gateway:13a3efdabaa01999e6b18d8bca59ff7a

Green, Deanna M, Johnson, C.P., Hagan, H., Corbett, A.H., 2003. The C-terminal domain of myosin-like protein 1 (Mlp1p) is a docking site for heterogeneous nuclear ribonucleoproteins that are required for mRNA export. Proc. Natl. Acad. Sci. U. S. A. 100, 1010–1015. https://doi.org/10.1073/pnas.0336594100

Grünwald, D., Singer, R.H., 2010. In vivo imaging of labelled endogenous β-actin mRNA during nucleocytoplasmic transport. Nature 467, 604–607. https://doi.org/10.1038/nature09438

Grüter, P., Tabernero, C., von Kobbe, C., Schmitt, C., Saavedra, C., Bachi, A., Wilm, M., Felber, B.K., Izaurralde, E., 1998. TAP, the Human Homolog of Mex67p, Mediates CTE-Dependent RNA Export from the Nucleus. Mol. Cell 1, 649–659. https://doi.org/10.1016/S1097-2765(00)80065-9

Gueroussov, S., Weatheritt, R.J., O'Hanlon, D., Lin, Z.-Y., Narula, A., Gingras, A.-C., Blencowe, B.J., 2017. Regulatory Expansion in Mammals of Multivalent hnRNP Assemblies that Globally Control Alternative Splicing. Cell 170, 324-339.e23. https://doi.org/10.1016/j.cell.2017.06.037

guez-Navarro, S.R., Hurt, E., 2011. Linking gene regulation to mRNA production and export. Curr. Opin. Cell Biol. 23, 302–309. https://doi.org/10.1016/j.ceb.2010.12.002

Guo, R., Zheng, L., Park, J.W., Lv, R., Chen, H., Jiao, F., Xu, W., Mu, S., Wen, H., Qiu, J., Wang, Z., Yang, P., Wu, F., Hui, J., Fu, X., Shi, X., Shi, Y.G., Xing, Y., Lan, F., Shi, Y., 2014. BS69/ZMYND11 reads and connects histone H3.3 lysine 36 trimethylation-decorated chromatin to regulated pre-mRNA processing. Mol. Cell 56, 298–310. https://doi.org/10.1016/j.molcel.2014.08.022

Guth, S., Martínez, C., Gaur, R.K., Valcárcel, J., 1999. Evidence for Substrate-Specific Requirement of the Splicing Factor U2AF35 and for Its Function after Polypyrimidine Tract Recognition by U2AF65. Mol. Cell. Biol. 19, 8263–8271. https://doi.org/10.1128/MCB.19.12.8263

Gwizdek, C., Hobeika, M., Kus, B., Ossareh-Nazari, B., Dargemont, C., Rodriguez, M.S., 2005. The mRNA Nuclear Export Factor Hpr1 Is Regulated by Rsp5-mediated Ubiquitylation. J. Biol. Chem. 280, 13401–13405. https://doi.org/10.1074/jbc.C500040200

Gwizdek, C., Iglesias, N., Rodriguez, M.S., Ossareh-Nazari, B., Hobeika, M., Divita, G., Stutz, F., Dargemont, C., 2006. Ubiquitin-associated domain of Mex67 synchronizes recruitment of the mRNA export machinery with transcription. Proc. Natl. Acad. Sci. U. S. A. 103, 16376–16381.

Häcker, S., Krebber, H., 2004. Differential Export Requirements for Shuttling Serine/Arginine-type mRNA-binding Proteins. J. Biol. Chem. 279, 5049–5052. https://doi.org/10.1074/jbc.C300522200

Hackmann, A., Wu, H., Schneider, U.-M., Meyer, K., Jung, K., Krebber, H., 2014. Quality control of spliced mRNAs requires the shuttling SR proteins Gbp2 and Hrb1. Nat. Commun. 5, 3123. https://doi.org/10.1038/ncomms4123

Harigaya, Y., Parker, R., 2012. Global analysis of mRNA decay intermediates in Saccharomyces cerevisiae. Proc. Natl. Acad. Sci. 109, 11764–11769. https://doi.org/10.1073/pnas.1119741109

Harlen, K.M., Trotta, K.L., Smith, E.E., Mosaheb, M.M., Fuchs, S.M., Churchman, L.S., 2016. Comprehensive RNA Polymerase II Interactomes Reveal Distinct and Varied Roles for Each Phospho-CTD Residue. CellReports 15, 2147–2158. https://doi.org/10.1016/j.celrep.2016.05.010

Haselbach, D., Komarov, I., Agafonov, D.E., Hartmuth, K., Graf, B., Dybkov, O., Urlaub, H., Kastner, B., Lührmann, R., Stark, H., 2018. Structure and Conformational Dynamics of the Human Spliceosomal Bact Complex. Cell 172, 454-464.e11. https://doi.org/10.1016/j.cell.2018.01.010

He, Y., Staley, J.P., Andersen, G.R., Nielsen, K.H., 2017. Structure of the DEAH/RHA ATPase Prp43p bound to RNA implicates a pair of hairpins and motif Va in translocation along RNA. RNA N. Y. N 23, 1110–1124. https://doi.org/10.1261/rna.060954.117

Herzel, L., Ottoz, D.S.M., Alpert, T., Neugebauer, K.M., 2017. Splicing and transcription touch base: co-transcriptional spliceosome assembly and function. Nat. Rev. Mol. Cell Biol. 5, 347. https://doi.org/10.1038/nrm.2017.63

Herzel, L., Straube, K., Neugebauer, K.M., 2018. Long-read sequencing of nascent RNA reveals coupling among RNA processing events. Genome Res. 28, 1008–1019. https://doi.org/10.1101/gr.232025.117

Hilleren, P.J., Parker, R., 2003. Cytoplasmic degradation of splice-defective pre-mRNAs and intermediates. Mol. Cell 12, 1453–1465.

Hirose, Y., Tacke, R., Manley, J.L., 1999. Phosphorylated RNA polymerase II stimulates pre-mRNA splicing. Genes Dev. 13, 1234–1239. https://doi.org/10.1101/gad.13.10.1234

Hollander, D., Naftelberg, S., Lev-Maor, G., Kornblihtt, A.R., Ast, G., 2016. How Are Short Exons Flanked by Long Introns Defined and Committed to Splicing? Trends Genet. 32, 596–606. https://doi.org/10.1016/j.tig.2016.07.003

Hoskins, A.A., Gelles, J., Moore, M.J., 2011. New insights into the spliceosome by single molecule fluorescence microscopy. Curr. Opin. Chem. Biol. 1–7. https://doi.org/10.1016/j.cbpa.2011.10.010

Hsin, J.-P., Manley, J.L., 2012. The RNA polymerase II CTD coordinates transcription and RNA processing. Genes Dev. 26, 2119–2137. https://doi.org/10.1101/gad.200303.112

Huang, S., Spector, D.L., 1996. Intron-dependent recruitment of pre-mRNA splicing factors to sites of transcription. J. Cell Biol. 133, 719–732. https://doi.org/10.1083/jcb.133.4.719

Huang, S., Spector, D.L., 1991. Nascent pre-mRNA transcripts are associated with nuclear regions enriched in splicing factors. Genes Dev. 5, 2288–2302. https://doi.org/10.1101/gad.5.12a.2288

Huang, Y., Steitz, J.A., 2005. SRprises along a Messenger's Journey. Mol. Cell 17, 613–615. https://doi.org/10.1016/j.molcel.2005.02.020

Huber, W., Carey, V.J., Gentleman, R., Anders, S., Carlson, M., Carvalho, B.S., Bravo, H.C., Davis, S., Gatto, L., Girke, T., Gottardo, R., Hahne, F., Hansen, K.D., Irizarry, R.A., Lawrence, M., Love, M.I., MacDonald, J., Obenchain, V., Oleś, A.K., Pagès, H., Reyes, A., Shannon, P., Smyth, G.K., Tenenbaum, D., Waldron, L., Morgan, M., 2015. Orchestrating high-throughput genomic analysis with Bioconductor. Nat. Methods 12, 115–121. https://doi.org/10.1038/nmeth.3252

