INVESTIGATING THE CONNECTIONS BETWEEN PRE-mRNA SPLICING, CHROMATIN REMODELING AND TRANSCRIPTION, AND 3'END PROCESSING

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

Laura-Oana Albulescu

May 2013

INVESTIGATING THE CONNECTIONS BETWEEN PRE-mRNA SPLICING, CHROMATIN REMODELING AND TRANSCRIPTION, AND 3'END PROCESSING

Laura-Oana Albulescu, Ph. D

Cornell University 2013

Pre-mRNA splicing is an essential eukaryotic pathway which controls gene expression. Increasing lines of evidence indicate links between splicing and other RNA processing pathways such as chromatin remodeling, transcription and 3’end processing, yet in many cases the specific proteins responsible for functionally connecting these pathways remain unclear.

To determine the full complement of factors which impact pre-mRNA splicing, I developed a genome-wide screen in *Saccharomyces cerevisiae* which allowed me to evaluate differences in splicing efficiency in the background of ~5500 unique gene mutations. By measuring expression changes in precursor levels by high-throughput quantitative PCR, I detected enrichment in several classes of genes, with very strong candidates mapping to the chromatin remodeling, transcription and 3’end processing classes. One of these candidates is the bromodomain protein Bdf1, a component of the transcription factor TFIID and also a member of the SWR-C chromatin remodeling complex. Splicing sensitive microarrays confirm that deletion of Bdf1 leads to a global splicing defect, while ChIP-qPCR data reveal a decrease in U1 snRNP recruitment at intron containing genes, suggesting an inhibitory effect on spliceosome assembly.

Conversely, Bdf1's homologue Bdf2 with which it is 46% identical, does not impact pre-mRNA splicing or spliceosome recruitment, consistent with my hypothesis that Bdf2 functions mainly in transcription.

To further characterize Bdf1 function, I modified the high-throughput screening approach described above and employed it in a forward genetic manner to enable a mutagenic analysis of the Bdf1 protein. This analysis revealed that the C-terminal tail which overlaps with the Taf7 interaction domain, and contains a conserved SEED region and one of the known phosphorylation sites in Bdf1, may be responsible for the splicing defect.

In opposition to the global splicing defect exhibited by Bdf1, mutations in 3'end processing factors such as Cft2 and Yth1 result in transcript-specific defects. My results highlight the cross-talk between 5' and 3'end processing factors and the spliceosome, and support a model in which the definition of terminal exons in the budding yeast is identical with the mechanism described in higher systems. Furthermore, the novel role of Bdf1 at the interface of transcription and pre-mRNA splicing suggests a new mechanism that underlies the coupling between these two RNA pathways.

BIOGRAPHICAL SKETCH

Laura was born on March 27th 1984 as the daughter of Olivia and Ovidiu Bud and grew up in Bucharest, Romania. From an early age she attended 2 schools - the primary school and the Music school where she learned how to play the piano. This arrangement taught her organization and how to best schedule her time to achieve multiple goals. Starting in the 7th Grade, she rounded off her musical education by singing in the Choir of the Romanian Radio House for a year. Laura then attended the "Mihai Viteazul" high school in the Romanian capital, in the Mathematics-Informatics section. It was here that her interest in science first manifested itself as a love for Chemistry. She participated in numerous Chemistry contests (Olympiads) which is where she met most of her closest friends to date, as well as in a mixed contest of Geography, Chemistry, Biology and Physics called "Earth Sciences" where she won an Honorable Mention. She then continued to integrate her interests in Chemistry and Biology, by attending the Biochemistry section at the Faculty of Biology, University of Bucharest. It was during this time that she started working in the Molecular Biology laboratory at the university as well as in two other Research institutes (one of them the Research Institute of the Romanian Academy) which allowed her to develop multiple skills and experiment with a breadth of techniques. This motivated her to more closely investigate gene expression and regulation and to pursue a PhD in the laboratory of Jeff Pleiss at Cornell University. At Cornell she developed new strategies to assess splicing changes in a high throughput manner and became immersed in RNA biology. Working in close association with her mentor, Laura also learned that in a new

laboratory a good part of the time is spent optimizing techniques before applying them to a biological question and that patience and an optimist outlook are very important in science. Beyond teaching biochemistry laboratories and giving conference talks, Laura got married in 2010 to Andrei, with whom she shares a passion for science, traveling, reading, music, history and games. Outside the lab, Laura can be found enjoying nature walks, traveling and experiencing ancient cultures, crafting or reading new fantasy novels.

To my husband, Andrei, who has always supported me in whatever I chose to try

To my parents who encouraged me to follow my dreams

To my friends who showed me that true friendship can withstand distance

And to my grandmother Maria who always strived for excellence

ACKNOWLEDGMENTS

I want to thank my advisor, Jeff Pleiss for guiding me in my graduate work and being very supportive, especially during times when my experiments were slow in yielding results. Jeff was always the optimistic person, who tried to give us (Ali, Mohan and I), his first graduate students, a lot of opportunities by taking us to conferences early on and exposing us to the scientific world. He encouraged me to present my data in front of large audiences and get feedback even if it wasn't publishable yet, and gave me and everyone in the lab a real "course" in what it means to give a good presentation. Jeff was always enthusiastic and ready to teach me new techniques – even if it meant programming the robot, coming in at 11 at night to show me how to load the samples onto a microarray, or working side by side on a weekend in the early stages when we were developing the screen. He always let me set my own schedule (which was very "standard" 9-5 and almost no weekends) without any complaints and understood my passion for science, as well as the one for traveling and seeing the world. I wish to thank Tom Fox and John Lis, my committee members for helpful discussions and for their guidance when I was rotating in their labs during the first year. I will always remember John having a sparkle in his eye at one of our committee meetings and finally asking very excitedly: "So, how do you think this works?" I wish to thank Andrew Grimson and the members of his lab for feedback and suggestions during our joint group meetings, and for his encouragement while navigating the frustrations of the publishing cycle. I am grateful to Susan Ely, for whom I TAed the biochemistry laboratory for being such an organized person and a great lecturer to work with. I

thank Nevin, for working with me to set up and optimize the screen, and for orienting me around the lab during my first days. I also with to thank Ali for helping me with the ChIP-Seq analysis and writing short custom programs to process the data, as well as giving me suggestions as to why the line of code I was trying to use didn't work as expected. And lastly, I wish to thank everyone in the Pleiss lab for making it such a friendly and warm environment to work in, as well as for all their feedback throughout the years. My work was funded by NIGMS grant GM098634 to JA Pleiss.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION: PRE-mRNA SPLICING AND ITS CONNECTIONS TO OTHER RNA PROCESSING PATHWAYS

1.1 Introduction

Pre-mRNA splicing is a major step in the processing of eukaryotic RNA precursor molecules which removes non-coding intervening sequences (introns) from the mature transcript and ligates the coding fragments (exons) together. The process consists of two subsequent transesterification reactions and is catalyzed by a massive macromolecular complex known as the spliceosome which includes 5 small nuclear RNAs (snRNAs) –U1, U2, U4, U5, U6 - and over 100 proteins. Whereas the snRNAs are thought to make up the catalytic core of the spliceosome [1], the proteins can be stably associated with the U1-U5 molecules forming small nuclear ribonucleoprotein (snRNP) particles, they can interact with the substrate or mediate the assembly of other spliceosomal factors, or can function as molecular motors which guide the rearrangements which take place at different steps in the splicing process (helicases, ATPases).

The splicing of a single precursor RNA molecule is a highly regulated process and requires the ordered assembly of several complexes at the level of the substrate. The substrate itself displays several conserved sequences which are specifically recognized such as the 5' splice site, the branchpoint, the 3'splice site and, in the case of higher systems the polypyrimidine tract (see Figure 1.1A and [2] for review) . Moreover, a

breadth of auxiliary SR proteins is known to act in higher systems to guide substrate recognition through interaction with splicing silencers and enhancers and influence alternative splicing decisions.

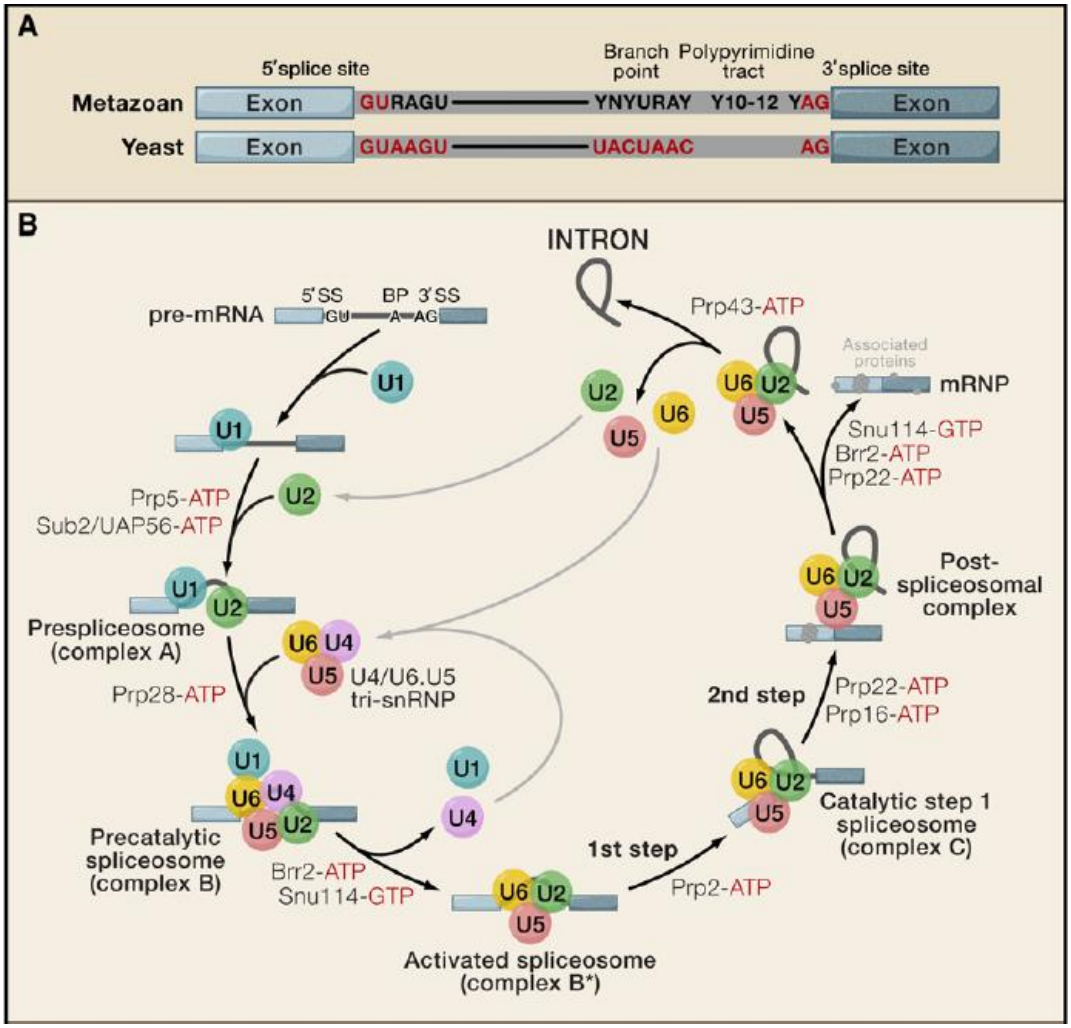

Figure 1.1: Intronic recognition sequences and the splicing cycle **[2]**

In the splicing cycle, the U1snRNP first base pairs with the 5' splice site and the branch point recognition protein (BBP in yeast or SF1 in higher systems) binds the adenosine residue which will later participate in the first transesterification reaction. In addition, in higher systems the U2 auxiliary factor (U2AF) which consists of two subunits assembles at level of the polypyrimidine tract (through U2AF65), and at the

3’ splice site (though U2AF35). This leads to the formation of the early complex or complex E which does not require ATP hydrolysis. Next, the U2 snRNP recognizes the branch point adenosine in an ATP-dependent fashion, displacing SF1/BBP and leads to the formation of the A complex. The U4/U5/U6 tri-snRNP is then recruited to the substrate, forming complex B, but this complex is not yet catalytically active. For this to occur, several rearrangements must take place which will allow for the formation of a base-pairing interaction between U2 and U6, thus creating the catalytic core of the spliceosome, and lead to the exclusion of U1 and U4 snRNPs from the complex (Figure 1.1B). The activated spliceosome, also known as complex B*, can now catalyze the first transesterification reaction in which the branch point adenosine attacks the phophodiester bond at the 5’ splice site. This leads to the formation of a free 5’exon and a second lariat-3’exon product and represents the spliceosomal complex C. Finally, the second transesterification reaction occurs when the 3'-hydroxyl on the free exon attacks the phosphodiester bond at the 3’ splice site and leads to the ligation of the two exons and the release of a circular lariat. Following this step, the spliceosome is disassembled and its subunits are recycled.

Historically, pre-mRNA splicing was studied in isolation from other RNA processing pathways. Nevertheless, a growing body of evidence has made it clear that cellular RNA processing pathways are very tightly interconnected. In the following sections, I will focus on several lines of evidence which link pre-mRNA splicing to the processing steps at the 5’end of the gene such as chromatin remodeling and transcription, as well as to the ones which occur at the 3’end: cleavage and polyadenylation.

1.2 Splicing and chromatin

Several recent studies have started to shed light on a connection between pre-mRNA splicing, nucleosome localization and chromatin marks [3-10]. Thus, the analysis of genome-wide nucleosome positioning and chromatin modification datasets from mammals, mouse, worms and flies has identified a strong correlation between nucleosome occupancy and internal exons, which is independent of their expression levels and is not a consequence of a higher GC content associated with coding regions. Nucleosomes appear to mark true exons, since they are depleted from pseudoexons in spite of their strong splice sites. In addition, there is an inverse correlation between the strength of the splice sites flanking the exons and nucleosome occupancy, which also correlates with exon inclusion [5]. This led to the hypothesis that nucleosomes are important for alternative splicing decisions and that they act as "speed bumps" since they have been shown to slow down RNA polymerase II elongation [11,12] and increase pausing. A slowed polymerase would allow for a "window of opportunity" in which the spliceosome can recognize suboptimal splice sites around constitutive exons and include them in the final spliced product.

Moreover, exons, but not introns, were found to be enriched in specific chromatin marks such as H3K36me3, K3K27me1, K3K27me2 or K3K27me3, H4K20me1, H3K79me1, H2BK5me1 [3]. Among these chromatin marks, H3K36me3 levels correlate with exon inclusion, are dependent on transcription and are enriched in the 3' portions of the genes [10]. One of the most interesting studies which showed a direct mechanism through which the H3K36me3 mark can influence alternative splicing decisions is the Luco *et al*. study [13]. In this study, the FGFR2 reporter gene was

monitored, in which exons IIIb and IIIc are alternatively spliced. H3K36me3 levels are increased in hMSC cells in which exon IIIb is excluded and decreased in PNT2 cells in which this exon is included. Interestingly, overexpression of the H3K36me3 methyltransferase Set2 [14,15] drives exon IIIb repression in PNT2 cells.. The exclusion of exon IIIb was shown to be dependent on MRG15, an adaptor protein which recognizes the H3K36me3 mark and physically interacts with PTB, the polypyrimidine tract binding protein which represses exon IIIb (Figure 1.2). This is the first example of an adapter system which can guide alternative splicing outcomes based on direct recognition of chromatin marks.

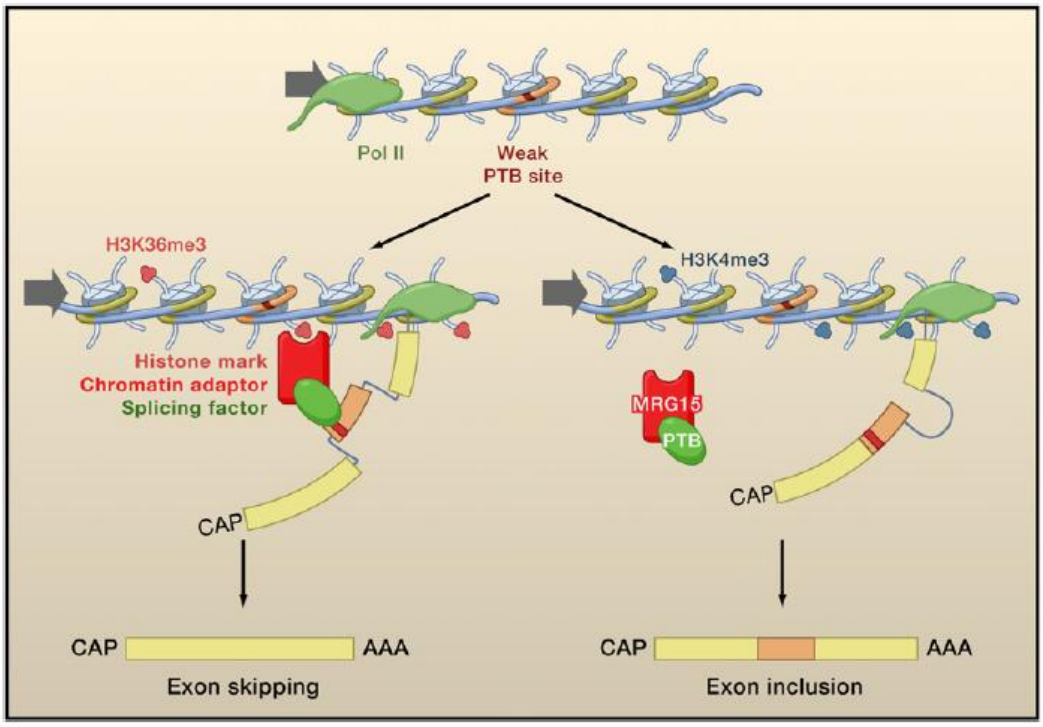

Figure 1.2: Modulation of splicing by H3K36me3 **[16]**

Further evidence which supports H3K36me3's link with splicing comes from a separate study [17] which showed that mutating the 3'SS in the second intron of a β-

globin reporter gene leads to a 3' shift in the H3K36me3 distribution on intron-containing genes. In addition, inhibiting splicing globally with spliceostatin A, (which blocks assembly after U2 binding) leads to the redistribution of the H3K36me3 mark in the same manner. Moreover, the Almeida *et al.* study [18] showed that the connection between the H3K36me3 chromatin mark deposition and splicing is bidirectional. Thus, inhibition of splicing with spliceostatin A inhibits SET2 recruitment and decreases H3K36me3 levels. Therefore, the H3K36me3 mark whose deposition depends on the phosphorylation status of RNA polymerase II [15,19], active elongation and which marks exons for splicing, is in return controlled by the spliceosome.

Besides the H3K36me3 mark, the H3K4me3 chromatin signature, which marks transcriptional starts sites, was shown to interact with CHD1 and change alternative splicing outcomes *in vivo* by regulating the interaction of U2snRNP with the substrate [20]. Moreover, interactions between U2snRNP and STAGA in human cells [21], a dependence on the Gcn5 HAT activity for the recruitment of U2snRNP components Lea1 and Msl1 [22], as well as synthetic lethal interactions between H2BK123ub1 and Lea1 and Msl1 in yeast [23] have been recently reported.

Lastly, several reports suggest that the chromatin to splicing connection can be often mediated by SR proteins [24 review]. SR proteins were shown to interact with the phosphorylated C-terminal domain (CTD) of RNA polymerase II [25, 26] and couple splicing to transcription [27]. Thus, the HP1 protein which recognizes the H3K9me3 mark was reported to coimmunoprecipitate with ASF/SF2 and hnRNP splicing repressors [28], while H3S10P regulates the interaction of SRp20 and ASF/SF2 with

mitotic chromosomes [29].

In conclusion, several avenues connect splicing to chromatin marks and new ways of regulation and communication between the two are currently under investigation.

1.3 Splicing and transcription

The view of RNA polymerase II as a transcriptional toolbox which interacts with capping, splicing, and polyadenylation factors has become the accepted model in the RNA processing world (Figure1.3). The CTD of the polymerase interacts with a host of factors based on its different phosphorylation statuses and coordinates the transitions between initiation, active elongation and splicing, and termination and 3'processing.

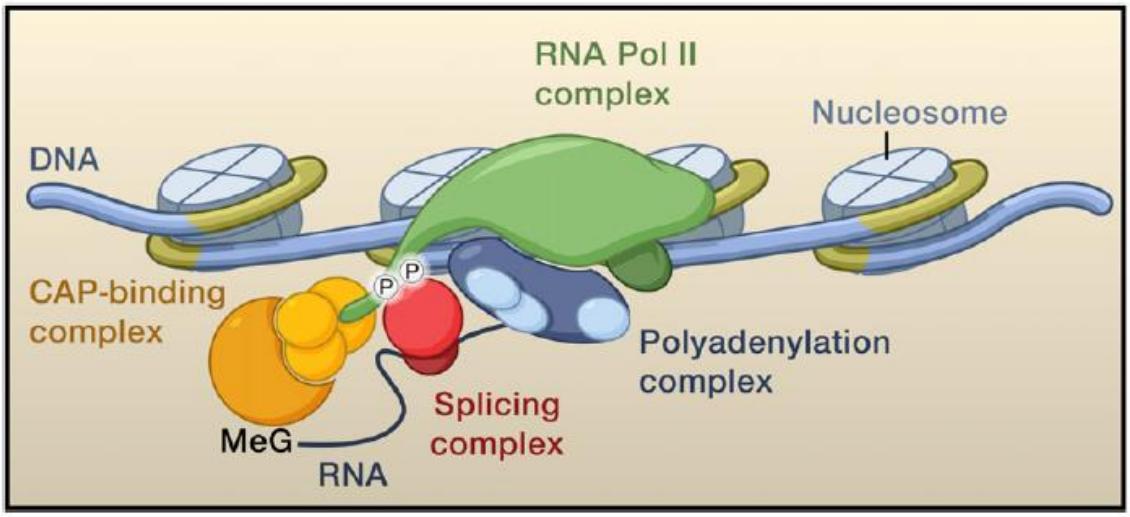

Figure 1.3: RNA polymerase links the RNA processing machineries **[16]**

Early EM studies [30] first observed co-transcriptional splicing on nascent transcripts in Drosophila. Today, the accepted view is that splicing reactions are initiated co-transcriptionally [31] and that a significant proportion of them (80% in higher systems

[32]) are finalized before the polymerase dissociates from the transcript [33,34]. RNA polymerase II was observed to pause in terminal exons ~200 bp after the 3'splice site and ~150 bp before the polyA signal in yeast and allow for a 30-60 second temporal window in which splicing could be completed [33]. Curiously, the genes which were fully spliced co-transcriptionally had shorter terminal exons and longer introns on average than the ones that were not. Moreover, a precise kinetic study [35] showed that the pause at the level of the 3' splice site is a cyclic process which coincides with the recruitment of splicing factors (U2, U5 snRNPs) and the first detection of spliced mRNA. RNA polymerase II pausing is splicing dependent and correlates with the Ser5 phosphorylation mark on the CTD, whereas when the pause is released the phophorylation mark switches to Ser2.

The positive feedback between splicing and transcription is supported by several lines of evidence, despite early *in vitro* studies that have shown that splicing and transcription can occur independently. Thus, gene expression is stimulated in the presence of an intron, and a 5' splice site can independently stimulate transcription [36]. Moreover, the recruitment of U1 is dependent on transcription [37], whereas the recognition of a 5'splice by U1snRNP promotes productive reinitiation [38].

1.3.1 The CTD of RNA polymerase II and pre-mRNA splicing

The coupling between transcription and pre-mRNA splicing involves the CTD of RNA polymerase II [39] in higher systems, since a deletion of the last 5 heptads in the CTD caused an inhibition of splicing and polyadenylation. It was further shown that the CTD phosporylation status was important for splicing stimulation [40, 41], while

the required domain was narrowed down to the last C-terminal 10 amino acids [42, 43]. Lastly, a recent study revealed that the coupling consists in the interaction of U2AF65 with the phosphorylated CTD, followed by the recruitment of the Prp19 complex through a physical interaction with U2AF65 [44]. Despite this however, it was recently shown that fusing the CTD onto a bacterial T7 polymerase or on RNA polymerase III does not enhance capping and splicing *in vivo* [45]. This could be due to promoter and 3'end processing differences between RNA polymerase II and these other polymerases, but might also suggest that the CTD is not sufficient for coupling transcription to splicing.

In opposition to the evidence in higher systems, the yeast CTD was shown not to be required for the splicing of the RP51A mini-gene [46]. Even if this piece of evidence cannot be generalized, it still remains to be seen if such a connection can be established in yeast. So far, the main candidate for coupling splicing to transcription is Prp40 [47] which is a stable U1snRNP component that interacts with the phosphorylated CTD through its WW and FF domains. However, a deletion of the WW domains does not cause lethality and was recently shown [48] not to impact U1 and U2 recruitment *in vivo*, but rather delay a later step of U5 recruitment. It remains to be seen if the yeast mechanism resembles the one in higher systems and if Prp40 is involved in this coupling.

1.3.2 Capping and pre-mRNA splicing

In addition to the coupling of splicing and transcription through the CTD of RNA polymerase II, numerous early studies have linked capping to splicing stimulation [49-

57]. Thus, the presence of a 7-methyl GpppG cap on an RNA molecule promoted splicing *in vivo* and *in vitro*. Furthermore, GpppG but not ApppG cap analogs were able to compete with capped molecules in *in vitro* assays. Interestingly, the stimulatory effect on splicing was observed only for the first intron within a substrate and was negligible for subsequent introns [49, 54, 57].

The splicing of a radiolabeled RNA precursor was also inhibited upon depletion of the cap-binding proteins CBC20 or CBC80 [50] in HeLa nuclear extracts, which correlated with a decrease in the rate of complex A formation. It was subsequently shown that the association of U1 with the 5'SS of the proximal intron is inefficient in extracts in which the cap binding complex has been depleted [57] and that both CBP proteins within the complex are required for efficient splicing stimulation. Support for this coupling was also found in yeast where Mud13, equivalent of mammalian CBP20, [58] exhibits synthetic lethality with a mutant allele of U1. Moreover, in a *mud13Δ* commitment complex formation is impaired. Lastly, additional interactions between nuclear cap binding proteins and hnRNPs [59] were described and shown to stimulate pre-mRNA splicing.

1.3.3 Pre-mRNA splicing and transcription elongation: the kinetic model

The RNA polymerase II elongation rate has been linked to alterative splicing choices. Thus, an increase in polymerase II elongation leads to increased skipping of exons with weaker splice sites, while a decrease in its rate is associated with higher alternative exon inclusion [60-64] see Figure 1.4)

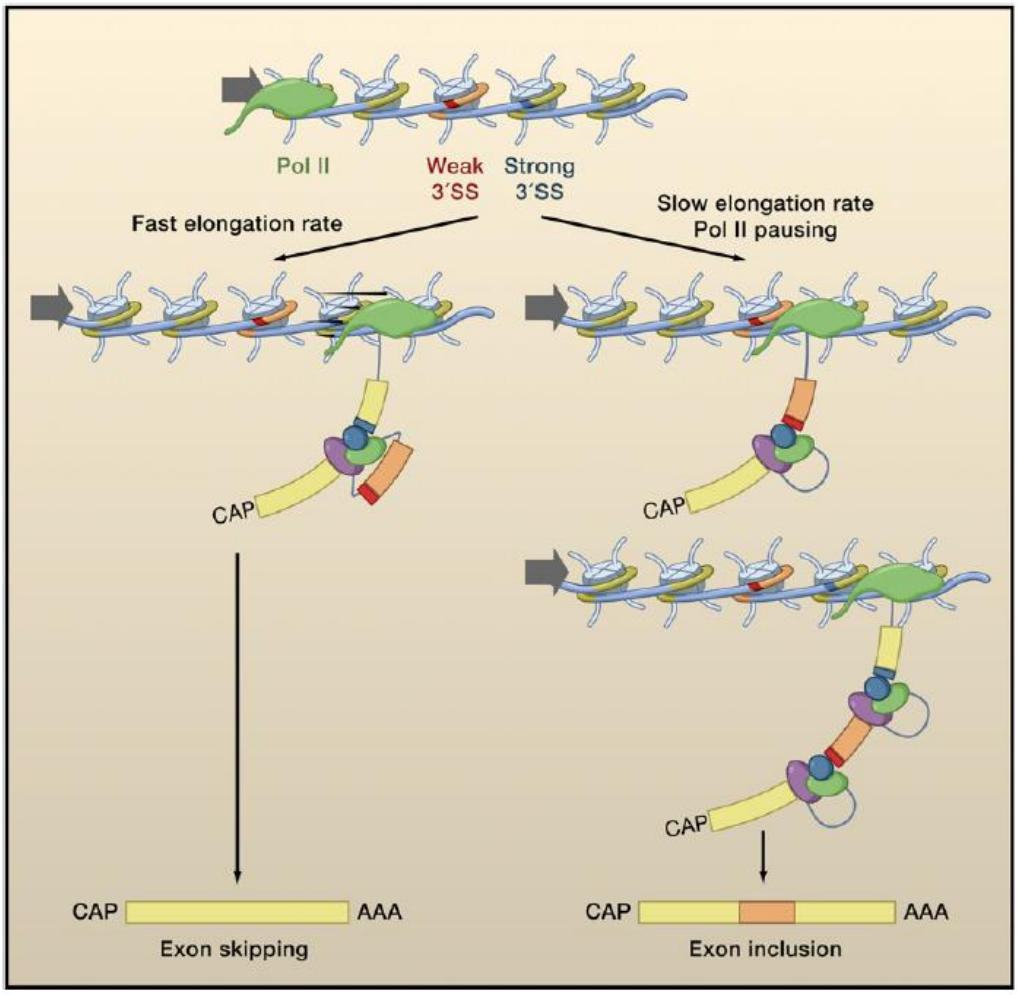

Figure 1.4: The kinetic model: RNA polymerase II modulates alternative splicing

[16]

One of the first reports [60] engineered a slow polymerase and observed an increased inclusion of the EDI exon of the fibronectin gene. Moreover, in the case of the E1a adenovirus substrate in which three alternative 5'splice sites can be coupled to a single 3'splice site, the slow polymerase caused the upstream 5'splice sites to be used more often. The reasoning behind this is that the transcription of the downstream preferred splice sites is delayed, and thus the spliceosome has time to recognize and act at the level of the suboptimal upstream splice sites. Similarly, a slow yeast RNA polymerase

(*rpb2-10*), as well as treatment with drugs which inhibit transcription elongation can suppress exon skipping *in vivo* and promote exon inclusion [65]. Furthermore, whereas inhibitors of transcription elongation increase exon inclusion, treatment with histone deacetylase inhibitors such as trichostatin A increase RNA polymerase II processivity and cause the opposite effect [62, 66]. Conversely, engineering of pause sites which delay the synthesis of repressive silencing elements in either the tropomyosin [67] or the FGFR2 [68] gene leads to the inclusion of the exon which is normally repressed.

Finally, in addition to the elongation rate, splicing choices can also be dictated by the promoter [69], since an exchange of the α-1 globin promoter with a CMV promoter led to EDI inclusion. This effect was not correlated with the strength of the promoter and was also seen in with a mutant fibronectin promoter, suggesting that elements within the promoter can play a role in splicing.

1.4 Splicing and 3' end processing

Given the toolbox model of RNA polymerase II, splicing and polyadenylation factors can interact and are brought together by the C-terminal domain of the polymerase. The CTD region of RNA polymerase II which is required for efficient cleavage and polyadenylation has been mapped [39, 42, 70] and it was shown [71] that the cleavage and polyadenylation specificity factor (CPSF) is recruited by TFIID and then transferred to the elongating polymerase.

Connections between splicing and the cleavage and polyadenylation machinery are supported at multiple levels. Thus, antibodies against Sm and some U1 antisera inhibit

polyadenylation *in vivo*, while a mutation in the 3' splice site of the β-globin intervening sequence IVS2 causes transcriptional read-through past the polyA site [72]. Moreover, early studies have shown that in coupled splicing and polyadenylation reactions mutating the polyA AAUAAA site to AAGAAA decreases the splicing of the terminal intron 5-10 fold [73], whereas mutating the 3' splice site of the terminal intron leads to a 2-3 fold decrease in polyadenylation [74].

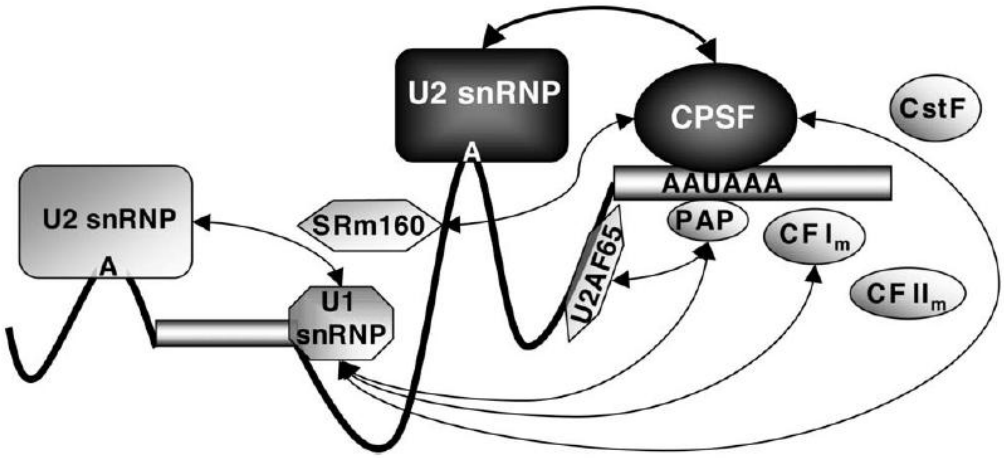

Figure 1.5: Interactions between splicing of the terminal exon and the cleavage and polyadenylation machinery **[75]**

Several physical interactions (Figure1.5) between splicing factors and the cleavage and polyadenlyation have been described. Thus, The RS domain of U2AF65 physically interacts with CFmI59 and promotes cleavage and polyadenylation [76], whereas physical interactions between the C-terminal end of polyA polymerase (PAP) and U2AF [77] stimulate U2AF binding to the upstream 3' splice site. Furthermore, direct interactions between U2 snRNP (SF3a, SF3b) subunits and CPSF have been established [75], and shown to potentiate positive feedback between the two reactions. Moreover, the N-terminus of the U1snRNP protein U1A interacts with the 160kDA

subunit of CPSF and leads to increased polyadenylation *in vitro* [78], while the 25kDa CFIm subunit binds to U1 70K [79]. Ultimately, the Keller group has recently shown that an allele of the cleavage and polyadenylation factor Ysh1, *ysh1-12*, can specifically impact splicing [80], thus reinforcing the link between the two RNA processing pathways.

1.5 The intron versus exon definition model

Recognition of intronic and exonic sequences by the spliceosome follows the rule of assembly over the shortest distance. Thus, in higher systems where introns are large (from 1 to hundreds of kilobases) and the exons small (~150bp), the spliceosome recognizes the substrate by what is known as "exon definition", in which the spliceosomal machinery assembles over the exon. In contrast, in budding yeast, the introns are small (with a median of ~150bp) whereas the exons are larger (with a median of ~450bp). In this case, the assembly of the spliceosome is thought to occur over the intron by following the "intron definition" model.

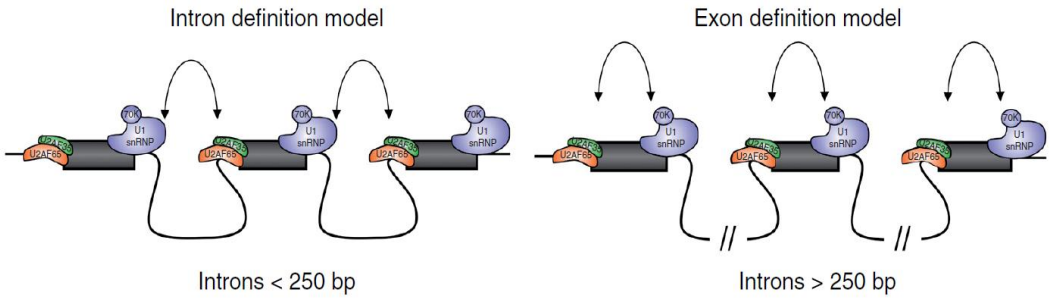

Figure 1.6: The intron versus exon definition models **[32]**

Several lines of evidence point to the different types of mechanisms that are used. Thus, in higher systems RNA substrates which contain complete exons only with a

preceding 3’ splice site and a downstream 5’ splice site can readily be integrated in spliceosomal A complexes *in vitro* [81]. Interestingly, the length of these units is critical, since exons longer than 300 bp are not recognized and this leads to the activation and usage of cryptic 3’splice sites *in vivo*. In addition, several studies have shown that short exons (<50bp) which make up about 4% of the human genome are usually skipped [82, 83], unless additional sequences such as strong polypyrimidine tracts are present upstream. Moreover, in HeLa cells, expanding the size of an exon from 18 to 109 bp leads to its constitutive inclusion, whereas the shortening of an exon to less than 50 bp leads to its skipping. Given that in higher systems the spliceosome assembles across an exon, the hypothesis is that recruitment of U1snRNP at the 5’ splice site of a downstream intron will recruit U2snRNP across the exon at the upstream 3’ splice site. This was shown experimentally, in that mutating the 5’ splice site in a multi-intron substrate, led to the inhibition of splicing of the upstream intron (for review [84]) and conversely, the strengthening of a weak 5’ splice site, increased the splicing of an upstream intron. Perhaps one of the "rules" of exon versus intron definition is the fact that mutation of a splice site in an exon that is governed by exon definition will lead to exon skipping, whereas the same mutation in an exon which follows the intron definition model will lead to intron retention.

One case of particular interest regards the definition of terminal exons. The 5’ exon contains a 5’ splice site but no consensus sequence upstream that could aid in its definition. However, the 5’ end of the transcript is capped and it was shown that the presence of the cap enhances splicing (see section 1.3.2). Conversely, terminal exons which are larger than internal ones (~600bp), contain a 3’ splice site and a downstream

polyadenylation signal. As mentioned before, there is a bidirectional link between the splicing of the last intron and cleavage and polyadenylation, where a mutation in the polyadenylation signal inhibits the splicing of the upstream intron and the mutation in the 3' splice site of this intron inhibits downstream cleavage and processing (see section 1.4).

Interestingly, this model of terminal exon definition also applies in lower eukaryotes, even though internal introns are recognized by the intron definition model. In the body of this thesis, I will present a genome-wide screen which allowed me to map global cellular factors which impact pre-mRNA splicing. Aside from broadening our understanding of how splicing is connected to other RNA processing pathways within the cell, several of these mutants which impact pre-mRNA splicing reinforce this model of terminal exon definition and belong either to 5' chromatin remodeling and transcription complexes, or the 3' cleavage and polyadenylation machinery.

REFERENCES

1. Valadkhan, S., & Manley, J. L. (2001). Splicing-related catalysis by protein-free snRNAs. *Nature*, *413*(6857), 701-707. doi: 10.1038/35099500

2. Wahl, M. C., Will, C. L., & Luhrmann, R. (2009). The spliceosome: Design principles of a dynamic RNP machine. *Cell*, *136*(4), 701-718. doi: 10.1016/j.cell.2009.02.009

3. Andersson, R., Enroth, S., Rada-Iglesias, A., Wadelius, C., & Komorowski, J. (2009). Nucleosomes are well positioned in exons and carry characteristic histone modifications. *Genome Research*, *19*(10), 1732–1741. doi:10.1101/gr.092353.109

4. Kolasinska-Zwierz, P., Down, T., Latorre, I., Liu, T., Liu, X. S., & Ahringer, J. (2009). Differential chromatin marking of introns and expressed exons by H3K36me3. *Nature Genetics*, *41*(3), 376–381. doi:10.1038/ng.322

5. Tilgner, H., Nikolaou, C., Althammer, S., Sammeth, M., Beato, M., Valcárcel, J., & Guigó, R. (2009). Nucleosome positioning as a determinant of exon recognition. *Nature Structural & Molecular Biology*, *16*(9), 996–1001. doi:10.1038/nsmb.1658

6. Schwartz, S., Meshorer, E., & Ast, G. (2009). Chromatin organization marks exon-intron structure. *Nature Structural & Molecular Biology*, *16*(9), 990–995. doi:10.1038/nsmb.1659

7. Spies, N., Nielsen, C. B., Padgett, R. A., & Burge, C. B. (2009). Biased Chromatin Signatures around Polyadenylation Sites and Exons. *Molecular Cell*, *36*(2), 245–254. doi:10.1016/j.molcel.2009.10.008

8. Nahkuri, S., Taft, R. J., & Mattick, J. S. (2009). Nucleosomes are preferentially positioned at exons in somatic and sperm cells. *Cell Cycle*, *8*(20), 3420–3424.

