IDENTIFICATION OF SPLICING PATHWAY MUTATIONS VIA TARGETED SEQUENCING

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Benjamin Jung Fair

August 2018

IDENTIFICATION OF SPLICING PATHWAY MUTATIONS VIA TARGETED SEQUENCING

Benjamin Jung Fair, Ph. D. Candidate

Cornell University 2018

The identification of splice sites and catalytic removal of introns via the spliceosome is an essential and regulated component of eukaryotic gene expression. While ever-increasing numbers of human genetic diseases can be linked to defects in the splicing pathway, our molecular understanding of how these mutations disrupt this complex process remains incomplete. To identify mutations which impact the splicing pathway I have developed a series of targeted-sequencing based quantitative genetic screens in *S. pombe*, a yeast species which is genetically tractable yet retains features of complex splicing patterns that have been lost in the *S. cerevisiae* lineage. Through these screens I have isolated conditional alleles of spliceosome components and investigated the mechanism by which some of these mutations disrupt splicing. Furthermore, I have identified genes in other nuclear processes, such as transcription and chromatin remodeling, which contribute to splicing outcomes. These findings suggest splicing is largely co-transcriptional and suggest specific genes by which transcription is functionally linked to splicing. Finally, I describe a novel approach for enriching RNA-sequencing libraries for splicing-informative reads. This technique allows for precise quantitation and discovery of rare splice products, including splice intermediates, which are poorly captured in other RNA-sequencing methods.

BIOGRAPHICAL SKETCH

Ben has always enjoyed the sciences from a young age. Biology and physics were his favorite subjects in secondary school – biology because he appreciates the rich diversity, color, and intricacy that the living world offers; and physics because he loved the epiphany moments, abstract thinking experiences, and creativity that came with solving the hardest problem sets. He attended Grand Valley State University in his home state of Michigan, originally following the academic path of a pre-medical student. As a junior at Grand Valley, he got a taste of the scientific process in a lab setting under the supervision of Dr. Martin Burg… Learn, think, design, generate data, learn, think, repeat… The epiphany moments, the abstract thinking, the creativity. He felt scientific process more stimulating and freeing than the observational and volunteer experiences he had accumulated at the hospital. He realized his passions were somewhat misaligned from many of his pre-med peers, who were passionate about their own noble cause. After some introspection, and discussions with a few of his undergraduate professors, he aspired to attend graduate school in the sciences rather than a career in medicine.

Ben was awarded the 'Most Outstanding Student in Biology' upon graduating from Grand Valley, before attending Cornell University for graduate school in molecular biology. There he worked in the lab of Jeff Pleiss, where he learned to combine molecular biology, genetics, and computational approaches to investigate pre-mRNA splicing biology. Although at the time of writing, he is still fuzzy on his precise career goals, he undoubtedly maintains a desire to learn, be creative, and problem solve. Upon receiving his PhD, Ben looks forward to expanding his scientific skills with a postdoctoral position at University of Chicago under the supervision of Dr. Yang Li and Dr. Yoav Gilad, where he will investigate human genetic variation as it relates to complex molecular phenotypes.

ACKNOWLEDGEMENTS

Firstly, I would like to express my gratitude to my advisor Dr. Jeffrey Pleiss for his continuous funding and support of my PhD research, as well as his and guidance towards my scientific development. I appreciate the working relationship we have developed, which has fostered my scientific independence while still providing insightful discussions and contributions to this work that I would have not have made alone.

I also thank my lab mates, some of whom are equal co-authors on parts of this work for their essential contributions. Each co-author's contributions, including my own, are noted at the end of each chapter. But for now, I want to specifically acknowledge the two co-authors who contributed the most. Amy Larson co-wrote much of Chapter 2, and we have worked together at the bench for many of the experiments described in Chapter 2 and Chapter 3. These chapters could not have been completed without the experiments she did, and it has been a pleasure collaborating with with her. Hansen Xu, an equal co-author on Chapter 4, developed the novel method (abbreviated as MPE-seq) and performed most experiments described in that chapter. My contributions to this work were merely in data analysis and writing, which is practice has been the less time-consuming portion of this project. I am thankful for the interesting collaboration we developed after his hard work towards developing the MPE-seq methodology through a long iterative process. The datasets he generated have provided the whole lab with a great supply of new and creative ideas for the future.

Finally, I am thankful to all those who have intellectually contributed to this and related work, whether through scientific discussions at lab meetings or over beers. This group includes my committee members - Haiyuan Yu, John Lis, and Paul Soloway – as well as Hojoong Kwak,

Andrew Grimson and his lab, and the fellow graduate students with whom I've most often discussed this work: Michael Gildea, Zach Dwyer, Madhura Raghavan, and Gregory Booth.

Table of Contents

Chapter 1: The power of fission: yeast as a tool for understanding complex splicing

Alternative citation for this chapter:

Fair BJ, Pleiss JA. (2017). The power of fission: yeast genetics as a tool for understand complex splicing. Current Genetics 63(3):375-380. doi: 10.1007/s00294-016-0647-6.

Abstract

Pre-mRNA splicing is an essential component of eukaryotic gene expression. Many metazoans, including humans, regulate alternative splicing patterns to generate expansions of their proteome from a limited number of genes. Importantly, a considerable fraction of human disease causing mutations manifest themselves through altering the sequences that shape the splicing patterns of genes. Thus, understanding the mechanistic bases of this complex pathway will be an essential component of combating these diseases. Dating almost to the initial discovery of splicing, researchers have taken advantage of the genetic tractability of budding yeast to identify the components and decipher the mechanisms of splicing. However, budding yeast lacks the complex splicing machinery and alternative splicing patterns most relevant to humans. More recently, many researchers have turned their efforts to studying the fission yeast, *Schizosaccharomyces pombe*, which has retained many features of complex splicing, including degenerate splice site sequences, the usage of exonic splicing enhancers, and SR-proteins. Here we review recent work using fission yeast genetics to examine pre-mRNA splicing, highlighting its promise for modeling the complex splicing seen in higher eukaryotes.

Main text

Eukaryotic genes often contain non-coding introns which must be spliced out of the pre-mRNA by the spliceosome in order to create a translatable mRNA. At its core, the spliceosome is composed of five highly conserved small nuclear RNA-protein complexes (snRNPs). Together with as many as 200 auxiliary proteins (Wahl *et al.* 2009; Will and Lührmann 2011b), the spliceosome assembles anew upon each intron in a highly dynamic process. The mechanisms by which the spliceosome faithfully assembles upon and activates the correct splice sites amongst the vast sequence space of the transcriptome is a complex question which will remain only partially understood for the foreseeable future. It is clear that the spliceosome requires guidance from many sources, including sequence elements within the pre-mRNA, the local chromatin environment, and the influence of auxiliary splicing components which assemble upon the pre-mRNA (Matlin *et al.* 2005; Lee and Rio 2015). Adding to the complexity and functional importance of splicing, the use of alternative splice sites in different cellular contexts can change the coding sequence of mRNAs, thus the alternative activation of splice sites in different cell types functions as a critical control point for regulating gene expression. Recent RNA-seq datasets suggest >95% of intron-containing human genes undergo alternative splicing (Barbosa-Morais *et al.* 2012; Bradley *et al.* 2012). Conversely, budding yeast (*S. cerevisiae*) has few examples of alternative splicing (Juneau *et al.* 2009; Marshall *et al.* 2013; Kawashima *et al.* 2014), most of which reflect intron-retention, and few of which are exon-skipping events, the type most prevalent in humans. Accordingly, the budding yeast contains a relatively reduced splicing apparatus, absent of many auxiliary components that control alternative splicing in humans (Käufer and Potashkin 2000). Nevertheless, the identification of many conserved core spliceosome components was accomplished through the use of the genetically tractable yeast

model system. For example, the PRP genes were identified by screening libraries of temperature sensitive yeast strains for pre-mRNA processing (PRP) mutant phenotypes (Lustig *et al.* 1986; Vijayraghavan *et al.* 1989; Noble and Guthrie 1996). Further biochemical studies using purified proteins from these mutants in *in vitro* splicing reactions have played a pivotal role in delineating the mechanisms of splicing catalysis (Meyer and Vilardell 2009; Hossain and Johnson 2014a).

To more fully understand the complexity of splicing in higher eukaryotes, our lab and others have turned to the fission yeast, *Schizosaccharomyces pombe*, a distantly related ascomycete yeast which is similarly tractable to budding yeast as a model organism, yet in many ways is more similar to humans from a splicing perspective. Some of these similarities become clear by comparing the genes and introns present in the genomes of budding yeast, fission yeast, and humans. Fission yeast genes are intron-dense compared to budding yeast (average of 0.9 introns per gene compared to 0.05 in budding yeast), though not as intron-dense as human genes (average of 8 introns per gene). Many fission yeast genes contain two or more introns, a pre-requisite for exon-skipping, some of which are interrupted by extremely short microexons reminiscent of those found in human genes (Scheckel and Darnell 2015). Moreover, a high degree of degeneracy is seen in the fission yeast splice site sequences that direct the spliceosome, more closely reflecting those seen in human transcripts (Fig. 1). The fission yeast genome also contains many orthologs of human splicing proteins which are not present in budding yeast (Käufer and Potashkin 2000; Kuhn and Käufer 2003), including two *bona fide* orthologs of human SR proteins, a class of auxiliary splicing proteins thought be master regulators of alternative splicing in plants and animals. Additionally, two recent publications (Bitton *et al.* 2015; Stepankiw *et al.* 2015) each concluded that the fission yeast spliceosome makes use of a previously unappreciated number of alternative splice sites. Both of these studies estimate that

about 2-3% of splicing events genome-wide contain an alternative (unannotated) splice site, though the alternative mRNA isoforms are generally unstable and only detectable in RNA decay mutants. While this is a higher usage of alternative splice sites than the <1% observed in similar RNA decay mutants in budding yeast (Kawashima *et al.* 2014), it pales in comparison to the extent of alternative splicing in humans.

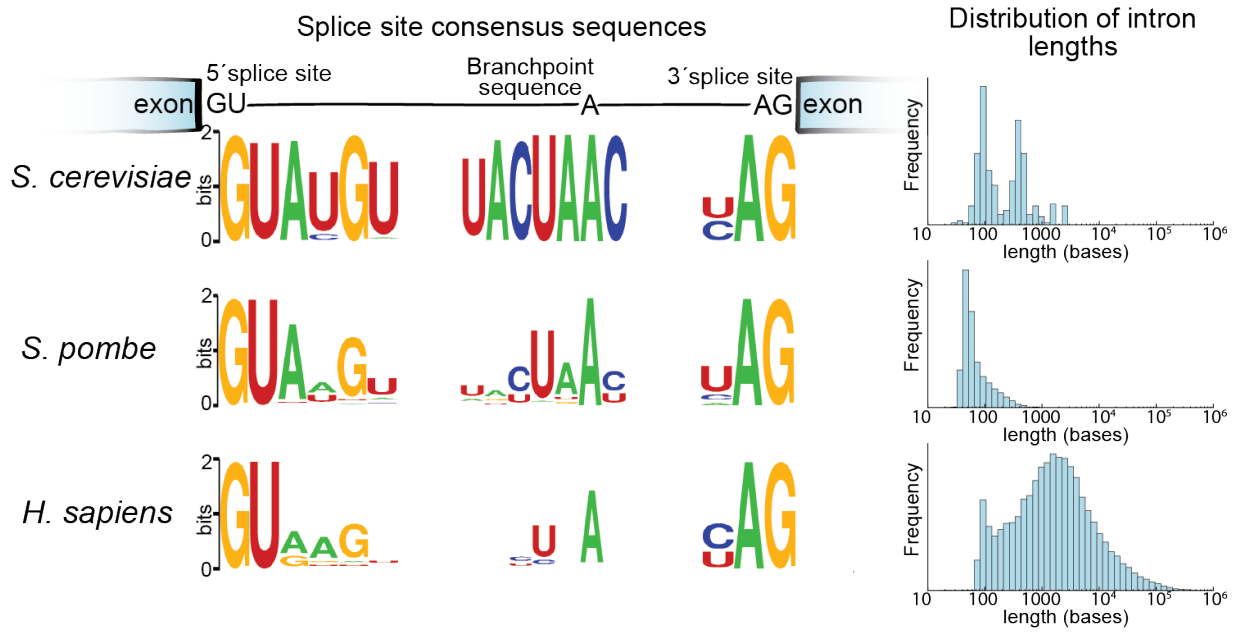


Figure 1. Comparison of intron features in yeasts and humans. A. Sequence logos depict the consensus sequence of splice site sequences. The total height at each position is proportional to the conservation at that position (Crooks *et al*. 2004). B. The distribution of intron lengths for each species is shown as a histogram. Intron lengths and splice site sequences were obtained from (James Kent *et al*. 2002; Wilhelm *et al*. 2008). Branchpoint sequences were obtained from (Clark *et al*. 2002; Gao *et al*. 2008; Stepankiw *et al*. 2015).

One insight as to why there isn't more alternative splicing in fission yeast is that the yeast spliceosomes are constrained to select splice sites within a relatively short distance of each other. The natural distribution of intron lengths in both yeasts, which favor shorter introns than humans, supports this notion (Fig. 1). The prevalence of short introns in the yeast genome but long introns in humans has led to consideration of two often juxtaposed models for the initial identification of splice site pairs: the intron-definition model and exon-definition model (Berget 1995; De Conti

et al. 2013). Importantly, the intron-definition and exon-definition models are not mutually exclusive means to explain spliceosome assembly, rather both may impact assembly upon the same splice sites. In the intron-definition model, which is certainly prevalent in yeasts (Romfo *et al.* 2000), splice sites are identified by pairing a 5′ splice site to the nearest downstream branchpoint and 3′ splice site. Because the splice sites are initially paired across the intron, the intron length is under selection to remain short to maintain these interactions. In the exon-definition model, splice sites are initially identified through interactions between a 5′ splice and an upstream 3′ splice site, across the exon (Fig. 2). Thus exons are under selection to remain short while introns may be long. These initial cross-exon interactions later need to be exchanged for cross-intron interactions for proper splicing catalysis. The molecular mechanisms by which the requisite cross-intron interactions are formed after the initial cross-exon interactions are established remains a critical, yet unanswered question in the field (De Conti *et al.* 2013; Hollander *et al.* 2016).

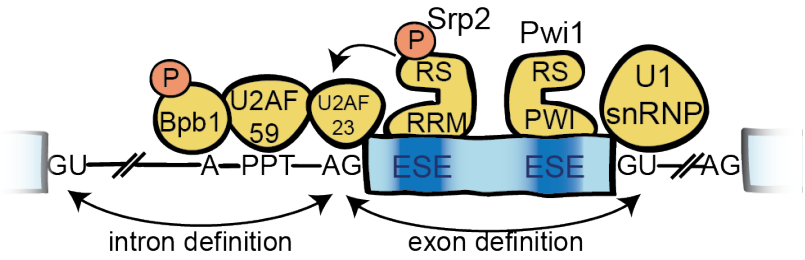


Figure 2. Mammalian-like mechanisms of splice site definition in fission yeast. Exonic-splicing-enhancer (ESE) sequences recruit SR and SR-related proteins, which recruit or stabilize early spliceosome assembly. Failure of these components to assemble across the exon may result in exon-skipping. Mutations in Bpb1, U2AF-59, and U2AF-23 result in exon-skipping, which can be suppressed by overexpression of Srp2. Naturally skipped exons have particularly non-consensus downstream 5' splice sites, suggesting cross-exon interactions may contribute to recognition of upstream 3' splice sites. Phosphorylation of Bpb1 is necessary for efficient splicing of introns with non-consensus branchpoint sequences. Pwi1 is an SR-related protein which is required for splicing of introns with non-consensus 5' splice sites in fission yeast. Though the PWI domain can bind RNA, it is not clear whether it directly binds the ESE during ESE-dependent splicing.

In higher eukaryotes, it is clear that initial establishment of cross-exon interactions is facilitated by members of the SR family of proteins. These modular proteins bind to exonic splicing enhancer (ESE) sequences on the pre-mRNA through their RNA-recognition motif (RRM) and recruit the spliceosome to nearby splice sites through their RS domain (Graveley 2000; Long and Caceres 2009). The RS domain is rich in arginine-serine dipeptides and the phosphorylation status of the serine residues is important for activating splicing. Although ESE-dependent splicing does not necessarily require cross-exon interactions, it is interesting to note that when mammalian-derived purine-rich ESE sequences were placed into fission yeast exons, they still functioned to recruit the yeast SR protein SRP2 to aid in identification of weak upstream 3′ splice sites (Webb *et al.* 2005a). RNA sequencing experiments have identified low-frequency alternative splicing events in the form on exon-skipping in fission yeast, the predominant form of alternative splicing predicted by the exon-definition model (Awan *et al.* 2013; Bitton *et al.* 2015; Stepankiw *et al.* 2015). Analysis of the splice site sequences surrounding the skipped exons revealed relatively consensus upstream branchpoint sequences, but weak downstream 5′ splice sites (Stepankiw *et al.* 2015), reminiscent of the observation that placement of a 5′ splice site sequence enhances recognition of upstream branchpoints in

mammals (Robberson *et al.* 1990). Some of these alternative splicing events are particularly sensitive to environmental cues, suggesting their splicing may be regulated.

Together these results suggest that some elements of exon definition existed in the ancestral spliceosome (Ram and Ast 2007b), and that these mechanisms persist to some extent in fission yeast, wherein SRP2 can function as a regulator of ESE-dependent splicing. Whereas loss-of-function studies on SR and SR-related proteins in humans are hampered by the high level of functional redundancy between the many SR proteins present in vertebrates, often resulting in relatively weak *in vivo* phenotypes (Pandit *et al.* 2013), this level of redundancy is not present in fission yeast. Yeast researchers are now in a strong position to tackle important questions about mammalian splicing mechanisms. What factors impact alternative splice site choice and early spliceosome assembly? How do SR proteins function, what contacts do they make to recruit the spliceosome? How do other nuclear processes such as transcription mechanistically affect the splicing process?

Several recent publications have made progress towards these goals. Haraguchi and co-workers used a reporter construct linked to a selectable auxotrophic marker in a classic forward genetic screen to identify factors which, when mutated, promote the production of exon-skipped splicing isoforms (Haraguchi *et al.* 2007; Sasaki-Haraguchi *et al.* 2015). This screen yielded point mutations in the essential proteins Bpb1, U2AF-59, and U2AF-23 (orthologs of the human SF1, U2AF-65, and U2AF-35 proteins, respectively), which bind to the branchpoint, polypyrimidine tract (PPT), and 3′ splice site respectively during early spliceosome assembly. The mechanisms by which these proteins are influenced by other auxiliary splicing proteins to selectively bind to *bona fide* branchpoint/3′splice sites is fundamental to alternative splice site choice (Shao *et al.* 2014). In humans, U2AF-35 mediates protein-protein interactions involved in

ESE-dependent splicing (Zuo and Maniatis 1996). Fission yeast U2AF-23 physically interacts with Srp2 (Webb and Wise 2004). Interestingly, the phenotypes of all three of the mutations identified by Haraguchi et al could be suppressed by overexpression of Srp2, consistent with a model wherein Srp2 contacts U2AF-23 to recruit Bpb1/U2AF-59/23 to the branchpoint/3′splice site (Fig. 2). Each of these proteins contain arginine-serine-rich regions with reasonable homology to the RS-domains found in SR proteins (Mazroui *et al.* 1999; Graveley 2000; Kielkopf *et al.* 2001). Lipp and co-workers offered additional insights into the function of the RS domains, first by identifying relevant phosphorylation sites by mass spectrometry, then by making serine to alanine mutations at those sites (Lipp *et al.* 2015). Using splicing-sensitive microarrays to probe splicing genome-wide, they observed a moderate defect when the Srp2 phosphorylation sites were mutated: wherein some introns were spliced less efficiently, others were spliced more efficiently. A much more striking phenotype was observed when they mutated sites in the RS-like domain of Bpb1. The dependence of Bpb1 on splicing was relieved when a non-consensus branchpoint sequence was mutated to a more optimal sequence, suggesting phosphorylation is functionally important to recruiting or activating Bpb1/SF1 at suboptimal branchpoints (Fig. 2).

More general high throughput genetic approaches have also revealed information valuable to the splicing community. For example, Vo and co-workers recently published a proteome-wide map of protein-protein interactions using arrays of yeast-two-hybrid strains (Vo *et al.* 2016). This map revealed a physical interaction network between proteins containing RS domains, specifically between Pwi1, Srp1, and Srp2 (orthologs of the human SRM160, SRSF2, and SRSF4 proteins, respectively). Although the yeast-two-hybrid interactions were performed with fission yeast proteins in the cellular context of budding yeast and may not have functional

relevance during splicing, these results reinforce the hypothesis that RS domains primarily interact with other RS domains. An interaction between Srp2 and Sap14 (a component of the U2 snRNP, and orthologous to human SF3b2), was also identified, further strengthening the argument for Srp2 assisting in early spliceosome assembly at the branchpoint. Patrick and co-workers took a complementary approach, using an epistatic miniarray profile (E-MAP) to identify loss-of-function mutations which genetically interact with splicing factor mutations (Patrick *et al.* 2015). Strong negative genetic interactions were found between the SWI/SNF chromatin remodeling complex and U2 snRNP components such as *sap145* (SF3b145 in humans). Additional ChIP experiments showed that SWI/SNF recruits (directly or indirectly) Cdc28 (orthologous to Prp2p in budding yeast, DHX16 in humans), the helicase responsible for displacing SF3 from the branchpoint prior to the first catalytic step of splicing. These results highlight the interconnections between spliceosome assembly and other nuclear processes, such as nucleosome remodeling, during co-transcriptional splicing.

Recently, we published the results of a quantitative screen for mutant strains which display splicing defects (Larson *et al.* 2016, Chapter2 of this dissertation). By using a sequencing-based approach, we were able to precisely and systematically measure the splicing efficiency of two endogenous transcripts in the background of ~3000 strains, each containing a deletion of a single, non-essential *S. pombe* gene. In addition to identifying known splicing factors, we found that various chromatin remodeling factors, including but not limited to SWI/SNF, resulted in as much as four-fold higher levels of unspliced pre-mRNA when deleted. Among the notable factors identified through this screen is the SR-related gene *pwi1* (SRM160 in humans). This protein contains a canonical RS-domain but has an atypical RNA-binding domain that contains a conserved PWI amino acid motif (Szymczyna *et al.* 2003). Mutation of

this factor resulted in a strong genome-wide splicing defect, wherein introns with weak 5′ splice sites were particularly affected (Vo *et al.* 2016). This suggests that *pwi1* may function similarly to its human ortholog by participating in ESE-dependent splicing that requires interactions with the U1 snRNP (Fig. 2) (Eldridge *et al.* 1999; Blencowe *et al.* 2000).

Moving past understanding splice site selection, the recent publications of atomic-resolution spliceosome structures in fission and budding yeast (Nguyen *et al.* 2015, 2016; Yan *et al.* 2015) will motivate enlightening experiments towards better understanding splicing catalysis. Much of our current understanding of splicing catalysis comes from biochemical approaches using mammalian or budding yeast extracts to purify splicing intermediates or reconstitute splicing reactions *in vitro*. Unfortunately, the lack of a successful *in vitro* splicing system has hampered biochemical studies in fission yeast. Hopefully the advancement of structural knowledge will help researchers understand why fission yeast splicing cannot be reconstituted *in vitro* and possibly lead to a solution, enabling additional mechanistic insights. For the time being, however, structure-function mapping experiments in fission yeast must be done using *in vivo* methods for assessing splicing-related phenotypes of mutations. The identification and characterization of splicing factor alleles containing discrete point mutations will be key in these types of experiments. We have recently used our quantitative screening approach to screen through a library of thousands of temperature sensitive strains to identify alleles with major *in vivo* splicing defects at the restrictive temperature (Chapter3 of this dissertation). This quantitative screening assay could also be used to assess thousands of variants within a protein of interest. This strategy, often referred to as saturating mutagenesis or deep mutational scanning (Araya and Fowler 2011; Fowler and Fields 2014; Gao *et al.* 2014), has not yet been applied to study the function of spliceosome components and could yield unanticipated findings. Surely the

improvement of genetic technologies, including but certainly not limited to CRISPR, will allow the splicing field to perform previously difficult genetic experiments without the need for a model organism. Perhaps with a cleverly designed experiment, deep mutational scanning experiments could be performed in human cells. However, if one were interested in deep mutational scanning analysis of an SR protein, fission yeast may still be the ideal organism to avoid the redundancy of the plethora of SR proteins present in humans. In an experimentally useful way, the fission yeast spliceosome sits at an intermediate point between the ultra-reductionist budding yeast spliceosome and the ultra-complex human spliceosome. So despite the coming of age for effective genetic manipulations in human cells, for many reasons, the splicing community will still benefit from using fission yeast as a stepping stone to decipher the human spliceosome's complexity.

Acknowledgements and Author Contributions

Authors wish to thank Tokio Tani and members of the Pleiss lab for helpful comments on this manuscript. This work was funded by National Institutes of Health grant GM098634. BJF wrote manuscript draft. JAP and BJF contributed to draft revisions.

Works Cited

Alahari, S. K., H. Schmidt, and N. F. Käufer, 1993 The fission yeast prp4+ gene involved in pre-mRNA splicing codes for a predicted serine/threonine kinase and is essential for growth. Nucleic Acids Res. 21: 4079–4083.

Albulescu, L. O., N. Sabet, M. Gudipati, N. Stepankiw, Z. J. Bergman *et al.*, 2012 A quantitative, high-throughput reverse genetic screen reveals novel connections between pre-mRNA splicing and 5′ and 3′ end transcript determinants. PLoS Genet. 8: e1002530.

Alexander, R. D., S. a. Innocente, J. D. Barrass, and J. D. Beggs, 2010 Splicing-Dependent RNA polymerase pausing in yeast. Mol. Cell 40: 582–593.

Ameur, A., A. Zaghlool, J. Halvardson, A. Wetterbom, U. Gyllensten *et al.*, 2011 Total RNA sequencing reveals nascent transcription and widespread co-transcriptional splicing in the human brain. Nat. Struct. Mol. Biol. 18: 1435–1440.

Andersson, R., S. Enroth, A. Rada-Iglesias, C. Wadelius, and J. Komorowski, 2009 Nucleosomes are well positioned in exons and carry characteristic histone modifications. Genome Res. 19: 1732–41.

Anver, S., A. Roguev, M. Zofall, N. J. Krogan, S. I. S. Grewal *et al.*, 2014 Yeast X-chromosome-associated protein 5 (Xap5) functions with H2A.Z to suppress aberrant transcripts. EMBO Rep. 15: 894–902.

Araya, C. L., and D. M. Fowler, 2011 Deep mutational scanning: Assessing protein function on a massive scale. Trends Biotechnol. 29: 435–442.

Armstrong, J., N. Bone, J. Dodgson, and T. Beck, 2007 The role and aims of the FYSSION project. Briefings Funct. Genomics Proteomics 6: 3–7.

Awan, A. R., A. Manfredo, and J. a Pleiss, 2013 Lariat sequencing in a unicellular yeast

identifies regulated alternative splicing of exons that are evolutionarily conserved with humans. Proc. Natl. Acad. Sci. U. S. A. 110: 12762–7.

Baejen, C., P. Torkler, S. Gressel, K. Essig, J. Söding *et al.*, 2014 Transcriptome maps of mRNP biogenesis factors define pre-mRNA recognition. Mol Cell 55: 745–757.

Bansal, V., 2010 A statistical method for the detection of variants from next-generation resequencing of DNA pools. Bioinformatics 26:.

Barbosa-Morais, N. L., M. Irimia, Q. Pan, H. Y. Xiong, S. Gueroussov *et al.*, 2012 The evolutionary landscape of alternative splicing in vertebrate species. Science (80-.). 338: 1587–1593.

Bayne, E. H., D. a Bijos, S. a White, F. de Lima Alves, J. Rappsilber *et al.*, 2014 A systematic genetic screen identifies new factors influencing centromeric heterochromatin integrity in fission yeast. Genome Biol. 15: 481.

Bayne, E. H., M. Portoso, A. Kagansky, I. C. Kos-Braun, T. Urano *et al.*, 2008 Splicing factors facilitate RNAi-directed silencing in fission yeast. Science 322: 602–606.

Berget, S. M., 1995 Minireviews : Exon Recognition in Vertebrate Splicing. 2411–2414.

Bitton, D. A., S. R. Atkinson, C. Rallis, G. C. Smith, D. A. Ellis *et al.*, 2015 Widespread exon-skipping triggers degradation by nuclear RNA surveillance in fission yeast. Genome Res. 884–896.

Blencowe, B. J., G. Baurén, A. G. Eldridge, R. Issner, J. A. Nickerson *et al.*, 2000 The SRm160/300 splicing coactivator subunits. RNA 6: 111–20.

Blomquist, T. M., E. L. Crawford, J. L. Lovett, J. Yeo, L. M. Stanoszek *et al.*, 2013 Targeted RNA-Sequencing with Competitive Multiplex-PCR Amplicon Libraries. PLoS One 8: e79120.

Boon, K.-L., T. Auchynnikava, G. Edwalds-Gilbert, J. D. Barrass, A. P. Droop *et al.*, 2006 Yeast Ntr1/Spp382 Mediates Prp43 Function in Postspliceosomes. Mol. Cell. Biol.

Booth, G. T., I. X. Wang, V. G. Cheung, and J. T. Lis, 2016 Divergence of a conserved elongation factor and transcription regulation in budding and fission yeast. Genome Res. 26: 799–811.

Bradley, R. K., J. Merkin, N. J. Lambert, and C. B. Burge, 2012 Alternative splicing of RNA triplets is often regulated and accelerates proteome evolution. PLoS Biol. 10: e1001229.

Burke, J., A. Longhurst, D. Merkurjev, J. Sales-Lee, B. Rao *et al.*, 2018 Spliceosome profiling visualizes the operations of a dynamic RNP in vivo at nucleotide resolution. Cell 173: 1014–1030.

Carey, M. F., C. L. Peterson, and S. T. Smale, 2013 The primer extension assay. Cold Spring Harb. Protoc. 8: 164–173.

Carrillo Oesterreich, F., S. Preibisch, and K. M. Neugebauer, 2010 Global analysis of nascent rna reveals transcriptional pausing in terminal exons. Mol. Cell 40: 571–581.

Cazzola, M., M. Rossi, and L. Malcovati, 2013 Biologic and clinical significance of somatic mutations of SF3B1 in myeloid and lymphoid neoplasms. Blood.

Chanfreau, G., P. Legrain, B. Dujon, and A. Jacquier, 1994 Interaction between the first and last nucleotides of pre-mRNA introns is a determinant of 3′ splice site selection in S.cerevisiae. Nucleic Acids Res.

Chen, W., J. Moore, H. Ozadam, H. P. Shulha, N. Rhind *et al.*, 2018 Transcriptome-wide interrogation of the functional intronome by spliceosome profiling. Cell 173: 1031–1044.

Chen, W., H. P. Shulha, A. Ashar-Patel, J. Yan, K. M. Green *et al.*, 2014 Endogenous U2·U5·U6 snRNA complexes in S. pombe are intron lariat spliceosomes. RNA 20: 308–20.

Ciurciu, A., L. Duncalf, V. Jonchere, N. Lansdale, O. Vasieva *et al.*, 2013 PNUTS/PP1 regulates RNAPII-mediated gene expression and is necessary for developmental growth. PLoS Genet. 9: e1003885.

Clark, T. a, C. W. Sugnet, and M. Ares, 2002 Genomewide analysis of mRNA processing in yeast using splicing-specific microarrays. Science 296: 907–910.

Collart, M. A., and S. Oliviero, 2001 Preparation of Yeast RNA, in *Current Protocols in Molecular Biology*,.

Conesa, A., P. Madrigal, S. Tarazona, D. Gomez-Cabrero, A. Cervera *et al.*, 2016 A survey of best practices for RNA-seq data analysis. Genome Biol. 17:.

De Conti, L., M. Baralle, and E. Buratti, 2013 Exon and intron definition in pre-mRNA splicing. Wiley Interdiscip. Rev. RNA 4: 49–60.

Coombes, C. E., and J. D. Boeke, 2005 An evaluation of detection methods for large lariat RNAs. RNA 11: 323–31.

Core, L. J., and J. T. Lis, 2008 Transcription regulation through promoter-proximal pausing of RNA polymerase II. Science 319: 1791–1792.

Crooks, G. E., G. Hon, J. M. Chandonia, and S. E. Brenner, 2004 WebLogo: A sequence logo generator. Genome Res. 14: 1188–1190.

David, C. J., and J. L. Manley, 2011 The RNA polymerase C-terminal domain: A new role in spliceosome assembly. Transcription 2: 221–225.

Depristo, M. A., E. Banks, R. Poplin, K. V. Garimella, J. R. Maguire *et al.*, 2011 A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat. Genet. 43: 491–501.

Dietrich, R. C., R. Incorvaia, and R. A. Padgett, 1997 Terminal intron dinucleotide sequences do

not distinguish between U2- and U12-dependent introns. Mol. Cell.

Dobin, A., C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski *et al.*, 2013 STAR: Ultrafast universal RNA-seq aligner. Bioinformatics 29: 15–21.

Dujardin, G., C. Lafaille, M. de la Mata, L. E. Marasco, M. J. Muñoz *et al.*, 2014 How Slow RNA Polymerase II Elongation Favors Alternative Exon Skipping. Mol. Cell 54: 683–690.

Dunn, E. A., and S. D. Rader, 2014 Preparation of yeast whole cell splicing extract. Methods Mol. Biol.

Egan, E. D., C. R. Braun, S. P. Gygi, and D. Moazed, 2014 Post-transcriptional regulation of meiotic genes by a nuclear RNA silencing complex. RNA 20: 867–81.

Eldridge, A. G., Y. Li, P. A. Sharp, and B. J. Blencowe, 1999 The SRm160/300 splicing coactivator is required for exon-enhancer function. Proc. Natl. Acad. Sci. U. S. A. 96: 6125–6130.

Engel, S. R., F. S. Dietrich, D. G. Fisk, G. Binkley, R. Balakrishnan *et al.*, 2014 The reference genome sequence of Saccharomyces cerevisiae: then and now. G3 (Bethesda). 4: 389–98.

Eser, P., L. Wachutka, K. C. Maier, C. Demel, M. Boroni *et al.*, 2016 Determinants of RNA metabolism in the Schizosaccharomyces pombe genome. Mol. Syst. Biol. 12: 857–857.

Evsyukova, I., S. S. Bradrick, S. G. Gregory, and M. a Garcia-Blanco, 2013 Cleavage and polyadenylation specificity factor 1 (CPSF1) regulates alternative splicing of interleukin 7 receptor (IL7R) exon 6. RNA 19: 103–15.

Fair, B. J., and J. A. Pleiss, 2017 The power of fission: yeast as a tool for understanding complex splicing. Curr. Genet.

Fetzer, S., J. Lauber, C. L. Will, and R. Lührmann, 1997 The [U4/U6.U5] tri-snRNP-specific 27K protein is a novel SR protein that can be phosphorylated by the snRNP-associated

protein kinase. RNA 3: 344–55.

Forsburg, S. L., and N. Rhind, 2006 Basic methods for fission yeast. Yeast 23: 173–183.

Fowler, D. M., and S. Fields, 2014 Deep mutational scanning: a new style of protein science. Nat. Methods 11: 801–7.

Galej, W. P., T. H. D. Nguyen, A. J. Newman, and K. Nagai, 2014 Structural studies of the spliceosome: Zooming into the heart of the machine. Curr. Opin. Struct. Biol.

Galej, W. P., C. Oubridge, A. J. Newman, and K. Nagai, 2013 Crystal structure of Prp8 reveals active site cavity of the spliceosome. Nature 493: 638–43.

Gao, J., F. Kan, J. L. Wagnon, A. J. Storey, R. U. Protacio *et al.*, 2014 Rapid, efficient and precise allele replacement in the fission yeast Schizosaccharomyces pombe. Curr. Genet. 60: 109–119.

Gao, K., A. Masuda, T. Matsuura, and K. Ohno, 2008 Human branch point consensus sequence is yUnAy. Nucleic Acids Res. 36: 2257–2267.

Gottschalk, A., B. Kastner, R. Lührmann, and P. Fabrizio The yeast U5 snRNP coisolated with the U1 snRNP has an unexpected protein composition and includes the splicing factor Aar2p.

Grate, L., and M. Ares, 2002 Searching yeast intron data at Ares lab web site. Methods Enzymol. 350: 380–392.

Graveley, B. R., 2000 Sorting out the complexity of SR protein functions. RNA 6: 1197–1211.

Grewal, S. I. S., 2010 RNAi-dependent formation of heterochromatin and its diverse functions. Curr. Opin. Genet. Dev. 20: 134–141.

Habara, Y., S. Urushiyama, T. Tani, and Y. Ohshima, 1998 The fission yeast prp10(+) gene involved in pre-mRNA splicing encodes a homologue of highly conserved splicing factor,

SAP155. Nucleic Acids Res. 26: 5662–5669.

Haraguchi, N., T. Andoh, D. Frendewey, and T. Tani, 2007 Mutations in the SF1-U2AF59-U2AF23 complex cause exon skipping in Schizosaccharomyces pombe. J. Biol. Chem. 282: 2221–2228.

Hartwell, L. H., C. S. McLaughlin, and J. R. Warner, 1970 Identification of ten genes that control ribosome formation in yeast. MGG Mol. Gen. Genet. 109: 42–56.

Hicks, M. J., B. J. Lam, and K. J. Hertel, 2005 Analyzing mechanisms of alternative pre-mRNA splicing using in vitro splicing assays. Methods.

Hollander, D., S. Naftelberg, G. Lev-Maor, A. R. Kornblihtt, and G. Ast, 2016 How Are Short Exons Flanked by Long Introns Defined and Committed to Splicing? Trends Genet. xx: 1–11.

Hossain, M. A., and T. L. Johnson, 2014a Spliceosomal Pre-mRNA Splicing (K. J. Hertel, Ed.). 1126: 285–298.

Hossain, M. A., and T. L. Johnson, 2014b Using yeast genetics to study splicing mechanisms. Methods Mol. Biol.

Hou, H., Y. Wang, S. P. Kallgren, J. Thompson, J. R. Yates *et al.*, 2010 Histone variant H2A.Z regulates centromere silencing and chromosome segregation in fission yeast. J. Biol. Chem. 285: 1909–1918.

Huang, T., J. Vilardell, and C. C. Query, 2002 Pre-spliceosome formation in S.pombe requires a stable complex of SF1-U2AF59-U2AF23. EMBO J. 21: 5516–5526.

Iannone, C., A. Pohl, P. Papasaikas, D. Soronellas, G. P. Vicent *et al.*, 2015 Relationship between nucleosome positioning and progesterone-induced alternative splicing in breast cancer cells. RNA 21: 360–374.

Inada, M., and J. a Pleiss, 2010 *Genome-wide approaches to monitor pre-mRNA splicing.* Elsevier Inc.

James Kent, W., C. W. Sugnet, T. S. Furey, K. M. Roskin, T. H. Pringle *et al.*, 2002 The human genome browser at UCSC. Genome Res. 12: 996–1006.

Jamieson, D. J., B. Rahe, J. Pringle, and J. D. Beggs, 1991 A suppressor of a yeast splicing mutation (prp8-1) encodes a putative ATP-dependent RNA helicase. Nature.

Juneau, K., C. Nislow, and R. W. Davis, 2009 Alternative splicing of PTC7 in Saccharomyces cerevisiae determines protein localization. Genetics 183: 185–194.

Kallgren, S. P., S. Andrews, X. Tadeo, H. Hou, J. J. Moresco *et al.*, 2014 The Proper Splicing of RNAi Factors Is Critical for Pericentric Heterochromatin Assembly in Fission Yeast. PLoS Genet. 10: 1–11.

Käufer, N. F., and J. Potashkin, 2000 Analysis of the splicing machinery in fission yeast: a comparison with budding yeast and mammals. Nucleic Acids Res. 28: 3003–3010.

Kawashima, T., S. Douglass, J. Gabunilas, M. Pellegrini, and G. F. Chanfreau, 2014 Widespread Use of Non-productive Alternative Splice Sites in Saccharomyces cerevisiae. PLoS Genet. 10:.

Khodor, Y. L., J. Rodriguez, K. C. Abruzzi, C. H. A. Tang, M. T. Marr *et al.*, 2011 Nascent-seq indicates widespread cotranscriptional pre-mRNA splicing in Drosophila. Genes Dev. 25: 2502–2512.

Kielkopf, C. L., N. A. Rodionova, M. R. Green, and S. K. Burley, 2001 A novel peptide recognition mode revealed by the X-ray structure of a core U2AF35/U2AF65 heterodimer. Cell 106: 595–605.

Kim, D.-U., J. Hayles, D. Kim, V. Wood, H.-O. Park *et al.*, 2010 Analysis of a genome-wide set

of gene deletions in the fission yeast Schizosaccharomyces pombe. Nat. Biotechnol. 28: 617–623.

Kim, D., B. Langmead, and S. L. Salzberg, 2015 HISAT: A fast spliced aligner with low memory requirements. Nat. Methods 12: 357–360.

Kim, S. H., and R. J. Lin, 1993 Pre-mRNA splicing within an assembled yeast spliceosome requires an RNA-dependent ATPase and ATP hydrolysis. Proc. Natl. Acad. Sci. U. S. A.

Kim, S. H., and R. J. Lin, 1996 Spliceosome activation by PRP2 ATPase prior to the first transesterification reaction of pre-mRNA splicing. Mol. Cell. Biol. 16: 6810–6819.

Kivioja, T., A. Vähärautio, K. Karlsson, M. Bonke, M. Enge *et al.*, 2012 Counting absolute numbers of molecules using unique molecular identifiers. Nat. Methods 9: 72–74.

Kuhn, A. N., and N. F. Käufer, 2003 Pre-mRNA splicing in Schizosaccharomyces pombe: regulatory role of a kinase conserved from fission yeast to mammals. Curr. Genet. 42: 241–251.

Kupfer, D. M., S. D. Drabenstot, K. L. Buchanan, H. Lai, H. Zhu *et al.*, 2004 Introns and splicing elements of five diverse fungi. Eukaryot. Cell 3: 1088–1100.

Kyburz, A., A. Friedlein, H. Langen, and W. Keller, 2006 Direct Interactions between Subunits of CPSF and the U2 snRNP Contribute to the Coupling of Pre-mRNA 3??? End Processing and Splicing. Mol. Cell 23: 195–205.

de la Mata, M., C. R. Alonso, S. Kadener, J. P. Fededa, M. Blaustein *et al.*, 2003 A Slow RNA Polymerase II Affects Alternative Splicing In Vivo. Mol. Cell 12: 525–532.

Larson, A., B. J. Fair, and J. A. Pleiss, 2016 Interconnections Between RNA-Processing Pathways Revealed by a Sequencing-Based Genetic Screen for Pre-mRNA Splicing Mutants in Fission Yeast. G3 (Bethesda). 6: 1513–23.

Lee, N. N., V. R. Chalamcharla, F. Reyes-Turcu, S. Mehta, M. Zofall *et al.*, 2013 Mtr4-like protein coordinates nuclear RNA processing for heterochromatin assembly and for telomere maintenance. Cell 155: 1–14.

Lee, Y., and D. C. Rio, 2015 Mechanisms and Regulation of Alternative Pre-mRNA Splicing. Annu. Rev. Biochem. 1–33.

Li, B., M. Carey, and J. L. Workman, 2007 The Role of Chromatin during Transcription. Cell 128: 707–719.

Li, H., and R. Durbin, 2009 Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25: 1754–60.

Libri, D., N. Graziani, C. Saguez, and J. Boulay, 2001 Multiple roles for the yeast SUB2/yUAP56 gene in splicing. Genes Dev.

Lim, L. P., and C. B. Burge, 2001 A computational analysis of sequence features involved in recognition of short introns. Proc. Natl. Acad. Sci. U. S. A. 98: 11193–8.

Lin, R. J., A. J. Lustig, and J. Abelson, 1987 Splicing of yeast nuclear pre-mRNA in vitro requires a functional 40S spliceosome and several extrinsic factors. Genes Dev.

Lin, P. C., and R. M. Xu, 2012 Structure and assembly of the SF3a splicing factor complex of U2 snRNP. EMBO J.

Lipp, J. J., M. C. Marvin, K. M. Shokat, and C. Guthrie, 2015 SR protein kinases promote splicing of nonconsensus introns. Nat. Struct. Mol. Biol. 22: 611–617.

Liu, H.-L., and S.-C. Cheng, 2012 The Interaction of Prp2 with a Defined Region of the Intron Is Required for the First Splicing Reaction. Mol. Cell. Biol.