Hülsmann, B.B., Labokha, A.A., Görlich, D., 2012. The Permeability of Reconstituted Nuclear Pores Provides Direct Evidence for the Selective Phase Model. Cell 150, 738–751. https://doi.org/10.1016/j.cell.2012.07.019

Hunter, J.D., 2007. Matplotlib: A 2D Graphics Environment. Comput. Sci. Eng. 9, 90–95. https://doi.org/10.1109/MCSE.2007.55

Hyung, D., Kim, J., Cho, S.Y., Park, C., 2018. ASpedia: a comprehensive encyclopedia of human alternative splicing. Nucleic Acids Res. 46, D58–D63. https://doi.org/10.1093/nar/gkx1014

Iannone, C., Pohl, A., Papasaikas, P., Soronellas, D., Vicent, G.P., Beato, M., ValcáRcel, J., 2015. Relationship between nucleosome positioning and progesterone-induced alternative splicing in breast cancer cells. RNA N. Y. N 21, 360–374. https://doi.org/10.1261/rna.048843.114

Ibrahim, E.C., Schaal, T.D., Hertel, K.J., Reed, R., Maniatis, T., 2005. Serine/arginine-rich protein-dependent suppression of exon skipping by exonic splicing enhancers. Proc. Natl. Acad. Sci. U. S. A. 102, 5002–5007. https://doi.org/10.1073/pnas.0500543102

Iglesias, N., Tutucci, E., Gwizdek, C., Vinciguerra, P., Von Dach, E., Corbett, A.H., Dargemont, C., Stutz, F., 2010. Ubiquitin-mediated mRNP dynamics and surveillance prior to budding yeast mRNA export. Genes Dev. 24, 1927–1938. https://doi.org/10.1101/gad.583310

Jimeno-González, S., Payán-Bravo, L., Muñoz-Cabello, A.M., Guijo, M., Gutierrez, G., Prado, F., Reyes, J.C., 2015. Defective histone supply causes changes in RNA polymerase II elongation rate and cotranscriptional pre-mRNA splicing. Proc. Natl. Acad. Sci. 112, 14840–14845. https://doi.org/10.1073/pnas.1506760112

Johnson, S.A., Cubberley, G., Bentley, D.L., 2009. Cotranscriptional Recruitment of the mRNA Export Factor Yra1 by Direct Interaction with the 3′ End Processing Factor Pcf11. Mol. Cell 33, 215–226. https://doi.org/10.1016/j.molcel.2008.12.007

Kannan, R., Hartnett, S., Voelker, R.B., Berglund, J.A., Staley, J.P., Baumann, P., 2013. Intronic sequence elements impede exon ligation and trigger a discard pathway that yields functional telomerase RNA in fission yeast. Genes Dev. 27, 627–638. https://doi.org/10.1101/gad.212738.112

Kannan, R., Helston, R.M., Dannebaum, R.O., Baumann, P., 2015. Diverse mechanisms for spliceosome-mediated 3′ end processing of telomerase RNA. Nat. Commun. 6, 6104. https://doi.org/10.1038/ncomms7104

Kassambara, A., 2020. ggpubr: "ggplot2" Based Publication Ready Plots.

Kastner, B., Will, C.L., Stark, H., Lührmann, R., 2019. Structural Insights into Nuclear pre-mRNA Splicing in Higher Eukaryotes. Cold Spring Harb. Perspect. Biol. a032417. https://doi.org/10.1101/cshperspect.a032417

Katahira, J., Sträßer, K., Podtelejnikov, A., Mann, M., Jung, J.U., Hurt, E., 1999. The Mex67p-mediated nuclear mRNA export pathway is conserved from yeast to human. EMBO J. 18, 2593–2609. https://doi.org/10.1093/emboj/18.9.2593

Keiper, S., Papasaikas, P., Will, C.L., Valcárcel, J., Girard, C., Lührmann, R., 2019. Smu1 and RED are required for activation of spliceosomal B complexes assembled on short introns. Nat. Commun. 10, 3639. https://doi.org/10.1038/s41467-019-11293-8

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

Khodor, Y.L., Rodriguez, J., Abruzzi, K.C., Tang, C.H.A., Marr, M.T., Rosbash, M., 2011. Nascent-seq indicates widespread cotranscriptional pre-mRNA splicing in Drosophila. Genes Dev. 25, 2502–2512. https://doi.org/10.1101/gad.178962.111

Kim, S.J., Fernandez-Martinez, J., Nudelman, I., Shi, Y., Zhang, W., Raveh, B., Herricks, T., Slaughter, B.D., Hogan, J.A., Upla, P., Chemmama, I.E., Pellarin, R., Echeverria, I., Shivaraju, M., Chaudhury, A.S., Wang, J., Williams, R., Unruh, J.R., Greenberg, C.H., Jacobs, E.Y., Yu, Z., Cruz, M.J. de la, Mironska, R., Stokes, D.L., Aitchison, J.D., Jarrold, M.F., Gerton, J.L., Ludtke, S.J., Akey, C.W., Chait, B.T., Sali, A., Rout, M.P., 2018. Integrative structure and functional anatomy of a nuclear pore complex. Nature 555, 475–482. https://doi.org/10.1038/nature26003

Kim, S.W., Taggart, A.J., Heintzelman, C., Cygan, K.J., Hull, C.G., Wang, J., Shrestha, B., Fairbrother, W.G., 2017. Widespread intra-dependencies in the removal of introns from human transcripts. Nucleic Acids Res. 45, 9503–9513. https://doi.org/10.1093/nar/gkx661

Kizer, K.O., Phatnani, H.P., Shibata, Y., Hall, H., Greenleaf, A.L., Strahl, B.D., 2005. A novel domain in Set2 mediates RNA polymerase II interaction and couples histone H3 K36 methylation with transcript elongation. Mol. Cell. Biol. 25, 3305–3316. https://doi.org/10.1128/MCB.25.8.3305-3316.2005

Köhler, A., Hurt, E., 2007. Exporting RNA from the nucleus to the cytoplasm. Nat. Rev. Mol. Cell Biol. 8, 761–773. https://doi.org/10.1038/nrm2255

Kolasinska-Zwierz, P., Down, T., Latorre, I., Liu, T., Liu, X.S., Ahringer, J., 2009. Differential chromatin marking of introns and expressed exons by H3K36me3. Nat. Genet. 41, 376–381. https://doi.org/10.1038/ng.322

Koodathingal, P., Novak, T., Piccirilli, J.A., Staley, J.P., 2010. The DEAH Box ATPases Prp16 and Prp43 Cooperate to Proofread 5′ Splice Site Cleavage during Pre-mRNA Splicing. Mol. Cell 39, 385–395. https://doi.org/10.1016/j.molcel.2010.07.014

Koodathingal, P., Staley, J.P., 2013. Splicing fidelity: DEAD/H-box ATPases as molecular clocks. RNA Biol. 10, 1073–1079. https://doi.org/10.4161/rna.25245

Kosova, B., Panté, N., Rollenhagen, C., Podtelejnikov, A., Mann, M., Aebi, U., Hurt, E., 2000. Mlp2p, a component of nuclear pore attached intranuclear filaments, associates with nic96p. J. Biol. Chem. 275, 343–350.