9. Hon, G., Wang, W., & Ren, B. (2009). Discovery and Annotation of Functional Chromatin Signatures in the Human Genome. (E. Segal, Ed.)*PLoS Computational Biology*, *5*(11), e1000566. doi:10.1371/journal.pcbi.1000566

10. Huff, J. T., Plocik, A. M., Guthrie, C., & Yamamoto, K. R. (2010). Reciprocal intronic and exonic histone modification regions in humans. *Nature Structural & Molecular Biology*, *17*(12), 1495–1499. doi:10.1038/nsmb.1924

11. Hodges, C., Bintu, L., Lubkowska, L., Kashlev, M., & Bustamante, C. (2009). Nucleosomal Fluctuations Govern the Transcription Dynamics of RNA Polymerase II. *Science*, *325*(5940), 626–628. doi:10.1126/science.1172926

12. Subtil-Rodríguez, A., & Reyes, J. C. (2010). BRG1 helps RNA polymerase II to overcome a nucleosomal barrier during elongation, in vivo. *EMBO reports*, *11*(10), 751–757. doi:10.1038/embor.2010.131

13. Luco, R. F., Pan, Q., Tominaga, K., Blencowe, B. J., Pereira-Smith, O. M., & Misteli, T. (2010). Regulation of Alternative Splicing by Histone Modifications. *Science*, *327*(5968), 996–1000. doi:10.1126/science.1184208

14. Fuchs, S. M., Kizer, K. O., Braberg, H., Krogan, N. J., & Strahl, B. D. (2011). RNA Polymerase II Carboxyl-terminal Domain Phosphorylation Regulates Protein Stability of the Set2 Methyltransferase and Histone H3 Di- and Trimethylation at Lysine 36. *Journal of Biological Chemistry*, *287*(5), 3249–3256. doi:10.1074/jbc.M111.273953

15. Krogan, N. J., Kim, M., Tong, A., Golshani, A., Cagney, G., Canadien, V., … Greenblatt, J. (2003). Methylation of Histone H3 by Set2 in Saccharomyces cerevisiae Is Linked to Transcriptional Elongation by RNA Polymerase II. *Molecular and Cellular Biology*, *23*(12), 4207–4218. doi:10.1128/MCB.23.12.4207-4218.2003

16. Luco, Reini F., Allo, M., Schor, I. E., Kornblihtt, A. R., & Misteli, T. (2011). Epigenetics in Alternative Pre-mRNA Splicing. *Cell*, *144*(1), 16–26. doi:10.1016/j.cell.2010.11.056

17. Kim, S., Kim, H., Fong, N., Erickson, B., & Bentley, D. L. (2011). Pre-mRNA splicing is a determinant of histone H3K36 methylation. *Proceedings of the National Academy of Sciences*, *108*(33), 13564–13569. doi:10.1073/pnas.1109475108

18. de Almeida, S. F., Grosso, A. R., Koch, F., Fenouil, R., Carvalho, S., Andrade, J., . . . Carmo-Fonseca, M. (2011). Splicing enhances recruitment of methyltransferase HYPB/Setd2 and methylation of histone H3 Lys36. *Nature Structural & Molecular Biology*, *18*(9), 977-983. doi: 10.1038/nsmb.2123; 10.1038/nsmb.2123

19. Li, J. (2002). Association of the Histone Methyltransferase Set2 with RNA Polymerase II Plays a Role in Transcription Elongation. *Journal of Biological Chemistry*, *277*(51), 49383–49388. doi:10.1074/jbc.M209294200

20. Sims, R. J., Millhouse, S., Chen, C.-F., Lewis, B. A., Erdjument-Bromage, H., Tempst, P., … Reinberg, D. (2007). Recognition of Trimethylated Histone H3 Lysine 4 Facilitates the Recruitment of Transcription Postinitiation Factors and Pre-mRNA Splicing. *Molecular Cell*, *28*(4), 665–676. doi:10.1016/j.molcel.2007.11.010

21. Martinez, E., Palhan, V. B., Tjernberg, A., Lymar, E. S., Gamper, A. M., Kundu, T. K., … Roeder, R. G. (2001). Human STAGA Complex Is a Chromatin-Acetylating Transcription Coactivator That Interacts with Pre-mRNA Splicing and DNA Damage-Binding Factors In Vivo. *Molecular and Cellular Biology*, *21*(20), 6782–6795. doi:10.1128/MCB.21.20.6782-6795.2001

22. Gunderson, Felizza Q., & Johnson, T. L. (2009). Acetylation by the Transcriptional Coactivator Gcn5 Plays a Novel Role in Co-Transcriptional Spliceosome Assembly. (H. D. Madhani, Ed.)*PLoS Genetics*, *5*(10), e1000682. doi:10.1371/journal.pgen.1000682

23. Shieh, G. S., Pan, C. H., Wu, J. H., Sun, Y. J., Wang, C. C., Hsiao, W. C., . . . Kao, C. F. (2011). H2B ubiquitylation is part of chromatin architecture that marks exon-intron structure in budding yeast. *BMC Genomics*, *12*, 627-2164-12-627. doi: 10.1186/1471-2164-12-627; 10.1186/1471-2164-12-627

24. Carrillo Oesterreich, F., Bieberstein, N., & Neugebauer, K. M. (2011). Pause locally, splice globally. *Trends in Cell Biology*, *21*(6), 328-335. doi: 10.1016/j.tcb.2011.03.002; 10.1016/j.tcb.2011.03.002

25. Kim, E., Du, L., Bregman, D. B., & Warren, S. L. (1997). Splicing factors associate with hyperphosphorylated RNA polymerase II in the absence of pre-mRNA. *The Journal of Cell Biology*, *136*(1), 19-28.

26. Du, L., & Warren, S. L. (1997). A functional interaction between the carboxy-terminal domain of RNA polymerase II and pre-mRNA splicing. *The Journal of cell biology*, *136*(1), 5–18.

27. Das, R., Yu, J., Zhang, Z., Gygi, M. P., Krainer, A. R., Gygi, S. P., & Reed, R. (2007). SR Proteins Function in Coupling RNAP II Transcription to Pre-mRNA Splicing. *Molecular Cell*, *26*(6), 867–881. doi:10.1016/j.molcel.2007.05.036

28. Saint-André, V., Batsché, E., Rachez, C., & Muchardt, C. (2011). Histone H3 lysine 9 trimethylation and HP1γ favor inclusion of alternative exons. *Nature Structural & Molecular Biology*, *18*(3), 337–344. doi:10.1038/nsmb.1995

29. Loomis, R. J., Naoe, Y., Parker, J. B., Savic, V., Bozovsky, M. R., Macfarlan, T., … Chakravarti, D. (2009). Chromatin Binding of SRp20 and ASF/SF2 and Dissociation from Mitotic Chromosomes Is Modulated by Histone H3 Serine 10 Phosphorylation. *Molecular Cell*, *33*(4), 450–461. doi:10.1016/j.molcel.2009.02.003

30. Beyer, A. L., & Osheim, Y. N. (1988). Splice site selection, rate of splicing, and alternative splicing on nascent transcripts. *Genes & Development*, *2*(6), 754–765. doi:10.1101/gad.2.6.754

31. Tardiff, D. F., Lacadie, S. A., & Rosbash, M. (2006). A genome-wide analysis indicates that yeast pre-mRNA splicing is predominantly posttranscriptional. *Molecular Cell, 24*(6), 917-929. doi: 10.1016/j.molcel.2006.12.002

32. De Conti, L., Baralle, M., & Buratti, E. (2013). Exon and intron definition in pre-mRNA splicing. *Wiley Interdisciplinary Reviews: RNA*, *4*(1), 49–60. doi:10.1002/wrna.1140

33. Carrillo Oesterreich, F., Preibisch, S., & Neugebauer, K. M. (2010). Global analysis of nascent RNA reveals transcriptional pausing in terminal exons. *Molecular Cell, 40*(4), 571-581. doi: 10.1016/j.molcel.2010.11.004; 10.1016/j.molcel.2010.11.004

34. Singh, J., & Padgett, R. A. (2009). Rates of in situ transcription and splicing in large human genes. *Nature Structural & Molecular Biology*, *16*(11), 1128–1133. doi:10.1038/nsmb.1666

35. Alexander, R. D., Innocente, S. A., Barrass, J. D., & Beggs, J. D. (2010). Splicing-Dependent RNA Polymerase Pausing in Yeast. *Molecular Cell*, *40*(4), 582–593. doi:10.1016/j.molcel.2010.11.005

36. Damgaard, C. K., Kahns, S., Lykke-Andersen, S., Nielsen, A. L., Jensen, T. H., & Kjems, J. (2008). A 5′ Splice Site Enhances the Recruitment of Basal Transcription Initiation Factors In Vivo. *Molecular Cell*, *29*(2), 271–278. doi:10.1016/j.molcel.2007.11.035

37. Kotovic, K. M., Lockshon, D., Boric, L., & Neugebauer, K. M. (2003). Cotranscriptional Recruitment of the U1 snRNP to Intron-Containing Genes in Yeast. *Molecular and Cellular Biology*, *23*(16), 5768–5779. doi:10.1128/MCB.23.16.5768-5779.2003

38. Kwek, K. Y., Murphy, S., Furger, A., Thomas, B., O'Gorman, W., Kimura, H., . . . Akoulitchev, A. (2002). U1 snRNA associates with TFIIH and regulates transcriptional initiation. *Nature Structural Biology, 9*(11), 800-805. doi: 10.1038/nsb862

39. McCracken, S., Fong, N., Yankulov, K., Ballantyne, S., Pan, G., Greenblatt, J., Bentley, D. L. (1997). The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. *Nature, 385*(6614), 357-361. doi: 10.1038/385357a0

40. Hirose, Yutaka, Tacke, R., & Manley, J. L. (1999). Phosphorylated RNA polymerase II stimulates pre-mRNA splicing. *Genes & development*, *13*(10), 1234–1239.

41. Bird, G., Zorio, D. A., & Bentley, D. L. (2004). RNA polymerase II carboxy-terminal domain phosphorylation is required for cotranscriptional pre-mRNA splicing and 3'-end formation. *Molecular and Cellular Biology, 24*(20), 8963-8969. doi: 10.1128/MCB.24.20.8963-8969.2004

42. Fong, N. (2001). Capping, splicing, and 3' processing are independently stimulated by RNA polymerase II: different functions for different segments of the CTD. *Genes & Development*, *15*(14), 1783–1795. doi:10.1101/gad.889101

43. Fong, Nova, Bird, G., Vigneron, M., & Bentley, D. L. (2003). A 10 residue motif at the C-terminus of the RNA pol II CTD is required for transcription, splicing and 3′ end processing. *The EMBO journal*, *22*(16), 4274–4282.

44. David, C. J., Boyne, A. R., Millhouse, S. R., & Manley, J. L. (2011). The RNA polymerase II C-terminal domain promotes splicing activation through recruitment of a U2AF65-Prp19 complex. *Genes & Development*, *25*(9), 972-983. doi: 10.1101/gad.2038011; 10.1101/gad.2038011

45. Natalizio, B. J., Robson-Dixon, N. D., & Garcia-Blanco, M. A. (2008). The Carboxyl-terminal Domain of RNA Polymerase II Is Not Sufficient to Enhance the Efficiency of Pre-mRNA Capping or Splicing in the Context of a Different Polymerase. *Journal of Biological Chemistry*, *284*(13), 8692–8702. doi:10.1074/jbc.M806919200

46. Licatalosi, D. D., Geiger, G., Minet, M., Schroeder, S., Cilli, K., McNeil, J. B., & Bentley, D. L. (2002). Functional interaction of yeast pre-mRNA 3′ end processing factors with RNA polymerase II. *Molecular cell*, *9*(5), 1101–1111.

47. Morris, D. P. (2000). The Splicing Factor, Prp40, Binds the Phosphorylated Carboxyl-terminal Domain of RNA Polymerase II. *Journal of Biological Chemistry*, *275*(51), 39935–39943. doi:10.1074/jbc.M004118200

48. Gornemann, J., Barrandon, C., Hujer, K., Rutz, B., Rigaut, G., Kotovic, K. M., . . . Seraphin, B. (2011). Cotranscriptional spliceosome assembly and splicing are independent of the Prp40p WW domain. *RNA (New York, N.Y.)*, *17*(12), 2119-2129. doi: 10.1261/rna.02646811; 10.1261/rna.02646811

49. Ohno, M., Sakamoto, H., & Shimura, Y. (1987). Preferential excision of the 5' proximal intron from mRNA precursors with two introns as mediated by the cap structure. *Proceedings of the National Academy of Sciences of the United States of America*, *84*(15), 5187-5191.

50. Izaurralde, E., Lewis, J., McGuigan, C., Jankowska, M., Darzynkiewicz, E., & Mattaj, I. W. (1994). A nuclear cap binding protein complex involved in pre-mRNA splicing. *Cell*, *78*(4), 657-668.

51. Edery, I., & Sonenberg, N. (1985). Cap-dependent RNA splicing in a HeLa nuclear extract. *Proceedings of the National Academy of Sciences of the United States of America*, *82*(22), 7590-7594.

52. Ohno, M., Kataoka, N., & Shimura, Y. (1990). A nuclear cap binding protein from HeLa cells. *Nucleic Acids Research, 18*(23), 6989-6995.

53. Patzelt, E., Thalmann, E., Hartmuth, K., Blaas, D., & Kuechler, E. (1987). Assembly of pre-mRNA splicing complex is cap dependent. *Nucleic Acids Research, 15*(4), 1387-1399.

54. Inoue, K., Ohno, M., Sakamoto, H., & Shimura, Y. (1989). Effect of the cap structure on pre-mRNA splicing in Xenopus oocyte nuclei. *Genes & Development*, *3*(9), 1472–1479. doi:10.1101/gad.3.9.1472

55. Konarska, M. M., Padgett, R. A., & Sharp, P. A. (1984). Recognition of cap structure in splicing in vitro of mRNA precursors. *Cell, 38*(3), 731-736.

56. O'Mullane, L., & Eperon, I. C. (1998). The pre-mRNA 5′ cap determines whether U6 small nuclear RNA succeeds U1 small nuclear ribonucleoprotein particle at 5′ splice sites. *Molecular and cellular biology*, *18*(12), 7510–7520.

57. Lewis, J. D., Izaurralde, E., Jarmolowski, A., McGuigan, C., & Mattaj, I. W. (1996). A nuclear cap-binding complex facilitates association of U1 snRNP with the cap-proximal 5' splice site. *Genes & Development, 10*(13), 1683-1698.

58. Colot, H. V., Stutz, F., & Rosbash, M. (1996). The yeast splicing factor Mud13p is a commitment complex component and corresponds to CBP20, the small subunit of the nuclear cap-binding complex. *Genes & Development, 10*(13), 1699-1708

59. Gamberi, C., Izaurralde, E., Beisel, C., & Mattaj, I. W. (1997). Interaction between the human nuclear cap-binding protein complex and hnRNP F. *Molecular and cellular biology*, *17*(5), 2587–2597.

60. De la Mata, M., Alonso, C. R., Kadener, S., Fededa, J. P., Blaustein, M., Pelisch, F., … Kornblihtt, A. R. (2003). A slow RNA polymerase II affects alternative splicing in vivo. *Molecular cell*, *12*(2), 525–532.

61. Muñoz, M. J., Santangelo, M. S. P., Paronetto, M. P., De la Mata, M., Pelisch, F., Boireau, S., … Kornblihtt, A. R. (2009). DNA Damage Regulates Alternative Splicing through Inhibition of RNA Polymerase II Elongation. *Cell*, *137*(4), 708–720. doi:10.1016/j.cell.2009.03.010

62. Nogues, G. (2002). Transcriptional Activators Differ in Their Abilities to Control Alternative Splicing. *Journal of Biological Chemistry*, *277*(45), 43110–43114. doi:10.1074/jbc.M208418200

63. Kadener, S., Fededa, J. P., Rosbash, M., & Kornblihtt, A. R. (2002). Regulation of alternative splicing by a transcriptional enhancer through RNA pol II elongation. *Proceedings of the National Academy of Sciences*, *99*(12), 8185–8190.

64. Ip, J. Y., Schmidt, D., Pan, Q., Ramani, A. K., Fraser, A. G., Odom, D. T., & Blencowe, B. J. (2010). Global impact of RNA polymerase II elongation inhibition on alternative splicing regulation. *Genome Research*, *21*(3), 390–401. doi:10.1101/gr.111070.110

65. Howe, K. J. (2003a). Perturbation of transcription elongation influences the fidelity of internal exon inclusion in Saccharomyces cerevisiae. *RNA*, *9*(8), 993–1006. doi:10.1261/rna.5390803

66. Schor, I. E., Rascovan, N., Pelisch, F., Alló, M., & Kornblihtt, A. R. (2009). Neuronal cell depolarization induces intragenic chromatin modifications affecting NCAM alternative splicing. *Proceedings of the National Academy of Sciences*, *106*(11), 4325–4330.

67. Gooding, C., Roberts, G. C., & Smith, C. W. (1998). Role of an inhibitory pyrimidine element and polypyrimidine tract binding protein in repression of a regulated alpha-tropomyosin exon. *Rna*, *4*(1), 85–100.

68. Robson-Dixon, N. D. (2004). MAZ Elements Alter Transcription Elongation and Silencing of the Fibroblast Growth Factor Receptor 2 Exon IIIb. *Journal of Biological Chemistry*, *279*(28), 29075–29084. doi:10.1074/jbc.M312747200

69. Cramer, Paula, Pesce, C. G., Baralle, F. E., & Kornblihtt, A. R. (1997). Functional association between promoter structure and transcript alternative splicing. *Proceedings of the National Academy of Sciences*, *94*(21), 11456–11460.

70. Ryan, K., Murthy, K. G. K., Kaneko, S., & Manley, J. L. (2002). Requirements of the RNA Polymerase II C-Terminal Domain for Reconstituting Pre-mRNA 3' Cleavage. *Molecular and Cellular Biology*, *22*(6), 1684–1692. doi:10.1128/MCB.22.6.1684-1692.2002

71. Dantonel, J.-C., Murthy, K. G., Manley, J. L., & Tora, L. (1997). Transcription factor TFIID recruits factor CPSF for formation of 3'end of mRNA. *Nature*, *389*(6649), 399–402.

72. Dye, M. J., & Proudfoot, N. J. (1999). Terminal exon definition occurs cotranscriptionally and promotes termination of RNA polymerase II. *Molecular cell*, *3*(3), 371–378.

73. Niwa, M., & Berget, S. M. (1991). Mutation of the AAUAAA polyadenylation signal depresses in vitro splicing of proximal but not distal introns. *Genes & Development*, *5*(11), 2086-2095.

74. Niwa, M., Rose, S. D., & Berget, S. M. (1990). In vitro polyadenylation is stimulated by the presence of an upstream intron. *Genes & Development*, *4*(9), 1552-1559.

75. Kyburz, A., Friedlein, A., Langen, H., & Keller, W. (2006). Direct interactions between subunits of CPSF and the U2 snRNP contribute to the coupling of pre-mRNA 3' end processing and splicing. *Molecular Cell*, *23*(2), 195-205. doi: 10.1016/j.molcel.2006.05.037

76. Millevoi, S., Loulergue, C., Dettwiler, S., Karaa, S. Z., Keller, W., Antoniou, M., & Vagner, S. (2006). An interaction between U2AF 65 and CF I(m) links the splicing and 3' end processing machineries. *The EMBO Journal*, *25*(20), 4854-4864. doi: 10.1038/sj.emboj.7601331

77. Vagner, S., Vagner, C., & Mattaj, I. W. (2000). The carboxyl terminus of vertebrate poly(A) polymerase interacts with U2AF 65 to couple 3'-end processing and splicing. *Genes & Development*, *14*(4), 403-413.

78. Lutz, C. S., Murthy, K. G., Schek, N., O'Connor, J. P., Manley, J. L., & Alwine, J. C. (1996). Interaction between the U1 snRNP-A protein and the 160-kD subunit of cleavage-polyadenylation specificity factor increases polyadenylation efficiency in vitro. *Genes & Development*, *10*(3), 325-337.

79. Awasthi, S., & Alwine, J. C. (2003). Association of polyadenylation cleavage factor I with U1 snRNP. *RNA (New York, N.Y.)*, *9*(11), 1400-1409.

80. Garas, M., Dichtl, B., & Keller, W. (2008). The role of the putative 3' end processing endonuclease Ysh1p in mRNA and snoRNA synthesis. *RNA (New York, N.Y.)*, *14*(12), 2671-2684. doi: 10.1261/rna.1293008; 10.1261/rna.1293008

81. Robberson, B. L., Cote, G. J., & Berget, S. M. (1990). Exon definition may facilitate splice site selection in RNAs with multiple exons. *Molecular and cellular biology*, *10*(1), 84–94.
82. Black, D. L. (1991). Does steric interference between splice sites block the splicing of a short c-src neuron-specific exon in non-neuronal cells? *Genes & Development*, *5*(3), 389–402. doi:10.1101/gad.5.3.389
83. Dominski, Z., & Kole, R. (1991). Selection of splice sites in pre-mRNAs with short internal exons. *Molecular and cellular biology*, *11*(12), 6075–6083.

84. Berget, S. M. (1995). Exon recognition in vertebrate splicing. *The Journal of Biological Chemistry*, *270*(6), 2411-2414.

CHAPTER 2

NOVEL CONNECTIONS BETWEEN PRE-mRNA SPLICING, 3' AND 5' END DETERMINANTS REVEALED BY A NOVEL HIGH-THROUGHPUT REVERSE GENETIC SCREEN[1]

2.1 Introduction

The coding portions of most eukaryotic genes are interrupted by non-coding introns which must be removed prior to the translation of their messenger RNAs (mRNA). Removal of introns from pre-mRNAs is catalyzed by the spliceosome, a large and dynamic ribonucleoprotein complex comprised of five small nuclear RNAs (snRNAs) and at least 100 proteins [1]. Much of our knowledge about the components that comprise the spliceosome as well as their mechanisms of action has been derived from experiments using the powerful genetic tools available in the budding yeast, *Saccharomyces cerevisiae*. Indeed, although the *RNA2 – RNA11* genes originally identified in Hartwell's forward genetic screen preceded the discovery of splicing [2], the mechanistic characterizations of these genes, since renamed *PRP2 – PRP11*, underlie current models of the splicing pathway. Importantly, because the core components of the spliceosome are highly conserved between budding yeast and humans, the mechanistic details derived from work in yeast have been instrumental in understanding mechanisms of pre-mRNA splicing in higher eukaryotes.

The modern view of pre-mRNA splicing acknowledges the integrated role of the spliceosome with several other aspects of RNA processing. Whereas the historical

[1] Part of this chapter, including all tables and figures 2.1-2.13 and 2.15-2.16 have been published in Albulescu *et al.*, PLoS. Genet. 2012.

view of splicing envisioned a cascade of temporal events initiated by transcription, followed by polyadenylation, and finalized with splicing and export of mRNAs from the nucleus, it is now clear that these pathways are not independent from one another but rather are functionally coupled. Strong evidence in both yeast and higher eukaryotes demonstrates that recruitment of the spliceosome to intron-containing transcripts occurs co-transcriptionally [3-6], mediated at least in part by physical associations between the C-terminal domain (CTD) of RNA polymerase II and the U1 snRNP [7]. A growing body of evidence also indicates that the landscape of chromatin modifications encountered by transcribing polymerase molecules can dictate the activity of the spliceosome at various splice sites. For example, recent work has identified an enrichment of methylated lysine-36 in the histone H3 protein specifically within exonic sequences, suggesting a possible mechanism for facilitating the identification of intron-exon boundaries [8, 9]. Similarly, the rate of transcription by RNA polymerase II, which can be impacted by chromatin marks, has also been shown to be critical for dictating alternative splicing decisions [10]. Furthermore, it is also clear that splicing is coupled to down-stream steps in RNA processing. For example, the yeast Ysh1 protein [11, 12], which is the homolog of CPSF73, the mammalian endonuclease required for 3' end processing, was originally identified as Brr5 in a cold-sensitive screen for mutants defective in pre-mRNA splicing [13]. Consistent with this observation, recent evidence suggests that transcriptional pausing near the 3' end of genes is a critical component of pre-mRNA splicing efficiency [14]. Despite the increasing evidence of the interconnectivity of these pathways, in many cases the mechanistic details which underlie these functional relationships remain unclear. Our

understanding of these mechanistic connections would benefit from a more complete characterization of the complement of factors through which splicing is connected to these cellular processes.

A variety of recent genome-wide approaches have provided important insights into the connections that exist between the spliceosome and other cellular processes. Two powerful approaches, Synthetic Genetic Array (SGA) Analysis [15] and Epistatic MiniArray Profiling (E-MAP) [16], leverage genetic tools available in yeast to systematically generate millions of double-mutant strains and then carefully quantitate their cellular fitness to determine an interaction score for every pair-wise mutation. On the basis of strong positive or negative genetic interaction scores these approaches have been successfully used to infer functional relationships between many cellular pathways, including several with pre-mRNA processing [17, 18]. Simultaneously, improvements in proteomic methodologies have enabled the direct analysis of protein complexes in organisms as diverse as humans and yeast, allowing for an assessment of all of the stably-bound proteins involved in pre-mRNA splicing in many organisms [19, 20]. While the combination of these and other approaches has provided a global picture of many of the cellular factors that influence the splicing pathway either directly or indirectly, an important question remains about the functional significance of these factors in the splicing of specific transcripts. Indeed, it has long been known that certain transcripts require the activity of unique accessory factors to facilitate their splicing [21]. Moreover, recent work supports the idea that different transcripts can have a greater or lesser dependence upon the activity of core spliceosomal components for their efficient splicing [22, 23].

Here I present the results of a novel approach that complements the genetic and physical approaches of others by allowing for a direct functional assessment of nearly every gene in the *S. cerevisiae* genome in the pre-mRNA splicing process. For this work, I developed automated methods that enabled the isolation of total cellular RNA from about 5500 unique strains, each of which contained a mutation in a single gene, and all of which were examined during exponential growth in liquid medium. Using a high-throughput quantitative PCR (QPCR) assay, the relative cellular level of nearly any RNA can be readily determined in the background of each of these strains. By assessing the levels of several different pre-mRNA species, I was able to identify not only those factors which are necessary for the splicing of many transcripts, but also factors that are specifically required for the splicing of a subset of intron-containing genes. Whereas this study specifically examines the levels of several cellular pre-mRNAs, the approach described herein can be easily adapted to study the level of nearly any RNA molecule of interest under a wide variety of cellular growth conditions.

2.2 Materials and Methods

2.2.1 High-throughput strain handling

All experiments were performed using haploid strains. To assess the function of non-essential genes, the mat a version of the haploid deletion library from Open Biosystems [55] was used (referred to herein as non-essential strains). Likewise, to assess the function of essential genes, a collection of strains provided by the Hieter lab

[25] was used (referred to herein as essential strains). In addition, a collection of strains containing previously characterized mutations in core spliceosomal components was used (from here on considered a part of the essential strains set). A complete list of the strains used in this work can be found online at *http://www.plosgenetics.org/article/info%3Adoi%2F10.1371%2Fjournal.pgen.1002530*. Unless otherwise indicated, all strains were grown in rich medium supplemented with 2% glucose (YPD) [56]. When appropriate, strains were recovered from frozen glycerol stocks on solid medium supplemented with 200µg/ml G418 grown at either 30°C (non-essential strains) or 25°C (essential strains). A manual pinning tool (V&P Scientific, cat.#: VP384FP6) was used to transfer cells from solid medium 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 900rpm 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 vary significantly [36], an approach was developed to enable the systematic collection of a similar number of rapidly dividing cells for each strain. An initial liquid culture was grown in 384-well plates for two days, allowing nearly all strains to reach saturation. Because all of the strains being used are derived from a common parental strain, the cell density for each of these strains is nearly identical at saturation, allowing us to effectively 'normalize' the cell numbers. Using a liquid handling robot (Biomek NX), 2 µl of saturated culture were used to inoculate a fresh 150 µl of YPD. This culture was allowed to grow for four hours, an amount of time which is sufficient to allow all strains to exit lag-phase

and begin exponential growth, but not so long as to result in a large variation in cell densities among the strains (Figure 2.1).

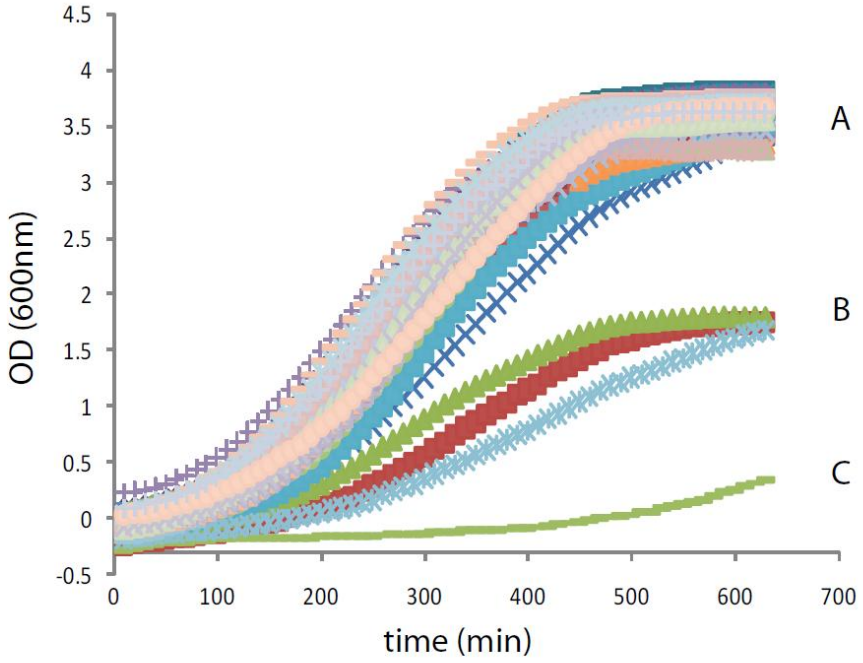

Figure 2.1: Growth rates of mutant strains in liquid culture

Growth curves for a subset of 96 mutant deletion strains over a 600 minute time interval. Whereas the majority of the strains (A) grow at a rate which is similar to wild type, a small number of strains (B) grow at a slightly reduced rate, while one strain (C) grows very slowly. On the basis of these data, I chose to harvest cells after 4 hours of outgrowth, which ensures that the majority of the strains are harvested when A_{600}~0.5.

For the non-essential strains, all growth was conducted at 30°C. For the essential strains, the initial growth was done at 25°C (a permissive temperature for all strains), but the saturated cells were back-diluted into plates containing media pre-warmed to 30°C (a non-permissive temperature for many, but not all, of the strains) and allowed to continue growing at 30°C for four hours. For both the non-essential and the

essential strains, two independent biological replicates were initiated from each saturated plate. After four hours of outgrowth, cells were harvested by centrifugation at 4000*xg* for five minutes. The cell pellets were flash frozen in liquid N_2 and stored at -80°C until further processing.

2.2.2 High-throughput RNA isolation

Isolation of total cellular RNA was performed using custom protocols written for a Biomek NX liquid handling system. To each frozen cell pellet collected as described above, 50µl of Acid Phenol: Chloroform (5:1, pH<5.5) and 25µl of AES buffer (50mM sodium acetate (pH5.3), 10mM EDTA, 1% SDS) were added. The plates were sealed with plastic CapMats (Greiner BioOne, cat.#: 384070) and vortexed for five minutes at top speed on a plate vortex. The plates were incubated for 30 minutes in a water bath at 65°C with intermittent vortexing. After incubation, the plates were spun for one minute at 1000x*g*. An additional 35µl of AES buffer was added to each well, and after mixing the organic and aqueous phases were separated by centrifugation for five minutes at 3000x*g*. Using a slow aspiration speed, 40µl of the upper phase containing the RNA were robotically transferred to a new 384-well microtiter plate. The transferred aqueous phase was mixed with 3 volumes of RNA Binding Buffer (2M Guanidine-HCl, 75% isopropanol) and passed through a 384-well glass fiber column (Whatman, cat.#: 7700-1101) by centrifugation for two minutes at 2000x*g*. The column was washed twice with two volumes of Wash Buffer (80% ethanol, 10mM Tris-HCl (pH8.0)), followed by a final dry spin for two minutes at 2000x*g*. To remove any contaminating genomic DNA, 5 µl of DNase Mix (1x DNase Buffer, 0.25

units of DNase I (Promega)) was added to each well and incubated at room temperature for 15 minutes. After the incubation, 80µl of RNA Binding Buffer was added to each well of the 384-well glass fiber plate and spun as before. After washing and drying as above, 15µl of sterile water was added to each well of the glass-fiber plate to elute the RNA into a clean 384-well microtiter plate (Greiner BioOne, cat. #: 781280). In general, this procedure yielded about 1µg of total cellular RNA from each cell pellet. The quality of the RNA produced by this protocol is equal to our conventionally purified samples, and the effectiveness of the DNase treatment is demonstrated in Figure 2.2.

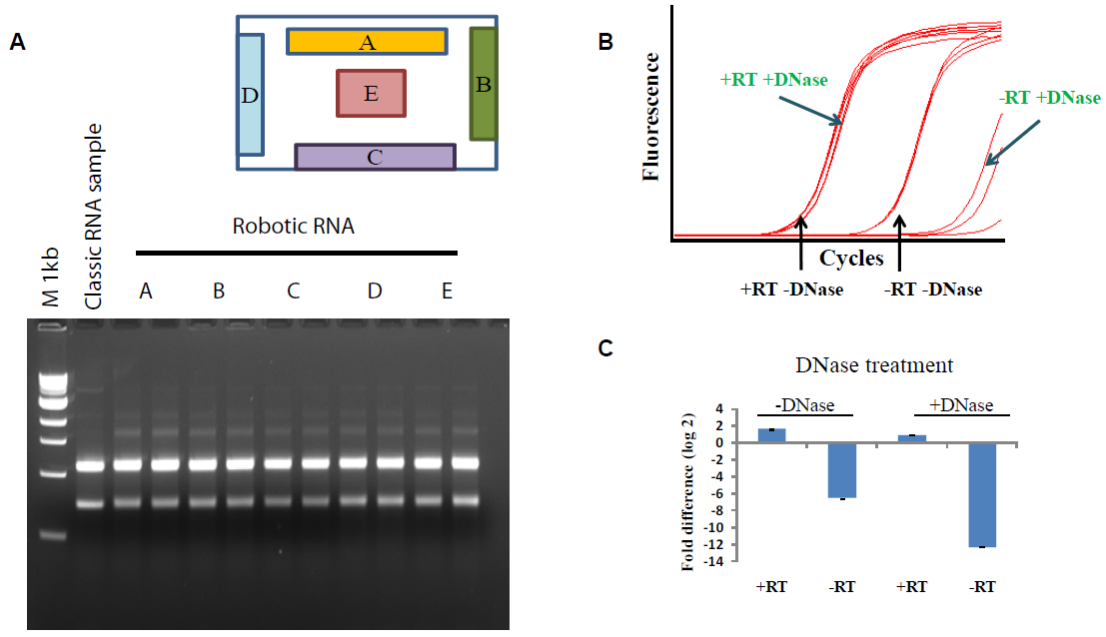

Figure 2.2*: *Assessing Total RNA and cDNA quality

A. A comparison of RNA quality between a "classic" phenol-extraction protocol and our robotic procedure. The diagram above the picture of the gel indicates the regions of a 384-well plate from which RNA was selected and run. Each lane contains about 300ng of total cellular RNA and duplicates from each region of the plate are shown.

B. Effectiveness of the DNase treatment as measured by quantitative RT-PCR, demonstrating the shift in fluorescence before and after DNase treatment. The green text indicates the samples that were treated with DNase, whereas the black text indicates the untreated samples.

C. The measured levels of Tef5 pre-mRNA in the indicated samples demonstrates that the contaminating genomic DNA is strongly depleted by DNase treatment.

2.2.3 Synthesis of cDNA and quantitative PCR

Total cellular RNA was converted into cDNA in 384-well microtiter plates. Of the 15µl of RNA purified as described above, 10µl were used in a cDNA synthesis reaction that had a total volume of 20µl and which contained 50mM Tris-HCl (pH8.3), 75mM KCl, 3mM $MgCl_2$, 10mM DTT, 0.5mM each dNTP, 5µg dN_9 primer, and 60ng M-MLV RT. Reactions were incubated overnight at 42°C.

The cDNA reactions were diluted 30-fold with water, giving a final concentration of ~ 1ng/µl based on the initial RNA concentration, and used without any further purification as templates in high-throughput quantitative PCR (QPCR) reactions. The QPCR reactions were performed in a reaction volume of 10µl, containing 5µl of template (~5ng of template), 10mM Tris-HCl (pH8.5), 50mM KCl, 1.5mM $MgCl_2$, 0.2mM each dNTP, 0.25x SYBR Green, 5% DMSO, 0.7ng *Taq* DNA polymerase, and 250nM forward and reverse primers. The sequences of the primers used for each targeted RNA are shown in Table 2.1.

Table 2.1: Forward and Reverse primer sequences used for QPCR for each target RNA

The list of forward and reverse primers used in this study.

Target	Forward primer	Reverse primer
U3 precursor (snR17a)	TAT GTA ATA TAC CCC AAA CAT TTT ACC CAC	AAA CTC TAG TAC CTA AGA CTT TTA GAT GCT
Tef5 pre-mRNA	GAT AGC ACA GAG CAG AGT ATC ATT A	CTG GAG AAT TCT GGG TAA GCA GAT T
Rpl31b pre-mRNA	ATT TCT CTG TGT TCT GCG ATC GAT	AGC GCC ATT ATA GTG TAA ACG TGA G
Ubc13 pre-mRNA	GAG CCG TCA CAC TCT ACA CTG TTT C	TTG CGT TAG TAT GTA AGT TCT AGC CA
Tub3 pre-mRNA	GCA AAT AAT TTG ACT TGT ACC AGC A	AGG TCA ATT ATG AGC TTT TTC GTC C
Yra1 pre-mRNA	ACA TTT TTA AGC TGG CGT ATT GTG T	TAG GAA CTA ACT ATC ACG GTC GAG G
Rec107 pre-mRNA	TCG CTA AAC AAG TTT AAC CAG CTA CT	AAA CTT GAG GGT ATG CAC TGT GTT
U1 snRNA	AAG TCC TAC TGA TCA AAC ATG CGC	GGA GTT TGC ATC AAT GAC TTC AAT G
Scr1 total	CAG GAG GCG TGA GGA ATC CGT CTC T	AAG GAC CCA GAA CTA CCT TGC CGC A
Tef5 total	CCA TTG TCA CTC TAG ATG TCA AGC C	GAT ACC GAA ACC AAT TGG GAT AAA T
Tub1 total	AGA AAC TTG GAT ATC CCA AGA CCA A	GCA ATT AGG TTG TTT AAG TTT GCA A
Faa1 total	TCT GCC CTA TGC TTA TTG GTT ACG	TAA CAC AAC CTG TTA GGT CAC CAG C
Srb2 total	CGC GGA TGT CTC CAA GCT CA	CGC GGA TGT CTC CAA GCT CA
Mud1 total	AAT AAG ACG CCT GAA GCA TAA GCT	CTA TCA ATC TCA GCC TCC TCT ACC TT

Standard curves were generated consisting of 4-fold serial dilutions of genomic DNA and covering a range of $1.6x10^5$ molecules. Each primer pair was well-behaved, showing an amplification efficiency of between 86% and 97% (Figure 2.3). Two technical replicates were measured for each biologically independent sample, generating four independent measurements for each of the ~5500 mutant strains.

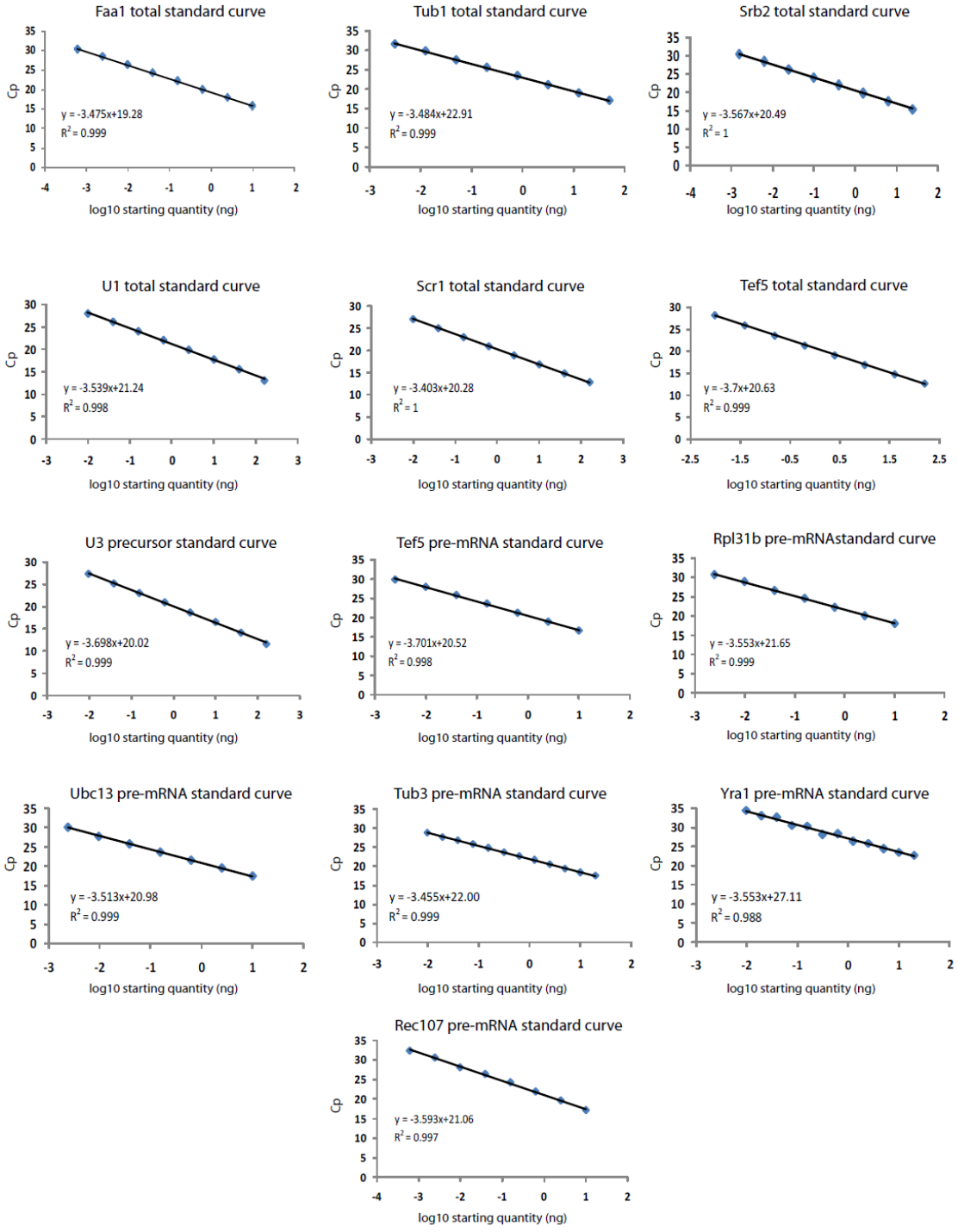

Figure 2.3: Standard curves for quantitative PCR (QPCR) for all primer pairs

Standard curves for all primer pairs used in this study. The Cp values (crossing points) are plotted against the logarithm of known starting sample amounts (in nanograms of genomic DNA). The equations that describe the curves as well as the R^2 values are specified for each of the primer pairs

2.2.4 Processing QPCR data

On the basis of standard curves generated using QPCR, relative nanogram quantities were calculated for every RNA transcript within each of the ~5500 strains tested. To assess reproducibility, coefficients of variation (CV) were determined for each primer pair and each strain. The vast majority of these were highly reproducible, both overall and on a per plate basis. As an initial quality filter, I chose to exclude any samples for which the CV was greater than 0.25.