Long, J. C., and J. F. Caceres, 2009 The SR protein family of splicing factors: master regulators of gene expression. Biochem. J. 417: 15–27.

Love, M. I., W. Huber, and S. Anders, 2014 Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15: 550.

Lucks, J. B., S. A. Mortimer, C. Trapnell, S. Luo, S. Aviran *et al.*, 2011 Multiplexed RNA structure characterization with selective 2'-hydroxyl acylation analyzed by primer extension sequencing (SHAPE-Seq). Proc. Natl. Acad. Sci. 108: 11063–11068.

Luco, R. F., M. Allo, I. E. Schor, A. R. Kornblihtt, and T. Misteli, 2011 Epigenetics in alternative pre-mRNA splicing. Cell 144: 16–26.

Lustig, a J., R. J. Lin, and J. Abelson, 1986 The yeast RNA gene products are essential for mRNA splicing in vitro. Cell 47: 953–63.

Lybarger, S., K. Beickman, V. Brown, N. Dembla-Rajpal, K. Morey *et al.*, 1999 Elevated levels of a U4/U6.U5 snRNP-associated protein, Spp381p, rescue a mutant defective in spliceosome maturation. Mol. Cell. Biol.

Mamanova, L., A. J. Coffey, C. E. Scott, I. Kozarewa, E. H. Turner *et al.*, 2010 Target-enrichment strategies for next- generation sequencing. Nat. Methods 7: 111–118.

Marshall, A. N., M. C. Montealegre, C. Jim??nez-L??pez, M. C. Lorenz, and A. van Hoof, 2013 Alternative Splicing and Subfunctionalization Generates Functional Diversity in Fungal Proteomes. PLoS Genet. 9:.

Matera, A. G., and Z. Wang, 2014 A day in the life of the spliceosome. Nat. Rev. Mol. Cell Biol.

Matlin, A. J., F. Clark, and C. W. J. Smith, 2005 Understanding alternative splicing: towards a cellular code. Nat. Rev. Mol. Cell Biol. 6: 386–98.

Mayas, R. M., H. Maita, and J. P. Staley, 2006 Exon ligation is proofread by the DExD/H-box ATPase Prp22p. Nat. Struct. Mol. Biol.

Mayerle, M., M. Raghavan, S. Ledoux, A. Price, N. Stepankiw *et al.*, 2017 Structural toggle in

the RNaseH domain of Prp8 helps balance splicing fidelity and catalytic efficiency. Proc. Natl. Acad. Sci. 114: 4739–4744.

Mazroui, R., a Puoti, and a Krämer, 1999 Splicing factor SF1 from Drosophila and Caenorhabditis: presence of an N-terminal RS domain and requirement for viability. RNA 5: 1615–31.

McLaren, W., L. Gil, S. E. Hunt, H. S. Riat, G. R. S. Ritchie *et al.*, 2016 The Ensembl Variant Effect Predictor. Genome Biol. 17:.

Mercer, T. R., M. B. Clark, J. Crawford, M. E. Brunck, D. J. Gerhardt *et al.*, 2014 Targeted sequencing for gene discovery and quantification using RNA CaptureSeq. Nat. Protoc. 9: 989–1009.

Mercer, T. R., D. J. Gerhardt, M. E. Dinger, J. Crawford, C. Trapnell *et al.*, 2011 Targeted RNA sequencing reveals the deep complexity of the human transcriptome. Nat. Biotechnol. 30: 99–104.

Merkin, J., C. Russell, P. Chen, and C. B. Burge, 2012 Evolutionary dynamics of gene and isoform regulation in Mammalian tissues. Science 338: 1593–9.

Meyer, M., and J. Vilardell, 2009 The quest for a message: budding yeast, a model organism to study the control of pre-mRNA splicing. Brief. Funct. Genomic. Proteomic. 8: 60–7.

Mikheyeva, I. V, P. J. R. Grady, F. B. Tamburini, D. R. Lorenz, and H. P. Cam, 2014 Multifaceted genome control by Set1 Dependent and Independent of H3K4 methylation and the Set1C/COMPASS complex. PLoS Genet. 10: e1004740.

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

Millhouse, S., and J. L. Manley, 2005 The C-terminal domain of RNA polymerase II functions as a phosphorylation-dependent splicing activator in a heterologous protein. Mol. Cell. Biol. 25: 533–544.

Misteli, T., and D. L. Spector, 1999 RNA polymerase II targets pre-mRNA splicing factors to transcription sites in vivo. Mol. Cell 3: 697–705.

Mitchell, P., and D. Tollervey, 2003 An NMD pathway in yeast involving accelerated deadenylation and exosome-mediated 3??????5??? degradation. Mol. Cell 11: 1405–1413.

Morris, D. P., and a. L. Greenleaf, 2000 The splicing factor, Prp40, binds the phosphorylated carboxyl-terminal domain of RNA Polymerase II. J. Biol. Chem. 275: 39935–39943.

Morrison, A. J., and X. Shen, 2009 Chromatin remodelling beyond transcription: the INO80 and SWR1 complexes. Nat. Rev. Mol. Cell Biol. 10: 373–384.

Muñoz, M. J., M. de la Mata, and A. R. Kornblihtt, 2010 The carboxy terminal domain of RNA polymerase II and alternative splicing. Trends Biochem. Sci. 35: 497–504.

Nakazawa, N., S. Harashima, and Y. Oshima, 1991 AAR2, a gene for splicing pre-mRNA of the MATa1 cistron in cell type control of Saccharomyces cerevisiae. Mol. Cell. Biol. 11: 5693–5700.

Network, T. C. G. A., 2012 Comprehensive molecular portraits of human breast tumors. Nature.

Nguyen, T. H. D., W. P. Galej, X. Bai, C. G. Savva, A. J. Newman *et al.*, 2015 The architecture of the spliceosomal U4/U6.U5 tri-snRNP. Nature 523: 47–52.

Nguyen, T. H. D., W. P. Galej, S. M. Fica, P. C. Lin, A. J. Newman *et al.*, 2016 CryoEM structures of two spliceosomal complexes: Starter and dessert at the spliceosome feast. Curr. Opin. Struct. Biol. 36: 48–57.

Noble, S. M., and C. Guthrie, 1996 Identification of novel genes required for yeast pre-mRNA

splicing by means of cold-sensitive mutations. Genetics 143: 67–80.

Nojima, T., T. Gomes, A. R. F. Grosso, H. Kimura, M. J. Dye *et al.*, 2015 Mammalian NET-Seq Reveals Genome-wide Nascent Transcription Coupled to RNA Processing. Cell 161: 526–540.

Noma, T., K. Sode, and K. Ikebukuro, 2006 Characterization and application of aptamers for Taq DNA polymerase selected using an evolution-mimicking algorithm. Biotechnol. Lett. 28: 1939–1944.

Ohi, M. D., A. J. Link, L. Ren, J. L. Jennings, W. H. McDonald *et al.*, 2002 Proteomics analysis reveals stable multiprotein complexes in both fission and budding yeasts containing Myb-related Cdc5p/Cef1p, novel pre-mRNA splicing factors, and snRNAs. Mol. Cell. Biol. 22: 2011–2024.

Padgett, R. A., M. M. Konarska, M. Aebi, H. Hornig, C. Weissmann *et al.*, 1985 Nonconsensus branch-site sequences in the in vitro splicing of transcripts of mutant rabbit beta-globin genes. Proc. Natl. Acad. Sci. U. S. A. 82: 8349–8353.

Pandit, S., Y. Zhou, L. Shiue, G. Coutinho-Mansfield, H. Li *et al.*, 2013 Genome-wide Analysis Reveals SR Protein Cooperation and Competition in Regulated Splicing. Mol. Cell 50: 223–235.

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

Parker, R., and P. G. Siliciano, 1993 Evidence for an essential non-Watson-Crick interaction between the first and last nucleotides of a nuclear pre-mRNA intron. Nature.

Patrick, K. L., C. J. Ryan, J. Xu, J. J. Lipp, K. E. Nissen *et al.*, 2015 Genetic Interaction Mapping Reveals a Role for the SWI/SNF Nucleosome Remodeler in Spliceosome Activation in

Fission Yeast. PLoS Genet. 11: e1005074.

Picelli, S., A. K. Björklund, B. Reinius, S. Sagasser, G. Winberg *et al.*, 2014 Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. Genome Res. 24: 2033–2040.

Potashkin, J., R. Li, and D. Frendewey, 1989 Pre-mRNA splicing mutants of Schizosaccharomyces pombe. EMBO J. 8: 551–9.

Qin, D., L. Huang, A. Wlodaver, J. Andrade, and J. P. Staley, 2016 Sequencing of lariat termini in S. cerevisiae reveals 5' splice sites, branch points, and novel splicing events. RNA.

Quesada, V., A. J. Ramsay, and C. Lopez-Otin, 2012 Chronic lymphocytic leukemia with SF3B1 mutation. N. Engl. J. Med.

Quinlan, A. R., and I. M. Hall, 2010 BEDTools: A flexible suite of utilities for comparing genomic features. Bioinformatics 26: 841–842.

Ram, O., and G. Ast, 2007a SR proteins: a foot on the exon before the transition from intron to exon definition. Trends Genet. 23: 5–7.

Ram, O., and G. Ast, 2007b SR proteins: a foot on the exon before the transition from intron to exon definition. Trends Genet. 23: 5–7.

Ramírez, F., D. P. Ryan, B. Grüning, V. Bhardwaj, F. Kilpert *et al.*, 2016 deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res. 44: W160–W165.

Rauhut, R., P. Fabrizio, O. Dybkov, K. Hartmuth, V. Pena *et al.*, 2016 Molecular architecture of the Saccharomyces cerevisiae activated spliceosome. Science (80-.).

Ren, L., J. R. McLean, T. R. Hazbun, S. Fields, C. Vander Kooi *et al.*, 2011 Systematic Two-Hybrid and Comparative Proteomic Analyses Reveal Novel Yeast Pre-mRNA Splicing

Factors Connected to Prp19. PLoS One 6: e16719.

Rhind, N., Z. Chen, M. Yassour, D. a Thompson, B. J. Haas *et al.*, 2011 Comparative functional genomics of the fission yeasts. Science 332: 930–6.

Robberson, B. L., G. J. Cote, and S. M. Berget, 1990 Exon definition may facilitate splice site selection in RNAs with multiple exons. Mol. Cell. Biol. 10: 84–94.

Romfo, C. M., C. J. Alvarez, W. J. van Heeckeren, C. J. Webb, and J. a Wise, 2000 Evidence for splice site pairing via intron definition in Schizosaccharomyces pombe. Mol. Cell. Biol. 20: 7955–7970.

Rosado-Lugo, J. D., and M. Hampsey, 2014 The Ssu72 phosphatase mediates the RNA polymerase II initiation-elongation transition. J. Biol. Chem. 289: 33916–26.

Rosenberg, G. H., S. K. Alahari, and N. F. Käufer, 1991 prp4 from Schizosaccharomyces pombe, a mutant deficient in pre-mRNA splicing isolated using genes containing artificial introns. MGG Mol. Gen. Genet. 226: 305–309.

Roshbash, M., P. K. Harris, J. L. Woolford, and J. L. Teem, 1981 The effect of temperature-sensitive RNA mutants on the transcription products from cloned ribosomal protein genes of yeast. Cell 24: 679–686.

Rouskin, S., M. Zubradt, S. Washietl, M. Kellis, and J. S. Weissman, 2014 Genome-wide probing of RNA structure reveals active unfolding of mRNA structures in vivo. Nature 505: 701–5.

Ryan, C. J., A. Roguev, K. Patrick, J. Xu, H. Jahari *et al.*, 2012 Hierarchical Modularity and the Evolution of Genetic Interactomes across Species. Mol. Cell 46: 691–704.

Sasaki-Haraguchi, N., T. Ikuyama, S. Yoshii, T. Takeuchi-Andoh, D. Frendewey *et al.*, 2015 Cwf16p associating with the nineteen complex ensures ordered exon joining in constitutive

Pre-mRNA splicing in fission yeast. PLoS One 10: 1–16.

Scheckel, C., and R. B. Darnell, 2015 Microexons—Tiny but mighty. EMBO J. 34: 273–274.

Schellenberg, M. J., T. Wu, D. B. Ritchie, S. Fica, J. P. Staley *et al.*, 2013 A conformational switch in PRP8 mediates metal ion coordination that promotes pre-mRNA exon ligation. Nat. Struct. Mol. Biol.

Schones, D. E., K. Cui, S. Cuddapah, T.-Y. Roh, A. Barski *et al.*, 2008 Dynamic regulation of nucleosome positioning in the human genome. Cell 132: 887–898.

Schwer, B., 2008 A Conformational Rearrangement in the Spliceosome Sets the Stage for Prp22-Dependent mRNA Release. Mol. Cell.

Schwer, B., and C. H. Gross, 1998 Prp22, a DExH-box RNA helicase, plays two distinct roles in yeast pre-mRNA splicing. EMBO J.

Schwer, B., and S. Shuman, 2015 Structure-function analysis and genetic interactions of the Yhc1, SmD3, SmB, and Snp1 subunits of yeast U1 snRNP and genetic interactions of SmD3 with U2 snRNP subunit Lea1. RNA 21: 1173–86.

Semlow, D. R., and J. P. Staley, 2012 Staying on message: Ensuring fidelity in pre-mRNA splicing. Trends Biochem. Sci.

Shao, C., B. Yang, T. Wu, J. Huang, P. Tang *et al.*, 2014 Mechanisms for U2AF to define 3' splice sites and regulate alternative splicing in the human genome. Nat. Struct. Mol. Biol. 21: 997–1005.

Sims, R. J., S. Millhouse, C. F. Chen, B. A. Lewis, H. Erdjument-Bromage *et al.*, 2007 Recognition of Trimethylated Histone H3 Lysine 4 Facilitates the Recruitment of Transcription Postinitiation Factors and Pre-mRNA Splicing. Mol. Cell 28: 665–676.

Spies, N., C. B. Nielsen, R. A. Padgett, and C. B. Burge, 2009 Biased Chromatin Signatures

around Polyadenylation Sites and Exons. Mol. Cell 36: 245–254.

Stepankiw, N., M. Raghavan, E. A. Fogarty, A. Grimson, and J. a. Pleiss, 2015 Widespread alternative and aberrant splicing revealed by lariat sequencing. Nucleic Acids Res. 43: 8488–501.

Storey, J. D., and R. Tibshirani, 2003 Statistical significance for genomewide studies. Proc. Natl. Acad. Sci. 100: 9440–9445.

Szymczyna, B. R., J. Bowman, S. McCracken, A. Pineda-Lucena, Y. Lu *et al.*, 2003 Structure and function of the PWI motif: a novel nucleic acid-binding domain that facilitates pre-mRNA processing. Genes Dev. 17: 461–475.

Taggart, A. J., A. M. DeSimone, J. S. Shih, M. E. Filloux, and W. G. Fairbrother, 2012 Large-scale mapping of branchpoints in human pre-mRNA transcripts in vivo. Nat. Struct. Mol. Biol. 19: 719–721.

The Gene Ontology Consortium, 2000 Gene Ontology: tool for the unification of biology. Nat. Genet. 25: 25–29.

The Gene Ontology Consortium, 2015 Gene Ontology Consortium: going forward. Nucleic Acids Res. 43: D1049–D1056.

Tilgner, H., C. Nikolaou, S. Althammer, M. Sammeth, M. Beato *et al.*, 2009 Nucleosome positioning as a determinant of exon recognition. Nat. Struct. Mol. Biol. 16: 996–1001.

Tsai, R. T., R. H. Fu, F. L. Yeh, C. K. Tseng, Y. C. Lin *et al.*, 2005a Spliceosome disassembly catalyzed by Prp43 and its associated components Ntr1 and Ntr2. Genes Dev. 19: 2991–3003.

Tsai, R. T., R. H. Fu, F. L. Yeh, C. K. Tseng, Y. C. Lin *et al.*, 2005b Spliceosome disassembly catalyzed by Prp43 and its associated components Ntr1 and Ntr2. Genes Dev.

Tseng, C. K., H. L. Liu, and S. C. Cheng, 2011 DEAH-box ATPase Prp16 has dual roles in remodeling of the spliceosome in catalytic steps. RNA.

Turunen, J. J., E. H. Niemelä, B. Verma, and M. J. Frilander, 2013 The significant other: Splicing by the minor spliceosome. Wiley Interdiscip. Rev. RNA.

Urushiyama, S., T. Tani, and Y. Ohshima, 1996 Isolation of novel pre-mRNA splicing mutants of Schizosaccharomyces pombe. Mol. Gen. Genet. 253: 118–127.

Vanoosthuyse, V., P. Legros, S. J. A. van der Sar, G. Yvert, K. Toda *et al.*, 2014 CPF-Associated Phosphatase Activity Opposes Condensin-Mediated Chromosome Condensation. PLoS Genet. 10: e1004415.

Verdel, A., S. Jia, S. Gerber, T. Sugiyama, S. Gygi *et al.*, 2004 RNAi-mediated targeting of heterochromatin by the RITS complex. Science 303: 672–676.

Vijayraghavan, U., M. Company, and J. Abelson, 1989 Isolation and characterization of pre-mRNA splicing mutants of Saccharomyces cerevisiae. Genes Dev. 3: 1206–1216.

Villa, T., and C. Guthrie, 2005 The Isy1p component of the NineTeen Complex interacts with the ATPase Prp16p to regulate the fidelity of pre-mRNA splicing. Genes Dev.

Vincent, M., P. Lauriault, M. F. Dubois, S. Lavoie, O. Bensaude *et al.*, 1996 The nuclear matrix protein p255 is a highly phosphorylated form of RNA polymerase II largest subunit which associates with spliceosomes. Nucleic Acids Res. 24: 4649–4652.

Vo, T. V., J. Das, M. J. Meyer, N. A. Cordero, N. Akturk *et al.*, 2016 A Proteome-wide Fission Yeast Interactome Reveals Network Evolution Principles from Yeasts to Human. Cell 164: 310–323.

Wahl, M. C., C. L. Will, and R. Lührmann, 2009 The Spliceosome: Design Principles of a Dynamic RNP Machine. Cell 136: 701–718.

Washington, K., T. Ammosova, M. Beullens, M. Jerebtsova, A. Kumar *et al.*, 2002 Protein phosphatase-1 dephosphorylates the C-terminal domain of RNA polymerase-II. J. Biol. Chem. 277: 40442–8.

Webb, C. J., C. M. Romfo, W. J. van Heeckeren, and J. A. Wise, 2005a Exonic splicing enhancers in fission yeast: functional conservation demonstrates an early evolutionary origin. Genes Dev. 19: 242–254.

Webb, C. J., C. M. Romfo, W. J. van Heeckeren, and J. A. Wise, 2005b Exonic splicing enhancers in fission yeast: functional conservation demonstrates an early evolutionary origin. Genes Dev. 19: 242–54.

Webb, C. J., and J. A. Wise, 2004 The Splicing Factor U2AF Small Subunit Is Functionally Conserved between Fission Yeast and Humans. Mol. Cell. Biol. 24: 4229–4240.

Wernersson, R., and H. B. Nielsen, 2005 OligoWiz 2.0 - Integrating sequence feature annotation into the design of microarray probes. Nucleic Acids Res. 33:.

Wilhelm, B., S. Marguerat, S. Aligianni, S. Codlin, S. Watt *et al.*, 2011 Differential patterns of intronic and exonic DNA regions with respect to RNA polymerase II occupancy, nucleosome density and H3K36me3 marking in fission yeast. Genome Biol 12: R82.

Wilhelm, B. T., S. Marguerat, S. Watt, F. Schubert, V. Wood *et al.*, 2008 Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution. Nature 453: 1239–43.

Wilkinson, M. E., S. M. Fica, W. P. Galej, C. M. Norman, A. J. Newman *et al.*, 2017 Postcatalytic spliceosome structure reveals mechanism of 3 ’-splice site selection. Science (80-.).

Will, C. L., and R. Lührmann, 2011a Spliceosome structure and function. Cold Spring Harb. Perspect. Biol. 3:.

Will, C. L., and R. Lührmann, 2011b Spliceosome structure and function. Cold Spring Harb. Perspect. Biol. 3: 1–2.

Xu, H., B. J. Fair, Z. Dwyer, M. Gildea, and J. A. Pleiss, 2018 Multiplexed Primer Extension Sequencing Enables High Precision Detection of Rare Splice Isoforms. BioRxiv.

Yan, C., J. Hang, R. Wan, M. Huang, C. C. L. Wong *et al.*, 2015 Structure of a yeast spliceosome at 3.6-angstrom resolution. Science 349: 1182–91.

Yan, C., R. Wan, R. Bai, G. Huang, and Y. Shi, 2016 Structure of a yeast activated spliceosome at 3.5 Å resolution. Science (80-.).

Yeh, T.-C., H.-L. Liu, C.-S. Chung, N.-Y. Wu, Y.-C. Liu *et al.*, 2011 Splicing Factor Cwc22 Is Required for the Function of Prp2 and for the Spliceosome To Escape from a Futile Pathway. Mol. Cell. Biol.

Zheng, W., L. M. Chung, and H. Zhao, 2011 Bias detection and correction in RNA-Sequencing data. BMC Bioinformatics 12:.

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

Zhu, Y. Y., E. M. Machleder, A. Chenchik, R. Li, and P. D. Siebert, 2001 Reverse transcriptase template switching: A SMART^{TM} approach for full-length cDNA library construction. Biotechniques 30: 892–897.

Zlatanova, J., and A. Thakar, 2008 H2A.Z: View from the Top. Structure 16: 166–179.

Zofall, M., T. Fischer, K. Zhang, M. Zhou, B. Cui *et al.*, 2009 Histone H2A.Z cooperates with RNAi and heterochromatin factors to suppress antisense RNAs. Nature 461: 419–422.

Zuo, P., and T. Maniatis, 1996 The splicing factor U2AF35 mediates critical protein-protein

interactions in constitutive and enhancer-dependent splicing. Genes Dev. 10: 1356–1368.

CHAPTER2: interconnections between RNA-processing pathways revealed by a sequencing-based genetic screen for pre-mRNA splicing mutants in fission yeast

Alternative citation for this chapter:

*Denotes equal contribution

Larson A*, Fair BJ*, & Pleiss JA. (2016). Interconnections Between RNA-Processing Pathways Revealed by a Sequencing-Based Genetic Screen for Pre-mRNA Splicing Mutants in Fission Yeast. G3 (Bethesda, Md.), 6(6), 1513–23.

Abstract

Pre-mRNA splicing is an essential component of eukaryotic gene expression and is highly conserved from unicellular yeasts to humans. Here we present the development and implementation of a sequencing based reverse genetic screen designed to identify non-essential genes that impact pre-mRNA splicing in the fission yeast *Schizosaccharomyces pombe*, an organism that shares many of the complex features of splicing in higher eukaryotes. Using a custom-designed barcoding scheme, we simultaneously queried ~3000 mutant strains for their impact on the splicing efficiency of two endogenous pre-mRNAs. A total of 61 non-essential genes were identified whose deletions resulted in defects in pre-mRNA splicing; enriched among these were factors encoding known or predicted components of the spliceosome. Included among the candidates identified here are genes with well-characterized roles in other RNA-processing pathways, including heterochromatic silencing and 3′ end processing. Splicing-sensitive microarrays confirm broad splicing defects for many of these factors, revealing novel functional connections between these pathways.

Introduction

The protein-coding regions of most eukaryotic genes are interrupted by non-coding introns which must be precisely removed from the pre-mRNA in order to generate a translatable message. This essential process is carried out by the spliceosome, a dynamic macromolecular machine that recognizes specific sequence elements within the pre-mRNA, such as the short consensus sequences that mark intron boundaries, and catalyzes intron removal (Will and Lührmann 2011a). At its core, the spliceosome is composed of five snRNA-protein complexes (snRNPs), each comprised of a single RNA and multiple core protein factors. In addition, the human spliceosome associates with upwards of 200 auxiliary splicing proteins that aid in proper recognition of splice sites and catalysis (Wahl *et al.* 2009).

Over the last decade, it has become increasingly clear that splicing is integrated with other steps of pre-mRNA synthesis and maturation. Studies from yeast to humans suggest that the majority of splicing occurs co-transcriptionally while the RNA is still tethered to the polymerase (Core and Lis 2008; Pandya-Jones and Black 2009; Carrillo Oesterreich *et al.* 2010; Ameur *et al.* 2011; Khodor *et al.* 2011). Accordingly, multiple lines of evidence support the idea that transcriptional dynamics influence the splicing process. Mutations that alter polymerase elongation rate yield different splicing patterns (de la Mata *et al.* 2003; Dujardin *et al.* 2014), suggesting a kinetic coupling between transcription rate and the ability of splicing factors to recognize and act upon splice sites. Genome-wide studies have also demonstrated that transcriptional pausing coincides with the splicing process (Core and Lis 2008; Alexander *et al.* 2010). In addition to a kinetic coupling of elongation and splicing, biochemical studies have shown that the C-terminal domain (CTD) of RNA polymerase II can directly interact with splicing components (Vincent *et al.* 1996; Misteli and Spector 1999) and post-translational

modifications of the CTD can differentially impact recruitment of splicing components (Morris and Greenleaf 2000; Muñoz *et al.* 2010; David and Manley 2011). The chromatin environment encountered by the transcribing polymerase can also influence splicing, and genome-wide nucleosome positioning data from fission yeast to humans reveal an enrichment of nucleosome density in exons over introns (Schones *et al.* 2008; Andersson *et al.* 2009; Spies *et al.* 2009; Tilgner *et al.* 2009; Wilhelm *et al.* 2011; Iannone *et al.* 2015). The mechanistic link between chromatin state and splicing could be explained in part by the relation between chromatin state and polymerase kinetics (Li *et al.* 2007; Luco *et al.* 2011), but might also reflect direct interactions between chromatin marks and splicing factors. For example, the H3K4me3 mark interacts with the U2 snRNP through interactions with the adapter protein CHD1 (Sims *et al.* 2007). In addition to chromatin-based interactions, some evidence suggest that the cleavage and polyadenylation machinery at the 3′ end of transcripts can interact with splicing components to influence splicing. In higher eukaryotes, identification of splice sites in terminal introns requires interactions between splicing components and the cleavage and polyadenylation machinery (Berget 1995). Recently, the cleavage and polyadenylation factor CPSF1 was found to regulate alternative splicing in human T-cells (Evsyukova *et al.* 2013). These interconnections between splicing and other nuclear processes underscore the need for unbiased genome-wide approaches to identify the full complement of factors that functionally impact spliceosomal activity.

The fission yeast, *Schizosaccharomyces pombe*, provides a powerful genetic system in which to examine the splicing pathway. Like the budding yeast, *Saccharomyces cerevisiae*, fission yeast is genetically tractable, allowing for easy manipulation of its genome. The *S. cerevisiae* genome, however, has shed most of its introns, with only ~300 introns remaining (Kupfer *et al.* 2004). In contrast, over 5000 introns have been identified in the *S. pombe* genome,

and over 1000 genes are interrupted by multiple introns (Wilhelm *et al.* 2008; Rhind *et al.* 2011). Furthermore, whereas splice site sequences in budding yeast introns tend to conform to a strict consensus sequence, *S. pombe* splice sites are characterized by a far higher level of degeneracy (Kupfer *et al.* 2004), more closely resembling the degeneracy seen in human splice sites. Perhaps accordingly, sequence homology identifies many auxiliary components of the human spliceosome, such as SR proteins, that are present in the *S. pombe* genome but have been lost in the *S. cerevisae* lineage (Käufer and Potashkin 2000; Webb *et al.* 2005b). These properties suggest that regulation of pre-mRNA splicing in *S. pombe* may be more similar to that seen in humans than in *S. cerevisiae* (Ram and Ast 2007a). Indeed, some transcripts in *S. pombe* are subject to mammalian-like, environmentally regulated exon skipping (Awan *et al.* 2013), and others respond to insertion of mammalian splicing enhancer elements (Webb *et al.* 2005a). Moreover, similar to observations in mammalian cells, widespread activation of cryptic splice sites has been demonstrated in *S. pombe*, highlighting the flexibility in the *S. pombe* spliceosome for selecting splice sites (Bitton *et al.* 2015; Stepankiw *et al.* 2015). Although these features highlight the potential of *S. pombe* to serve as a model for understanding the complex splicing seen in higher eukaryotes, the precise factors responsible for regulating these splicing events remain largely unknown.

Components of the *S. pombe* spliceosome have been identified using a variety of approaches. Genetic screening of randomly mutagenized strains identified numerous core splicing factors (Potashkin *et al.* 1989; Rosenberg *et al.* 1991; Alahari *et al.* 1993; Habara *et al.* 1998) and biochemical purifications followed by mass spectrometry have greatly added to the list of components (Ohi *et al.* 2002; Chen *et al.* 2014). Although these strategies successfully identified core components of the spliceosome, they have been less effective at identifying

factors that functionally connect splicing with other nuclear processes. More recently, a high-throughput genetic interaction mapping strategy examining non-essential *S. pombe* genes identified strong genetic interactions between U2 snRNP components of the spliceosome and chromatin remodeling enzymes such as the SWI/SNF complex (Patrick *et al.* 2015), suggesting that a mechanistic coupling between chromatin modification and splicing also exists in *S. pombe*. In addition, recent systematic genome-wide yeast-two-hybrid interaction mapping strategies have correctly identified a handful of *S. pombe* genes as factors in the splicing pathway based on physical interactions with known spliceosome components (Ren *et al.* 2011; Vo *et al.* 2016). These high-throughput genetic and physical interaction strategies can yield a wealth of information and strongly hint at a gene's involvement in the splicing pathway, but they do not provide a direct functional test of a factor's impact on splicing.

We have previously described a reverse-genetic screening methodology in *S. cerevisiae* that couples high-throughput sample processing with quantitative RT-PCR to enable direct measurements of the splicing efficiency of endogenous pre-mRNA transcripts in the background of thousands of mutant strains (Albulescu *et al.* 2012). In addition to identifying the majority of known splicing mutants, this work successfully identified splicing defects associated with components of the SWI/SNF complex as well as with components of the cleavage and polyadenylation machinery, confirming both the sensitivity of this approach and the evolutionarily conserved nature of these functional interactions. Here we present the results of a study designed to identify non-essential genes in the *S. pombe* genome whose deletion impacts the splicing efficiency of endogenous transcripts. We have developed and implemented a sequencing-based approach for monitoring splicing efficiency in the background of thousands of *S. pombe* strains, and describe the functional significance of those mutants identified.

Materials and Methods:

Strains and cell growth:

All strains examined here were from the haploid deletion library from Bioneer (Kim *et al.* 2010), representing 3020 individual gene deletions. All strains were grown in supplemented rich growth medium (YES) at 32°C, according to standard procedures (Forsburg and Rhind 2006). Strains were recovered from glycerol stocks on solid media supplemented with 200 µg/mL G418. A manual pinning tool (V&P Scientific, cat.#: VP384FP6) was used to transfer cells from solid media into 384-well microtiter plates (Greiner BioOne, cat.#: 781271) for growth in liquid media. Liquid cultures were grown in an Infors HT Multitron plate shaker at 900 rpm with 80% constant humidity. Breathable adhesive tape (VWR, cat.#: 60941-086) was used to seal the plates and reduce evaporation. Because the growth rates of the strains being used varied substantially, an approach was developed to enable the collection of a similar number of actively growing cells for every strain. Initial cultures (150 µL) of all strains were grown in microtiter plates for two days, allowing nearly all strains to reach saturation. The cell density for most strains is similar at saturation, allowing us to effectively 'normalize' cell numbers. Using a liquid handling robot (Biomek NX), 2 µL of saturated culture was used to inoculate 150 µL of fresh media in duplicate to create biological replicate cultures for each strain. After inoculation, cells were allowed to grow for eight hours, allowing most strains to reach A_{600} values near 0.5. Cells were harvested by centrifugation at 5000x*g* for 5 min, and pellets were flash frozen in liquid N_2 and stored at -80°C until further processing.

cDNA synthesis and library preparation:

RNA was isolated from cell pellets and cDNA was synthesized using random ninemers for primers, as previously described (Albulescu *et al.* 2012). The resulting cDNA was amplified

by two sequential PCR reactions to generate products compatible with the Illumina HiSeq2000 Flow Cell as follows. For each cDNA sample, a 12 µL PCR reaction was prepared containing 1x Phusion HF buffer (New England Biolabs), 1x Phusion enzyme, 250 nM forward primer with custom plate-specific barcodes, 250 nM reverse primer, and 1% of the cDNA reaction. The plate specific barcodes sequences were designed as previously described (Mamanova *et al.* 2010). Cycling conditions for this first PCR reaction were as follows: 95°C for 3 min, then an empirically determined number of cycles of 98°C for 15 s, 62°C for 20 s, 72°C for 30 s. The number of amplification cycles required was determined in a separate QPCR reaction as the minimum number of cycles necessary to generate a detectable fluorescence signal; required cycle numbers varied from 18 to 21 for the different primers and plates. The products resulting from this first PCR contained plate-specific barcodes, but no well-specific barcodes (see Figure 1). For each target, the products from each plate of this first PCR reaction were pooled into a single 384-well microtiter plate, and 0.5 µL was used to seed a second PCR reaction (15 µL), during which well-specific Illumina-Nextera barcodes and Illumina Flow Cell binding sites were appended. This reaction contained 1x Phusion HF buffer (New England Biolabs), 1x Phusion enzyme, 200 nM forward Nextera index primer, and 200 nM Nextera reverse index primer. Cycling conditions were as follows: 95°C for 3 min, then five cycles of 98°C for 15 s, 68°C for 60 s. The PCR products were pooled, concentrated via ethanol precipitation, purified using glass fiber spin columns (Zymo Research), and separated on a 6% native acrylamide gel. Materials of the expected molecular weight ranges were excised from the gel and recovered by soaking crushed gel bits in 0.3 M sodium acetate followed by ethanol precipitation. The resulting DNA precipitate was dissolved in 25 µL water and sequenced on the Illumina HiSeq2000 with the assistance of the Cornell University Biotechnology Resource Center.

Data Processing:

Reads were demultiplexed using a combination of Nextera-specific indices and custom plate-specific barcodes (Mamanova *et al.* 2010) within the insert read. The bwa-mem (Li and Durbin 2009) aligner was then used to align reads to a custom index containing both the spliced and unspliced isoforms of the two targets. A splice-index (SI) was calculated for *fet5*_intron1 and *pwi1*_intron2 in each sample by comparing the number of reads mapped to the unspliced isoform versus the spliced isoform as follows:

$$SI = \frac{unspliced\ read\ count}{spliced\ read\ count}$$

To determine SI relative to wild type while accounting for plate to plate variation, we assumed that the median SI within each 384-well plate was representative of wild-type. Therefore, the relative SI was calculated as:

$$SI_{relative} = \frac{SI}{SI_{PlateMedian}}$$

After determining the SI for each biological replicate, we filtered our dataset to include only those samples for which the standard deviation between the $\log_2(SI_{relative})$ was less than 1, and for which the combined read count was greater than 1000. A total of 3007 and 3005 strains (99.6% and 99.5%) passed these quality scores for the *fet5*_intron1 and *pwi1*_intron2 datasets, respectively. In order to determine the subset of strains which exhibit a $\log_2(SI_{relative})$ that was statistically significantly different from wild-type, we considered how the precision of our SI_{rel} measurements varied as a function of read count. In concept, this approach has similarities to algorithms commonly used for RNA-seq analysis which empirically estimate noise within a dataset to identify significant changes in gene-expression (Love *et al.* 2014). The $\log_2$-transformed $SI_{relative}$ values were plotted as a function of log-transformed read count for each

sample (see Figure 2). The dataset was then divided into 20 equal sized bins based on read count. Using the mean and standard deviation within the bins as data points, spline interpolation was used to estimate the $\log_2$-transformed mean ($\mu_{interpolated}$) and standard deviation ($\sigma_{interpolated}$) of SI_{rel} measurements at any given read count under the null hypothesis. For each strain, a *p*-value was then estimated by defining a Z-score as follows:

$$Z = \frac{log_2(SI_{relative}) - \mu_{interpolated}}{\sigma_{interpolated}}$$

Strains were called as significant if the Benjamini-Hochberg corrected *p*-value was below 0.05. The 95% confidence intervals in Figure 3 represent $\log_2(SI_{relative}) \pm 2\sigma_{interpolated}$ given the read depth of that strain.

Splicing sensitive microarrays

All microarrays were performed as two-color arrays comparing mutant and wild type strains, each grown under identical conditions. Briefly, strains were grown to saturation at 30°C, then back-diluted in 25 mL cultures and allowed to grow at 30°C until they reached an optical density of A_{600}~0.5. Total cellular RNA samples were isolated, converted into cDNA, fluorescently labeled, and hybridized to the array as previously described (Inada and Pleiss 2010). Biological replicate microarrays were performed for most mutant strains, with average expression measurements between biological replicates being presented in the figures. Both raw and processed microarray data are available through GEO using accession number GSE79153.

Results and Discussion:

Here we sought to identify the full complement of non-essential genes that impact pre-mRNA splicing efficiency in *S. pombe*, an organism whose splicing properties closely resemble those seen in higher eukaryotes (Ram and Ast 2007a). To quantitatively measure the impact of mutations on the splicing pathway, we designed an assay that would allow for high-sensitivity detection of both spliced and unspliced isoforms in thousands of unique samples (Figure 1). Briefly, cDNA from a given sample was used as template for a PCR reaction using primers that flank a splicing event, enabling amplification of both spliced and unspliced isoforms. By appending appropriately barcoded sequences, the resulting material was subjected to deep sequencing to count the number of molecules corresponding to both the spliced and unspliced isoforms for each sample. To demonstrate that this approach could provide a quantitative representation of the underlying species, we measured isoform ratios for samples that contained known ratios of different spliced isoforms. Across a large range of relative isoform abundances, this sequencing-based approach gave results that were both highly accurate and precise (Figure S1).

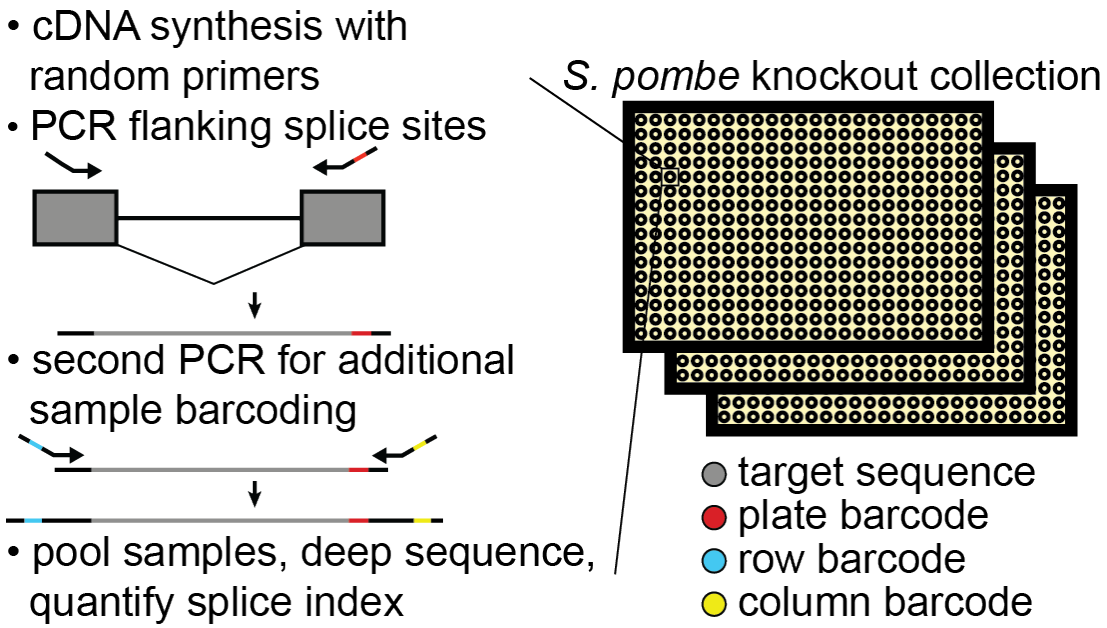


Figure 1: Schematic of workflow for quantitatively measuring splicing in the fission yeast deletion collection. After cell growth, RNA isolation, and cDNA synthesis with random primers, consecutive PCR reactions are performed using primers that flank an intron to amplify both spliced and unspliced RNA while appending sample specific barcodes and Illumina compatible ends. Estimates of splicing efficiency in each strain are determined by counting the number of spliced and unspliced sequencing reads derived from each well.

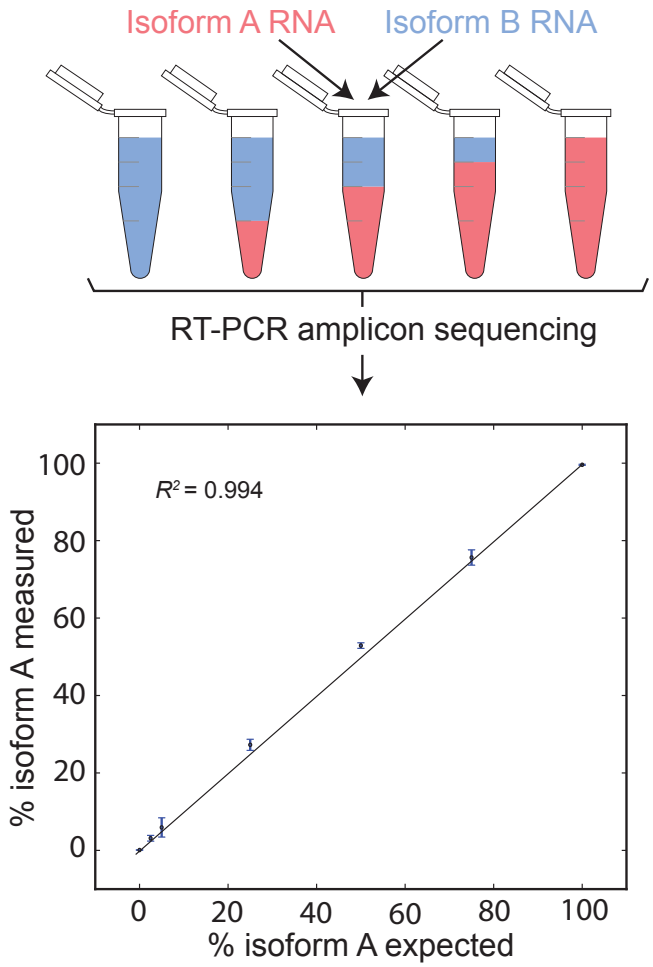


Figure S1: RT-PCR amplicon sequencing accurately measures splicing of in RNA mixtures of containing different relative isoform abundances. Total cellular RNA was collected from two strains, each carrying a different splice isoform of the *pwi1* gene ORF in place of the native *pwi1* gene. Isoforms A and B differed by a length of 90 nucletotides. RNA from each strain was mixed in different proportions, and the relative abundance of each isoform was measured in each mixture by RT-PCR amplicon sequencing. Errorbars represent the standard deviation between three technical replicate RT-PCR reactions.

After determining that this sequencing approach was sensitive and quantitative, we turned to examining each of the ~3000 deletion strains available within the *S. pombe* haploid deletion collection (Kim *et al.* 2010) to identify novel factors whose disruption impacts splicing. Primers were designed that would allow for the determination of the splicing efficiency of two introns: the single intron in the *fet5* transcript, a predicted GTPase involved in RNA polymerase localization, and the second intron in the *pwi1* transcript, a splicing co-activator. The *fet5* intron

resembles a typical intron in *S. pombe*, in that the *fet5* transcript contains just a single intron with canonical 5' splice site (GUAAGU), and branch point (UGCUAAU) sequences, and whose length (45 nt) is close to the median intron length (56 nt). The second intron in *pwi1* is also of typical length (59 nt) for an *S. pombe* intron, and has a typical branch point sequence (CAUUAAU) but has an atypical 5′ splice site sequence (GUACAA) which significantly deviates from the canonical sequence. Importantly, because these two introns are short, the PCR amplification efficiency of both the spliced and unspliced isoforms should be similar, reducing the likelihood of artifacts derived from amplification bias. In total, ~12,000 samples were generated, corresponding to each of these targets within each of these strains with biological replicates. As a convenient measure of splicing efficiency, we define the splice-index (SI) as the ratio of unspliced to spliced reads and looked to identify mutants that caused significant changes to the SI. Importantly, because this assay measures the steady state abundances of specific RNA species, a high SI could indicate a defect in pre-mRNA splicing, or alternatively, a change in the relative stabilities of spliced or unspliced RNA. For both the *fet5*_intron1 and *pwi1*_intron2, the unspliced pre-mRNA was present at about 2% of the spliced mRNA in the background of most strains, with respective median SI values of 0.018 and 0.025 (Figures S2A and B), consistent with the expectation that splicing occurs efficiently and the vast majority of these transcripts are present as the spliced isoform. Moreover, for the vast majority of strains, the measured SI was relatively close to the median value, with interquartile ranges across all samples of 0.004 and 0.011 for the *fet5*_intron1 and *pwi1*_intron2 targets, respectively (Figure S2A and B), consistent with the expectation that most genes examined here do not impact the splicing pathway. Across all strains, the biological replicates were correlated with R^2 values of 0.37 and 0.23 for *fet5*_intron1 and *pwi1*_intron2, respectively (Figure S2C and D).