Köster, J., Rahmann, S., 2012. Snakemake—a scalable bioinformatics workflow engine. Bioinformatics 28, 2520–2522. https://doi.org/10.1093/bioinformatics/bts480

Krismer, K., Bird, M.A., Varmeh, S., Handly, E.D., Gattinger, A., Bernwinkler, T., Anderson, D.A., Heinzel, A., Joughin, B.A., Kong, Y.W., Cannell, I.G., Yaffe, M.B., 2020. Transite: A Computational Motif-Based Analysis Platform That Identifies RNA-Binding Proteins Modulating Changes in Gene Expression. Cell Rep. 32, 108064. https://doi.org/10.1016/j.celrep.2020.108064

Lacadie, S.A., 2006. In vivo commitment to yeast cotranscriptional splicing is sensitive to transcription elongation mutants. Genes Dev. 20, 2055–2066. https://doi.org/10.1101/gad.1434706

Lacadie, S.A., Rosbash, M., 2005. Cotranscriptional Spliceosome Assembly Dynamics and the Role of U1 snRNA:5′ss Base Pairing in Yeast. Mol. Cell 19, 65–75. https://doi.org/10.1016/j.molcel.2005.05.006

Lallena, M.J., Chalmers, K.J., Llamazares, S., Lamond, A.I., Valcárcel, J., 2002. Splicing regulation at the second catalytic step by Sex-lethal involves 3’ splice site recognition by SPF45. Cell 109, 285–296.

Le Hir, H., Saulière, J., Wang, Z., 2016. The exon junction complex as a node of post-transcriptional networks. Nat. Rev. Mol. Cell Biol. 17, 41–54. https://doi.org/10.1038/nrm.2015.7

Lee, E.S., Wolf, E.J., Smith, H.W., Emili, A., Palazzo, A.F., 2019. TPR is required for the nuclear export of mRNAs and lncRNAs from intronless and intron-poor genes. bioRxiv 740498. https://doi.org/10.1101/740498

Lee, M.S., Henry, M., Silver, P.A., 1996. A protein that shuttles between the nucleus and the cytoplasm is an important mediator of RNA export. Genes Dev. 10, 1233–1246.

Legrain, P., Rosbash, M., 1989. Some cis- and trans-acting mutants for splicing target pre-mRNA to the cytoplasm. Cell 57, 573–583.

LeMaire, M.F., Thummel, C.S., 1990. Splicing precedes polyadenylation during Drosophila E74A transcription. Mol. Cell. Biol. 10, 6059–6063.

Lemaire, S., Fontrodona, N., Aubé, F., Claude, J.-B., Polvèche, H., Modolo, L., Bourgeois, C.F., Mortreux, F., Auboeuf, D., 2019. Characterizing the interplay between gene nucleotide composition bias and splicing. Genome Biol. 20, 259. https://doi.org/10.1186/s13059-019-1869-y

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., Subgroup, 1000 Genome Project Data Processing, 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Li, P., Banjade, S., Cheng, H.-C., Kim, S., Chen, B., Guo, L., Llaguno, M., Hollingsworth, J.V., King, D.S., Banani, S.F., Russo, P.S., Jiang, Q.-X., Nixon, B.T., Rosen, M.K., 2012. Phase transitions in the assembly of multivalent signalling proteins. Nature 483, 336–340. https://doi.org/10.1038/nature10879

Li, Y.I., Knowles, D.A., Humphrey, J., Barbeira, A.N., Dickinson, S.P., Im, H.K., Pritchard, J.K., 2018. Annotation-free quantification of RNA splicing using LeafCutter. Nat. Genet. 50, 151–158. https://doi.org/10.1038/s41588-017-0004-9

Lin, D.H., Correia, A.R., Cai, S.W., Huber, F.M., Jette, C.A., Hoelz, A., 2018. Structural and functional analysis of mRNA export regulation by the nuclear pore complex. Nat. Commun. 9, 2319. https://doi.org/10.1038/s41467-018-04459-3

Listerman, I., Sapra, A.K., Neugebauer, K.M., 2006. Cotranscriptional coupling of splicing factor recruitment and precursor messenger RNA splicing in mammalian cells. Nat. Struct. Mol. Biol. 13, 815. https://doi.org/10.1038/nsmb1135

Liu, Z., Liu, Q., Yang, X., Zhang, Y., Norris, M., Chen, X., Cheema, J., Ding, Y., 2019. In vivo nuclear RNA structurome reveals RNA-structure regulation of mRNA processing in plants. bioRxiv 839506. https://doi.org/10.1101/839506

Logsdon, G.A., Vollger, M.R., Eichler, E.E., 2020. Long-read human genome sequencing and its applications. Nat. Rev. Genet. 1–18. https://doi.org/10.1038/s41576-020-0236-x

Lorenz, R., Bernhart, S.H., Höner zu Siederdissen, C., Tafer, H., Flamm, C., Stadler, P.F., Hofacker, I.L., 2011. ViennaRNA Package 2.0. Algorithms Mol. Biol. 6, 26. https://doi.org/10.1186/1748-7188-6-26

Lubas, M., Christensen, M.S., Kristiansen, M.S., Domanski, M., Falkenby, L.G., Lykke-Andersen, S., Andersen, J.S., Dziembowski, A., Jensen, T.H., 2011. Interaction Profiling Identifies the Human Nuclear Exosome Targeting Complex. Mol. Cell 43, 624–637. https://doi.org/10.1016/j.molcel.2011.06.028

Luco, R.F., Pan, Q., Tominaga, K., Blencowe, B.J., Pereira-Smith, O.M., Misteli, T., 2010. Regulation of Alternative Splicing by Histone Modifications. Science 327, 996–1000. https://doi.org/10.1126/science.1184208

Lund, M.K., Guthrie, C., 2005. The DEAD-Box Protein Dbp5p Is Required to Dissociate Mex67p from Exported mRNPs at the Nuclear Rim. Mol. Cell 20, 645–651. https://doi.org/10.1016/j.molcel.2005.10.005

Lundberg, S.M., Lee, S.-I., 2017. A Unified Approach to Interpreting Model Predictions, in: Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R. (Eds.), Advances in Neural Information Processing Systems 30. Curran Associates, Inc., pp. 4765–4774.

Ma, W.K., Cloutier, S.C., Tran, E.J., 2013. The DEAD-box Protein Dbp2 Functions with the RNA-Binding Protein Yra1 to Promote mRNP Assembly. J. Mol. Biol. https://doi.org/10.1016/j.jmb.2013.05.016

Mackereth, C.D., Madl, T., Bonnal, S., Simon, B., Zanier, K., Gasch, A., Rybin, V., Valcárcel, J., Sattler, M., 2011. Multi-domain conformational selection underlies pre-mRNA splicing regulation by U2AF. Nature 475, 408–411. https://doi.org/10.1038/nature10171

Manning, K.S., Cooper, T.A., 2017. The roles of RNA processing in translating genotype to phenotype. Nat. Rev. Mol. Cell Biol. 18, 102–114. https://doi.org/10.1038/nrm.2016.139

Marfatia, K.A., Crafton, E.B., Green, D.M., Corbett, A.H., 2003. Domain analysis of the Saccharomyces cerevisiae heterogeneous nuclear ribonucleoprotein, Nab2p. Dissecting the requirements for Nab2p-facilitated poly(A) RNA export. J. Biol. Chem. 278, 6731–6740. https://doi.org/10.1074/jbc.M207571200

Martin, M., 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal 17, 10–12. https://doi.org/10.14806/ej.17.1.200

Mauger, D.M., Lin, C., Garcia-Blanco, M.A., 2008. hnRNP H and hnRNP F complex with Fox2 to silence fibroblast growth factor receptor 2 exon IIIc. Mol. Cell. Biol. 28, 5403–5419. https://doi.org/10.1128/MCB.00739-08

Mayas, R.M., Maita, H., Semlow, D.R., Staley, J.P., 2010. Spliceosome discards intermediates via the DEAH box ATPase Prp43p. Proc. Natl. Acad. Sci. U. S. A. 107, 10020–10025. https://doi.org/10.1073/pnas.0906022107/-/DCSupplemental