Because no simple mechanism exists to normalize for variability in each of the experimental steps, I instead chose to measure the levels of six different RNAs in each of the samples and use these to determine a composite normalization value to account for the overall yield in our procedure. The six RNAs were: U1 snRNA, Scr1 (SRP) RNA, Tef5 mRNA, Tub1 mRNA, Srb2 mRNA and Faa1 mRNA. These RNAs were chosen because their biological functions are diverse and their cellular levels vary over a broad range (~300-fold, Table 2.2).

Table 2.2: Relative abundances of RNA species as measured by QPCR

RNA species for which expression levels were measured and their corresponding levels. The values are normalized to the Tub3 pre-mRNA level, which was the lowest abundance transcript measured in our experiments.

RNA	Relative abundance
U1 total	1800
Scr1 total	1200
Tef5 total	280
Yra1 pre-mRNA	110
U3 precursor	70
Tub1 total	35
Faa1 total	12

Srb2 total	5
Rec107 pre-mRNA	3
Rpl31b pre-mRNA	3
Tef5 pre-mRNA	3
Ubc13 pre-mRNA	2
Tub3 pre-mRNA	1

For both independent biological replicates of every strain, a composite normalization constant, $C_{norm_{x_i}}$, was calculated according the following formula:

$$C_{norm_{x_i}} = 2^{\left(\frac{\log_2\left(\frac{U1_{ng_{x_i}}}{U1_{ng_{MED_{Plate}}}}\right)+\log_2\left(\frac{SCR1_{ng_{x_i}}}{SCR1_{ng_{MED_{Plate}}}}\right)+\log_2\left(\frac{FAA1_{ng_{x_i}}}{FAA1_{ng_{MED_{Plate}}}}\right)+\log_2\left(\frac{SRB2_{ng_{x_i}}}{SRB2_{ng_{MED_{Plate}}}}\right)+\log_2\left(\frac{TEF5_{ng_{x_i}}}{TEF5_{ng_{MED_{Plate}}}}\right)+\log_2\left(\frac{TUB1_{ng_{x_i}}}{TUB1_{ng_{MED_{Plate}}}}\right)}{6} \right)}$$

For each primer pair, ng_{x_i} represents the relative nanogram quantity calculated for an individual sample. Similarly, $ng_{MED_{Plate}}$ represents the median value determined for a given primer pair on an individual QPCR plate run. Because of the subtle variations that are apparent from one plate run to the next, I found that this per plate normalization using $ng_{MED_{Plate}}$ gave the most robust data. By determining the ratio of $\frac{ng_{x_i}}{ng_{MED_{Plate}}}$ for every primer pair, a relative abundance of total RNA can be calculated for every sample. As seen in Figure 2.4, a histogram of $C_{norm_{x_i}}$ values follows a normal distribution in $\log_2$ space with a variance of 1.5 units. A second filter at the level of $C_{norm_{x_i}}$ values was introduced which allowed for the filtering of samples with very low amounts of cDNA.

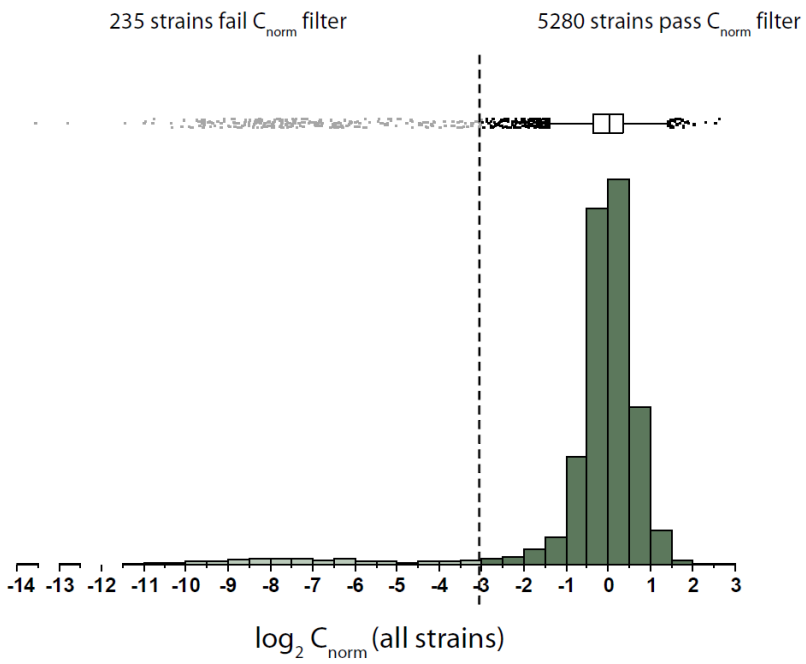

Figure 2.4*: *Distribution of C_{norm} values

The distribution of the normalization constants (C_{norm}) in $\log_2$ space for all samples in the biological replicate A dataset. The area highlighted in dark green comprises C_{norm} values which have passed the C_{norm} filter. All $\log_2(C_{norm})$ values less than -3 (an 8-fold decrease from the median) were eliminated from the dataset. Above the histogram, a box and outlier plot underlines the same distribution of C_{norm} values

For strains that passed this filter, the normalized levels of a given RNA were determined according to the following formula:

$$RNA_{ng_{norm_{x_i}}} = \frac{RNA_{ng_{x_i}}}{C_{norm_{x_i}}}$$

The relative amount of RNA in a given strain was then determined according to the following formula:

$$RNA_{rel_{x_i}} = \log_2\left(\frac{RNA_{ng_{norm_{x_i}}}}{RNA_{ng_{norm_{x_{MEDPlate}}}}}\right)$$

For each primer pair, $RNA_{ng_{norm_{xMEDPlate}}}$ represents the median value of the normalized RNA levels determined within a given QPCR plate. For each biological replicate of every strain, both the ng_{x_i} and the $RNA_{rel_{x_i}}$ values are available through Gene Expression Omnibus (GEO) using accession number GSE34330.

2.2.5 Significance analysis of QPCR data

To determine the subset of strains that cause a statistically significant increase or decrease in precursor levels, we employed the *Significance Analysis of Microarrays*, or SAM, program [37]. While this software was originally designed for the analysis of microarray data, a significance analysis of our QPCR data is subject to similar concerns regarding multiple hypothesis testing. For each RNA, SAM analysis was performed on the four $RNA_{rel_{x_i}}$ values, comprised of both technical and biological replicates that were generated for each of the ~5500 strains. For each transcript, a one class SAM analysis was performed where the Δ value was adjusted to minimize the false discovery rate (FDR), yielding the following values: for the U3 precursor using Δ=0.983, FDR=0.045; for the Tub3 pre-mRNA using Δ=0.91, FDR=0; for the Rpl31b pre-mRNA using Δ=1.061, FDR=0.003; for the Ubc13 pre-mRNA using Δ=0.978, FDR=0.002; and for the Tef5 pre-mRNA using Δ=0.99, FDR=0.

2.2.6 Splicing-sensitive microarrays

The candidate non-essential deletion strains were grown to saturation in YPD at 30°C, then back diluted in 25ml cultures in flasks at a starting A_{600}~0.2 and allowed to grow

at 30°C until they reached an optical density of between A_{600}=0.5 and A_{600}=0.7. The candidate essential strains were initially grown at 25°C in YPD, then shifted to the indicated temperatures for 15 minutes after they reached an optical density of between A_{600}=0.5 and A_{600}=0.7. In parallel with the collection of the mutant strains, wild type isogenic controls were grown and collected under the same conditions as the mutant strains examined. Total cellular RNA samples were isolated, converted into cDNA, and fluorescently labeled as previously described [31]. All microarrays were performed as two-color arrays comparing mutant and wild type strains, each grown under identical conditions. Both raw and processed microarray data are available through GEO using accession number GSE34330.

2.2.7 Chromatin immunoprecipitation

The U1C-Tap *bdf1Δ* and U1C-Tap *bdf2Δ* strains were generated by deleting the appropriate genes in the background of the U1C-Tap strain from Open Biosystems [57] using standard techniques. The strains were grown at 30°C in rich medium supplemented with 2% glucose (YPD) until they reached an optical density of A_{600}~0.7. The chromatin was cross-linked with 1% formaldehyde for 2 minutes at 30°C. Glycine was added at a final concentration of 125mM and the cultures were left shaking for another 5 minutes. Cell pellets from 50 ml of culture were collected by centrifugation at 1620xg for 3 minutes, then washed with 25ml ice-cold 1X PBS and the pellet stored at -80°C. The pellets were resuspended in 1ml Lysis buffer (50mM Hepes pH 7.5, 140mM NaCl, 1mM EDTA, 1% TritonX-100, 0.1% Na-deoxycholate supplemented with protease inhibitors) and lysed in the presence of 500µl 0.5mm

glass beads in a beat beater. The lysate was collected by centrifugation at 1000xg for 1 minute, and then pre-cleared by spinning for 15 minutes at 14000rpm in a tabletop centrifuge at 4°C. The pellet was re-suspended in another 1ml of Lysis buffer and the chromatin was sheared to an average size of 300bp (range 100-500bp) by means of a Bioruptor sonicator. The sample was clarified by 2 cycles of centrifugation at 14000rpm for 15 minutes in a tabletop centrifuge at 4°C and the resultant chromatin solution frozen and stored at -80C. From the chromatin samples a 1% Input sample was retained, and then each sample was split equally between a Mock IP and an IP sample. The IP samples were incubated with 5µl 0.5mg/ml anti-Tap Antibody (Thermo Scientific, CAB1001). After 2 hours at 4°C on a rotator, 25µl of protein A/G-agarose resin (#Sc-2003Santa Cruz) was added to all samples and they were further incubated for another 2h at 4°C. The resin was washed twice with 1ml Lysis buffer, twice with 1ml Wash buffer (10mM Tris-HCl, 25mM LiCl, 0.5% NP-40, 0.5%Na-deoxycholate, 1mM EDTA) supplemented with 360mM NaCl, twice with 1ml Wash buffer, and finally twice with 1ml TE. The first wash was a brief one, followed by a 15 minute incubation of the samples on a rotator at 4°C for the second wash. In between washes, the resin was collected by short spins at 2000rpm in a tabletop centrifuge. The resin was resuspended in 100ul Elution buffer (50mM Tris-HCl pH8.0, 5mM EDTA, 1% SDS) and the immunoprecipitated material was eluted from the beads by incubating at 65°C for 30 minutes with occasional tapping. To reverse crosslinks, the IPs and the 1% Input samples were incubated overnight in a 65°C water bath. The next day, the samples were treated with 12.5µl 20mg/ml Proteinase K solution and incubated at 42° for 2h. The DNA was then purified by using a Cycle Pure Kit

(Omega Bio-Tek, D6492-01) following the manufacturer's instructions and eluted in a final volume of 120µl.

Quantitative real-time PCR was performed on a Roche Light Cycler 480 machine as described above, using the 1% Input sample to generate a standard curve for each of the primer pairs we used. For the primers used in the screen, the sequences are available in Table 2.1. The primers for the different regions of actin gene and the PMA1 gene are the same as previously published [5]. For each sample the Mock IP value calculated as percent input was subtracted from the IP value (in percent input). Then, a fold enrichment value was calculated, by dividing these values by the PMA1 value.

2.2.8 Mud1 overexpression

An overexpressing plasmid containing a full-length copy of the Mud1 gene including ~500bp up- and down-stream of the ORF was transformed into BY4741 (Open Biosystems). This strain and a control strain containing the empty vector were grown in 25 ml minimal media until they reached an optical density of A_{600}~0.5-0.6. RNA isolation was performed as previously described [31], and cDNA synthesis and Q-PCR were performed as described above. The primer sequences are found in Table 2.1.

2.3 Results

2.3.1 A high-throughput method for measuring cellular levels of specific RNA species

To identify the comprehensive network of cellular factors that lead to a change in splicing efficiency, I developed a high-throughput reverse genetic screen that allowed me to readily assess changes in pre-mRNA levels in the background of ~5500 *Saccharomyces cerevisiae* strains, each of which contained a mutation in a single gene. The library of strains contained deletions of non-essential genes [24] as well as conditional mutations in essential genes [25], accounting for mutational access to over 93% of known yeast genes. Using a liquid-handling robot, protocols were developed (see Materials and Methods) that allowed for the simultaneous collection of each of these strains under exponential growth conditions in liquid medium in 384-well plates. Total cellular RNA was isolated robotically from each of these strains using a phenol extraction protocol [23] followed by a glass-fiber purification step [26]. After converting this RNA into cDNA using a random-priming strategy, QPCR was used to directly measure the level of a given RNA species within each strain. Because of the inherent variability between the samples in the cell collection, RNA isolation, and cDNA synthesis steps, the levels of six different RNA species were measured in each of the samples in order to calculate a normalization constant. On the basis of this normalization constant, the relative level of virtually any cellular RNA species can be determined in each of the mutant strains.

As an initial test of the approach I sought to identify the full complement of factors

involved in pre-mRNA splicing by determining the relative levels of unspliced U3 small nucleolar RNA (snoRNA) present in each of the mutant strains. The U3 snoRNA is unique in the *S. cerevisiae* genome in that it is the only known non-coding RNA that is interrupted by a spliceosomal intron [27]. Nevertheless, the U3 transcript has been widely used historically as a splicing reporter, owing to its relatively high basal expression level and the strong accumulation of U3 precursor levels observed in the background of canonical splicing mutants [13,28,29]. As shown in Figure 2.5, the U3 precursor levels are unaffected in the vast majority of the strains examined, with levels varying by less than 1.35-fold from one another for 95% of the strains.

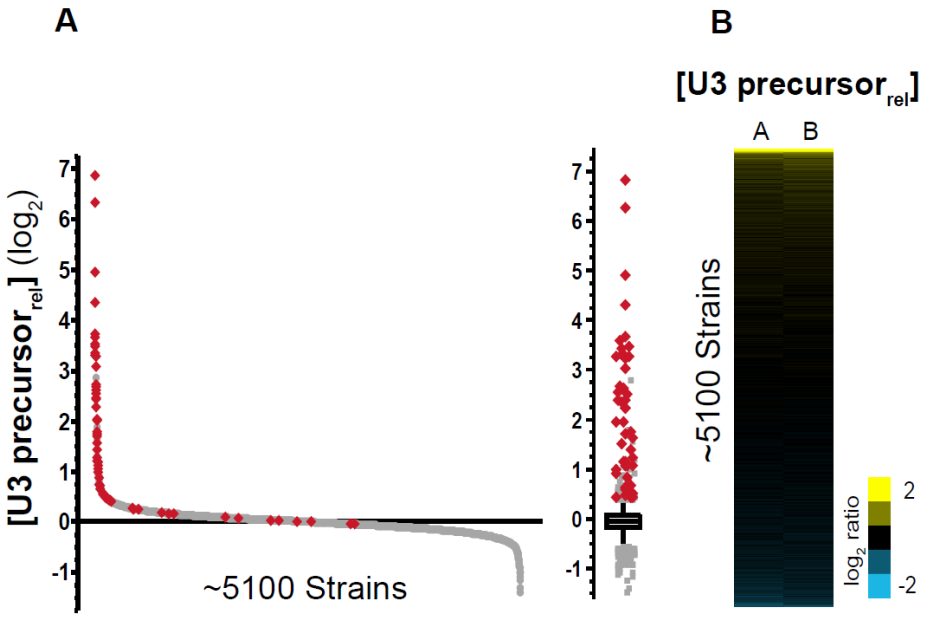

Figure 2.5: A subset of gene disruptions causes an increase in the unspliced levels of U3 snoRNA

A. The relative level of U3 snoRNA precursor in 5122 strains, ordered from the highest to lowest. Red data points highlight the strains containing mutations in a factor within the GO PROCESS: *RNA Splicing* category (see also Figure 2.6). To the right, a box and outlier plot for the same dataset.

B. False color representation of both the A (left) and B (right) biological replicates. The values are ordered from high to low, mirroring the data in (A).

Indeed, only ~200 of the ~5100 strains that passed the quality filters (see Materials and Methods) showed a change in the relative U3 precursor levels of more than ~30% from the median value (~0.35 in $\log_2$-transformed space), consistent with the expectation that mutations in most genes will have little or no effect on cellular pre-mRNA splicing efficiency. The tight distribution of relative U3 precursor levels seen within this dataset demonstrates the high precision with which these measurements can be made, and suggests a low false discovery rate for this approach.

2.3.2 Known spliceosomal components dominate the top subset of mutations

To characterize the data generated by this approach I sought to define the biological significance of those strains that showed increased levels of U3 precursor. As an initial analysis, I examined the U3 precursor levels in those strains containing mutations in known splicing genes. Using the GO PROCESS: *RNA Splicing* as a guide [30], a total of 71 strains in our library were classified as containing mutations in canonical splicing factors (Figure 2.6), of which 68 passed our quality filters for the U3 precursor dataset (see Materials and Methods).

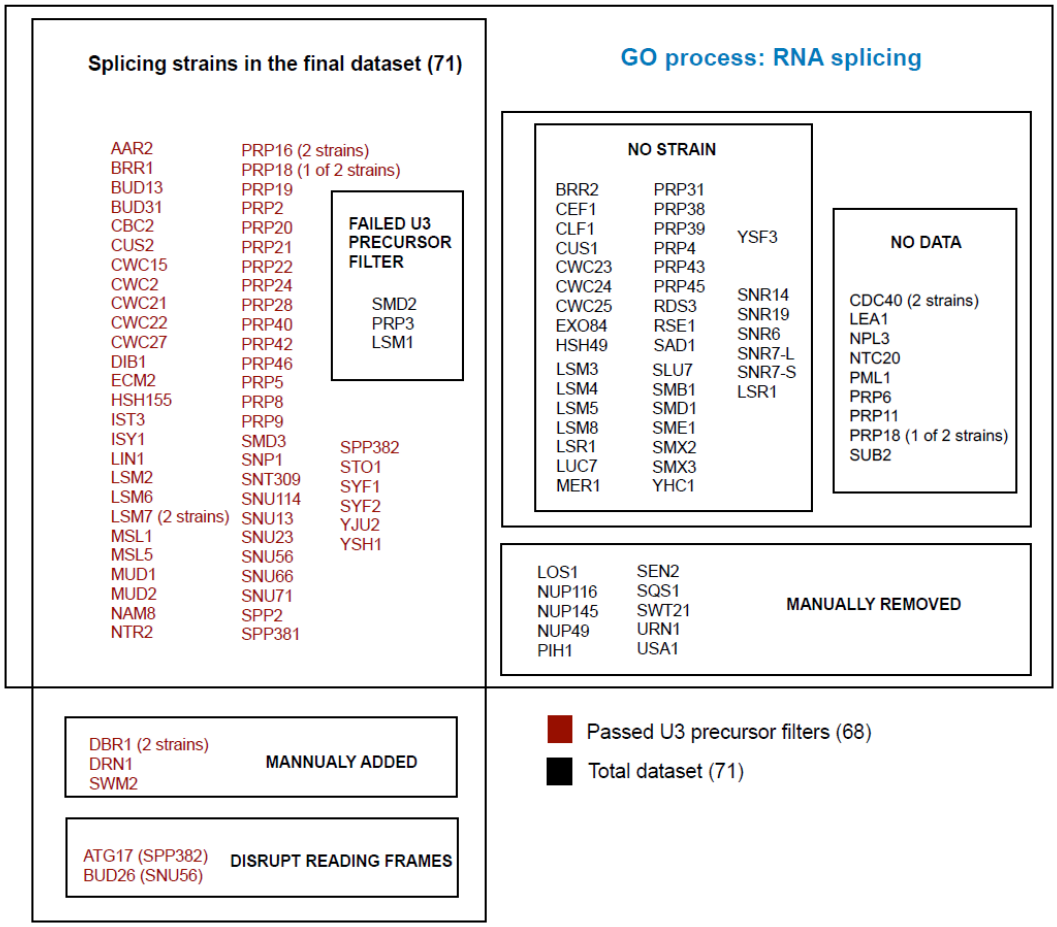

Figure 2.6: Splicing factors included within strain library

Within the "GO process: RNA splicing" box are listed all splicing factors as annotated by Gene Ontology [30]. These factors have been subdivided based on the presence or absence of a strain within the library, and further by the presence or absence of data from the final filtered dataset. The subset highlighted in red represents the strains which passed all the filters in the U3 snoRNA precursor screen. Strains for which there is "NO DATA" that passed the quality filters are listed separately. Factors for which "NO STRAIN" was present in the library are also noted. I "MANUALLY REMOVED" several strains that are included by Gene Ontology but that I felt are either are not part of the spliceosome or have a yet uncharacterized function. In addition, I have "MANUALLY ADDED" three factors which are known to impact splicing. Two more strains that partially "DISRUPT READING FRAMES" of specific splicing factors (indicated in the parentheses) were also added to the dataset.

A strong overrepresentation of these splicing factors can be seen within the set of

strains showing an enrichment of U3 precursor (Figure 2.5A). Of the 68 strains containing splicing mutations that passed the quality filters: 53 are found within the top 200 strains (p=9.28E-64, Fisher's exact test); 38 are found in the top 50 strains (p=1.33E-66); and the top 14 strains all belong to this list (p=1.27E-27). Taken together, these data argue strongly that the candidates identified by this approach will be characterized by a high true positive discovery rate.

By contrast, out of the 68 strains containing mutations in known splicing factors for which I obtained high quality data, 15 failed to show an enrichment of U3 precursor levels in this dataset, suggesting either that mutations in these genes don't cause an increase in U3 precursor levels (true negative), or that our approach incorrectly failed to detect the accumulation of unspliced U3 (false negative). To better resolve these possibilities I chose to more completely examine the global splicing fitness of some of these strains using splicing-sensitive microarrays. For every intron-containing gene in the genome, these custom-designed microarrays contain at least three probes (Figure 2.7A) that allow us to distinguish between spliced and unspliced isoforms [31]. I used these microarrays to assess the global splicing defects of four mutants: two canonical splicing mutants that showed strong U3 precursor accumulation (*snt309Δ* and *lsm6Δ*, Figure 2B), and two that showed little or no accumulation (*mud2Δ* and *cus2Δ*, Figure 2.7C).

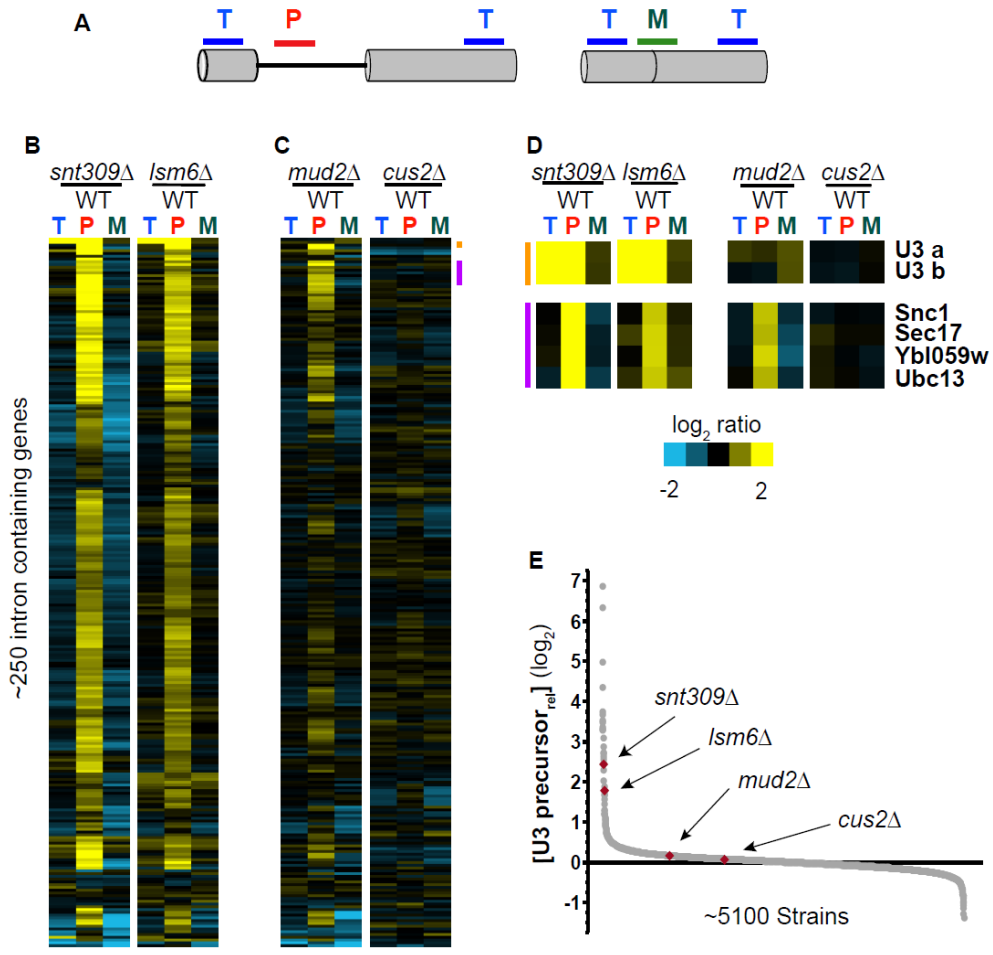

Figure 2.7*: *Not all mutations in splicing factors result in defective splicing for all intron-containing genes

A. Splicing-sensitive microarrays contain probes that target precursor (P), mature (M), and total (T) mRNA levels.

B. Global splicing profiles of the *snt309Δ* and *lsm6Δ* strains, showing a strong splicing defect for both mutants.

C. Global splicing profiles of the *mud2Δ* and *cus2Δ* strains, illustrating different splicing defects for these mutants. Whereas many transcripts show a splicing defect in the *mud2Δ* strain, many continue to be efficiently spliced. Similarly, most intron-containing genes show no splicing defect in the *cus2Δ* strain.

D. The behavior of specific transcripts in the different mutant backgrounds.

E. Locations within the U3 precursor dataset of the mutant strains for which microarrays are shown in (B) and (C).

As expected, and consistent with previous work from others [32], the *snt309Δ* and

lsm6Δ strains demonstrate a broad splicing defect, with most intron-containing genes displaying an increase in precursor levels accompanied by a decrease in the amount of spliced mRNA. By contrast, the global splicing profiles of the *mud2Δ* and *cus2Δ* strains are markedly different. In the *cus2Δ* background, few intron-containing genes display a splicing defect: very little precursor accumulation is observed, and there is little if any detectable loss in mature mRNA. The *mud2Δ* mutation does cause a splicing defect for some intron-containing genes, whereas little change in splicing efficiency is seen for many others. Notably, as seen in Figure 2.7D, the microarrays of both the *snt309Δ* and *lsm6Δ* strains show a strong accumulation of U3 precursor levels, whereas the *mud2Δ* and *cus2Δ* strains show almost no accumulation, consistent with the QPCR screen results. It is worth noting that in my experience the behavior of the U3 transcript differs from the other intron-containing genes in that every splicing mutation I have examined that causes an increase in the U3 precursor levels also results in an *increase* in the total level of U3; the reason for this apparent discrepancy is currently under investigation. Nevertheless, these microarray data demonstrate that the failure to detect an increase in U3 precursor levels in the *mud2Δ* and *cus2Δ* strains does not represent a failure of the approach, but rather that these are true negative results.

2.3.3 Assessing transcript-specificity of the mutants

To better assess the total complement of genes that can impact the splicing of any precursor transcript, I chose to expand the analysis by measuring the precursor levels of several additional intron-containing genes. I chose to examine four 'canonical' intron-containing genes (*RPL31B, UBC13, TUB3* and *TEF5*) that vary in terms of intron size, transcriptional frequency, biological function, and the presence or absence of an intron-encoded snoRNA. In spite of these differences, these transcripts are similar to one another in so much as they each contain splice site and branch point sequences that conform to consensus sequences. In addition to these four genes, I chose to examine two intron-containing genes (*YRA1* and *REC107/MER2*) that are known to be poorly spliced under standard growth conditions [21, 33, 34]; as such, I expected the behavior of these two transcripts to be distinct from the efficiently spliced transcripts. For all six of these genes, the precursor levels were measured in all ~5500 strains. As an initial analysis of this data set, I considered the behavior of the 71 strains containing mutations in spliceosomal components (Figure 2.8).

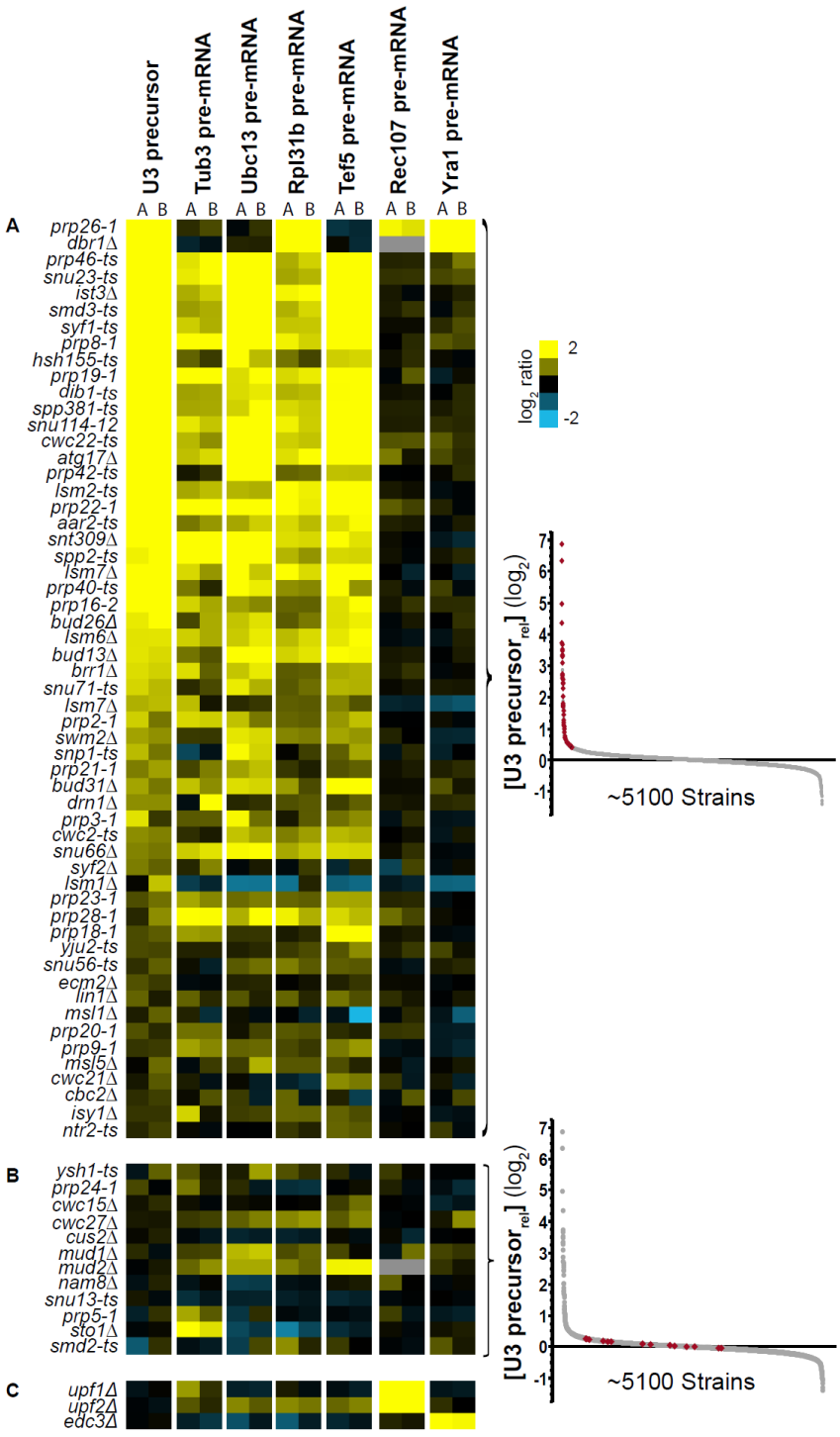

Figure 2.8: Mutations in most known splicing factors lead to increased precursor levels of canonical splicing substrates

Relative levels of the indicated RNAs are shown in the background of all strains

containing mutations in known splicing factors. The biological replicates (A and B) are shown for each RNA. Precursor levels for all transcripts are ordered based on the average expression values of U3 precursor, from high to low values. Gene disruptions are indicated on the left (-*ts* indicates a temperature sensitive allele). Insets to right indicate the location of the data in the U3 precursor dataset.

A. Mutations that lead to an increase in the U3 snoRNA.

B. Mutations that do not affect U3 precursor levels but may affect the levels of other intron containing genes.

C. The increase of Rec107 precursor levels seen in the *ufp1Δ* and *upf2Δ* backgrounds, and the increase of Yra1 precursor levels seen in the *edc3Δ* background are consistent with their well-characterized degradation pathways.

As expected, precursor accumulation can be detected for each of the canonical intron-containing transcripts in the background of nearly all of the splicing mutations. While all four canonical precursors accumulate in the *mud2Δ* background, consistent with the microarray data, no precursor accumulation is detected for any of these transcripts in the *cus2Δ* strain (Figure 2.8B). In addition, several of the splicing mutants that failed to cause an increase in the U3 precursor levels do cause a splicing defect for these other transcripts. Importantly, the behavior of the Rec107 and Yra1 pre-mRNAs within this subset of strains differs significantly from that seen for the canonically spliced transcripts. Splicing of the Rec107 pre-mRNA shows a strong accumulation in the *upf1Δ* and *upf2Δ* strains (Figure 2.8C), consistent with its known degradation via the nonsense-mediated decay pathway [35]. Because the Rec107 pre-mRNA does not engage the spliceosome during vegetative growth [21], no precursor accumulation is expected in strains containing spliceosomal mutations. Likewise, the Yra1 pre-mRNA shows a strong accumulation in the *edc3Δ* strain [34], consistent with its previously characterized cytoplasmic degradation pathway. The failure to detect Yra1 pre-mRNA

accumulation in strains containing spliceosomal mutations presumably reflects the inherently high levels of unspliced Yra1 transcript present in a wild type cell. Taken together, these data strongly support the capacity of this approach to successfully identify mutations that impact pre-mRNA splicing with low false positive and false negative rates of discovery.

2.3.4 Global splicing efficiency is impacted by many cellular mutations

To expand this analysis beyond previously characterized splicing factors, I sought to identify novel mutations that caused an increase in precursor levels in most, if not all, of the canonical intron-containing genes I analyzed. By determining the rank order of precursor accumulation in each strain for each of the five canonical splicing substrates (U3, Rpl31b, Tef5, Tub3, and Ubc13 precursors), a composite rank order of each strain was calculated as the average of these independent measurements (Figure 2.9).

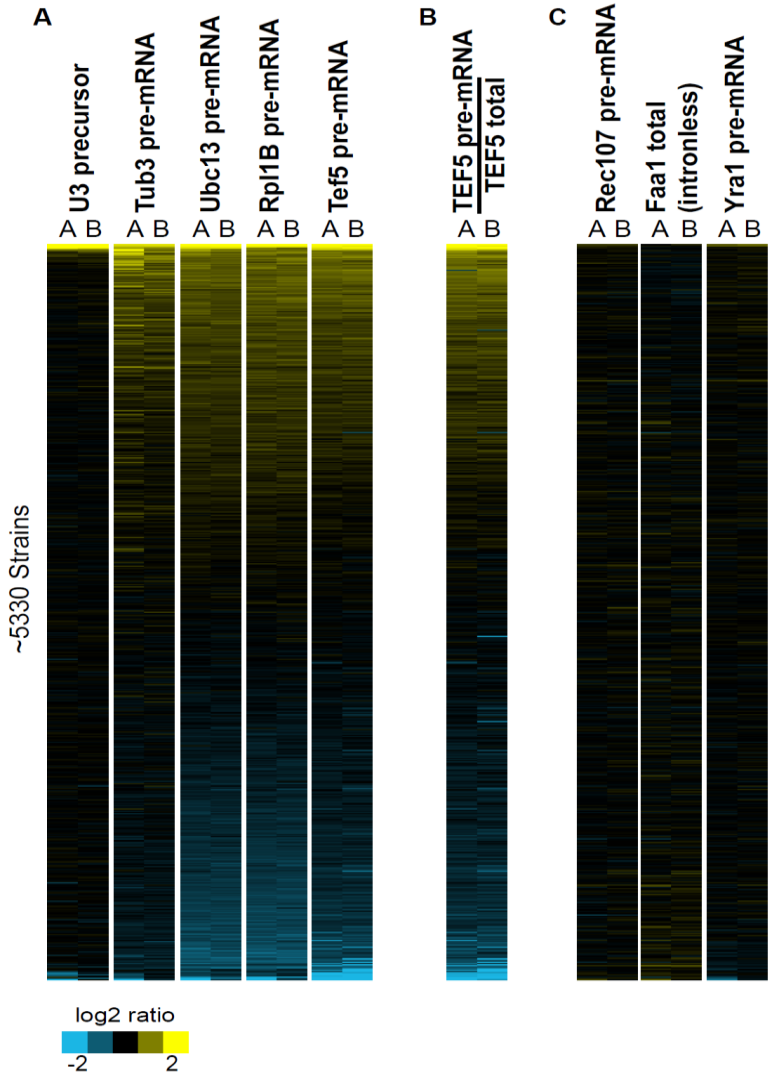

Figure 2.9*: *Increases in precursor levels correlate with a decrease in splicing efficiency

A. Relative levels of the indicated RNAs are shown in the background of 5334 strains. The data are ordered from high to low values based on a composite rank order, determined on the basis of the expression levels of the five canonical splicing substrates.

B. The relative splicing efficiency of the Tef5 transcript in each of these strains, calculated as the relative Tef5 precursor levels divided by the relative Tef5 total RNA levels.

C. The relative levels of the low abundance Rec107 pre-mRNA and Faa1 transcripts are shown along with the inefficiently spliced Yra1 pre-mRNA, none of which shows a similar pattern to the splicing defects seen in (A).

Remarkably, while the majority of the mutations examined cause little or no change in precursor levels of these four transcripts, the subset of mutations which do cause detectable increases in precursor levels is larger for some of the coding mRNAs than was seen for U3. Interestingly, although there is variation in the number of strains that cause pre-mRNA accumulation of the different transcripts, with Tub3<Tef5<Ubc13~Rpl31b, strong overlap can nevertheless be identified across the four transcripts. For example, the majority of the strains that cause an increase in the Tub3 pre-mRNA also display an increase in the pre-mRNA levels of the other three transcripts. By contrast, many strains cause a strong accumulation of the Ubc13 and Rpl31b pre-mRNAs without causing a significant change in the Tub3 or Tef5 pre-mRNA levels.

Because the absolute levels of the Rpl31b, Tef5, Tub3 and Ubc13 pre-mRNAs are significantly lower than the U3 precursor levels in most strain backgrounds (Table 2.3), I considered the possibility that these results reflected a technical artifact associated with measuring the cellular levels of low abundance RNA species in certain strain backgrounds. Importantly, however, the relative levels of the Rec107 pre-mRNA, whose normal cellular level is similar to these other pre-mRNAs, is largely unchanged in the vast majority of the strains examined (Figure 2.9). Likewise, an analysis of the cellular levels of the Faa1 mRNA, an intronless gene whose transcript abundance is of a similar magnitude as the Rpl31b, Tef5, Tub3 and Ubc13 pre-mRNAs, also shows a nearly constant level in all of the examined strains, further suggesting that there is no inherent bias in detecting low level transcripts. Finally, the Yra1 pre-mRNA, which is inefficiently spliced and has a higher endogenous level than

most pre-mRNAs, also shows very little change in the examined strains. Taken together, these results strongly support the conclusion that the levels of the Rpl31b, Tef5, Tub3 and Ubc13 pre-mRNAs are increased in these strains.