Fig S2

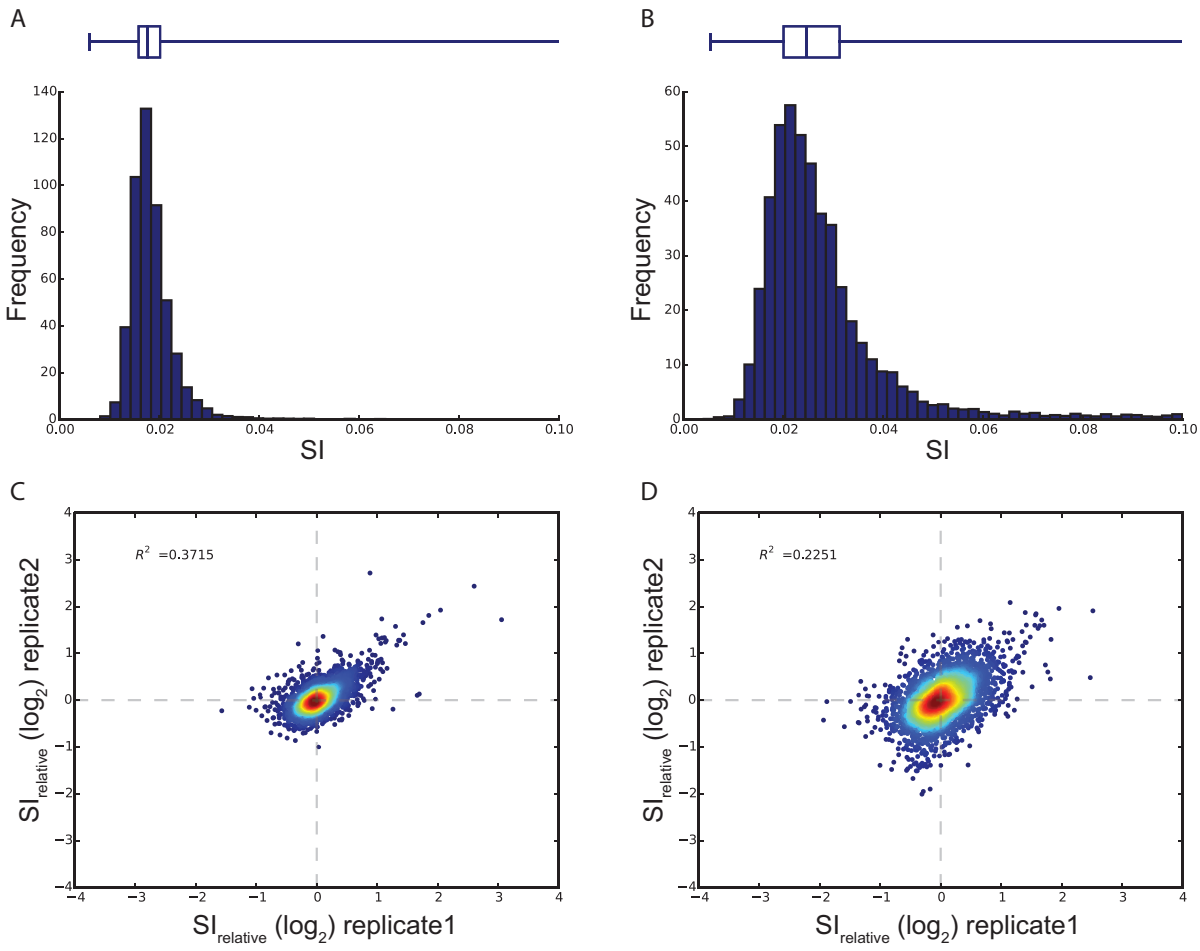


Figure S2: Splice index measurements for *fet5*-intron1 and *pwi1*-intron2 are reproducible. Histograms and box-whisker plots of the measured splice index for *fet5*-intron1 (A) and *pwi1*-intron2 (B) reveal that we measure a relatively consistent splice index in all samples. The splice index relative to the median is was calculated for each biological replicate for each deletion strain and a replicate scatter plots (C, D) show that the relative splice index measurements are largely reproducible among the 3000+ strains in the strain collection.

As with any RNA-sequencing experiment, the statistical power to identify changes in expression increases with greater read depth. In order to identify the subset of strains that exhibited a significant change in splicing, we developed a statistical test that assessed the observed change in SI as a function of read depth, while addressing empirically observed biases that are not addressed in more standard differential splicing statistical tests (see Materials and Methods, Figure 2). Using this approach, statistically significant changes in SI were identified

for 57 and 18 deletion strains for the *fet5*_intron1 and *pwi1*_intron2, respectively (see Figure 3). Importantly, of the 18 strains that affected splicing *pwi1*_intron2, 14 were also found to affect splicing of *fet5*_intron1. This significant degree of overlap (p<3.81e-22, Fisher's exact test) suggests that the splicing defect observed in many of the strains is not specific to a single gene.

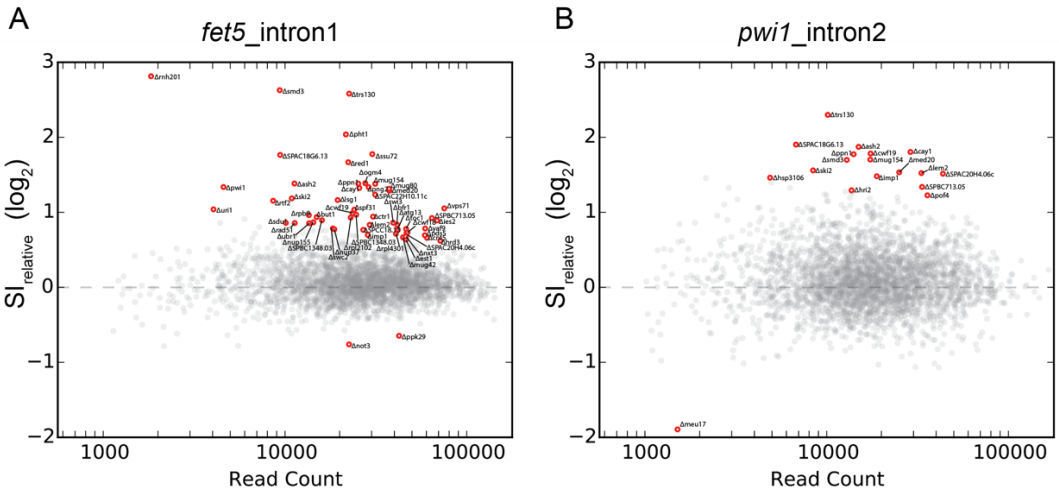


Figure 2: Relative splice index measurements in deletion strains. The relative splice index for *fet5*_intron1 (A) or *pwi1*_intron2 (B) is plotted as a function of read count for each of the ~3000 strains examined. Strains which significantly differed from wild-type after multiple hypothesis correction are colored red and labeled. A total of 61 strains were identified as having a significantly different splice index measurement for either *fet5*_intron1 or *pwi*_intron2.

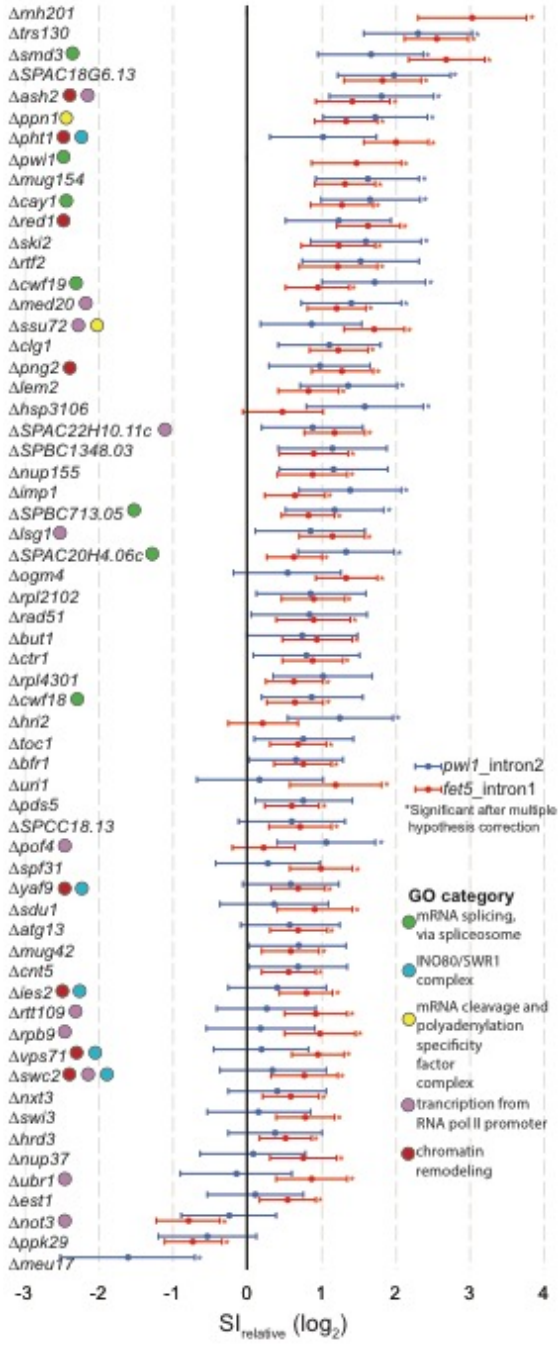


Figure 3: Gene deletions which result in significant splice index changes in either *fet5*_intron1 or *pwi1*_intron2. The measured relative splice index is shown for *fet5*_intron1 and *pwi1*_intron2 with 95% confidence intervals for the 61 gene deletion strains which were significantly different than wild-type for at least one of splicing events examined. Notable Gene Ontology (GO) categories are indicated.

To better understand the functional significance of the genes identified through this screen, we asked whether there was enrichment for factors involved in similar pathways by analyzing their Gene Ontology (GO) (The Gene Ontology Consortium 2000, 2015). Appropriately, the most highly enriched biological process identified was 'mRNA splicing, via spliceosome' (p<6.63e-3), confirming the ability of the method to positively identify known splicing factors. Consistent with our previous results in *S. cerevisiae*, not all deletions of known splicing factors resulted in a measurable change in splicing efficiency of either of the tested introns. Although these might represent false negative discoveries, on the basis of our experience in *S. cerevisiae* we expect the more likely explanation is that these factors are not strictly required for efficient splicing of these specific introns under the conditions tested. Interestingly, significant overrepresentation of components of the SWR1 nucleosome remodeling complex was also uncovered (p<6.75e-3), consistent with previous reports describing the role of SWR1 components in splicing (Patrick *et al.* 2015). Other GO categories that are well represented in the list of significant genes include 'transcription from polymerase II promoter', 'mRNA cleavage and polyadenylation specificity factor complex', and 'chromatin remodeling' (Figure 3).

Several of the genes identified here belong to seemingly unrelated GO categories. It seems important to reiterate that the approach implemented here doesn't measure splicing defects *per se*, but rather changes in the relative steady state levels of spliced and unspliced isoforms. As such, while some of these candidates may represent false positive discoveries, it seems likely that many are true positives which impact splicing isoform abundances either through non-splicing related pathways, or by modulating the activity of *bona fide* splicing factors. For example, deletions of either *ski2* or *trs130* resulted in some of the most significant increases in SI for

either of the tested splicing events. Ski2 is an RNA helicase and member of the SKI complex, a highly conserved complex necessary for 3′ to 5′ degradation of transcripts subject to the nonsense-mediated decay (NMD) pathway (Mitchell and Tollervey 2003). The unspliced isoforms of *fet5* and *pwi1* contain premature stop codons and would be predicted targets of the NMD pathway, providing a plausible explanation for their accumulation in the Δ*ski2* strain. The *trs130* gene, by contrast, is involved in vesicle transport from the endoplasmic reticulum; the mechanism by which it might relate to altered splice isoform abundances is less clear. While no physical interactions have been documented between Trs130 and splicing-related proteins, epistatic genetic interactions between *trs130* and essential *bona-fide* splicing factors have been documented (Ryan *et al.* 2012).

To better understand the evolutionary nature of the genes that we identified, we examined whether homologs could be identified in either *S. cerevisiae* or humans. Of the 61 candidates we identified, 17 have clear human homologs but appear to lack an *S. cerevisiae* homolog, underscoring the potential that *S. pombe* provides as a model system for understanding the complex splicing seen in mammalian systems. Four of these genes, *cay1, cwf18, cwf19,* and *pwi1* have previously been annotated as splicing factors, while two others, *SPAC20H4.06c* and *SPBC713.05*, have been implicated in the splicing pathway on the basis of homology to human counterparts. Here, we provide experimental evidence that these protein products functionally impact pre-mRNA splicing.

Known splicing factors identified here display global splicing defects.

To better understand the impact of the genes identified here, splicing sensitive microarrays were used to determine the global changes in pre-mRNA splicing that result from

their deletions. These microarrays contain probes that target an exonic region of every protein coding gene in the *S. pombe* genome, as well as probes targeting every intron and its corresponding exon-exon junction, allowing for measurements of changes in total expression, pre-mRNA, and mature mRNA levels, respectively (Figure 4A).

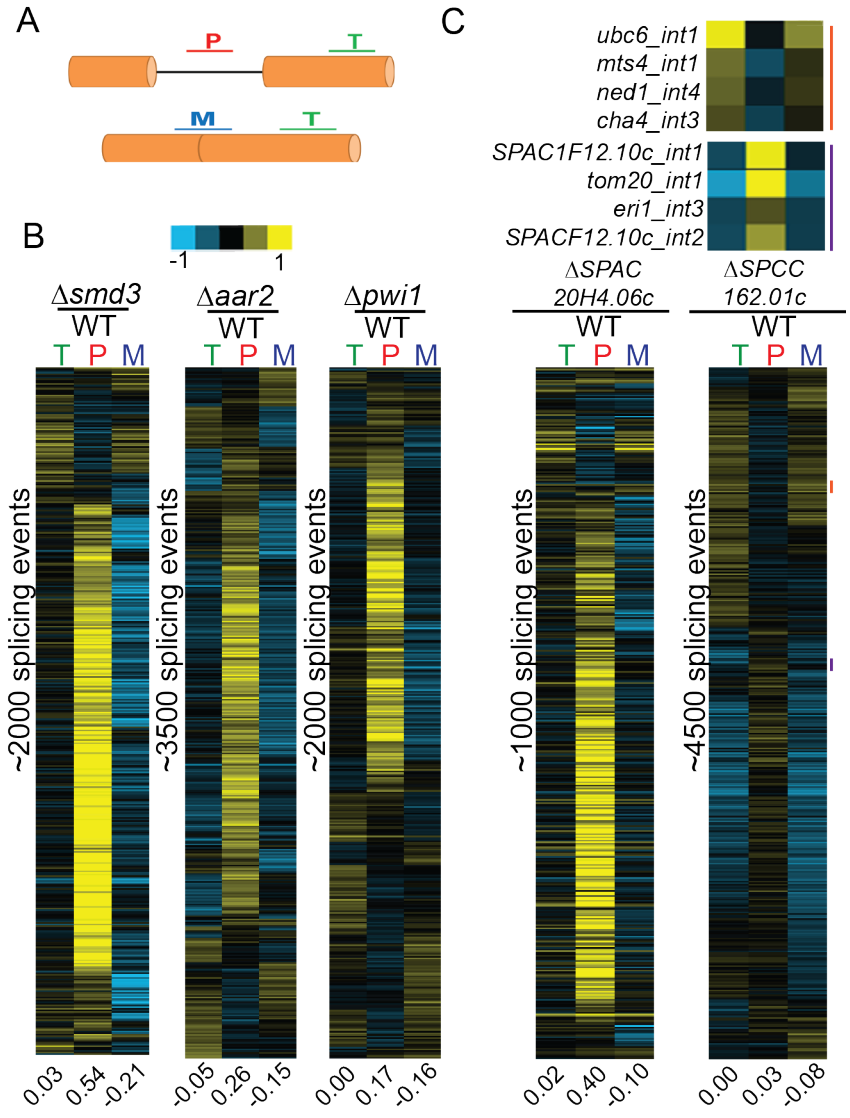


Figure 4: Known and predicted splicing factors display global splicing defects. (A) Splicing sensitive microarrays contain probes for quantification of total (T), pre-mRNA (P), and mature (M) mRNA levels. (B) Deletion of known splicing factors *smd3*, *aar2*, and *pwi1* each display global splicing defects. Each row represents the relative measurements for total, pre-mRNA, and mature mRNA for a particular splicing event. Numbers below each column represent the median value within the column. The data from each sample are

independently organized by hierarchical clustering and display only those events for which data were available for all three probes. (C) Global splicing phenotypes of the *ΔSPAC20H4.06c* and *ΔSPCC162.01c* strains. The orange and purple bars highlight specific splicing events showing decreases or increases in splicing efficiency, respectively.

As an initial test, we chose to examine strains harboring deletions in three known splicing factors: (1) *smd3*, a core component of the SM complex in the U1, U2, U4, and U5 snRNPs; (2) *aar2*, a component of the U5 snRNP; and (3) *pwi1*, a splicing co-activator. For both *smd3* and *aar2*, clear homologs exist in both *S. cerevisiae* and humans, and their specific roles in the splicing pathways have been well characterized (Gottschalk *et al.*; Nakazawa *et al.* 1991; Schwer and Shuman 2015). Moreover, the *Δsmd3* strain showed a statistically significant increase in our screen for both the *fet5*_intron1 and the *pwi1*_intron2 pre-mRNAs, while the *Δaar2* strain showed increased pre-mRNA levels for both transcripts, albeit just below our cutoff for statistical significance. The third gene, *pwi1*, is the homolog of the human SRRM1 gene, a member of the SR-like family of proteins (Graveley 2000). Unlike *smd3* and *aar2*, there is no homolog of *pwi1* in the *S. cerevisiae* genome. In our screen data, deletion of *pwi1* caused a statistically significant increase in the SI of the *fet5*_intron1. Deletion of any of these three genes resulted in global defects in pre-mRNA splicing, albeit each with unique properties (Figure 4B). For each of the mutants, increased levels of pre-mRNAs were detected for a majority of the events observed, and concomitant decreases were seen for many of the mature mRNA species, consistent with our expectations for a *bona fide* splicing mutant. Furthermore, similar to our screen data, the *Δaar2* strain showed levels of pre-mRNA accumulation that were overall lower than in the *Δsmd3* strain. Nevertheless, a similar number of splicing events was impaired by all three deletions.

Predicted splicing factors also display global splicing defects

Among the genes we identified in our screen whose deletions negatively impacted the splicing of either *fet5*_intron1 or *pwi1*_intron2 were several that are predicted based on homology studies to be involved in the splicing pathway. We chose to focus on two of these mutants, deletions of *SPAC20H4.06c*, a RNA-binding protein, and *SPCC162.01c*, a putative tri-snRNP component. Deletion of *SPAC20H4.06c* resulted in a statistically significant increase in both *fet5*_intron1 and *pwi1*_intron2 pre-mRNA levels, whereas deletion of *SPCC162.01c* also caused an increase in both pre-mRNAs, although just below our significance cutoff. As with *pwi1*, there are no apparent *S. cerevisiae* homologs for either *SPAC20H4.06c* or *SPCC162.01c*, but apparent human homologs do exist. For *SPAC20H4.06c*, the human homolog is GPATCH1, a member of the G-patch containing family of proteins. G-patch containing proteins have been previously implicated in splicing (Tsai *et al.* 2005a), yet no direct evidence appears to exist that specifically couples GPATCH1 to splicing. By contrast, the human homolog of *SPCC162.01c* is SNRNP27, a component of the U4/U6·U5 tri-snRNP complex and has a direct role in splicing (Fetzer *et al.* 1997). Interestingly, human SNRNP27 was previously shown to contain an N-terminal domain with strong homology to the RS domain of U170K; however, unlike U170K, SNRNP27 lacks an RNA-binding domain.

The global splicing defects of Δ*SPAC20H4.06c* and Δ*SPCC162.01c* revealed remarkably different phenotypes (Figure 4C). Deletion of *SPAC20H4.06c* showed a canonical splicing defect with broad increases in pre-mRNA species and decreases in mature mRNA species. The level to which pre-mRNAs accumulated is similar to that seen upon deletion of the canonical splicing factor *smd3* (Figure 3B). These data strongly suggest that the *SPAC20H4.06c* gene product is participating in the splicing pathway. By contrast, the global splicing profile resulting from

deletion of *SPCC162.01c* looked quite different from the other splicing mutants examined here. Whereas a subset of splicing events appeared to be negatively affected by deletion of *SPCC162.01c*, as evidenced by the accumulation of pre-mRNA and loss of mature mRNA, a nearly equal number of splicing events seemed to be positively, albeit weakly, affected by its deletion, with pre-mRNA levels decreasing and mature mRNA levels increasing for these transcripts. These results are consistent with a model where SR proteins can function to either enhance or repress splice site activation at different introns. These data also suggest that a large number of *S. pombe* introns are spliced at suboptimal efficiency in wild type cells. Additional experiments will be necessary to understand the mechanistic basis by which this SNRNP27 homolog can impart these phenotypes.

Genes involved in heterochromatin formation show a range of genome-wide splicing defects

In examining the list of candidates identified in our screen, we were struck by the number of components with known roles involved in RNA silencing and heterochromatin formation. It has long been known that RNA plays a critical role in silencing in *S. pombe* via the RNA-induced initiation of transcriptional gene silencing (RITS) complex (Verdel *et al.* 2004).While it has been suggested that splicing components facilitate RITS function (Bayne *et al.* 2008), it remains unclear whether these effects are direct or indirect (Kallgren *et al.* 2014). Two groups recently described purifications of two related complexes involved in silencing: MTREC, which is involved in assembling heterochromatin (Lee *et al.* 2013); and the Nuclear RNA Silencing complex, NURS (Egan *et al.* 2014). While these complexes each contain unique elements, they share in common both the essential RNA helicase Mtl1 and the non-essential zinc-finger protein Red1. In our work, deletion of *red1* resulted in a statistically significant decrease in the splicing

efficiency of both tested splicing events. In addition, affinity purification of Mtl1 as part of the MTREC complex co-purified Ctr1 (Lee *et al.* 2013), whereas affinity purification of Red1 as part of the NURS complex co-purified SPAC18G6.13 (Egan *et al.* 2014). In our screen, deletions of *ctr1* and *SPAC18G6.13* both resulted in statistically significant decreases in splicing efficiency of both target pre-mRNAs.

Ctr1 was previously implicated in splicing of TER1, the RNA component of telomerase (Lee *et al.* 2013). Moreover, Ctr1 had been shown to physically interact with components of the Prp19 complex, including Cwf10, Cwf11, and Prp19. To determine whether Ctr1 had a more global effect on the splicing pathway we again turned to microarray analysis. Deletion of *ctr1* resulted in a striking global splicing defect, strongly resembling that of a canonical splicing mutant (Figure 5A). A recent RNA-seq analysis of *Δctr1* and other MTREC mutants also revealed a global increase in unspliced transcript levels (Zhou *et al.* 2015). Interestingly, whereas our data reveal a broad decrease in mature mRNA concomitant with the increase in unspliced isoform, the Zhou *et al.* study reported largely unchanged levels of spliced transcripts. Owing at least in part to this observation, Zhou and colleagues proposed that Ctr1/MTREC plays a role in targeting unspliced transcripts to the nuclear exosome, and that the pre-mRNA accumulation phenotype of the Δ*ctr1* strain did not reflect a direct role for MTREC on splicing. The broad decreases in mature mRNA demonstrated by our microarray experiments are more consistent with a direct role for Ctr1 in the splicing pathway; additional experiments will be necessary to understand the discrepancy between these results, and the functional significance of Ctr1 in the pre-mRNA splicing pathway.

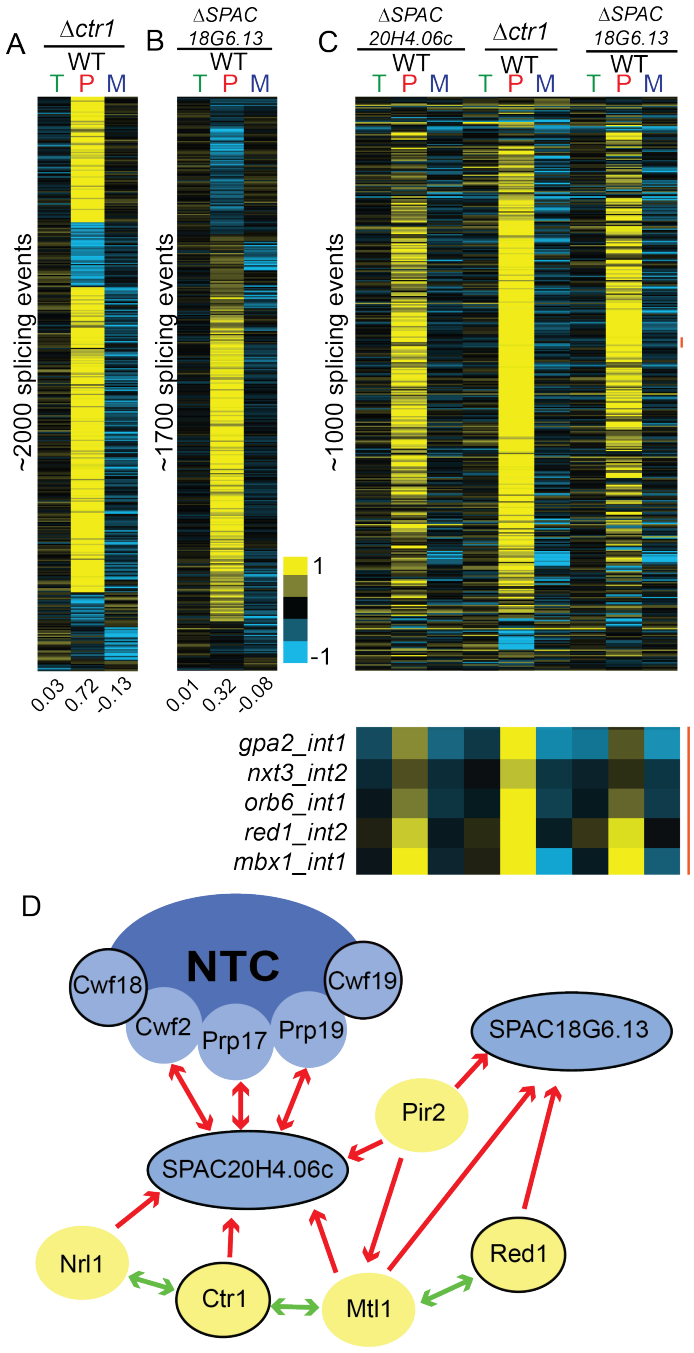


Figure 5. Deletion of factors involved in heterochromatin formation strongly impact global splicing. Splicing sensitive microarrays for *Δctr1* (A) and *ΔSPAC18G6.13* (B) reveal global splicing defects for each. Datasets for each mutant were organized independently using hierarchical clustering and display only those events for which data were available for all three probes. (C) A comparison of the splicing defects on common targets reveals a large overlap among all three of these deletion strains, with a subset of highlighted events marked by the orange bar. (D) Known physical interactions between several components of the silencing pathway and the splicing pathway. Red arrows indicate previously published one-way physical interactions. Green arrows

indicate two-way interaction. Blue ovals represent splicing factors, while yellow ovals represent members of the NURS and/or MTREC complexes. Black outlines note the components whose deletions caused splicing defects in this study.

By contrast with Ctr1, far less is known about the relationship between SPAC18G6.13 and splicing. Whereas physical interactions have been described between *SPAC18G6.13* and some splicing factors (Chen *et al.* 2014), the functional relevance of these interactions has not been previously described. Using microarray analysis, we showed that deletion of *SPAC18G6.13* also resulted in a broad increase in unspliced messages (Figure 5B). Interestingly, whereas SPAC18G6.13 was co-purified with Red1 as part of the NURS complex, the same study also demonstrated that Mtl1 co-purifies with SPAC20H4.06c, homolog of the human GPATCH1 gene described in the section above. When the global splicing defects of the *Δctr1*, *ΔSPAC18G6.13*, and *ΔSPAC20H4.06c* strains were analyzed together, the overlap in genome-wide splicing patterns was striking (Figure 5C). The physical interactions both between these proteins themselves, and with other components of the spliceosome as observed by others, as well as the splicing phenotypes we observed here in these mutants suggest that the functional relationship between splicing and heterochromatin formation may be more bi-directional than previously thought (Figure 5D).

In addition to the NURS complex, heterochromatic silencing is accomplished in part through cooperation between the RNAi machinery and the heterochromatic factors Clr4 and the histone variant H2A.Z (Zofall *et al.* 2009; Grewal 2010; Hou *et al.* 2010; Anver *et al.* 2014). While H2A.Z is generally thought of as a repressive mark, it is also associated with promoters and may play roles in recruiting RNAP II to genes (Zlatanova and Thakar 2008). Interestingly, among our list of mutants that affected splicing of our targets were *Δpht1*, the fission yeast homolog of H2A.Z, as well as many components of the INO80/SWR1complex, which is

responsible for catalyzing H2A/H2A.Z exchange (Morrison and Shen 2009), including *Δyaf9*, *Δies2*, *Δvps71,* and *Δswc2*. Similarly, while the Set1C complex is responsible for catalyzing the addition of H3K4me marks, it is also necessary for proper silencing of subtelomeric regions in *S. pombe* (Mikheyeva *et al.* 2014). Deletion of two components of the Set1complex, *ash2* and *swd1*, were identified in our screen as causing decreases in pre-mRNA splicing efficiency, although the *Δswd1* effect was just below our significance cutoff.

To determine the effect that loss of these heterochromatic factors have on the splicing pathway, we again assessed the global splicing profiles of the *Δpht1*, *Δash2*, and *Δswd1* strains by microarray. On the basis of these experiments alone, it is difficult to say whether deletion of any of these factors is impacting pre-mRNA splicing. While small groups of transcripts can be seen to exhibit a canonical splicing defect, the large changes in total gene expression, both increases and decreases, that are associated with these mutations complicates their analysis. Further studies will be necessary to better characterize the impact on pre-mRNA splicing of deletion of these genes.

3′ end processing factors impact the splicing of both terminal and non-terminal introns

Here we identified two genes involved in the cleavage and polyadenylation pathway, *ssu72* and *ppn1*, whose deletions resulted in pre-mRNA splicing defects. The 3′ end processing and splicing pathways have been previously demonstrated to be functionally coupled together (Kyburz *et al.* 2006; Millevoi *et al.* 2006). Components of the U2 snRNP co-purify with cleavage and polyadenylation specificity factor, CPSF, demonstrating a physical interaction between the two pre-mRNA processing pathways (Kyburz *et al.* 2006). In addition, CPSF is necessary for efficient splicing activity, while binding of the U2 snRNP promotes efficient

cleavage at the 3′ end (Kyburz *et al.* 2006). Importantly, Ppn1 and Ssu72 have been shown to physically interact with each other in *S. pombe* and to co-purify with the 3′end processing machinery (Vanoosthuyse *et al.* 2014). We determined the global splicing profiles of these two mutants using microarrays: deletion of both *ssu72* and *ppn1* resulted in pre-mRNA splicing defects for a large fraction of the events monitored (Figure 6A). The defects seen for these mutants was similar to those seen for deletion of the canonical splicing mutant *smd3*, both in terms of the number of transcripts affected and the levels of pre-mRNA accumulation.

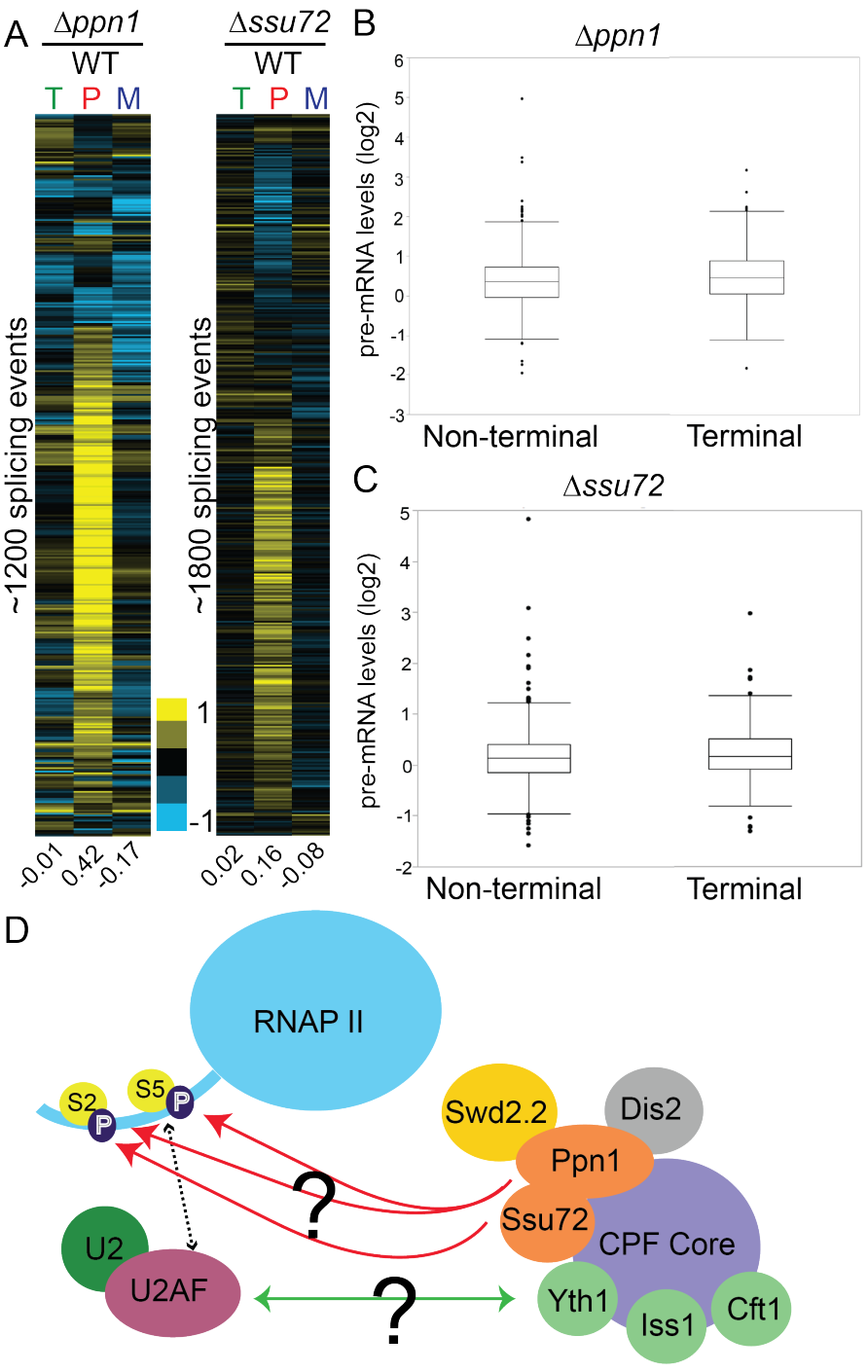


Figure 6. Deletions of 3′ end processing factors result in global splicing defects. Splicing sensitive microarrays for *Δppn1* (A) and *Δssu72* (B) strains each show broad splicing defects. Splicing events from each array were clustered independently using hierarchical clustering and displayed for only those events for which data were available for all three probe types. (C) The pre-mRNA levels of terminal and non-terminal introns within multi-intronic genes were compared for each mutant, revealing no obvious difference between their

behaviors. (D) From our data, it is unclear how the CPF factors Ppn1 and Ssu72 are impacting splicing. Here we depict two possible explanations: deletion of these factors could either prevent proper phosphorylation of the CTD tail and thus disrupt the interaction between U2AF and the CTD tail, or their absence from the CPF complex could disrupt the physical interactions between Yth1 and U2AF. Orange ovals represent CPF factors that cause significant splicing defects in our screen. Yellow ovals indicate factors that caused increases in pre-mRNA levels but were not significant. Green circles represent factors that have been shown in *S. cerevisiae* to cause splicing defects upon deletion.

In higher eukaryotes, where exons are short and introns can be extraordinarily long, spliceosome assembly is hypothesized to occur by exon definition, wherein recognition of a downstream 5′ splice site can facilitate recognition of an upstream 3′ splice site by cross-exon interactions. For terminal introns, where no downstream 5′ splice site exists, it has been demonstrated that components of the cleavage and polyadenylation machinery can serve to facilitate recognition of the terminal 3′ splice site in a process termed terminal exon definition. Although it has been thought that the short introns in yeast would not require cross-exon interactions for efficient splicing, several studies have demonstrated that components of the cleavage and polyadenylation machinery do impact pre-mRNA splicing in yeast (Albulescu *et al.* 2012; Baejen *et al.* 2014). Given the large number of multi-intronic genes in *S. pombe*, we sought to determine whether the extent to which the pre-mRNA increases detected in these 3′ end mutants were dependent upon the locations of the intron. Each intron in the genome was classified as being either the last annotated intron (terminal) or not the last (non-terminal). The pre-mRNA increases we observed for terminal introns was not obviously different than the increases seen for non-terminal introns, neither in the *Δssu72* nor the *Δppn1* strains (Figures 6B, 6C).

The mechanistic bases by which Ssu72 and PPn1 influence pre-mRNA splicing remain unclear. Although they are physically parts of the CPSF complex, both Ssu72 and Ppn1 are

phosphatases that target the CTD of RNA Pol II. Ssu72 preferentially targets the Ser5P modification (Rosado-Lugo and Hampsey 2014), while Ppn1 acts upon both Ser2P and Ser5P via the PP1 Nuclear Targeting Subunit (PNUTS) complex (Washington *et al.* 2002; Ciurciu *et al.* 2013). Phosphorylation of Ser5 is generally associated with promoter proximal pausing, and its dephosphorylation is important for escape into productive elongation (Rosado-Lugo and Hampsey 2014). The Ser5 mark of the CTD has been shown to be important for efficient splicing in yeast and humans, perhaps by slowing or pausing the polymerase so as to allow more time for co-transcriptional splicing to occur (Millhouse and Manley 2005; Alexander *et al.* 2010; Nojima *et al.* 2015).Given these roles for Ssu72 and Ppn1, the changes in splicing efficiency that accompany their deletions may not be a result of defects in cleavage and polyadenylation activity, *per se*, but rather changes in the CTD phosphorylation state. Alternatively, our understanding of the interactions between the cleavage and polyadenylation machinery and splicing may be incomplete, such that the interactions known to be important for terminal exon definition may in fact be important for general spliceosome assembly. In budding yeast, where introns are strongly biased towards the 5' end of transcripts, mutations in the endonuclease Brr5/Ysh1, ortholog of human CPSF-73, yield a strong splicing defect (Noble and Guthrie 1996), highlighting the capacity of *bona fide* 3' end processing factors to influence splicing at distances far removed from locations of cleavage and polyadenlation. Moreover, affinity capture and mass-spectrometry analysis of the *S. pombe* cleavage and polyadenylation factor complex reveals physical interactions between PPN1 and both of the SR-protein orthologs in the *S. pombe* genome (Vanoosthuyse *et al.* 2014). More experiments will be necessary to understand the mechanistic bases by which 3′ end processing and splicing impact one another in *S. pombe*.

Conclusion

Here we described the development and implementation of a sequencing-based reverse genetic screen to identify the complement of non-essential genes in the fission yeast *S. pombe* that impact pre-mRNA splicing. Our ability to positively identify both known and predicted splicing factors demonstrates the ability of this approach to identify splicing mutants among a collection of thousands of diverse strains. Moreover, the identification here of scores of factors previously unknown to impact splicing highlights the potential of this approach for *de novo* discovery. As with all genetic screens, further characterization of the individual factors identified here will be necessary to understand the mechanistic bases by which each of them impacts the splicing pathway. Nevertheless, the broad interconnectivity of RNA-processing pathways revealed in this work is testimony to the utility of *S. pombe* as a genetic system for studying these processes. Moreover, the recently solved EM structure of the *S. pombe* spliceosome (Yan *et al.* 2015) significantly enhances the ability of genetic data to inform about the mechanistic underpinnings of this process. Importantly, because many known components of the spliceosome are themselves essential, they have not been assayed in the screen described here. The creation of a temperature sensitive strain collection and subsequent screening using methods similar to those described here will present the opportunity to explore those essential genes, and thus provide greater insight into the mechanisms of more complex splicing.

Acknowledgements and Author Contributions

BJF designed sequencing assay and data processing scheme. AL and BJF performed screen. AL performed and analyzed microarray experiments. BJF designed figures 1-3. AL designed figures 4-6. BJF, AL, and JAP co-wrote chapter.

Works Cited

Alahari, S. K., H. Schmidt, and N. F. Käufer, 1993 The fission yeast prp4+ gene involved in pre-mRNA splicing codes for a predicted serine/threonine kinase and is essential for growth. Nucleic Acids Res. 21: 4079–4083.

Albulescu, L. O., N. Sabet, M. Gudipati, N. Stepankiw, Z. J. Bergman *et al.*, 2012 A quantitative, high-throughput reverse genetic screen reveals novel connections between pre-mRNA splicing and 5′ and 3′ end transcript determinants. PLoS Genet. 8: e1002530.

Alexander, R. D., S. a. Innocente, J. D. Barrass, and J. D. Beggs, 2010 Splicing-Dependent RNA polymerase pausing in yeast. Mol. Cell 40: 582–593.

Ameur, A., A. Zaghlool, J. Halvardson, A. Wetterbom, U. Gyllensten *et al.*, 2011 Total RNA sequencing reveals nascent transcription and widespread co-transcriptional splicing in the human brain. Nat. Struct. Mol. Biol. 18: 1435–1440.

Andersson, R., S. Enroth, A. Rada-Iglesias, C. Wadelius, and J. Komorowski, 2009 Nucleosomes are well positioned in exons and carry characteristic histone modifications. Genome Res. 19: 1732–41.

Anver, S., A. Roguev, M. Zofall, N. J. Krogan, S. I. S. Grewal *et al.*, 2014 Yeast X-chromosome-associated protein 5 (Xap5) functions with H2A.Z to suppress aberrant transcripts. EMBO Rep. 15: 894–902.

Araya, C. L., and D. M. Fowler, 2011 Deep mutational scanning: Assessing protein function on a massive scale. Trends Biotechnol. 29: 435–442.

Armstrong, J., N. Bone, J. Dodgson, and T. Beck, 2007 The role and aims of the FYSSION project. Briefings Funct. Genomics Proteomics 6: 3–7.

Awan, A. R., A. Manfredo, and J. a Pleiss, 2013 Lariat sequencing in a unicellular yeast

identifies regulated alternative splicing of exons that are evolutionarily conserved with humans. Proc. Natl. Acad. Sci. U. S. A. 110: 12762–7.

Baejen, C., P. Torkler, S. Gressel, K. Essig, J. Söding *et al.*, 2014 Transcriptome maps of mRNP biogenesis factors define pre-mRNA recognition. Mol Cell 55: 745–757.

Bansal, V., 2010 A statistical method for the detection of variants from next-generation resequencing of DNA pools. Bioinformatics 26:.

Barbosa-Morais, N. L., M. Irimia, Q. Pan, H. Y. Xiong, S. Gueroussov *et al.*, 2012 The evolutionary landscape of alternative splicing in vertebrate species. Science (80-.). 338: 1587–1593.

Bayne, E. H., D. a Bijos, S. a White, F. de Lima Alves, J. Rappsilber *et al.*, 2014 A systematic genetic screen identifies new factors influencing centromeric heterochromatin integrity in fission yeast. Genome Biol. 15: 481.

Bayne, E. H., M. Portoso, A. Kagansky, I. C. Kos-Braun, T. Urano *et al.*, 2008 Splicing factors facilitate RNAi-directed silencing in fission yeast. Science 322: 602–606.

Berget, S. M., 1995 Minireviews : Exon Recognition in Vertebrate Splicing. 2411–2414.

Bitton, D. A., S. R. Atkinson, C. Rallis, G. C. Smith, D. A. Ellis *et al.*, 2015 Widespread exon-skipping triggers degradation by nuclear RNA surveillance in fission yeast. Genome Res. 884–896.