Mayas, R.M., Maita, H., Staley, J.P., 2006. Exon ligation is proofread by the DExD/H-box ATPase Prp22p. Nat. Struct. Mol. Biol. 13, 482–490. https://doi.org/10.1038/nsmb1093

Mayer, A., di Iulio, J., Maleri, S., Eser, U., Vierstra, J., Reynolds, A., Sandstrom, R., Stamatoyannopoulos, J.A., Churchman, L.S., 2015. Native elongating transcript sequencing reveals human transcriptional activity at nucleotide resolution. Cell 161, 541–554. https://doi.org/10.1016/j.cell.2015.03.010

Mayerle, M., Raghavan, M., Ledoux, S., Price, A., Stepankiw, N., Hadjivassiliou, H., Moehle, E.A., Mendoza, S.D., Pleiss, J.A., Guthrie, C., Abelson, J., 2017. Structural toggle in the RNaseH domain of Prp8 helps balance splicing fidelity and catalytic efficiency. Proc. Natl. Acad. Sci. U. S. A. 542, 201701462–6. https://doi.org/10.1073/pnas.1701462114

McCracken, S., Fong, N., Yankulov, K., Ballantyne, S., Pan, G., Greenblatt, J., Patterson, S.D., Wickens, M., Bentley, D.L., 1997. The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. Nature 385, 357–361. https://doi.org/10.1038/385357a0

Mccracken, S., Rosonina, E., Fong, N., Sikes, M., Beyer, A., O'hare, K., Shuman, S., Bentley, D., 1998. Role of RNA Polymerase II Carboxy-terminal Domain in Coordinating Transcription with RNA Processing. Cold Spring Harb. Symp. Quant. Biol. 63, 301–310. https://doi.org/10.1101/sqb.1998.63.301

McManus, C.J., Graveley, B.R., 2011. RNA structure and the mechanisms of alternative splicing. Curr. Opin. Genet. Dev. 21, 373–379. https://doi.org/10.1016/j.gde.2011.04.001

Meinel, D.M., Burkert-Kautzsch, C., Kieser, A., O'Duibhir, E., Siebert, M., Mayer, A., Cramer, P., Söding, J., Holstege, F.C.P., Sträßer, K., 2013. Recruitment of TREX to the Transcription Machinery by Its Direct Binding to the Phospho-CTD of RNA Polymerase II. PLOS Genet. 9, e1003914. https://doi.org/10.1371/journal.pgen.1003914

Meola, N., Domanski, M., Karadoulama, E., Chen, Y., Gentil, C., Pultz, D., Vitting-Seerup, K., Lykke-Andersen, S., Andersen, J.S., Sandelin, A., Jensen, T.H., 2016. Identification of a Nuclear Exosome Decay Pathway for Processed Transcripts. Mol. Cell 64, 520–533. https://doi.org/10.1016/j.molcel.2016.09.025

Mercer, T.R., Clark, M.B., Andersen, S.B., Brunck, M.E., Haerty, W., Crawford, J., Taft, R.J., Nielsen, L.K., Dinger, M.E., Mattick, J.S., 2015. Genome-wide discovery of human splicing branchpoints. Genome Res. 25, 290–303. https://doi.org/10.1101/gr.182899.114

Merkin, J., Russell, C., Chen, P., Burge, C.B., 2012. Evolutionary dynamics of gene and isoform regulation in Mammalian tissues. Science 338, 1593–1599. https://doi.org/10.1126/science.1228186

Michaud, S., Reed, R., 1993. A functional association between the 5' and 3' splice site is established in the earliest prespliceosome complex (E) in mammals. Genes Dev. 7, 1008–1020. https://doi.org/10.1101/gad.7.6.1008

Misteli, T., Cáceres, J.F., Spector, D.L., 1997. The dynamics of a pre-mRNA splicing factor in living cells. Nature 387, 523–527.

Moore, M.J., Proudfoot, N.J., 2009. Pre-mRNA processing reaches back to transcription and ahead to translation. Cell 136, 688–700. https://doi.org/10.1016/j.cell.2009.02.001

Morris, D.P., Greenleaf, A.L., 2000. The Splicing Factor, Prp40, Binds the Phosphorylated Carboxyl-terminal Domain of RNA Polymerase II. J. Biol. Chem. 275, 39935–39943. https://doi.org/10.1074/jbc.M004118200

Nasim, F.H., Spears, P.A., Hoffmann, H.M., Kuo, H.C., Grabowski, P.J., 1990. A Sequential splicing mechanism promotes selection of an optimal exon by repositioning a downstream 5' splice site in preprotachykinin pre-mRNA. Genes Dev. 4, 1172–1184. https://doi.org/10.1101/gad.4.7.1172

Nasim, F.-U.H., Hutchison, S., Cordeau, M., Chabot, B., 2002. High-affinity hnRNP A1 binding sites and duplex-forming inverted repeats have similar effects on 5' splice site selection in support of a common looping out and repression mechanism. RNA N. Y. N 8, 1078–1089. https://doi.org/10.1017/s1355838202024056

Neugebauer, K.M., 2019. Nascent RNA and the Coordination of Splicing with Transcription. Cold Spring Harb. Perspect. Biol. 11, a032227. https://doi.org/10.1101/cshperspect.a032227

Newman, A.J., Norman, C., 1992. U5 snRNA interacts with exon sequences at 5′ and 3′ splice sites. Cell 68, 743–754. https://doi.org/10.1016/0092-8674(92)90149-7

Niepel, M., Molloy, K.R., Williams, R., Farr, J.C., Meinema, A.C., Vecchietti, N., Cristea, I.M., Chait, B.T., Rout, M.P., Strambio-De-Castillia, C., 2013. The nuclear basket proteins Mlp1p and Mlp2p are part of a dynamic interactome including Esc1p and the proteasome. Mol. Biol. Cell 24, 3920–3938. https://doi.org/10.1091/mbc.E13-07-0412

Niño, C.A., Hérissant, L., Babour, A., Dargemont, C., 2013. mRNA Nuclear Export in Yeast. Chem. Rev. 113, 8523–8545. https://doi.org/10.1021/cr400002g

Nojima, T., Gomes, T., Grosso, A.R.F., Kimura, H., Dye, M.J., Dhir, S., Carmo-Fonseca, M., Proudfoot, N.J., 2015. Mammalian NET-Seq Reveals Genome-wide Nascent Transcription Coupled to RNA Processing. Cell 161, 526–540. https://doi.org/10.1016/j.cell.2015.03.027

Nojima, T., Rebelo, K., Gomes, T., Grosso, A.R., Proudfoot, N.J., Carmo-Fonseca, M., 2018. RNA Polymerase II Phosphorylated on CTD Serine 5 Interacts with the Spliceosome during Co-transcriptional Splicing. Mol. Cell 72, 369-379.e4. https://doi.org/10.1016/j.molcel.2018.09.004

Oeffinger, M., Zenklusen, D., 2012. To the pore and through the pore: A story of mRNA export kinetics. BBA - Gene Regul. Mech. 1–13. https://doi.org/10.1016/j.bbagrm.2012.02.011

Olthof, A.M., Hyatt, K.C., Kanadia, R.N., 2019. Minor intron splicing revisited: identification of new minor intron-containing genes and tissue-dependent retention and alternative splicing of minor introns. BMC Genomics 20, 686. https://doi.org/10.1186/s12864-019-6046-x

Pai, A.A., Henriques, T., McCue, K., Burkholder, A., Adelman, K., Burge, C.B., 2017. The kinetics of pre-mRNA splicing in the Drosophila genome and the influence of gene architecture. eLife 6, 1123. https://doi.org/10.7554/eLife.32537

Palazzo, A.F., Lee, E.S., 2018. Sequence Determinants for Nuclear Retention and Cytoplasmic Export of mRNAs and lncRNAs. Front. Genet. 9. https://doi.org/10.3389/fgene.2018.00440

Pan, Q., Shai, O., Lee, L.J., Frey, B.J., Blencowe, B.J., 2008. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. Nat. Genet. 40, 1413–1415. https://doi.org/10.1038/ng.259

Pandya-Jones, A., Black, D.L., 2009. Co-transcriptional splicing of constitutive and alternative exons. RNA N. Y. N 15, 1896–1908. https://doi.org/10.1261/rna.1714509

Pawlicki, J.M., Steitz, J.A., 2010. Nuclear networking fashions pre-messenger RNA and primary microRNA transcripts for function. Trends Cell Biol. 20, 52–61. https://doi.org/10.1016/j.tcb.2009.10.004

Pedersen, T.L., 2019. patchwork: The Composer of Plots.