2.3.5 Increases in precursor levels correlate with a decrease in splicing efficiency

Because this approach, as described so far, directly measures the cellular levels of precursor RNA but does not directly determine the efficiency of splicing *per se*, those mutations which cause an increase in the precursor levels could be doing so simply by increasing the transcriptional frequency of these genes rather than by directly impacting their splicing. To distinguish this possibility from a true splicing defect, I chose to directly calculate the splicing efficiency of the Tef5 transcript by measuring the total cellular level of Tef5 mRNA by QPCR in each strain and using this value to calculate the ratio of unspliced:spliced RNA in the cell, a classical measure of splicing efficiency. Consistent with a splicing rather than transcriptional cause for precursor accumulation, the measured levels of total Tef5 transcript showed little variation across nearly the entire set of strains (Figure 2.10). Indeed, nearly every strain that showed an increase in Tef5 pre-mRNA levels also showed a decrease in the splicing efficiency of the Tef5 transcript (Figure 2.9), suggesting that those mutations affect the splicing of this transcript rather than its transcription. These results strongly suggest that the increased pre-mRNA levels observed in these strains largely reflect changes in pre-mRNA splicing.

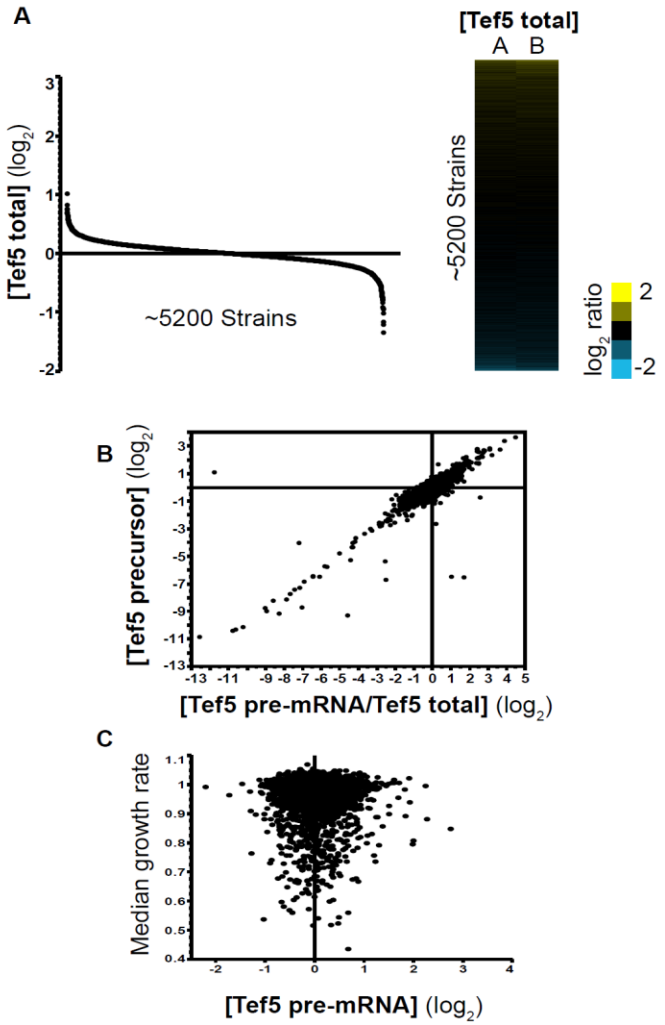

Figure 2.10*: *Global changes in Tef5 RNA levels

A. Total RNA levels of the Tef5 transcript in 5198 strains which passed the quality filters. The total RNA levels are presented in $\log_2$ space as a composite behavior of both biological replicates, and are ordered from the highest (left) to the lowest (right) values. On the right side of the figure the data are presented as a heat map, with both biological datasets (A and B) shown. The data in the heat map are ordered from the highest to the lowest values, similar to the representation on the left.

B. A comparison of the levels of Tef5 pre-mRNA versus the splicing efficiency of this transcript (calculated as a ratio of precursor Tef5 to total Tef5 levels across the entire dataset) shows a strong correlation.

C. A comparison of the relative growth rate of the non-essential library strains [36] versus the Tef5 pre-mRNA levels reveals no correlation between cellular fitness and splicing efficiency.

To assess the functional significance of the strains displaying increased pre-mRNA levels, I sought to rule out the possibility that mutations which cause a change in overall cellular fitness might indirectly lead to a decrease in overall splicing efficiency. To test this, I compared our precursor accumulation levels with recently described strain fitness data calculated for each of the ~5000 non-essential genes [36]. This comparison yielded no correlation between precursor accumulation and cellular fitness (Figure 2.10C), suggesting that cellular growth rate alone is insufficient to explain the observed increase in pre-mRNA levels.

2.3.6 Assessing the statistical significance of the precursor accumulation

While the precursor accumulation seen for each of the canonical transcripts in the known splicing mutants lends strong empirical support for the overall robustness of this approach, additional analysis was needed to assess the statistical significance of the data. Towards this end, I employed a statistical approach originally developed for analysis of microarray data called Significance Analysis of Microarrays [37], or SAM (see also Materials and Methods). I chose this software because, conceptually, the data generated by the QPCR approach are orthogonal to those from a microarray experiment: whereas a microarray experiment examines the behavior of thousands of mRNAs in a single strain, here it examines the behavior of a single RNA in thousands of different strains. Because similar concerns regarding multiple hypothesis testing apply to both types of data [38], I used this software as a tool for assessing the quality of the data. The results of the SAM analysis were consistent with the qualitative results seen in Figure 2.9, in so much as the number of strains causing a statistically

significant increase in the levels of each precursor species varied depending upon the precursor mRNA in question. A total of 224 strains caused a statistically significant increase in the Rpl31b pre-mRNA levels, 209 strains caused a significant increase in Ubc13 pre-mRNA levels, 146 strains caused a significant increase in U3 precursor levels, 83 strains caused a significant increase in Tef5 pre-mRNA levels, and 78 strains caused a significant increase in Tub3 pre-mRNA levels. Importantly, many of the SAM-identified strains are found to cause a significant enrichment of the precursor levels of all five of these RNAs, including the majority of strains with mutations in canonical splicing factors.

Interestingly, for some of the species examined, a small number of strains were identified which showed *decreased* levels of precursor RNA. In certain instances these reflected expected outcomes: a large decrease in the Ubc13 precursor was identified in the *ubc13Δ* strain, for example. However, in other cases these may indicate important biological phenomena. For example, both the *xrn1Δ* and the *tfg2Δ* strains cause a significant decrease in the U3 precursor levels. It has been previously shown that deletion of the Xrn1 nuclease paradoxically leads to decreases in many precursor RNAs [39], although the mechanism by which this occurs remains unknown. Likewise, it is unclear whether the decreased precursor resulting from deletion of the TFIIF component Tfg2 reflects an overall decrease in transcription of this gene, or whether this in fact reflects increased splicing efficiency perhaps resulting from a decreased transcription elongation rate [10].

2.3.7 Top screen candidates predict specific links between splicing and several pre-mRNA processing pathways.

To better characterize the factors that impact pre-mRNA splicing, we examined the lists of SAM-identified candidates for factors that are not canonical components of the spliceosome. As an initial approach, I asked whether any functional categories of proteins were statistically overrepresented within this set of strains. For this analysis, I ordered the strains according to the largest precursor accumulation that they affected for any of the RNA species. I then used the *GO::Term Finder* program [40] to identify overrepresented classes of genes. As expected, when considering the 50 strains that caused the largest precursor accumulations, a strong enrichment for splicing factors was seen with 30 out of 50 strains containing mutations in genes belonging to the GO PROCESS: *RNA Splicing* category (p=1.3E-40 with Bonferroni correction). Interestingly, when the top 100 strains are considered, significant enrichment can also be seen for strains with mutations in factors belonging to the GO PROCESS: *Chromatin Remodeling* category, with eight different mutants causing precursor accumulation (*arp5Δ*, *arp8Δ*, *bdf1Δ*, *npl6Δ*, *rsc2Δ*, *rsc9-ts*, *vps72Δ*, and *yaf9Δ*; p=1.5E-03). Expanding the analysis to the top 200 candidates increases the enrichment of this category to include twelve factors (adding *arp6Δ*, *swc5Δ*, *swr1Δ*, and *taf14Δ*; p=1.8E-04). Interestingly, within the top 200 candidates, significant enrichment is also seen for the GO PROCESS: *RNA Catabolic Process* category, with 13 different factors being present (*ccr4Δ*, *dis3-ts*, *dbr1Δ*, *kem1Δ*, *lsm2-ts*, *lsm6Δ*, *lsm7Δ*, *prp18Δ*, *rrp6Δ*, *rtt101Δ*, *ski3Δ*, *ssn2Δ*, and *upf3Δ*; p=8.5E-03). Whereas some of these factors, such as *lsm2-ts*, *lsm6Δ*, *lsm7Δ*, and *prp18Δ* are known to directly function in pre-

mRNA splicing, the identification of many of these factors presumably reflects their defects in degradation pathways for unspliced pre-mRNAs.

One of the top factors I identified that bridges chromatin remodeling with transcription initiation is the bromodomain factor Bdf1. Bdf1 is a member of the SWR1 complex and, along with its homolog Bdf2, has been shown to interact with the TFIID component of RNA polymerase II [41]. Moreover, *BDF1* and *BDF2* have been demonstrated to be genetically redundant with one another. Whereas the SAM analysis indicated that the *bdf1Δ* caused a statistically significant accumulation of most of the canonical precursor species in our experiments, the *bdf2Δ* strain showed little or no detectable increase in the levels of any of the precursors tested (Figure 2.11), and was not considered by SAM analysis to be significant for increases in any of the precursor RNAs. To better characterize the global splicing profile of these two mutants, we again turned to splicing-sensitive microarrays. Remarkably, a dramatic splicing defect can be seen in the *bdf1Δ* strain for most intron-containing genes, as evidenced by an increase in the precursor transcript levels with a concomitant decrease in the mature and total transcript levels (Figure 2.11A). By comparison, the *bdf2Δ* mutation has almost no effect on cellular splicing, strongly corroborating the specific identification of Bdf1 in our screen.

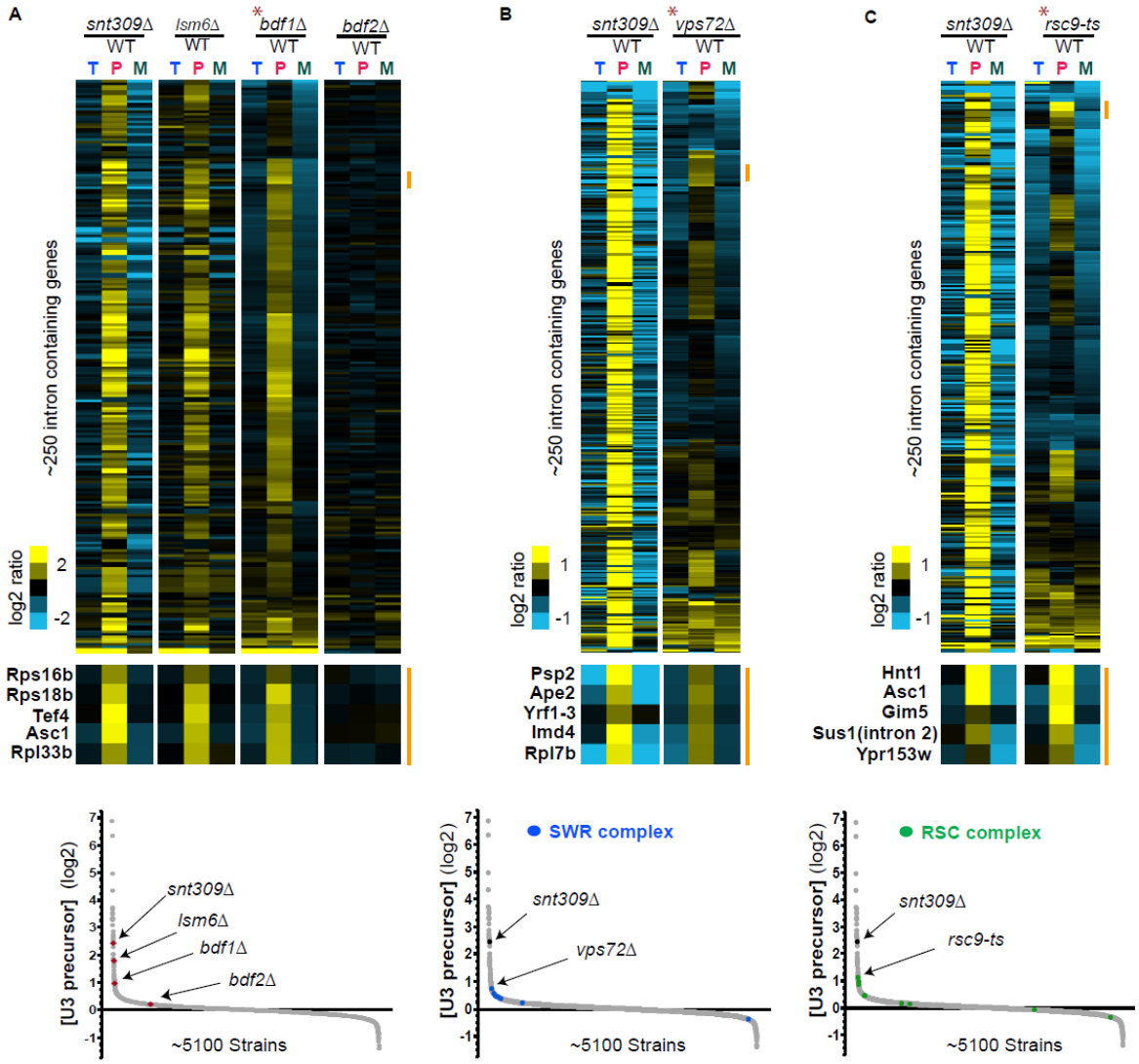

Figure 2.11*: *Mutations in chromatin remodeling factors can impact splicing in a global or transcript-specific manner

Splicing sensitive microarrays of candidate factors (A) *bdf1Δ* and *bdf2Δ*, (B) *vps72Δ*, and (C) *rsc9-ts*, as compared to the phenotypes of known splicing mutants *snt309Δ* and *lsm6Δ*. For each panel, the asterisk indicates the strain by which the data have been organized using hierarchical clustering, with the other data sets in those panels sharing an identical gene ordering. The orange bar highlights the location of specific subsets of transcripts showing splicing defects. The bottom insets show the location of each of the candidates within the U3 precursor dataset. Also, the locations of each of the components of the SWR complex are shown in blue, and each of the components of the RSC complex are shown in green.

To better assess the mechanism by which Bdf1 impacts pre-mRNA splicing, we monitored U1 snRNP recruitment in the background of wild-type, *bdf1Δ*, and *bdf2Δ* strains using chromatin immunoprecipitation coupled to QPCR (ChIP-QPCR). As seen in Figure 2.12, these experiments show that the deletion of Bdf1 but not of Bdf2 decreases the occupancy of U1snRNP at several intron-containing genes, suggesting impairment of co-transcriptional spliceosomal recruitment in the *bdf1Δ* strain. A more comprehensive ChIP-Seq experiment will be required to fully characterize the global landscape of genes impacted by the deletion of Bdf1 and further characterize the roles of Bdf1 and Bdf2 in transcription and splicing.

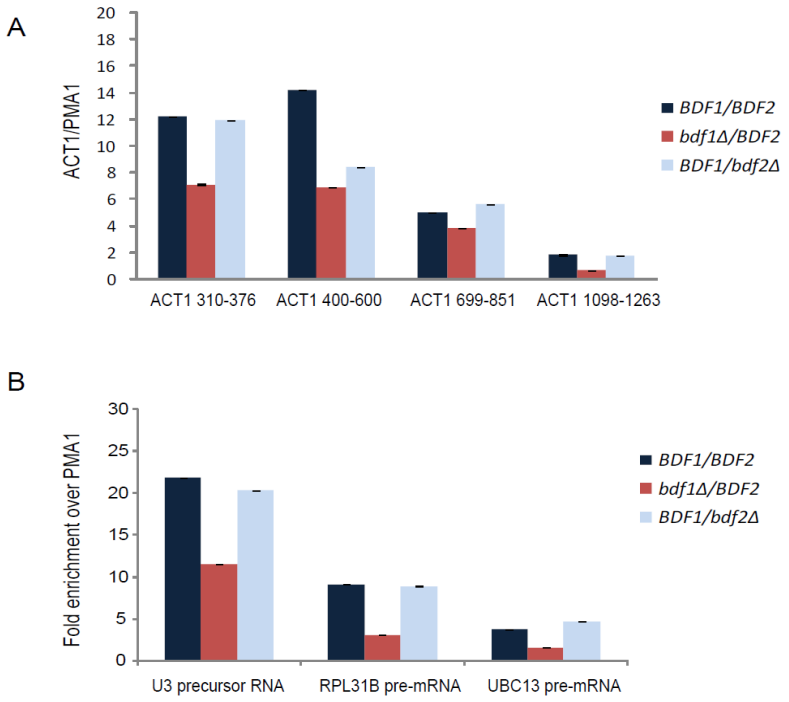

Figure 2.12*: *U1 snRNP recruitment is diminished upon Bdf1 deletion

Chromatin immunoprecipitation (ChIP) was performed using a Tap tagged version of Yhc1 (U1C) in wild type, *bdf1Δ*, and *bdf2Δ* strains to assess the co-transcriptional occupancy of the U1 snRNP. A) Primers that had been previously used in a similar

assay [5] allowed us to monitor by quantitative PCR the amount of U1 snRNP associated with different genomic regions of the ACT1 gene. The plotted values were calculated as the percent of input signal detected at given region within the actin gene divided by the percent of input signal observed for the intronless gene PMA1. The error bars represent the standard deviations of technical replicates. B) The ChIP samples described above were assayed with the same primers used in our screening which targeted intronic regions of the U3 snoRNA, Rpl31B and Ubc13 pre-mRNAs. As above, the values are presented as fold enrichment over the intronless gene PMA1 and the error bars are indicate standard deviation of technical replicates. For all four intron-containing genes, decreased levels of U1 snRNP are detected in the *bdf1Δ* strain relative to both the wild type and *bdf2Δ* strains.

I also chose to further examine several factors that were identified in the screen that are more classically connected with chromatin remodeling. The lower panels of Figure 2.11B and 2.11C show the locations within the U3 precursor dataset of all of the strains containing mutations in components of the SWR1 complex and the RSC complex respectively. Notably, mutations in many but not all of the components of these complexes cause a splicing defect of the U3 transcript. Moreover, each of the five precursor species that I examined shows a slightly different susceptibility to the different components of these complexes. I chose to examine the global splicing defects of strains containing mutations in two of these components: Vps72, a member of the SWR1 complex; and Rsc9, a member of the RSC complex. Splicing-sensitive microarrays of the *vps72Δ* and *rsc9-ts* strains, respectively, reveal a splicing defect in each strain (Figures 2.11B and 2.11C). However, unlike the *bdf1Δ* strain, the *vps72Δ* and *rsc9-ts* strains cause a splicing defect in only distinct subsets of intron-containing genes. Interestingly, the affected subsets of transcripts are neither completely overlapping nor completely independent of one another; rather the microarray data are consistent with the QPCR data in suggesting that mutations in specific chromatin-

modifying components can result in aberrant splicing of specific pre-mRNA transcripts.

While an ontology-based approach can successfully identify entire pathways that display enrichment, I was also interested in considering those factors which showed strong pre-mRNA accumulation but whose functional categories were not statistically over-represented at the top of our dataset. Remarkably, while the GO PROCESS: *RNA 3' end Processing* wasn't significantly overrepresented as a category within our dataset (9 out of top 200, p=0.09), several strains with mutations in factors belonging to this category resulted in a strong, statistically-significant accumulation of multiple precursor species. Included among these were: *yth1-ts*, a zinc-finger containing protein that is the homolog of human CPSF-30; *cft2-ts*, the homolog of human CPSF-100; and *fip1-ts*, a component of the polyadenylation factor PF I. To further examine the global splicing defects of each of these mutants, microarrays were performed comparing mutant and wild type behavior after shifting them to both elevated and reduced temperatures. Of the three mutants, the profile seen in the *yth1-ts* mutation most closely resembles a canonical splicing defect, with more than half of the genes showing an increase of precursor and loss of mature RNA (Figure 2.13A).

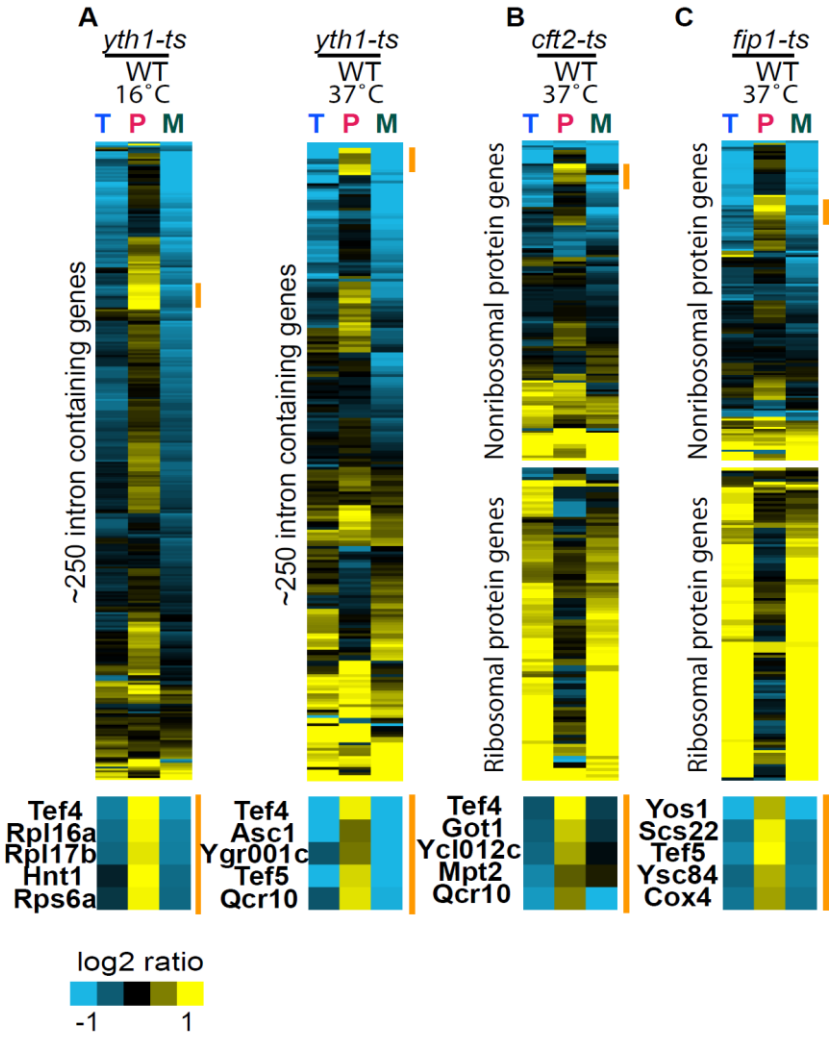

Figure 2.13: Mutations in 3' end processing mutants result in transcript-specific splicing defects

Splicing sensitive microarrays for the cleavage and polyadenylation factor mutants *yth1-ts* (A), *cft2-ts* (B), and *fip1-ts* (C). For each panel, the data have been independently organized using hierarchical clustering. The orange bar highlights the location of specific subsets of transcripts showing splicing defects.

Interestingly, the splicing defect is strongest at reduced temperatures even though this strain has only a subtle low-temperature growth defect (Figure 2.14).

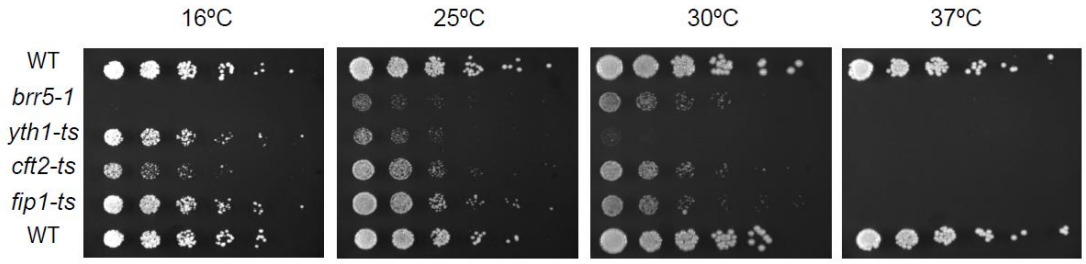

Figure 2.14: Growth phenotypes for 3'end mutants

Serial four-fold dilutions of temperature sensitive alleles at various temperatures. The *brr5-1* allele exhibits the well known cold-sensitive phenotype, as well as a growth defect at 37°C.The temperature sensitive alleles *yth1-ts, cft2-ts and fip1-ts* all show growth defects at 37°C, while displaying slower growth at 30°C. The most pronounced growth defect at 30°C as well as at the permissive temperature is seen for *yth1-ts*.

By comparison, neither the *cft2-ts* nor the *fip1-ts* strains showed a strong splicing defect at low temperature (not shown), but each mutant was characterized by an unusual phenotype at elevated temperatures. As seen in Figures 2.13B and 2.13C, two distinct types of behavior are seen in the *cft2-ts* and *fip1-ts* mutants, respectively, that are largely defined by whether or not the affected transcript encodes a ribosomal protein gene (RPG). For a subset of the non-RPG transcripts a canonical splicing defect is apparent, consistent with our QPCR results. Interestingly, the subset of affected non-RPG transcripts is different between the two mutant strains. By comparison, nearly all of the RPG transcripts show a dramatic increase in both the mature and total mRNA levels, with little or no detectable change in precursor levels. The strong increases caused by these mutants suggest that the RPG transcripts may be subject to regulatory control at their 3’ ends. Interestingly, while it has long been known that RPG introns are, in general, longer than non-RPG introns [42], whereas

the second exons of RPGs tend to be shorter than non-RPGs [5], I nevertheless find no strong correlation between either intron or second exon length and the strength of the splicing defect seen for these 3' end mutants (data not shown). The mechanisms by which these 3' end factors impact pre-mRNA splicing are currently under investigation.

2.3.8 Top screen candidates do not cause changes in the cellular mRNA levels of most spliceosomal components

In considering the mechanisms by which candidate factors may be functioning, I sought to determine whether any of the candidates we examined might be indirectly affecting pre-mRNA splicing by changing the cellular levels of known spliceosomal components. Although the splicing-sensitive microarrays were designed primarily to interrogate the splicing status of the ~300 intron-containing genes in *S. cerevisiae*, they also contain probes against all ~6000 protein-coding genes and ~200 RNA genes, including the spliceosomal snRNAs. Figure 2.15 shows the relative RNA levels for each of the canonical spliceosomal components, including the snRNAs, in the background of each of the different strains we examined, as determined from our microarray analyses. While these results positively recapitulate the expected changes (for example, the decreases in Snt309 and Lsm6 mRNA levels in the *snt309Δ* and *lsm6Δ* strains, respectively), with only a few exceptions, most spliceosomal components appear unchanged in most of the mutants we examined.

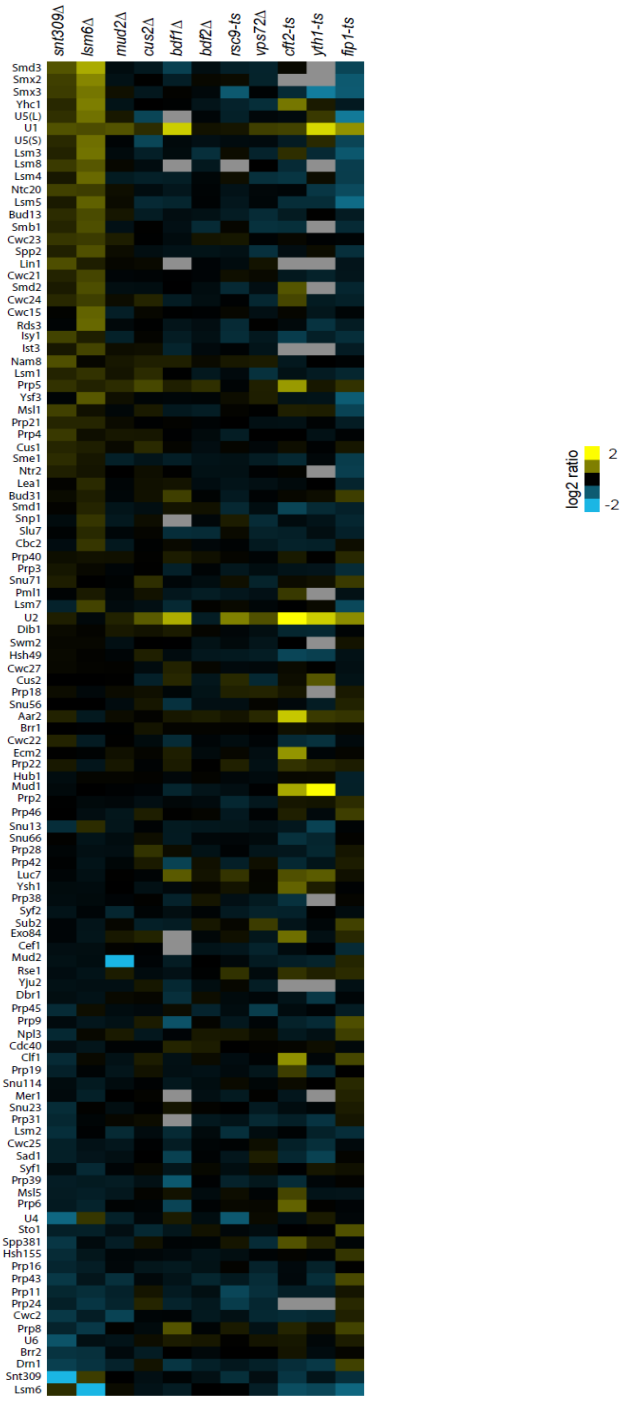

Figure 2.15: The RNA levels of most spliceosomal factors are unchanged in most mutants

Total RNA levels for all splicing factors in the background of different gene deletions or point mutations for which microarrays were performed. The data are organized on the basis of the highest to the lowest average change in the *snt309Δ* and *lsm6Δ* strains.

Importantly, the transcript encoding the Mud1 protein showed dramatic mis-regulation in both the *yth1-ts* and *cft2-ts* strains, increasing by more than 10-fold in each background. To test whether Mud1 overexpression might be causing the splicing defects observed in these strains, a strain was constructed where the wild type Mud1 transcript was encoded on a high-copy plasmid. As seen in Figure 2.16, in spite of the over 30-fold increase in Mud1 levels in this strain, there is no detectable change in pre-mRNA splicing. Therefore, while the mis-regulation of Mud1 levels in these 3' end mutants suggests that, similar to its human homolog, Mud1 levels in yeast may be subject to negative regulation via its 3' end processing [43], it nevertheless appears that the splicing defect observed in these strains is not a consequence of Mud1 overexpression.

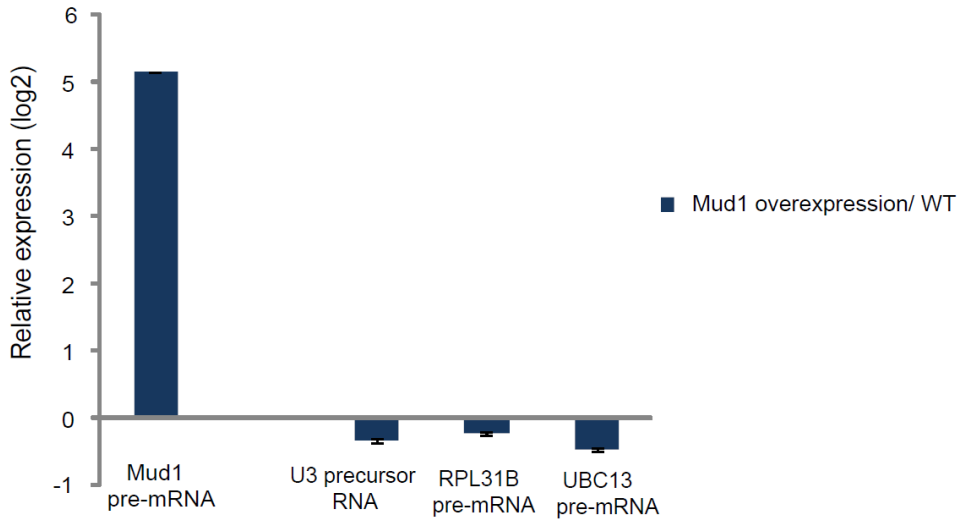

Figure 2.16*: *Mud 1 overexpression does not cause increases in precursor levels

A high copy plasmid (2-micron) containing the Mud1 gene was transformed into an otherwise wild type strain in order to affect its overexpression. The expression levels of Mud1 and several precursor RNAs were monitored by quantitative real-time PCR and compared to a wild type strain containing an empty vector. The data were

normalized to the expression of the intronless Faa1 transcript to account for loading differences in the samples. While a >30-fold increase is apparent for the Mud1 transcript, no increase is detected in the precursor levels of any of the RNAs surveyed, suggesting that their splicing is unaffected by Mud1 overexpression. The error bars represent the standard deviation of technical replicates.

Interestingly, several of the strains, including *bdf1Δ*, *yth1-ts*, *cft2-ts*, and *fip1-ts*, showed an ~2-fold *increase* in the levels of both the U1 and U2 snRNAs. Although spliceosomes function as an equimolar complex of all five snRNAs, the total cellular levels of the snRNAs vary: in yeast, the U2 snRNA is the most abundant [44], while in mammals the U1 snRNA is most abundant [45]. While recent work demonstrates the cellular defects associated with decreased levels of snRNA [46], it is less clear whether increases in their levels will impart a defect on global splicing. Nevertheless, because each of these strains shows a similar increase in these snRNA levels but distinct splicing defects, it seems unlikely that the changes in snRNA levels alone can explain the observed splicing phenotypes. However, it is not inconceivable that small changes in levels for one or more of these transcripts could lead to the observed splicing defects. As such, additional work will be necessary to determine the functional consequences of these mutations.

2.4 Discussion

2.4.1 A high-throughput, reverse genetic approach to measure the cellular levels of specific RNA species

Here I present the results of a global survey designed to identify the full subset of cellular factors in the budding yeast, *Saccharomyces cerevisiae*, that impact the efficiency of pre-mRNA splicing. As a complement to other recently described genetic and physical genome-wide approaches, in this work I have developed an approach that allows for a direct readout of the accumulation of specific RNA species in the background of thousands of different mutant strains. An important strength of a genome-wide screen such as this is its unbiased approach. By directly measuring the splicing efficiency of endogenous transcripts, this method avoids bias generated using reporter constructs. Moreover, the ability to examine numerous different transcripts allowed to distinguish the natural variation in the spliceosomal factors that are required for efficient splicing of different intron-containing transcripts. Indeed, by systematically examining the precursor levels in the background of each strain, mutations can be identified which result in a change in splicing efficiency regardless of their previously described functions. In the work described here, mutations in scores of genes with no previously known role in splicing were identified, some of which impacted the splicing of all five canonical transcripts examined and some of which impacted only a subset of them. While some of these factors have been further characterized and discussed here, many have not. To be sure, as is the case with all genetic screens, it is impossible on the basis of these screen data alone to ascribe a direct role for any of these candidate factors in the splicing pathway. Rather, the

identification of these different factors can be seen as generating a rich dataset from which hypotheses can be generated and tested for their mechanistic underpinnings.

2.4.2 The 5' end: connecting splicing with chromatin remodeling and transcription initiation

Beyond known splicing factors, the most highly over-represented set of factors identified in this work function in chromatin remodeling. One particularly interesting mutation that was identified was the *bdf1Δ* mutant. In budding yeast, Bdf1 has been demonstrated to play a role precisely at the interface of transcription initiation and chromatin remodeling. Based in part on its physical interaction with the Taf7 subunit of TFIID, yeast Bdf1 has been proposed to function as the missing C-terminal portion of the higher eukaryotic $TAF_{II}250$ [41], the largest subunit of the TFIID complex. More recently, it has become clear that Bdf1 interacts with Swr1 and functions in recruiting the entire SWR1 chromatin remodeling complex to nucleosomes. A recent genome-wide study demonstrates that Bdf1 is enriched on the +1 and +2 nucleosomes of actively transcribed genes [47], and that it coincides with the localization of Vps72, another component of the SWR1 complex, and another component which was identified in our screen (Figure 2.11).

Remarkably, the splicing of nearly every intron-containing gene is negatively affected in a *bdf1Δ* strain, and the quantitative defect seen in this mutant rivals that seen for canonical splicing mutants. Given its role in global gene expression, one possible explanation for our results in the *bdf1Δ* strain is that the transcription of some key

splicing factor is repressed by this mutation, causing a decrease in splicing efficiency. Indeed, early work on Bdf1 from the Séraphin lab suggested a role in global transcription, including transcription of the spliceosomal snRNA genes [48]. However, my microarray analyses show essentially normal RNA levels of all known splicing components in the *bdf1Δ* strain (see Figure 2.15). Moreover, the microarray data assessing the snRNA levels themselves are entirely consistent with Séraphin's original observations and demonstrate that none of the five wild type snRNAs are decreased in cellular level during growth at 30°C in the *bdf1Δ* mutant; rather, there are subtle increases in the U1 and U2 snRNA levels. Importantly, the ChIP-QPCR experiments in the *bdf1Δ* strain demonstrate a decreased occupancy of the U1 snRNP on all four intron-containing genes that were tested, suggesting the intriguing possibility that Bdf1 plays a direct role in connecting pre-mRNA splicing with chromatin remodeling and transcription initiation.

In considering such a role for Bdf1, it is important to note that the yeast *BDF1* gene has a close sequence homolog in the *BDF2* gene. These two genes are genetically redundant, in so much as both single gene deletions are viable but the double mutant *bdf1Δ/bdf2Δ* is lethal. Moreover, it has been shown that these two genes evolved from a single ancestral gene following a whole-genome duplication event [49]. Yet surprisingly, unlike the *bdf1Δ* strain, the *bdf2Δ* strain showed no signs of a splicing defect either in the screen or when examined by splicing sensitive microarrays. Moreover, unlike the *bdf1Δ* strain, there was no apparent decrease in U1 snRNP ChIP-QPCR signal in the *bdf2Δ* strain. In considering a mechanism whereby Bdf1 connects transcription initiation and chromatin remodeling with pre-mRNA splicing, it is worth

noting that, unlike human genes, the majority of yeast genes do not contain an intron. As such, co-transcriptional recruitment of the spliceosome is unnecessary for most yeast genes. I am considering the possibility that, in the time since the duplication event, Bdf2 has evolved to a point where it retains the capacity to recruit RNA polymerase but has lost the ability to efficiently connect splicing with transcription. Such a scenario would explain the differences observed between the *bdf1Δ* and *bdf2Δ* microarrays and U1 snRNP ChIP-QPCR data. It would also likely explain the previously published results that Bdf1 shows higher sequence conservation with the C-terminal domain of human $TAF_{II}250$ than does Bdf2 [41]. Given such a model for the divergence of Bdf1 and Bdf2 functions, the differences in protein sequence between these two proteins may prove informative for deciphering the mechanism of Bdf1 activity.

2.4.3 The 3' end: connecting splicing with cleavage and polyadenylation

In addition to the over-representation of factors marking the 5' end of genes, the screen identified a number of factors involved in the 3' end processing of mRNAs. Splicing-sensitive microarrays confirm a broad splicing defect in a mutant of Yth1, the homolog of human CPSF30, and transcript-specific splicing defects in mutants of Cft2 and Fip 1, the homolog of human CPSF100 and a component of the polyadenylation factor complex PF I, respectively. In higher eukaryotes, components of the 3' end processing machinery have been shown to physically associate with components of the U2 snRNP [50] and U2AF65 [51]; moreover, *in vitro* studies demonstrate a functional link between the pre-mRNA splicing and 3' end processing pathways [52]. The

interactions between these two pathways in mammalian systems have led to the proposal that the 3' end machinery plays an important role in terminal exon definition. Whereas the exon-definition model for mammalian spliceosome assembly posits that internal exons are defined by interactions between U1 and U2 snRNP components across an exon [53], definition of terminal exons is achieved by interactions between the 3' end processing machinery and the U2 snRNP (Figure 2.17), imposing a functional connection between the pathways. Yet because of the relatively short length of *S. cerevisiae* introns and the limited number of genes that are interrupted by multiple introns, splicing in yeast has long been considered to proceed through a model of intron-definition. Nevertheless, the Keller lab recently demonstrated that some conditional alleles of *YSH1*/*BRR5* lead to a decrease in splicing efficiency [54]. By identifying additional pre-mRNA splicing defects in the background of other mutants in 3' end processing suggests the intriguing possibility that some of the basic interactions that facilitate exon-definition in higher systems may also be present in budding yeast. Indeed, further characterizing the mechanism by which these 3' end processing factors are affecting splicing in yeast may provide important insights into the mechanisms by which exon-definition is accomplished in higher eukaryotes.