Blencowe, B. J., G. Baurén, A. G. Eldridge, R. Issner, J. A. Nickerson *et al.*, 2000 The SRm160/300 splicing coactivator subunits. RNA 6: 111–20.

Blomquist, T. M., E. L. Crawford, J. L. Lovett, J. Yeo, L. M. Stanoszek *et al.*, 2013 Targeted RNA-Sequencing with Competitive Multiplex-PCR Amplicon Libraries. PLoS One 8: e79120.

Boon, K.-L., T. Auchynnikava, G. Edwalds-Gilbert, J. D. Barrass, A. P. Droop *et al.*, 2006 Yeast Ntr1/Spp382 Mediates Prp43 Function in Postspliceosomes. Mol. Cell. Biol.

Booth, G. T., I. X. Wang, V. G. Cheung, and J. T. Lis, 2016 Divergence of a conserved elongation factor and transcription regulation in budding and fission yeast. Genome Res. 26: 799–811.

Bradley, R. K., J. Merkin, N. J. Lambert, and C. B. Burge, 2012 Alternative splicing of RNA triplets is often regulated and accelerates proteome evolution. PLoS Biol. 10: e1001229.

Burke, J., A. Longhurst, D. Merkurjev, J. Sales-Lee, B. Rao *et al.*, 2018 Spliceosome profiling visualizes the operations of a dynamic RNP in vivo at nucleotide resolution. Cell 173: 1014–1030.

Carey, M. F., C. L. Peterson, and S. T. Smale, 2013 The primer extension assay. Cold Spring Harb. Protoc. 8: 164–173.

Carrillo Oesterreich, F., S. Preibisch, and K. M. Neugebauer, 2010 Global analysis of nascent rna reveals transcriptional pausing in terminal exons. Mol. Cell 40: 571–581.

Cazzola, M., M. Rossi, and L. Malcovati, 2013 Biologic and clinical significance of somatic mutations of SF3B1 in myeloid and lymphoid neoplasms. Blood.

Chanfreau, G., P. Legrain, B. Dujon, and A. Jacquier, 1994 Interaction between the first and last nucleotides of pre-mRNA introns is a determinant of 3′ splice site selection in S.cerevisiae. Nucleic Acids Res.

Chen, W., J. Moore, H. Ozadam, H. P. Shulha, N. Rhind *et al.*, 2018 Transcriptome-wide interrogation of the functional intronome by spliceosome profiling. Cell 173: 1031–1044.

Chen, W., H. P. Shulha, A. Ashar-Patel, J. Yan, K. M. Green *et al.*, 2014 Endogenous U2·U5·U6 snRNA complexes in S. pombe are intron lariat spliceosomes. RNA 20: 308–20.

Ciurciu, A., L. Duncalf, V. Jonchere, N. Lansdale, O. Vasieva *et al.*, 2013 PNUTS/PP1 regulates RNAPII-mediated gene expression and is necessary for developmental growth. PLoS Genet. 9: e1003885.

Clark, T. a, C. W. Sugnet, and M. Ares, 2002 Genomewide analysis of mRNA processing in yeast using splicing-specific microarrays. Science 296: 907–910.

Collart, M. A., and S. Oliviero, 2001 Preparation of Yeast RNA, in *Current Protocols in Molecular Biology*,.

Conesa, A., P. Madrigal, S. Tarazona, D. Gomez-Cabrero, A. Cervera *et al.*, 2016 A survey of best practices for RNA-seq data analysis. Genome Biol. 17:.

De Conti, L., M. Baralle, and E. Buratti, 2013 Exon and intron definition in pre-mRNA splicing. Wiley Interdiscip. Rev. RNA 4: 49–60.

Coombes, C. E., and J. D. Boeke, 2005 An evaluation of detection methods for large lariat RNAs. RNA 11: 323–31.

Core, L. J., and J. T. Lis, 2008 Transcription regulation through promoter-proximal pausing of RNA polymerase II. Science 319: 1791–1792.

Crooks, G. E., G. Hon, J. M. Chandonia, and S. E. Brenner, 2004 WebLogo: A sequence logo generator. Genome Res. 14: 1188–1190.

David, C. J., and J. L. Manley, 2011 The RNA polymerase C-terminal domain: A new role in spliceosome assembly. Transcription 2: 221–225.

Depristo, M. A., E. Banks, R. Poplin, K. V. Garimella, J. R. Maguire *et al.*, 2011 A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat. Genet. 43: 491–501.

Dietrich, R. C., R. Incorvaia, and R. A. Padgett, 1997 Terminal intron dinucleotide sequences do

not distinguish between U2- and U12-dependent introns. Mol. Cell.

Dobin, A., C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski *et al.*, 2013 STAR: Ultrafast universal RNA-seq aligner. Bioinformatics 29: 15–21.

Dujardin, G., C. Lafaille, M. de la Mata, L. E. Marasco, M. J. Muñoz *et al.*, 2014 How Slow RNA Polymerase II Elongation Favors Alternative Exon Skipping. Mol. Cell 54: 683–690.

Dunn, E. A., and S. D. Rader, 2014 Preparation of yeast whole cell splicing extract. Methods Mol. Biol.

Egan, E. D., C. R. Braun, S. P. Gygi, and D. Moazed, 2014 Post-transcriptional regulation of meiotic genes by a nuclear RNA silencing complex. RNA 20: 867–81.

Eldridge, A. G., Y. Li, P. A. Sharp, and B. J. Blencowe, 1999 The SRm160/300 splicing coactivator is required for exon-enhancer function. Proc. Natl. Acad. Sci. U. S. A. 96: 6125–6130.

Engel, S. R., F. S. Dietrich, D. G. Fisk, G. Binkley, R. Balakrishnan *et al.*, 2014 The reference genome sequence of Saccharomyces cerevisiae: then and now. G3 (Bethesda). 4: 389–98.

Eser, P., L. Wachutka, K. C. Maier, C. Demel, M. Boroni *et al.*, 2016 Determinants of RNA metabolism in the Schizosaccharomyces pombe genome. Mol. Syst. Biol. 12: 857–857.

Evsyukova, I., S. S. Bradrick, S. G. Gregory, and M. a Garcia-Blanco, 2013 Cleavage and polyadenylation specificity factor 1 (CPSF1) regulates alternative splicing of interleukin 7 receptor (IL7R) exon 6. RNA 19: 103–15.

Fair, B. J., and J. A. Pleiss, 2017 The power of fission: yeast as a tool for understanding complex splicing. Curr. Genet.

Fetzer, S., J. Lauber, C. L. Will, and R. Lührmann, 1997 The [U4/U6.U5] tri-snRNP-specific 27K protein is a novel SR protein that can be phosphorylated by the snRNP-associated

protein kinase. RNA 3: 344–55.

Forsburg, S. L., and N. Rhind, 2006 Basic methods for fission yeast. Yeast 23: 173–183.

Fowler, D. M., and S. Fields, 2014 Deep mutational scanning: a new style of protein science. Nat. Methods 11: 801–7.

Galej, W. P., T. H. D. Nguyen, A. J. Newman, and K. Nagai, 2014 Structural studies of the spliceosome: Zooming into the heart of the machine. Curr. Opin. Struct. Biol.

Galej, W. P., C. Oubridge, A. J. Newman, and K. Nagai, 2013 Crystal structure of Prp8 reveals active site cavity of the spliceosome. Nature 493: 638–43.

Gao, J., F. Kan, J. L. Wagnon, A. J. Storey, R. U. Protacio *et al.*, 2014 Rapid, efficient and precise allele replacement in the fission yeast Schizosaccharomyces pombe. Curr. Genet. 60: 109–119.

Gao, K., A. Masuda, T. Matsuura, and K. Ohno, 2008 Human branch point consensus sequence is yUnAy. Nucleic Acids Res. 36: 2257–2267.

Gottschalk, A., B. Kastner, R. Lührmann, and P. Fabrizio The yeast U5 snRNP coisolated with the U1 snRNP has an unexpected protein composition and includes the splicing factor Aar2p.

Grate, L., and M. Ares, 2002 Searching yeast intron data at Ares lab web site. Methods Enzymol. 350: 380–392.

Graveley, B. R., 2000 Sorting out the complexity of SR protein functions. RNA 6: 1197–1211.

Grewal, S. I. S., 2010 RNAi-dependent formation of heterochromatin and its diverse functions. Curr. Opin. Genet. Dev. 20: 134–141.

Habara, Y., S. Urushiyama, T. Tani, and Y. Ohshima, 1998 The fission yeast prp10(+) gene involved in pre-mRNA splicing encodes a homologue of highly conserved splicing factor,

SAP155. Nucleic Acids Res. 26: 5662–5669.

Haraguchi, N., T. Andoh, D. Frendewey, and T. Tani, 2007 Mutations in the SF1-U2AF59-U2AF23 complex cause exon skipping in Schizosaccharomyces pombe. J. Biol. Chem. 282: 2221–2228.

Hartwell, L. H., C. S. McLaughlin, and J. R. Warner, 1970 Identification of ten genes that control ribosome formation in yeast. MGG Mol. Gen. Genet. 109: 42–56.

Hicks, M. J., B. J. Lam, and K. J. Hertel, 2005 Analyzing mechanisms of alternative pre-mRNA splicing using in vitro splicing assays. Methods.

Hollander, D., S. Naftelberg, G. Lev-Maor, A. R. Kornblihtt, and G. Ast, 2016 How Are Short Exons Flanked by Long Introns Defined and Committed to Splicing? Trends Genet. xx: 1–11.

Hossain, M. A., and T. L. Johnson, 2014a Spliceosomal Pre-mRNA Splicing (K. J. Hertel, Ed.). 1126: 285–298.

Hossain, M. A., and T. L. Johnson, 2014b Using yeast genetics to study splicing mechanisms. Methods Mol. Biol.

Hou, H., Y. Wang, S. P. Kallgren, J. Thompson, J. R. Yates *et al.*, 2010 Histone variant H2A.Z regulates centromere silencing and chromosome segregation in fission yeast. J. Biol. Chem. 285: 1909–1918.

Huang, T., J. Vilardell, and C. C. Query, 2002 Pre-spliceosome formation in S.pombe requires a stable complex of SF1-U2AF59-U2AF23. EMBO J. 21: 5516–5526.

Iannone, C., A. Pohl, P. Papasaikas, D. Soronellas, G. P. Vicent *et al.*, 2015 Relationship between nucleosome positioning and progesterone-induced alternative splicing in breast cancer cells. RNA 21: 360–374.

Inada, M., and J. a Pleiss, 2010 *Genome-wide approaches to monitor pre-mRNA splicing.* Elsevier Inc.

James Kent, W., C. W. Sugnet, T. S. Furey, K. M. Roskin, T. H. Pringle *et al.*, 2002 The human genome browser at UCSC. Genome Res. 12: 996–1006.

Jamieson, D. J., B. Rahe, J. Pringle, and J. D. Beggs, 1991 A suppressor of a yeast splicing mutation (prp8-1) encodes a putative ATP-dependent RNA helicase. Nature.

Juneau, K., C. Nislow, and R. W. Davis, 2009 Alternative splicing of PTC7 in Saccharomyces cerevisiae determines protein localization. Genetics 183: 185–194.

Kallgren, S. P., S. Andrews, X. Tadeo, H. Hou, J. J. Moresco *et al.*, 2014 The Proper Splicing of RNAi Factors Is Critical for Pericentric Heterochromatin Assembly in Fission Yeast. PLoS Genet. 10: 1–11.

Käufer, N. F., and J. Potashkin, 2000 Analysis of the splicing machinery in fission yeast: a comparison with budding yeast and mammals. Nucleic Acids Res. 28: 3003–3010.

Kawashima, T., S. Douglass, J. Gabunilas, M. Pellegrini, and G. F. Chanfreau, 2014 Widespread Use of Non-productive Alternative Splice Sites in Saccharomyces cerevisiae. PLoS Genet. 10:.

Khodor, Y. L., J. Rodriguez, K. C. Abruzzi, C. H. A. Tang, M. T. Marr *et al.*, 2011 Nascent-seq indicates widespread cotranscriptional pre-mRNA splicing in Drosophila. Genes Dev. 25: 2502–2512.

Kielkopf, C. L., N. A. Rodionova, M. R. Green, and S. K. Burley, 2001 A novel peptide recognition mode revealed by the X-ray structure of a core U2AF35/U2AF65 heterodimer. Cell 106: 595–605.

Kim, D.-U., J. Hayles, D. Kim, V. Wood, H.-O. Park *et al.*, 2010 Analysis of a genome-wide set

of gene deletions in the fission yeast Schizosaccharomyces pombe. Nat. Biotechnol. 28: 617–623.

Kim, D., B. Langmead, and S. L. Salzberg, 2015 HISAT: A fast spliced aligner with low memory requirements. Nat. Methods 12: 357–360.

Kim, S. H., and R. J. Lin, 1993 Pre-mRNA splicing within an assembled yeast spliceosome requires an RNA-dependent ATPase and ATP hydrolysis. Proc. Natl. Acad. Sci. U. S. A.

Kim, S. H., and R. J. Lin, 1996 Spliceosome activation by PRP2 ATPase prior to the first transesterification reaction of pre-mRNA splicing. Mol. Cell. Biol. 16: 6810–6819.

Kivioja, T., A. Vähärautio, K. Karlsson, M. Bonke, M. Enge *et al.*, 2012 Counting absolute numbers of molecules using unique molecular identifiers. Nat. Methods 9: 72–74.

Kuhn, A. N., and N. F. Käufer, 2003 Pre-mRNA splicing in Schizosaccharomyces pombe: regulatory role of a kinase conserved from fission yeast to mammals. Curr. Genet. 42: 241–251.

Kupfer, D. M., S. D. Drabenstot, K. L. Buchanan, H. Lai, H. Zhu *et al.*, 2004 Introns and splicing elements of five diverse fungi. Eukaryot. Cell 3: 1088–1100.

Kyburz, A., A. Friedlein, H. Langen, and W. Keller, 2006 Direct Interactions between Subunits of CPSF and the U2 snRNP Contribute to the Coupling of Pre-mRNA 3??? End Processing and Splicing. Mol. Cell 23: 195–205.

de la Mata, M., C. R. Alonso, S. Kadener, J. P. Fededa, M. Blaustein *et al.*, 2003 A Slow RNA Polymerase II Affects Alternative Splicing In Vivo. Mol. Cell 12: 525–532.

Larson, A., B. J. Fair, and J. A. Pleiss, 2016 Interconnections Between RNA-Processing Pathways Revealed by a Sequencing-Based Genetic Screen for Pre-mRNA Splicing Mutants in Fission Yeast. G3 (Bethesda). 6: 1513–23.

Lee, N. N., V. R. Chalamcharla, F. Reyes-Turcu, S. Mehta, M. Zofall *et al.*, 2013 Mtr4-like protein coordinates nuclear RNA processing for heterochromatin assembly and for telomere maintenance. Cell 155: 1–14.

Lee, Y., and D. C. Rio, 2015 Mechanisms and Regulation of Alternative Pre-mRNA Splicing. Annu. Rev. Biochem. 1–33.

Li, B., M. Carey, and J. L. Workman, 2007 The Role of Chromatin during Transcription. Cell 128: 707–719.

Li, H., and R. Durbin, 2009 Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25: 1754–60.

Libri, D., N. Graziani, C. Saguez, and J. Boulay, 2001 Multiple roles for the yeast SUB2/yUAP56 gene in splicing. Genes Dev.

Lim, L. P., and C. B. Burge, 2001 A computational analysis of sequence features involved in recognition of short introns. Proc. Natl. Acad. Sci. U. S. A. 98: 11193–8.

Lin, R. J., A. J. Lustig, and J. Abelson, 1987 Splicing of yeast nuclear pre-mRNA in vitro requires a functional 40S spliceosome and several extrinsic factors. Genes Dev.

Lin, P. C., and R. M. Xu, 2012 Structure and assembly of the SF3a splicing factor complex of U2 snRNP. EMBO J.

Lipp, J. J., M. C. Marvin, K. M. Shokat, and C. Guthrie, 2015 SR protein kinases promote splicing of nonconsensus introns. Nat. Struct. Mol. Biol. 22: 611–617.

Liu, H.-L., and S.-C. Cheng, 2012 The Interaction of Prp2 with a Defined Region of the Intron Is Required for the First Splicing Reaction. Mol. Cell. Biol.

Long, J. C., and J. F. Caceres, 2009 The SR protein family of splicing factors: master regulators of gene expression. Biochem. J. 417: 15–27.

Love, M. I., W. Huber, and S. Anders, 2014 Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15: 550.

Lucks, J. B., S. A. Mortimer, C. Trapnell, S. Luo, S. Aviran *et al.*, 2011 Multiplexed RNA structure characterization with selective 2'-hydroxyl acylation analyzed by primer extension sequencing (SHAPE-Seq). Proc. Natl. Acad. Sci. 108: 11063–11068.

Luco, R. F., M. Allo, I. E. Schor, A. R. Kornblihtt, and T. Misteli, 2011 Epigenetics in alternative pre-mRNA splicing. Cell 144: 16–26.

Lustig, a J., R. J. Lin, and J. Abelson, 1986 The yeast RNA gene products are essential for mRNA splicing in vitro. Cell 47: 953–63.

Lybarger, S., K. Beickman, V. Brown, N. Dembla-Rajpal, K. Morey *et al.*, 1999 Elevated levels of a U4/U6.U5 snRNP-associated protein, Spp381p, rescue a mutant defective in spliceosome maturation. Mol. Cell. Biol.

Mamanova, L., A. J. Coffey, C. E. Scott, I. Kozarewa, E. H. Turner *et al.*, 2010 Target-enrichment strategies for next- generation sequencing. Nat. Methods 7: 111–118.

Marshall, A. N., M. C. Montealegre, C. Jim??nez-L??pez, M. C. Lorenz, and A. van Hoof, 2013 Alternative Splicing and Subfunctionalization Generates Functional Diversity in Fungal Proteomes. PLoS Genet. 9:.

Matera, A. G., and Z. Wang, 2014 A day in the life of the spliceosome. Nat. Rev. Mol. Cell Biol.

Matlin, A. J., F. Clark, and C. W. J. Smith, 2005 Understanding alternative splicing: towards a cellular code. Nat. Rev. Mol. Cell Biol. 6: 386–98.

Mayas, R. M., H. Maita, and J. P. Staley, 2006 Exon ligation is proofread by the DExD/H-box ATPase Prp22p. Nat. Struct. Mol. Biol.

Mayerle, M., M. Raghavan, S. Ledoux, A. Price, N. Stepankiw *et al.*, 2017 Structural toggle in

the RNaseH domain of Prp8 helps balance splicing fidelity and catalytic efficiency. Proc. Natl. Acad. Sci. 114: 4739–4744.

Mazroui, R., a Puoti, and a Krämer, 1999 Splicing factor SF1 from Drosophila and Caenorhabditis: presence of an N-terminal RS domain and requirement for viability. RNA 5: 1615–31.

McLaren, W., L. Gil, S. E. Hunt, H. S. Riat, G. R. S. Ritchie *et al.*, 2016 The Ensembl Variant Effect Predictor. Genome Biol. 17:.

Mercer, T. R., M. B. Clark, J. Crawford, M. E. Brunck, D. J. Gerhardt *et al.*, 2014 Targeted sequencing for gene discovery and quantification using RNA CaptureSeq. Nat. Protoc. 9: 989–1009.

Mercer, T. R., D. J. Gerhardt, M. E. Dinger, J. Crawford, C. Trapnell *et al.*, 2011 Targeted RNA sequencing reveals the deep complexity of the human transcriptome. Nat. Biotechnol. 30: 99–104.

Merkin, J., C. Russell, P. Chen, and C. B. Burge, 2012 Evolutionary dynamics of gene and isoform regulation in Mammalian tissues. Science 338: 1593–9.

Meyer, M., and J. Vilardell, 2009 The quest for a message: budding yeast, a model organism to study the control of pre-mRNA splicing. Brief. Funct. Genomic. Proteomic. 8: 60–7.

Mikheyeva, I. V, P. J. R. Grady, F. B. Tamburini, D. R. Lorenz, and H. P. Cam, 2014 Multifaceted genome control by Set1 Dependent and Independent of H3K4 methylation and the Set1C/COMPASS complex. PLoS Genet. 10: e1004740.

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

Millhouse, S., and J. L. Manley, 2005 The C-terminal domain of RNA polymerase II functions as a phosphorylation-dependent splicing activator in a heterologous protein. Mol. Cell. Biol. 25: 533–544.

Misteli, T., and D. L. Spector, 1999 RNA polymerase II targets pre-mRNA splicing factors to transcription sites in vivo. Mol. Cell 3: 697–705.

Mitchell, P., and D. Tollervey, 2003 An NMD pathway in yeast involving accelerated deadenylation and exosome-mediated 3??????5??? degradation. Mol. Cell 11: 1405–1413.

Morris, D. P., and a. L. Greenleaf, 2000 The splicing factor, Prp40, binds the phosphorylated carboxyl-terminal domain of RNA Polymerase II. J. Biol. Chem. 275: 39935–39943.

Morrison, A. J., and X. Shen, 2009 Chromatin remodelling beyond transcription: the INO80 and SWR1 complexes. Nat. Rev. Mol. Cell Biol. 10: 373–384.

Muñoz, M. J., M. de la Mata, and A. R. Kornblihtt, 2010 The carboxy terminal domain of RNA polymerase II and alternative splicing. Trends Biochem. Sci. 35: 497–504.

Nakazawa, N., S. Harashima, and Y. Oshima, 1991 AAR2, a gene for splicing pre-mRNA of the MATa1 cistron in cell type control of Saccharomyces cerevisiae. Mol. Cell. Biol. 11: 5693–5700.

Network, T. C. G. A., 2012 Comprehensive molecular portraits of human breast tumors. Nature.

Nguyen, T. H. D., W. P. Galej, X. Bai, C. G. Savva, A. J. Newman *et al.*, 2015 The architecture of the spliceosomal U4/U6.U5 tri-snRNP. Nature 523: 47–52.

Nguyen, T. H. D., W. P. Galej, S. M. Fica, P. C. Lin, A. J. Newman *et al.*, 2016 CryoEM structures of two spliceosomal complexes: Starter and dessert at the spliceosome feast. Curr. Opin. Struct. Biol. 36: 48–57.

Noble, S. M., and C. Guthrie, 1996 Identification of novel genes required for yeast pre-mRNA

splicing by means of cold-sensitive mutations. Genetics 143: 67–80.

Nojima, T., T. Gomes, A. R. F. Grosso, H. Kimura, M. J. Dye *et al.*, 2015 Mammalian NET-Seq Reveals Genome-wide Nascent Transcription Coupled to RNA Processing. Cell 161: 526–540.

Noma, T., K. Sode, and K. Ikebukuro, 2006 Characterization and application of aptamers for Taq DNA polymerase selected using an evolution-mimicking algorithm. Biotechnol. Lett. 28: 1939–1944.

Ohi, M. D., A. J. Link, L. Ren, J. L. Jennings, W. H. McDonald *et al.*, 2002 Proteomics analysis reveals stable multiprotein complexes in both fission and budding yeasts containing Myb-related Cdc5p/Cef1p, novel pre-mRNA splicing factors, and snRNAs. Mol. Cell. Biol. 22: 2011–2024.

Padgett, R. A., M. M. Konarska, M. Aebi, H. Hornig, C. Weissmann *et al.*, 1985 Nonconsensus branch-site sequences in the in vitro splicing of transcripts of mutant rabbit beta-globin genes. Proc. Natl. Acad. Sci. U. S. A. 82: 8349–8353.

Pandit, S., Y. Zhou, L. Shiue, G. Coutinho-Mansfield, H. Li *et al.*, 2013 Genome-wide Analysis Reveals SR Protein Cooperation and Competition in Regulated Splicing. Mol. Cell 50: 223–235.

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

Parker, R., and P. G. Siliciano, 1993 Evidence for an essential non-Watson-Crick interaction between the first and last nucleotides of a nuclear pre-mRNA intron. Nature.

Patrick, K. L., C. J. Ryan, J. Xu, J. J. Lipp, K. E. Nissen *et al.*, 2015 Genetic Interaction Mapping Reveals a Role for the SWI/SNF Nucleosome Remodeler in Spliceosome Activation in

Fission Yeast. PLoS Genet. 11: e1005074.

Picelli, S., A. K. Björklund, B. Reinius, S. Sagasser, G. Winberg *et al.*, 2014 Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. Genome Res. 24: 2033–2040.

Potashkin, J., R. Li, and D. Frendewey, 1989 Pre-mRNA splicing mutants of Schizosaccharomyces pombe. EMBO J. 8: 551–9.

Qin, D., L. Huang, A. Wlodaver, J. Andrade, and J. P. Staley, 2016 Sequencing of lariat termini in S. cerevisiae reveals 5' splice sites, branch points, and novel splicing events. RNA.

Quesada, V., A. J. Ramsay, and C. Lopez-Otin, 2012 Chronic lymphocytic leukemia with SF3B1 mutation. N. Engl. J. Med.

Quinlan, A. R., and I. M. Hall, 2010 BEDTools: A flexible suite of utilities for comparing genomic features. Bioinformatics 26: 841–842.

Ram, O., and G. Ast, 2007a SR proteins: a foot on the exon before the transition from intron to exon definition. Trends Genet. 23: 5–7.

Ram, O., and G. Ast, 2007b SR proteins: a foot on the exon before the transition from intron to exon definition. Trends Genet. 23: 5–7.

Ramírez, F., D. P. Ryan, B. Grüning, V. Bhardwaj, F. Kilpert *et al.*, 2016 deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res. 44: W160–W165.

Rauhut, R., P. Fabrizio, O. Dybkov, K. Hartmuth, V. Pena *et al.*, 2016 Molecular architecture of the Saccharomyces cerevisiae activated spliceosome. Science (80-.).

Ren, L., J. R. McLean, T. R. Hazbun, S. Fields, C. Vander Kooi *et al.*, 2011 Systematic Two-Hybrid and Comparative Proteomic Analyses Reveal Novel Yeast Pre-mRNA Splicing

Factors Connected to Prp19. PLoS One 6: e16719.

Rhind, N., Z. Chen, M. Yassour, D. a Thompson, B. J. Haas *et al.*, 2011 Comparative functional genomics of the fission yeasts. Science 332: 930–6.

Robberson, B. L., G. J. Cote, and S. M. Berget, 1990 Exon definition may facilitate splice site selection in RNAs with multiple exons. Mol. Cell. Biol. 10: 84–94.

Romfo, C. M., C. J. Alvarez, W. J. van Heeckeren, C. J. Webb, and J. a Wise, 2000 Evidence for splice site pairing via intron definition in Schizosaccharomyces pombe. Mol. Cell. Biol. 20: 7955–7970.

Rosado-Lugo, J. D., and M. Hampsey, 2014 The Ssu72 phosphatase mediates the RNA polymerase II initiation-elongation transition. J. Biol. Chem. 289: 33916–26.

Rosenberg, G. H., S. K. Alahari, and N. F. Käufer, 1991 prp4 from Schizosaccharomyces pombe, a mutant deficient in pre-mRNA splicing isolated using genes containing artificial introns. MGG Mol. Gen. Genet. 226: 305–309.

Roshbash, M., P. K. Harris, J. L. Woolford, and J. L. Teem, 1981 The effect of temperature-sensitive RNA mutants on the transcription products from cloned ribosomal protein genes of yeast. Cell 24: 679–686.

Rouskin, S., M. Zubradt, S. Washietl, M. Kellis, and J. S. Weissman, 2014 Genome-wide probing of RNA structure reveals active unfolding of mRNA structures in vivo. Nature 505: 701–5.

Ryan, C. J., A. Roguev, K. Patrick, J. Xu, H. Jahari *et al.*, 2012 Hierarchical Modularity and the Evolution of Genetic Interactomes across Species. Mol. Cell 46: 691–704.

Sasaki-Haraguchi, N., T. Ikuyama, S. Yoshii, T. Takeuchi-Andoh, D. Frendewey *et al.*, 2015 Cwf16p associating with the nineteen complex ensures ordered exon joining in constitutive

Pre-mRNA splicing in fission yeast. PLoS One 10: 1–16.

Scheckel, C., and R. B. Darnell, 2015 Microexons—Tiny but mighty. EMBO J. 34: 273–274.

Schellenberg, M. J., T. Wu, D. B. Ritchie, S. Fica, J. P. Staley *et al.*, 2013 A conformational switch in PRP8 mediates metal ion coordination that promotes pre-mRNA exon ligation. Nat. Struct. Mol. Biol.

Schones, D. E., K. Cui, S. Cuddapah, T.-Y. Roh, A. Barski *et al.*, 2008 Dynamic regulation of nucleosome positioning in the human genome. Cell 132: 887–898.

Schwer, B., 2008 A Conformational Rearrangement in the Spliceosome Sets the Stage for Prp22-Dependent mRNA Release. Mol. Cell.

Schwer, B., and C. H. Gross, 1998 Prp22, a DExH-box RNA helicase, plays two distinct roles in yeast pre-mRNA splicing. EMBO J.

Schwer, B., and S. Shuman, 2015 Structure-function analysis and genetic interactions of the Yhc1, SmD3, SmB, and Snp1 subunits of yeast U1 snRNP and genetic interactions of SmD3 with U2 snRNP subunit Lea1. RNA 21: 1173–86.

Semlow, D. R., and J. P. Staley, 2012 Staying on message: Ensuring fidelity in pre-mRNA splicing. Trends Biochem. Sci.

Shao, C., B. Yang, T. Wu, J. Huang, P. Tang *et al.*, 2014 Mechanisms for U2AF to define 3' splice sites and regulate alternative splicing in the human genome. Nat. Struct. Mol. Biol. 21: 997–1005.

Sims, R. J., S. Millhouse, C. F. Chen, B. A. Lewis, H. Erdjument-Bromage *et al.*, 2007 Recognition of Trimethylated Histone H3 Lysine 4 Facilitates the Recruitment of Transcription Postinitiation Factors and Pre-mRNA Splicing. Mol. Cell 28: 665–676.

Spies, N., C. B. Nielsen, R. A. Padgett, and C. B. Burge, 2009 Biased Chromatin Signatures

around Polyadenylation Sites and Exons. Mol. Cell 36: 245–254.

Stepankiw, N., M. Raghavan, E. A. Fogarty, A. Grimson, and J. a. Pleiss, 2015 Widespread alternative and aberrant splicing revealed by lariat sequencing. Nucleic Acids Res. 43: 8488–501.

Storey, J. D., and R. Tibshirani, 2003 Statistical significance for genomewide studies. Proc. Natl. Acad. Sci. 100: 9440–9445.

Szymczyna, B. R., J. Bowman, S. McCracken, A. Pineda-Lucena, Y. Lu *et al.*, 2003 Structure and function of the PWI motif: a novel nucleic acid-binding domain that facilitates pre-mRNA processing. Genes Dev. 17: 461–475.

Taggart, A. J., A. M. DeSimone, J. S. Shih, M. E. Filloux, and W. G. Fairbrother, 2012 Large-scale mapping of branchpoints in human pre-mRNA transcripts in vivo. Nat. Struct. Mol. Biol. 19: 719–721.

The Gene Ontology Consortium, 2000 Gene Ontology: tool for the unification of biology. Nat. Genet. 25: 25–29.

The Gene Ontology Consortium, 2015 Gene Ontology Consortium: going forward. Nucleic Acids Res. 43: D1049–D1056.

Tilgner, H., C. Nikolaou, S. Althammer, M. Sammeth, M. Beato *et al.*, 2009 Nucleosome positioning as a determinant of exon recognition. Nat. Struct. Mol. Biol. 16: 996–1001.

Tsai, R. T., R. H. Fu, F. L. Yeh, C. K. Tseng, Y. C. Lin *et al.*, 2005a Spliceosome disassembly catalyzed by Prp43 and its associated components Ntr1 and Ntr2. Genes Dev. 19: 2991–3003.

Tsai, R. T., R. H. Fu, F. L. Yeh, C. K. Tseng, Y. C. Lin *et al.*, 2005b Spliceosome disassembly catalyzed by Prp43 and its associated components Ntr1 and Ntr2. Genes Dev.

Tseng, C. K., H. L. Liu, and S. C. Cheng, 2011 DEAH-box ATPase Prp16 has dual roles in remodeling of the spliceosome in catalytic steps. RNA.

Turunen, J. J., E. H. Niemelä, B. Verma, and M. J. Frilander, 2013 The significant other: Splicing by the minor spliceosome. Wiley Interdiscip. Rev. RNA.

Urushiyama, S., T. Tani, and Y. Ohshima, 1996 Isolation of novel pre-mRNA splicing mutants of Schizosaccharomyces pombe. Mol. Gen. Genet. 253: 118–127.

Vanoosthuyse, V., P. Legros, S. J. A. van der Sar, G. Yvert, K. Toda *et al.*, 2014 CPF-Associated Phosphatase Activity Opposes Condensin-Mediated Chromosome Condensation. PLoS Genet. 10: e1004415.

Verdel, A., S. Jia, S. Gerber, T. Sugiyama, S. Gygi *et al.*, 2004 RNAi-mediated targeting of heterochromatin by the RITS complex. Science 303: 672–676.

Vijayraghavan, U., M. Company, and J. Abelson, 1989 Isolation and characterization of pre-mRNA splicing mutants of Saccharomyces cerevisiae. Genes Dev. 3: 1206–1216.

Villa, T., and C. Guthrie, 2005 The Isy1p component of the NineTeen Complex interacts with the ATPase Prp16p to regulate the fidelity of pre-mRNA splicing. Genes Dev.

Vincent, M., P. Lauriault, M. F. Dubois, S. Lavoie, O. Bensaude *et al.*, 1996 The nuclear matrix protein p255 is a highly phosphorylated form of RNA polymerase II largest subunit which associates with spliceosomes. Nucleic Acids Res. 24: 4649–4652.

Vo, T. V., J. Das, M. J. Meyer, N. A. Cordero, N. Akturk *et al.*, 2016 A Proteome-wide Fission Yeast Interactome Reveals Network Evolution Principles from Yeasts to Human. Cell 164: 310–323.

Wahl, M. C., C. L. Will, and R. Lührmann, 2009 The Spliceosome: Design Principles of a Dynamic RNP Machine. Cell 136: 701–718.

Washington, K., T. Ammosova, M. Beullens, M. Jerebtsova, A. Kumar *et al.*, 2002 Protein phosphatase-1 dephosphorylates the C-terminal domain of RNA polymerase-II. J. Biol. Chem. 277: 40442–8.

Webb, C. J., C. M. Romfo, W. J. van Heeckeren, and J. A. Wise, 2005a Exonic splicing enhancers in fission yeast: functional conservation demonstrates an early evolutionary origin. Genes Dev. 19: 242–254.

Webb, C. J., C. M. Romfo, W. J. van Heeckeren, and J. A. Wise, 2005b Exonic splicing enhancers in fission yeast: functional conservation demonstrates an early evolutionary origin. Genes Dev. 19: 242–54.

Webb, C. J., and J. A. Wise, 2004 The Splicing Factor U2AF Small Subunit Is Functionally Conserved between Fission Yeast and Humans. Mol. Cell. Biol. 24: 4229–4240.

Wernersson, R., and H. B. Nielsen, 2005 OligoWiz 2.0 - Integrating sequence feature annotation into the design of microarray probes. Nucleic Acids Res. 33:.

Wilhelm, B., S. Marguerat, S. Aligianni, S. Codlin, S. Watt *et al.*, 2011 Differential patterns of intronic and exonic DNA regions with respect to RNA polymerase II occupancy, nucleosome density and H3K36me3 marking in fission yeast. Genome Biol 12: R82.

Wilhelm, B. T., S. Marguerat, S. Watt, F. Schubert, V. Wood *et al.*, 2008 Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution. Nature 453: 1239–43.

Wilkinson, M. E., S. M. Fica, W. P. Galej, C. M. Norman, A. J. Newman *et al.*, 2017 Postcatalytic spliceosome structure reveals mechanism of 3 ’-splice site selection. Science (80-.).

Will, C. L., and R. Lührmann, 2011a Spliceosome structure and function. Cold Spring Harb. Perspect. Biol. 3:.

Will, C. L., and R. Lührmann, 2011b Spliceosome structure and function. Cold Spring Harb. Perspect. Biol. 3: 1–2.

Xu, H., B. J. Fair, Z. Dwyer, M. Gildea, and J. A. Pleiss, 2018 Multiplexed Primer Extension Sequencing Enables High Precision Detection of Rare Splice Isoforms. BioRxiv.

Yan, C., J. Hang, R. Wan, M. Huang, C. C. L. Wong *et al.*, 2015 Structure of a yeast spliceosome at 3.6-angstrom resolution. Science 349: 1182–91.

Yan, C., R. Wan, R. Bai, G. Huang, and Y. Shi, 2016 Structure of a yeast activated spliceosome at 3.5 Å resolution. Science (80-.).

Yeh, T.-C., H.-L. Liu, C.-S. Chung, N.-Y. Wu, Y.-C. Liu *et al.*, 2011 Splicing Factor Cwc22 Is Required for the Function of Prp2 and for the Spliceosome To Escape from a Futile Pathway. Mol. Cell. Biol.

Zheng, W., L. M. Chung, and H. Zhao, 2011 Bias detection and correction in RNA-Sequencing data. BMC Bioinformatics 12:.

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

Zhu, Y. Y., E. M. Machleder, A. Chenchik, R. Li, and P. D. Siebert, 2001 Reverse transcriptase template switching: A SMART^{TM} approach for full-length cDNA library construction. Biotechniques 30: 892–897.

Zlatanova, J., and A. Thakar, 2008 H2A.Z: View from the Top. Structure 16: 166–179.

Zofall, M., T. Fischer, K. Zhang, M. Zhou, B. Cui *et al.*, 2009 Histone H2A.Z cooperates with RNAi and heterochromatin factors to suppress antisense RNAs. Nature 461: 419–422.

Zuo, P., and T. Maniatis, 1996 The splicing factor U2AF35 mediates critical protein-protein

interactions in constitutive and enhancer-dependent splicing. Genes Dev. 10: 1356–1368.

Chapter 3: Identification and characterization of novel conditional splicing alleles in fission yeast

Alternative citation for this chapter:

*Denotes equal contribution

Fair BJ*, Larson A*, Dwyer Z, Armstrong J, Bone N, Pleiss JA. Isolation of scores of novel temperature-sensitive splicing pathway mutations in fission yeast (In preparation)

Abstract

To identify conditional alleles which disrupt the splicing pathway, a collection of ~2000 chemically mutagenized temperature-sensitive fission yeast isolates were screened for intron accumulation in three introns, including a naturally occurring U2-dependent AT-AC intron. We identified 54 strains which accumulate unspliced message at the non-permissive temperature. High resolution mapping revealed missense mutations in 6 core splicing genes which function at various steps of spliceosome assembly and activation: *sap61, ntr1, spp42, sap114, cwf22,* and *prp22*. Genome-wide analysis of the splicing defects in these mutations revealed intron-specific defects which infer mechanisms of action.

Introduction

Pre-mRNA splicing is catalyzed by the spliceosome, a dynamic multi-megadalton complex composed of five small nuclear ribonucleoprotein complexes (snRNPs) and hundreds of auxiliary proteins (reviewed in Will and Lührmann 2011; Matera and Wang 2014). Each snRNP (U1, U2, U4, U5 and U6) is composed of a small nuclear RNA (snRNA) complexed with proteins. The snRNPs assemble anew on each intron in a stepwise manner dependent on conserved sequence elements of introns, namely the 5'splice site (5'ss), the 3' splice site (3'ss),

and the branchpoint sequence (BPS) which typically lies 10-40 nucleotides upstream of the 3'ss (Taggart *et al*. 2012; Qin *et al*. 2016). Spliceosome assembly begins with binding of the U1 snRNP at the 5'ss, which nearly always begins with a GU dinucleotide and base-pairs with the 5' end of U1 snRNA. A rare class of introns, aptly named AT-AC introns, contain AU at the 5'ss and AC at the 3'ss. These introns are often spliced in a non-canonical pathway by an analogous set snRNPs (U11, U12, U4atac, U5, U6atac) which are collectively called the minor spliceosome (Turunen *et al*. 2013). In the canonical pathway, after U1 binding, the U2 snRNP assembles on the pre-mRNA in an ATP-dependent manner, resulting in a base-paired duplex between the BPS and the U2 snRNA wherein a conserved catalytic adenosine in the BPS is bulged. This duplex is stabilized by SF3A and SF3B, heteromeric protein subcomplexes that associate with the U2 snRNP. The SF3B complex shields the reactive bulged adenosine until catalytic activation (Rauhut *et al*. 2016). The pre-formed U4/U5/U6 tri-snRNP assembles on the intron, and a cascade of structural and base-pairing rearrangements are catalyzed by ATP-dependent RNA-helicases. U4 snRNP and U1 snRNP dissociate, and the U1:5'ss interaction is replaced by U6:5'ss base-pairing. Concomitantly, the NineTeen Complex (NTC), a complex of proteins associated with Prp19, joins the spliceosome. The RNA helicase Prp2 disassociates the SF3A/B complexes (Kim and Lin 1993; Liu and Cheng 2012), allowing for transesterification between the reactive bulged adenosine and the 5'ss, resulting in a lariat intermediate. The RNA helicases Prp16 and Prp22 facilitate a structural rearrangement that leads to the second tranesterification between the upstream exon and the 3'ss, resulting in exon-ligation and excision of the lariat (Schwer and Gross 1998; Tseng *et al*. 2011). Spliceosome disassembly is catalyzed by the helicase Prp43 and its cofactors, Ntr1 and Ntr2, resulting in separation of U2, U5, U6, NTC and the excised lariat (Tsai *et al*. 2005b; Boon *et al*. 2006).

Many of the genes in the splicing pathway were initially identified by genetic screens for *S. cerevisiae* strains which harbor defects in pre-mRNA processing, thus the naming of the Prp genes (Hartwell *et al*. 1970; Roshbash *et al*. 1981; Vijayraghavan *et al*. 1989; Noble and Guthrie 1996; Hossain and Johnson 2014b). Often these screens result in the isolation of conditional splicing alleles: cold-sensitive or temperature-sensitive (ts) alleles of splicing genes which are active at a permissive temperature (typically 25°C) but inactive and inviable for cell growth at the non-permissive temperature. These conditional alleles have allowed for temperature-selectable genetic screens for extragenic suppressor mutations, leading to identification of additional factors (Jamieson *et al*. 1991; Lybarger *et al*. 1999; Villa and Guthrie 2005). Furthermore, these conditional alleles serve as in-activateable splicing factors for *in vitro* splicing assays to delineate the biochemical requirements for each ordered step in the pathway (Lustig *et al*. 1986; Lin *et al*. 1987; Libri *et al*. 2001). Most recently, the cryo-EM structure of a late-stage yeast spliceosome was achieved by stalling the spliceosome immediately after to exon-ligation through utilization of a dominant negative ts-allele of Prp22 (Schwer 2008; Wilkinson *et al*. 2017).

The distantly related fission yeast *Schizosaccharomyces pombe* has also been used to identify conditional mutants via Northern blot screening for pre-mRNA accumulation in libraries of ts-strains (Potashkin *et al*. 1989; Urushiyama *et al*. 1996). *S. pombe* is similarly genetically tractable as *S. cerevisiae*, yet retains many features of splicing that have been lost in *the S. cerevisiae* lineage (Kuhn and Käufer 2003; Fair and Pleiss 2017), including many spliceosome genes for which there exists a human homolog but not an *S. cerevisiae* homolog (Käufer and Potashkin 2000; Webb and Wise 2004). We previously screened a haploid non-essential gene-deletion library in *S. pombe* for genes which affect the splicing pathway (Larson *et al*. 2016). As

most known splicing pathway genes in fission yeast are essential (Kim *et al*. 2010), we sought to gain mutational access to essential splicing pathway genes. Here we apply a similar quantitative screening methodology to a library of ~2000 ts-strains and report the identification of novel ts-alleles of known core splicing factors. After characterizing the genome-wide *in vivo* splicing phenotypes of some of these core factor mutations, we describe the relative dependencies of different subsets of introns for different splicing factors.

