THE UNIVERSITY
of EDINBURGH

THE UNIVERSITY
of EDINBURGH

Using the auxin-inducible degron to study the spliceosome cycle and splicing fidelity

Gonzalo Ismael Mendoza-Ochoa

August, 2017
A thesis presented for the degree of Doctor of Philosophy

Declaration

I declare that this thesis has been composed by myself and that the work has not been submitted for any other degree or professional qualification. My contribution and those of the other authors to this work have been explicitly indicated below. I confirm that appropriate credit has been given within this thesis where reference has been made to the work of others.

_____

Gonzalo Ismael Mendoza-Ochoa

Edinburgh, August 2017

Abstract

I investigated two aspects of *in vivo* splicing that are poorly understood: spliceosome disassembly and recycling, and proofreading. To this end, I used the auxin-inducible degron (AID) to individually deplete several splicing factors in budding yeast and then I measured the effect on co-transcriptional spliceosome assembly through chromatin immunoprecipitation. In addition, using RNA next-generation sequencing, I measured the frequency of splicing errors following depletion or mutation of the fidelity factor, Prp22. I show that formation of the pre-spliceosome (the first stage of spliceosome assembly) is rapidly inhibited by global defects in late stages of spliceosome assembly. I demonstrate that this is due to the accumulation of arrested spliceosomes that sequester the splicing machinery and, as a result, causes a recycling defect. This suggests that spliceosomes that lack essential splicing factors are not always properly disassembled and recycled *in vivo*, and warns about potential systematic secondary effects when perturbing single components of the spliceosome. Secondly, I describe the development of a new version of the AID system for budding yeast, called the B-estradiol AID. To the best of my knowledge, an AID system for budding yeast that is fast-acting, tightly-controlled and gratuitous, was lacking until now. Lastly, I show that absence of Prp22 protein, which was previously proposed to play a role in splicing fidelity, correlates with more mistakes in 3'ss selection of many endogenous intron-containing transcripts *in vivo*. This provides indirect evidence to suggest that Prp22-dependent splicing proofreading is physiologically important. The data from this analysis will be useful in ongoing studies to try to identify common features that could improve our understanding of the mechanism of Prp22's function in splicing proofreading.

Lay Summary

Every living organism is made up of one or multiple cells, and every cell stores within its DNA instructions on how to survive and reproduce. DNA is a very long molecule that contains information that resembles the "letters" and "words" of language. For these molecular "words" to be read and interpreted by the cell, a copy of them is made in the form of RNA. There are many types of RNAs. One very special type of RNA is called messenger RNA (mRNA), which contains specific information ("words") on how to build a protein. Proteins are essential for cells and there are thousands of different proteins. Every single protein is built by reading and translating the information contained in mRNAs. However, before an mRNA can be read, there are bits of it that should be removed. If these bits are not removed, the message contained in the RNA does not make sense to the cell. This process is called splicing.

Splicing is essential for complex organisms, such as ourselves or yeast, and errors in this process can cause diseases, such as spinal muscular atrophy or retinitis pigmentosa – which leads to blindness. This is why it is important to understand the details of how splicing works, so we can find good ways of treating splicing-related diseases. Splicing is carried out by the spliceosome. The spliceosome is a very large molecular complex composed of more than 150 different parts. One special aspect of the spliceosome is that for every splicing event, it is assembled and then disassembled, through a very complicated cycle. Just think of it as "cutting and pasting" machine that for every single editing event, it is assembled from all its individual pieces in a step-wise fashion and, after the process is complete, it falls apart again, so its pieces can be recycling into more rounds of "cutting and pasting", through a cycle that goes on and on.

One common strategy that researchers use to study the function of cell components (for example, RNAs or proteins), is to artificially deplete an individual component and then observe what happens to the cells. As part of the experimental research I did during my PhD, my first objective was to use this depletion strategy to study the function of several protein components of the spliceosome. Unexpectedly, I observed that this depletion strategy is not always capable of answering the research question that was originally asked. My data suggests that this is because depletion of spliceosome components, can inhibit the recycling process and cause unwanted secondary effects. This discovery is important because it has implications in the way we study the function of cell components that are part of large complexes, such as the spliceosome. The message is that one has to be careful on how the data is interpreted when following a depletion strategy (or other similar ones), which suggests that, in some cases, additional controls must be implemented to reach an accurate conclusion. Another contribution I made was to generate new molecular tools for depleting protein targets in yeast. Compared to previous versions, these tools allow very efficient depletion, while minimizing unwanted secondary effects. I constructed these tools as an extension of a powerful research technology for protein depletion published a few years ago.