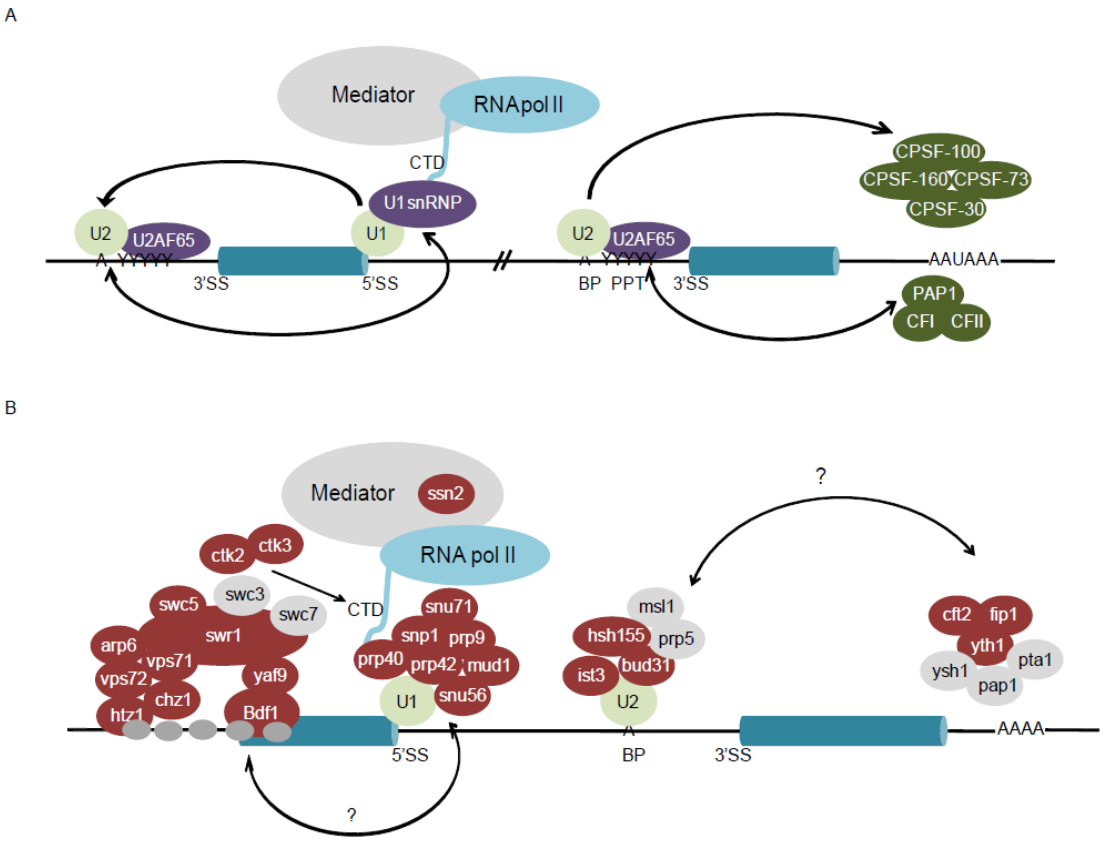

Figure 2.17: The location of mutants that disrupt splicing suggests a similarity to exon-definition in mammalian splicing

A. Spliceosome assembly in mammalian systems is thought to occur via exon-definition, where interactions between U1 and U2 snRNP components across an internal exon, or CPSF components and U2 snRNP across a terminal exon, are necessary for efficient processing of the upstream intron.

B. Functional location of a subset of the mutants examined in this work, presented on a model transcript. Factors shown in red caused a statistically significant increase in the precursor levels of at least one transcript in our screen, whereas those in grey did not affect the splicing of any of the transcripts.

2.4.4 A global perspective

An important strength of an approach such as this is the genome-wide perspective that it provides. Figure 2.17B shows a model of an idealized transcript along with the

functional location of a subset of the factors that have been examined in the screen. It is striking to note that many of the factors identified here function both during transcription initiation (Bdf1 and others) and termination (Yth1, Cft2, Fip1, and others), thereby defining the beginning and ends of the first and last exons, respectively. In this work, I have identified not only those factors whose disruption leads to a functional defect in splicing efficiency, but in many cases the specific transcripts whose splicing is affected.

More broadly, the work presented here demonstrates the feasibility of quantitating cellular RNA levels in the background of large mutant strain collections. While the current approach examined splicing efficiency in the context of optimized growth conditions, a similar approach could be applied to identify factors necessary for efficient splicing under varying cellular or developmental growth states. Likewise, although my work focused on the levels of several pre-mRNA species, this methodology should be directly applicable to assessing the levels of nearly any cellular RNA of interest.

REFERENCES

1. Wahl MC, Will CL, Lührmann R (2009) The Spliceosome: Design Principles of a Dynamic RNP Machine. Cell 136: 701-718. doi:10.1016/j.cell.2009.02.009

2. Hartwell LH, McLaughlin CS, Warner JR (1970) Identification of ten genes that control ribosome formation in yeast. Molecular and General Genetics MGG 109: 42-56. doi:10.1007/BF00334045

3. Kotovic KM, Lockshon D, Boric L, Neugebauer KM (2003) Cotranscriptional recruitment of the U1 snRNP to intron-containing genes in yeast. Mol. Cell. Biol 23: 5768-5779.

4. Moore MJ, Schwartzfarb EM, Silver PA, Yu MC (2006) Differential recruitment of the splicing machinery during transcription predicts genome-wide patterns of mRNA splicing. Mol. Cell 24: 903-915. doi:10.1016/j.molcel.2006.12.006

5. Tardiff DF, Lacadie SA, Rosbash M (2006) A genome-wide analysis indicates that yeast pre-mRNA splicing is predominantly posttranscriptional. Mol. Cell 24: 917-929. doi:10.1016/j.molcel.2006.12.002

6. Kornblihtt AR, de la Mata M, Fededa JP, Munoz MJ, Nogues G (2004) Multiple links between transcription and splicing. RNA 10: 1489-1498. doi:10.1261/rna.7100104

7. Phatnani HP, Jones JC, Greenleaf AL (2004) Expanding the Functional Repertoire of CTD Kinase I and RNA Polymerase II: Novel PhosphoCTD-Associating Proteins in the Yeast Proteome†. Biochemistry 43: 15702-15719. doi:10.1021/bi048364h

8. Schwartz S, Meshorer E, Ast G (2009) Chromatin organization marks exon-intron structure. Nat. Struct. Mol. Biol 16: 990-995. doi:10.1038/nsmb.1659

9. Huff JT, Plocik AM, Guthrie C, Yamamoto KR (2010) Reciprocal intronic and exonic histone modification regions in humans. Nat. Struct. Mol. Biol. 17: 1495-1499. doi:10.1038/nsmb.1924

10. Kornblihtt AR (2007) Coupling transcription and alternative splicing. Adv. Exp. Med. Biol 623: 175-189.

11. Chanfreau G, Noble SM, Guthrie C (1996) Essential yeast protein with unexpected similarity to subunits of mammalian cleavage and polyadenylation specificity factor (CPSF). Science 274: 1511-1514.

12. Jenny A, Minvielle-Sebastia L, Preker PJ, Keller W (1996) Sequence similarity between the 73-kilodalton protein of mammalian CPSF and a subunit of yeast polyadenylation factor I. Science 274: 1514-1517.

13. Noble SM, Guthrie C (1996) Identification of novel genes required for yeast pre-mRNA splicing by means of cold-sensitive mutations. Genetics 143: 67-80.

14. Oesterreich FC, Preibisch S, Neugebauer KM (2010) Global Analysis of Nascent RNA Reveals Transcriptional Pausing in Terminal Exons. Molecular Cell 40: 571-581. doi:10.1016/j.molcel.2010.11.004

15. Tong AH, Evangelista M, Parsons AB, Xu H, Bader GD, et al. (2001) Systematic genetic analysis with ordered arrays of yeast deletion mutants. Science 294: 2364-2368. doi:10.1126/science.1065810

16. Schuldiner M, Collins SR, Thompson NJ, Denic V, Bhamidipati A, et al. (2005) Exploration of the function and organization of the yeast early secretory pathway through an epistatic miniarray profile. Cell 123: 507-519. doi:10.1016/j.cell.2005.08.031

17. Wilmes GM, Bergkessel M, Bandyopadhyay S, Shales M, Braberg H, et al. (2008) A genetic interaction map of RNA-processing factors reveals links between Sem1/Dss1-containing complexes and mRNA export and splicing. Mol. Cell 32: 735-746. doi:10.1016/j.molcel.2008.11.012

18. Costanzo M, Baryshnikova A, Bellay J, Kim Y, Spear ED, et al. (2010) The genetic landscape of a cell. Science 327: 425-431. doi:10.1126/science.1180823

19. Jurica MS, Moore MJ (2003) Pre-mRNA Splicing: Awash in a Sea of Proteins. Molecular Cell 12: 5-14. doi:10.1016/S1097-2765(03)00270-3

20. Bessonov S, Anokhina M, Will CL, Urlaub H, Luhrmann R (2008) Isolation of an active step I spliceosome and composition of its RNP core. Nature 452: 846-850. doi:10.1038/nature06842

21. Engebrecht JA, Voelkel-Meiman K, Roeder GS (1991) Meiosis-specific RNA splicing in yeast. Cell 66: 1257-1268.

22. Park JW, Parisky K, Celotto AM, Reenan RA, Graveley BR (2004) Identification of alternative splicing regulators by RNA interference in Drosophila. Proc. Natl. Acad. Sci. U.S.A 101: 15974-15979. doi:10.1073/pnas.0407004101

23. Pleiss JA, Whitworth GB, Bergkessel M, Guthrie C (2007) Transcript specificity in yeast pre-mRNA splicing revealed by mutations in core spliceosomal components. PLoS Biol 5: e90. doi:10.1371/journal.pbio.0050090

24. Winzeler EA, Shoemaker DD, Astromoff A, Liang H, Anderson K, et al. (1999) Functional characterization of the S. cerevisiae genome by gene deletion and parallel analysis. Science 285: 901-906.

25. Ben-Aroya S, Coombes C, Kwok T, O'Donnell KA, Boeke JD, et al. (2008) Toward a comprehensive temperature-sensitive mutant repository of the essential genes of Saccharomyces cerevisiae. Mol. Cell 30: 248-258. doi:10.1016/j.molcel.2008.02.021

26. Chomczynski P, Sacchi N (1987) Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Analytical Biochemistry 162: 156-159. doi:10.1016/0003-2697(87)90021-2

27. Myslinski E, Ségault V, Branlant C (1990) An intron in the genes for U3 small nucleolar RNAs of the yeast Saccharomyces cerevisiae. Science 247: 1213-1216.

28. Zavanelli MI, Ares M (1991) Efficient association of U2 snRNPs with pre-mRNA requires an essential U2 RNA structural element. Genes Dev 5: 2521-2533.

29. Liao XC, Tang J, Rosbash M (1993) An enhancer screen identifies a gene that encodes the yeast U1 snRNP A protein: implications for snRNP protein function in pre-mRNA splicing. Genes Dev 7: 419-428.

30. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, et al. (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat. Genet 25: 25-29. doi:10.1038/75556

31. Inada M, Pleiss JA (2010) Genome-Wide Approaches to Monitor Pre-mRNA Splicing. In: Guide to Yeast Genetics: Functional Genomics, Proteomics, and Other Systems Analysis. Academic Press, Vol. Volume 470. pp. 51-75. Available: http://www.sciencedirect.com/science/article/B7CV2-4YH5590-8/2/d7fcf2150257f447b70d92d5dd24567c. Accessed 4 Mar 2010.

32. Clark TA, Sugnet CW, Ares M (2002) Genomewide analysis of mRNA processing in yeast using splicing-specific microarrays. Science 296: 907-910. doi:10.1126/science.1069415

33. Preker PJ, Guthrie C (2006) Autoregulation of the mRNA export factor Yra1p requires inefficient splicing of its pre-mRNA. RNA 12: 994-1006. doi:10.1261/rna.6706

34. Dong S, Li C, Zenklusen D, Singer RH, Jacobson A, et al. (2007) YRA1 autoregulation requires nuclear export and cytoplasmic Edc3p-mediated degradation of its pre-mRNA. Mol. Cell 25: 559-573. doi:10.1016/j.molcel.2007.01.012

35. He F, Li X, Spatrick P, Casillo R, Dong S, et al. (2003) Genome-Wide Analysis of mRNAs Regulated by the Nonsense-Mediated and 5' to 3' mRNA Decay Pathways in Yeast. Molecular Cell 12: 1439-1452. doi:10.1016/S1097-2765(03)00446-5

36. Breslow DK, Cameron DM, Collins SR, Schuldiner M, Stewart-Ornstein J, et al. (2008) A comprehensive strategy enabling high-resolution functional analysis of the yeast genome. Nat. Methods 5: 711-718. doi:10.1038/nmeth.1234

37. Tusher VG, Tibshirani R, Chu G (2001) Significance analysis of microarrays applied to the ionizing radiation response. Proc. Natl. Acad. Sci. U.S.A 98: 5116-5121. doi:10.1073/pnas.091062498

38. Miller RG (1981) Simultaneous statistical inference. New York: Springer-Verlag. 299 p.

39. Pleiss JA, Whitworth GB, Bergkessel M, Guthrie C (2007) Rapid, transcript-specific changes in splicing in response to environmental stress. Mol. Cell 27: 928-937. doi:10.1016/j.molcel.2007.07.018

40. Boyle E, Weng S, Gollub J, Jin H, Botstein D, et al. (2004) GO::TermFinder--open source software for accessing Gene Ontology information and finding significantly enriched Gene Ontology terms associated with a list of genes. Bioinformatics 20: 3710-3715.

41. Matangkasombut O, Buratowski RM, Swilling NW, Buratowski S (2000) Bromodomain factor 1 corresponds to a missing piece of yeast TFIID. Genes Dev 14: 951-962.

42. Spingola M, Grate L, Haussler D, Ares M Jr (1999) Genome-wide bioinformatic and molecular analysis of introns in Saccharomyces cerevisiae. RNA 5: 221-234.

43. Boelens WC, Jansen EJ, van Venrooij WJ, Stripecke R, Mattaj IW, et al. (1993) The human U1 snRNP-specific U1A protein inhibits polyadenylation of its own pre-mRNA. Cell 72: 881-892.

44. Brenner TJ, Guthrie C (2006) Assembly of Snu114 into U5 snRNP requires Prp8 and a functional GTPase domain. RNA 12: 862 -871. doi:10.1261/rna.2319806

45. Baserga SJ, Steitz JA (1993) The Diverse World of Small Ribonucleoproteins. In: Gesteland RF, Atkins JF, editors. The RNA World. Cold Spring Harbor: Cold Spring Harbor Laboratory Press. pp. 359-381.

46. Kaida D, Berg MG, Younis I, Kasim M, Singh LN, et al. (2010) U1 snRNP protects pre-mRNAs from premature cleavage and polyadenylation. Nature 468: 664-668. doi:10.1038/nature09479

47. Koerber RT, Rhee HS, Jiang C, Pugh BF (2009) Interaction of transcriptional regulators with specific nucleosomes across the Saccharomyces genome. Mol. Cell 35: 889-902. doi:10.1016/j.molcel.2009.09.011

48. Lygerou Z, Conesa C, Lesage P, Swanson RN, Ruet A, et al. (1994) The yeast BDF1 gene encodes a transcription factor involved in the expression of a broad class of genes including snRNAs. Nucleic Acids Res 22: 5332-5340.

49. Scannell DR, Byrne KP, Gordon JL, Wong S, Wolfe KH (2006) Multiple rounds of speciation associated with reciprocal gene loss in polyploid yeasts. Nature 440: 341-345. doi:10.1038/nature04562

50. Kyburz A, Friedlein A, Langen H, Keller W (2006) Direct Interactions between Subunits of CPSF and the U2 snRNP Contribute to the Coupling of Pre-mRNA 3' End Processing and Splicing. Molecular Cell 23: 195-205. doi:10.1016/j.molcel.2006.05.037

51. Millevoi S, Loulergue C, Dettwiler S, Karaa SZ, Keller W, et al. (2006) An interaction between U2AF 65 and CF Im links the splicing and 3[prime] end processing machineries. EMBO J 25: 4854-4864. doi:10.1038/sj.emboj.7601331

52. Rigo F, Martinson HG (2008) Functional coupling of last-intron splicing and 3'-end processing to transcription in vitro: the poly(A) signal couples to splicing before committing to cleavage. Mol. Cell. Biol 28: 849-862. doi:10.1128/MCB.01410-07

53. Berget SM (1995) Exon recognition in vertebrate splicing. J. Biol. Chem 270: 2411-2414.

54. Garas M, Dichtl B, Keller W (2008) The role of the putative 3 ' end processing endonuclease Ysh1p in mRNA and snoRNA synthesis. RNA-A PUBLICATION OF THE RNA SOCIETY 14: 2671-2684. doi:10.1261/rna.1293008

55. Giaever G, Chu AM, Ni L, Connelly C, Riles L, et al. (2002) Functional profiling of the Saccharomyces cerevisiae genome. Nature 418: 387-391. doi:10.1038/nature00935

56. Sherman F (2002) Getting started with yeast. Meth. Enzymol 350: 3-41.

57. Ghaemmaghami S, Huh W-K, Bower K, Howson RW, Belle A, et al. (2003) Global analysis of protein expression in yeast. Nature 425: 737-741. doi:10.1038/nature02046

CHAPTER 3

A HIGH-THROUGHPUT MUTAGENESIS STRATEGY IDENTIFIES RESIDUES IN THE C-TERMINAL TAIL OF BDF1 THAT MAY BE IMPORTANT FOR ITS ROLE IN PRE-mRNA SPLICING

3.1 Introduction

The eukaryotic cell is a complex environment in which functional pathways are closely interconnected to allow for optimal decision making processes under normal growth conditions or in response to environmental cues. The study of a single protein within a pathway involves both the detailed study of its functional domains as well as its integration in the broader network of interactions it makes with other cellular factors. Several genome-wide techniques have started to characterize global genetic and physical interaction networks, giving us a better perspective on how different factors can link multiple cellular pathways. In yeast, current methods like synthetic genetic arrays (SGA) [1] and Epistatic Miniarray profiling (E-MAPs) [2] aim to uncover the underlying genetic interactions based on the cellular fitness of double mutants, whereas proteomic studies [3-5] seek to place a certain factor within a complex by mapping its physical interactions. The more targeted analysis of functional domains has historically involved deleting or exchanging protein domains, identifying the interactors of a certain domain by yeast two hybrid analyses and affinity purification, and then using specific mutations to assess the function of important residues within the domain. Nevertheless, creating deletion mutants and

chimeric constructs can be time consuming may not readily offer answers when a combination of domains within a protein are required for a specific function.

The Bdf1 protein functions at the border of transcription and chromatin remodeling, as a subunit of yeast TFIID [6]. Through its interaction with Taf7, Bdf1 recruits RNA polymerase II to TATA-less [7, 8] promoters, which represent 80% of yeast promoters. Bdf1 maps to the C-terminal part of its mammalian homologue TafII250 and is a protein of the BET family, containing 2 bromodomains and a C-terminal ET (extra terminal) domain. Through its two bromodomains which recognize acetylated lysines on histone tails [9, 10], Bdf1 specifically binds hyperacetylayed histone H4 *in vitro* [11, 12] and *in vivo* [13, 14]. In addition to its function in transcription, Bdf1 co-purifies with the yeast SWR-C complex [15-17], which is responsible for exchanging histone H2A with histone H2A.Z (Htz1 in yeast) at promoters genome-wide. It was shown that Bdf1, Taf1 (TFIID) and Htz1 global occupancy is highly correlated [18] and that this correlation is dependent on the Esa1 acetyltransferase. Moreover, Bdf1 colocalizes with the +1 and +2 nucleosomes genome-wide [19], together with another SWR-C component, Vps72.

Interestingly, Bdf2 is a Bdf1 homologue which emerged following a whole-genome duplication event [20]. Deletion of either Bdf1 or Bdf2 is viable, but deletion of both copies leads to a synthetic lethal phenotype. Similarly to Bdf1, Bdf2 can interact with Taf7. However, whereas Bdf1 prefers H4 hyperacetylated tails, Bdf2 binds with equal affinity to all H4 species [11, 12]. Moreover, Bdf1 and Bdf2 occupy distinct genomic locations, although Bdf2 is upregulated upon Bdf1 loss and can bind some Bdf1 specific promoters [7].

In the previous chapter, I reported that that *bdf1Δ* but not *bdf2Δ* leads to global splicing inhibition in budding yeast, suggesting that while both proteins play a role in transcription, only Bdf1 maintains a role in pre-mRNA splicing. Here I present a modified version of the high-throughput assay I employed, which allows for fast screening through thousands of mutations generated in a protein of interest and leads to the rapid identification of protein regions involved in a particular process. In a pilot test, I employed this method to screen through multiple Bdf1 mutants and identify specific mutations that cause the pre-mRNA splicing defect in a *bdf1Δ* strain. I show that these mutations are distributed to two main regions of the protein – at the N and C-terminus. While most of the N-terminal mutations fall in the region of a predicted coiled coil which is more permissive to mutations, the C-terminal mutations lead to the truncation or modification through frameshift of the C-terminal acidic domain. Fascinatingly, this region overlaps with Bdf1's interaction domain with Taf7 and, moreover, with a region of Bdf1 phosphorylation and the C-terminal acidic SEED sequence.

3.2 Materials and Methods

3.2.1 Plasmid and strain construction

Bdf2 containing the ORF and about 500 base pairs of the 5' and 3' UTR was amplified off wild-type yBY4741 [21]. *S. cerevisiae* genomic DNA using the high fidelity Phusion polymerase (NEB cat#: M0530S) as per manufacturer's instructions (Denature: 98C 30sec; 30 cycles of 98C 10sec, 60C 20sec, 72C 1min; Final extension 72C 10min). The primers used to amplify the fragment are listed in Table 3.1 and were

phophorylated with T4 polynucleotide kinase (NEB cat#: M0201S) prior to the PCR reaction. Similarly, the Bdf1 ORF containing 111bp of the 5'UTR and 48bp of the 3'UTR (a 2259 bp product) was amplified using phosphorylated primers off the same yBY4741 genomic DNA. The URA marked pRS416 (see the ATCC collection, www.ATCC.org) CEN and pRS426 2-micron plasmids were linearized using SmaI (NEB cat#: R0141S), dephosphorylated with Antarctic phosphatase (M0289S) and purified off a 1% agarose gel. Similarly, the Bdf1 and Bdf2 fragments were gel purified. The Bdf1 fragment was cloned into the pRS416 vector, whereas the Bdf2 fragment was inserted into the pRS426 backbone using T4 ligase (M0202S). The plasmids were extracted out of bacteria and sequenced to confirm the quality of the integration.

Table 3.1: Primer sequences used in this study for cloning and qPCR

Target	Forward primer	Reverse primer
Bdf1 111bp upstream, 48bp downstream of ORF (mutazyme primers)	TAATTAAGCCTACTGGGTCGCTCC	GGTGCTCATTCTTCTCAGTCGTTG
Bdf2 ~500bp upstream and downstream of ORF	AGTATAATGTGACGTGGTTACCTTGG	TTAGATCGAGCTCGAGTAGTGAACAT
Bdf1-SmaI (1st fragment)	AACTGTACCTGAGTGGCTCAAATAGC	GAAGATAATCAAATTCAAAATTCAGCCCGGGTACGTACGTTTATCAGAGCCGTTGA
Bdf1-SmaI (2st fragment)	TCAACGGCTCTGATAAACGTACGTACCCGGGCTGAATTTTGAATTTGATTATCTTC	GAACTGCTACTGAATCCCTCAGACAT
Bdf2 qPCR	AATACCTTTTAAGCTCCATTAAGGCGACC	GGCTTCAGAAACGGTCTAGCATCTTTTAAT

In budding yeast, mutagenized templates can be introduced into a backbone which

contains sequences which are homologous to the ends of the fragment through spontaneous recombination. This process is known as gapped repair [22] and occurs when both the fragment of interest and the linearized backbone are co-transformed into a yeast cell. In this approach, a pool of such fragments were transformed into a population of yeast cells, thus generating a pool of strains in which each strain contains a unique mutated version of Bdf1 on a plasmid.

To obtain a recombination template for gapped repair (from now on referred to as Bdf1-SmaI:pRS413), ~500 bp upstream and downstream of the Bdf1 ORF were amplified and stitched together by PCR, using the overlapping primers in Table 3.1 to create a SmaI restriction site in the middle of the fragment. Two independent PCR reactions (Bdf1-SmaI) amplifying ~500bp of the 5'UTR and the ~500bp of the 3'UTR were run with Phusion polymerase following the manufacturer's instructions for 30 cycles then the products were gel extracted and 2ng of each product mixed together. A 10 cycle 20 µl reaction with the mixed template but lacking the external primers was run under the same conditions as above to allow for the annealing and extension of the fused template. Then, 1µl of this reaction was diluted in a 50 µl reaction with Phusion polymerase in which external phosphorylated primers were used to amplify the stitched fragment. The pRS413 vector (ATCC) was linearized with SmaI, dephosphorylated and used to clone in the fragment containing ~500bp of the 5' and 3'UTR of Bdf1. Selection was performed on plates lacking histidine.

In addition, a Bdf1-Kan fragment was amplified using the same Bdf1 (mutazyme) primers listed in Table 3.1 from *bdf1Δ* genomic DNA and introduced into Bdf1-SmaI:pRS413 by gap repair. To this end, the Bdf1-SmaI:pRS413 vector was cut with

SmaI, dephosphorylated and purified by gel extraction, then mixed stoichiometrically with a purified Bdf1-Kan fragment. This mix was used in a regular yeast transformation reaction in a yBY4741 (WT) background and colonies were selected on SD-His media. The plasmids were then extracted out of saturated yeast cultures, by first degrading the cell wall with lysozyme and then by performing a customary bacterial prep. These were then transformed into competent DH5α strains and a bacterial stock of the plasmid prepared.

The yBY4741 *bdf2Δ* MATα strain from Open Biosystems was used to construct the *bdf1Δ bdf2Δ* Bdf2:pRS426 strain in which the gap repair reactions for obtaining Bdf1 mutants were performed. To this end, first the *KanMX* marker cassette which marks the Bdf2 deletion was exchanged with the *NatMX* (nourseothricin) marker by using the standard lithium acetate transformation procedure (ref). The *bdf2Δ::NatMX* Bdf2:pRS426 strain was generated by standard transformation techniques and colonies were selected on minimal media lacking uracil. The Bdf1-Kan:pRS413 vector was restriction digested with BamHI (R0136S) and EcoRV-HF (R3195S) and 1-2 µg amounts were used to knockdown Bdf1 in the *bdf2Δ::NatMX* Bdf2:pRS426 background. The *bdf1Δ::KanMX bdf2Δ::NatMX* Bdf2:pRS426 strain was selected on YPD (ref) +G418+Nat and then tested on SD-Ura and 5-FOA.

The Bdf1:pRS416 vector was used as a template in the Mutazyme PCR (Agilent technologies cat#:200550). The reactions were run as indicated in the manufacturer's manual, by testing different input amounts and varying the number of cycles. Thus, for the "low" mutation rate reaction 5ng of a purified Bdf1 fragment (111bp upstream, 48bp downstream of ORF) were used and 20 PCR cycles were run. Conversely, for

the “medium” and “high” mutation rate 100 and 300ng of template were used respectively and 30 cycles of amplification were performed. In parallel, a control reaction using Phusion polymerase was run, and 5% of the amplification products run on a 1% agarose gel to check the correct size.

The PCR products were then mixed with DNA binding buffer (5M Guanidine-HCl, 30% Isopropanol), bound to a Zymo5 (cat#: C1003-250) column and washed twice with Wash Buffer (80% Ethanol, 10mM Tris-HCl). The column purified products were integrated by gap repair into the Bdf1-SmaI:pRS413 vector, leading to the formation of a Bdf1 mutagenized ORF flanked by ~500bp of native UTR sequences on either side. The gap repair reactions were initially performed in a yBY4741 strain in order to empirically check the mutation rate under the 3 conditions. In this case, the yBY4741 strain was grown in rich media (YPD) at 30°C and the cells were allowed to recover O/N in 5ml SD-His media before plating onto SD-His agar plates. To obtain Bdf1 mutants for the screen, the gap repair was performed in the *bdf1Δ::KanMX bdf2Δ::NatMX* Bdf2:pRS426 strain. Given that the cell needs to retain a functional Bdf copy for viability (in this case Bdf2), this strain was also grow in rich media but at 25°C (*bdf1Δ* is temperature sensitive) and allowed to recover overnight (26h) in 60ml of SD-His media. Then 5ml aliquots were spun down and plates onto 5-FOA plates to achieve the shuffling out of the Bdf2:pRS426 plasmid.

Serial dilutions of strains were done by growing them until mid-log phase OD_{600nm}=0.5 in rich media (YPD) then, diluting them in sterile MilliQ water to OD_{600nm}= 0.1 and making 6 4-fold serial dilutions. The strains were then pinned onto YPD or selective media and allowed to grow at different temperatures.

3.2.2 High throughput strain handling, RNA extraction, cDNA conversion and qPCR

Bdf1:pRS413 *bdf1Δ bdf2Δ* strains together with a couple of *bdf1Δ* positive and *bdf2Δ* negative controls were hand-picked off 5-FOA selection plates and set to grow to saturation in a 96-well plate (USA Scientific cat#:1896-1110) in 500ml YPD at 25°C for 2 days. At this time, a glycerol stock at a 15% final glycerol concentration was made and the plate was stored at -80°C. The strains were pinned onto solid YPD media using a 96-well pinner tool, allowed to grow for 1-2 days in the 25°C incubator, then transferred into positions A1 and A2 (every other row) of a 384-well plate (Greiner BioOne cat#:781271) containing 150 µl YPD. The rest of the plate was filled with a diluted yBY4741 (WT) culture. The plate was sealed with adhesive breathable tape (VWR cat#: 600941-086) and allowed to grow for 2 days at 25°C 250rpm. Using robotic protocols previously developed (see Chapter 2, Methods), the saturated cultures were mixed and 2 µl of the saturated culture was freshly diluted into a new 384-well plate containing 150 µl YPD/well. The cultures were allowed to grow for 4h at 30°C 900rpm, then collected through centrifugation at 4000*g for 5 minutes. The plates were flash frozen in liquid nitrogen and stored at -80°C.

Robotic RNA isolation, cDNA conversion into cDNA and high-throughput qPCR was performed as described in chapter 2, with the exception that all protocols were modified to accommodate one plate per extraction instead of two. qPCR was performed using the U3 precursor primers to measure the species of interest and FAA1 primers as a control. Data analysis was performed as described in Chapter 2 with the difference that the CV filter was applied only to the FAA1 primer pair and the FAA1

RNA levels were used to normalize the RNA amount in each well instead of a composite normalization constant. The SAM analysis to determine statistical significance was performed as described before and the delta values was chosen to minimize FDR (Δ=1.38, FDR=0).

3.2.3 Culture growth, RNA extraction and cDNA synthesis for small scale qPCR

The strains were grown overnight in rich media (YPD) at the permissive temperature of 25°C, then back-diluted the following morning at an OD_{600nm}=0.05 and outgrown in YPD at 30°C until they had undergone at least two doublings (OD_{600nm}=0.3-0.5). The strains were collected by filtration under vacuum using Millipore filters (Millipore cat #:HAWP04700), flash frozen in liquid nitrogen and stored at -80°C.

Total RNA was extracted using the hot phenol method. The pellets were resuspended in equal amounts of Acid phenol (pH<5.5) and AES buffer (50mM NaAcetate pH 5.3, 10mM EDTA, 1% SDS), and incubated in a 65°C water bath for 7 minutes, vortexing every minute. The samples were then loaded onto prespun 15mL PhaseLock Gel Tubes (Fisher Sci cat#: FP2302850) and spun in an X 15R/SX4750A rotor centrifuge at 4750rpm for 5 min. The RNA was back-extracted with Phenol:Chloroform:IAA (25:24:1), followed by Chloroform. Total RNA was precipitated with 3M NaOAc pH 5.3 in the presence of 100% isopropanol then spun for 20min at 14K at 4°C in a tabletop centrifuge. The pellet was washed twice with 70% ethanol then dried in a vacuum concentrator. The pellets were resuspended in sterile MilliQ water to a concentration of 1-2µg/µl.

5 µg of total RNA was DNase treated at room temperature for 15 min with 2 µl of

DNase Mix (1x DNase Buffer, 0.25 units of DNase I (Promega)). The samples were then mixed with RNA biding buffer (2M Guanidine-HCl, 75% isopropanol) and bound to Zymo-5 columns (C1003-250). The columns were spun at 14K for 1 minute in a tabletop centrifuge then washed twice with RNA wash buffer (80% Ethanol 10mM Tris-HCl). A final dry spin prior to elution was performed for 2 min at 14K. Total DNase trated RNA was eluted into 10 µl sterile MilliQ water. The 10µl were used in a cDNA synthesis reaction that had a total volume of 40µl and which contained 50mM Tris-HCl (pH8.3), 75mM KCl, 3mM $MgCl_2$, 10mM DTT, 0.5mM each dNTP, 5µg dN_9 primer, and 60ng M-MLV RT. Reactions were incubated for 2h (>99% reaction completion) or overnight at 42°C.

3.2.4 qPCR

The cDNA reactions were diluted to 2ng/µl based on the initial RNA concentration and used without any further purification as templates in quantitative PCR (qPCR) reactions. The qPCR reactions were performed in a reaction volume of 15µl, containing 5µl of template (~10ng of template), 10mM Tris-HCl (pH8.5), 50mM KCl, 1.5mM $MgCl_2$, 0.2mM each dNTP, 0.25x SYBR Green, 5% DMSO, 0.7ng *Taq* DNA polymerase, and 250nM forward and reverse primers. The sequences of the primers used for measuring Bdf2 levels are shown in table 3.1, whereas the sequences for amplifying the U3 precursor and the FAA1, SCR1 and Tub1 normalization genes were discussed in Chapter 2.

3.3 Results

3.3.1 A forward genetic high-throughput method to identify mutants which impact total transcript levels

In the previous chapter, I described a quantitative approach I have developed to assess global changes in transcript levels in the budding yeast *Saccharomyces cerevisiae*. This method allowed me to screen through a known collection of mutants comprising over 93% of the budding yeast genome and identify trans-acting factors involved in pre-mRNA splicing. Among the top candidates, I identified factors involved in chromatin remodeling and transcription, and proteins belonging to the 3'end processing machinery which cause global or gene-specific splicing defects. One of the most promising candidates is Bdf1, a factor which has been characterized to actively function in transcription by recruiting RNA polymerase II, as well as in chromatin remodeling at the level of the promoter, by interacting with the SWR-C complex. In our study, Bdf1 caused a significant increase in the precursor levels of the splicing reporter U3 snoRNA (Figure 3.1A) and led to global defects in splicing as shown by splicing sensitive microarrays (Figure 3.1B). Intriguingly, Bdf1's homologue Bdf2 with which it shares significant homology (35% identity, 67% similarity) did not impact U3 precursor levels, nor did it cause detectable mis-splicing of intron containing genes on a genome-wide basis as determined by microarray analysis. Indeed, Bdf2, which resulted following the whole-genome duplication event in yeast, shares the domain structure of Bdf1, including the two bromodomains (52-56% identity, 86-91% similarity) and the C-terminal ET (extra terminal) acidic domain. Nevertheless, despite these similarities, Bdf1 and Bdf2 have diverged and contain

several stretches of amino acids which are present in one protein, but not the other. Most significantly, these differences in protein structure enabled Bdf1 to maintain a role in pre-mRNA splicing which is absent in Bdf2. Therefore, one important question that remains to be addressed is how Bdf1 functions in pre-mRNA splicing and what residues within the protein are essential for this function.

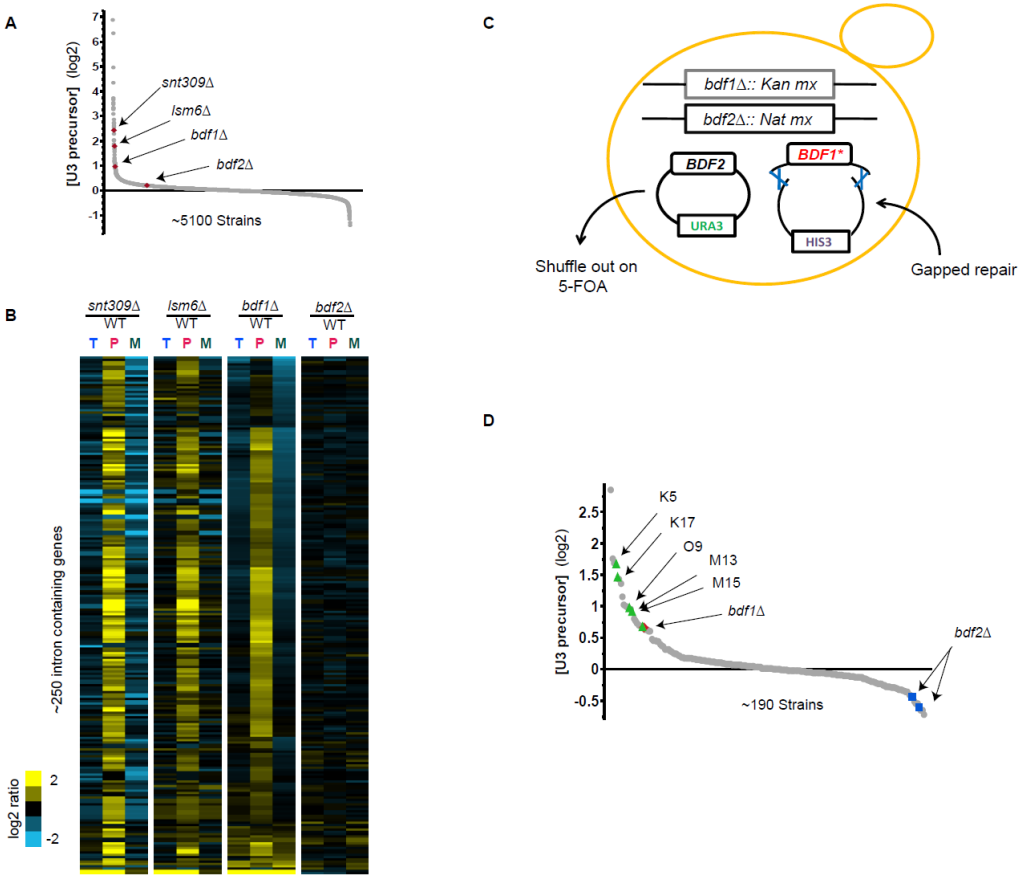

Figure 3.1: Mutagenesis rationale and identification of Bdf1 mutants which mimic bdf1Δ

A. *bdf1Δ* upregulates U3 precursor levels similar to the bona-fide splicing mutants *snt309Δ* and *lsm6Δ*, in contrast to *bdf2Δ*. The data were generated in the context of ~5500 yeast mutants (see chapter 2) and are presented in log2 space.
B. Splicing sensitive microarrays depicting the whole-genome splicing defect in bdf1Δ in comparison to the two splicing mutants *snt309Δ* and *lsm6Δ,* and in contrast with *bdf2Δ*. The genes are organized based on the clustering within the *bdf1Δ* array, and each row in all 4 arrays represents the same gene.

C. Schematic representation of the mutagenesis rationale to create yeast strains containing only a single mutagenized Bdf1 protein-.
D. Pilot data from a plate containing 96 mutants and 96-wild type strains. The data was processed as described in Materials and Methods, and the expression values (in log2 space) were ordered from high to low. The data points within the S-curve represent the average of 4 experimental values. The top ranking highly-reproducible candidates which passed the SAM analysis are highlighted in green. The location of the *bdf1Δ* positive control and the *bdf2Δ* negative control are shown in red and blue respectively.

To identify the residues in Bdf1 which make it crucial for pre-mRNA splicing, I employed our previously published method in a forward genetic manner. To this end, I first mutagenized the Bdf1 gene using Mutazyme II (see Materials and Methods). The mutagenized Bdf1 product will be referred to from now on as Bdf1*. Mutazyme II is a mixture of enzymes which allows for substitutions of bases within the mutagenized fragment to occur with broadly equal preference for the four bases. Compared to other PCR mutagenesis approaches, the concentrations of dNTPs or Mg2+ are not varied. Instead, the amount of template in the reaction and the number of PCR cycles determine the mutation rates in the final product. To empirically test mutation rates, I used three different mutagenesis conditions – "high", "medium" and "low" – and utilized the fragments in a gap repair reaction in order to construct Bdf1*:pRS413 plasmids (HIS marked). Following recombination, these plasmids should contain ~500 bp of the native 5' and 3'UTR surrounding the mutagenized ORF. The mutation rates I empirically determined closely match the expected values (see Table 3.2), whereas a control reaction using Phusion polymerase led to zero mutations/product. In addition, the mutations were comprised of 13 transitions and 7 transversions.