Results

Screening for splicing defects in canonical and non-canonical (AT-AC) introns identifies 54 ts mutant strains

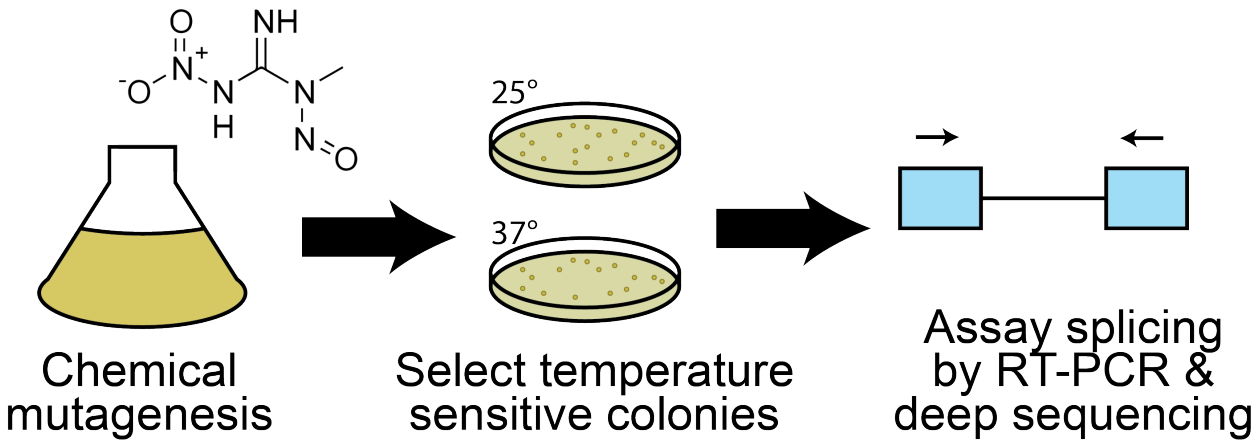


Figure1: Workflow of a forward genetic screen for ts-alleles which display splicing defects. Yeast cells were randomly mutagenized with nitrosoguanidine and isolates were picked after replica plating at 27°C and 37°C to obtain an arrayed library of ~2000 ts-isolates. Strains were assayed for splicing defects by RT-PCR with primers that flank an intron, followed by deep sequencing the RT-PCR products to measure the relative amount of intron retention.

An arrayed library of ~2000 ts-isolates was obtained by chemical mutagenesis and replica plating, selecting for isolates which are viable at 27°C but not at 37°C (Armstrong *et al*. 2007). To identify the ts-strains harboring splicing defects, we performed RT-PCR and deep sequencing of the RT-PCR products to quantify the splicing status of naturally occurring introns in each strain after a 15-minute shift to the non-permissive temperature (Figure1, methods). To capture

potential substrate-specific splicing defects, we chose three introns with varying features to quantify splicing in each strain. Firstly, we screened for strains with increased intron retention in a ribosomal protein gene, *rpl39*_intron1. This intron is of normal length for fission yeast (62bp) and contains splice sites that match the consensus splice site motifs for fission yeast. Importantly, this intron lies within the 3'UTR and thus, intron retention would not interrupt the open reading frame. We therefore expect accumulation of this intron retention isoform to specifically reflect defects in splicing, rather than defects in the nonsense mediated decay (NMD) pathway which strongly affects the abundance of most other intron-retained transcripts (Bitton *et al*. 2015). Consistent with this intron retention event being a weak target for surveillance by the NMD pathway, we observe a relatively high natural abundance, ~7%, of intron-retained transcript (Fig2A). We identified a significant increase in intron-retention in 42 strains, with intron retention rates as high as 40 percent-spliced-in (PSI). Given the relatively high sequencing depth targeted at a single locus in this screening approach, we were additionally able to detect ultra-low-frequency, unannotated splicing events that occur between the RT-PCR primers. For example, we found an AG dinucleotide 17 bases downstream of the annotated 3'ss that gets utilized at a mere 0.001% of the annotated 3'ss, suggesting very strong constraints for the distance between the branchpoint and the 3'ss when multiple AG dinucleotides are present.

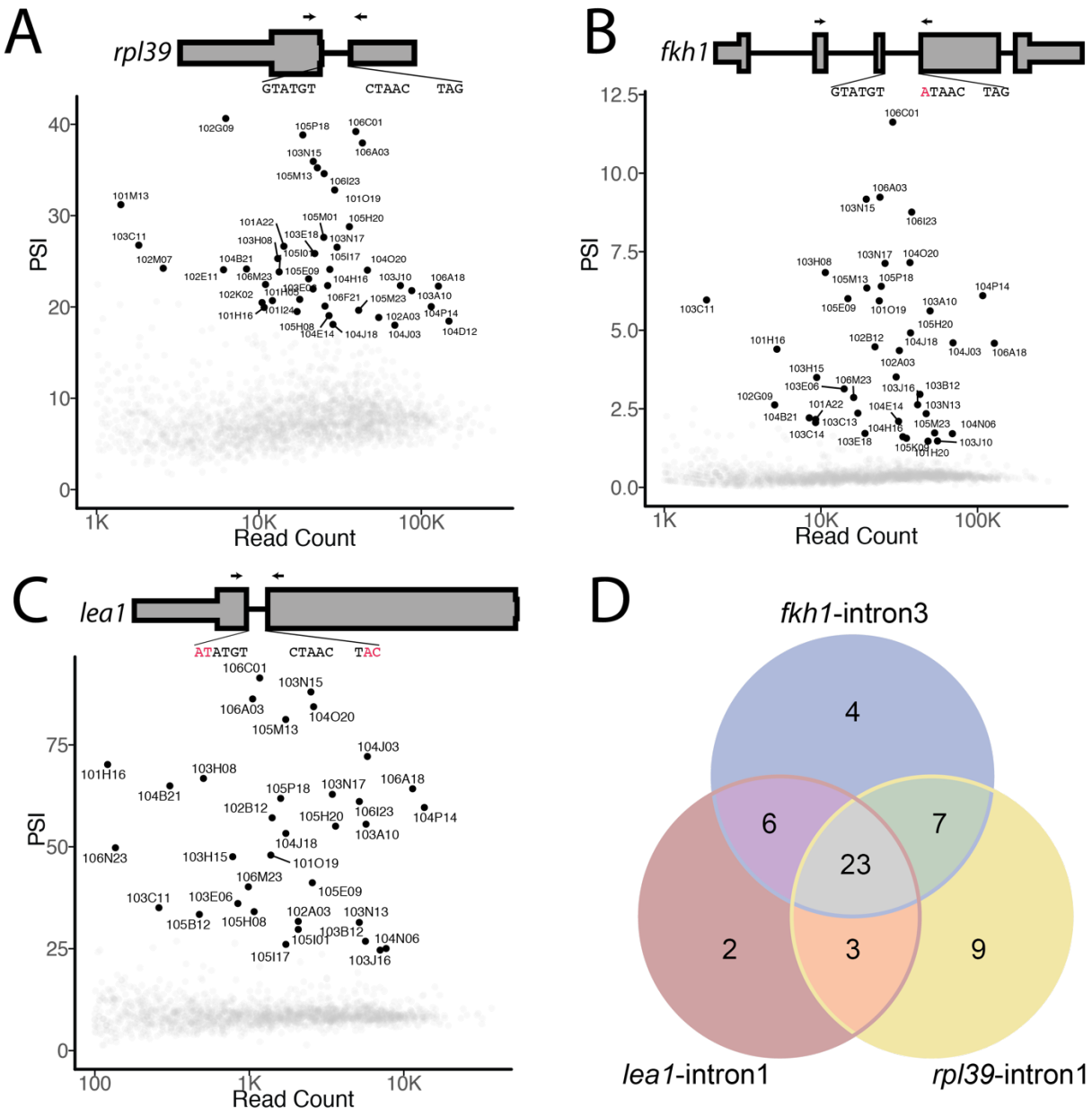
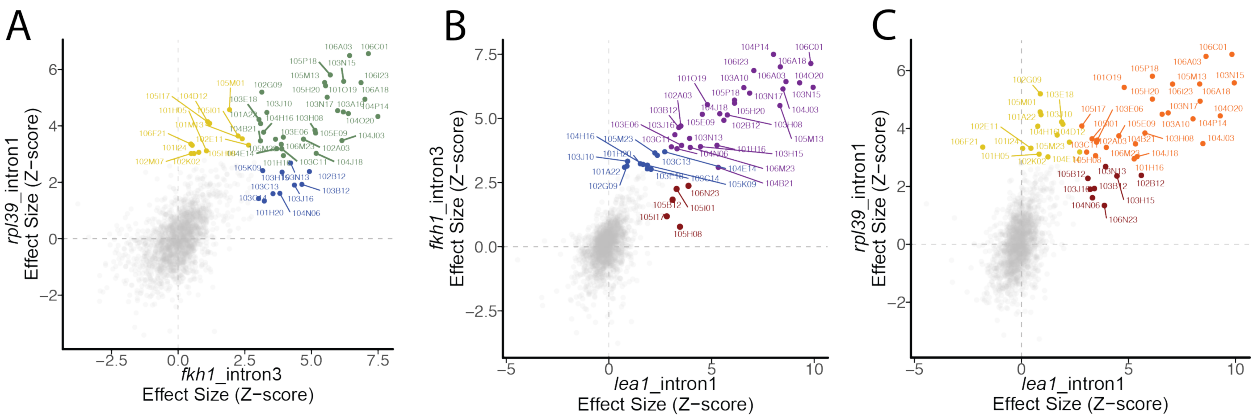


Figure2: Screening for intron retention defects in different intron substrates reveals different sets of strains. The intron retention percent spliced-in (PSI), a measure of the percent of transcripts that have the intron retained, is plotted for three different substrates for each of the ~2000 strains screened. Each strain is represented by a dot. Strains which have a significant increase in PSI, given read depth, are bolded and labelled according to an arbitrary strain identifier. The three intron substrates screened against are **(A)** *rpl39*-intron1, which has consensus motifs for the 5'ss, branchpoint motif and 3'ss; **(B)** *fkh1*-intron3, which has a non-consensus branchpoint motif; **(C)** and *lea1*-intron1 which is a major-spliceosome-dependent AT-AC intron. **(D)** The overlap among the 54 strains identified as exhibiting a significant splicing defect

in the three substrates screened.

Next, we screened for strains which have increased intron retention in intron 3 of *fkh1*, a peptidylprolyl isomerase. This intron has a notably non-consensus BPS, due to the adenosine at position -3, a position reserved for pyrimidines in 99.5% of *S. pombe* introns (Fig1B). Nonetheless, this intron is efficiently spliced (PSI<1%) in most strains. Of the 40 strains identified with significant increases in intron-retained transcript, 30 were also identified in the screen against *rpl39*_intron1 (Fig2D). Changes in intron retention in rpl39_intron1 and *fkh1*_intron3 correlate well, suggesting most of the identified mutations affect splicing of many or all introns, albeit with slightly different effect sizes for different introns (FigS1).



FigureS1: The effect size of splicing defect as measured by thee different introns. The measured effect size for intron retention in each strain is plotted as a point in pairwise scatter plots for the three introns assayed. Each point is a strain. Strains with significant intron retention on either of the axes are bolded. Points are colored red for significant intron retention in *lea1*_intron1, yellow for *rpl39*_intron1, and blue for *fkh1*_intron3. The intersection of two significant sets of strains is the shown as green, purple, or orange.

Lastly, in view of the recent discovery of an AT-AC intron (with an unusual AU at the 5'ss and AC at the 3'ss) in the *lea1* gene (Chen *et al*. 2018), we screened for intron retention in this unusual splicing substrate. While considering alternative introns to screen against, we discovered an additional AT-AC intron in the *taf2* gene (data not shown). These introns, although unusual in their 5'ss and 3'ss sequence, must be substrates of the canonical

U1/U2/U4/U5/U6 spliceosome, as fission yeast does not contain the U11/U12/U4atac/U5/U6atac minor spliceosome machinery known to catalyze splicing of many AT-AC introns in plants and humans (Turunen *et al*. 2013). Furthermore, the *lea1*_intron1 was found physically associated with the post-catalytic U2-U5-U6 spliceosome as an excised lariat (Chen *et al*. 2018). In our screen, we identified only subtle differences in the mutant strains identified by screening against *lea1*_intron1; 32 of the 34 strains with significant increases in *lea1* intron retention (Fig2C, 2D) were also identified in the screens against more canonical intron substrates, consistent with this AT-AC intron being spliced by the canonical set of splicoesome factors. The two strains uniquely identified as significant for *lea1*_intron1 retention also had clear effects in *rpl39*_intron1 and *fkh1*_intron3 that just missed our significance thresholds for *lea1*_intron1 (FigS1). An more in-depth examination of differences in the relative effect sizes of different spliceosome factor mutations on this intron versus more canonical introns may reveal insights towards mechanistic distinctions in the splicing of GT-AG and AT-AC introns.

High resolution mapping of the causative mutations identifies novel alleles of core splicing factors

We identified a total of 54 strains carrying a significant increase in intron retention among at least one of the three introns examined. We initially picked three strains, each referred to hereafter by an arbitrary identifier, to map the causal mutation(s) for: 103A10, 103H15, and 103N15. Under the assumption that the causal mutation(s) for the ts-phenotype are also causing the observed splicing defects, we turned to bulk segregant analysis with whole-genome sequencing for high resolution mapping of the alleles that segregate into ts and non-ts phenotypes. Each strain was outcrossed and F1 progeny from each cross were separated into bulk populations based on the progeny's ts-phenotype. In each cross, the F1 progeny appeared in

roughly equal numbers of ts and non-ts, suggesting a single gene is responsible (data not shown). Whole genome sequencing of each bulk comprehensively identified the mutations enriched in the ts- bulk and de-enriched in the non-ts-bulk. Genome-wide association analysis revealed *prp22*-G917R;V801M as the causative mutation(s) in strain 103A10 (Fig3A,B). The only other mutation with significant association signal is a non-coding mutation tightly linked to *prp22*. Given that the only significant coding mutations were in *prp22*, a DEAH-box RNA-helicase required for the second catalytic step of splicing (Mayas *et al.* 2006), we conclude that the splicing-defect and ts-phenotype are both caused by this single-gene mutation. We obtained a similarly simple genetic explanation for the phenotype of strain 103N15, where the significant association signal could be mapped to a single amino acid substitution in *sap61*, a component of the U2 snRNP, orthologous to S. cerevisiae Prp19 and human SF3A3. We did not find any significant association signal to explain the phenotype of strain 103H15, possibly owing to insufficient sequencing coverage to adequately genotype allele frequencies in bulk populations at the causal locus, despite ~30X genome coverage in each bulk.

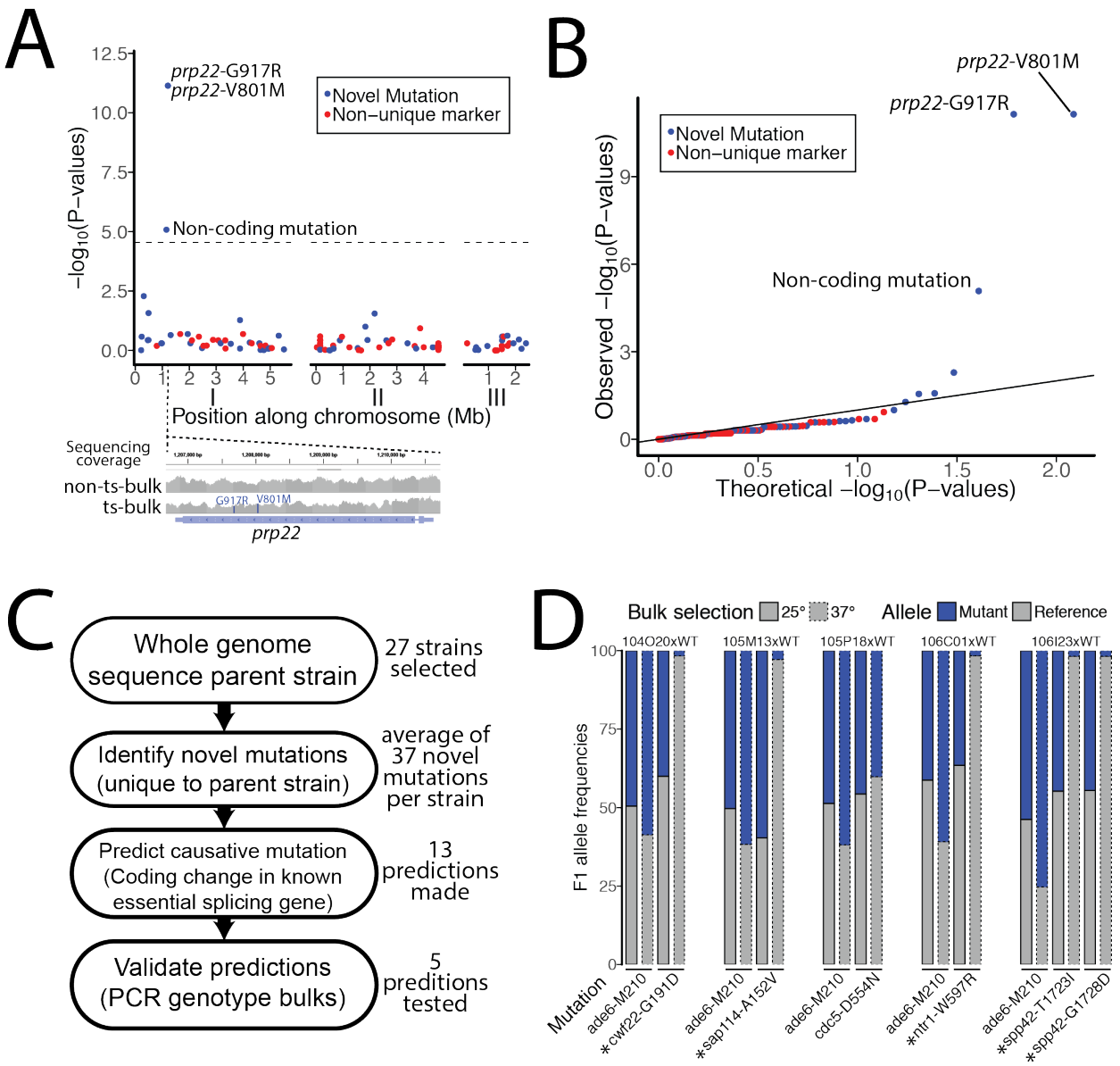
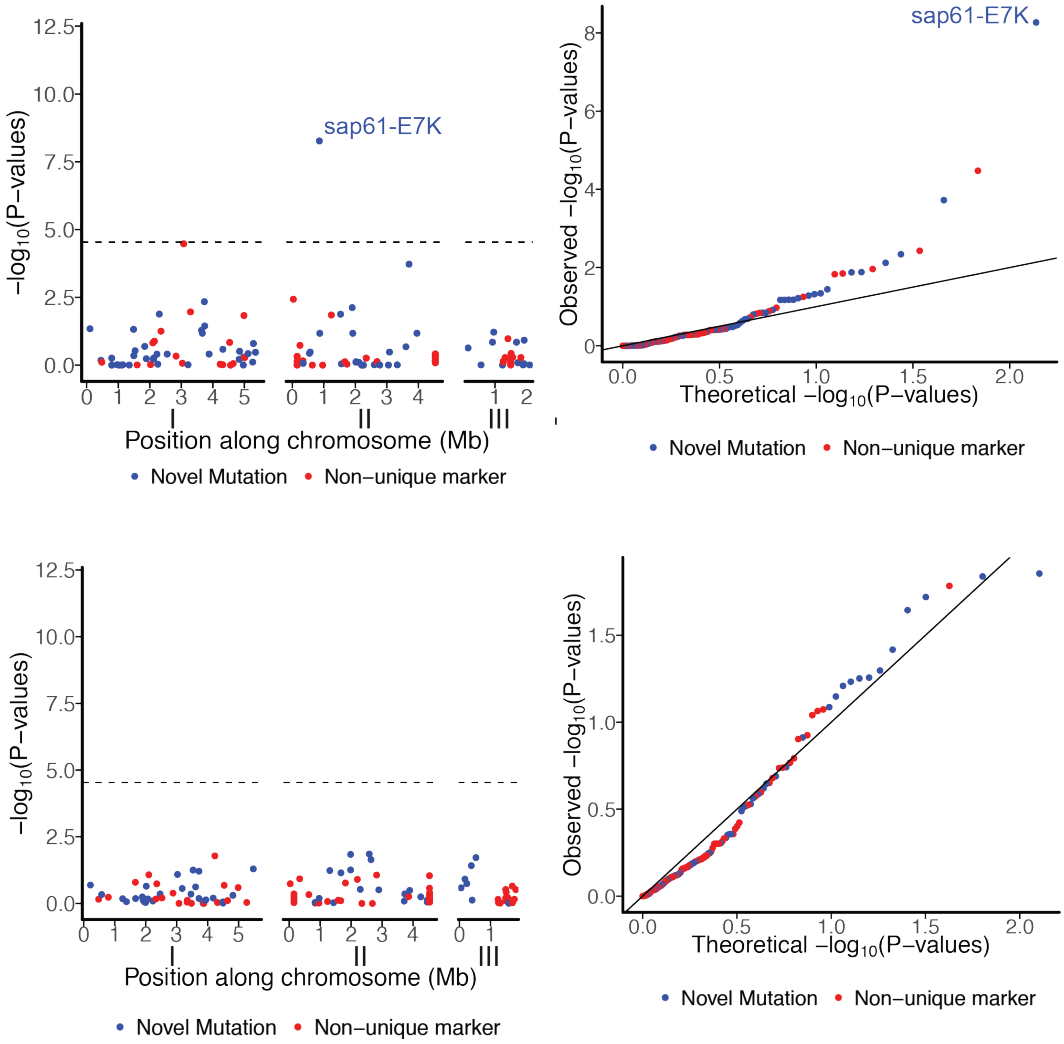


Figure3: High resolution mapping strategies employed to identify the causative mutation of splicing-defective strains. (A) Manhattan plot of associated mutations of the F1 progeny (separated into ts and non-ts bulks) of a 103A10 x wildtype cross. Mutations which were specific to this cross are novel mutations, while mutations that are shared across different crosses are inferred as non-unique markers which must have been present in either parental strain prior to mutagenesis. Mutations that are significantly enriched in the ts-bulk and depleted from the non-ts bulk are above the Bonferonni-corrected significance threshold (dotted line). Raw read coverage with mismatches at the *prp22* locus is shown in the inset. **(B)** QQ-plot of theoretical and observed P-values of the association analysis in (A). **(C)** An alternative strategy that was used to map causative loci involves whole-genome-sequencing the strain of interest, and making predictions on which of the ~40 mutations could be causative, assuming simple the ts-phenotype and splicing defect are simple traits caused by the same mutation. **(D)** Validation of some of the predicted candidate mutations by bulk

segregant analysis followed by targeted sequencing to genotype allele frequencies in each bulk. Allele frequencies in each bulk are indicated by stacked bar chart. As a control, an unrelated locus, ade6-M210, was genotyped in each bulk. Asterisks indicate mutations which are closely linked to the ts-phenotype.



FigureS2: High resolution mapping strategies employed to identify the causative mutation of splicing-defective strains. (A) Manhattan plot of associated mutations of the F1 progeny (separated into ts and non-ts bulks) of a 103N15 x wildtype cross. Mutations which were specific to this cross are novel mutations, while mutations that are shared across different crosses are inferred as non-unique markers which must have been present in either parental strain prior to mutagenesis. Mutations that are significantly enriched in the ts-bulk and depleted from the non-ts bulk are above the Bonferonni-corrected significance threshold (dotted line). **(B)** QQ-plot of theoretical and observed P-values of the association analysis in (A). **(C)** Manhattan plot of association analysis on 103H15. **(D)** QQ-plot of theoretical and observed P-values of the association

analysis in (C).

To overcome the high sequencing coverage required for high-resolution mapping by whole genome sequencing large bulks, we developed an alternative strategy to mapping mutations. Given that whole genome sequencing of bulks detected only 50 to 70 novel mutations in each ts-strain (Fig3A,3B, FigSMapping), of which only about half are coding mutations, we surmised that the causative mutation could be predicted by whole genome sequencing the parental (F0) ts-strain and using gene annotations to make verifiable predictions (Fig 3C). To demonstrate this approach, we performed whole genome sequencing on each of the strains identified as carrying a splicing defect the screen, of which, 26 of these strains had sufficient coverage (median 9X genome coverage) to comprehensively identify mutations genome-wide (TableS1). Given that causal mutations that were introduced during mutagenesis should be unique to the strain of interest, we excluded mutations common to many strains from further analysis, leaving a median of 36.5 novel mutations per strain. There was a median of 12 genes containing novel coding mutations per strain, with a median of 4.5 of these mutated genes being essential for viability. In 13 of these strains, one of the mutated essential genes was in a known splicing factor, which we predicted to be the causative mutation. Of the 5 of strains selected for validation of the candidate causal mutation, 4 of the candidate mutations were tightly linked to the ts-phenotype, confirming our prediction (Fig3D). The strain for which the predicted mutation was not linked to the ts-phenotype (105P18) could represent an incorrect prediction for the splicing-defect mutation, or alternatively, a strain which contains a splicing-phenotype and a ts-phenotype which are unlinked. In total we identified, and confirmed via linkage, 6 novel alleles in core splicing factors: prp22-V801M;G917R, sap61-E7K; cwf22-G191D; sap114-A152V; ntr1-W597R and spp42-T1723I;G1728D. Additionally, we have strong predictions on the causative mutation for 6 strains, including one strain which harbors a predicted causative

mutation in *prp10*, the ortholog of human SF3B1 which is found commonly mutated in patients with chronic lymphocytic leukemia, myelodysplastic syndrome, and breast cancer (Network 2012; Quesada *et al*. 2012; Cazzola *et al*. 2013). The mutation identified in this study, *prp10*-R379H, maps to a region of the protein that contacts the U2 snRNA:BPS duplex formed prior to the first step of splicing (Fig4A).

TableS1: Causative mutations predicted and validated in strains subject to whole genome sequencing

Strain	Median base coverage in whole genome sequencing	Percent bases covered by at least 2 reads	Number of novel mutations identified	Genes with coding mutations	Essential genes with coding mutations	Essential, splicing genes with coding mutations	Causative mutation predicted without linkage analysis data	Mutations experimentally validated as linked to ts-phenotype
101A22	9	97.438	34	13	2	1	prp1-G705D	None tested
101I24	6	91.6416	26	9	0	0	No prediction	None tested
102A03	11	99.3911	43	6	1	0	No prediction	None tested
102B12	10	98.6822	61	23	5	2	prp10-R379H	None tested
102M07	15	99.3303	11	1	0	0	No prediction	None tested
103A10	4	97.4002	9	2	0	0	No prediction	prp22-V801M;G917R
103H08	4	95.355	48	18	5	1	cwf22-G468D	None tested
103H15	NA	NA	NA	NA	NA	NA	NA	Tested whole genome, none identified
103N15	8	97.4083	39	19	4	1	sap61-E7K	sap61-E7K
103B12	9	97.9299	57	17	5	0	No prediction	None tested
103C11	6	94.2691	60	27	6	0	No prediction	None tested
103C13	7	95.2521	16	7	2	0	No prediction	None tested
103J10	9	97.5016	27	7	5	1	brr2-G1527D	None tested
103J16	12	99.1053	11	4	1	0	No prediction	None tested
103N13	16	99.6166	48	21	7	1	prp1-G705D	None tested
104J03	9	98.3559	102	46	16	2	prp28-P383S	None tested
104J18	11	98.9073	30	8	3	0	No prediction	None tested
104O20	7	95.9661	26	10	7	1	cwf22-G191D	cwf22-G191D
104P14	11	98.7842	42	8	2	0	No prediction	None tested
104N06	6	92.5183	13	5	3	0	No prediction	None tested
105H20	9	97.6107	25	12	1	0	No prediction	None tested
105M13	8	97.2089	40	18	7	1	sap114-A152V	sap114-A152V
105P18	10	98.6747	43	16	6	1	cdc5-D554N	Tested 1 locus, none identified
106A03	11	99.1203	45	21	5	0	No prediction	None tested
106C01	12	99.2234	34	14	5	1	ntr1-W597R	ntr1-W597R
106I23	10	98.4374	28	7	4	1	spp42-T1723I;G1728D	spp42-T1723I;G1728D
106A18	19	99.7531	60	25	9	1	prp1-G705D	None tested

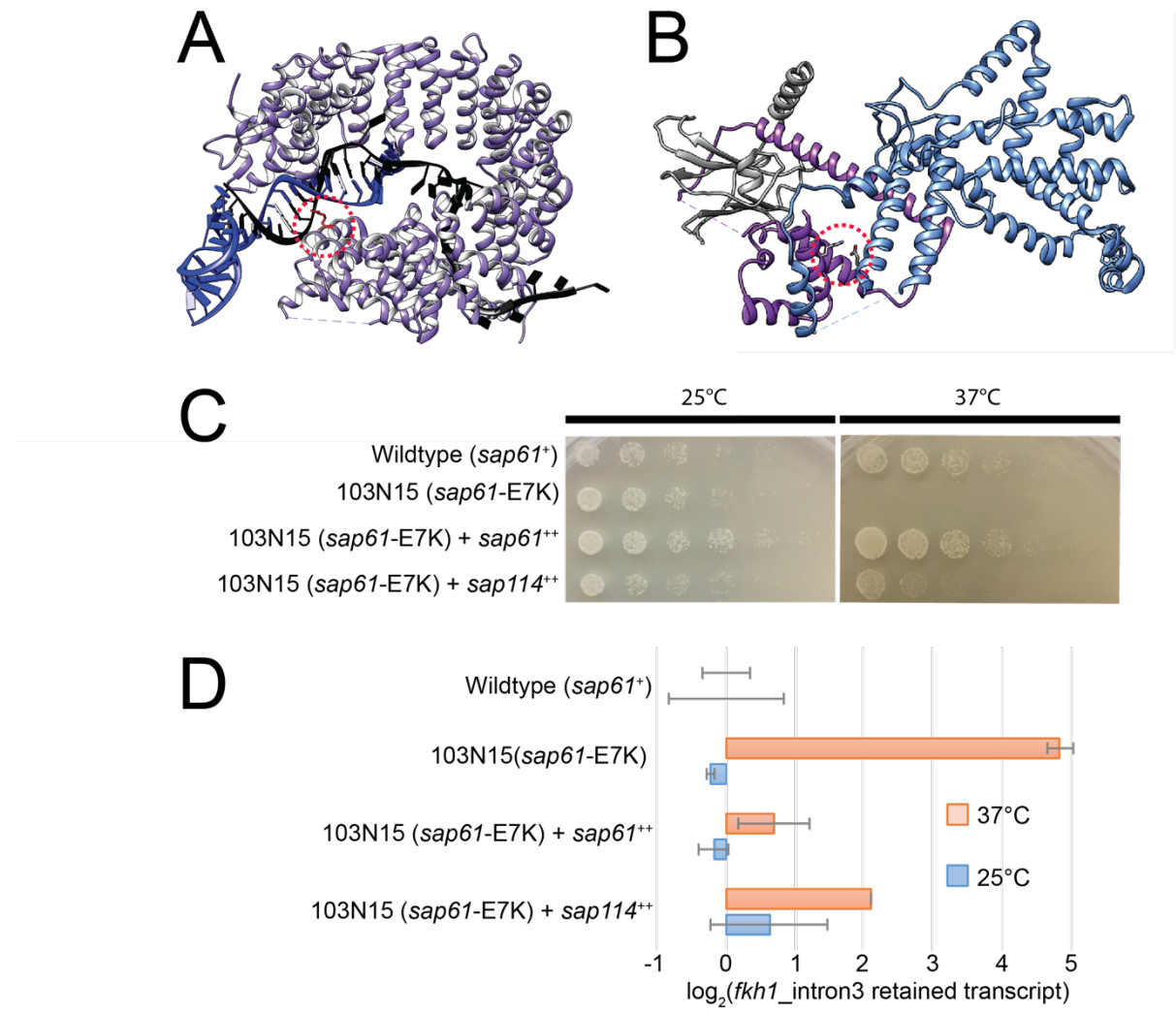


Figure4: Structural insights to the nature of conditional mutations. (A) The published structure of S. cerevisiae HSH155 (orthologous to S. pombe *prp10*, human SF3B1) is shown covering the U2 snRNA:pre-mRNA duplex formed at the branch site in the B^{act} spliceosome complex. The residue orthologous to *prp10*-R379 is highlighted in red. U2 snRNA in blue. pre-mRNA in black. **(B)** The published structure of S. cerevisiae SF3A complex (Lin and Xu 2012), composed of Prp9 (blue), Prp21 (purple), and Prp11 (gray). The residue orthologous to S. pombe sap61-E7 (Prp9-E5) is shown making a salt bridge, highlighted in red, with Prp21 (*sap114*). **(C)** Yeast serial dilution growth assays of ts-strain 105N15 with various plasmids (indicated by + after the genotype) demonstrate that the ts-phenotype can be suppressed by exogenous over-expression (plasmid genotype marked as $^{++}$) of *sap61*, as well as *sap114* **(D)** qRT-PCR measurement of *fkh1*-intron3-retained transcript levels demonstrate that the splicing defect at the non-permissive temperature of strain 105M13 can be suppressed by expression of *sap61*, as well as sap114.

Mapping the location of the mutated residues of these ts-alleles onto recent atomic-level structures (Lin and Xu 2012; Yan *et al*. 2015) suggested the mechanistic basis for splicing

defects. For example, *sap61*-E7K disrupts a salt bridge between the *sap61* (SF3A3 in humans) and *sap114* (SF3A1 in humans) genes of the hetero-trimeric SF3A sub-complex of the U2 snRNP (Fig4B). Armed with this structural knowledge, we hypothesized that over-expression of *sap114* could restore sufficient functional SF3A complex. Consistent with this hypothesis, exogenous overexpression of *sap114* in the sap61-E7K strain suppressed ts-phenotype (Fig4C) and partially suppressed the splicing defect as measured by qRT-PCR of fkh1_intron3 (Fig4D).

Genome-wide analysis of core splicing factor mutations reveal distinct *in vivo* splicing signatures

To better understand the splicing defects caused by the identified mutations, we used RNA-seq to measure genome-wide patterns in 6 of the mutant strains and two replicates of wildtype yeast after a 15-minute shift to the non-permissive temperature. All of the mutant strains increased intron retention isoforms genome-wide (Fig5A). The global levels of other forms of alternative splicing (Fig5A) were in <1% all strains examined and no strains had a clear preference for any particular form of alternative splicing (alternative 5'ss, alternative 3'ss, etc.). The global levels of intron retention ranged from ~30% PSI in strain 106A18 (mutation in *prp1*-G705D) to ~75% PSI in 106C01 (mutation in *ntr1*-W597R). *prp1* (orthologous to *S. cerevisiae* Prp6) is a snRNP maturation factor which aids in bridging interactions between U4/U6 and U5 snRNPs to form the functional tri-snRNP. Given the relatively short (15 minute) inactivation time at the non-permissive temperature, it is perhaps unsurprising that 106C01 (*prp1*-G705D) has the weakest of the genome-wide intron-accumulation phenotype compared to the other strains which presumably directly affect splicing reactions catalyzed by pre-formed tri-snRNPs. A longer inactivation period in this strain may be required to observe a stronger molecular phenotype.

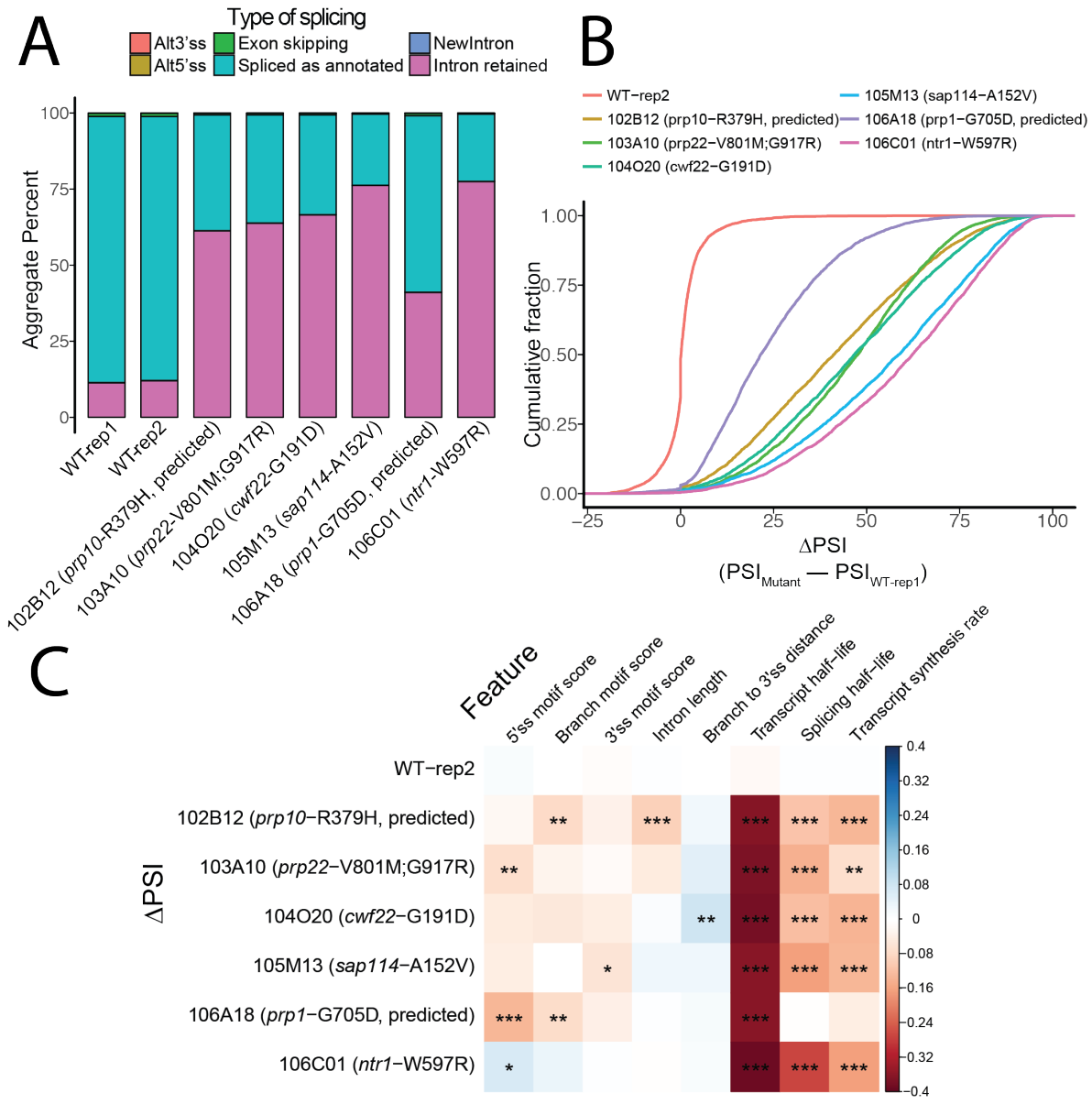
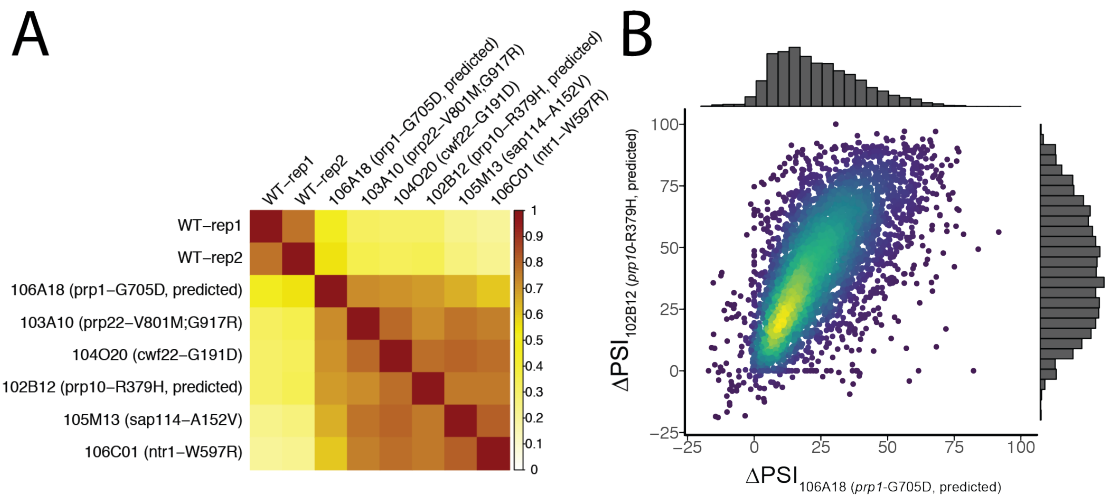


Figure5: RNA-seq reveals distinct global splicing defects. (A) Stacked bar charts indicate the percentage of splice events of various alternative splicing categories as measured by RNA-seq after 15 minutes of cells shifted to the non-permissive temperature. Genotypes inferred as causative are labelled in parenthesis and marked as 'predicted' if tight linkage to the ts-phenotype was not demonstrated for that genotype **(B)** Cumulative distribution plots of the change in intron retention ($\Delta PSI = PSI_{Mutant} - PSI_{WT\text{-}rep1}$) for annotated introns in each of the mutants as well as the other wildtype strain (WT-rep2) as a control. All strains generally increase intron retention, though the change in intron retention varies over a wide range. **(C)** Heatmap of pairwise spearman correlation coefficients between ΔPSI and features of the introns or their parent transcripts. Asterisks indicate levels of significance after Bonferroni-Holm correction: P<0.05*, 0.001**, 0.00001***.

Although global intron-retention levels increased in all 6 mutant strains, the changes in intron retention for each intron varied greatly from relatively unchanged (low ΔPSI) to greatly changed (high ΔPSI), with a median interquartile ΔPSI range of 35% amongst the 6 mutant strains (Fig5B). Interestingly, even for strains with different levels of global intron, such as 103B12 and 106A18 (FigS3), there was a high degree of correlation in the change in intron-retention as compared to wildtype (FigS3). In other words, generally, the same set of introns are most affected in each of the mutant strains, suggesting there exists common features that explain intron-to-intron variation of splicing defects in all strains. To identify these features, we tested pairwise correlations of ΔPSI versus various features of the intron, including splice site motifs, intron length, and published rates of transcription, mRNA degradation, and intron splicing rate (Eser *et al*. 2016). None of these features significantly correlated when comparing the ΔPSI of two replicates, confirming that the biological and technical noise between samples does not substantially bias this type of association analysis (Fig5C). The most significant correlating feature for all strains was mRNA degradation rate, wherein shortly lived transcripts are associated with the greatest changes in intron retention. This is consistent with the idea that the ability to measure a change in splicing after a short *in vivo* inactivation period is limited by the length of time for which the over-abundant spliced mRNA is degraded, as well as the rate at which new, initially unspliced, transcripts can be synthesized. Splicing time and transcription rate, also previously measured by a metabolic labeling time course in wildtype cells (Eser *et al*. 2016), also correlated strongly to changes ΔPSI. The *prp1*-G705D strain was unique in that changes in splicing were not significantly correlated to splicing time or synthesis rate, consistent with this mutation affecting splicing in a time-delayed manner.



FigureS3: All ts-alleles examined display similar profile of global intron retention. **(A)** Hierarchal clustered spearman correlation matrix of intron retention (PSI) values genomewide. All mutant strains cluster similarly **(B)** Scatterplot and marginal histograms of the changes in intron retention compared to wildtype (ΔPSI) for all introns in strain 106A18 and strain 102B12. Even though 106A18 experiences a lower degree of intron retention, the degree to which each intron experiences a change (ΔPSI) is highly correlated with 102B12.

Some other distinctive correlations that we identified also suggest specific mechanisms of splicing defects. For example, consistent with the observation that the 102B12 strain (causative mutation predicted in *prp10*) was identified in the screen for rpl39_intron1 retention (Fig1A) which carries a consensus branch motif, but not the screen for fkh1_intron3 retention (Fig1B) which has a non-consensus branch motif, the genome-wide ΔPSI measurements for this strain negatively correlate with branch motif strength (Fig5C). This perhaps is unsurprising, given the mutated residue in *prp10* contacts the U2-snRNA:BPS duplex at about the -3 to -4 position from the bulged adenosine (Fig3A). Changes in intron retention in strain 103A10 (*prp22*-V801M) negatively correlate with 5'ss strength. Studies in *S. cerevisiae* suggest that Prp22, can proofread and reject substrates with mutations in any of the consensus splice site motifs, leading to discard of lariat intermediates (Mayas *et al*. 2006; Semlow and Staley 2012). However these results

(Fig5C) suggest that *prp22* proofreading is dependent more so on the relative strength of the 5'ss motif, rather than the branch or 3'ss motifs in *S. pombe*.