Picard toolkit, 2019. . Broad Inst. GitHub Repos.

Pickrell, J.K., Pai, A.A., Gilad, Y., Pritchard, J.K., 2010. Noisy splicing drives mRNA isoform diversity in human cells. PLoS Genet. 6, e1001236. https://doi.org/10.1371/journal.pgen.1001236.g001

Pineda, J.M.B., Bradley, R.K., 2018. Most human introns are recognized via multiple and tissue-specific branchpoints. Genes Dev. 32, 577–591. https://doi.org/10.1101/gad.312058.118

Piovesan, A., Caracausi, M., Antonaros, F., Pelleri, M.C., Vitale, L., 2016. GeneBase 1.1: a tool to summarize data from NCBI gene datasets and its application to an update of human gene statistics. Database J. Biol. Databases Curation 2016. https://doi.org/10.1093/database/baw153

Plaschka, C., Newman, A.J., Nagai, K., 2019. Structural Basis of Nuclear pre-mRNA Splicing: Lessons from Yeast. Cold Spring Harb. Perspect. Biol. a032391. https://doi.org/10.1101/cshperspect.a032391

Pradeepa, M.M., Sutherland, H.G., Ule, J., Grimes, G.R., Bickmore, W.A., 2012. Psip1/Ledgf p52 binds methylated histone H3K36 and splicing factors and contributes to the regulation of alternative splicing. PLoS Genet. 8, e1002717. https://doi.org/10.1371/journal.pgen.1002717

Query, C.C., Strobel, S.A., Sharp, P.A., 1996. Three recognition events at the branch-site adenine. EMBO J. 15, 1392–1402.

Quinlan, A.R., Hall, I.M., 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26, 841–842. https://doi.org/10.1093/bioinformatics/btq033

R Core Team, 2013. R: A Language and Environment for Statistical Computing.

Raj, A., Tyagi, S., 2010. Detection of individual endogenous RNA transcripts in situ using multiple singly labeled probes. Methods Enzymol. 472, 365–386. https://doi.org/10.1016/S0076-6879(10)72004-8

Raj, B., Irimia, M., Braunschweig, U., Sterne-Weiler, T., O'Hanlon, D., Lin, Z.-Y., Chen, G.I., Easton, L.E., Ule, J., Gingras, A.-C., Eyras, E., Blencowe, B.J., 2014. A Global Regulatory

Mechanism for Activating an Exon Network Required for Neurogenesis. Mol. Cell 56, 90–103. https://doi.org/10.1016/j.molcel.2014.08.011

Ramírez, F., Dündar, F., Diehl, S., Grüning, B.A., Manke, T., 2014. deepTools: a flexible platform for exploring deep-sequencing data. Nucleic Acids Res. 42, W187–W191. https://doi.org/10.1093/nar/gku365

Rauhut, R., Fabrizio, P., Dybkov, O., Hartmuth, K., Pena, V., Chari, A., Kumar, V., Lee, C.-T., Urlaub, H., Kastner, B., Stark, H., Lührmann, R., 2016. Molecular architecture of the Saccharomyces cerevisiae activated spliceosome. Science 353, 1399–1405. https://doi.org/10.1126/science.aag1906

Ray, D., Kazan, H., Cook, K.B., Weirauch, M.T., Najafabadi, H.S., Li, X., Gueroussov, S., Albu, M., Zheng, H., Yang, A., Na, H., Irimia, M., Matzat, L.H., Dale, R.K., Smith, S.A., Yarosh, C.A., Kelly, S.M., Nabet, B., Mecenas, D., Li, W., Laishram, R.S., Qiao, M., Lipshitz, H.D., Piano, F., Corbett, A.H., Carstens, R.P., Frey, B.J., Anderson, R.A., Lynch, K.W., Penalva, L.O.F., Lei, E.P., Fraser, A.G., Blencowe, B.J., Morris, Q.D., Hughes, T.R., 2013. A compendium of RNA-binding motifs for decoding gene regulation. Nature 499, 172–177. https://doi.org/10.1038/nature12311

Reimer, K.A., Mimoso, C., Adelman, K., Neugebauer, K.M., 2020. Rapid and Efficient Co-Transcriptional Splicing Enhances Mammalian Gene Expression. bioRxiv 2020.02.11.944595. https://doi.org/10.1101/2020.02.11.944595

Reynolds, D.J., Hertel, K.J., 2019. Ultra-deep sequencing reveals pre-mRNA splicing as a sequence driven high-fidelity process. PloS One 14, e0223132. https://doi.org/10.1371/journal.pone.0223132

Rosonina, E., Blencowe, B.J., 2004. Analysis of the requirement for RNA polymerase II CTD heptapeptide repeats in pre-mRNA splicing and 3'-end cleavage. RNA N. Y. N 10, 581–589. https://doi.org/10.1261/rna.5207204

Rothbart, S.B., Dickson, B.M., Raab, J.R., Grzybowski, A.T., Krajewski, K., Guo, A.H., Shanle, E.K., Josefowicz, S.Z., Fuchs, S.M., Allis, C.D., Magnuson, T.R., Ruthenburg, A.J., Strahl, B.D., 2015. An Interactive Database for the Assessment of Histone Antibody Specificity. Mol. Cell 59, 502–511. https://doi.org/10.1016/j.molcel.2015.06.022

Saldi, T., Cortazar, M.A., Sheridan, R.M., Bentley, D.L., 2016. Coupling of RNA Polymerase II Transcription Elongation with Pre-mRNA Splicing. J. Mol. Biol. 428, 2623–2635. https://doi.org/10.1016/j.jmb.2016.04.017

Saroufim, M.-A., Bensidoun, P., Raymond, P., Rahman, S., Krause, M.R., Oeffinger, M., Zenklusen, D., 2015. The nuclear basket mediates perinuclear mRNA scanning in budding yeast. J. Cell Biol. 211, 1131–1140. https://doi.org/10.1083/jcb.201503070

Saulière, J., Sureau, A., Expert-Bezançon, A., Marie, J., 2006. The polypyrimidine tract binding protein (PTB) represses splicing of exon 6B from the beta-tropomyosin pre-mRNA by directly

interfering with the binding of the U2AF65 subunit. Mol. Cell. Biol. 26, 8755–8769. https://doi.org/10.1128/MCB.00893-06

Sayani, S., Chanfreau, G.F., 2012. Sequential RNA degradation pathways provide a fail-safe mechanism to limit the accumulation of unspliced transcripts in Saccharomyces cerevisiae. RNA 18, 1563–1572. https://doi.org/10.1261/rna.033779.112

Schlackow, M., Nojima, T., Gomes, T., Dhir, A., Carmo-Fonseca, M., Proudfoot, N.J., 2017. Distinctive Patterns of Transcription and RNA Processing for Human lincRNAs. Mol. Cell 65, 25–38. https://doi.org/10.1016/j.molcel.2016.11.029

Schlautmann, L.P., Gehring, N.H., 2020. A Day in the Life of the Exon Junction Complex. Biomolecules 10. https://doi.org/10.3390/biom10060866

Schmid, M., Jensen, T.H., 2018. Controlling nuclear RNA levels. Nat. Rev. Genet. 19, 518–529. https://doi.org/10.1038/s41576-018-0013-2

Schneider, M., Will, C.L., Anokhina, M., Tazi, J., Urlaub, H., Lührmann, R., 2010. Exon Definition Complexes Contain the Tri-snRNP and Can Be Directly Converted into B-like Precatalytic Splicing Complexes. Mol. Cell 38, 223–235. https://doi.org/10.1016/j.molcel.2010.02.027

Schwer, B., 2008. A Conformational Rearrangement in the Spliceosome Sets the Stage for Prp22-Dependent mRNA Release. Mol. Cell 30, 743–754. https://doi.org/10.1016/j.molcel.2008.05.003

Schwer, B., Guthrie, C., 1992. A conformational rearrangement in the spliceosome is dependent on PRP16 and ATP hydrolysis. EMBO J. 11, 5033–5039.