Table 3.2: Expected versus observed mutation rates under different mutagenesis conditions

	Mutation frequency (mut/kb)		
Mutation rate	**Expected**	**Observed**	**Mutations/Bdf1 gene (2.1kb)**
Low	0-4.5	0-3	0-5
Medium	4.5-9	1.5-4.3	3-9
High	9-16	4.3-7	9-15

To screen for Bdf1 mutants which behave like a Bdf1 null, I chose the "medium" mutagenesis conditions to allow for a sufficient number of mutations within the 2.1kb Bdf1 gene and constructed Bdf1*:pRS413 *bdf1Δ bdf2Δ* strains. Since the double *bdf1Δ bdf2Δ* is not viable in yeast cells, a multi-copy 2-micron version of the Bdf2 gene was introduced on a URA marked vector in a *bdf2Δ* background, after which the wild type chromosomal copy of Bdf1 was knocked out in this strain (see Materials and Methods). The Bdf2:pRS426 *bdf1Δ bdf2Δ* strain thus created was verified for its ability to grow on plates lacking uracil, and the double deletion was confirmed through resistance to both G418 and Nourseothricin. Next, I conducted gapped repair with a mutagenized Bdf1 pool and a HIS marked vector backbone in this strain background, followed by the shuffling out of the Bdf2 copy onto 5-FOA (see Figure 3.1C and Materials and Methods). This approach ensures that the mutagenized Bdf1* is the only functional Bdf copy present in these cells, whose presence was confirmed by growth on plates lacking histidine. In addition, the strains were checked for the lack of ability to grow on plates lacking uracil, thus ensuring that they have indeed shuffled out Bdf2. These mutants were grown in 384-well plates together with wild type, positive (*bdf1Δ*) and negative (*bdf2Δ*) controls. The robotic methods discussed in

chapter 2 were employed to extract RNA, convert it into cDNA, and quantitatively measure U3 precursor level in all of these mutants by qPCR. Thus, our previously reported method can be employed to rapidly identify mutants within a target protein which cause changes in cellular RNA levels.

3.3.2 Testing and validation of Bdf1 mutants which cause U3 precursor upregulation

In the original screen, in which I was globally searching for trans-acting factors which impacted pre-mRNA splicing, the general expectation was that the majority of ~5500 yeast mutants play no role in splicing. In the current screen the hypothesis is similar, in that the majority of mutations would change residues in regions of Bdf1 which are conserved with Bdf2 and thus would have no role in splicing. Figure 3.1D shows the distribution of U3 precursor levels in 96 Bdf1*:pRS413 *bdf1Δ bdf2Δ* mutants and 96 wild-type strains. As expected, the vast majority of the data sit within a ±1.5 fold-change, with only a few candidates reproducibly causing a larger accumulation in U3precursor levels. The *bdf1Δ* positive control strain shows a 60% increase, which is lower than previously observed, while the *bdf2Δ* negative controls appear towards the bottom of this distribution and show a 50-60% decrease in total levels. Furthermore, I employed the SAM software (Significance Analysis of Microarrays, see Chapter 2) to test the statistical significance of the data which led to five candidates being called as causing a reproducible increase in U3 precursor levels at least at the level of *bdf1Δ*: K5, K17, O9, M13 and M15 (Table 3.3).

Table 3.3: SAM generated list of the top reproducible candidates

The Score(d) is calculated by dividing the Numerator value (fold change) to the Denominator (a measure of the standard deviation within the replicates)

Mutant	Score(d)	Numerator(r)	Denominator(s+s0)
O9	11.4273	0.9708	0.0850
M15	10.9732	0.6783	0.0618
M13	10.3015	0.9158	0.0889
K5	4.9730	1.5840	0.3185
K17	4.2383	1.2803	0.3021

To further validate these results, I re-measured the U3 precursor levels by small scale qPCR (Figure 3.2A). All of these candidates except for one (M13) recapitulated the 2-fold increase in U3 precursor levels observed in the screen. Upon sequencing, the candidates actually revealed three different mutation patterns, with two of the positive candidates having the same sequence (O9 and K17).

Besides the candidate strains, I tested several strains which did not impact U3 precursor levels in my pilot screen (negatives). Strains A1, A2, A3 and A5 did not show any increase in U3 precursor levels in the screen or led to a slight (10-20%) decrease from the median, showing a similar phenotype to the *bdf2Δ* strain. As expected, none of these strains showed an increase in U3 precursor levels larger than 2-fold. However, A1 did exhibit an upregulation in U3 precursor levels of 0.75 on a log2 scale which is slightly larger than the variation I observed for *bdf2Δ* in multiple experiments. This begs the question of what degree of upregulation really constitutes a splicing mutant capable of mimicking the *bdf1Δ* phenotype and will need to be investigated with different assays. In addition to the strains from this particular experiment, I tested several negatives from a separate experiment (bottom panels, figure 3.2). Neither of the H4, D8 or L3 mutants caused a change in U3 precursor

levels larger than ~50%.

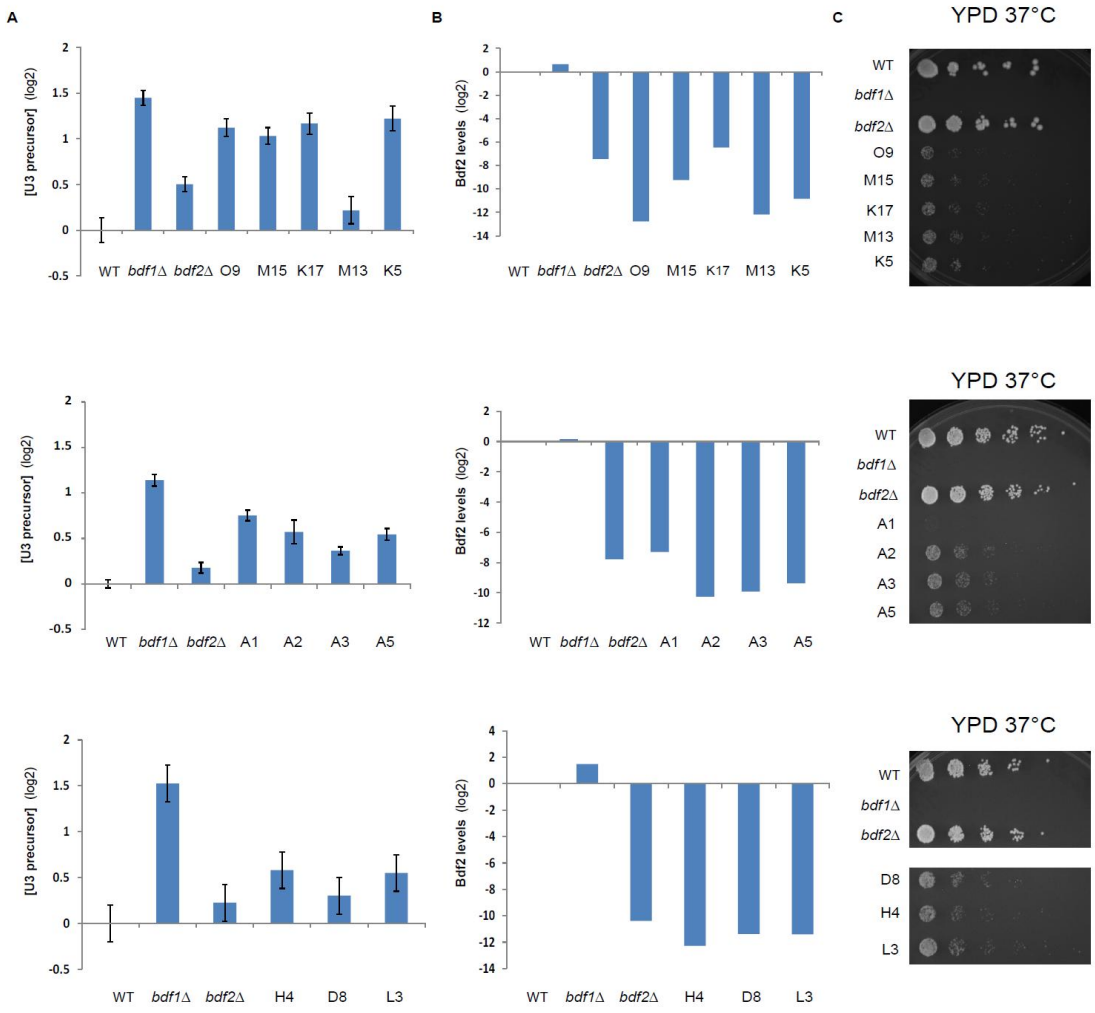

Fig 3.2: Confirmation of candidate strains and negatives by small scale qPCR

A. qPCR to measure U3 precursor levels in the candidates and negative strains. The values are normalized to a normalization constant calculated based on the FAA1, Tub1 and SCR1 levels and expressed in log2 space. Representative experiments are shown and the standard error represents the error of technical replicates.
B. qPCR to measure and Bdf2 levels
C. Serial dilution of yeast strains grow onto YPD media at 37°C. The two panels at the bottom right show strains on the same plate.

Further validation indicated that the Bdf2 levels in all strains were depleted at least to the extent of the *bdf2Δ* strain, indicating that the strains were indeed lacking a Bdf2 copy (figure 3.2B). Moreover, Bdf2 was consistently upregulated ~2-fold in a *bdf1Δ*. All candidate strains were His+ Ura- (data not shown) and grew slowly at 37°C, regardless if the mutation caused an increase in U3 precursor levels or not (figure 3.4C). Furthermore, the A1 mutant displayed a more severe sickness at 37°C than any of the other Bdf1 mutants. These observations suggest that Bdf1's role in splicing and the temperature sensitiveness of Bdf1 mutants are not correlated. This lack of correlation is not entirely surprising in the light of other Bdf1 mutants which were described by the Buratowski group. Thus, they have shown that multiple mutations which fall within the bromodomains cause ts phenotypes [12], as can mutations of serine residues which are normally phosphorylated in Bdf1 [23]. The mutations they describe fall in different regions of the Bdf1 protein - bromodomains versus IPR (internal phosphorylation region) and CPR (C-terminal phosphorylation region) - and do not seem to be linked to any one particular function of Bdf1.

3.3.3 The Bdf1 mutations which mimic the *bdf1Δ* phenotype affect the distal C-terminal region

In order to determine the significance of certain mutations within a particular sequence context I took a comparative approach: I was interested in mutations which affect highly unique residues in Bdf1 when comparing Bdf1 with Bdf2, and which show high conservation with other Bdf1 homologues in sister *Saccharomyces* species. In addition, I was interested how these mutations map within different domains of the

Bdf1 protein (Figure 3.3A) and what was the likelihood of their disrupting the function of that domain. Thus, I first generated an alignment of Bdf1 and Bdf2. As shown in Figure 3.3B, despite the significant homology between the two homologues, there are certain regions in Bdf1 which are absent in Bdf2, as well as regions of lower conservation between the two, such as the N- and C-termini. Next, I aligned all 7 *Saccharomyces* Bdf1 homologues using the information on the UCSC platform and derived a conservation trace (Figure 3.3C). Finally, I used both the Swiss Model [24, 25, 26] and I-TASSER [27, 28] programs to generate the secondary structure prediction of Bdf1 and superimposed it with the conservation traces (Figure 3.3A). I then mapped the location of the mutants within all three contexts (Figure 3.3D).

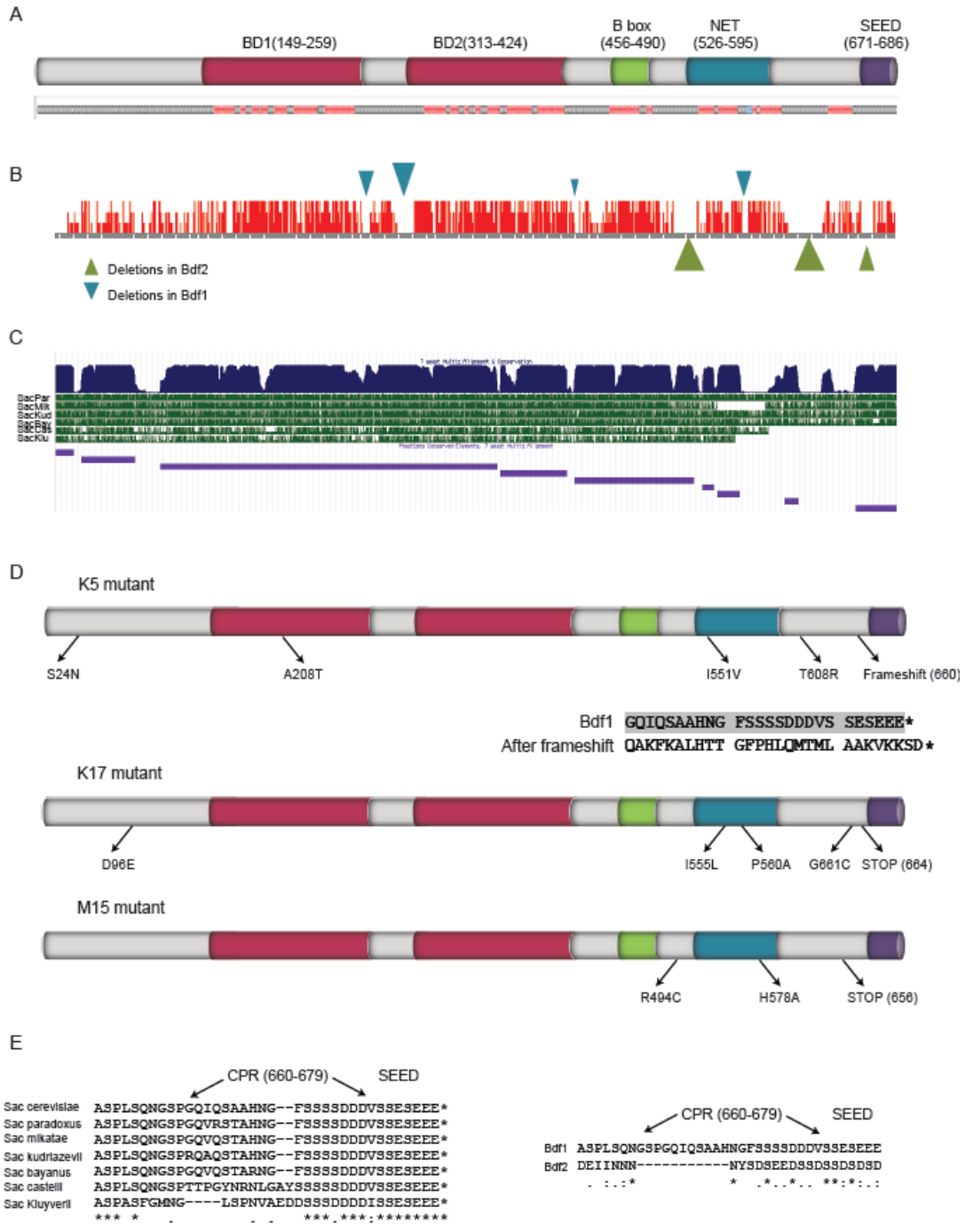

Figure 3.3: The mutations in Bdf1 in the context of its primary and secondary structure

A. Domain structure of Bdf1 - the domains are defined in terms of position and are colored, as opposed to the gray coiled coil area. Underneath, the secondary structure prediction from I-TASSER is aligned, where red represents an α-helix, blue a β-sheet and black the coiled coil.

B. Conservation diagram at the amino acid level for Bdf1 and Bdf2. The height of the graph is directly proportional to the conservation between the two proteins. Blue triangles represent deletions in Bdf1, whereas green triangles stand for deletions in Bf2.
C. Conservation trace generated by the UCSC platform, representing the alignment between *S. cerevisiae* Bdf1 and 6 other *Saccharomyce*s species.
D. Location of the mutations in the three top candidates with respect to Bdf1's protein structure. The altered sequence as a result of the frameshift in the K5 mutant is presented underneath the grey Bdf1 native sequence.
E. Conservation of the C-terminal end comprising the CPR (660-670) and SEED regions in Bdf1 homologues as opposed to Bdf1/Bdf2.

A comprehensive list of mutations (including silent mutations) generated by Sanger sequencing of all plasmids is provided in Table 3.4.

Table 3.4: The complete list of Bdf1 mutants

All mutated amino acids are reported. If the mutants contain a change in a specific residue the change is noted and is compared with the residue in Bdf2. The conservation within Bdf1 and Bdf2 for a residue is noted, as well as the underlying secondary structure (coil or helix) and whether the residue falls within a specific domain. The IPR (internal phosphorylation region) occurs right after the second bromodomain, but before the B box, whereas the CPR (C-terminal phosphorylation region occurs after the NET domain but before the terminal SEED sequence (also see Figure 3.3) Silent mutations which occur in the wobble position of the codon are noted in grey.

Mutant	Position	Amino acid	Mutation	Bdf2 equivalent	Conserved in Bdf1	Conserved in Bdf2	Secondary structure	Domain
K5	24	S	N	A	5/5	5/5	C	No
	62	L	L					
	69	L	L					
	208	A	T	V	7/7	4V/T/2A	C	BD1 (coil between αA and αB)
	228	D	D					
	551	I	V	I	5I/C/V	4I/V	H	NET
	608	T	R	no	6T/V		C	No
	660	P	Frameshift					acidic tail; becomes more positive/ hydrophobic/ excludes CPR
K17	96	D	E	L	D/K/3E	5/5	C	No
	555	I	L	I	7/7	5/5	H	NET
	560	P	A	P	7/7	5/5	C	NET (between 2 α helices)
	661	G	C	no	4G/R/T		C	CPR
	664	Q	STOP	no	4Q/R/G		C	cuts off very acidic tail excludes CPR
	671	F	F					
M15	49	L	L					
	74	I	I					
	252	L	L					
	494	R	C	H	7/7	H/R/S/2Y	C	No (after B box, before NET)
	578	H	Q	A	4H/Y/E/S (within conserved domain)	A/E/2G/R	H	NET
	656	Q	STOP	N	6Q/F	4/4	C	No; cuts off acidic tail excludes CPR
A1	245	A	D	K	7/7	K/2T/L/I/2A	H	BD1 (αC)
M13	61	A	T	G	4A/I	4G/D	C	No
	76	G	D	D	3G/D/S	5/5	C	No
	159	L	F	S	4L/S/M/T	2S/3N/M/T	H	BD1
	233	N	K	E	5N/2T less conserved	5E/2T	C	BD1, between αB and αC

					within the bromodomain			
	358	V	I	V	7/7	3V/4I	C	BD2 (within αA)
	377	Q	H	K	6Q/E	3K/2E/S/Q	C	BD2 (right before αB)
	669	N	N					
A2 (same as M13 plus one)	685	E	H	S	7/7	4/4	C	final acidic domain
A3	16	N	N					
	28	R	K	A	5/5	5/5	C	No
	38	G	V	S	5/5	3/3	C	No
	426	D	N	E	3D/4E	5E/P/D	C	No (right after BD2, before IPR)
	465	E	G	E	7/7	7/7	H	in the very conserved B box domain
	563	S	C	S	6S/G	4S/P	C	between H and a beta sheet in NET
A5	211	V	I	S	3V/F/A/T/S	6S/T	C	BD1 (right before αB)
	665	S	R	no	5S/I/L		C	in CPR
D8	41	E	V	S	5/5	3/3	C	No
	49	L	L					
	58	D	E	D	D/4N	4D/1V	C	No
	69	L	R	D	4L/1P	5/5	C	No
	373	D	D					
	621	S	S					
H4	107	G	R	S	2G/2E/missing	5/5	C	No
	685	E	STOP	S	7/7	4/4	C	final acidic tail
L3	194	P	P					
	531	D	G	D	6D/1N	5/5	H	NET

The K5 mutant contains 5 non-silent mutations (figure 3.3D). The first one, S24N occurs in a coiled coil terminal region that is not conserved between Bdf1 and Bdf2 (S versus A), but in which the S and the A are conserved in Bdf1 and Bdf2 respectively.

However, as discussed below, this N-terminal coiled coil region seems more permissive to mutations that do not impact U3 precursor levels. The A208T mutation is located in a coiled coil region between helix αA and αB in bromodomain 1 (BD1). It is highly conserved in Bdf1 whereas the Bdf2 homologue has a valine at this position. Nevertheless, two other species of Saccharomyces (*S. castelli* and *S. kluyveri*) also have an alanine in this position in Bdf2, which could suggest that this residue has a similar role in Bdf1 and Bdf2. Another mutation in this mutant is I551V. However, Bdf1 and Bdf2 share the same residue at this position and therefore this mutation is unlikely to be relevant from a splicing perspective. The fourth mutation in this mutant occurs at position 608, T608R, which switches a hydrophobic residue to a positively charged one. The threonine residue is highly conserved in Bdf1, but has no correspondence in Bdf2. Given that this residue maps to a deletion in Bdf2, it is - a strong candidate to be of functional significance for the role of Bdf1 in splicing. Lastly, there is a frameshift at position 660 in Bdf1 which deletes one of the phosphorylation regions in Bdf1 (CPR) and also the SEED terminal domain which is highly conserved in Bdf1 and is enriched in E, D and S residues.

The K17 mutant contains a D96E mutation which leads to no change in the charge of this residue. In addition, this residue is not highly conserved between Bdf1 species, where D and E are highly interchangeable. Moreover, two other mutations present at I555L and P560A are unlikely to be of importance from a splicing point of view, because these residues are identical in Bdf1 and Bdf2 and highly conserved in both homologues. Therefore, a mutation in one of these residues would likely cause identical phenotypes in Bdf1 and Bdf2. More interestingly, two other mutations in this

Bdf1 mutant fall in a region of Bdf1 that maps to a missing region in Bdf2. Thus, G661C maps to a relatively conserved region within the CPR, while the other mutation inserts a stop codon at position 664. This stop codon cleaves off part of the CPR and the highly conserved SEED region.

Finally, there are three mutations in the M15 mutant. The first one, R494C, occurs in a highly conserved residue within a coiled coil after bromodomain 2 (BD2). The residue is not conserved within Bdf2 homologues, although the *S. paradoxus* Bdf2 homologue contains an R at this position. Furthermore, a second mutation, H578Q, occurs in a non-conserved residue within the NET domain of Bdf1, while the third mutation at position 656 introduces a stop codon, again cleaving off the CPR and the acidic tail.

Thus, all three mutants that cause an upregulation in U3 precursor levels contain either a non-sense mutation or a frameshift that leads to the removal of the C-terminal phosphorylation domain and of the highly acidic SEED domain.

Among the negatives, D8 has several mutations within the N-terminal coiled coil which either occur in very conserved (E41V, L69A) or in less conserved residues (D58E). This suggests that the N-terminal coiled coil is more permissive to mutations and that this region is likely not very important for Bdf1's involvement in splicing. Furthermore, the H4 mutant contains a mutation in a non-conserved residue, G107A, in the N-terminal coiled coil and a second one which leads to a stop codon in the next to last amino acid of Bdf1. Despite its position in the very conserved SEED sequence, this mutation only shortens the protein by one amino acid. The third mutant, L3, contains a single mutation, D531G, at a well conserved position in the NET helix. Nevertheless, the same residue is highly conserved in Bdf2 as well, and it is likely to

therefore affect both proteins in a similar fashion.

M13 is the only false positive mutant, which failed to re-test in the small scale qPCR. It contains three mutations within the N-terminal coiled coil (A61T, G76D and L159F). It also contains one mutation within BD1 (N233K) and two mutations within BD2 (V358I and Q377H), all of which appear to have no impact on U3 precursor levels. Moreover, the A2 mutant is nearly identical with M13, containing only one additional mutation, E685H, which was shown in the H4 mutant to be aphenotypic.

The last negative strain, A3, contains mutations that either maintain the original charge of the amino acid (R28K), or that occur in residues that are highly conserved between Bdf1 and Bdf2 (E465G, S563C) and therefore are unlikely to impact a Bdf1 specific function. Another mutation in the N-terminal coiled coil (G38V), as well as one after BD2 (D426N), are likely not important for the Bdf1 splicing phenotype.

Interestingly, the A5 mutant does not cause a significant increase in U3 precursor levels, despite a mutation (S665R) in the CPR domain.

Lastly, mutant A1 contains a single mutation (A245D), within the αC helix of BD1, in a highly conserved residue in Bdf1. The importance of this mutation will need to be addressed in future experiments, since a 2-fold increase in U3 precursor levels is not a clear cut measurement which guarantees a splicing phenotype. Therefore, it is unclear if a mutation in the first bromodomain of Bdf1 can cause splicing defect.

In addition to all the previously discussed mutations, the silent mutations observed in some of these mutants might interfere with translational rates, despite failing to change the amino acid at a particular position. However, until further validation of these phenotypes is achieved, it is unclear what their role is. In conclusion, mutations in the

C-terminal phosphorylation region and the SEED domain of Bdf1, which belong to the region of Taf7 interaction (amino acids 500-686 in Bdf1), are so far the most likely candidates in defining a domain which might be important for pre-mRNA splicing.

3.4 Discussion

Bdf1 is a 686 amino acid protein with multiple functions in budding yeast. Aside from its role in transcription and chromatin remodeling through its interaction with SWR-C, it was initially described as a factor associated with chromatin which is required for sporulation [29, 30]. In addition, a separate function for Bdf1 was described in inhibiting the spreading of heterochromatin from subtelomeric regions into areas characterized by active gene expression, by competing with the Sir2 deacetylase on histone H4 tails [13, 31, 32]. Bdf1 and Bdf2 are homologues which resulted following the whole genome duplication event which took place in an ancestor of *S. cerevisiae* ~100 million years ago. Despite sharing significant sequence homology, Bdf2 plays a more limited role in the cell. Bdf2 can interact with Taf7 and recruit RNA polymerase in the absence of Bdf1, but was not purified together with the SWR-C complex [15, 33] and has equal preference in binding histone tails. Moreover, Bdf2 is not required for sporulation and does not appear to exhibit any other function aside from its involvement in transcription. This could be partially explained by the fact that Bdf1 is the more abundant cellular isoform among the two (~8100 copies/cell as opposed to ~2930 copies/cell for Bdf2 [5]). Nevertheless, overexpression of Bdf2 in a *bdf1*Δ background did not rescue the over-accumulation of U3 precursor levels which are observed in this strain (data not shown), suggesting that it is likely unable to suppress

the splicing defect exhibited in *bdf1Δ*. All these lines of evidence suggest that Bdf1 plays multiple roles, only part of which are fulfilled by Bdf2.

In this study, I followed up on a previous observation that *bdf1*Δ leads to a general splicing phenotype in *S. cerevisiae*, whereas *bdf2*Δ is aphenotypic. By applying the method described in Chapter two in a forward genetic manner, I have mapped residues within Bdf1 which when mutated lead to a *bdf1*Δ phenotype, measured an increase in the accumulation of U3 precursor. Although this approach was only applied on a small subset of mutants as part of a pilot test, I was able to identify several candidates, which suggest that the C-terminal tail of Bdf1 may be linked to its function in pre-mRNA splicing. Even though these candidates await characterization by splicing sensitive microarrays, in the future this type of approach could be expanded to more precisely determine "hotspots" within the protein which are impacted when the splicing related function of Bdf1 is lost.

Given that the Bdf1 gene is 2100 nucleotides long, and assuming that a mutation at each position would generate a viable mutant, one would require 2100 unique mutants to generate 1X coverage of this gene. In my current test, I can assume a median of ~5 mutations/gene. Thus, screening of about 2000 mutants (about 5-6 384-well plates) would generate ~5X coverage of the Bdf1 ORF. Nevertheless, with an increased number of mutants comes the complicated task of assessing how important the mutated residues really are for Bdf1's role in pre-mRNA splicing. In addition, since the mutazyme strategy relies on the number of PCR cycles for achieving the desired mutation rate, an increased number of PCR cycles will most likely lead to an oversampling of certain mutations or mutation combinations and to the under

sampling of others. If one assumes that one third of the mutations will fall within the wobble position of the codon and thus most likely lead to silent mutations, whereas some mutations will not change the charge on the native amino acid, the number of strains that would be required to capture all these mutants would increase significantly. Thus, an approach that would use a low mutation rate (1-3 mutations/gene) combined with a number of strains which can be easily handled but would allow for 2-3X coverage of the Bdf1 gene would be a good starting compromise. A low mutation rate would have the advantage of more unique mutations being sampled, together with a far simpler way of identifying the relevant residues. While some combinations of mutations will certainly be lethal (for example the two mutations in the active sites of the two bromodomains), most will still likely not impact a Bdf1 specific function, making this a good starting point for a larger screen. Nevertheless, the current test shows that a medium number of mutations can pinpoint important domains within the Bdf1 protein which are more likely to indicate a specific function. One could envision that in parallel with the mutagenesis approach that involves the entire Bdf1 ORF, smaller screens that are targeted towards specific domains could be easily employed, and would allow for a better characterization of definite Bdf1 regions.

In order to identify unique residues in Bdf1 that might offer insights into its specialized role in splicing, it is important to examine these mutations in the context of Bdf1's protein domains which undoubtedly underlie its modularity. Bdf1, as a member of the BET family [34] contains two N-terminal bromodomains and a C-terminal acidic tail. Within this C-terminal tail, it possesses a region of conservation with other

BET members known as a B2 box (see figure 3.3A), a NET domain (N-terminal ET domain) followed by a less conserved intervening sequence and a distal SEED domain. All of these domains are conserved with Bdf2 which is a member of the same protein family. However, the residues within some of these domains are highly dissimilar and moreover, there are regions in Bdf1 which are absent from Bdf2 (see figure 3.3B). Therefore, when searching for mutations which confer specific Bdf1 functions, the lack of conservation with Bdf2 and the high conservation with Bdf1 homologues from other yeast species can be highly relevant. Given that negative strains which contain all the mutations in a candidate strain and lack only the ones which are important for Bdf1's function in splicing are unlikely to be encountered, a better approach is to search for hotspots within the protein which are more likely to exhibit a splicing defect when mutagenized.

The mutants identified in this screen which cause an upregulation in U3 snoRNA precursor levels higher than 2-fold map primarily to two regions of the Bdf1 protein – the N-terminal coiled coil which lies upstream of the bromodomains and the C-terminus past the second bromodomain. A single mutation of questionable importance maps to the first bromodomain. Only a couple of mutations in the candidate strains map to the N-terminal coiled coil - S24N in the K5 mutant and D96E in the K17 mutant, whereas the M15 mutant displays all 3 mutations in the C-terminal region. It could be argued that the D96E mutation doesn't cause a major change in the protein environment, since it exchanges a negatively charged amino acid with one of the same charge which is only slightly larger. Moreover, it occurs in a residue of Bdf1 which is not highly conserved. While no such argument can be made for S24N, the N-terminal

coiled coil appears to be more permissive to mutation than other Bdf1 domains. The evidence for this is that several of the mutations in the negative strains which do not lead to a significant upregulation of U3 precursor levels occur at a higher rate within this coiled coil region, and that these mutations are as likely to occur in very conserved residues (5/5 yeast species, e.g E41V in mutant D8) as in less conserved ones (G76D in mutant M13). In addition, the fact that the M15 mutant does not harbor any mutation in this domain but still significantly upregulates U3 precursor levels suggests that this is not the major domain responsible for Bdf1's role in splicing.

Four of the mutations in the confirmed candidates are located in the NET domain. This domain is conserved in all BET proteins (for example: human RING3, murine MCAP, Drosophila *fsh* (for review see [34, 35]) and is relatively well conserved between Bdf1 and Bdf2, though not at the extent of the conservation seen at the level of the bromodomains. The NET domain has a helix-helix loop-extended beta sheet-helix (H-H-L-E-H) structure, although in the case of Bdf1 the beta sheet is quite short. While the mutations I551V (in K5) and I555L (in K17), which occur in the second helix of this domain, are less likely to be linked to the splicing phenotype because they are in residues that are identical in Bdf1 and Bdf2, the H578Q mutation occurs in the second helix of the Net domain, but in a region where Bdf1 and Bdf2 are dissimilar. However, it is unclear if this mutation is necessary for the upregulation of U3 precursor levels since it occurs in a non-conserved residue.

So far, a common denominator for the three mutants which mimic the *bdf1Δ* phenotype is that they lead to the modification of the very C-terminal end of the protein, usually past amino acid 600, either through the insertion of a premature

termination codon or through frameshift. This region is important because it encompasses the C-terminal phosphorylation domain (known as CPR, amino acids 660-679) in Bdf1 as described by Sawa *et al*. [23] and also the terminal SEED domain. The SEED domain is enriched in negatively charged amino acids and serine residues and is believed to be a putative ATP binding motif. Moreover, the SEED domain is distinct in Bdf2. Therefore, this region is interesting threefold: 1) it falls within the region of interaction with Taf7 (amino acids 500-686); 2) it overlaps with an area of Bdf1 modification (phosphorylation) whose importance for Bdf1 function is unknown and 3) part of the CPR domain overlaps with a deletion in Bdf2, while the SEED sequence is highly conserved in both Bdf1 and Bdf2 homologues but not between Bdf1 and Bdf2. An attractive hypothesis which remains to be tested is that this terminal region may be sufficient to account for Bdf1's role in pre-mRNA splicing. An easy test of this hypothesis would be to construct plasmid versions of the candidate strains which retain all other mutations, but for which the non-sense mutations or the frameshift are reversed, and to determine if the accumulation in U3 precursor levels still persists.

While not a mutation hotspot which correlates with significant increases in U3 precursor levels, the first bromodomain contains a mutation A208T in the coil between helix A and B in the K5 mutant. Moreover, the A1 mutant which causes 70% increase in U3 precursor levels compared to wild-type contains an A245D mutation in helix C of BD1. It will remain to be seen if mutations in this bromodomain are necessary or sufficient (in the case of the A1 mutant) to cause a splicing defect and if they correlate with any charge change in the binding pocket of the acetyl lysine. Since the 2-fold

increase in U3 precursor levels is somewhat arbitrary, a different assay, such as splicing sensitive microarrays will be required to fully characterize the mutations which are sufficient to cause splicing defects and test the magnitude of these defects.

In contrast, the mutations in the negatives either cluster in the N-terminal coiled coil where they equally affect conserved or non-conserved residues, or predominantly occur in regions of lower Bdf1 conservation. Some of these mutations also occur in regions where Bdf1 and Bdf2 are highly conserved, where they may presumably impact the transcription function which is common to Bdf1 and Bdf2. Moreover they tend not to change the hydrophobicity or relative charge of the original residue, or can lead to premature stop codons in the last but one amino acid, which is likely too distal to affect Bdf1's interaction with a target factor. While these guidelines are generally true, they cannot account for every amino acid change which occurs within these mutants. As before, in the case of the confirmed candidates, additional proof will be required to definitively declare that a specific residue is unimportant for Bdf1's role in splicing. Since a combination of mutations can lead to distinct phenotypic outcomes, our approach is more likely to pinpoint protein domains which are important for this function, rather than ones that are not. Nevertheless, at present, the C-terminal region encompassing the CPR and the SEED domain appears the most likely candidate for connecting Bdf1's transcription role with splicing. In the future, it will be important to further address the mechanism behind Bdf1's role in splicing. To this end, identifying the interactions which are missing in a Bdf1 mutant which causes a global splicing defect versus a wild-type, will represent a first step in defining more targeted interaction domains in Bdf1.

REFERENCES

1. Tong, A. H., Evangelista, M., Parsons, A. B., Xu, H., Bader, G. D., Page, N., . . . Boone, C. (2001). Systematic genetic analysis with ordered arrays of yeast deletion mutants. *Science (New York, N.Y.), 294*(5550), 2364-2368. doi: 10.1126/science.1065810

2. Schuldiner, M., Collins, S. R., Thompson, N. J., Denic, V., Bhamidipati, A., Punna, T., . . . Krogan, N. J. (2005). Exploration of the function and organization of the yeast early secretory pathway through an epistatic miniarray profile. *Cell, 123*(3), 507-519. doi: 10.1016/j.cell.2005.08.031

3. Costanzo, M., Baryshnikova, A., Bellay, J., Kim, Y., Spear, E. D., Sevier, C. S., . . . Boone, C. (2010). The genetic landscape of a cell. *Science (New York, N.Y.), 327*(5964), 425-431. doi: 10.1126/science.1180823; 10.1126/science.1180823

4. Collins, S. R., Kemmeren, P., Zhao, X. C., Greenblatt, J. F., Spencer, F., Holstege, F. C., . . . Krogan, N. J. (2007). Toward a comprehensive atlas of the physical interactome of saccharomyces cerevisiae. *Molecular & Cellular Proteomics : MCP, 6*(3), 439-450. doi: 10.1074/mcp.M600381-MCP200

5. Ghaemmaghami, S., Huh, W. K., Bower, K., Howson, R. W., Belle, A., Dephoure, N., . . . Weissman, J. S. (2003). Global analysis of protein expression in yeast. *Nature, 425*(6959), 737-741. doi: 10.1038/nature02046

6. Matangkasombut, O., Buratowski, R. M., Swilling, N. W., & Buratowski, S. (2000). Bromodomain factor 1 corresponds to a missing piece of yeast TFIID. *Genes & Development, 14*(8), 951-962.

7. Durant, M., & Pugh, B. F. (2007). NuA4-directed chromatin transactions throughout the saccharomyces cerevisiae genome. *Molecular and Cellular Biology, 27*(15), 5327-5335. doi: 10.1128/MCB.00468-07

8. Martinez-Campa, C., Politis, P., Moreau, J. L., Kent, N., Goodall, J., Mellor, J., & Goding, C. R. (2004). Precise nucleosome positioning and the TATA box dictate requirements for the histone H4 tail and the bromodomain factor Bdf1. *Molecular Cell, 15*(1), 69-81. doi: 10.1016/j.molcel.2004.05.022

9. Jeanmougin, F., Wurtz, J. M., Le Douarin, B., Chambon, P., & Losson, R. (1997). The bromodomain revisited. *Trends in Biochemical Sciences, 22*(5), 151-153.

10. Zeng, L., & Zhou, M. M. (2002). Bromodomain: An acetyl-lysine binding domain. *FEBS Letters, 513*(1), 124-128.

11. Pamblanco, M., Poveda, A., Sendra, R., Rodriguez-Navarro, S., Perez-Ortin, J. E., & Tordera, V. (2001). Bromodomain factor 1 (Bdf1) protein interacts with histones. *FEBS Letters, 496*(1), 31-35.

12. Matangkasombut, O., & Buratowski, S. (2003). Different sensitivities of bromodomain factors 1 and 2 to histone H4 acetylation. *Molecular Cell, 11*(2), 353-363.

13. Ladurner, A. G., Inouye, C., Jain, R., & Tjian, R. (2003). Bromodomains mediate an acetyl-histone encoded antisilencing function at heterochromatin boundaries. *Molecular Cell, 11*(2), 365-376.

14. Kurdistani, S. K., Tavazoie, S., & Grunstein, M. (2004). Mapping global histone acetylation patterns to gene expression. *Cell, 117*(6), 721-733. doi: 10.1016/j.cell.2004.05.023

15. Krogan, N. J., Keogh, M. C., Datta, N., Sawa, C., Ryan, O. W., Ding, H., . . . Greenblatt, J. F. (2003). A Snf2 family ATPase complex required for recruitment of the histone H2A variant Htz1. *Molecular Cell, 12*(6), 1565-1576.

16. Kobor, M. S., Venkatasubrahmanyam, S., Meneghini, M. D., Gin, J. W., Jennings, J. L., Link, A. J., . . . Rine, J. (2004). A protein complex containing the conserved Swi2/Snf2-related ATPase Swr1p deposits histone variant H2A.Z into euchromatin. *PLoS Biology, 2*(5), E131. doi: 10.1371/journal.pbio.0020131

17. Zhang, H., Roberts, D. N., & Cairns, B. R. (2005). Genome-wide dynamics of Htz1, a histone H2A variant that poises repressed/basal promoters for activation through histone loss. *Cell, 123*(2), 219-231. doi: 10.1016/j.cell.2005.08.036

18. Durant, M., & Pugh, B. F. (2006). Genome-wide relationships between TAF1 and histone acetyltransferases in saccharomyces cerevisiae. *Molecular and Cellular Biology, 26*(7), 2791-2802. doi: 10.1128/MCB.26.7.2791-2802.2006

19. Koerber, R. T., Rhee, H. S., Jiang, C., & Pugh, B. F. (2009). Interaction of transcriptional regulators with specific nucleosomes across the saccharomyces genome. *Molecular Cell, 35*(6), 889-902. doi: 10.1016/j.molcel.2009.09.011; 10.1016/j.molcel.2009.09.011

20. Scannell, D. R., Byrne, K. P., Gordon, J. L., Wong, S., & Wolfe, K. H. (2006). Multiple rounds of speciation associated with reciprocal gene loss in polyploid yeasts. *Nature, 440*(7082), 341-345. doi: 10.1038/nature04562

21. Giaever, G., Chu, A. M., Ni, L., Connelly, C., Riles, L., Veronneau, S., Johnston, M. (2002). Functional profiling of the saccharomyces cerevisiae genome. *Nature, 418*(6896), 387-391. doi: 10.1038/nature00935

22. Muhlrad, D., Hunter, R., & Parker, R. (1992). A rapid method for localized mutagenesis of yeast genes. *Yeast (Chichester, England), 8*(2), 79-82. doi: 10.1002/yea.320080202

23. Sawa, C., Nedea, E., Krogan, N., Wada, T., Handa, H., Greenblatt, J., & Buratowski, S. (2004). Bromodomain factor 1 (Bdf1) is phosphorylated by protein kinase CK2. *Molecular and Cellular Biology, 24*(11), 4734-4742. doi: 10.1128/MCB.24.11.4734-4742.2004

24. Arnold, K., Bordoli, L., Kopp, J., & Schwede, T. (2006). The SWISS-MODEL workspace: A web-based environment for protein structure homology modelling. *Bioinformatics (Oxford, England), 22*(2), 195-201. doi: 10.1093/bioinformatics/bti770

25. Bordoli, L., Kiefer, F., Arnold, K., Benkert, P., Battey, J., & Schwede, T. (2009). Protein structure homology modeling using SWISS-MODEL workspace. *Nature Protocols, 4*(1), 1-13. doi: 10.1038/nprot.2008.197; 10.1038/nprot.2008.197

26. Schwede, T., Kopp, J., Guex, N., & Peitsch, M. C. (2003). SWISS-MODEL: An automated protein homology-modeling server. *Nucleic Acids Research, 31*(13), 3381-3385.