Intron retention in 104O20 (*cwf22*-G191D) is associated with longer distances between the branch and 3'ss. A relatively weaker pairwise associations was observed between 105M13 (*sap114*-A152V) intron retention and weaker 3'ss motifs. Interestingly, 106C01 (ntr1-W597) which exhibited the most intron retention globally (Fig5A,B), also exhibited a slight positive association with strength of splice site motifs. Additionally, this strain has a uniquely strong association with splicing time as measured in wildtype cells (Fig5C). Some of the intron features presented in this association analysis, namely 5'ss motif score and branch motif score, independently correlated with splicing time, or transcript synthesis and degradation rates (Eser *et al*. 2016). Therefore, the significant pairwise positive correlation between 5'ss strength and intron retention in this strain may not be a causal association, but rather a result of confounding or interacting variables. This would be most consistent with *ntr1*'s role as a spliceosome-disassembly factor acting on the excised-intron-lariat post-spliceosome complex (Tsai *et al*. 2005b; Boon *et al*. 2006) wherein *ntr1* would have limited ability to discriminate splicing of substrates with varying splice site motifs. None of the strains examined here had obvious unique affects on the *lea1*_intron1 or *taf2*_intron1 (data not shown), the only two AT-AC introns *S. pombe*. However, given the effect sizes observed for other intron features, and the fact that there are only two instances of AT-AC introns, the power to detect a true association between splicing factor mutations and a tolerance for AT-AC splicing is severely limited.

Discussion

Here we screened an S. pombe library of ts-isolates for splicing defects in three introns, including an intron with canonical splice site motif, an intron with a weak BPS, and a U1/U2

dependent AT-AC intron. There was a high degree of correlation between all screens (Fig2D, FigS1), suggesting there were not any factors completely specific to splicing of the AT-AC intron. Together these screens identified 54 ts-strains with significant splicing defects, and from these we mapped with high-resolution six novel ts-alleles in splicing factors (tableS1), including mutations in *prp22*, *cwf22*, *sap61*, *sap114*, *ntr1*, and *spp42* (orthologous to budding yeast Prp22, Cwc22, Prp21, Spp382, and Prp8, respectively). Additionally, we made predictions on the causative alleles in an additional 6-ts isolates with splicing defects, and demonstrated a 4/5 validation rate for our prediction method (Fig3C, Fig3D). The strains which did not harbor candidate mutations in known splicing genes may have mutations in periphery or unknown splicing-related genes. The application of more traditional mapping strategies, like gene complementation plasmid libraries, may lead to the identification of a more complete set of trans-factors involved in splicing.

The mutation we isolated in *sap61*, a component of the SF3A complex, disrupts the binding interface with complex member *sap114*. Overexpression of *sap114* suppressed the *sap61* mutant phenotype. We expect that identification of additional extragenic suppressors of these alleles will serve to clarify functional relationships between spliceosome components. The mutation identified in *spp42*, the largest spliceosome protein, maps to a functional hotspot in the highly conserved RNAseH domain of the protein. Alleles in this region, often with opposing phenotypes, have been well studied in *S. cerevisiae*. These studies suggest multiple conformations of this domain in mediating catalysis and proofreading of the 1st and 2nd steps of splicing and suggest dynamic conformational states for this region during different steps of splicing (Schellenberg *et al*. 2013; Galej *et al*. 2013, 2014; Mayerle *et al*. 2017). Our identification of a similarly positioned allele in *S. pombe*, an organism with a larger variation in

splicing substrate motifs, may allow for a more extensive *in vivo* analysis of how the RNAseH domain proofreads substrates at the 1st or 2nd step of splicing.

Genome-wide assessment of splicing in the mutant strains resolved substrate-specific defects caused by mutations in *prp10*, *prp22*, *cwf22*, *sap114*, *prp1*, and *ntr1*. Unlike the cancer-associated mutations which naturally occur most often in the SF3B1 alpha-helical HEAT domain repeats 6-8, the *prp10*-R379H allele in alpha-helical HEAT domain repeat 2 does not result any noticeable increase in alternative 3'ss selection (data not shown). Though, as with all of the mutations we examined, it remains to be seen if orthologous mutations would similarly manifest themselves in the human spliceosome which is more permissive to alternative splicing. Unsurprisingly, the *prp10*-R379H allele, which structurally maps to the -3 to -4 position of the BPS:U2snRNA duplex in activated spliceosome, causes intron retention preferentially in substrates with a weak BPS. Interestingly, this mutation also correlates with increased intron retention of long-introns. It is yet unclear if this association is explained by an *in vivo* confounding factor such, as turnover rate of affected transcripts, or by a mechanism wherein the intron length is sensed by *prp10*. We also observed an association wherein a mutation in *cwf22* (*S. cerevisiae* Cwc22), a protein loosely associated with NTC and essential for Prp2-mediated displacement of the SF3A/SF3B complexes (Yeh *et al*. 2011), causes intron retention preferentially in introns with a long distance between the BPS and 3'ss. Given the placement of *cwf22* in the activated spliceosome (Yan *et al*. 2016), a BPS-to-3'ss length-sensing mechanism for cwf22 is plausible, though it may be mediated by additional proteins such as the non-essential N-terminal region of *prp10* (Habara *et al*. 1998).

Recently, the atomic structure of a spliceosome immediately after exon-ligation revealed the non-Watson-Crick base-pair interactions that must occur between the 5'ss G(+1) and the 3'ss

G(-1) for exon ligation (Wilkinson *et al*. 2017) and suggesting a mechanism by which Prp22 may proofread and reject non-optimal substrates at this interaction site. This structure also suggested how 5'ss A(+1) and 3'ss C(-1) mutations could similarly satisfy these non-Watson-Crick interactions, reconciling the decades-old observation that the major spliceosome can splice AT-AC introns (Parker and Siliciano 1993; Chanfreau *et al*. 1994; Dietrich *et al*. 1997). Interestingly, we do not see unique accumulation of *lea1*_intron1 or *taf2*_intron1 retention in the *prp22*-V801 allele, consistent with *prp22* tolerating AT-AC introns. We observed preferential intron accumulation in introns with weak 5'ss, consistent with previous observations on the ability of *prp22* to proofread and discard substrates at the lariat intermediate stage (Mayas *et al*. 2006; Semlow and Staley 2012). If *prp22* does not reject AT-AC introns, it remains to be elucidated which steps of spliceosome assembly and/or catalysis and which factors are most responsible for rejection of potential AT-AC splice sites, as there likely exists many more potential AT-AC splice site pairs than actually get utilized by the cell. A simple explanation is that these introns fail to efficiently recruit U1 snRNP to the 5'ss. An estimation of the prevalence (or absence) of AT-AC intron lariat intermediates, relative to the prevalence of potential AT-AC introns, may clarify at which step AT-AC segments are rejected from the splicing pathway. More generally, an investigation to the degree of accumulation of canonical (GT-AG) intron splice intermediates in optimal and non-optimal introns may be necessary to better understand the proofreading capabilities of the identified factors at different steps in splicing.

Historically, similar analyses of proofreading capabilities have often been performed by measuring the relative abundances of unspliced, intermediate, and spliced product of a labelled pre-mRNA substrate after *in vitro* reconstitution of the splicing reaction from yeast or human cell extracts (Hicks *et al*. 2005; Dunn and Rader 2014). However, fission yeast extracts largely

contains late-stage spliceosomes which have proven thus far incompetent for full reconstitution of the splicing reaction *in vitro* (Huang *et al*. 2002; Dunn and Rader 2014). Use of conditional alleles to stall spliceosomes may be lead to insights as to why fission yeast extracts are incompetent for splicing. Recently published methods, including spliceosome profiling (Burke *et al*. 2018; Chen *et al*. 2018) and splice-isoform targeted sequencing (Xu *et al*. 2018), were utilized to simultaneously measure *in vivo* splice intermediates and splice products on a genome-wide scale in yeast. These methods may also yield higher-resolution insights to the proofreading capabilities of these mutants in various spliceosome stages.

Methods

Screening library for splicing defects

An arrayed library of approximately 2000 ts-isolates, originally created by the group of P. Nurse via nitrosoguanidine mutagenesis and replica plating, was obtained from J. Armstrong (Armstrong *et al*. 2007) and re-arrayed and stocked into 384-well format. We then used robotically assisted protocols to grow strains and perform RT-PCR and deep sequencing as previously described (Larson *et al*. 2016) with the following exceptions to accommodate the growth characteristics of ts-strains: All strains were grown at the permissive temperature of 25°C for 3 days, rather than 2 days at 32°C, for initial pinning from glycerol stocks onto solid YES media. Cells were grown for 3 days at 25°C to saturation in liquid media to normalize cell density. After cultures reached saturation, cultures were back-diluted to OD_{600}~0.1 and grown for ~12 hours at 25°C (until OD_{600} ~0.8) before shifting cells to the non-permissive temperature of 37°C for 15 minutes. This temperature shift was achieved by mixing 100uL of culture to 100uL of pre-warmed culture plates containing 100uL of 45°C liquid YES media. Cells were then

incubated with shaking at 37°C for an additional 15 minutes prior to cell collection by centrifugation as previously described (Larson *et al*. 2016).

The RT-PCR primers and thermocycling conditions used for screening the splicing phenotype of *fkh1*-intron3, *rpl39*-intron1, and *lea1*-intron1 are listed in tableS1. We took careful care to prevent primer dimer formation during PCR reactions by employing empirically determined PCR conditions to be as follows: The initial PCR reaction containing gene-specific primers with plate-specific barcoded primers contained 10mM Tris (pH 8.3), 50mM KCl, 1.5mM $MgCl_2$ 0.2mM dNTPs, 0.25x SYBR Green I, 250nM Fwd and Rev primer, 150nM Hot-Start aptamer (Noma *et al*. 2006) (TableS1) and 1x Taq polymerase. The first PCRs for each gene-target were pooled into a single 384-well plate, preserving 384-well position, and cleaned via glass-fiber columns (Whatman cat#7700-1101) by adding 2 volumes DNA binding buffer (5M Guanidinium HCl, 30% isopropanol, 90mM KOH, 150mM acetic acid) to the pooled samples followed by applying samples to the columns by centrifugation (2min, 2000xg) and 2 sequential wash steps with wash buffer (80% ethanol, 10mM Tris) and a dry spin. Samples were eluted in The original sample volume and diluted 5 fold prior to the next PCR which was used to append Nextera v2 indices to samples in a well-specific manner. This PCR was performed with 10mM Tris (pH 8.3), 50mM KCl, 1.5mM $MgCl_2$ 0.2mM dNTPs, 0.25x SYBR Green I, 250nM Fwd and Rev primer, and 1x Taq with cycling conditions described in tableS1. PCR reactions for each gene-target were pooled, concentrated via ethanol precipitation, cleaned via Zymo-25 glass fiber column, and resuspended in water. Libraries were sequenced to ~50M reads per gene target (~10,000 reads per individual sample per gene target) on a NextSeq lane with 75bp single-end reads.

Data processing for screen

Reads corresponding to individual samples were demultiplexed into separate fastq files via a custom script which takes into account Illumina dual index reads as well as the 8-12 base plate-barcode that is part of the insert read. Fastq files were aligned with the STAR aligner (Dobin *et al*. 2013) to a manually generated genome file containing only the targeted genes with their splice sites annotated. We had to manually re-annotate the splice sites in the *lea1* gene, as the Ensembl genome annotations (ASM294v2) erroneously list GT-AG splice sites for this gene, while we and others (Chen *et al*. 2018) empirically determined that splicing occurs at AT-AG splice sites. We used default STAR aligner parameters with the following exceptions: {--clip5pNbases 23 --alignEndsType EndToEnd --clip3pAdapterSeq {ADAPTERSEQ} --genomeLoad LoadAndKeep --limitBAMsortRAM 2000000000 --alignIntronMax 290}. For each sample, spliced read counts were taken from the SJ.out.tab file created by the aligner, while read counts for unspliced read counts were determined using bedtools coverage (Quinlan and Hall 2010) to sum the number of reads which read into the intron. Spliced and unspliced read counts for biological replicates were combined, and strains which had a combined (unspliced + spliced) read count under 1000 (*fkh1*-intron3), 1000 (*rpl39*-intron1), or 100 (*lea1*-intron1) were discarded from further analysis. We determined the strains which have a significantly changed splice index (SI, equal to unspliced/spliced read count), using the statistical approach previously described to account for biases that change as a function of read depth and gene-target. Unlike our previous description of this statistical approach, in this work we performed inference using a right-sided Z-test from the interpolated Z-score, as we observed that the strains which were inferred as significantly increasing splicing efficiency (left-sided significant) are likely false-positives as they were generally unreproducible between the biological replicates. We attribute this to stochastic noise between biological replicates with low reads counts of the minor (unspliced)

isoform which is unaccounted for in the statistical model. Multiple hypothesis correction was performed by converting P-values to Q-values (Storey and Tibshirani 2003), controlling for FDR at 0.05. For purposes of presenting the screen data in an easily interpretable manner (Figure2), we transformed the relative splice index (ratio of unspliced/spliced reads) into an estimate of percent spliced in (PSI) by the following function:

$$PSI = 100\left(\frac{SI_{relative}SI_{median}}{1 + SI_{relative}SI_{median}}\right)$$

where $SI_{relative}$ is a ratio of a particular sample's SI to an estimate of the true wildtype SI given read a particular depth ($\mu_{interpolated}$). SI_{median} is the median SI across all samples at all read depths. The $\mu_{interpolated}$ is used as a normalization factor to account for the variation in SI that is a function of read count, possibly originating from length-dependent and RNA-input-amount-dependent PCR biases. Calculation of $\mu_{interpolated}$ has been previously described previously (Larson *et al*. 2016).

Mapping mutations via 'bulk-segregant analysis first' approach

For 3 strains, we performed bulk segregant analysis to isolate the causative allele(s) to separate bulks of ts and non-ts phenotype, followed by genotyping the bulks by whole genome sequencing. For this approach, we first outcrossed the ts-strain to a WT strain (ED666, Bioneer) and isolated spores via glusalase treatment as described previously (Forsburg and Rhind 2006) with the exception that we found empirically that glusalase treatment required only 0.2% glusalase (Perkin Elmer cat#NEE154001EA) as the final concentration. We diluted and plated the spores to single colonies to isolate ~48 F1 isolates for each mating. Each isolate was then replica plated in quadruplicate at 25°C and 37°C to assay for ts-phenotype. Spores which were ts and non-ts were pooled into separate bulks of about *n*=20 and grown to saturation in a 2mL culture at 25°C. After cell collection and DNA isolation (Lucigen/Epicentre cat#MPY80200), we

produced WGS libraries (TruSeq PCR free kit) to genotype the bulks. Bulks were sequenced to ~30X coverage using 75bp single-end reads on the Illumina NextSeq platform. Reads were aligned to the reference genome (ASM294v2) using bwa-mem aligner (Li and Durbin 2009) with default settings. Variant calling was performed using CRISP (Bansal 2010) with an appropriate pool size (*n*) used as the --poolsize parameter for pooled genotyping. Read counts for the reference and non-reference allele were summed for each variant for each bulk.

We then performed genome-wide association analysis. In order to properly account for noise due to random sampling of both reads and spores in the bulk, we first estimated allele counts in each bulk as follows: If the read count at a variant is larger than the bulk size, the sampling error is likely driven mostly by the bulk size. Therefore, to estimate allele counts in the bulk, we coerced the allelic-ratio of read counts for each variant to the number of individuals in the bulk via simple rounding. For example, an allelic ratio in read counts of 18p:22q will correspond to an allele counts of 9p:11q if the bulk size is 20 individuals. If the read count at a variant is smaller than the bulk size, the random sampling error is likely driven by low read count so we simply used the allelic ratio of read counts as the number of allele counts in the population. Variants with a minor allele frequency less than 0.15 were discarded. Variants which were present in bulks from separate matings were categorized as markers which may be useful for mapping the causative loci but are unlikely to be the exact causative variant from a novel mutation. To look for statistical significant enrichments, we used the allele counts of each variant to create a contingency table for a one-sided Fisher exact test, looking for enrichment of the non-reference allele into the ts-bulk.

Mapping mutations via 'whole genome sequencing first' approach

Each strain of the 52 strains with significantly changed splicing was grown to saturation at 25°C in 100uL cultures and DNA was isolated as described above with proportionally scaled down volumes and whole genome sequencing libraries were generated via a tagmentation protocol similar to previously published protocols (Picelli *et al*. 2014) using ~5ng DNA as input: 5ng of DNA in Tagmentation buffer and Tn5, and indexed adapters were appended via PCR. Libraries were sequenced either as 65+10 paired end reads on a Illumina NextSeq platform or as 38+38 paired end reads on an Illumina NextSeq platform to ~10X coverage of the genome. PCR duplicates (non-unique reads with respect to every sequenced base) were removed from the dataset prior to alignment, and reads were aligned using HiSat2 (Kim *et al*. 2015). Variant calling was performed using the GATK pipeline according to their 'best practices' protocols (Depristo *et al*. 2011) using the HaplotypeCaller tool with {–ploidy 1}. Variants that were shared by 4 or more strains were discarded as they likely do not represent novel mutations that are plausible candidates for the causative mutation. We choose 4 as a threshold to allow for mutations that could be present in 2 or 3 strains due to cross-contamination of the strain library. Of the novel mutations, we used the VariantEffectPredictor software (McLaren *et al*. 2016) to annotate the consequence of each mutation as coding or non-coding. We made predictions on the causative mutation (TableS2) for 13/27 strains if there existed a gene containing coding mutations that is a known splicing gene [belongs to the gene ontology category 'mRNA cis splicing, via spliceosome' (GO:0045292)] and is essential [associated with the PomBase gene annotation 'inviable cell population' (FYPO:0002059)]. In the case of 102B12 and 104J03 there existed two mutations in essential splicing genes, but in both strains one of the mutations was a shared mutation in the *cwf2* gene. Reasoning that the causative gene is more likely to be a unique

mutation, the causative mutation prediction listed in tableSX for these two strains is the other splicing essential gene mutation.

Confirming predicted mutations via bulk segregant analysis

We outcrossed the ts-strain of interest to a wildtype strain (ED666, Bioneer) and isolated spores via glusalase treatment as described above. The glusalase-treated spores were divided into equal volumes, each of which was used to inoculate a large bulks. Plating assays suggest that the population size of these bulks is >10000 individuals. The bulks were then grown at either 25°C or 37°C in 2mL of YES media to either maintain the initial allele frequency in the bulk or to select against the causative mutation. Cells were collected after 3 days of growth or until saturated (OD_{600} ~ 12) and DNA was isolated as described above. We genotyped the candidate mutation in each bulk by PCR-targeted deep sequencing. PCR primers with Illumina overhangs that flank the locus of interest were designed and PCR was performed with the same reaction conditions as used for screening, with the exception that each PCR was with the following thermocycler conditions listed in tableSX. As with the screen protocol, PCRs were diluted and used in a sequential PCR for indexing with identical reaction conditions and cleanup procedures. Reads were deep sequenced on a MiSeq platform with 150bp single end reads to a depth of ~20000 reads per sample. After alignment to the genome, allelic-ratio of read counts was determined and linkage of the candidate mutation to the ts-phenotype was inferred if the allele is relatively depleted (0-10% allele frequency) from the 37°C bulk.

RNA-seq

25mL cultures of each strain were grown in a single replicate, with the exception of the wildtype strain (ED666, Bioneer) which was grown in two 25mL replicates. Cultures were grown at 25°C to OD_{600} ~ 0.8 and subsequently shifted to a shaking water bath incubator set at

37°C for 15 minutes before cell collection by vacuum filtration and snap frozen until later processing. RNA was isolated by hot-phenol extraction. 1ug of RNA, without DNAse treatment, was used as input into a poly-A selection protocol (NEB cat#E7490S) and a stranded RNA-seq protocol (NEB cat#E7760S), following the manufacturer's instructions. Libraries were sequenced on a NextSeq platform with 38+38 paired end reads to a depth of 20M reads per sample. RNA-seq data is deposited with accession code ###.

Reads were aligned with the STAR aligner in two-pass mode with the following parameters: {--alignIntronMin 10 --alignIntronMax 1000 --sjdbFileChrStartEnd $junctions } where $junctions is a file from the aggregated first-pass spliced alignments to allow for sensitive detection of novel splice junctions. Counts of spliced reads at annotated and unannotated junctions were taken from the SJ.out.tab file generated by the aligner. Unspliced reads which infer an intron retention event were counted by using bedtools intersect to sum the counts of reads that aligned >=3 bases of an annotated intron. Spliced reads were categorized as either annotated, alt3'ss (annotated 5'ss, unannotated 3'ss), alt5'ss (unannotated 5'ss, annotated 3'ss), exon-skipping (annotated 5'ss and annotated 3'ss but in novel pairing) or new intron (unannotated 5'ss, unannotated 3'ss). Read counts of spliced and intron-retention events were normalized to the length the feature's potential mapping space, which in the case of intron retention reads is 2(R-O)+L where R is read length, O is minimum overhang length (3) and L is length of the intron. In the case of spliced reads, the mapping space is 2(R-O). Length-normalized read counts were used for all further analysis.

PSI is calculated for each intron as the length-normalized count of intron retention reads divided by the sum of length-normalized intron retention and annotated spliced reads. ΔPSI was calculated for each sample (less WT-rep1) as the difference in PSI values between PSI_{Mutant} and

$PSI_{WT\text{-}rep1}$. Motif scores were calculated for the 5'ss, branchsite, and 3'ss using a position weight matrix and log-odds scoring scheme (Lim and Burge 2001) using bases (-2 to +7) for 5'ss, (-4 to +2) for the branchsite, and (-6 to +2) for 3'ss. As branchsites for each intron are not annotated on PomBase, creating a position weight matrix first required identification branchsites. To accomplish this, we first created a position weight matrix using branch motif frequencies (PomBase) to search for the best match within a window (-40 to 0 from the 3'ss). From these branch locations, the new position weight matrix was created (-2 to +7). Estimates of transcription synthesis, splicing half-life, and transcript half-life were obtained from a published dataset (Eser *et al*. 2016).

Acknowledgements and Author Contributions

JA and NB donated the ts-library of strains. AL and BJF collected cells and isolated RNA for screen. BJF performed PCR, and data analysis for screen. AL performed bulk segregant analysis experiments for whole genome association mapping. BJF analyzed, performed all other experiments, made figures and wrote chapter.

Works Cited

Armstrong, J., N. Bone, J. Dodgson, and T. Beck, 2007 The role and aims of the FYSSION project. Briefings Funct. Genomics Proteomics 6: 3–7.

Bansal, V., 2010 A statistical method for the detection of variants from next-generation resequencing of DNA pools. Bioinformatics 26:.

Bitton, D. A., S. R. Atkinson, C. Rallis, G. C. Smith, D. A. Ellis *et al.*, 2015 Widespread exon-skipping triggers degradation by nuclear RNA surveillance in fission yeast. Genome Res. 884–896.

Boon, K.-L., T. Auchynnikava, G. Edwalds-Gilbert, J. D. Barrass, A. P. Droop *et al.*, 2006 Yeast Ntr1/Spp382 Mediates Prp43 Function in Postspliceosomes. Mol. Cell. Biol.

Burke, J., A. Longhurst, D. Merkurjev, J. Sales-Lee, B. Rao *et al.*, 2018 Spliceosome profiling visualizes the operations of a dynamic RNP in vivo at nucleotide resolution. Cell 173: 1014–1030.

Cazzola, M., M. Rossi, and L. Malcovati, 2013 Biologic and clinical significance of somatic mutations of SF3B1 in myeloid and lymphoid neoplasms. Blood.

Chanfreau, G., P. Legrain, B. Dujon, and A. Jacquier, 1994 Interaction between the first and last nucleotides of pre-mRNA introns is a determinant of 3′ splice site selection in S.cerevisiae. Nucleic Acids Res.

Chen, W., J. Moore, H. Ozadam, H. P. Shulha, N. Rhind *et al.*, 2018 Transcriptome-wide interrogation of the functional intronome by spliceosome profiling. Cell 173: 1031–1044.

Depristo, M. A., E. Banks, R. Poplin, K. V. Garimella, J. R. Maguire *et al.*, 2011 A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat. Genet. 43: 491–501.

Dietrich, R. C., R. Incorvaia, and R. A. Padgett, 1997 Terminal intron dinucleotide sequences do not distinguish between U2- and U12-dependent introns. Mol. Cell.

Dobin, A., C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski *et al*., 2013 STAR: Ultrafast universal RNA-seq aligner. Bioinformatics 29: 15–21.

Dunn, E. A., and S. D. Rader, 2014 Preparation of yeast whole cell splicing extract. Methods Mol. Biol.

Eser, P., L. Wachutka, K. C. Maier, C. Demel, M. Boroni *et al*., 2016 Determinants of RNA metabolism in the Schizosaccharomyces pombe genome. Mol. Syst. Biol. 12: 857–857.

Fair, B. J., and J. A. Pleiss, 2017 The power of fission: yeast as a tool for understanding complex splicing. Curr. Genet.

Forsburg, S. L., and N. Rhind, 2006 Basic methods for fission yeast. Yeast 23: 173–183.

Galej, W. P., T. H. D. Nguyen, A. J. Newman, and K. Nagai, 2014 Structural studies of the spliceosome: Zooming into the heart of the machine. Curr. Opin. Struct. Biol.

Galej, W. P., C. Oubridge, A. J. Newman, and K. Nagai, 2013 Crystal structure of Prp8 reveals active site cavity of the spliceosome. Nature 493: 638–43.

Habara, Y., S. Urushiyama, T. Tani, and Y. Ohshima, 1998 The fission yeast prp10(+) gene involved in pre-mRNA splicing encodes a homologue of highly conserved splicing factor, SAP155. Nucleic Acids Res. 26: 5662–5669.

Hartwell, L. H., C. S. McLaughlin, and J. R. Warner, 1970 Identification of ten genes that control ribosome formation in yeast. MGG Mol. Gen. Genet. 109: 42–56.

Hicks, M. J., B. J. Lam, and K. J. Hertel, 2005 Analyzing mechanisms of alternative pre-mRNA splicing using in vitro splicing assays. Methods.

Hossain, M. A., and T. L. Johnson, 2014 Using yeast genetics to study splicing mechanisms.

Methods Mol. Biol.

Huang, T., J. Vilardell, and C. C. Query, 2002 Pre-spliceosome formation in S.pombe requires a stable complex of SF1-U2AF59-U2AF23. EMBO J. 21: 5516–5526.

Jamieson, D. J., B. Rahe, J. Pringle, and J. D. Beggs, 1991 A suppressor of a yeast splicing mutation (prp8-1) encodes a putative ATP-dependent RNA helicase. Nature.

Käufer, N. F., and J. Potashkin, 2000 Analysis of the splicing machinery in fission yeast: a comparison with budding yeast and mammals. Nucleic Acids Res. 28: 3003–3010.

Kim, D.-U., J. Hayles, D. Kim, V. Wood, H.-O. Park *et al*., 2010 Analysis of a genome-wide set of gene deletions in the fission yeast Schizosaccharomyces pombe. Nat. Biotechnol. 28: 617–623.

Kim, D., B. Langmead, and S. L. Salzberg, 2015 HISAT: A fast spliced aligner with low memory requirements. Nat. Methods 12: 357–360.

Kim, S. H., and R. J. Lin, 1993 Pre-mRNA splicing within an assembled yeast spliceosome requires an RNA-dependent ATPase and ATP hydrolysis. Proc. Natl. Acad. Sci. U. S. A.

Kuhn, A. N., and N. F. Käufer, 2003 Pre-mRNA splicing in Schizosaccharomyces pombe: regulatory role of a kinase conserved from fission yeast to mammals. Curr. Genet. 42: 241–251.

Larson, A., B. J. Fair, and J. A. Pleiss, 2016 Interconnections Between RNA-Processing Pathways Revealed by a Sequencing-Based Genetic Screen for Pre-mRNA Splicing Mutants in Fission Yeast. G3 (Bethesda). 6: 1513–23.

Li, H., and R. Durbin, 2009 Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25: 1754–60.

Libri, D., N. Graziani, C. Saguez, and J. Boulay, 2001 Multiple roles for the yeast

SUB2/yUAP56 gene in splicing. Genes Dev.

Lim, L. P., and C. B. Burge, 2001 A computational analysis of sequence features involved in recognition of short introns. Proc. Natl. Acad. Sci. U. S. A. 98: 11193–8.

Lin, R. J., A. J. Lustig, and J. Abelson, 1987 Splicing of yeast nuclear pre-mRNA in vitro requires a functional 40S spliceosome and several extrinsic factors. Genes Dev.

Lin, P. C., and R. M. Xu, 2012 Structure and assembly of the SF3a splicing factor complex of U2 snRNP. EMBO J.

Liu, H.-L., and S.-C. Cheng, 2012 The Interaction of Prp2 with a Defined Region of the Intron Is Required for the First Splicing Reaction. Mol. Cell. Biol.

Lustig, a J., R. J. Lin, and J. Abelson, 1986 The yeast RNA gene products are essential for mRNA splicing in vitro. Cell 47: 953–63.

Lybarger, S., K. Beickman, V. Brown, N. Dembla-Rajpal, K. Morey *et al*., 1999 Elevated levels of a U4/U6.U5 snRNP-associated protein, Spp381p, rescue a mutant defective in spliceosome maturation. Mol. Cell. Biol.

Matera, A. G., and Z. Wang, 2014 A day in the life of the spliceosome. Nat. Rev. Mol. Cell Biol.

Mayas, R. M., H. Maita, and J. P. Staley, 2006 Exon ligation is proofread by the DExD/H-box ATPase Prp22p. Nat. Struct. Mol. Biol.

Mayerle, M., M. Raghavan, S. Ledoux, A. Price, N. Stepankiw *et al*., 2017 Structural toggle in the RNaseH domain of Prp8 helps balance splicing fidelity and catalytic efficiency. Proc. Natl. Acad. Sci. 114: 4739–4744.

McLaren, W., L. Gil, S. E. Hunt, H. S. Riat, G. R. S. Ritchie *et al*., 2016 The Ensembl Variant Effect Predictor. Genome Biol. 17:.

Network, T. C. G. A., 2012 Comprehensive molecular portraits of human breast tumors. Nature.

Noble, S. M., and C. Guthrie, 1996 Identification of novel genes required for yeast pre-mRNA splicing by means of cold-sensitive mutations. Genetics 143: 67–80.

Noma, T., K. Sode, and K. Ikebukuro, 2006 Characterization and application of aptamers for Taq DNA polymerase selected using an evolution-mimicking algorithm. Biotechnol. Lett. 28: 1939–1944.

Parker, R., and P. G. Siliciano, 1993 Evidence for an essential non-Watson-Crick interaction between the first and last nucleotides of a nuclear pre-mRNA intron. Nature.

Picelli, S., A. K. Björklund, B. Reinius, S. Sagasser, G. Winberg *et al*., 2014 Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. Genome Res. 24: 2033–2040.

Potashkin, J., R. Li, and D. Frendewey, 1989 Pre-mRNA splicing mutants of Schizosaccharomyces pombe. EMBO J. 8: 551–9.

Qin, D., L. Huang, A. Wlodaver, J. Andrade, and J. P. Staley, 2016 Sequencing of lariat termini in S. cerevisiae reveals 5' splice sites, branch points, and novel splicing events. RNA.

Quesada, V., A. J. Ramsay, and C. Lopez-Otin, 2012 Chronic lymphocytic leukemia with SF3B1 mutation. N. Engl. J. Med.

Quinlan, A. R., and I. M. Hall, 2010 BEDTools: A flexible suite of utilities for comparing genomic features. Bioinformatics 26: 841–842.

Rauhut, R., P. Fabrizio, O. Dybkov, K. Hartmuth, V. Pena *et al*., 2016 Molecular architecture of the Saccharomyces cerevisiae activated spliceosome. Science (80-.).

Roshbash, M., P. K. Harris, J. L. Woolford, and J. L. Teem, 1981 The effect of temperature-sensitive RNA mutants on the transcription products from cloned ribosomal protein genes of yeast. Cell 24: 679–686.

Schellenberg, M. J., T. Wu, D. B. Ritchie, S. Fica, J. P. Staley *et al*., 2013 A conformational switch in PRP8 mediates metal ion coordination that promotes pre-mRNA exon ligation. Nat. Struct. Mol. Biol.

Schwer, B., 2008 A Conformational Rearrangement in the Spliceosome Sets the Stage for Prp22-Dependent mRNA Release. Mol. Cell.

Schwer, B., and C. H. Gross, 1998 Prp22, a DExH-box RNA helicase, plays two distinct roles in yeast pre-mRNA splicing. EMBO J.

Semlow, D. R., and J. P. Staley, 2012 Staying on message: Ensuring fidelity in pre-mRNA splicing. Trends Biochem. Sci.

Storey, J. D., and R. Tibshirani, 2003 Statistical significance for genomewide studies. Proc. Natl. Acad. Sci. 100: 9440–9445.

Taggart, A. J., A. M. DeSimone, J. S. Shih, M. E. Filloux, and W. G. Fairbrother, 2012 Large-scale mapping of branchpoints in human pre-mRNA transcripts in vivo. Nat. Struct. Mol. Biol. 19: 719–721.

Tsai, R. T., R. H. Fu, F. L. Yeh, C. K. Tseng, Y. C. Lin *et al*., 2005 Spliceosome disassembly catalyzed by Prp43 and its associated components Ntr1 and Ntr2. Genes Dev.

Tseng, C. K., H. L. Liu, and S. C. Cheng, 2011 DEAH-box ATPase Prp16 has dual roles in remodeling of the spliceosome in catalytic steps. RNA.

Turunen, J. J., E. H. Niemelä, B. Verma, and M. J. Frilander, 2013 The significant other: Splicing by the minor spliceosome. Wiley Interdiscip. Rev. RNA.

Urushiyama, S., T. Tani, and Y. Ohshima, 1996 Isolation of novel pre-mRNA splicing mutants of Schizosaccharomyces pombe. Mol. Gen. Genet. 253: 118–127.

Vijayraghavan, U., M. Company, and J. Abelson, 1989 Isolation and characterization of pre-

mRNA splicing mutants of Saccharomyces cerevisiae. Genes Dev. 3: 1206–1216.

Villa, T., and C. Guthrie, 2005 The Isy1p component of the NineTeen Complex interacts with the ATPase Prp16p to regulate the fidelity of pre-mRNA splicing. Genes Dev.

Webb, C. J., and J. A. Wise, 2004 The Splicing Factor U2AF Small Subunit Is Functionally Conserved between Fission Yeast and Humans. Mol. Cell. Biol. 24: 4229–4240.

Wilkinson, M. E., S. M. Fica, W. P. Galej, C. M. Norman, A. J. Newman *et al*., 2017 Postcatalytic spliceosome structure reveals mechanism of 3 ’-splice site selection. Science (80-.).

Will, C. L., and R. Lührmann, 2011 Spliceosome structure and function. Cold Spring Harb. Perspect. Biol. 3:.

Xu, H., B. J. Fair, Z. Dwyer, M. Gildea, and J. A. Pleiss, 2018 Multiplexed Primer Extension Sequencing Enables High Precision Detection of Rare Splice Isoforms. BioRxiv.

Yan, C., J. Hang, R. Wan, M. Huang, C. C. L. Wong *et al*., 2015 Structure of a yeast spliceosome at 3.6-angstrom resolution. Science 349: 1182–91.

Yan, C., R. Wan, R. Bai, G. Huang, and Y. Shi, 2016 Structure of a yeast activated spliceosome at 3.5 Å resolution. Science (80-.).

Yeh, T.-C., H.-L. Liu, C.-S. Chung, N.-Y. Wu, Y.-C. Liu *et al*., 2011 Splicing Factor Cwc22 Is Required for the Function of Prp2 and for the Spliceosome To Escape from a Futile Pathway. Mol. Cell. Biol.

Chapter4: Multiplexed Primer Extension Sequencing Enables High Precision Detection of Rare Splice Isoforms

Alternative citation for this chapter:

*Denotes equal contribution

Xu H*, Fair BJ*, Dwyer Z, Gildea M, Pleiss JA. Multiplexed primer extension sequencing enables high precision detection of rare splice isoforms. 2018. bioRxiv 331629; dos: https://doi.org/10.1101/331629.

Abstract

Targeted RNA-sequencing aims to focus coverage on areas of interest that are inadequately sampled in standard RNA-sequencing experiments. Here we present a novel approach for targeted RNA-sequencing that uses complex pools of reverse transcription primers to enable sequencing enrichment at user-selected locations across the genome. We demonstrate this approach by targeting pre-mRNA splice junctions in *S. cerevisiae*, revealing high-precision detection of splice isoforms, including rare pre-mRNA splicing intermediates.

Main text

RNA sequencing (RNA-Seq) has greatly expanded our understanding of the variety of splice isoforms that can be generated within a cell. Identification of the small subset of reads that span exon-exon junctions within transcripts has enabled the unambiguous detection of vast numbers of novel splice isoforms in scores of organisms (Barbosa-Morais *et al.* 2012; Merkin *et al.* 2012). Yet in spite of the power presented by this approach, the sequencing depth necessary to quantitatively detect many splicing events is significantly higher than most experiments

generate. While this limitation of whole-transcriptome profiling has been addressed in part by methods that utilize antisense probes (Mercer *et al.* 2011, 2014) or PCR enrichment (Blomquist *et al.* 2013) to target sequencing coverage to genomic regions of interest, a deeper understanding of the basic mechanisms by which splicing is regulated, and the pathological consequences of its mis-regulation, will be facilitated by methods that enable higher resolution and precise detection of splicing states within cells. Towards this end, we have designed and implemented a novel targeted sequencing method that enhances splice junction detection and allows for genome-wide resolution of splicing intermediates. Building upon the historically validated use of primer extension as a tool for assessing splicing status, we demonstrate the ability to multiplex primer extension assays and evaluate the products by deep sequencing, an approach we hereafter refer to as Multiplexed Primer Extension sequencing, or MPE-seq (Fig1A).

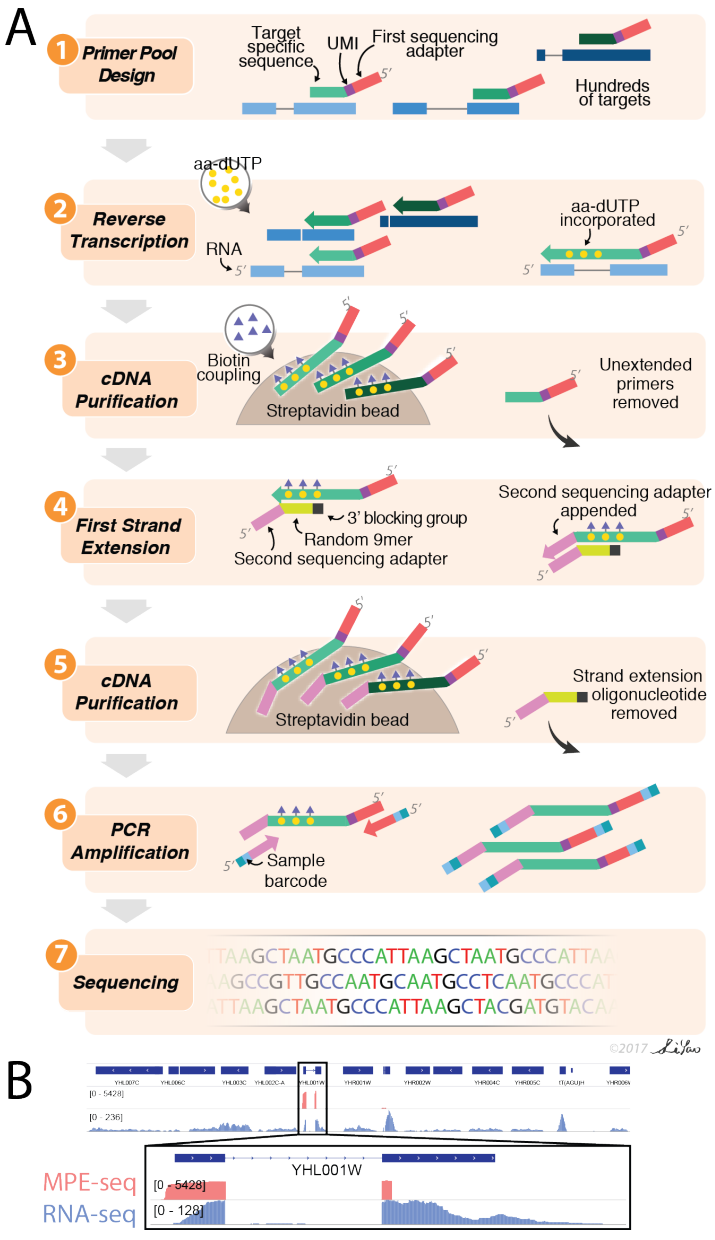
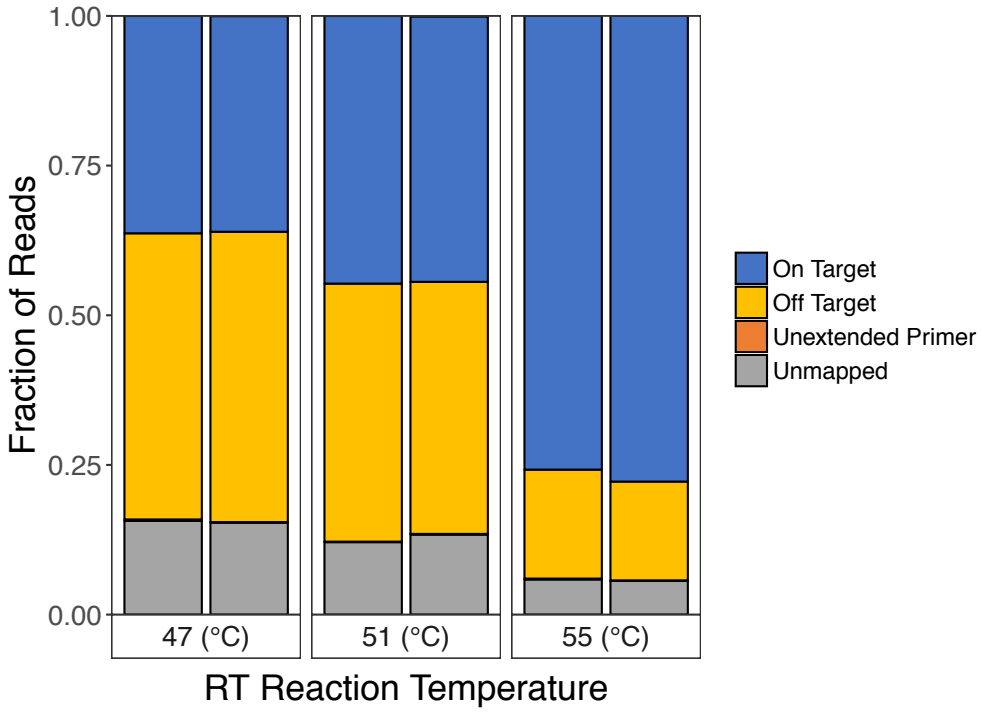

A
1 Primer Pool Design
Target specific sequence
UMI
First sequencing adapter
Hundreds of targets
2 Reverse Transcription
aa-dUTP
RNA
aa-dUTP incorporated
3 cDNA Purification
Biotin coupling
Streptavidin bead
Unextended primers removed
4 First Strand Extension
3' blocking group
Random 9mer
Second sequencing adapter
Second sequencing adapter appended
5 cDNA Purification
Streptavidin bead
Strand extension oligonucleotide removed
6 PCR Amplification
Sample barcode
7 Sequencing
TAAGCTAATGCCCATTAAGCTAATGCCCATTAA
AGCCGTTGCCAATGCAATGCCTCAATGCCCAT
ATTAAGCTAATGCCCATTAAGCTACGATGTACA
©2017
B
YHL001W
MPE-seq
RNA-seq
[0 - 5428]
[0 - 128]

Figure1: MPE-seq uses complex pools of reverse transcription primers to target sequencing to regions of interest (A) MPE-seq comprises the following steps: (1) Primers containing sequencing adapter overhangs and a target-specific region are designed for genomic regions of interest. (2) Primers are pooled and used to reverse transcribe total cellular RNA in the presence of amino-allyl dUTP (aa-dUTP), which allows (3) biotin coupling and purification of cDNAs from other reaction components. (4) A second sequencing adapter is appended at the 3' terminus of the cDNA through first strand extension using Klenow. Libraries are further purified (5), PCR amplified (6) and sequenced (7). (B) Genome browser screenshot of a targeted region in MPE-seq (pink) and conventional RNA-seq (purple)

The method we developed harbors two straightforward yet key features. First, user-selected primers are used to generate complementary DNA (cDNA) during a reverse transcription reaction, enabling targeting of RNA regions of interest. Each primer is appended with a next-generation sequencing adapter, as well as a unique molecular identifier (UMI) to alleviate artifacts associated with PCR amplification during library preparation (Kivioja *et al.* 2012). The use of elevated temperatures during the reverse transcription reaction minimizes non-specific primer annealing (FigS1), while the inclusion of derivatized nucleotides allows for the efficient purification of extended products and removal of excess primers. Secondly, a strand-extension step similar to template-switching (Zhu *et al.* 2001) is used to append the second sequencing adapter onto the 3' terminus of the cDNA molecules. Coupling this approach with paired-end sequencing allows for the simultaneous querying of the 5' and 3' ends of the cDNAs from targeted regions (see Methods for full details).