Segref, A., Sharma, K., Doye, V., Hellwig, A., Hurt, E., 1997. Mex67p, a novel factor for nuclear mRNA export, binds to both poly (A)+ RNA and nuclear pores. EMBO ….

Selth, L.A., Sigurdsson, S., Svejstrup, J.Q., 2010. Transcript Elongation by RNA Polymerase II. Annu. Rev. Biochem. 79, 271–293. https://doi.org/10.1146/annurev.biochem.78.062807.091425

Semlow, D.R., Blanco, M.R., Walter, N.G., Staley, J.P., 2016. Spliceosomal DEAH-Box ATPases Remodel Pre-mRNA to Activate Alternative Splice Sites. Cell 164, 985–998. https://doi.org/10.1016/j.cell.2016.01.025

Semlow, D.R., Staley, J.P., 2012. Staying on message: ensuring fidelity in pre-mRNA splicing. Trends Biochem. Sci. 37, 263–273. https://doi.org/10.1016/j.tibs.2012.04.001

Shah, R.N., Grzybowski, A.T., Cornett, E.M., Johnstone, A.L., Dickson, B.M., Boone, B.A., Cheek, M.A., Cowles, M.W., Maryanski, D., Meiners, M.J., Tiedemann, R.L., Vaughan, R.M., Arora, N., Sun, Z.-W., Rothbart, S.B., Keogh, M.-C., Ruthenburg, A.J., 2018. Examining the Roles of H3K4 Methylation States with Systematically Characterized Antibodies. Mol. Cell 72, 162-177.e7. https://doi.org/10.1016/j.molcel.2018.08.015

Shao, C., Yang, B., Wu, T., Huang, J., Tang, P., Zhou, Y., Zhou, J., Qiu, J., Jiang, L., Li, H., Chen, G., Sun, H., Zhang, Y., Denise, A., Zhang, D.-E., Fu, X.-D., 2014. Mechanisms for U2AF to define 3′ splice sites and regulate alternative splicing in the human genome. Nat. Struct. Mol. Biol. 21, 997–1005. https://doi.org/10.1038/nsmb.2906

Shao, W., Ding, Z., Zheng, Z.-Z., Shen, J.-J., Shen, Y.-X., Pu, J., Fan, Y.-J., Query, C.C., Xu, Y.-Z., 2020. Prp5−Spt8/Spt3 interaction mediates a reciprocal coupling between splicing and transcription. Nucleic Acids Res. 48, 5799–5813. https://doi.org/10.1093/nar/gkaa311

Shao, W., Kim, H.-S., Cao, Y., Xu, Y.-Z., Query, C.C., 2012. A U1-U2 snRNP interaction network during intron definition. Mol. Cell. Biol. 32, 470–478. https://doi.org/10.1128/MCB.06234-11

Shibata, S., Matsuoka, Y., Yoneda, Y., 2002. Nucleocytoplasmic transport of proteins and poly(A)+ RNA in reconstituted Tpr-less nuclei in living mammalian cells. Genes Cells Devoted Mol. Amp Cell. Mech. 7, 421–434.

Silla, T., Karadoulama, E., Mąkosa, D., Lubas, M., Jensen, T.H., 2018. The RNA Exosome Adaptor ZFC3H1 Functionally Competes with Nuclear Export Activity to Retain Target Transcripts. Cell Rep. 23, 2199–2210. https://doi.org/10.1016/j.celrep.2018.04.061

Slowikowski, K., 2020. ggrepel: Automatically Position Non-Overlapping Text Labels with "ggplot2."

Smith, T.S., Heger, A., Sudbery, I., 2017. UMI-tools: Modelling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. Genome Res. gr.209601.116. https://doi.org/10.1101/gr.209601.116

Soheilypour, M., Mofrad, M.R.K., 2018. Quality control of mRNAs at the entry of the nuclear pore: Cooperation in a complex molecular system. Nucleus 9, 202–211. https://doi.org/10.1080/19491034.2018.1439304

Sorek, R., Ast, G., 2003. Intronic sequences flanking alternatively spliced exons are conserved between human and mouse. Genome Res. 13, 1631–1637. https://doi.org/10.1101/gr.1208803

Spies, N., Nielsen, C.B., Padgett, R.A., Burge, C.B., 2009. Biased Chromatin Signatures Around Polyadenylation Sites and Exons. Mol. Cell 36, 245–254. https://doi.org/10.1016/j.molcel.2009.10.008

Staley, J.P., Guthrie, C., 1999. An RNA switch at the 5' splice site requires ATP and the DEAD box protein Prp28p. Mol. Cell 3, 55–64.

Staley, J.P., Guthrie, C., 1998. Mechanical devices of the spliceosome: review motors, clocks, springs, and things. Cell 92, 315–326.

Stewart, M., 2019. Polyadenylation and nuclear export of mRNAs. J. Biol. Chem. 294, 2977–2987. https://doi.org/10.1074/jbc.REV118.005594

Strambio-de-Castillia, C., Blobel, G., Rout, M.P., 1999. Proteins connecting the nuclear pore complex with the nuclear interior. J. Cell Biol. 144, 839–855.

Sträßer, K., Masuda, S., Mason, P., Pfannstiel, J., Oppizzi, M., Rodríguez-Navarro, S., Rondón, A.G., Aguilera, A., Struhl, K., Reed, R., Hurt, E., 2002. TREX is a conserved complex coupling transcription with messenger RNA export. Nature 417, 304–308. https://doi.org/10.1038/nature746

Sun, L., Fazal, F.M., Li, P., Broughton, J.P., Lee, B., Tang, L., Huang, W., Kool, E.T., Chang, H.Y., Zhang, Q.C., 2019. RNA structure maps across mammalian cellular compartments. Nat. Struct. Mol. Biol. 26, 322–330. https://doi.org/10.1038/s41594-019-0200-7

Sun, Y., Hamilton, K., Tong, L., 2020. Recent molecular insights into canonical pre-mRNA 3'-end processing. Transcription 11, 83–96. https://doi.org/10.1080/21541264.2020.1777047

Sung, M.-K., Porras-Yakushi, T.R., Reitsma, J.M., Huber, F.M., Sweredoski, M.J., Hoelz, A., Hess, S., Deshaies, R.J., 2016. A conserved quality-control pathway that mediates degradation of unassembled ribosomal proteins. - PubMed - NCBI. eLife 5, 3429. https://doi.org/10.7554/eLife.19105

Taggart, A.J., Lin, C.-L., Shrestha, B., Heintzelman, C., Kim, S., Fairbrother, W.G., 2017. Large-scale analysis of branchpoint usage across species and cell lines. Genome Res. 27, 639–649. https://doi.org/10.1101/gr.202820.115

Talhouarne, G.J.S., Gall, J.G., 2018. Lariat intronic RNAs in the cytoplasm of vertebrate cells. Proc. Natl. Acad. Sci. U. S. A. 237, 201808816. https://doi.org/10.1073/pnas.1808816115