Acknowledgements

I deeply thank my supervisor Jean Beggs for welcoming me into her group, giving me freedom to follow my research interests while showing me how to do good science, and for always having her door kindly open. I greatly appreciate all former and current lab-mates (Jane R., David B., Susana d L., Keerthi C., Darek A., Eve H., Ema S., Vahid A., Barbara T., Charlotte C., Karen T., Jim B. and Edward W.), for their fantastic attitude of always happy to help, and for contributing to the friendly and collaborative environment of the lab. Especially, I thank Jane, Ema and David for nicely teaching me some powerful lab techniques; and my office-mate Vahid Aslanzadeh, always easy-going and a great friend, who taught me how to get NGS data analysis started. I sincerely express my gratitude to *Consejo Nacional de Ciencia y Tecnología* (CONACYT) and the School of Biological Sciences (UoE) for sponsoring my studies. Their generous support has allowed me to launch my career in science – wherever it might take me, while learning from the experience of living in this foreign and wonderful place. I am grateful for the work done by my thesis committee (Jean B., Sander G., David T. and Robin A.) and my examination committee (Greg K. and Andy N.). I thank them for paying attention to my work and giving useful feedback, and a boost of optimism when it was most needed. I thank my wife and love of my life Myri – the most beautiful soul I have met, for joining me in this adventure that has brought us, amongst other things, a little person that is the joy of our world. Also, I would like to express my appreciation for the community of Mexicans in Edinburgh, their network support and friendship helped me enormously throughout these years. Last but not least, I would not be here and be the person I am if not for the unconditional support and love of my parents, sisters and grandparents, I am incredibly lucky to have them as family.

Abbreviations

%	percentage
◦C	Celsius degrees
4tU	4-thio-uracil
5-FOA	5-fluoroorotic acid
A_{260}	absorbance at wavelength of 260 nM
A_{280}	absorbance at wavelength of 280 nM
AA	Amino acid
AID	auxin-inducible degron
ATP	Adenosine triphosphate
BP	Branchpoint
BS	Branchsite
Cat. No.	catalogue number
cDNA	complementary deoxyribonucleic acid
ChIP	chromatin immunoprecipitation
cs	cold sensitive
ddH_2O	double distilled H_2O
DMSO	Dimethyl sulfoxide
DNA	deoxyribonucleic acid
DNase	Deoxy ribonuclease
EDTA	Ethylenediamine tetraacetic acid
HA	Hemagglutinin
HEPES	(4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
hs	heat sensitive
IgG	Immunoglobulin G
IP	immunoprecipitation
LB	Lysogeny broth
M	Molar
mg	milligrams
min	minutes
mL	milliliters
mM	millimolar
mRNA	messenger ribonucleic acid
N/S	not Specified
NaCl	Sodium chloride
ng	nanograms
NP-40	nonidet P40 detergent
OD_{600}	optical density at wavelength of 600 nM
ORF	open reading frame
PCR	polymerase chain reaction
PEG	Polyethylene glycol
POL II	RNA Polymerase II
pre-mRNA	precursor messenger ribonucleic acid
qPCR	quantitative polymerase chain reaction

RIP	RNA immunoprecipitation
RNA	ribonucleic acid
Rnase	ribonuclease
RP	ribosomal protein
RPG	ribosomal protein gene
rpm	revolutions per minute
rRNA	ribosomal ribonucleic acid
RT	reverse transcription
SDS	Sodium dodecyl sulfate
snRNA	small nuclear ribonucleic acid
snRNP	small nuclear ribonucleic protein particle
SOC	super optimal broth with catabolite repression
SSDNA	salmon sperm deoxyribonucleic acid
TAE	Tris acetate buffer
TE	Tris EDTA buffer
tRNA	transfer ribonucleic acid
ts	temperature sensitive
U	Units
uL	microliters
uM	micromolar
V	Volts
w/o	without
WB	western blot
WCCB	Welcome Centre for Cell Biology
YMM	yeast minimal media
YNB	yeast nitrogen base
YPDA	yeast/peptone/dextrose/adenine

Contents

Chapter 1: Introduction

pre-mRNA Splicing

pre-mRNA splicing is the process of removing the "*intr*agenic regi*on*" (intron) and ligating the "*ex*pressed regi*ons*" (exons) of the precursor messenger RNA (pre-mRNA). Together with capping and polyadenylation, it is one of the main post-transcriptional modifications required for stable expression and translation of mRNA. Pre-mRNA splicing is carried out in the nucleus by the spliceosome. There are other types of splicing that occur by different mechanisms, such as self-splicing introns and tRNA splicing, and are not described here. This thesis is exclusively focused on spliceosomal splicing.