FigureS1: Elevated temperatures in reverse transcription reactions increase specificity. The fraction of on-target and off-target reads from replicate MPE-seq libraries generated from reverse transcription reactions performed at various temperatures. A small fraction of reads were categorized as "Unextended primer" which corresponds to short primer extension products (0-5 bases extended past the primer) and thus they were neither categorized as cDNAs derived from RNA targets or unamappable.

As an initial demonstration of MPE-seq, we examined pre-mRNA splicing in the budding yeast *Saccharomyces cerevisiae*. For each of the 309 annotated introns in the yeast genome, primers were systematically designed within a 50nt window immediately downstream of the 3' splice site, ensuring that short extensions would be sufficient to cross the splice junctions. Primers were pooled at equimolar concentration and MPE-seq libraries were generated using total cellular RNA from wildtype yeast and sequenced to a depth of only ~5 million reads. As a comparative reference, standard RNA-seq libraries were generated using poly-A selected RNA and sequenced to ~40 million reads. Whereas the standard RNA-seq libraries yielded read coverage that comprised full gene bodies across the transcriptome, MPE-seq coverage was

focused on the selected genes, precisely targeted to the regions upstream of the designed primers (Fig1B). Just over 75% of sequenced fragments from MPE-seq mapped to targeted regions (Fig2A), resulting on average in a greater than 100-fold enrichment in sequencing depth at these regions when compared with RNA-seq (Fig2B, FigS2). The fold-enrichment was similar across transcripts with a wide range of expression levels (FigS2), and from these data we extrapolate that a standard RNA-seq experiment would require ~500 million sequencing reads to achieve a similar level of coverage over the targeted regions as these 5 million MPE-seq reads provided.

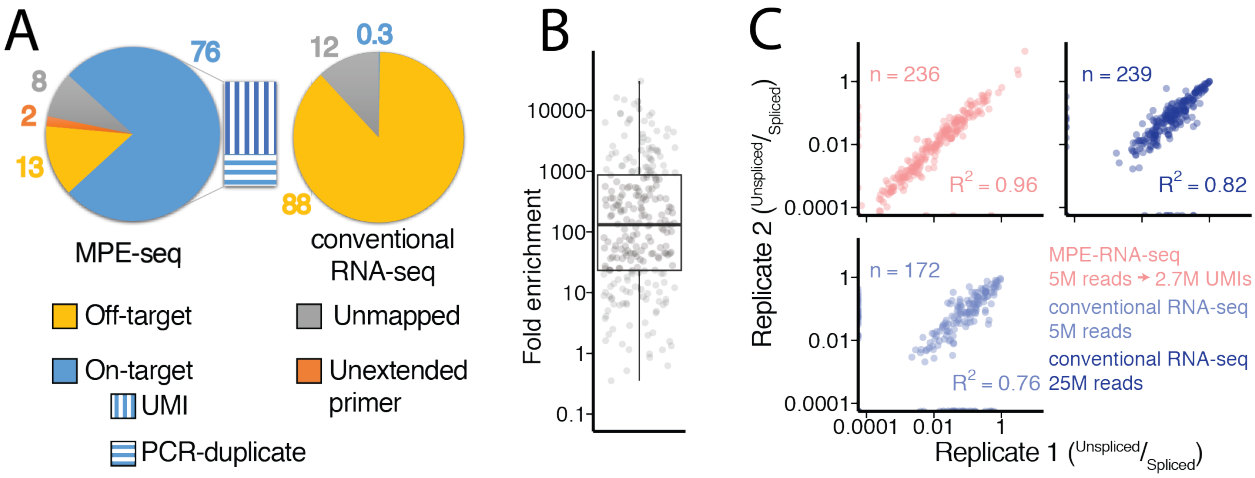
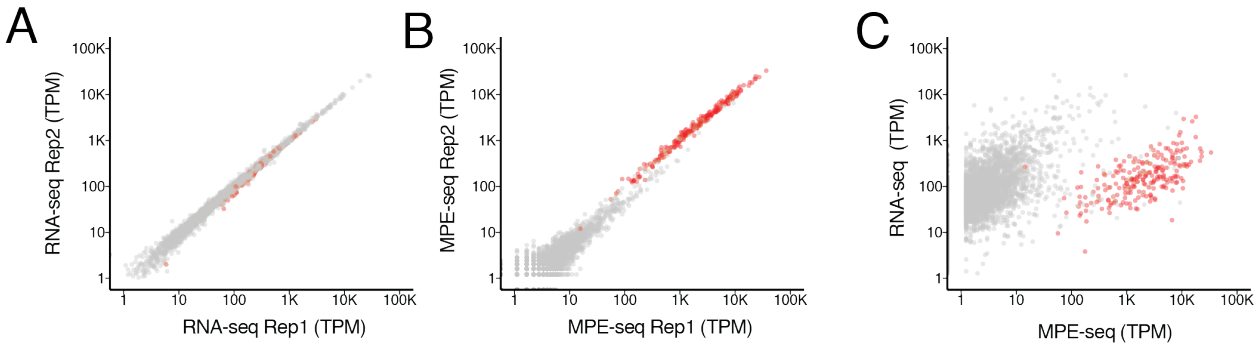


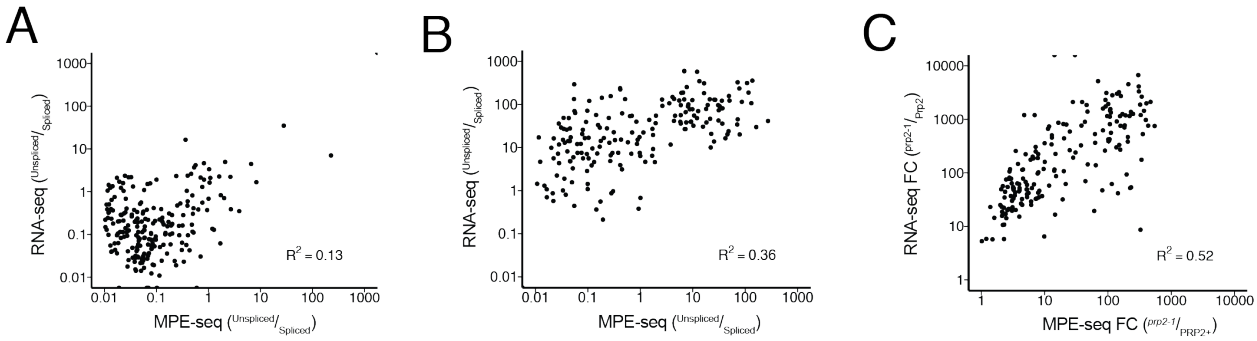
Figure2: MPE-seq enrichment enables high-precision measurements of splicing. (A) The percentage of reads mapped to target and off-target regions is depicted for MPE-seq and conventional RNA-seq. In MPE-seq a small fraction of reads were categorized as "Unextended primer" which corresponds to short primer extension products (0-5 bases extended past the primer) and thus they were not categorized as cDNAs derived from RNA targets. (B) Each point represents the fold enrichment of a target region in MPE-seq over conventional RNA-seq. (C) Scatter plots depict intron-retention measurements in replicate libraries in MPE-seq and conventional RNA-seq at matched or greater read depth.



FigS2: Expression measurements as determined by MPE-seq and RNA-seq. (A) A scatter plot depicts gene expression measurements (RNA-seq) in replicate datasets. Genes containing splice-events that were among those chosen for targeted sequencing are depicted in red. These targeted genes range in expression levels by orders of magnitude. (B) A scatter plot depicts gene expression measurements in replicate MPE-seq datasets (red). Similar to conventional RNA-seq, expression measurements in MPE-seq are highly reproducible between replicates, even for the small proportion of mis-priming events that map to off-target locations (grey). (C) A scatter plot depicts gene expression measurements in RNA-seq and MPE-seq. The right shift of targeted genes reflects successful enrichment of targets by orders of magnitude. The observation that even highly expressed genes as measured by RNA-seq are proportionally highly expressed in MPE-seq suggests that primers are not limiting during reverse transcription.

Given the increased read depth achieved over targeted regions using MPE-seq, we asked how well rare splice events were sampled. Because each primer extension event corresponded to a single RNA molecule, determining relative isoform expression was simplified because read counts did not need to be adjusted for mapping space as in standard analyses of RNA-seq data (see Methods). Importantly, the levels of intron retention determined from replicate libraries using MPE-seq showed superior internal reproducibility compared with the larger, replicate RNA-seq libraries (Fig2C), likely reflecting the sampling noise associated with RNA-seq data with reduced sequencing depth over the targeted regions. Moreover, while MPE-seq is not amenable to *de novo* discovery of novel splicing events across the entire genome, it did allow for the identification of scores of rare, previously unannotated splicing events at the targeted regions (data not shown), consistent with the significantly increased sensitivity of this approach. Nevertheless, while MPE-seq provided increased sensitivity and reproducibility of splicing

measurements, the intron retention levels determined from MPE-seq in a wildtype strain only modestly correlated with those determined by RNA-seq (FigS3A, FigS3B). Notably, this correlation improved when comparing how these techniques measured *changes* in splicing between samples assayed by the same methodology (FigS3C), presumably reflecting inherent biases (Zheng *et al.* 2011) (in fragmentation, ligation, PCR amplification, library size selection, etc.) present in one or both approaches that are internally well controlled.



FigS3: Splicing measurements as determined by MPE-seq and RNA-seq. **(A)** A scatter plot depicts intron-retention measurements in MPE-seq and conventional RNA-seq using wildtype (Prp2) RNA. **(B)** A scatter plot depicts intron-retention measurements in MPE-seq and conventional RNA-seq using RNA from a splicing mutant strain (*prp2-1*). **(C)** A scatter plot depicts the fold-change (*prp2-1*/Prp2) in intron-retention as measured by MPE-seq and conventional RNA-seq

With the increased resolution provided by this approach, we sought to determine whether we could detect splicing intermediates using MPE-seq. By identifying the locations of reverse transcription stops, primer extension reactions have historically been used to map a variety of biological features, including among others transcription start sites (TSSs) (Carey *et al.* 2013), and the locations of branch sites within the lariat intermediate species of the pre-mRNA splicing reaction (Padgett *et al.* 1985; Coombes and Boeke 2005) (Fig3A).

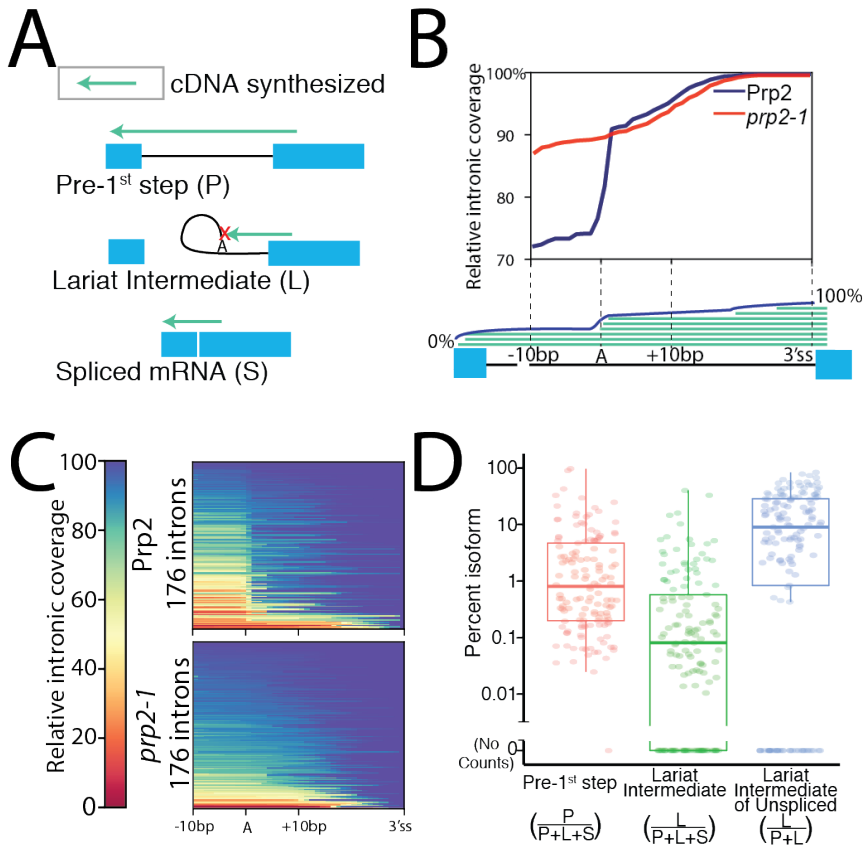
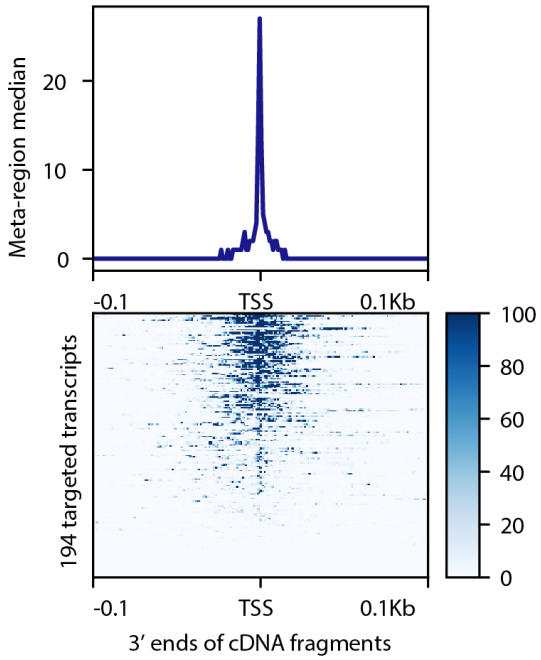


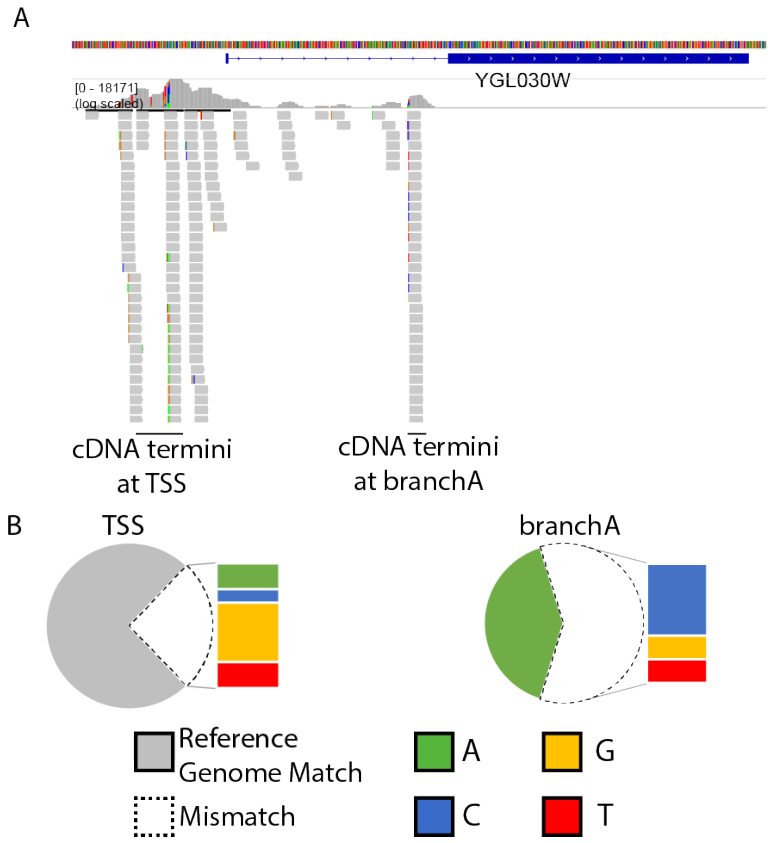
Figure3: MPE-seq enables genome-wide profiling of lariat intermediates. **(A)** Schematic depicting cDNA products derived from pre 1st step (P), lariat intermediate (L) and spliced mRNA (S) isoforms. **(B)** Meta-intron coverage plot surrounding predicted branchpoints in a wild-type (Prp2) and step1 splicing mutant strain (*prp2-1*). The region between the +10 position downstream of the annotated branchpoint and the 3' splice site (3'ss) was re-scaled for each intron. **(C)** Heatmap plots showing the relative coverage at each intron for which reads were detected. (D)Estimates of the relative abundance of each isoform for each targeted intron for which reads were detected.

Because the approach we developed anticipated the possibility of mapping the 3' ends of the cDNA molecules, we examined the locations of those generated by MPE-seq. As expected, the 3' ends of many cDNAs accumulated at the TSSs as determined by an orthologous method (Booth *et al.* 2016) (FigS4), indicating that reverse transcription generally proceeded to the 5' terminus of the RNA. Importantly, we also observed many cDNAs which terminated at or near the annotated branchpoint motifs within introns, with decreased read coverage upstream of the motifs, consistent with the inability of reverse transcriptase to read past the branched adenosine

in the lariat intermediate species (Fig3B, Fig3C). This drop in read coverage was not apparent in MPE-seq libraries generated from a strain harboring a conditional mutation in Prp2, an RNA helicase required for catalyzing the 1st step of splicing (Kim and Lin 1996), corroborating that many of these cDNAs originate from lariat intermediates. We noted that these lariat-intermediate derived cDNAs contain a unique signature of mismatches incorporated by reverse transcriptase at the branched adenosine (FigS5), which may serve as a unique mark for *de novo* identification of branch sites in organisms with less well annotated branch sites.



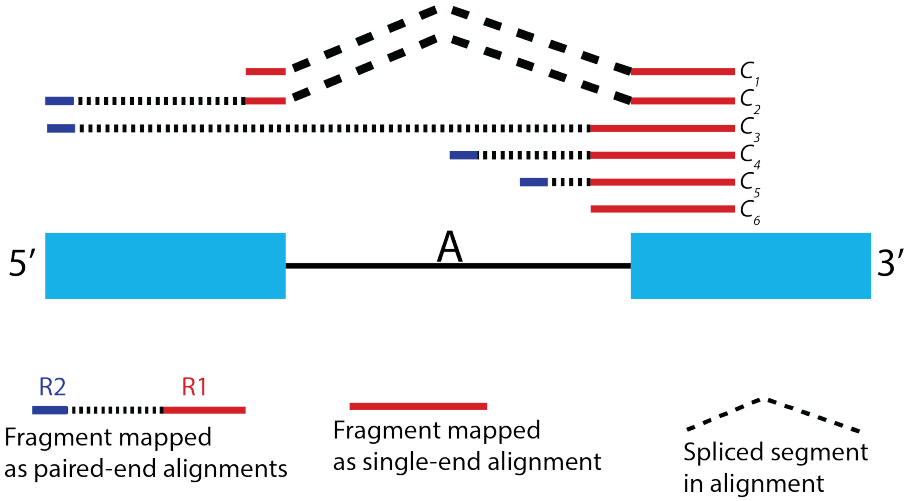
FigS4: Transcription start site profiling by MPE-seq. Metagene profile of 3' ends mapped by MPE-seq, centered around transcription start sites (TSS) as determined by PRO-cap, an orthologous method for mapping transcription start sites. The high abudance of read ends that pile up at TSSs indicates that MPE-seq can be used to profile cDNA termini.



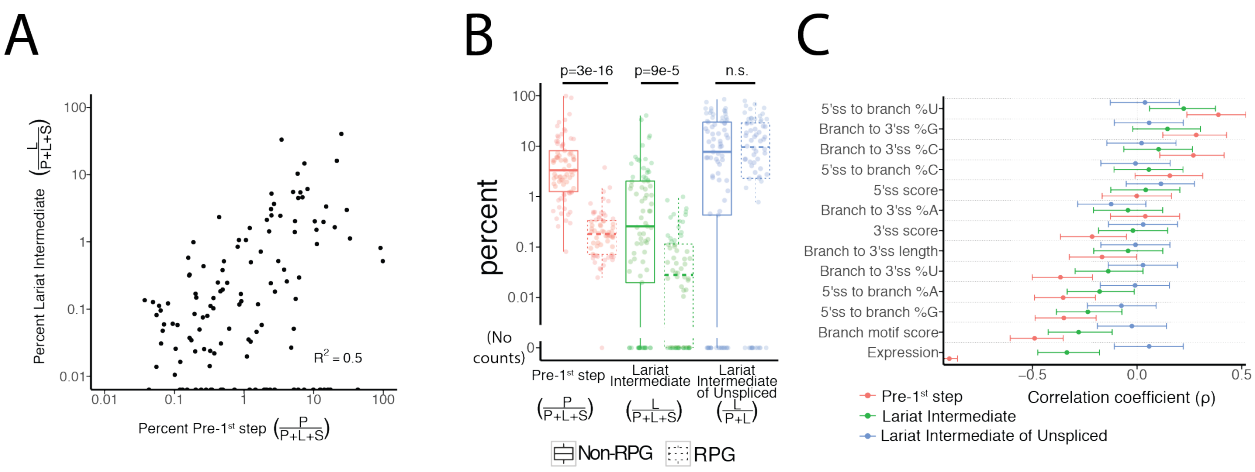
FigS5: Lariat-intermediate derived cDNAs contain a unique signature of mismatches incorporated by reverse transcriptase at the branched adenosine. (A) Genome browser screenshot of 3' reads from paired-end sequenced fragments illustrates the unique signature of non-templated base incorporation by reverse transcriptase at a branched-adenosine vs the 5' RNA terminus. (B) Genome-wide quantification of the mismatch frequencies at 3' cDNA termini near the TSS (left) versus at the annotated branchpoint (right).

The ability of MPE-seq to differentiate between unspliced isoforms in unfractionated cellular RNA provides a unique opportunity to investigate the relative efficiencies with which transcripts undergo the 1st and 2nd chemical steps in the splicing pathway. Our data revealed that ~10% of unspliced pre-mRNAs at steady state conditions were present in the lariat intermediate form genome-wide (see FigS6, methods), albeit with significant variation between individual pre-mRNAs (Fig3D). Remarkably, a strong correlation was observed between the relative levels of pre-mRNA and lariat intermediate species for a given transcript (FigS7A). Correlations were also observed between splice intermediate levels and transcript expression level, branch motif

strength, and host gene function (FigS7B, FigS7C); however no correlation was seen when considering the ratio of pre-1st step intermediate to lariat intermediate, a metric which we expect to reflect the relative catalysis rates of the 1st and 2nd step of splicing. A complete understanding of the determinants of *in vivo* splicing efficiency will require measurements of the kinetics of these individual steps rather than their steady state levels. The ability of MPE-seq to robustly differentiate pre-mRNA isoforms provides a powerful new opportunity to do just this.



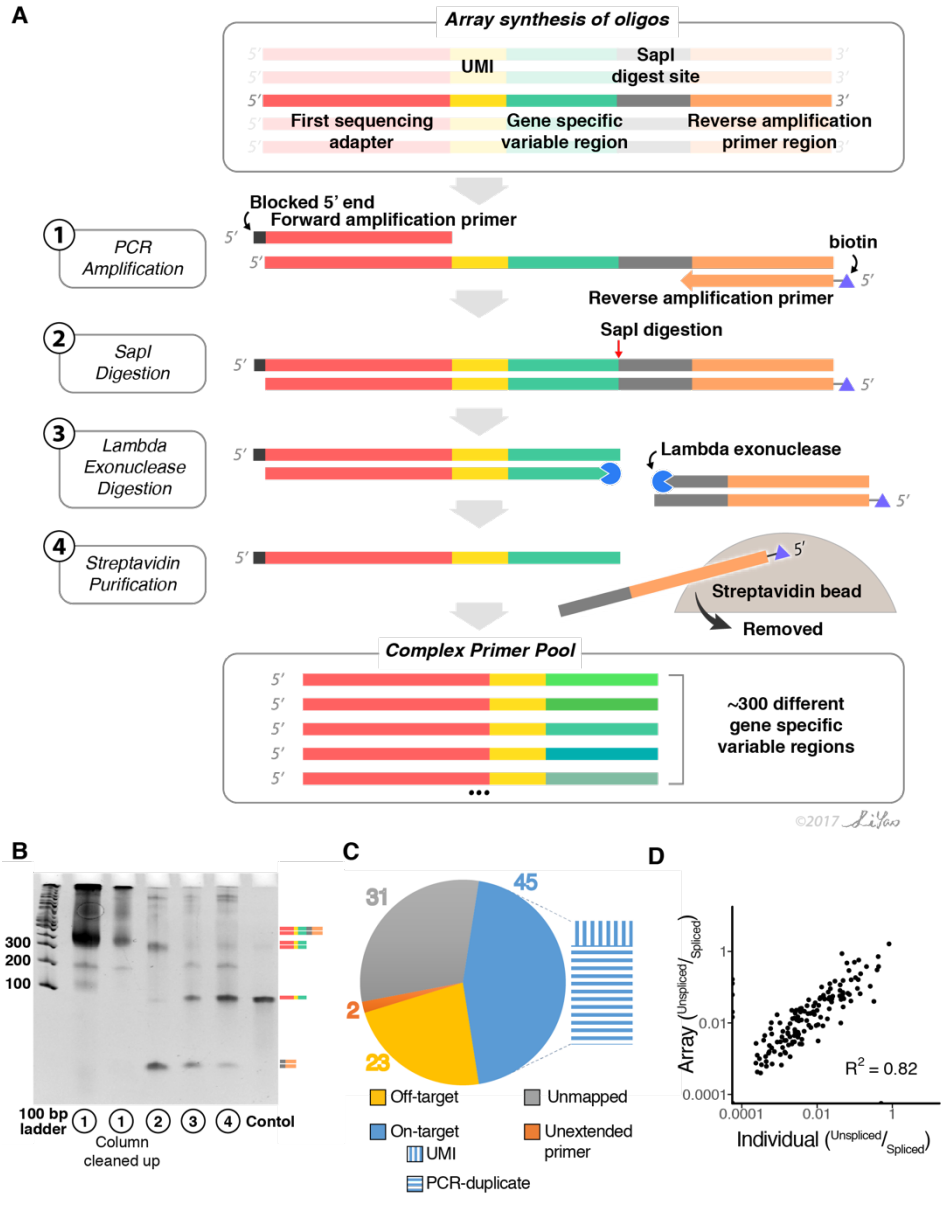
FigS6: Schematic for assigning reads to splice intermediate isoforms. To quantify the abundance of pre 1st step, lariat intermediate, and spliced isoforms for each targeted splice event, we categorized fragments into six classes based on paired-end read alignments. Fragments containing a splice junction (C_1 and C_2) are indicative of a spliced RNA (S). Fragments that are unspliced and traverse the branchpoint region (C_3) are classified as pre-1st step RNA (P). Fragments that are unspliced but terminate within a -3 to +5bp window from the previously determined branchpoint (C_4) are classified as lariat intermediate (L). Fragments that are unspliced and either terminate downstream of the branchpoint (C_5) or the terminus could not be mapped (C_6) are ambiguous between P and L. Therefore, for accounting purposes, the counts for these fragments were coerced into P and L classifications based on the ratio of P and L determined by unambiguous mappings (C_3 and C_4). See methods for more details.



FigS7: Transcript features that correlate with the abundance of lariat intermediate. (A) Scatter plot depicting the correlation between the abundance of pre-1st step RNA and the abundance of lariat intermediate **(B)** The abundance of pre-1st step RNA and lariat intermediate RNA is significantly correlated with classification of introns into those that are in ribosomal protein genes (RPG) and non-RPG. However, the abundance of lariat intermediate relative to pre-1st step RNA, a metric of the efficiency of the second step of splicing, does not correlate. **(C)** Spearman correlations of various features to the abundance of pre-1st step RNA, lariat intermediate, the abundance of lariat intermediate relative to pre-1st step RNA. Errorbars indicate 95% confidence intervals as estimated by Fisher transformation of Spearman's correlation coefficient.

Whereas our initial experiments were performed using individually synthesized oligonucleotides as primers, we sought to increase the utility of this approach by examining methods that would facilitate an increase in the number of targeted regions. Many commercial sources exist that allow for the cost-effective, array-based synthesis of pools of thousands of individual oligonucleotide sequences, so we developed an approach for the pooled synthesis of an equivalent set of the previously described primer sequences in order to test the effectiveness with which they could be used in MPE-seq. By appending a common sequence onto the 3' end of the desired primers and then using a protocol that included PCR amplification, restriction digestion, and targeted strand degradation (FigS8A, FigS8B), we readily prepared a sufficient quantity of single-stranded oligonucleotides with which to generate MPE-seq libraries. Importantly, MPE-seq libraries generated using primers synthesized by this approach also showed strong enrichment for the targeted regions, with levels on par with what we observed

using individually synthesized oligonucleotide primers (fig. FigS8C). Moreover, genome-wide splicing efficiencies determined from MPE-seq libraries generated using primers from this pooled synthesis were highly correlated with splice efficiencies derived from individually synthesized oligos (FigS8D), confirming the capacity of this approach to generate large numbers of unique primers for use in MPE-seq experiments.



FigS8: Array-based oligonucleotide synthesis can be used to generate primer pools for use in MPE-seq. (A) Obtaining adequate amounts of primer pools for MPE-seq from cost-effective array-based oligonucleotide synthesis can be achieved in four steps: (1) PCR amplification of the oligonucleotide synthesis

pool using a 5' blocked sense primer and a biotinylated antisense primer. (2) Restriction digestion to cleave off the PCR primer handle. (3) Lambda exonuclease digestion of free 5' ends. (4) Streptavidin purification of biotinylated PCR handle. The unbound fraction is the desired primer pool product. **(B)** Steps during the amplification and purification of array-synthesized primer pools are monitored via native gel electrophoresis. The control lane represents a pool of individually synthesized MPE-seq primers which did not require amplification and purification. **(C)** The percentage of reads mapped to target and off-target regions is depicted for MPE-seq using array-synthesized primers. In MPE-seq a small fraction of reads were categorized as "Unextended primer" which corresponds to short primer extension products (0-5 bases extended past the primer) and thus they were not categorized as cDNAs derived from RNA targets. **(D)** A scatter plot depicts intron-retention measurements in MPE-seq libraries which used individually synthesized primer pools and array-based synthesis of primer pools

Our work here demonstrates the capacity of MPE-seq to facilitate examinations of pre-mRNA splicing status in a targeted, cost-effective way that improves the precision and sensitivity of splice isoform detection. The improved sensitivity of this approach is perhaps best exemplified by our ability to detect the lariat intermediate products of the pre-mRNA splicing pathway. Though other methods, including splicoesome profiling, have reported large-scale detection of upstream exon splicing intermediates (Nojima *et al.* 2015; Burke *et al.* 2018; Chen *et al.* 2018), MPE-seq uniquely detects the lariat intermediate from unfractionated cellular RNA, and does not require laborious purification of complexes from cellular extracts. Our demonstration that oligonucleotides derived from pooled commercial syntheses can be used in this approach expands the types of applications to which MPE-seq could be applied in a cost-effective manner. While we see no *de facto* limitation to the species or number of unique primer sequences that could be used for MPE-seq, with increasing numbers of primers comes an increasing potential for their cross-reactivity with undesirable RNA targets, highlighting the importance of specificity in primer design. Similarly, the level of enrichment provided by this approach would vary as a function not only of the number of regions being targeted, but also of the distribution of the expression levels of those targets. We expect that the improved sensitivity, precision and flexibility of this approach will enable a higher-resolution understanding of the

pre-mRNA splicing pathway. Likewise, primer extension assays have also been used to assay RNA secondary structure after *in vitro* (Lucks *et al.* 2011) or *in vivo* (Rouskin *et al.* 2014) chemical probing, and we expect that MPE-seq could readily be adapted to RNA structure interrogation and other approaches where primer-extension assays or targeted RNA sequencing is applicable.

Methods

Strain Maintenance and Growth Conditions. Unless otherwise indicated, all experiments used the wild type (WT) *S. cerevisiae* strain BY4741 (*MATa*, *his2Δ1*, *leu2Δ0*, *met15Δ0*, *ura3Δ0*). Single colonies were inoculated into liquid YPD media and grown overnight at 30℃. Overnight cultures were then inoculated into fresh liquid YPD, seeding cultures at OD_{600} ~0.05. Cells were collected by vacuum filtration once cultures reached OD_{600} ~0.7 immediately followed by flash freezing in liquid nitrogen. Cell pellets were then stored at -80℃. For the temperature sensitive mutant *prp2-1*(Hartwell *et al.* 1970) we grew cultures as described above except at 25℃. Once cultures reached OD_{600} ~0.7, an equal volume of fresh 50℃ YPD media was added to shift cells to the non-permissive temperature of 37℃. The cultures were then maintained at 37℃ for 15 minutes before cell collection as described above.

MPE-seq Primer and Oligo Design.

Gene specific reverse transcription primer design. For each of the 309 annotated spliceosomal introns within the budding yeast genome (annotations obtained from UCSC SacCer3), a reverse transcription primer was designed within the first 50 nucleotides downstream of the intron. Targeting to this region ensured that short-read sequencing of the products generated from reverse transcription with these primers would cross the upstream exon-exon or exon-intron boundaries, enabling determination of the spliceing status. Primers were designed using

OligoWiz, a program initially developed for microarray probe design, but which enables the selection of primer sequences optimized for target specificity relative to a designated genomic background(Wernersson and Nielsen 2005)**.** We used the standalone version of OligoWiz with default parameters for short (24-26bp) oligo design to obtain optimal sequences within each 50bp window. To the 5' end of each of these sequences was appended two additional sequence elements: a random 7-nucleotide unique molecular index (UMI) which allows for the detection and removal of amplification artifacts arising from library preparation (Kivioja *et al.* 2012); and the P5XX region of the Illumina sequencing primer to enable the sequencing of the reverse transcription products. Each of these primers was individually synthesized by Integrated DNA Technologies (IDT).

Complex oligo mix amplification method: The above described oligo primers were batch synthesized as a pool on a OligoMix microarray by LC Sciences. These oligos are synthesized at vastly lower quantities than is required for cDNA synthesis in MPE-seq. To generate a quantity of primer pool that is sufficient, PCR amplification, along with several processing steps were used (FigS8A). This was enabled by addition of 2 key sequence elements appended onto the 3' end of the individually synthesized oligo primers detailed above. From the 5' to 3' direction: 1.) a SapI restriction site, and 2.) a PCR amplification sequence. First, the oligos were amplified in a standard PCR using Phusion polymerase with 14 amplification cycles. This 400 µL PCR reaction contained: 1% of the pooled oligonucleotides from LCSciences as a template, a forward primer (oHX093) containing a C3 spacer at its 5' end, and a reverse amplification primer (oHX094) containing a biotin-label at its 5' end. Cycling conditions were as follows: denaturation at 95°C for 10 sec; annealing at 60°C for 20 sec; and extension at 72°C for 30 sec. Upon completion of this initial reaction, the entire reaction was used as a template to seed a

larger (40 mL) PCR reaction. For efficient amplification, this large reaction was performed in four 96-well plates with 100μL in each well. Reaction conditions were identical to those described for the first reaction, and a total of 15 cycles were performed for this second amplification. Reactions were purified and concentrated by isopropanol precipitation. To generate single stranded primers for use in MPE-seq, the double-stranded amplicons were first digested using SapI (NEB R0569) in a 150 μL reaction containing 30 μL of enzyme. The reaction was incubated at 37°C overnight, after which the reaction products were concentrated by ethanol precipitation. Next, the 5' to 3' lambda exonuclease (NEB M0262) was used to preferentially degrade the two strands containing unmodified 5' ends. This reaction was performed at 37°C for 2 hours according to the manufacturer's protocol. The products of this reaction were then purified using Zymo columns using 7X volume binding buffer (2 M guanidinium-HCl, 75% isopropanol). After this step, the remaining DNA consisted of the desired single stranded RT primer, and an undesired single stranded section containing the SapI site plus the amplification primer. Making use of the 5' biotin tag on the amplification primer, these undesired oligos were removed by affinity capture with streptavidin beads. Specifically, this was accomplished by using 50 μL of Dynabeads MyOne Streptavidin C1 according to the manufacturer's protocol. The unbound supernatant fraction was retained as it contains the desired products. The recovered material was precipitated and verified using 6% native PAGE stained with SyBr Gold.

1st Strand extension template oligo design. The oligos were designed with 3 key features from the 5' to 3' end of the oligo: 1.) A portion of the Nextera P7XX sequencing adapter. Of the entirety of the P7XX adapter, the region 3' of the I7 barcode was used. This allowed for barcoding and amplification of the sequencing libraries. 2.) a dN9 anchor on the 3' end to

randomly anneal to cDNA products. 3.) A 3' carbon block modification. This was done so that the oligo may only be used as a template to append the Nextera sequencing adapter to the end of cDNAs, rather than as a primer.

MPE-Seq Library Prep

cDNA synthesis. RNA was isolated following a hot acid phenol extraction protocol(Collart and Oliviero 2001).10ug of total RNA was used to generate each library. cDNA was synthesized by mixing 1 µg of the gene specific primer pool described above with each RNA sample in 50 mM Tris-HCl (pH 8.5), 75 mM KCl in a 20 µL volume. The primers were then annealed in a thermocycler with the following cycle; 70°C for 1 minute, 65°C for 5 minutes, hold at 47°C. An equivalent volume of MMLV reverse transcriptase enzyme mix containing 1 mM dATP, 1 mM dGTP, 1 mM dCTP, 0.4 mM aminoallyl-dUTP, 0.6 mM dTTP, 50 mM Tris-HCl (pH 8.5), 150 mM KCl, 6 mM $MgCl_2$, 10 mM DTT was pre-heated to 47°C and added to the primer-annealed RNA mix resulting in a total reaction volume of 40uL. Maintaining the samples at 47°C was essential for reducing off-target cDNA synthesis. Reactions were incubated at 47°C for 3 hours, followed by heat inactivation at 85°C for 5 minutes. Remaining RNA was hydrolyzed by addition of ½ volume of 0.3 M NaOH, 0.03 M EDTA and incubated at 65°C for 15 minutes. After neutralization with ½ (original) volume of 0.3 M HCl, the cDNA was purified with a Zymo-5 column using 7X volume of binding buffer (2 M guanidinium HCl, 75% isopropanol). Purified cDNA samples were dried in a SpeedVac until all liquid had evaporated.

NHS ester biotin coupling. Dried cDNA samples were resuspended in 18 µL of fresh 0.1 M Sodium Bicarbonate (pH 9.0), to which 2 µL of 0.1 mg/µL NHS-biotin (ThermoFisher 20217) was added. Reactions were incubated at 65°C for 1 hour followed by purification of biotin

coupled cDNA from unreacted NHS-biotin by using Zymo-5 columns using 7X volume of binding buffer (2 M guanidinium HCl, 75% isopropanol).

Streptavidin-biotin purification. 20 μL of Dynabeads MyOne Streptavidin C1 (ThermoFisher 65602) per sample were pre-washed twice in 500 μL of 1X bind and wash buffer (5 mM Tris-HCl (pH 7.5), 0.5 mM EDTA and 1 M NaCl) as per manufacturer's protocol. Washed beads were resuspended in 50 μL of 2X bind and wash buffer per sample and 50 μL was combined with each 50 μL purified cDNA sample. Biotin-streptavidin binding was allowed to proceed for 30 minutes at room temperature with rotation. Bound material was washed twice with 500 μL of 1X bind and wash buffer, followed by an additional wash with 100 μL of 1X SSC. To ensure purification of only single-stranded cDNAs, beads were then incubated with 0.1 M NaOH for two consecutive room temperature washes for 10 minutes and 1 minute, respectively. Finally, the bound material was washed 3 times with 100 μL 1X TE. cDNA was eluted from the beads by heating samples to 90°C for 2 minutes in the presence of 100 μL of 95% formamide, 10 mM EDTA. The eluate was then purified using Zymo-5 columns as described above. cDNA was eluted from columns in 40 μL H_20.

First strand extension. Primers were annealed to purified cDNA by combining: 1uL 1^{st} strand extension oligo (100 μM), 5 μL 10x NEB buffer 2, 40 μL purified cDNA sample, and 1 μL of 10 mM (each) dNTP mix. Samples were then incubated at 65°C for 5 minutes, followed by cooling to room temperature on the bench top. To each sample was added 3 μL of Klenow exo- fragment (NEB M0212) and reactions were incubated for 5 minutes at room temperature, after which they were moved to 37°C for 30 minutes. Samples were subsequently purified via streptavidin beads following the protocol described above. Samples were then purified and concentrated via Zymo-5 columns. Samples were then eluted in 33 μL H_20.

PCR amplification. Amplification of the reaction products was accomplished by using 10 µL of the purified material generated in the 1st strand extension reaction as a template in a PCR reaction. Illumina Nextera (i5) and (i7) indexing primers were used in a standard 50 µL PCR reaction with Phusion polymerase (ThermoFisher F530S). Cycling conditions were as follows: denaturation at 95°C for 10 sec; annealing at 62°C for 20 sec; and extension at 72°C for 30 sec. Libraries typically required between 14 and 20 cycles of amplification, depending upon the amount of starting material. Libraries were then purified via PAGE. Each 50 µL PCR reaction was run on a 6% native poly-acrylamide gel and DNA was resolved by staining with SyBr gold. Libraries were size selected from 200bp to 800bp and DNA was extracted from gel fragments via passive diffusion overnight in 0.3 M sodium acetate (pH 5.3). Libraries were then ethanol precipitated and quantified.

cDNA synthesis temperature experiment. Due to the target specific nature of MPE-seq cDNA synthesis, any reverse transcription (RT) events at non-target sites will reduce the fraction of on-target reads. Indeed, these off-target events contribute significantly to the nonspecific class reads in a typical MPE-seq experiment (Fig2A). One way to reduce off-target RT events is through increasing the specificity of the RT primers. We assessed this by testing the effect of increased temperature during the RT reaction on off-target sequencing reads. MPE-seq libraries were generated using the above described protocol with one primary difference: Increased reaction temperatures required the use of a thermostable enzyme. For this reason, Superscript III (ThermoFisher) was used along with the manufacturer supplied buffer (reaction concentration: 50 mM Tris-HCl pH 8.3, 75 mM KCl, 3 mM MgCl2). Primer annealing and reactions were carried out at 47°C, 51°C, and 55°C in replicate.

MPE-Seq Data Anlysis

Sequencing and alignment. MPE-seq libraries were sequenced on the NextSeq platform by the BRC Genomics Facility at Cornell university using 60bp (P5) +15bp (P7) paired-end chemistry. PCR duplicates were removed from the dataset by filtering out non-unique reads with respect to all base calls in both reads, including the 7bp UMI. In other words, for each set of identical paired-end reads, a single read-pair was retained for analysis. MPE-seq Reads were aligned to the yeast genome (Reference genome assembly R64-1-1(Engel *et al.* 2014)) using the STAR aligner(Dobin *et al.* 2013) with the following alignment parameters: {--alignEndsType EndToEnd --alignIntronMin 20 --alignIntronMax 1000 --alignMatesGapMax 400 --alignSplicedMateMapLmin 16 --alignSJDBoverhangMin 1 --outSAMmultNmax 1 --outFilterMismatchNmax 3 --clip3pAdapterSeq CTGTCTCTTATACACATCTCCGAGCCCACGAGAC --clip5pNbases 7 0}. Alignment files were filtered to exclude read mappings deriving from inserts less than 30bases. We believe these short fragments represent unextended reverse transcription primers that were retained in the sequencing libraries. These small fragments can sometimes erroneously map to splice junctions or target introns, even though we believe they are not derived from cellular RNA.

RNA-seq libraries were sequenced on an Illumina HiSeq2500 by the BRC Genomics Facility at Cornell University using 100bp single end reads. Reads were aligned using the STAR aligner with the following alignment parameters: {--alignEndsType EndToEnd --alignIntronMin 20 --alignIntronMax 1000 --alignSJDBoverhangMin 1 --outSAMmultNmax 1 --outFilterMismatchNmax 3 --clip3pAdapterSeq CTGTCTCTTATACACATCTCCGAGCCCACGAGAC}.

When applicable, replicate libraries were combined prior to alignment. However, to assess technical reproducibility of MPE-seq, replicate libraries were subsampled to varying read depths, aligned separately, and compared to RNA-seq libraries also subsampled to varying read depths.

Estimating Splice isoform abundances from MPE-Seq data.