Tanaka, E., Bailey, T.L., Keich, U., 2014. Improving MEME via a two-tiered significance analysis. Bioinformatics 30, 1965–1973. https://doi.org/10.1093/bioinformatics/btu163

Teng, T., Tsai, J.H., Puyang, X., Seiler, M., Peng, S., Prajapati, S., Aird, D., Buonamici, S., Caleb, B., Chan, B., Corson, L., Feala, J., Fekkes, P., Gerard, B., Karr, C., Korpal, M., Liu, X., T Lowe, J., Mizui, Y., Palacino, J., Park, E., Smith, P.G., Subramanian, V., Wu, Z.J., Zou, J., Yu, L., Chicas, A., Warmuth, M., Larsen, N., Zhu, P., 2017. Splicing modulators act at the branch point adenosine binding pocket defined by the PHF5A-SF3b complex. Nat. Commun. 8, 15522. https://doi.org/10.1038/ncomms15522

Testa, S.M., Disney, M.D., Turner, D.H., Kierzek, R., 1999. Thermodynamics of RNA−RNA Duplexes with 2- or 4-Thiouridines: Implications for Antisense Design and Targeting a Group I Intron. Biochemistry 38, 16655–16662. https://doi.org/10.1021/bi991187d

Tieg, B., Krebber, H., 2013. Dbp5 - from nuclear export to translation. Biochim. Biophys. Acta 1829, 791–798. https://doi.org/10.1016/j.bbagrm.2012.10.010

Tilgner, H., Knowles, D.G., Johnson, R., Davis, C.A., Chakrabortty, S., Djebali, S., Curado, J., Snyder, M., Gingeras, T.R., Guigó, 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. https://doi.org/10.1101/gr.134445.111

Tran, E.J., Zhou, Y., Corbett, A.H., Wente, S.R., 2007. The DEAD-box protein Dbp5 controls mRNA export by triggering specific RNA: protein remodeling events. Mol. Cell 28, 850–859.

Tremblay, B.J.-M., 2019. universalmotif: Import, Modify, and Export Motifs with R.

Tudek, A., Schmid, M., Jensen, T.H., 2019. Escaping nuclear decay: the significance of mRNA export for gene expression. Curr. Genet. 65, 473–476. https://doi.org/10.1007/s00294-018-0913-x

Tutucci, E., Stutz, F., 2011. Keeping mRNPs in check during assembly and nuclear export. Nat. Rev. Mol. Cell Biol. 12, 377–384. https://doi.org/10.1038/nrm3119

Ule, J., Blencowe, B.J., 2019. Alternative Splicing Regulatory Networks: Functions, Mechanisms, and Evolution. Mol. Cell 76, 329–345. https://doi.org/10.1016/j.molcel.2019.09.017

Ule, J., Stefani, G., Mele, A., Ruggiu, M., Wang, X., Taneri, B., Gaasterland, T., Blencowe, B.J., Darnell, R.B., 2006. An RNA map predicting Nova-dependent splicing regulation. Nature 444, 580–586. https://doi.org/10.1038/nature05304

Valencia, P., Dias, A.P., Reed, R., 2008. Splicing promotes rapid and efficient mRNA export in mammalian cells. Proc. Natl. Acad. Sci. 105, 3386–3391. https://doi.org/10.1073/pnas.0800250105

Van Nostrand, E.L., Freese, P., Pratt, G.A., Wang, X., Wei, X., Xiao, R., Blue, S.M., Chen, J.-Y., Cody, N.A.L., Dominguez, D., Olson, S., Sundararaman, B., Zhan, L., Bazile, C., Bouvrette, L.P.B., Bergalet, J., Duff, M.O., Garcia, K.E., Gelboin-Burkhart, C., Hochman, M., Lambert, N.J., Li, H., McGurk, M.P., Nguyen, T.B., Palden, T., Rabano, I., Sathe, S., Stanton, R., Su, A., Wang, R., Yee, B.A., Zhou, B., Louie, A.L., Aigner, S., Fu, X.-D., Lécuyer, E., Burge, C.B., Graveley, B.R., Yeo, G.W., 2020. A large-scale binding and functional map of human RNA-binding proteins. Nature 583, 711–719. https://doi.org/10.1038/s41586-020-2077-3

Van Nostrand, E.L., Pratt, G.A., Shishkin, A.A., Gelboin-Burkhart, C., Fang, M.Y., Sundararaman, B., Blue, S.M., Nguyen, T.B., Surka, C., Elkins, K., Stanton, R., Rigo, F., Guttman, M., Yeo, G.W., 2016. Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). Nat. Methods 13, 508–514. https://doi.org/10.1038/nmeth.3810

Vargas, D.Y., Shah, K., Batish, M., Levandoski, M., Sinha, S., Marras, S.A.E., Schedl, P., Tyagi, S., 2011. Single-Molecule Imaging of Transcriptionally Coupled and Uncoupled Splicing. Cell 147, 1054–1065. https://doi.org/10.1016/j.cell.2011.10.024

Veloso, A., Kirkconnell, K.S., Magnuson, B., Biewen, B., Paulsen, M.T., Wilson, T.E., Ljungman, M., 2014. Rate of elongation by RNA polymerase II is associated with specific gene features and epigenetic modifications. Genome Res. 24, 896–905. https://doi.org/10.1101/gr.171405.113

Vinciguerra, P., Vinciguerra, P., Iglesias, N., Iglesias, N., Camblong, J., Camblong, J., Zenklusen, D., Zenklusen, D., Stutz, F., Stutz, F., 2005. Perinuclear Mlp proteins downregulate gene expression in response to a defect in mRNA export. EMBO J. 24, 813–823. https://doi.org/10.1038/sj.emboj.7600527

Wachutka, L., Caizzi, L., Gagneur, J., Cramer, P., 2019. Global donor and acceptor splicing site kinetics in human cells. eLife 8, e45056. https://doi.org/10.7554/eLife.45056

Wagih, O., 2017. ggseqlogo: a versatile R package for drawing sequence logos. Bioinformatics 33, 3645–3647. https://doi.org/10.1093/bioinformatics/btx469

Wahl, M.C., Will, C.L., Lührmann, R., 2009. The spliceosome: design principles of a dynamic RNP machine. Cell 136, 701–718. https://doi.org/10.1016/j.cell.2009.02.009

Wang, E.T., Sandberg, R., Luo, S., Khrebtukova, I., Zhang, L., Mayr, C., Kingsmore, S.F., Schroth, G.P., Burge, C.B., 2008. Alternative isoform regulation in human tissue transcriptomes. Nature 456, 470–476. https://doi.org/10.1038/nature07509

Wang, Z., Burge, C.B., 2008. Splicing regulation: From a parts list of regulatory elements to an integrated splicing code. RNA 14, 802–813. https://doi.org/10.1261/rna.876308

Warnasooriya, C., Feeney, C.F., Laird, K.M., Ermolenko, D.N., Kielkopf, C.L., 2020. A splice site-sensing conformational switch in U2AF2 is modulated by U2AF1 and its recurrent myelodysplasia-associated mutation. Nucleic Acids Res. 48, 5695–5709. https://doi.org/10.1093/nar/gkaa293

Weirich, C.S., Erzberger, J.P., Flick, J.S., Berger, J.M., Thorner, J., Weis, K., 2006. Activation of the DExD/H-box protein Dbp5 by the nuclear-pore protein Gle1 and its coactivator InsP6 is required for mRNA export. Nat. Cell Biol. 8, 668–676. https://doi.org/10.1038/ncb1424

Werner, M.S., Ruthenburg, A.J., 2015. Nuclear Fractionation Reveals Thousands of Chromatin-Tethered Noncoding RNAs Adjacent to Active Genes. CellReports 12, 1089–1098. https://doi.org/10.1016/j.celrep.2015.07.033

Westholm, J.O., Lai, E.C., 2011. Mirtrons: microRNA biogenesis via splicing. Biochimie 93, 1897–1904. https://doi.org/10.1016/j.biochi.2011.06.017

Wetterberg, I., Baurén, G., Wieslander, L., 1996. The intranuclear site of excision of each intron in Balbiani ring 3 pre-mRNA is influenced by the time remaining to transcription termination and different excision efficiencies for the various introns. RNA 2, 641–651.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L.D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T.L., Miller, E., Bache, S.M., Müller, K., Ooms, J., Robinson, D., Seidel, D.P., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K., Yutani, H., 2019. Welcome to the tidyverse. J. Open Source Softw. 4, 1686. https://doi.org/10.21105/joss.01686

Wilke, C.O., 2019. cowplot: Streamlined Plot Theme and Plot Annotations for “ggplot2.”