What is the function of introns?

The most widespread role of introns is to expand protein diversity, by allowing a single gene to encode for a range of protein isoforms, through a process called alternative splicing. This generally happens when an exon is skipped or an intron is included from the final spliced mRNA product, or from the use of alternative splice sites. Alternative splicing occurs widely in eukaryotes, especially in higher eukaryotes and is highly regulated. It is estimated that in humans about 95% of transcripts from multi-intron genes are alternatively spliced (Pan et al., 2008), and it has been predicted that more than half of alternative splicing events are differentially regulated between tissues (Wang et al., 2008). Alternative splicing is mediated by RNA-binding proteins that can enhance (*e.g.* SR proteins) or silence (*e.g.* hnRNP) splicing events.

Some introns may have functions unrelated to alternative splicing. For example, they have been associated with enhanced expression, 3' end formation, nonsense-mediated decay, nuclear export and cytoplasmic

localization [reviewed by (Chorev and Carmel, 2012)]. In addition, many introns encode non-coding RNAs, such as snoRNAs and miRNAs. One model that aims to explain the origin of introns suggests that they appeared early in the evolutionary history of eukaryotes, initially as functionless selfish elements (Koonin, 2006). According to this hypothesis, it was only later that they acquired a diverse range of functions, both on the DNA and RNA level. Their tremendous evolutionary success as regulatory elements, is owed in great part to their freedom to evolve independently of the coding sequence of the gene they inhabit.

Splice sites and the chemistry of splicing

Introns are defined by three short consensus sequences called splice sites. These splice sites are located at the start (5' splice site; 5'ss), at the end (3' splice site;3'ss) and near the 3' end (branchsite;BS) of the intron. The consensus sequences of the splice sites vary between organisms. For example, in humans only the -GT (5'ss) and AG- (3'ss) motifs are fully conserved, with only a weakly conserved consensus sequence for the branchsite, but contain a polypyrimidine tract immediately upstream of the 3'ss. In contrast, budding yeast have a strong consensus 5'ss (-GTATGT) and branchsite (ACTAAC), a poor consensus 3'ss (T/CAG-), with polypyrimidine tract sequences being absent in many budding yeast introns. Due to the well conserved mechanisms of splicing catalysis, almost all eukaryotic intron-containing genes contain a -GT, A and AG- at the 5'ss, branchsite and 3'ss, respectively. One exception to this are the -AT/AC- introns that are common in introns spliced by a subtype of spliceosome called the U12-dependent spliceosome.

Splicing happens by two consecutive transesterification reactions. In the first, the 2'OH of the A within the branchsite makes a nucleophilic attack on the

phosphate group of the first nucleotide of the intron, which is the conserved -G of the 5'ss. This forms the lariat intron-exon intermediate. Then, the 3'OH end of the first exon attacks the phosphate group of the first nucleotide of the second exon, which results in ligation of the two exons and separation of the lariat intron. Although these chemical reactions are primarily catalysed by RNA, splicing of pre-mRNA requires the help of spliceosomal proteins.

Splicing in budding yeast

Splicing in budding yeast is unusual in several respects. Firstly, only about 5% of its protein coding genes contain an intron. However, this does not mean that splicing is rare in this organism, as most of its ribosomal protein (RP) transcripts are spliced (71%) and, because RPs are very abundant, these transcripts represent 24% of the mRNA mass within the cell (Ares et al., 1999). So, highly expressed genes of budding yeast tend to have introns.

Secondly, the vast majority of its intron-containing genes only contain a single intron, and there are only a couple of cases described of functional alternative splicing (Hossain et al., 2016; Juneau et al., 2009). So, virtually all intron-containing transcripts of this organism are constitutively spliced. Nevertheless, the yeast spliceosome is remarkably similar to those of higher eukaryotes (Fabrizio et al., 2009; Will and Lührmann); in terms of protein and U snRNA composition, the assembly pathway, and the mechanisms of splicing catalysis. Probably most differences in protein composition between the budding yeast and human spliceosomes, are due to the lack of proteins that regulate alternative splicing in the former. For this reason, budding yeast is a good model organism to study the mechanisms of splicing.
Thirdly, budding yeast introns are not essential when cells are grown under normal conditions. A large-scale study, where every intron was individually