For each intron, the relative abundance of unspliced and spliced isoforms was determined by counting spliced and unspliced reads. Spliced reads (*S*) were counted using the SJ.out.tab file created by the aligner. Unspliced reads were counted using bedtools (Quinlan and Hall 2010) to count the number of reads that cover any part of the intron, considering only the first read of the paired-end reads. Unspliced read counts were further categorized as deriving from a lariat intermediate (*L*) or pre-1st step RNA (*P*) by considering the mapping location of the second read of the paired end reads, which we observed to often terminate near the TSS, or in the case of a lariat-intermediate-derived cDNA, near the branchpoint-A of the intron. Based on paired end mapping locations, each fragment was categorized into one of six categories (See FigS6) and the counts within those six categories were used to calculate *S*, *P*, and *L* as follows:

$$S = C_1 + C_2 \qquad P = C_3\left(1 + \frac{C_5 + C_6}{C_3 + C_4}\right) \qquad L = C_4\left(1 + \frac{C_5 + C_6}{C_3 + C_4}\right)$$

Locations of branchpoints were determined by consolidating the most used branchpoint from lariat sequencing data (Mayerle *et al.* 2017) and previously described branch locations based on sequence motif searches (Grate and Ares 2002).

Heatmaps and meta-gene plots: To generate metagene plots which illustrate read coverage around features of interest, we used the deepTools ComputeMatrix command(Ramírez *et al.* 2016) in conjunction with a BigWig coverage file of the 3' terminating bases and a bedfile containing TSS-positions as determined by PRO-cap (Booth *et al.* 2016) or a bedfile containing the annotated branchpoint regions detailed above. Importantly, this bedfile was filtered to only

include branchpoint regions that would produce a lariat intermediate that is within the size range captured by library size-selection of MPE-seq libraries.

RNA-seq Experiments

Library prep: For each RNA-seq library, 1 µg of total RNA was input into the "NEBNext Ultra Directional RNA Library Prep Kit for Illumina". Libraries were prepared following the manufacturer's protocol.

Estimating Splice isoform abundances from RNA-seq data: Similar to MPE-seq data, spliced reads from target introns were counted using the SJ.out.tab file created by the aligner. Unspliced reads were counted using the bedtools software package (Quinlan and Hall 2010) to count the number of reads which overlapped an intron. Spliced and unspliced read counts for each intron were then length normalized for the feature's potential mapping space. The potential mapping space for a spliced read is equal to 2 x read length minus the minimum splice junction overhang length. The potential mapping space for an unspliced read is equal to the 2 x read length minus minimum splice junction overhang length plus length of the intron. Reads counts assigned to each feature were then divided by the length. Fraction unspliced was calculated for each intron as the quotient of length normalized unspliced reads and spliced reads. Relative transcript expression was calculated via transcripts per million (TPM) normalization (Conesa *et al.* 2016), only considering exonic reads and exonic gene-lengths.

Acknowledgements and Author Contributions

HX designed MPE-seq protocol and prepared most sequencing libraries. MG performed experiments and analysis involved in figure S1. BJF analyzed data and made all other figures. BJF and HX wrote manuscript draft. BJF, HX, JAP, ZD, and MG contributed to draft revisions. Li Yao drafted Fig1A.

Works Cited

Alahari, S. K., H. Schmidt, and N. F. Käufer, 1993 The fission yeast prp4+ gene involved in pre-mRNA splicing codes for a predicted serine/threonine kinase and is essential for growth. Nucleic Acids Res. 21: 4079–4083.

Albulescu, L. O., N. Sabet, M. Gudipati, N. Stepankiw, Z. J. Bergman *et al.*, 2012 A quantitative, high-throughput reverse genetic screen reveals novel connections between pre-mRNA splicing and 5′ and 3′ end transcript determinants. PLoS Genet. 8: e1002530.

Alexander, R. D., S. a. Innocente, J. D. Barrass, and J. D. Beggs, 2010 Splicing-Dependent RNA polymerase pausing in yeast. Mol. Cell 40: 582–593.

Ameur, A., A. Zaghlool, J. Halvardson, A. Wetterbom, U. Gyllensten *et al.*, 2011 Total RNA sequencing reveals nascent transcription and widespread co-transcriptional splicing in the human brain. Nat. Struct. Mol. Biol. 18: 1435–1440.

Andersson, R., S. Enroth, A. Rada-Iglesias, C. Wadelius, and J. Komorowski, 2009 Nucleosomes are well positioned in exons and carry characteristic histone modifications. Genome Res. 19: 1732–41.

Anver, S., A. Roguev, M. Zofall, N. J. Krogan, S. I. S. Grewal *et al.*, 2014 Yeast X-chromosome-associated protein 5 (Xap5) functions with H2A.Z to suppress aberrant transcripts. EMBO Rep. 15: 894–902.

Araya, C. L., and D. M. Fowler, 2011 Deep mutational scanning: Assessing protein function on a massive scale. Trends Biotechnol. 29: 435–442.

Armstrong, J., N. Bone, J. Dodgson, and T. Beck, 2007 The role and aims of the FYSSION project. Briefings Funct. Genomics Proteomics 6: 3–7.

Awan, A. R., A. Manfredo, and J. a Pleiss, 2013 Lariat sequencing in a unicellular yeast

identifies regulated alternative splicing of exons that are evolutionarily conserved with humans. Proc. Natl. Acad. Sci. U. S. A. 110: 12762–7.

Baejen, C., P. Torkler, S. Gressel, K. Essig, J. Söding *et al.*, 2014 Transcriptome maps of mRNP biogenesis factors define pre-mRNA recognition. Mol Cell 55: 745–757.

Bansal, V., 2010 A statistical method for the detection of variants from next-generation resequencing of DNA pools. Bioinformatics 26:.

Barbosa-Morais, N. L., M. Irimia, Q. Pan, H. Y. Xiong, S. Gueroussov *et al.*, 2012 The evolutionary landscape of alternative splicing in vertebrate species. Science (80-.). 338: 1587–1593.

Bayne, E. H., D. a Bijos, S. a White, F. de Lima Alves, J. Rappsilber *et al.*, 2014 A systematic genetic screen identifies new factors influencing centromeric heterochromatin integrity in fission yeast. Genome Biol. 15: 481.

Bayne, E. H., M. Portoso, A. Kagansky, I. C. Kos-Braun, T. Urano *et al.*, 2008 Splicing factors facilitate RNAi-directed silencing in fission yeast. Science 322: 602–606.

Berget, S. M., 1995 Minireviews : Exon Recognition in Vertebrate Splicing. 2411–2414.

Bitton, D. A., S. R. Atkinson, C. Rallis, G. C. Smith, D. A. Ellis *et al.*, 2015 Widespread exon-skipping triggers degradation by nuclear RNA surveillance in fission yeast. Genome Res. 884–896.

Blencowe, B. J., G. Baurén, A. G. Eldridge, R. Issner, J. A. Nickerson *et al.*, 2000 The SRm160/300 splicing coactivator subunits. RNA 6: 111–20.

Blomquist, T. M., E. L. Crawford, J. L. Lovett, J. Yeo, L. M. Stanoszek *et al.*, 2013 Targeted RNA-Sequencing with Competitive Multiplex-PCR Amplicon Libraries. PLoS One 8: e79120.

Boon, K.-L., T. Auchynnikava, G. Edwalds-Gilbert, J. D. Barrass, A. P. Droop *et al.*, 2006 Yeast Ntr1/Spp382 Mediates Prp43 Function in Postspliceosomes. Mol. Cell. Biol.

Booth, G. T., I. X. Wang, V. G. Cheung, and J. T. Lis, 2016 Divergence of a conserved elongation factor and transcription regulation in budding and fission yeast. Genome Res. 26: 799–811.

Bradley, R. K., J. Merkin, N. J. Lambert, and C. B. Burge, 2012 Alternative splicing of RNA triplets is often regulated and accelerates proteome evolution. PLoS Biol. 10: e1001229.

Burke, J., A. Longhurst, D. Merkurjev, J. Sales-Lee, B. Rao *et al.*, 2018 Spliceosome profiling visualizes the operations of a dynamic RNP in vivo at nucleotide resolution. Cell 173: 1014–1030.

Carey, M. F., C. L. Peterson, and S. T. Smale, 2013 The primer extension assay. Cold Spring Harb. Protoc. 8: 164–173.

Carrillo Oesterreich, F., S. Preibisch, and K. M. Neugebauer, 2010 Global analysis of nascent rna reveals transcriptional pausing in terminal exons. Mol. Cell 40: 571–581.

Cazzola, M., M. Rossi, and L. Malcovati, 2013 Biologic and clinical significance of somatic mutations of SF3B1 in myeloid and lymphoid neoplasms. Blood.

Chanfreau, G., P. Legrain, B. Dujon, and A. Jacquier, 1994 Interaction between the first and last nucleotides of pre-mRNA introns is a determinant of 3′ splice site selection in S.cerevisiae. Nucleic Acids Res.

Chen, W., J. Moore, H. Ozadam, H. P. Shulha, N. Rhind *et al.*, 2018 Transcriptome-wide interrogation of the functional intronome by spliceosome profiling. Cell 173: 1031–1044.

Chen, W., H. P. Shulha, A. Ashar-Patel, J. Yan, K. M. Green *et al.*, 2014 Endogenous U2·U5·U6 snRNA complexes in S. pombe are intron lariat spliceosomes. RNA 20: 308–20.

Ciurciu, A., L. Duncalf, V. Jonchere, N. Lansdale, O. Vasieva *et al.*, 2013 PNUTS/PP1 regulates RNAPII-mediated gene expression and is necessary for developmental growth. PLoS Genet. 9: e1003885.

Clark, T. a, C. W. Sugnet, and M. Ares, 2002 Genomewide analysis of mRNA processing in yeast using splicing-specific microarrays. Science 296: 907–910.

Collart, M. A., and S. Oliviero, 2001 Preparation of Yeast RNA, in *Current Protocols in Molecular Biology*,.

Conesa, A., P. Madrigal, S. Tarazona, D. Gomez-Cabrero, A. Cervera *et al.*, 2016 A survey of best practices for RNA-seq data analysis. Genome Biol. 17:.

De Conti, L., M. Baralle, and E. Buratti, 2013 Exon and intron definition in pre-mRNA splicing. Wiley Interdiscip. Rev. RNA 4: 49–60.

Coombes, C. E., and J. D. Boeke, 2005 An evaluation of detection methods for large lariat RNAs. RNA 11: 323–31.

Core, L. J., and J. T. Lis, 2008 Transcription regulation through promoter-proximal pausing of RNA polymerase II. Science 319: 1791–1792.

Crooks, G. E., G. Hon, J. M. Chandonia, and S. E. Brenner, 2004 WebLogo: A sequence logo generator. Genome Res. 14: 1188–1190.

David, C. J., and J. L. Manley, 2011 The RNA polymerase C-terminal domain: A new role in spliceosome assembly. Transcription 2: 221–225.

Depristo, M. A., E. Banks, R. Poplin, K. V. Garimella, J. R. Maguire *et al.*, 2011 A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat. Genet. 43: 491–501.

Dietrich, R. C., R. Incorvaia, and R. A. Padgett, 1997 Terminal intron dinucleotide sequences do

not distinguish between U2- and U12-dependent introns. Mol. Cell.

Dobin, A., C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski *et al.*, 2013 STAR: Ultrafast universal RNA-seq aligner. Bioinformatics 29: 15–21.

Dujardin, G., C. Lafaille, M. de la Mata, L. E. Marasco, M. J. Muñoz *et al.*, 2014 How Slow RNA Polymerase II Elongation Favors Alternative Exon Skipping. Mol. Cell 54: 683–690.

Dunn, E. A., and S. D. Rader, 2014 Preparation of yeast whole cell splicing extract. Methods Mol. Biol.

Egan, E. D., C. R. Braun, S. P. Gygi, and D. Moazed, 2014 Post-transcriptional regulation of meiotic genes by a nuclear RNA silencing complex. RNA 20: 867–81.

Eldridge, A. G., Y. Li, P. A. Sharp, and B. J. Blencowe, 1999 The SRm160/300 splicing coactivator is required for exon-enhancer function. Proc. Natl. Acad. Sci. U. S. A. 96: 6125–6130.

Engel, S. R., F. S. Dietrich, D. G. Fisk, G. Binkley, R. Balakrishnan *et al.*, 2014 The reference genome sequence of Saccharomyces cerevisiae: then and now. G3 (Bethesda). 4: 389–98.

Eser, P., L. Wachutka, K. C. Maier, C. Demel, M. Boroni *et al.*, 2016 Determinants of RNA metabolism in the Schizosaccharomyces pombe genome. Mol. Syst. Biol. 12: 857–857.

Evsyukova, I., S. S. Bradrick, S. G. Gregory, and M. a Garcia-Blanco, 2013 Cleavage and polyadenylation specificity factor 1 (CPSF1) regulates alternative splicing of interleukin 7 receptor (IL7R) exon 6. RNA 19: 103–15.

Fair, B. J., and J. A. Pleiss, 2017 The power of fission: yeast as a tool for understanding complex splicing. Curr. Genet.

Fetzer, S., J. Lauber, C. L. Will, and R. Lührmann, 1997 The [U4/U6.U5] tri-snRNP-specific 27K protein is a novel SR protein that can be phosphorylated by the snRNP-associated

protein kinase. RNA 3: 344–55.

Forsburg, S. L., and N. Rhind, 2006 Basic methods for fission yeast. Yeast 23: 173–183.

Fowler, D. M., and S. Fields, 2014 Deep mutational scanning: a new style of protein science. Nat. Methods 11: 801–7.

Galej, W. P., T. H. D. Nguyen, A. J. Newman, and K. Nagai, 2014 Structural studies of the spliceosome: Zooming into the heart of the machine. Curr. Opin. Struct. Biol.

Galej, W. P., C. Oubridge, A. J. Newman, and K. Nagai, 2013 Crystal structure of Prp8 reveals active site cavity of the spliceosome. Nature 493: 638–43.

Gao, J., F. Kan, J. L. Wagnon, A. J. Storey, R. U. Protacio *et al.*, 2014 Rapid, efficient and precise allele replacement in the fission yeast Schizosaccharomyces pombe. Curr. Genet. 60: 109–119.

Gao, K., A. Masuda, T. Matsuura, and K. Ohno, 2008 Human branch point consensus sequence is yUnAy. Nucleic Acids Res. 36: 2257–2267.

Gottschalk, A., B. Kastner, R. Lührmann, and P. Fabrizio The yeast U5 snRNP coisolated with the U1 snRNP has an unexpected protein composition and includes the splicing factor Aar2p.

Grate, L., and M. Ares, 2002 Searching yeast intron data at Ares lab web site. Methods Enzymol. 350: 380–392.

Graveley, B. R., 2000 Sorting out the complexity of SR protein functions. RNA 6: 1197–1211.

Grewal, S. I. S., 2010 RNAi-dependent formation of heterochromatin and its diverse functions. Curr. Opin. Genet. Dev. 20: 134–141.

Habara, Y., S. Urushiyama, T. Tani, and Y. Ohshima, 1998 The fission yeast prp10(+) gene involved in pre-mRNA splicing encodes a homologue of highly conserved splicing factor,

SAP155. Nucleic Acids Res. 26: 5662–5669.

Haraguchi, N., T. Andoh, D. Frendewey, and T. Tani, 2007 Mutations in the SF1-U2AF59-U2AF23 complex cause exon skipping in Schizosaccharomyces pombe. J. Biol. Chem. 282: 2221–2228.

Hartwell, L. H., C. S. McLaughlin, and J. R. Warner, 1970 Identification of ten genes that control ribosome formation in yeast. MGG Mol. Gen. Genet. 109: 42–56.

Hicks, M. J., B. J. Lam, and K. J. Hertel, 2005 Analyzing mechanisms of alternative pre-mRNA splicing using in vitro splicing assays. Methods.

Hollander, D., S. Naftelberg, G. Lev-Maor, A. R. Kornblihtt, and G. Ast, 2016 How Are Short Exons Flanked by Long Introns Defined and Committed to Splicing? Trends Genet. xx: 1–11.

Hossain, M. A., and T. L. Johnson, 2014a Spliceosomal Pre-mRNA Splicing (K. J. Hertel, Ed.). 1126: 285–298.

Hossain, M. A., and T. L. Johnson, 2014b Using yeast genetics to study splicing mechanisms. Methods Mol. Biol.

Hou, H., Y. Wang, S. P. Kallgren, J. Thompson, J. R. Yates *et al.*, 2010 Histone variant H2A.Z regulates centromere silencing and chromosome segregation in fission yeast. J. Biol. Chem. 285: 1909–1918.

Huang, T., J. Vilardell, and C. C. Query, 2002 Pre-spliceosome formation in S.pombe requires a stable complex of SF1-U2AF59-U2AF23. EMBO J. 21: 5516–5526.

Iannone, C., A. Pohl, P. Papasaikas, D. Soronellas, G. P. Vicent *et al.*, 2015 Relationship between nucleosome positioning and progesterone-induced alternative splicing in breast cancer cells. RNA 21: 360–374.

Inada, M., and J. a Pleiss, 2010 *Genome-wide approaches to monitor pre-mRNA splicing.* Elsevier Inc.

James Kent, W., C. W. Sugnet, T. S. Furey, K. M. Roskin, T. H. Pringle *et al.*, 2002 The human genome browser at UCSC. Genome Res. 12: 996–1006.

Jamieson, D. J., B. Rahe, J. Pringle, and J. D. Beggs, 1991 A suppressor of a yeast splicing mutation (prp8-1) encodes a putative ATP-dependent RNA helicase. Nature.

Juneau, K., C. Nislow, and R. W. Davis, 2009 Alternative splicing of PTC7 in Saccharomyces cerevisiae determines protein localization. Genetics 183: 185–194.

Kallgren, S. P., S. Andrews, X. Tadeo, H. Hou, J. J. Moresco *et al.*, 2014 The Proper Splicing of RNAi Factors Is Critical for Pericentric Heterochromatin Assembly in Fission Yeast. PLoS Genet. 10: 1–11.

Käufer, N. F., and J. Potashkin, 2000 Analysis of the splicing machinery in fission yeast: a comparison with budding yeast and mammals. Nucleic Acids Res. 28: 3003–3010.

Kawashima, T., S. Douglass, J. Gabunilas, M. Pellegrini, and G. F. Chanfreau, 2014 Widespread Use of Non-productive Alternative Splice Sites in Saccharomyces cerevisiae. PLoS Genet. 10:.

Khodor, Y. L., J. Rodriguez, K. C. Abruzzi, C. H. A. Tang, M. T. Marr *et al.*, 2011 Nascent-seq indicates widespread cotranscriptional pre-mRNA splicing in Drosophila. Genes Dev. 25: 2502–2512.

Kielkopf, C. L., N. A. Rodionova, M. R. Green, and S. K. Burley, 2001 A novel peptide recognition mode revealed by the X-ray structure of a core U2AF35/U2AF65 heterodimer. Cell 106: 595–605.

Kim, D.-U., J. Hayles, D. Kim, V. Wood, H.-O. Park *et al.*, 2010 Analysis of a genome-wide set

of gene deletions in the fission yeast Schizosaccharomyces pombe. Nat. Biotechnol. 28: 617–623.

Kim, D., B. Langmead, and S. L. Salzberg, 2015 HISAT: A fast spliced aligner with low memory requirements. Nat. Methods 12: 357–360.

Kim, S. H., and R. J. Lin, 1993 Pre-mRNA splicing within an assembled yeast spliceosome requires an RNA-dependent ATPase and ATP hydrolysis. Proc. Natl. Acad. Sci. U. S. A.

Kim, S. H., and R. J. Lin, 1996 Spliceosome activation by PRP2 ATPase prior to the first transesterification reaction of pre-mRNA splicing. Mol. Cell. Biol. 16: 6810–6819.

Kivioja, T., A. Vähärautio, K. Karlsson, M. Bonke, M. Enge *et al.*, 2012 Counting absolute numbers of molecules using unique molecular identifiers. Nat. Methods 9: 72–74.

Kuhn, A. N., and N. F. Käufer, 2003 Pre-mRNA splicing in Schizosaccharomyces pombe: regulatory role of a kinase conserved from fission yeast to mammals. Curr. Genet. 42: 241–251.

Kupfer, D. M., S. D. Drabenstot, K. L. Buchanan, H. Lai, H. Zhu *et al.*, 2004 Introns and splicing elements of five diverse fungi. Eukaryot. Cell 3: 1088–1100.

Kyburz, A., A. Friedlein, H. Langen, and W. Keller, 2006 Direct Interactions between Subunits of CPSF and the U2 snRNP Contribute to the Coupling of Pre-mRNA 3??? End Processing and Splicing. Mol. Cell 23: 195–205.

de la Mata, M., C. R. Alonso, S. Kadener, J. P. Fededa, M. Blaustein *et al.*, 2003 A Slow RNA Polymerase II Affects Alternative Splicing In Vivo. Mol. Cell 12: 525–532.

Larson, A., B. J. Fair, and J. A. Pleiss, 2016 Interconnections Between RNA-Processing Pathways Revealed by a Sequencing-Based Genetic Screen for Pre-mRNA Splicing Mutants in Fission Yeast. G3 (Bethesda). 6: 1513–23.

Lee, N. N., V. R. Chalamcharla, F. Reyes-Turcu, S. Mehta, M. Zofall *et al.*, 2013 Mtr4-like protein coordinates nuclear RNA processing for heterochromatin assembly and for telomere maintenance. Cell 155: 1–14.

Lee, Y., and D. C. Rio, 2015 Mechanisms and Regulation of Alternative Pre-mRNA Splicing. Annu. Rev. Biochem. 1–33.

Li, B., M. Carey, and J. L. Workman, 2007 The Role of Chromatin during Transcription. Cell 128: 707–719.

Li, H., and R. Durbin, 2009 Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25: 1754–60.

Libri, D., N. Graziani, C. Saguez, and J. Boulay, 2001 Multiple roles for the yeast SUB2/yUAP56 gene in splicing. Genes Dev.

Lim, L. P., and C. B. Burge, 2001 A computational analysis of sequence features involved in recognition of short introns. Proc. Natl. Acad. Sci. U. S. A. 98: 11193–8.

Lin, R. J., A. J. Lustig, and J. Abelson, 1987 Splicing of yeast nuclear pre-mRNA in vitro requires a functional 40S spliceosome and several extrinsic factors. Genes Dev.

Lin, P. C., and R. M. Xu, 2012 Structure and assembly of the SF3a splicing factor complex of U2 snRNP. EMBO J.

Lipp, J. J., M. C. Marvin, K. M. Shokat, and C. Guthrie, 2015 SR protein kinases promote splicing of nonconsensus introns. Nat. Struct. Mol. Biol. 22: 611–617.

Liu, H.-L., and S.-C. Cheng, 2012 The Interaction of Prp2 with a Defined Region of the Intron Is Required for the First Splicing Reaction. Mol. Cell. Biol.

Long, J. C., and J. F. Caceres, 2009 The SR protein family of splicing factors: master regulators of gene expression. Biochem. J. 417: 15–27.

Love, M. I., W. Huber, and S. Anders, 2014 Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15: 550.

Lucks, J. B., S. A. Mortimer, C. Trapnell, S. Luo, S. Aviran *et al.*, 2011 Multiplexed RNA structure characterization with selective 2'-hydroxyl acylation analyzed by primer extension sequencing (SHAPE-Seq). Proc. Natl. Acad. Sci. 108: 11063–11068.

Luco, R. F., M. Allo, I. E. Schor, A. R. Kornblihtt, and T. Misteli, 2011 Epigenetics in alternative pre-mRNA splicing. Cell 144: 16–26.

Lustig, a J., R. J. Lin, and J. Abelson, 1986 The yeast RNA gene products are essential for mRNA splicing in vitro. Cell 47: 953–63.

Lybarger, S., K. Beickman, V. Brown, N. Dembla-Rajpal, K. Morey *et al.*, 1999 Elevated levels of a U4/U6.U5 snRNP-associated protein, Spp381p, rescue a mutant defective in spliceosome maturation. Mol. Cell. Biol.

Mamanova, L., A. J. Coffey, C. E. Scott, I. Kozarewa, E. H. Turner *et al.*, 2010 Target-enrichment strategies for next- generation sequencing. Nat. Methods 7: 111–118.

Marshall, A. N., M. C. Montealegre, C. Jim??nez-L??pez, M. C. Lorenz, and A. van Hoof, 2013 Alternative Splicing and Subfunctionalization Generates Functional Diversity in Fungal Proteomes. PLoS Genet. 9:.

Matera, A. G., and Z. Wang, 2014 A day in the life of the spliceosome. Nat. Rev. Mol. Cell Biol.

Matlin, A. J., F. Clark, and C. W. J. Smith, 2005 Understanding alternative splicing: towards a cellular code. Nat. Rev. Mol. Cell Biol. 6: 386–98.

Mayas, R. M., H. Maita, and J. P. Staley, 2006 Exon ligation is proofread by the DExD/H-box ATPase Prp22p. Nat. Struct. Mol. Biol.

Mayerle, M., M. Raghavan, S. Ledoux, A. Price, N. Stepankiw *et al.*, 2017 Structural toggle in

the RNaseH domain of Prp8 helps balance splicing fidelity and catalytic efficiency. Proc. Natl. Acad. Sci. 114: 4739–4744.

Mazroui, R., a Puoti, and a Krämer, 1999 Splicing factor SF1 from Drosophila and Caenorhabditis: presence of an N-terminal RS domain and requirement for viability. RNA 5: 1615–31.

McLaren, W., L. Gil, S. E. Hunt, H. S. Riat, G. R. S. Ritchie *et al.*, 2016 The Ensembl Variant Effect Predictor. Genome Biol. 17:.

Mercer, T. R., M. B. Clark, J. Crawford, M. E. Brunck, D. J. Gerhardt *et al.*, 2014 Targeted sequencing for gene discovery and quantification using RNA CaptureSeq. Nat. Protoc. 9: 989–1009.

Mercer, T. R., D. J. Gerhardt, M. E. Dinger, J. Crawford, C. Trapnell *et al.*, 2011 Targeted RNA sequencing reveals the deep complexity of the human transcriptome. Nat. Biotechnol. 30: 99–104.

Merkin, J., C. Russell, P. Chen, and C. B. Burge, 2012 Evolutionary dynamics of gene and isoform regulation in Mammalian tissues. Science 338: 1593–9.

Meyer, M., and J. Vilardell, 2009 The quest for a message: budding yeast, a model organism to study the control of pre-mRNA splicing. Brief. Funct. Genomic. Proteomic. 8: 60–7.

Mikheyeva, I. V, P. J. R. Grady, F. B. Tamburini, D. R. Lorenz, and H. P. Cam, 2014 Multifaceted genome control by Set1 Dependent and Independent of H3K4 methylation and the Set1C/COMPASS complex. PLoS Genet. 10: e1004740.

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

Millhouse, S., and J. L. Manley, 2005 The C-terminal domain of RNA polymerase II functions as a phosphorylation-dependent splicing activator in a heterologous protein. Mol. Cell. Biol. 25: 533–544.

Misteli, T., and D. L. Spector, 1999 RNA polymerase II targets pre-mRNA splicing factors to transcription sites in vivo. Mol. Cell 3: 697–705.

Mitchell, P., and D. Tollervey, 2003 An NMD pathway in yeast involving accelerated deadenylation and exosome-mediated 3??????5??? degradation. Mol. Cell 11: 1405–1413.

Morris, D. P., and a. L. Greenleaf, 2000 The splicing factor, Prp40, binds the phosphorylated carboxyl-terminal domain of RNA Polymerase II. J. Biol. Chem. 275: 39935–39943.

Morrison, A. J., and X. Shen, 2009 Chromatin remodelling beyond transcription: the INO80 and SWR1 complexes. Nat. Rev. Mol. Cell Biol. 10: 373–384.

Muñoz, M. J., M. de la Mata, and A. R. Kornblihtt, 2010 The carboxy terminal domain of RNA polymerase II and alternative splicing. Trends Biochem. Sci. 35: 497–504.

Nakazawa, N., S. Harashima, and Y. Oshima, 1991 AAR2, a gene for splicing pre-mRNA of the MATa1 cistron in cell type control of Saccharomyces cerevisiae. Mol. Cell. Biol. 11: 5693–5700.

Network, T. C. G. A., 2012 Comprehensive molecular portraits of human breast tumors. Nature.

Nguyen, T. H. D., W. P. Galej, X. Bai, C. G. Savva, A. J. Newman *et al.*, 2015 The architecture of the spliceosomal U4/U6.U5 tri-snRNP. Nature 523: 47–52.

Nguyen, T. H. D., W. P. Galej, S. M. Fica, P. C. Lin, A. J. Newman *et al.*, 2016 CryoEM structures of two spliceosomal complexes: Starter and dessert at the spliceosome feast. Curr. Opin. Struct. Biol. 36: 48–57.

Noble, S. M., and C. Guthrie, 1996 Identification of novel genes required for yeast pre-mRNA

splicing by means of cold-sensitive mutations. Genetics 143: 67–80.

Nojima, T., T. Gomes, A. R. F. Grosso, H. Kimura, M. J. Dye *et al.*, 2015 Mammalian NET-Seq Reveals Genome-wide Nascent Transcription Coupled to RNA Processing. Cell 161: 526–540.

Noma, T., K. Sode, and K. Ikebukuro, 2006 Characterization and application of aptamers for Taq DNA polymerase selected using an evolution-mimicking algorithm. Biotechnol. Lett. 28: 1939–1944.

Ohi, M. D., A. J. Link, L. Ren, J. L. Jennings, W. H. McDonald *et al.*, 2002 Proteomics analysis reveals stable multiprotein complexes in both fission and budding yeasts containing Myb-related Cdc5p/Cef1p, novel pre-mRNA splicing factors, and snRNAs. Mol. Cell. Biol. 22: 2011–2024.

Padgett, R. A., M. M. Konarska, M. Aebi, H. Hornig, C. Weissmann *et al.*, 1985 Nonconsensus branch-site sequences in the in vitro splicing of transcripts of mutant rabbit beta-globin genes. Proc. Natl. Acad. Sci. U. S. A. 82: 8349–8353.

Pandit, S., Y. Zhou, L. Shiue, G. Coutinho-Mansfield, H. Li *et al.*, 2013 Genome-wide Analysis Reveals SR Protein Cooperation and Competition in Regulated Splicing. Mol. Cell 50: 223–235.

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

Parker, R., and P. G. Siliciano, 1993 Evidence for an essential non-Watson-Crick interaction between the first and last nucleotides of a nuclear pre-mRNA intron. Nature.

Patrick, K. L., C. J. Ryan, J. Xu, J. J. Lipp, K. E. Nissen *et al.*, 2015 Genetic Interaction Mapping Reveals a Role for the SWI/SNF Nucleosome Remodeler in Spliceosome Activation in

Fission Yeast. PLoS Genet. 11: e1005074.

Picelli, S., A. K. Björklund, B. Reinius, S. Sagasser, G. Winberg *et al.*, 2014 Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. Genome Res. 24: 2033–2040.

Potashkin, J., R. Li, and D. Frendewey, 1989 Pre-mRNA splicing mutants of Schizosaccharomyces pombe. EMBO J. 8: 551–9.

Qin, D., L. Huang, A. Wlodaver, J. Andrade, and J. P. Staley, 2016 Sequencing of lariat termini in S. cerevisiae reveals 5' splice sites, branch points, and novel splicing events. RNA.

Quesada, V., A. J. Ramsay, and C. Lopez-Otin, 2012 Chronic lymphocytic leukemia with SF3B1 mutation. N. Engl. J. Med.

Quinlan, A. R., and I. M. Hall, 2010 BEDTools: A flexible suite of utilities for comparing genomic features. Bioinformatics 26: 841–842.

Ram, O., and G. Ast, 2007a SR proteins: a foot on the exon before the transition from intron to exon definition. Trends Genet. 23: 5–7.

Ram, O., and G. Ast, 2007b SR proteins: a foot on the exon before the transition from intron to exon definition. Trends Genet. 23: 5–7.

Ramírez, F., D. P. Ryan, B. Grüning, V. Bhardwaj, F. Kilpert *et al.*, 2016 deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res. 44: W160–W165.

Rauhut, R., P. Fabrizio, O. Dybkov, K. Hartmuth, V. Pena *et al.*, 2016 Molecular architecture of the Saccharomyces cerevisiae activated spliceosome. Science (80-.).

Ren, L., J. R. McLean, T. R. Hazbun, S. Fields, C. Vander Kooi *et al.*, 2011 Systematic Two-Hybrid and Comparative Proteomic Analyses Reveal Novel Yeast Pre-mRNA Splicing

Factors Connected to Prp19. PLoS One 6: e16719.

Rhind, N., Z. Chen, M. Yassour, D. a Thompson, B. J. Haas *et al.*, 2011 Comparative functional genomics of the fission yeasts. Science 332: 930–6.

Robberson, B. L., G. J. Cote, and S. M. Berget, 1990 Exon definition may facilitate splice site selection in RNAs with multiple exons. Mol. Cell. Biol. 10: 84–94.

Romfo, C. M., C. J. Alvarez, W. J. van Heeckeren, C. J. Webb, and J. a Wise, 2000 Evidence for splice site pairing via intron definition in Schizosaccharomyces pombe. Mol. Cell. Biol. 20: 7955–7970.

Rosado-Lugo, J. D., and M. Hampsey, 2014 The Ssu72 phosphatase mediates the RNA polymerase II initiation-elongation transition. J. Biol. Chem. 289: 33916–26.

Rosenberg, G. H., S. K. Alahari, and N. F. Käufer, 1991 prp4 from Schizosaccharomyces pombe, a mutant deficient in pre-mRNA splicing isolated using genes containing artificial introns. MGG Mol. Gen. Genet. 226: 305–309.

Roshbash, M., P. K. Harris, J. L. Woolford, and J. L. Teem, 1981 The effect of temperature-sensitive RNA mutants on the transcription products from cloned ribosomal protein genes of yeast. Cell 24: 679–686.

Rouskin, S., M. Zubradt, S. Washietl, M. Kellis, and J. S. Weissman, 2014 Genome-wide probing of RNA structure reveals active unfolding of mRNA structures in vivo. Nature 505: 701–5.

Ryan, C. J., A. Roguev, K. Patrick, J. Xu, H. Jahari *et al.*, 2012 Hierarchical Modularity and the Evolution of Genetic Interactomes across Species. Mol. Cell 46: 691–704.

Sasaki-Haraguchi, N., T. Ikuyama, S. Yoshii, T. Takeuchi-Andoh, D. Frendewey *et al.*, 2015 Cwf16p associating with the nineteen complex ensures ordered exon joining in constitutive

Pre-mRNA splicing in fission yeast. PLoS One 10: 1–16.

Scheckel, C., and R. B. Darnell, 2015 Microexons—Tiny but mighty. EMBO J. 34: 273–274.

Schellenberg, M. J., T. Wu, D. B. Ritchie, S. Fica, J. P. Staley *et al.*, 2013 A conformational switch in PRP8 mediates metal ion coordination that promotes pre-mRNA exon ligation. Nat. Struct. Mol. Biol.

Schones, D. E., K. Cui, S. Cuddapah, T.-Y. Roh, A. Barski *et al.*, 2008 Dynamic regulation of nucleosome positioning in the human genome. Cell 132: 887–898.

Schwer, B., 2008 A Conformational Rearrangement in the Spliceosome Sets the Stage for Prp22-Dependent mRNA Release. Mol. Cell.

Schwer, B., and C. H. Gross, 1998 Prp22, a DExH-box RNA helicase, plays two distinct roles in yeast pre-mRNA splicing. EMBO J.

Schwer, B., and S. Shuman, 2015 Structure-function analysis and genetic interactions of the Yhc1, SmD3, SmB, and Snp1 subunits of yeast U1 snRNP and genetic interactions of SmD3 with U2 snRNP subunit Lea1. RNA 21: 1173–86.

Semlow, D. R., and J. P. Staley, 2012 Staying on message: Ensuring fidelity in pre-mRNA splicing. Trends Biochem. Sci.

Shao, C., B. Yang, T. Wu, J. Huang, P. Tang *et al.*, 2014 Mechanisms for U2AF to define 3' splice sites and regulate alternative splicing in the human genome. Nat. Struct. Mol. Biol. 21: 997–1005.

Sims, R. J., S. Millhouse, C. F. Chen, B. A. Lewis, H. Erdjument-Bromage *et al.*, 2007 Recognition of Trimethylated Histone H3 Lysine 4 Facilitates the Recruitment of Transcription Postinitiation Factors and Pre-mRNA Splicing. Mol. Cell 28: 665–676.

Spies, N., C. B. Nielsen, R. A. Padgett, and C. B. Burge, 2009 Biased Chromatin Signatures

around Polyadenylation Sites and Exons. Mol. Cell 36: 245–254.

Stepankiw, N., M. Raghavan, E. A. Fogarty, A. Grimson, and J. a. Pleiss, 2015 Widespread alternative and aberrant splicing revealed by lariat sequencing. Nucleic Acids Res. 43: 8488–501.

Storey, J. D., and R. Tibshirani, 2003 Statistical significance for genomewide studies. Proc. Natl. Acad. Sci. 100: 9440–9445.

Szymczyna, B. R., J. Bowman, S. McCracken, A. Pineda-Lucena, Y. Lu *et al.*, 2003 Structure and function of the PWI motif: a novel nucleic acid-binding domain that facilitates pre-mRNA processing. Genes Dev. 17: 461–475.

Taggart, A. J., A. M. DeSimone, J. S. Shih, M. E. Filloux, and W. G. Fairbrother, 2012 Large-scale mapping of branchpoints in human pre-mRNA transcripts in vivo. Nat. Struct. Mol. Biol. 19: 719–721.

The Gene Ontology Consortium, 2000 Gene Ontology: tool for the unification of biology. Nat. Genet. 25: 25–29.

The Gene Ontology Consortium, 2015 Gene Ontology Consortium: going forward. Nucleic Acids Res. 43: D1049–D1056.

Tilgner, H., C. Nikolaou, S. Althammer, M. Sammeth, M. Beato *et al.*, 2009 Nucleosome positioning as a determinant of exon recognition. Nat. Struct. Mol. Biol. 16: 996–1001.

Tsai, R. T., R. H. Fu, F. L. Yeh, C. K. Tseng, Y. C. Lin *et al.*, 2005a Spliceosome disassembly catalyzed by Prp43 and its associated components Ntr1 and Ntr2. Genes Dev. 19: 2991–3003.

Tsai, R. T., R. H. Fu, F. L. Yeh, C. K. Tseng, Y. C. Lin *et al.*, 2005b Spliceosome disassembly catalyzed by Prp43 and its associated components Ntr1 and Ntr2. Genes Dev.

Tseng, C. K., H. L. Liu, and S. C. Cheng, 2011 DEAH-box ATPase Prp16 has dual roles in remodeling of the spliceosome in catalytic steps. RNA.

Turunen, J. J., E. H. Niemelä, B. Verma, and M. J. Frilander, 2013 The significant other: Splicing by the minor spliceosome. Wiley Interdiscip. Rev. RNA.

Urushiyama, S., T. Tani, and Y. Ohshima, 1996 Isolation of novel pre-mRNA splicing mutants of Schizosaccharomyces pombe. Mol. Gen. Genet. 253: 118–127.

Vanoosthuyse, V., P. Legros, S. J. A. van der Sar, G. Yvert, K. Toda *et al.*, 2014 CPF-Associated Phosphatase Activity Opposes Condensin-Mediated Chromosome Condensation. PLoS Genet. 10: e1004415.

Verdel, A., S. Jia, S. Gerber, T. Sugiyama, S. Gygi *et al.*, 2004 RNAi-mediated targeting of heterochromatin by the RITS complex. Science 303: 672–676.

Vijayraghavan, U., M. Company, and J. Abelson, 1989 Isolation and characterization of pre-mRNA splicing mutants of Saccharomyces cerevisiae. Genes Dev. 3: 1206–1216.

Villa, T., and C. Guthrie, 2005 The Isy1p component of the NineTeen Complex interacts with the ATPase Prp16p to regulate the fidelity of pre-mRNA splicing. Genes Dev.

Vincent, M., P. Lauriault, M. F. Dubois, S. Lavoie, O. Bensaude *et al.*, 1996 The nuclear matrix protein p255 is a highly phosphorylated form of RNA polymerase II largest subunit which associates with spliceosomes. Nucleic Acids Res. 24: 4649–4652.

Vo, T. V., J. Das, M. J. Meyer, N. A. Cordero, N. Akturk *et al.*, 2016 A Proteome-wide Fission Yeast Interactome Reveals Network Evolution Principles from Yeasts to Human. Cell 164: 310–323.

Wahl, M. C., C. L. Will, and R. Lührmann, 2009 The Spliceosome: Design Principles of a Dynamic RNP Machine. Cell 136: 701–718.

Washington, K., T. Ammosova, M. Beullens, M. Jerebtsova, A. Kumar *et al.*, 2002 Protein phosphatase-1 dephosphorylates the C-terminal domain of RNA polymerase-II. J. Biol. Chem. 277: 40442–8.

Webb, C. J., C. M. Romfo, W. J. van Heeckeren, and J. A. Wise, 2005a Exonic splicing enhancers in fission yeast: functional conservation demonstrates an early evolutionary origin. Genes Dev. 19: 242–254.

Webb, C. J., C. M. Romfo, W. J. van Heeckeren, and J. A. Wise, 2005b Exonic splicing enhancers in fission yeast: functional conservation demonstrates an early evolutionary origin. Genes Dev. 19: 242–54.

Webb, C. J., and J. A. Wise, 2004 The Splicing Factor U2AF Small Subunit Is Functionally Conserved between Fission Yeast and Humans. Mol. Cell. Biol. 24: 4229–4240.

Wernersson, R., and H. B. Nielsen, 2005 OligoWiz 2.0 - Integrating sequence feature annotation into the design of microarray probes. Nucleic Acids Res. 33:.

Wilhelm, B., S. Marguerat, S. Aligianni, S. Codlin, S. Watt *et al.*, 2011 Differential patterns of intronic and exonic DNA regions with respect to RNA polymerase II occupancy, nucleosome density and H3K36me3 marking in fission yeast. Genome Biol 12: R82.

Wilhelm, B. T., S. Marguerat, S. Watt, F. Schubert, V. Wood *et al.*, 2008 Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution. Nature 453: 1239–43.

Wilkinson, M. E., S. M. Fica, W. P. Galej, C. M. Norman, A. J. Newman *et al.*, 2017 Postcatalytic spliceosome structure reveals mechanism of 3 ’-splice site selection. Science (80-.).

Will, C. L., and R. Lührmann, 2011a Spliceosome structure and function. Cold Spring Harb. Perspect. Biol. 3:.

Will, C. L., and R. Lührmann, 2011b Spliceosome structure and function. Cold Spring Harb. Perspect. Biol. 3: 1–2.

Xu, H., B. J. Fair, Z. Dwyer, M. Gildea, and J. A. Pleiss, 2018 Multiplexed Primer Extension Sequencing Enables High Precision Detection of Rare Splice Isoforms. BioRxiv.

Yan, C., J. Hang, R. Wan, M. Huang, C. C. L. Wong *et al.*, 2015 Structure of a yeast spliceosome at 3.6-angstrom resolution. Science 349: 1182–91.

Yan, C., R. Wan, R. Bai, G. Huang, and Y. Shi, 2016 Structure of a yeast activated spliceosome at 3.5 Å resolution. Science (80-.).

Yeh, T.-C., H.-L. Liu, C.-S. Chung, N.-Y. Wu, Y.-C. Liu *et al.*, 2011 Splicing Factor Cwc22 Is Required for the Function of Prp2 and for the Spliceosome To Escape from a Futile Pathway. Mol. Cell. Biol.

Zheng, W., L. M. Chung, and H. Zhao, 2011 Bias detection and correction in RNA-Sequencing data. BMC Bioinformatics 12:.

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

Zhu, Y. Y., E. M. Machleder, A. Chenchik, R. Li, and P. D. Siebert, 2001 Reverse transcriptase template switching: A SMART^{TM} approach for full-length cDNA library construction. Biotechniques 30: 892–897.

Zlatanova, J., and A. Thakar, 2008 H2A.Z: View from the Top. Structure 16: 166–179.

Zofall, M., T. Fischer, K. Zhang, M. Zhou, B. Cui *et al.*, 2009 Histone H2A.Z cooperates with RNAi and heterochromatin factors to suppress antisense RNAs. Nature 461: 419–422.

Zuo, P., and T. Maniatis, 1996 The splicing factor U2AF35 mediates critical protein-protein

interactions in constitutive and enhancer-dependent splicing. Genes Dev. 10: 1356–1368.