27. Roy, A., Kucukural, A., & Zhang, Y. (2010). I-TASSER: A unified platform for automated protein structure and function prediction. *Nature Protocols, 5*(4), 725-738. doi: 10.1038/nprot.2010.5; 10.1038/nprot.2010.5

28. Zhang, Y. (2008). I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics, 9*, 40-2105-9-40. doi: 10.1186/1471-2105-9-40; 10.1186/1471-2105-9-40

29. Chua, P., & Roeder, G. S. (1995). Bdf1, a yeast chromosomal protein required for sporulation. *Molecular and Cellular Biology, 15*(7), 3685-3696.

30. Lygerou, Z., Conesa, C., Lesage, P., Swanson, R. N., Ruet, A., Carlson, M., . . . Seraphin, B. (1994). The yeast BDF1 gene encodes a transcription factor involved in the expression of a broad class of genes including snRNAs. *Nucleic Acids Research, 22*(24), 5332-5340.

31. Raisner, R. M., Hartley, P. D., Meneghini, M. D., Bao, M. Z., Liu, C. L., Schreiber, S. L., . . . Madhani, H. D. (2005). Histone variant H2A.Z marks the 5' ends of both active and inactive genes in euchromatin. *Cell, 123*(2), 233-248. doi: 10.1016/j.cell.2005.10.002

32. Raisner, R. M., & Madhani, H. D. (2008). Genomewide screen for negative regulators of sirtuin activity in saccharomyces cerevisiae reveals 40 loci and links to metabolism. *Genetics, 179*(4), 1933-1944. doi: 10.1534/genetics.108.088443; 10.1534/genetics.108.088443

33. Wu, W. H., Wu, C. H., Ladurner, A., Mizuguchi, G., Wei, D., Xiao, H., . . . Wu, C. (2009). N terminus of Swr1 binds to histone H2AZ and provides a platform for subunit assembly in the

chromatin remodeling complex. *The Journal of Biological Chemistry, 284*(10), 6200-6207. doi: 10.1074/jbc.M808830200; 10.1074/jbc.M808830200

34. Florence, B., & Faller, D. V. (2001). You bet-cha: A novel family of transcriptional regulators. *Frontiers in Bioscience : A Journal and Virtual Library, 6*, D1008-18.

35. Wu, S. Y., & Chiang, C. M. (2007). The double bromodomain-containing chromatin adaptor Brd4 and transcriptional regulation. *The Journal of Biological Chemistry, 282*(18), 13141-13145. doi: 10.1074/jbc.R700001200

CHAPTER 4

MUD1 AUTOREGULATES ITS EXPRESSION BY A DIFFERENT MECHANISM THAN ITS HUMAN HOMOLOGUE U1A

4.1 Introduction

The spliceosome is a massive macromolecular machine responsible for the excision of introns and ligation of exons from eukaryotic precursor mRNA substrates. At twice the size of the ribosome, this complex consists of the 5 small nuclear RNAs (snRNAs) U1, U2, U4, U5 and U6 and between 150 and 300 proteins [1]. Substrate recognition and catalysis of the two transesterification reactions central to pre-mRNA splicing are the result of the careful interplay between the snRNAs which are packaged with core spliceosomal proteins in snRNP particles and the numerous non-snRNP proteins.

The dynamics of the spliceosome are extensive. In order for a single transcript to be processed, the splicing machinery must gradually assemble onto the substrate. The first snRNP to recognize the substrate through base pairing interactions with the 5' splice site is U1 which leads to the formation of the commitment complex, followed by the ATP-dependent recognition of the branch point adenosine by U2snRNP (see figure 1.1 Introduction). Next, the tri-snRNP consisting of U4, U5 and U6 snRNPs assembles and leads to rearrangements which will allow for the formation of the catalytically active U2-U6snRNP core [2, 3] and lead to the exclusion and recycling of the U1 and U4 snRNPs. Subsequent reorganization of the catalytically active spliceosome is required for the transitioning through the two transesterification

reactions as well as for the final recycling of the spliceosomal subunits. Considering this requirement for very swift and precise dynamics, it is crucial that each step of the reaction if very carefully controlled. To this end, a breadth of RNA-dependent ATPases and helicases are recruited to ensure that the numerous rearrangements lead to the correct processing of the pre-mRNA substrate.

Given the absolute need for very precise regulation, it is not surprising that this regulation extends to the level of the individual spliceosomal subunits. Such an example is the U1snRNP protein U1A in higher systems, which binds loop B on U1snRNA. In mammals, U1A has been shown to down-regulate its expression through a polyadenlylation inhibition (PIE) element located in its 3'UTR (see figure 4.1) [4].

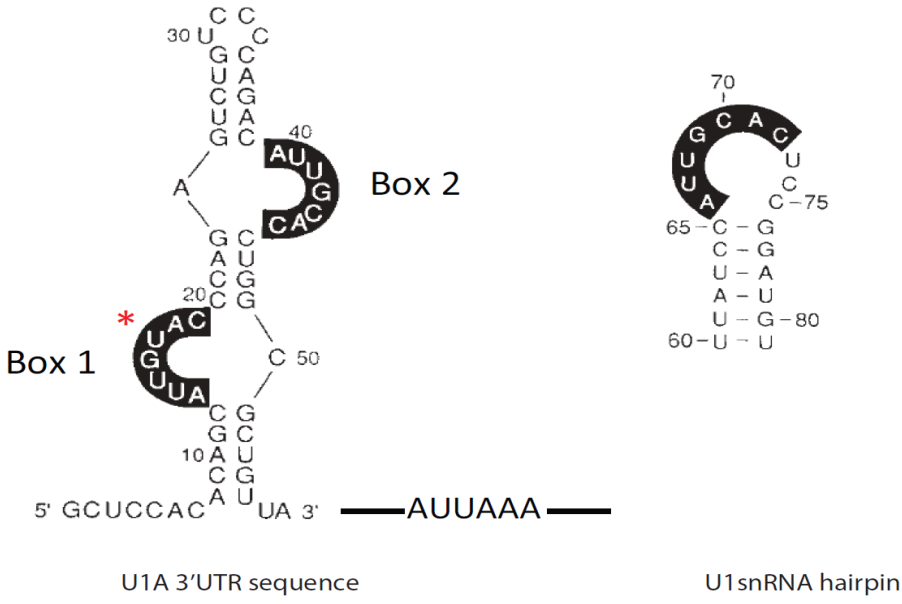

Figure 4.1: Comparison of the U1A 3'UTR sequence with the U1snRNA binding hairpin which is recognized by U1A (adapted from Oubridge et.al [13])

This bipartite element contains one box that is identical to the binding site of U1A on

loop B (Box2) of the U1snRNP and a second of high sequence similarity. The N-terminal RNA recognition motif (RRM) of U1A binds the PIE element [5] and inhibits polyadenylation through a direct interaction with the C-terminal 20 residues of the mammalian, but not yeast, polyA polymerase [6,7]. Furthermore, U1A contacts the PIE element upon dimerization through residues 102-115 [7-9] and it was shown that the binding surface has to be accessible on the same side of the RNA. Evidence from several NMR [10,11] and X-ray crystallography studies [12,13] supports this model, by which U1A binding to the PIE element induces a conformational fit which favors the formation of a dimerization surface.

Although the mechanism through which U1A fine-tunes its expression has been thoroughly investigated for higher systems, such a mechanism has not yet been reported in budding yeast. The yeast homologue, Mud1 (Mutant U1 Die) was discovered through an enhancer screen for mutants which exhibited synthetic lethal phenotypes in the absence of a functional U1snRNA [14]. Mud1 is a non-essential protein which was shown to bind U1snRNA and is hypothesized to participate in the proper folding of this RNA. Mud1, like its mammalian counterpart U1A, has 2 RRMs and shows high conservation at the protein level with its homologue, especially at the level of the C-terminal RRM (36% identity, 64% similarity). While the deletion of Mud1 does not affect cellular growth rate, the *mud1-1* mutant which truncates the protein at amino acid 84 leads to faster migrating commitment complexes and decreases their formation, as well as inhibits the splicing of inefficiently spliced reporters. Since the *mud1-1*truncation is in the middle of the first RRM, the *mud1-1* mutation is thought to lead to little or no functional Mud1 protein.

Interestingly, while the secondary structure of U1 snRNA is conserved between yeast and higher systems, at the sequence level the differences are more pronounced. Thus, in *S. cerevisiae*, U1snRNA contains a fungal domain which is conserved among yeast species (Figure 4.2) [15].

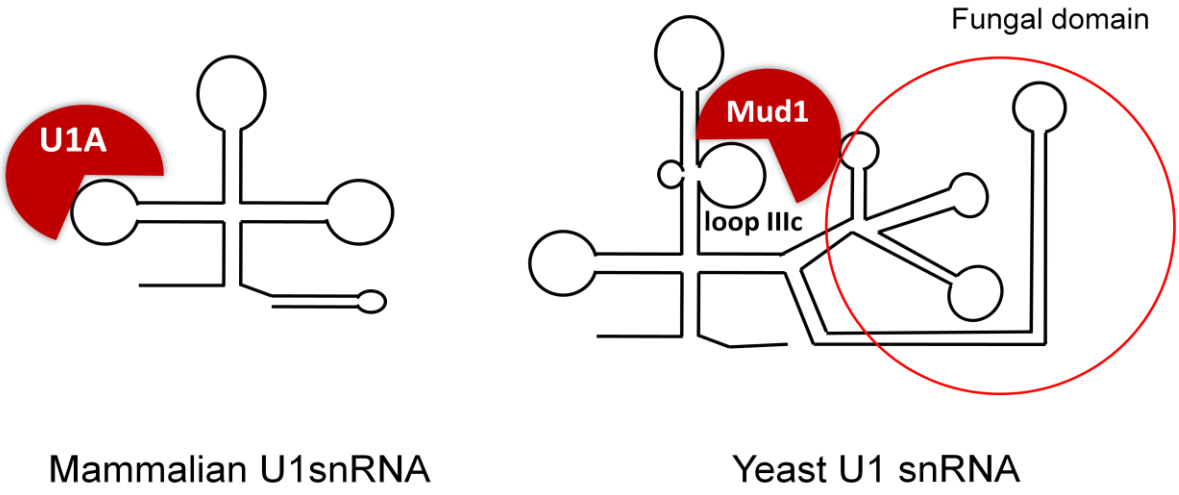

Figure 4.2: Schematic of the binding of U1A to the mammalian, versus the yeast U1 snRNA

Moreover, there is a lack of conservation between the yeast and mammalian loop B, the corresponding yeast counterpart having a longer and more varied stem-loop structure. By using *in vivo* dimethyl sulfate (DMS) footprinting, Tang *et al*. [16], showed that the N-terminal RRM of Mud1 binds loop IIIC of the yeast U1snRNP. Furthermore, they showed that deletion of the Mud1 C-terminal RRM which has no known function in higher systems impairs splicing in a CUP1 reporter system. This suggests that in yeast, the conserved C-terminal domain might play an independent role in splicing which is supported by evidence showing that the mammalian C-terminal RRM cannot complement this function. Therefore, while the binding of U1A and Mud1 to U1snRNP through the N-terminal RRM is a conserved process, the exact

sequences which are used differ between organisms. Moreover, the C-terminal domain appears to have an additional function in yeast.

Here I report that mutations in the cleavage and polyadenylation factors *cft2-ts* and *yth1-ts* lead to misregulation of Mud1 levels in *Saccharomyces cerevisiae*, suggesting a defect in the 3'end processing and decay of these species. This is not a general phenotype caused by mutations in the cleavage and polyadenylation machinery, since mutations in another 3'end cleavage and polyadenylation factor – *fip1-ts* have no effect on Mud1 levels. In addition, I show that overexpression of a cDNA version of Mud1 causes down-regulation of endogenous Mud1 levels. However, upon replacement of the endogenous 3'UTR, no misregulation of Mud1 levels is observed, suggesting that the 3'UTR is not sufficient for this regulation. Therefore, unlike in higher systems, the mechanisms of Mud1 autoregulation cannot be limited to an element in its 3'UTR. Since Mud1 is an intron containing gene, and there is ample evidence of coupling between CPSF and splicing of the last intron [17-24], I considered that Mud1 might represent a good model to study this coupling in budding yeast. Below, I discuss my findings in connection with new data from Madhura Raghavan who is investigating this regulation further.

4.2 Materials and Methods

4.2.1 Plasmids and strains

Temperature sensitive alleles from the Hieter collection [25], the *brr5-1* mutant [26], the Tap-tagged version of Mud1 [27] and the wild type strain yBY4741 [28] were used. Plasmids containing different versions of Mud1 were cloned into pRS416 (CEN) or pRS426 (2 micron) URA marked vectors using standard cloning procedures. The cDNA version of Mud1 contains ~200 bp in the 5'UTR of the gene and ~600bp of the 3'UTR, and was obtained by stitching PCR fragments to omit the intron. Unless otherwise specified, the strains lacking a plasmid were grown in rich YPD media supplemented with 2% glucose [29] at permissive temperature (25 or 30°C). The strains harboring plasmids were grown in SD-Ura media at 25°C and compared to empty vector controls grown under the same conditions.

4.2.2 RNA extraction, cDNA synthesis and qPCR

RNA extraction, DNase treatment and cDNA synthesis and qPCR were performed as previously described in chapter 3. The primers used for detecting Mud1 total and Mud1 precursor levels as well as endogenous Mud1 (TAP) levels are shown in Table 4.1

Table 4.1: Forward and reverse primer sequences used for amplifying the target gene in qPCR

The list of forward and reverse primers used:

Target	Forward primer	Reverse primer
Mud1 total	AAT AAG ACG CCT GAA GCA TAA GCT	CTA TCA ATC TCA GCC TCC TCT ACC TT
Mud1 precursor	AAAAAGCTGCTATCAACAAAAGACG	AAGATTTAAGGAAACGTAAATTAC
Faa1 total	TCT GCC CTA TGC TTA TTG GTT ACG	TAA CAC AAC CTG TTA GGT CAC CAG C
Mud1-TAP	TCGTAACCTAGCTTTCGTGGAATAC	CTGCTGAGACGGCTATGAAATTCTT

4.3 Results

4.3.1 Mud1 levels are misregulated in 3'end processing mutants

In the genome-wide screen described in chapter 2, I identified *cft2-ts* and *yth1-ts*, two components of the cleavage and polyadenylation CPSF complex which cause targeted splicing defects. Interestingly, upon genome-wide characterization of these mutants by means of splicing sensitive microarrays, Mud1 was the only transcript to exhibit a very pronounced misregulation in transcript levels, where all measured species (total, precursor or mature) displayed an 8-16 fold increase (Figure 4.3A). Even though increases in Mud1 levels are shown in figure 4.3A at the non-permissive temperature of 37°C, this misregulation is also apparent at the permissive temperature of 25°C or under cold shock (Figure 4.3B). These phenotypes do not correlate with the robustness of strain growth. Figure 2.14 shows that while *cft2-ts* is sick at 16°C, the growth of *yth1-ts* is not significantly different from wild-type.

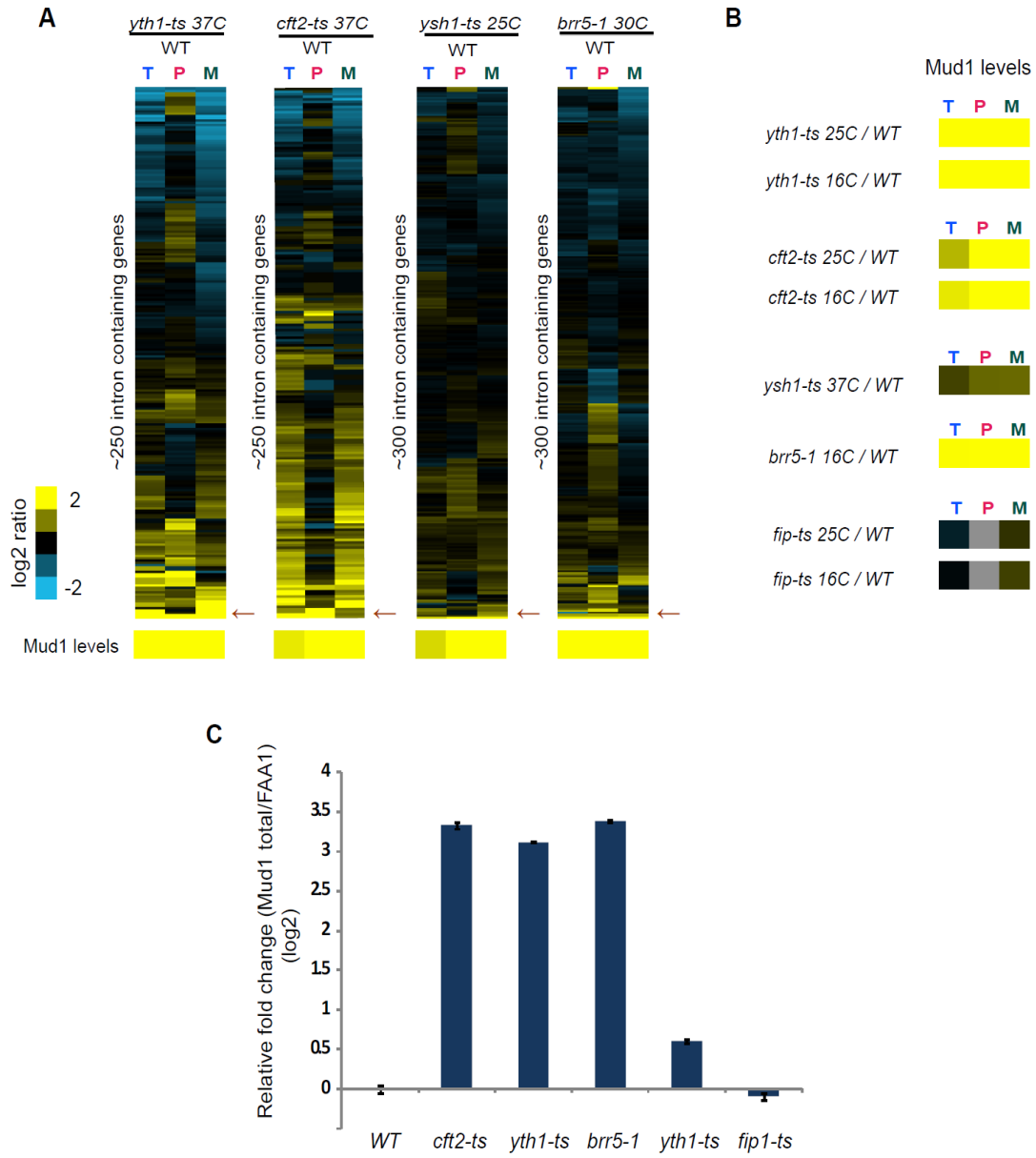

Figure 4.3: Mutations in cleavage and polyadenlyation factors cause Mud1 misregulation

A. Splicing sensitive microarrays for the cleavage and polyadenylation factor mutants *yth1-ts*, *cft2-ts*, *ysh1-ts* and *brr5-ts* and blow-up of the Mud1 levels shown below each array. Each of the arrays was clustered independently. The red arrows indicate the position of Mud1 on each of the arrays.

B. Blow-up of array data for cleavage and polyadenlyation mutants at either the permissive or non-permissive temperatures

C. q(RT)PCR assessing the expression levels of the total Mud1 transcript relative to a wild-type control. All strains were grown at 25°C. The values are reported on a log2 scale.

Yth1-ts is essential in yeast and is the homologue of mammalian CPSF30 [30] to which it is 40% identical. Mutants in different regions of Yth1 have been shown to be defective in both cleavage [31] and polyadenylation. Yth1 physically interacts with other members of CPSF such as Ysh1 and Fip1 through its C-terminal domain [31, 32] and binds pre-mRNA close to the cleavage site. Similarly, Cft2 is the homologue of mammalian CPSF100 and binds to uncleaved substrates in an ATP-dependent fashion [33]. Mutations in Cft2 lead to the inhibition of both the cleavage and polyadenylation reactions [34, 35]. Cft2 is generally involved in providing interaction surfaces within the CPSF complex and it also interacts with the CTD of RNA polymerase II. Moreover, in yeast, deletion of the RNAPII CTD was shown to lead to a decrease in cleavage efficiency [36, 37].

Since I observed misregulation in Mud1 levels in two CPSF factors, I wanted to determine if this was a general phenotype observed upon mutation of the cleavage and polyadenylation machinery. Thus, I chose to monitor Mud1 expression levels in other cleavage and polyadenylation mutants: *fip1-ts* and *ysh1-ts/brr5-1*. Fip1 functions as an anchor between the CPSF complex (through Yth1) and the polyA polymerase (Pap1) [38, 39] and controls polyadenylation. In contrast, *ysh1*-ts and *brr5-1* are mutants in the same CPSF factor, the yeast homologue of mammalian CPSF73. Ysh1 is the most conserved factor of the CPSF complex at 53% identity with its bovine homologue [40] and it is still debated if it represents the endonuclease within the complex [41]. The *ysh1-ts* allele used in this study was created by the Hieter lab and is lethal at 34°C. *brr5-1* was first discovered as a cold sensitive splicing mutant which is required for

3'end processing [42] and fails to grow at 16℃ (see Figure 2.14).

Mud1 levels are indeed up-regulated in the *ysh1-ts* and *brr5-1* mutants at the permissive temperature (25℃ for *ysh1-ts* and 30℃ for *brr5-1*) as shown in figure 4.3A. Similar to the phenotypes seen for *cft2-ts* and *yth1-ts*at 37°C, the Mud1 transcript is the only one which displays such a profound misregulation and clusters independently at the bottom of the array. In addition, *brr5-1* displays a ~16-fold increase in Mud1 levels at the non-permissive temperature of 16℃, whereas the *ysh1-ts*allele shows an increase of only about 2-fold at 37°C. Nevertheless, in contrast to the three CPSF mutants I analyzed, *fip1-ts* doesn't cause any change in total or mature Mud1 transcript levels. The data suggests that while not all mutations in components of the cleavage and polyadenylation machinery impact Mud1 transcript levels, mutations in CPSF components in particular might cause misregulation of Mud1 expression. However, M. Raghavan has shown that a mutation in *pta1-ts,* which is the fourth factor of CFII together with Cft1, Cft2 and Ysh1 [43], does not cause misregulation of Mud1 levels (data not shown). While it cannot be excluded that the phenotypes are allele specific, this indicates that specific mutations in cleavage and polyadenylation factors impact Mud1 levels.

It is important to note that the 8-16 fold increase in Mud1 levels reported in the *cft2-ts, yth1-ts, ysh1-ts* or *brr5-1* strains partially reflects the upper limit of detection for the microarrays. To get a more accurate determination of the change in transcript levels and to verify these results, I used small scale quantitative real-time PCR (qPCR). Mud1 total levels were increased 8-10 fold in the *cft2-ts, yth1-ts* and *brr5-1* backgrounds (Figure 4.3C). In my hands, *ysh1-ts* did not show more than a 50%

increase in total levels. However, M. Raghavan reported at least a 3-fold increase in Mud1 total levels for the same experiment. As expected, *fip1-ts* did not cause any misregulation in Mud1 expression.

4.3.2 Mud1 overexpression regulates the endogenous copy

Since the mammalian homologue of Mud1, U1A, is capable of autoregulating its expression, I asked whether expression of Mud1 from an ectopic location could influence the expression of Mud1 in its chromosomal context. To test this, I constructed a cDNA version of Mud1 on a high copy vector and expressed it in a Mud1-Tap strain. The Mud1-Tap strain expresses a Tap tagged version of Mud1 and contains the endogenous Mud1 5'UTR and the Mud1 intron, but has an ADH1 3'UTR. Overexpression of an exogenous Mud1 cDNA within this strain increased the total Mud1 levels above 35-fold (Figure 4.4) and caused a ~2-fold downregulation of the endogenous Mud1 precursor and total levels (measured as Mud1-Tap). Curiously, the splicing efficiency of the endogenous transcript increased by ~50% (not shown). In addition, M. Raghavan's data indicates that Mud1 overexpression also leads to a 2-fold decrease in Mud1-Tap tagged protein. Taken together, these results support a model in which an excess of Mud1 protein is capable of decreasing the endogenous production of Mud1, so that the Mud1 homeostasis is maintained.

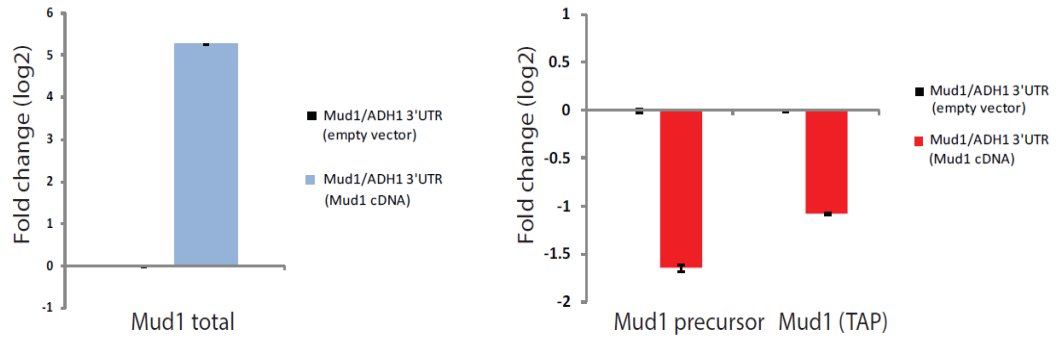

Figure 4.4: Mud1 overexpression down-regulates the chromosomal Mud1 copy

Mud1 total (left) and Mud1 precursor and Mud1 (TAP) endogenous levels measured by qPCR. The values are normalized to an empty vector control and presented in log2 space. A representative experiment is shown and the error bars indicate standard deviation of technical replicates

4.3.3 The Mud1 3'UTR is not sufficient for regulation

To test the similarity between U1A's and Mud1's mechanisms of autoregulation, I asked if Mud1 levels were controlled through a negative regulatory element in the Mud1 3'UTR. Unlike in the case of the human U1A, I was not able to detect a sequence element in the Mud1 3'UTR reminiscent of the U1snRNA loop sequence which Mud1 binds as part of the U1snRNP. In addition, the Mud1-Tap endogenous copy was repressed upon Mud1 overexpression, even though it lacked its native 3'UTR. This suggests that the mechanism of Mud1 regulation might be independent of its 3'UTR. To test this prediction, I analyzed Mud1 expression levels by qPCR in the Mud1-Tap strain. In this strain the endogenous Mud1 3'UTR is replaced with the ADH1 3'UTR, whereas the endogenous Mud1 3'UTR is shifted about 3kb downstream of the STOP codon and thus is believed not to be functional. If the 3'UTR is indeed sufficient for regulation, one would expect to see the same misregulation in Mud1

levels in the Mud1-Tap strain as in the 3'end mutants. Mud1 total levels were unchanged from the wild-type control in the Mud1-Tap background (Figure 4.5). Moreover, a Blast analysis identified no similarity between the Mud1 and Adh1 3'UTRs and M. Raghavan's data suggests that the Tap-tag does not interfere with Mud1 processing (data not shown). In conclusion, the endogenous 3'UTR does not appear to be important for Mud1 regulation.

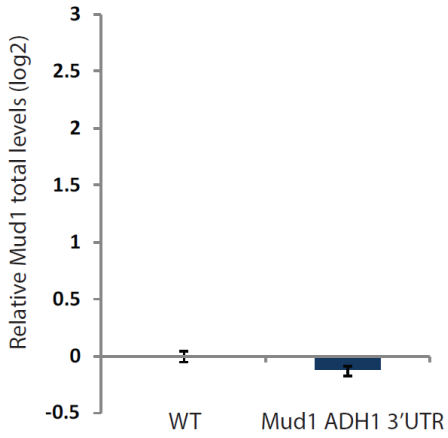

Figure 4.5: The absence of the native Mud1 3'UTR does not misregulate Mud1 levels
Mud1 total levels in log 2 space as measured by qPCR. The Mud1ADH1 3'UTR levels are normalized to the WT (yBY4741) strain.

4.3 Discussion

The model through which U1A inhibits its own expression is one of the best characterized models of regulations in mammals. Here I present evidence that Mud1, the yeast homologue of U1A is also subject to autoregulation, though the mechanism through which this happens is currently unknown. My first observation, that Mud1 levels are misregulated in 3'end processing mutants (Figure 4.3) suggested that Mud1

might use a similar mechanism to its human counterpart. However, it soon became apparent that Mud1 lacks sequences in its 3'UTR that resemble the binding sequence Mud1 binds on U1snRNA, as is the case with U1A. Moreover, a chromosomal version of Mud1-Tap which has the 3'UTR replaced with an ADH1 3'UTR can be downregulated at the RNA and protein level upon Mud1 overexpression (Figure 4.4). Since the two 3'UTR sequences are very dissimilar from one another, this suggested that the negative regulatory sequence lies outside the 3'UTR. Lastly, the Mud1-Tap strain by itself did not show any increase in Mud1 total levels (figure 4.5), as did the *cft2-ts* and *yth1-ts* mutants, further indicating that the sought negative regulatory element is not in this region.

Since Mud1 does not appear to share the same regulatory mechanism with U1A, I considered an alternative model in which Mud1 might be a sensitized transcript which would allow us to study splicing-3'end connections. There is plentiful evidence about the interplay between splicing of the last intron and 3'end processing. Early studies in mammalian systems showed that the presence of an upstream intron with a functional 3' splice site enhanced polyadenylation [22] and conversely, mutations in the polyadenylation signals inhibited splicing of the upstream intron [23]. More recently, several physical interactions between spliceosomal subunits and the cleavage and polyadenylation machinery have been described. Thus, the UAF65 physically interacts with CFmI59 [21] to promote cleavage and polyadenylation *in vitro*, while the presence of the wild-type but not mutated polyadenylation signal stimulates U2AF65 recruitment to the polypyrimidine tract of the upstream intron [24]. Furthermore, the 25kDa subunit of CFI coimmunoprecipitates with U1 snRNP protein 70K [17],

whereas the N-terminus of U1A interacts with the 160kDa subunit of CPSF and stabilizes its interaction with the polyA signal sequence *in vitro* [20]. Also, data from the Keller lab indicates a direct interaction between U2snRNP and CPSF in mammals [19]. Finally, I have identified several mutations in cleavage and polyadenylation components in yeast which cause targeted splicing defects (Chapter 2), while the Keller lab reported a *ysh1-12* allele defective in splicing [44]. All these lines of evidence underline the extensive coupling between splicing and 3'end processing.

Mud1 is a very good substrate to study this coupling because it is an intron containing gene with a suboptimal intron. Within the 89bp Mud1 intron, the branch point is unusually located in the middle of this sequence. Genome-wide ChIP-chip data [45] also indicates that U2 is suboptimally recruited to this branchpoint, which given the coupling between splicing and CPSF, would make Mud1 intron more sensitive to any variation in CPSF recruitment. One could easily imagine that the absence of a functional CPSF which has mutant versions of Cft2, Yth1 or Ysh1 would inhibit recruitment of U2snNRP. In turn, spliceosome assembly would be hindered, since the U1 snRNP recruited at the 5' splice site would have no U2 to interact with. The Vagner *et al.* study [24], showed that at least in mammals, U1 snRNP is not involved in the recruitment of U2AF to the polypyrimidine tract. In addition, two studies have shown that in mammalian systems, U1snRNP depletion can cause premature cleavage and polyadenylationin introns [46, 47]. These studies work under the hypothesis that U1 can base pair to cryptic sites within the genome and protect the transcript from usage of cryptic polyadenylation signals (PAS). However, polyA site recognition is inhibited by the presence of a 5'splice site [48], and it is thought this inhibition occurs

through an interaction between U1 70K and PAP [49]. Thus, an alternative hypothesis is that if a functional spliceosome can be assembled, then the U1 to U2 "hand-off" can protect the intron from premature cleavage and polyadenylation. If early spliceosome assembly cannot occur, then the cleavage and polyadenylation machinery can act upon the substrate and cause cleavage at cryptic sites. This model is in agreement with the toolbox model of the RNA polymerase II CTD which interacts with U1 and CPSF [50, 51]. The Mud1 transcript was a particularly attractive model to test this hypothesis, since PolyA-Seq data [52] suggests that there are several sites in Mud1 intron which can lead to premature termination and thus function to down-regulate its expression. Nevertheless, this model was hindered by two observations: 1) that the 3'UTR is not required for Mud1 regulation and 2) that the intron is not necessary either. M. Raghavan has shown that a Mud1 intronless version containing the native 3'UTR expressed at 1-2 copies/cell in a *mud1Δ* strain causes no misregulation of Mud1 levels, while the same plasmid in a *mud1Δ yth1-ts* background can entirely recapitulate the misregulation seen in the *yth1-ts* or *cft2-ts* background. This suggests that the intronic sequence does not contain the element required for Mud1 repression.

If the Mud1 element required for its regulation is not present in the 3'UTR or in the intron where is it located? The creation of an intronless Mud1 ADH1 3'UTR that can still recapitulate the misregulation in the *mud1Δ yth1-ts* background will still be required and would suggest that the regulatory element is contained fully within the remainder of the Mud1 sequence, including the 5'UTR. Moreover, given the polyA-seq data which suggests that cleavage and polyadenylation occurs within the intron, at a time perhaps prior to exon 2 synthesis, it is likely that exon 2 does not contain the

regulatory element. Nevertheless, the way this regulation is achieved and its interplay with CPSF remain to be discovered.

REFERENCES

1. Wahl, M. C., Will, C. L., &Luhrmann, R. (2009). The spliceosome: Design principles of a dynamic RNP machine. *Cell, 136*(4), 701-718. doi: 10.1016/j.cell.2009.02.009

2. Valadkhan, S. (2007). The spliceosome: A ribozyme at heart? *Biological Chemistry, 388*(7), 693-697. doi: 10.1515/BC.2007.080

3. Valadkhan, S., & Manley, J. L. (2001). Splicing-related catalysis by protein-free snRNAs. *Nature, 413*(6857), 701-707. doi: 10.1038/35099500

4. Boelens, W. C., Jansen, E. J., van Venrooij, W. J., Stripecke, R., Mattaj, I. W., & Gunderson, S. I. (1993). The human U1 snRNP-specific U1A protein inhibits polyadenylation of its own pre-mRNA. *Cell, 72*(6), 881-892.

5. van Gelder, C. W., Gunderson, S. I., Jansen, E. J., Boelens, W. C., Polycarpou-Schwarz, M., Mattaj, I. W., & van Venrooij, W. J. (1993). A complex secondary structure in U1A pre-mRNA that binds two molecules of U1A protein is required for regulation of polyadenylation. *The EMBO Journal, 12*(13), 5191-5200

6. Gunderson, S. I., Beyer, K., Martin, G., Keller, W., Boelens, W. C., &Mattaj, L. W. (1994). The human U1A snRNP protein regulates polyadenylation via a direct interaction with poly(A) polymerase. *Cell, 76*(3), 531-541.

7. Gunderson, S. I., Vagner, S., Polycarpou-Schwarz, M., &Mattaj, I. W. (1997). Involvement of the carboxyl terminus of vertebrate poly(A) polymerase in U1A autoregulation and in the coupling of splicing and polyadenylation. *Genes & Development, 11*(6), 761-773.

8. Guan, F., Palacios, D., Hussein, R. I., & Gunderson, S. I. (2003). Determinants within an 18-amino-acid U1A autoregulatory domain that uncouple cooperative RNA binding, inhibition of polyadenylation, and homodimerization. *Molecular and Cellular Biology, 23*(9), 3163-3172.

9. Klein Gunnewiek, J. M., Hussein, R. I., van Aarssen, Y., Palacios, D., de Jong, R., van Venrooij, W. J., & Gunderson, S. I. (2000). Fourteen residues of the U1 snRNP-specific U1A protein are required for homodimerization, cooperative RNA binding, and inhibition of polyadenylation. *Molecular and Cellular Biology, 20*(6), 2209-2217.

10. Allain, F. H., Gubser, C. C., Howe, P. W., Nagai, K., Neuhaus, D., &Varani, G. (1996). Specificity of ribonucleoprotein interaction determined by RNA folding during complex formulation. *Nature, 380*(6575), 646-650. doi: 10.1038/380646a0

11. Varani, L., Gunderson, S. I., Mattaj, I. W., Kay, L. E., Neuhaus, D., &Varani, G. (2000). The NMR structure of the 38 kDa U1A protein - PIE RNA complex reveals the basis of cooperativity in regulation of polyadenylation by human U1A protein. *Nature Structural Biology, 7*(4), 329-335. doi: 10.1038/74101

12. Nagai, K., Oubridge, C., Jessen, T. H., Li, J., & Evans, P. R. (1990). Crystal structure of the RNA-binding domain of the U1 small nuclear ribonucleoprotein A. *Nature, 348*(6301), 515-520. doi: 10.1038/348515a0

13. Oubridge, C., Ito, N., Evans, P. R., Teo, C. H., & Nagai, K. (1994). Crystal structure at 1.92 A resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin. *Nature, 372*(6505), 432-438. doi: 10.1038/372432a0

14. Liao, X. C., Tang, J., &Rosbash, M. (1993). An enhancer screen identifies a gene that encodes the yeast U1 snRNPA protein: Implications for snRNP protein function in pre-mRNA splicing. *Genes & Development, 7*(3), 419-428.

15. Kretzner, L., Krol, A., &Rosbash, M. (1990). Saccharomyces cerevisiae U1 small nuclear RNA secondary structure contains both universal and yeast-specific domains. *Proceedings of the National Academy of Sciences of the United States of America, 87*(2), 851-855.

16. Tang, J., &Rosbash, M. (1996). Characterization of yeast U1 snRNP A protein: Identification of the N-terminal RNA binding domain (RBD) binding site and evidence that the C-terminal RBD functions in splicing. *RNA (New York, N.Y.), 2*(10), 1058-1070.

17. Awasthi, S., &Alwine, J. C. (2003). Association of polyadenylation cleavage factor I with U1 snRNP. *RNA (New York, N.Y.), 9*(11), 1400-1409.

18. Garas, M., Dichtl, B., & Keller, W. (2008). The role of the putative 3' end processing endonuclease Ysh1p in mRNA and snoRNA synthesis. *RNA (New York, N.Y.), 14*(12), 2671-2684. doi: 10.1261/rna.1293008; 10.1261/rna.1293008

19. Kyburz, A., Friedlein, A., Langen, H., & Keller, W. (2006). Direct interactions between subunits of CPSF and the U2 snRNP contribute to the coupling of pre-mRNA 3' end processing and splicing. *Molecular Cell, 23*(2), 195-205. doi: 10.1016/j.molcel.2006.05.037

20. Lutz, C. S., Murthy, K. G., Schek, N., O'Connor, J. P., Manley, J. L., &Alwine, J. C. (1996). Interaction between the U1 snRNP-A protein and the 160-kD subunit of cleavage-polyadenylation specificity factor increases polyadenylation efficiency in vitro. *Genes & Development, 10*(3), 325-337.

21. Millevoi, S., Loulergue, C., Dettwiler, S., Karaa, S. Z., Keller, W., Antoniou, M., &Vagner, S. (2006). An interaction between U2AF 65 and CF I(m) links the splicing and

3' end processing machineries. *The EMBO Journal, 25*(20), 4854-4864. doi: 10.1038/sj.emboj.7601331

22. Niwa, M., &Berget, S. M. (1991). Mutation of the AAUAAA polyadenylation signal depresses in vitro splicing of proximal but not distal introns. *Genes & Development, 5*(11), 2086-2095.

23. Niwa, M., Rose, S. D., &Berget, S. M. (1990). In vitro polyadenylation is stimulated by the presence of an upstream intron. *Genes & Development, 4*(9), 1552-1559.

24. Vagner, S., Vagner, C., &Mattaj, I. W. (2000). The carboxyl terminus of vertebrate poly(A) polymerase interacts with U2AF 65 to couple 3'-end processing and splicing. *Genes & Development, 14*(4), 403-413.

25. Ben-Aroya, S., Coombes, C., Kwok, T., O'Donnell, K. A., Boeke, J. D., &Hieter, P. (2008). Toward a comprehensive temperature-sensitive mutant repository of the essential genes of saccharomyces cerevisiae. *Molecular Cell, 30*(2), 248-258. doi: 10.1016/j.molcel.2008.02.021; 10.1016/j.molcel.2008.02.021

26. Noble, S. M., & Guthrie, C. (1996). Identification of novel genes required for yeast pre-mRNA splicing by means of cold-sensitive mutations. *Genetics, 143*(1), 67-80.

27. Ghaemmaghami, S., Huh, W. K., Bower, K., Howson, R. W., Belle, A., Dephoure, N., . . .Weissman, J. S. (2003). Global analysis of protein expression in yeast. *Nature, 425*(6959), 737-741. doi: 10.1038/nature02046

28. Giaever, G., Chu, A. M., Ni, L., Connelly, C., Riles, L., Veronneau, S., Johnston, M. (2002). Functional profiling of the saccharomyces cerevisiae genome. *Nature, 418*(6896), 387-391. doi: 10.1038/nature00935

29. Sherman, F. (2002). Getting started with yeast. *Methods in Enzymology, 350*, 3-41.

30. Barabino, S. M., Hubner, W., Jenny, A., Minvielle-Sebastia, L., & Keller, W. (1997). The 30-kD subunit of mammalian cleavage and polyadenylation specificity factor and its yeast homolog are RNA-binding zinc finger proteins. *Genes & Development, 11*(13), 1703-1716.