deleted, revealed that any intron can be deleted without affecting cell fitness. However, when cells were exposed to stress, they discovered that 20% of the intron deletion strains grew slower than wild-type (Parenteau et al., 2008). Furthermore, a similar study associating intron deletion and cell behaviour under stress conditions, found that in 40% of the intron deletion strains, expression of the intron-deleted gene was reduced (Parenteau et al., 2011). Interestingly, they found that about half of the introns of duplicated RP genes (from a total of 118 genes = 59 pairs) are required for normal expression level of the paralog RP gene. In line with this, evidence suggests that splicing regulates the expression of the 13 meiosis-specific intron-containing genes (Juneau et al., 2007), probably through an unusual mechanism, in which pre-mRNA species compete for a limited pool of splicing components (Munding et al., 2013). Therefore, although there is still no consensus of the general function of introns in budding yeast, it seems that they are important during meiosis or when cells are grown under stress conditions, and for RP genes in particular.

The spliceosome cycle

The spliceosome is a highly dynamic complex composed of U1, U2, U4, U5 and U6 uridine-rich small nuclear RNAs (U snRNA) and about 150 proteins [reviewed by (Matera and Wang, 2014; Will and Lührmann)]. Each U snRNA is bound to seven Sm or Lsm proteins and to several specific proteins. These RNA-protein complexes are referred to as U small nuclear ribonuclear protein particles (U snRNP). The spliceosome is assembled from U snRNPs, plus the nineteen complex (NTC) and other non-snRNP splicing factors, through a process that involves major structural and compositional rearrangements of its RNA and protein components.

In summary, the classical model of splicing proceeds as follows (Figure 1.1): 1) the U1 and U2 snRNPs bind the 5'ss and the branchsite (BS), respectively, constituting the pre-spliceosome 2) then the tri-snRNP (U4/U6.U5) and the NTC join in, and U1 and U4 snRNPs are displaced, forming B activated complex, 3) other rearrangements follow that allow for the two catalytic reactions to happen and finally 4) the post-spliceosome is disassembled, releasing the spliced mRNA and the spliceosome components that are recycled into new rounds of splicing.

In the following sections I will describe the key stages of the spliceosome cycle of budding yeast, most of which have been recently modelled by high resolution CryoEM (Fica et al., 2017; Galej et al., 2014, 2016; Hang et al., 2015; Nguyen et al., 2015, 2016b; Wan et al., 2016a; Yan et al., 2015, 2016). snRNP biogenesis is a separate aspect of the splicing process, which I will not describe here.

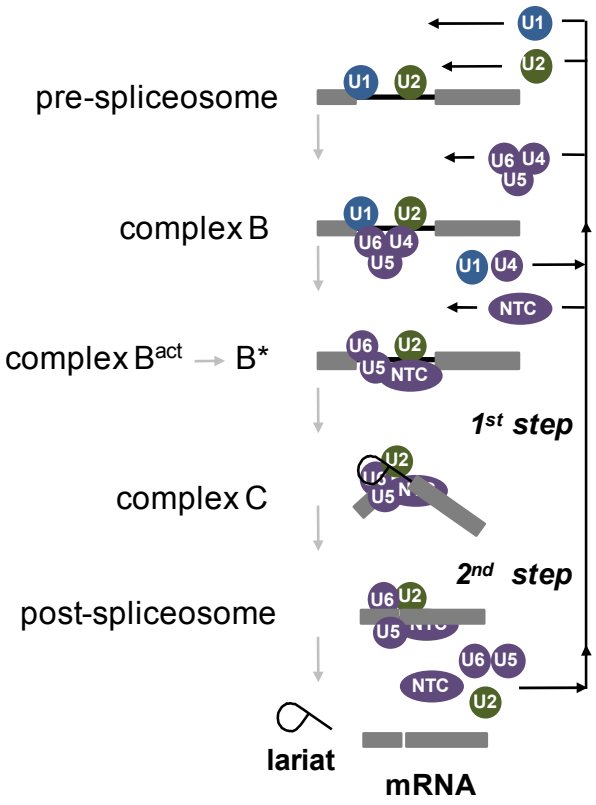


Figure 1.1 Schematic representation of the key stages of the spliceosome cycle.

Pre-spliceosome

The first step of spliceosome assembly is the binding of U1 snRNP to the pre-mRNA, through base pairing between the U1 snRNA and the 5'ss. In parallel, protein Msl5 (Branchpoint Binding protein; Bbp1) and the non-essential protein Mud2 bind the BS and interact with U1 snRNP protein Prp40 (Abovich and Rosbash, 1997). It is believed that this interaction forms a bridge that puts in close proximity the ends of the intron. This stage was given the name commitment complex, as it was suggested that at this point the intron has been selected and committed for splicing.