Wilkinson, M.E., Charenton, C., Nagai, K., 2020. RNA Splicing by the Spliceosome. Annu. Rev. Biochem. 89, 359–388. https://doi.org/10.1146/annurev-biochem-091719-064225

Wissink, E.M., Vihervaara, A., Tippens, N.D., Lis, J.T., 2019. Nascent RNA analyses: tracking transcription and its regulation. Nat. Rev. Genet. 1–19. https://doi.org/10.1038/s41576-019-0159-6

Wu, J.Y., Maniatis, T., 1993. Specific interactions between proteins implicated in splice site selection and regulated alternative splicing. Cell 75, 1061–1070. https://doi.org/10.1016/0092-8674(93)90316-I

Wu, S., Romfo, C.M., Nilsen, T.W., Green, M.R., 1999. Functional recognition of the 3′ splice site AG by the splicing factor U2AF 35. Nature 402, 832–835. https://doi.org/10.1038/45590

Wuarin, J., Schibler, U., 1994. Physical isolation of nascent RNA chains transcribed by RNA polymerase II: evidence for cotranscriptional splicing. Mol. Cell. Biol. 14, 7219–7225. https://doi.org/10.1128/MCB.14.11.7219

Xiao, X., Wang, Z., Jang, M., Nutiu, R., Wang, E.T., Burge, C.B., 2009. Splice site strength–dependent activity and genetic buffering by poly-G runs. Nat. Struct. Mol. Biol. 16, 1094–1100. https://doi.org/10.1038/nsmb.1661

Xie, Y., Ren, Y., 2019. Mechanisms of nuclear mRNA export: A structural perspective. Traffic 20, 829–840. https://doi.org/10.1111/tra.12691

Xu, X.M., Rose, A., Muthuswamy, S., Jeong, S.Y., Venkatakrishnan, S., Zhao, Q., Meier, I., 2007. NUCLEAR PORE ANCHOR, the Arabidopsis homolog of Tpr/Mlp1/Mlp2/megator, is involved in mRNA export and SUMO homeostasis and affects diverse aspects of plant development. Plant Cell 19, 1537–1548. https://doi.org/10.1105/tpc.106.049239

Xu, Y.-Z., Query, C.C., 2007. Competition between the ATPase Prp5 and Branch Region-U2 snRNA Pairing Modulates the Fidelity of Spliceosome Assembly. Mol. Cell 28, 838–849. https://doi.org/10.1016/j.molcel.2007.09.022

Yan, C., Wan, R., Shi, Y., 2019. Molecular Mechanisms of pre-mRNA Splicing through Structural Biology of the Spliceosome. Cold Spring Harb. Perspect. Biol. 11. https://doi.org/10.1101/cshperspect.a032409

Yang, F., Wang, X.-Y., Zhang, Z.-M., Pu, J., Fan, Y.-J., Zhou, J., Query, C.C., Xu, Y.-Z., 2013. Splicing proofreading at 5’ splice sites by ATPase Prp28p. Nucleic Acids Res. 41, 4660–4670. https://doi.org/10.1093/nar/gkt149

Yang, X., Coulombe-Huntington, J., Kang, S., Sheynkman, G.M., Hao, T., Richardson, A., Sun, S., Yang, F., Shen, Y.A., Murray, R.R., Spirohn, K., Begg, B.E., Duran-Frigola, M., MacWilliams, A., Pevzner, S.J., Zhong, Q., Trigg, S.A., Tam, S., Ghamsari, L., Sahni, N., Yi, S., Rodriguez, M.D., Balcha, D., Tan, G., Costanzo, M., Andrews, B., Boone, C., Zhou, X.J., Salehi-Ashtiani, K., Charloteaux, B., Chen, A.A., Calderwood, M.A., Aloy, P., Roth, F.P., Hill, D.E., Iakoucheva, L.M., Xia, Y., Vidal, M., 2016. Widespread Expansion of Protein Interaction

Capabilities by Alternative Splicing. Cell 164, 805–817. https://doi.org/10.1016/j.cell.2016.01.029

Yee, B.A., Pratt, G.A., Graveley, B.R., Nostrand, E.L.V., Yeo, G.W., 2019. RBP-Maps enables robust generation of splicing regulatory maps. RNA 25, 193–204. https://doi.org/10.1261/rna.069237.118

Yeo, G., Burge, C.B., 2004. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. J. Comput. Biol. J. Comput. Mol. Cell Biol. 11, 377–394. https://doi.org/10.1089/1066527041410418

Ying, Y., Wang, X.-J., Vuong, C.K., Lin, C.-H., Damianov, A., Black, D.L., 2017. Splicing Activation by Rbfox Requires Self-Aggregation through Its Tyrosine-Rich Domain. Cell 170, 312-323.e10. https://doi.org/10.1016/j.cell.2017.06.022

Yoh, S.M., Lucas, J.S., Jones, K.A., 2008. The Iws1:Spt6:CTD complex controls cotranscriptional mRNA biosynthesis and HYPB/Setd2-mediated histone H3K36 methylation. Genes Dev. 22, 3422–3434. https://doi.org/10.1101/gad.1720008

Zander, G., Hackmann, A., Bender, L., Becker, D., Lingner, T., Salinas, G., Krebber, H., 2016. mRNA quality control is bypassed for immediate export of stress-responsive transcripts. Nature 540, 593–596. https://doi.org/10.1038/nature20572

Zeisel, A., Köstler, W.J., Molotski, N., Tsai, J.M., Krauthgamer, R., Jacob-Hirsch, J., Rechavi, G., Soen, Y., Jung, S., Yarden, Y., Domany, E., 2011. Coupled pre-mRNA and mRNA dynamics unveil operational strategies underlying transcriptional responses to stimuli. Mol. Syst. Biol. 7, 529. https://doi.org/10.1038/msb.2011.62

Zeng, C., Berget, S.M., 2000. Participation of the C-Terminal Domain of RNA Polymerase II in Exon Definition during Pre-mRNA Splicing. Mol. Cell. Biol. 20, 8290–8301.

Zenklusen, D., Vinciguerra, P., Wyss, J.-C., Stutz, F., 2002. Stable mRNP formation and export require cotranscriptional recruitment of the mRNA export factors Yra1p and Sub2p by Hpr1p. Mol. Cell. Biol. 22, 8241–8253. https://doi.org/10.1128/MCB.22.23.8241-8253.2002

Zhang, L., Vielle, A., Espinosa, S., Zhao, R., 2019. RNAs in the spliceosome: Insight from cryoEM structures. Wiley Interdiscip. Rev. RNA 10, e1523. https://doi.org/10.1002/wrna.1523

Zubradt, M., Gupta, P., Persad, S., Lambowitz, A.M., Weissman, J.S., Rouskin, S., 2017. DMS-MaPseq for genome-wide or targeted RNA structure probing in vivo. Nat. Methods 14, 75–82. https://doi.org/10.1038/nmeth.4057