31. Barabino, S. M., Ohnacker, M., & Keller, W. (2000). Distinct roles of two Yth1p domains in 3'-end cleavage and polyadenylation of yeast pre-mRNAs. *The EMBO Journal, 19*(14), 3778-3787. doi: 10.1093/emboj/19.14.3778

32. Tacahashi, Y., Helmling, S., & Moore, C. L. (2003). Functional dissection of the zinc finger and flanking domains of the Yth1 cleavage/polyadenylation factor. *Nucleic Acids Research, 31*(6), 1744-1752.

33. Zhao, J., Kessler, M. M., & Moore, C. L. (1997). Cleavage factor II of saccharomyces cerevisiae contains homologues to subunits of the mammalian cleavage/ polyadenylation specificity factor and exhibits sequence-specific, ATP-dependent interaction with precursor RNA. *The Journal of Biological Chemistry, 272*(16), 10831-10838.

34. Kyburz, A., Sadowski, M., Dichtl, B., & Keller, W. (2003). The role of the yeast cleavage and polyadenylation factor subunit Ydh1p/Cft2p in pre-mRNA 3'-end formation. *Nucleic Acids Research, 31*(14), 3936-3945.

35. Preker, P. J., Ohnacker, M., Minvielle-Sebastia, L., & Keller, W. (1997). A multisubunit 3' end processing factor from yeast containing poly(A) polymerase and homologues of the subunits of mammalian cleavage and polyadenylation specificity factor. *The EMBO Journal, 16*(15), 4727-4737. doi: 10.1093/emboj/16.15.4727

36. Licatalosi, D. D., Geiger, G., Minet, M., Schroeder, S., Cilli, K., McNeil, J. B., & Bentley, D. L. (2002). Functional interaction of yeast pre-mRNA 3' end processing factors with RNA polymerase II. *Molecular Cell, 9*(5), 1101-1111.

37. McNeil, J. B., Agah, H., & Bentley, D. (1998). Activated transcription independent of the RNA polymerase II holoenzyme in budding yeast. *Genes & Development, 12*(16), 2510-2521.

38. Ezeokonkwo, C., Zhelkovsky, A., Lee, R., Bohm, A., & Moore, C. L. (2011). A flexible linker region in Fip1 is needed for efficient mRNA polyadenylation. *RNA (New York, N.Y.), 17*(4), 652-664. doi: 10.1261/rna.2273111; 10.1261/rna.2273111

39. Helmling, S., Zhelkovsky, A., & Moore, C. L. (2001). Fip1 regulates the activity of poly(A) polymerase through multiple interactions. *Molecular and Cellular Biology, 21*(6), 2026-2037. doi: 10.1128/MCB.21.6.2026-2037.2001

40. Jenny, A., Minvielle-Sebastia, L., Preker, P. J., & Keller, W. (1996). Sequence similarity between the 73-kilodalton protein of mammalian CPSF and a subunit of yeast polyadenylation factor I. *Science (New York, N.Y.), 274*(5292), 1514-1517.

41. Ryan, K., Calvo, O., & Manley, J. L. (2004). Evidence that polyadenylation factor CPSF-73 is the mRNA 3' processing endonuclease. *RNA (New York, N.Y.), 10*(4), 565-573.

42. Chanfreau, G., Noble, S. M., & Guthrie, C. (1996). Essential yeast protein with unexpected similarity to subunits of mammalian cleavage and polyadenylation specificity factor (CPSF). *Science (New York, N.Y.), 274*(5292), 1511-1514.

43. Zhao, J., Kessler, M., Helmling, S., O'Connor, J. P., & Moore, C. (1999). Pta1, a component of yeast CF II, is required for both cleavage and poly(A) addition of mRNA precursor. *Molecular and Cellular Biology, 19*(11), 7733-7740.

44. Garas, M., Dichtl, B., & Keller, W. (2008). The role of the putative 3' end processing endonuclease Ysh1p in mRNA and snoRNA synthesis. *RNA (New York, N.Y.), 14*(12), 2671-2684. doi: 10.1261/rna.1293008; 10.1261/rna.1293008

45. Tardiff, D. F., Lacadie, S. A., &Rosbash, M. (2006). A genome-wide analysis indicates that yeast pre-mRNA splicing is predominantly posttranscriptional. *Molecular Cell, 24*(6), 917-929. doi: 10.1016/j.molcel.2006.12.002

46. Berg, M. G., Singh, L. N., Younis, I., Liu, Q., Pinto, A. M., Kaida, D., . . .Dreyfuss, G. (2012). U1 snRNP determines mRNA length and regulates isoform expression. *Cell, 150*(1), 53-64. doi: 10.1016/j.cell.2012.05.029; 10.1016/j.cell.2012.05.029

47. Kaida, D., Berg, M. G., Younis, I., Kasim, M., Singh, L. N., Wan, L., &Dreyfuss, G. (2010). U1 snRNP protects pre-mRNAs from premature cleavage and polyadenylation. *Nature, 468*(7324), 664-668. doi: 10.1038/nature09479; 10.1038/nature09479

48. Proudfoot, N. (1996). Ending the message is not so simple. *Cell, 87*(5), 779-781.

49. Gunderson, S. I., Polycarpou-Schwarz, M., &Mattaj, I. W. (1998). U1 snRNP inhibits pre-mRNA polyadenylation through a direct interaction between U1 70K and poly(A) polymerase. *Molecular Cell, 1*(2), 255-264.

50. Dantonel, J. C., Murthy, K. G., Manley, J. L., &Tora, L. (1997). Transcription factor TFIID recruits factor CPSF for formation of 3' end of mRNA. *Nature, 389*(6649), 399-402. doi: 10.1038/38763

51. McCracken, S., Fong, N., Yankulov, K., Ballantyne, S., Pan, G., Greenblatt, J., Bentley, D. L. (1997). The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. *Nature, 385*(6614), 357-361. doi: 10.1038/385357a0

52. Ozsolak, F., Kapranov, P., Foissac, S., Kim, S. W., Fishilevich, E., Monaghan, A. P., . . . Milos, P. M. (2010). Comprehensive polyadenylation site maps in yeast and human reveal pervasive alternative polyadenylation. *Cell, 143*(6), 1018-1029. doi: 10.1016/j.cell.2010.11.020; 10.1016/j.cell.2010.11.020

CHAPTER 5

FUTURE DIRECTIONS

5.1 Mapping spliceosome assembly genome-wide in a wild type, bdf1Δ and bdf2Δ strain

Loss of Bdf1 but not of its homologue Bdf2 leads to a global misregulation in pre-mRNA splicing. Both Bdf1 and Bdf2 are capable of interacting with Taf7 and recruiting RNA polymerase II, and it was shown that in the absence of Bdf1, Bdf2 can be recruited to some Bdf1 specific promoters and substitute for its transcriptional function. Thus, my working hypothesis has been that, while both Bdf1 and Bdf2 can interchangeably function in transcription, only Bdf1 maintains a role in pre-mRNA splicing. Given that only 5% of yeast genes contain an intron, loss of Bdf1 would lead to a global splicing defect, which would not cause lethality if intron-containing genes could still be inefficiently spliced.

Consistent with this hypothesis, only *bdf1Δ* decreased the recruitment of the first spliceosomal factor to recognize the substrate, U1 snRNP, for several intron-containing genes in a small scale ChIP-qPCR experiment. Intriguingly, the same 2-fold decrease that was observed for U1 snRNP, was also observed for RNA polymerase II levels. However, one cannot simply dismiss this as purely a transcription defect, since if RNA polymerase recruitment is halved, one would expect a two-fold decrease in transcription, but no *a priori* defect in splicing, even if the spliceosome was recruited at half its normal levels. A decrease in transcription on a

splicing sensitive microarray would appear as a concomitant decrease in the total, precursor and mature levels when compared to a wild-type strain. This is not consistent with what I observed on a *bdf1Δ* array in which precursor levels were significantly upregulated, suggesting a secondary splicing defect.

Since so far I have only tested this hypothesis on a small subset of genes, it would be interesting to expand this study to map spliceosome recruitment genome-wide. Such an analysis would allow me to address several questions: (1) Is the decrease in U1 recruitment observed in *bdf1Δ* a global phenotype that is present on all intron-containing genes or does it affect specific classes of genes? (2) Is only the recruitment of U1 impacted, or are the other snRNAs less efficiently recruited as well? (3) Is the correlation between U1 and RNA polymerase II recruitment a global phenotype and if not, what types of genes is it associated with? (4) Is the defect in recruitment simply a quantitative feature or can delays in recruitment be observed in the body of the genes? (5) Since Bdf1 is actively recruited at the level of the +1 and +2 nucleosomes, are these nucleosomes more stable in a *bdf1Δ* and does that impact RNA polymerase II profiles? (6) Are any of the observed changes in a *bdf1Δ* strain correlated with expression levels, and exonic or intronic features? These and many other questions could be addressed by a ChIP-Seq study which globally maps RNA polymerase and nucleosome occupancy together with snRNP recruitment in a wild-type strain in comparison with a *bdf1Δ* and a *bdf2Δ* strain.

To this end, I have generated several strains in which RNA polymerase or spliceosomal proteins that are stably bound to U1, U2 or U5 snRNPs are TAP-tagged in a background that is either wild-type, *bdf1Δ* or *bdf2Δ*. I have tested these strains to

ensure they match the *bdf1Δ* or *bdf2Δ* parent phenotypes, including the presence or absence of an increase in U3 snoRNA reporter levels. Unfortunately, the first ChIP-Seq dataset that would start to address these questions did not have a high-enough resolution at the level of mapping snRNP recruitment and will have to be repeated with more input material. Nevertheless, the strains and methods are currently in place to monitor global spliceosome patterns in these backgrounds.

5.2 Identifying and characterizing Bdf1 interactors

Mutagenesis of Bdf1 coupled with a forward-genetic screen for mutants that mimic *bdf1Δ* phenotypes has generated a subset of mutants which map to the C-terminal end of Bdf1. Given the difficulty in precisely filtering out mutations which do not impact the splicing phenotype, one would ideally wish to map hotspots within the protein which are consistently mutated in true candidates. Expanding the number of mutants which are screened would broaden the number of mutations and better aid in mapping residues which are truly important for the splicing function of Bdf1.

Nevertheless, despite the difficulty in assigning function to specific residues, and the complication which arises when combinations of mutations are present in a single strain, mutations which fall in highly conserved Bdf1 residues which are absent from Bdf2 can be highly relevant for indicating a Bdf1 specific function. Once a domain or residues have been narrowed down, a first step would be to re-confirm the splicing phenotype though an alternative method such as splicing sensitive microarrays, or by testing the splicing of several intron containing genes which are impacted in the *bdf1Δ* array with small scale qPCR. Next, one can use these mutants to address the important

question of which interactions are disrupted in a Bdf1 mutant that mimics *bdf1Δ*. Using a method such as SILAC, the interactors of Bdf1 and its mutant can be comparatively mapped, to the level of changes in their post-translational status.

Bdf1 displays several physical interactions with histone H4 (Hhf1, Hhf2), TAFs, SWR-C components (Swr1, Htz1) and Bdf2 (BioGRID). While none of its genetic or primary physical interactions involve splicing components, Prp9 (through Skp1), Prp43 (through Taf6, Cka2 or Htz1), Prp19 (through Swr1 and Ies1), Npl3 (through Hhf2, Spt16) or Prp31 and Cwc2 (through Rvb1) are just some of the secondary interactions which could link Bdf1 to its role in splicing. More importantly, these interactions are not shared with Bdf2. Interestingly, aside from Prp9 which corresponds to the mammalian U2 factor SF3a, most of these factors are involved in the later stages of spliceosome assembly together with the U4/U6/U5 snRNP or in spliceosome recycling (Prp43), and it will be interesting to see how they reconcile with the early defect in U1 recruitment. Identifying specific interactions which are disrupted in the Bdf1 mutants would aid in narrowing down this list of factors and shed more light on the mechanism behind Bdf1's role in splicing. Once such an interactor is identified, more specific interaction domains and its mode of action can be thoroughly examined.

APPENDIX

PRE5, A PROTEASOMAL SUBUNIT INVOLVED IN PRE-mRNA SPLICING?

A.1 Introduction

The yeast proteasome is a large 2.5MDa multiprotein machine involved in the degradation of poly-ubiquitylated protein substrates. The 26S proteasome consists of a 20S barrel-shaped core particle and two 19S regulatory subunits. Within the 20S core, there are 14 subunits – 7 alpha and 7 beta subunits which make up the 2 inner (β) and 2 outer (α) rings [1,2]. All subunits except for one alpha subunit (Pre9) are essential in *Saccharomyces cerevisiae*. For a targeted protein to be degraded by the proteasome, it needs to be first unfolded by the ATPase subunits within the 19S regulatory particle and then fed through the 20S barrel towards the β subunits which possess catalytic activity.

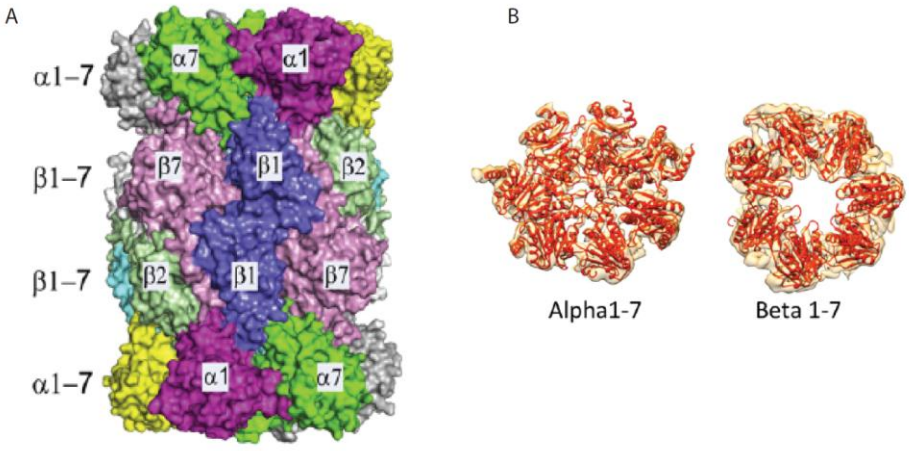

Figure A.1: The yeast 20S proteasome (adapted from [3] and [1])

Pre5 is the alpha 6 subunit of the yeast proteasome, a 234 amino acid protein whose

function is unknown [4]. It has conserved domains which are characteristic of the α-type proteasome subunits and is homologous to the Drosophila PROS35, human C2 and Arabidopsis PMS30 proteins. Little more is known about Pre5, such as that its repression leads to abnormal mitochondrial morphology [5] and that a reduction in Pre5 levels causes telomere shortening [6].

Interestingly, recent studies have uncovered a role for ubiquitin in pre-mRNA splicing [7]. Thus, an *in vitro* study has shown that mutant alleles of ubiquitin, as well as treatment with ubistatin, promotes disassembly of the U4/U6.U5 tri-snRNP complex and inhibits splicing. Tri-snRNP stability is dependent on the ubiquitination status of Prp8, a protein which is central to spliceosome activation and a member of the triple snRNP. The current hypothesis beyond this model is that a conformational shift of Prp8 which itself has a Jab1/MPN ubiquitin-interacting domain [8, 9] regulates the unwinding of the tri-snRNP, thus timing the transitions between the different steps of spliceosome rearrangements. Moreover, Prp19, a member of the NineTeen complex which stabilizes the association of U5 and U6 snRNPs with the activated spliceosome is itself a U-box E3 ubiquitin ligase [10, 11]. Investigation of Prp19 targets is currently underway, but the first one – Prp3, which is a component of the U4 snRNP – has been identified [12]. Prp19 modifies Prp3 with non-proteolytic K63-linked ubiquitin chains and increases its affinity for Prp8. Moreover, asides from studies in *S. cerevisiae*, two studies in S pombe [13, 14] identified that the mutation or deletion of Hub1, which also encodes a ubiquitin-like protein and interacts with the spliceosomal factor Snu66 inhibits pre-mRNA splicing. Furthermore, it was shown that depletion of the human orthologue of Prp19, SNEV [15], impairs splicing in nuclear extracts. SNEV is also a

U3-box ubiquitin ligase [16] which, like Prp19, was shown to physically associate with the β7 subunit of the proteasome, although the purpose for this is currently unknown. Lastly, the mouse Prp19 was also shown to interact with SUG1, a regulatory subunit of the 26S proteasome [17]. Although the role of ubiquitin in regulating spliceosome dynamics is becoming clearer, the significance of the physical interactions between Prp19 and proteasome subunits and its importance in promoting degradation or in a different context remains to be addressed.

Pre5 emerged as a candidate in my initial genome-wide screen for factors which impact pre-mRNA splicing. Microarray analysis showed that the *pre5-ts* allele [18] leads to a global inhibition of splicing, a novel phenotype so far unreported. While this phenotype does not appear to be an indirect effect caused by the altered expression of a splicing factor, it could arguably be linked to an indirect change in the protein levels of such a factor. Nevertheless, the expression of most genes in *pre5-ts* is similar to their expression in other proteasome mutants – *pre1-ts* and *pre10-ts*, as well as published datasets. However, these other mutants do not exhibit any splicing defects, suggesting that any putative role in splicing is specific to *pre5-ts*.

A.2 Results and discussion

In the genome-wide screen presented in chapter 2, *pre5-ts* caused an 8-fold upregulation in U3 precursor levels (Figure 2A), suggesting that it may play a role in pre-mRNA splicing. This upregulation was particular to the *pre5-ts* mutant, since none of the other proteasome mutants, either in the core or in the 19S regulatory subunit, showed this effect. To investigate this phenotype further, I employed splicing sensitive

microarrays to study the *pre5-ts* mutant and two other core proteasome mutants in the library: *pre1-ts* and *pre10-ts*. Pre1 is the β4 proteasome subunit and has an inferred endopeptidase activator activity [19], whereas Pre10 is the α7 subunit and was shown to bind RNA [20]. In our splicing sensitive microarrays, true splicing mutants like *snt309*Δ are characterized by an increase in precursor levels for most intron containing genes, as well as a lack of change or a slight decrease in the total and mature levels. Also, since our microarrays are always comparing the mutant version to a wild-type version, all changes are measured relative to the wild-type.

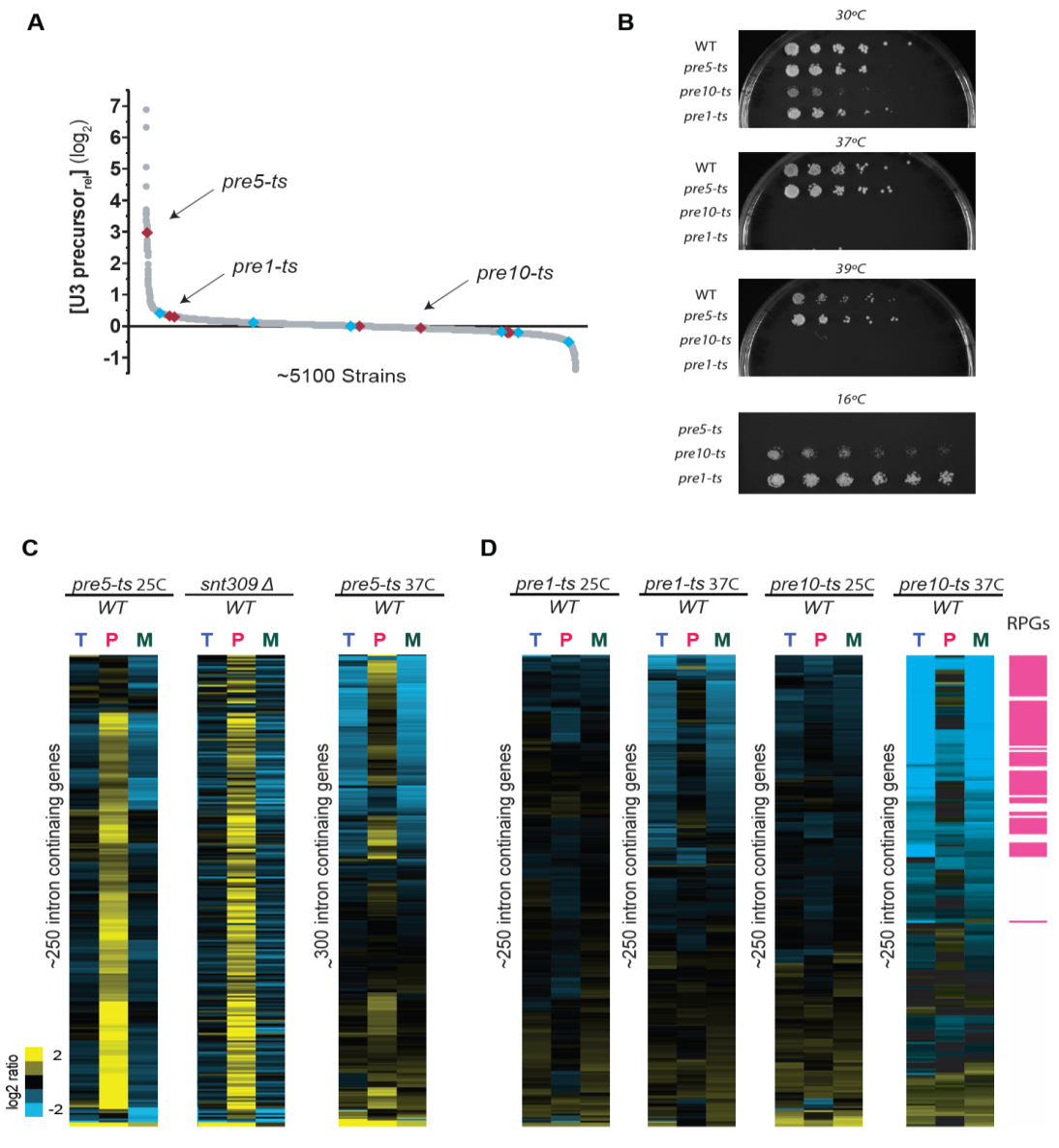

Figure A.2: **pre5-ts** ***impacts splicing genome wide***

A. The impact of proteasome mutants on U3 precursor levels. The data is shown on a log2 scale. The core proteasome mutants are colored in red, whereas the mutants in the 19S regulatory subunit are shown in blue.
B. Growth phenotypes for the proteasome mutants under different temperature conditions
C. Splicing sensitive microarrays in the pre5-ts background at permissive (25°C) or restrictive temperature. The *pre5-ts* datasets are clustered individually, whereas the snt309Δ mutant is ordered the same way as the data in the *pre5-ts* 25°C mutant.

D. Microarrays for the *pre1-ts* and *pre10-ts* strains, clustered independently. The *pre10-ts* strain at 37°C shows a significant downregulation of ribosomal protein genes (right).

The *pre5-ts* allele causes misregulation in precursor levels for most intron containing genes (figure 2C), displaying a phenotype very similar to the *bona-fide* splicing mutant *snt309Δ*. Curiously, this phenotype was more pronounced in the *pre5-ts* allele at the permissive temperature of 25°C, rather than after a 15 minute shift to the non-permissive temperature of 37°C. Conversely, the two other proteasome mutants didn't impact pre-mRNA splicing at all (figure 2D) – either at the permissive or the non-permissive temperatures, which is consistent with the screen results (figure 2A). Of note is that the *pre10-ts* allele leads to a significant global decrease in the expression of ribosomal protein genes (RPGs) at 37°C, which is consistent with a role for the proteasome in controlling RPG expression levels [21, 22]. The expression of RPGs was previously observed to be decreased upon proteasome inhibition and in the presence of proteasome mutants, which correlated with a decrease in the recruitment of proteasome components and RNA polymerase II on these genes.

Since I was intrigued that *pre5-ts* would show a more pronounced splicing defect at 25°C rather that at the non-permissive temperature [18], I decided to examine the growth phenotypes of the proteasome mutants I was investigating. As seen in figure 2B, the *pre5-ts* allele did not show any growth defect at the non-permissive temperature of 37°C, nor did it fail to grow at 39°C (data not shown). On the contrary, *pre5-ts* did not grow at cold temperatures, which is reminiscent of other splicing mutants. Conversely, *pre1-ts* and *pre10-ts* exhibit phenotypes which are consistent

with the ones reported by Ben-Aroya et.al (figure 2B).

Given the contradiction between the expected phenotype and the one I observed in what is now a *pre5-cs* allele, one needs to be careful in investigating this phenotype further. A first step would be to clone out and sequence this allele, and then check if insertion in an otherwise wild-type cell reconstitutes the splicing defect. Conversely, if the mutation is indeed found at the *pre5* locus, one could back-cross the strain to ensure that any other putative secondary mutations are selected against. If the splicing defect persists, then the phenotype is caused by the mutation in the *pre5* allele and is not due to some other spurious effect.

Nevertheless, a couple of other lines of evidence tend to suggest that the splicing phenotype that I see in *pre5-ts* is valid. One is that the genome-wide regulation of expression patterns is very similar among the three proteasome mutants I examined (figure 3A).

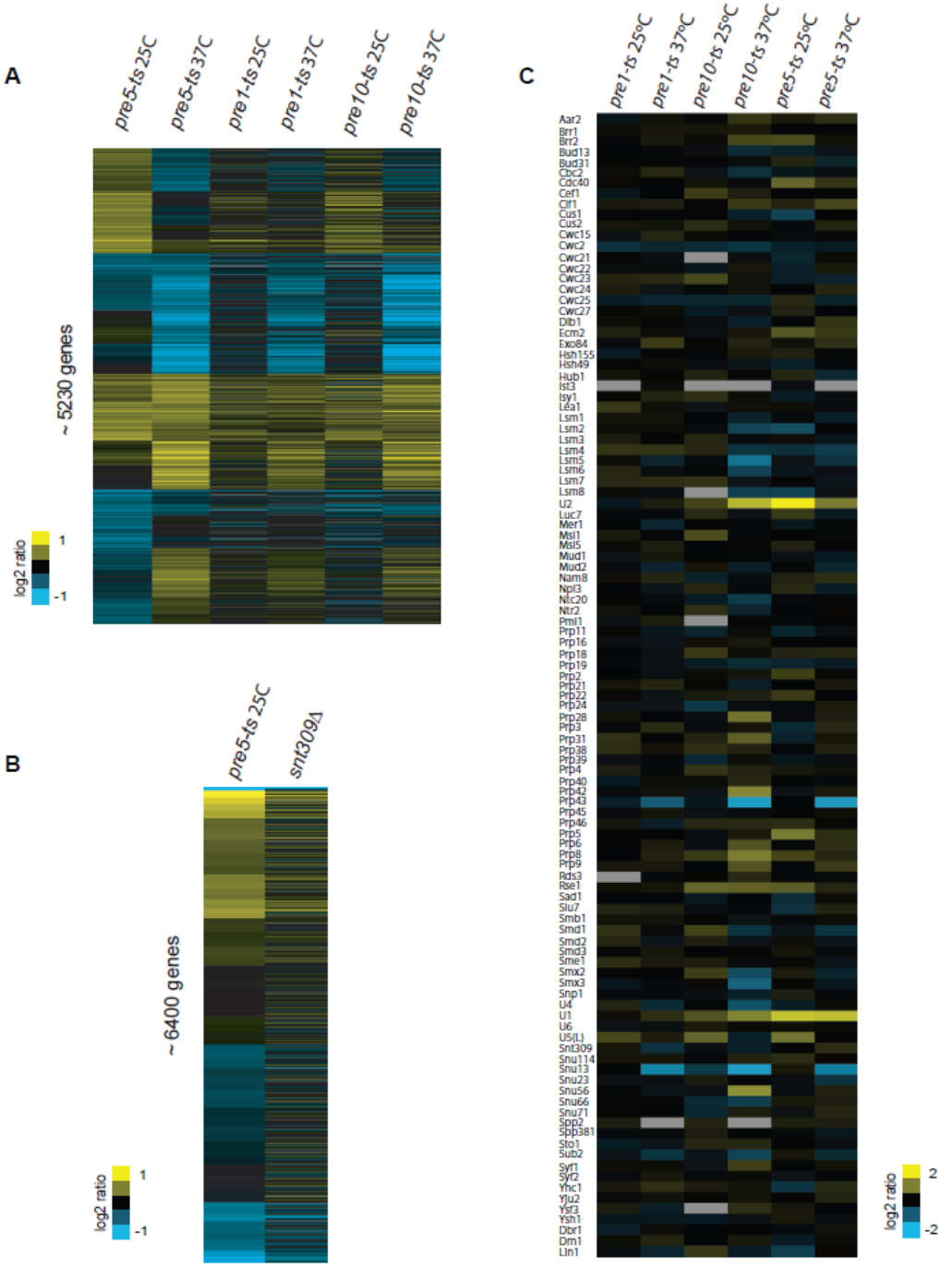

Figure A.3: General expression patterns in proteasome mutants

A. Ordering of all exons in the proteasome mutants at permissive and restrictive temperatures based on *pre5-ts* clustering at both 25°C and 37°C. The data is shown on a log2 scale.
B. Comparison of the *pre5-ts* and *snt309*Δ exon levels genome-wide reveals little similarity

C. Expression profiles of all splicing factors in the proteasome mutants.

The ~5200 genes shown in figure 3A are ordered hierarchically by organizing and grouping the *pre5-ts* expression patterns at 25°C and 37°C. As one can see, the expression patterns in *pre1-ts* and *pre10-ts* closely mirror the ones in *pre5-ts*, with specific genes being highly up- or down regulated in all arrays. Moreover, the regulation of different classes of genes within these arrays matches published datasets of other proteasome mutants [21]. Thus, genes involved in mating type regulation (alpha1, alpha2) as well as a significant proportion of ribosomal protein genes (87 out of most 200 down-regulated genes in *pre5-ts* at non-permissive temperature), are generally downregulated in all mutants I examined. Furthermore, expression of genes involved in fatty acid metabolism (Ole1, Fas1, Fas2) is decreased in all arrays, whereas an increase in expression for catabolism-related and mitocondrial genes is observed, similar to published data. In addition, there are clear clusters of genes which are similarly regulated at 37°C, and which are most probably involved in the stress response at high temperature. In contrast, when I compared the *snt309Δ* and *pre5-ts* gene expression patterns, these were highly dissimilar (figure 3B). This suggests that the *pre5-ts* mutant, despite being cold sensitive, exhibits the same type of gene regulation as other proteasome mutants, suggesting that it might indeed function in the same metabolic pathway.

A second piece of evidence suggests that the splicing defects I observed in the *pre5-ts* mutant are unlikely to be caused by indirect misregulation of the expression of another spliceosomal factor (figure 3C). Whereas the U1 and U2 snRNA levels are increased in the *pre5-ts* strain, they are also up-regulated in the *pre10-ts* mutant, which suggests

that the upregulation of these factors is unlikely to cause the general splicing defect seen in *pre5-ts*. In addition, no other factors are increased or decreased more than 2-fold in *pre5-ts* alone. However, this does not preclude an indirect effect caused by misregulation of protein levels in a splicing factor, which is a concern that will need to be addressed.

Pre5 is thus interesting in that it is a core proteasome subunit which may play a separate role in pre-mRNA splicing. Evidence that this is not a global phenotype caused by mutations in proteasome components comes from the other two mutants I examined *pre1-ts* and *pre10-ts,* which do not affect splicing of yeast genes. While the proteasome can regulate gene expression indirectly, by down-regulating the levels of specific transactivators, it has more recently been shown to associate with the majority of yeast genes [21, 23]. Several 19S components were also shown to have roles which are independent of the main degradation function. Thus, the Rpt6 subunit of the 19S complex was shown to physically interact with FACT1, an elongation factor [24], and its depletion inhibited transcription [25]. In addition, the proteasomal ATPases in the 19S complex facilitate the recruitment of SAGA to yeast promoters [26] and regulate the function of mono-ubiquitylated activators [27]. However, these alternative roles are independent of the 20S core subunit. To my knowledge, this is the first report of a 20S α-subunit which might play a separate role in pre-mRNA splicing.

Lastly, Sem1, a component of the 19S lid subcomplex was shown to form two other complexes in which it functions independently of the proteasome [28]. As a part of the Sac3-Thp1 complex, Sem1 is involved in pre-mRNA export. Sem1 also associates with Csn12, a COP9 signalosome protein which plays a separate role in pre-mRNA

splicing. Thus, it is not inconceivable that Pre5 may play a proteasome-independent role. While Pre1, Pre5 and Pre10 share several interactors, only Pre5 interacts with Sub2 [29] and Sus1 [30] suggesting that it might play a role in coupling splicing to pre-mRNA export. However, extensive testing of models must be put off until the true nature of the *pre5-ts* mutation and its phenotype are confirmed.

REFERENCES

1. Beck, F., Unverdorben, P., Bohn, S., Schweitzer, A., Pfeifer, G., Sakata, E., et al. (2012). Near-atomic resolution structural model of the yeast 26S proteasome. *Proceedings of the National Academy of Sciences of the United States of America, 109*(37), 14870-14875.

2. Groll, M., Ditzel, L., Lowe, J., Stock, D., Bochtler, M., Bartunik, H. D., et al. (1997). Structure of 20S proteasome from yeast at 2.4 A resolution. *Nature, 386*(6624), 463-471.

3. Stadtmueller, B. M., & Hill, C. P. (2011). Proteasome activators. *Molecular Cell, 41*(1), 8-19. doi: 10.1016/j.molcel.2010.12.020; 10.1016/j.molcel.2010.12.020

4. Heinemeyer, W., Trondle, N., Albrecht, G., & Wolf, D. H. (1994). PRE5 and PRE6, the last missing genes encoding 20S proteasome subunits from yeast? Indication for a set of 14 different subunits in the eukaryotic proteasome core. *Biochemistry, 33*(40), 12229-12237.

5. Altmann, K., &Westermann, B. (2005). Role of essential genes in mitochondrial morphogenesis in saccharomyces cerevisiae. *Molecular Biology of the Cell, 16*(11), 5410-5417.

6. Ungar, L., Yosef, N., Sela, Y., Sharan, R., Ruppin, E., &Kupiec, M. (2009). A genome-wide screen for essential yeast genes that affect telomere length maintenance. *Nucleic Acids Research, 37*(12), 3840-3849.

7. Bellare, P., Small, E. C., Huang, X., Wohlschlegel, J. A., Staley, J. P., & Sontheimer, E. J. (2008). A role for ubiquitin in the spliceosome assembly pathway. *Nature Structural & Molecular Biology, 15*(5), 444-451. doi: 10.1038/nsmb.1401; 10.1038/nsmb.1401

8. Bellare, P., Kutach, A. K., Rines, A. K., Guthrie, C., & Sontheimer, E. J. (2006). Ubiquitin binding by a variant Jab1/MPN domain in the essential pre-mRNA splicing factor Prp8p. *RNA (New York, N.Y.), 12*(2), 292-302. doi: 10.1261/rna.2152306

9. Pena, V., Liu, S., Bujnicki, J. M., Luhrmann, R., & Wahl, M. C. (2007). Structure of a multipartite protein-protein interaction domain in splicing factor prp8 and its link to retinitis pigmentosa. *Molecular Cell, 25*(4), 615-624. doi: 10.1016/j.molcel.2007.01.023

10. Hatakeyama, S., Yada, M., Matsumoto, M., Ishida, N., & Nakayama, K. I. (2001). U box proteins as a new family of ubiquitin-protein ligases. *The Journal of Biological Chemistry, 276*(35), 33111-33120. doi: 10.1074/jbc.M102755200

11. Ohi, M. D., Vander Kooi, C. W., Rosenberg, J. A., Chazin, W. J., & Gould, K. L. (2003). Structural insights into the U-box, a domain associated with multi-ubiquitination. *Nature Structural Biology*, *10*(4), 250-255. doi: 10.1038/nsb906

12. Song, E. J., Werner, S. L., Neubauer, J., Stegmeier, F., Aspden, J., Rio, D., . . . Rape, M. (2010). The Prp19 complex and the Usp4Sart3 deubiquitinating enzyme control reversible ubiquitination at the spliceosome. *Genes & Development, 24*(13), 1434-1447. doi: 10.1101/gad.1925010; 10.1101/gad.1925010

13. Yashiroda, H., & Tanaka, K. (2004). Hub1 is an essential ubiquitin-like protein without functioning as a typical modifier in fission yeast. *Genes to Cells : Devoted to Molecular & Cellular Mechanisms*, *9*(12), 1189-1197. doi: 10.1111/j.1365-2443.2004.00807.x

14. Wilkinson, C. R., Dittmar, G. A., Ohi, M. D., Uetz, P., Jones, N., & Finley, D. (2004). Ubiquitin-like protein Hub1 is required for pre-mRNA splicing and localization of an essential splicing factor in fission yeast. *Current Biology : CB*, *14*(24), 2283-2288. doi: 10.1016/j.cub.2004.11.058

15. Grillari, J., Ajuh, P., Stadler, G., Loscher, M., Voglauer, R., Ernst, W., . . . Katinger, H. (2005). SNEV is an evolutionarily conserved splicing factor whose oligomerization is necessary for spliceosome assembly. *Nucleic Acids Research*, *33*(21), 6868-6883. doi: 10.1093/nar/gki986

16. Loscher, M., Fortschegger, K., Ritter, G., Wostry, M., Voglauer, R., Schmid, J. A., . . . Grillari, J. (2005). Interaction of U-box E3 ligase SNEV with PSMB4, the beta7 subunit of the 20 S proteasome. *The Biochemical Journal*, *388*(Pt 2), 593-603. doi: 10.1042/BJ20041517

17. Sihn, C. R., Cho, S. Y., Lee, J. H., Lee, T. R., & Kim, S. H. (2007). Mouse homologue of yeast Prp19 interacts with mouse SUG1, the regulatory subunit of 26S proteasome. *Biochemical and Biophysical Research Communications*, *356*(1), 175-180. doi: 10.1016/j.bbrc.2007.02.134

18. Ben-Aroya, S., Coombes, C., Kwok, T., O'Donnell, K. A., Boeke, J. D., &Hieter, P. (2008). Toward a comprehensive temperature-sensitive mutant repository of the essential genes of saccharomyces cerevisiae. *Molecular Cell*, *30*(2), 248-258. doi: 10.1016/j.molcel.2008.02.021; 10.1016/j.molcel.2008.02.021

19. Chen, P., &Hochstrasser, M. (1996). Autocatalytic subunit processing couples active site formation in the 20S proteasome to completion of assembly. *Cell*, *86*(6), 961-972.

20. Scherrer, T., Mittal, N., Janga, S. C., & Gerber, A. P. (2010). A screen for RNA-binding proteins in yeast indicates dual functions for many enzymes.*PloS One*, *5*(11), e15499.

21. Auld, K. L., Brown, C. R., Casolari, J. M., Komili, S., & Silver, P. A. (2006). Genomic association of the proteasome demonstrates overlapping gene regulatory activity with transcription factor substrates. *Molecular Cell*, *21*(6), 861-871.

22. Auld, K. L., & Silver, P. A. (2006). Transcriptional regulation by the proteasome as a mechanism for cellular protein homeostasis. *Cell Cycle (Georgetown, Tex.), 5*(14), 1503-1505.

23. Sikder, D., Johnston, S. A., &Kodadek, T. (2006). Widespread, but non-identical, association of proteasomal 19 and 20 S proteins with yeast chromatin. *The Journal of Biological Chemistry, 281*(37), 27346-27355.

24. Sun, L., Johnston, S. A., &Kodadek, T. (2002). Physical association of the APIS complex and general transcription factors. *Biochemical and Biophysical Research Communications, 296*(4), 991-999.

25. Ferdous, A., Gonzalez, F., Sun, L., Kodadek, T., & Johnston, S. A. (2001). The 19S regulatory particle of the proteasome is required for efficient transcription elongation by RNA polymerase II. *Molecular Cell, 7*(5), 981-991.

26. Lee, D., Ezhkova, E., Li, B., Pattenden, S. G., Tansey, W. P., & Workman, J. L. (2005). The proteasome regulatory particle alters the SAGA coactivator to enhance its interactions with transcriptional activators. *Cell, 123*(3), 423-436.

27. Ferdous, A., Sikder, D., Gillette, T., Nalley, K., Kodadek, T., & Johnston, S. A. (2007). The role of the proteasomalATPases and activator monoubiquitylation in regulating Gal4 binding to promoters. *Genes & Development, 21*(1), 112-123.

28. Wilmes, G. M., Bergkessel, M., Bandyopadhyay, S., Shales, M., Braberg, H., Cagney, G., et al. (2008). A genetic interaction map of RNA-processing factors reveals links between Sem1/Dss1-containing complexes and mRNA export and splicing. *Molecular Cell, 32*(5), 735-746.

29. Cagney, G., Uetz, P., & Fields, S. (2001). Two-hybrid analysis of the saccharomyces cerevisiae 26S proteasome. *Physiological Genomics, 7*(1), 27-34.

30. Lee, K. K., Sardiu, M. E., Swanson, S. K., Gilmore, J. M., Torok, M., Grant, P. A., et al. (2011). Combinatorial depletion analysis to assemble the network architecture of the SAGA and ADA chromatin remodeling complexes. *Molecular Systems Biology, 7*, 503.