Then, the U2 snRNP binds the intron through base pairing of the U2 snRNA with the BS which requires 1) RNA helicase Sub2 to displace the Msl5/Mud2 heterodimer from the BS and 2) stabilization of U2 snRNA into the appropriate

structural conformation through the ATPase activity of RNA helicase Prp5. This stage is called the pre-spliceosome. In the canonical model of spliceosome assembly, U2 is recruited after U1 snRNP, which has been confirmed *in vivo* (Tardiff and Rosbash, 2006). However, by single-molecule spectroscopy, it was observed U2 recruited before U1 in about a third of the productive assembly events, suggesting that there may be alternative assembly pathways of the spliceosome (Shcherbakova et al., 2013).

B and Bact complexes

The next stage of assembly starts with the incorporation of the tri-snRNP to the pre-spliceosome or A complex. The tri-snRNP (Nguyen et al., 2015) is the name given to the pre-assembled particle formed by U4, U5 and U6 snRNP, where U4 and U6 snRNAs are extensively base paired. This particle contains more that 30 proteins, including Prp8, which is a very large (280 kDa) and well conserved splicing factor that plays a central role in the spliceosome (Grainger and Beggs, 2005). This intermediate complex of the spliceosome that contains all five snRNPs is called B complex.

Soon afterwards, the U4/U6 duplex is unwounded by RNA helicase Brr2 (Laggerbauer et al., 1998; Raghunathan and Guthrie, 1998a), and the U1 snRNA/5'ss duplex is destabilized by RNA helicase Prp28 (Chen et al., 2001; Staley and Guthrie, 1999), in an ATP dependent manner. These rearrangements allow U6 snRNA to base-pair with the 5'ss and with U2 snRNA, and leads to the dissociation of U1 and U4 snRNPs from the spliceosome and incorporation of the NTC. This stage is called B activated complex (B^{act}).

Nineteen complex

The NTC is composed of up to 16 proteins (Ohi et al., 2002). 8 of these are referred to as NTC core components, including Prp19 and 7 other proteins that were found to be tightly associated with it (Tarn et al., 1994), including Prp46, Cef1 and Syf1. The other 8 NTC components were identified later (Ohi et al., 2002), and are now referred to as NTC-associated or NTC-related, which includes Prp45.

The NTC is recruited to the B complex during or immediately after dissociation of U4 snRNPs, (Tarn et al., 1993) and was shown to be required for stable binding of U5 and U6 snRNPs to the spliceosome and the 5'ss (Chan and Cheng, 2005; Chan et al., 2003). One can now observe in the high resolution Cryo-EM models, that the NTC adopts an extended and interconnected architecture within the spliceosome, and occupies much of the periphery and part of the catalytic core (Yan et al., 2015). Thus, it seems that most of the NTC components work as scaffolds that provide structure and flexibility to the spliceosome.

Interestingly, it could be that some NTC components are recruited to the spliceosome even before U2 snRNP recruitment. This is mainly based on evidence that in mammals the NTC interacts with U2AF65, homologue of Mud2 (David et al., 2011) that is displaced during the commitment complex to pre-spliceosome transition. In addition, in budding yeast the NTC component Syf3 directly interacts with Mud2 and Prp40 (U1) (Chung et al., 1999). However, as NTC components have not been generally found associated to the commitment complex or the pre-spliceosome, it could be that its potential early recruitment happens only on some gene transcripts.

B* and the first catalytic step of splicing

The final step before catalysis is the Prp2-mediated (RNA helicase) release of the SF3a/b U2 snRNP proteins and RES complex (Lardelli et al., 2010; Warkocki et al., 2009). Most of these proteins surround the U2/BS duplex (Schneider et al., 2015; Yan et al., 2016), and in their absence, Cwc25 and Yju2 bind that BS and position it at the 5'ss for splicing catalysis (Chiang and Cheng, 2013; Galej et al., 2016; Wan et al., 2016a). This stage, where the catalytic centre of the spliceosome is ready to splice, is referred to as catalytically activated B complex (B* complex).

The key components of the catalytic centre, are loop I of U5 snRNA that holds-on to the end of the first exon, the U6 snRNA loop bound to the 5'ss, the U2 snRNA bound to the BS, the intramolecular stem loop (ISL) of U6 snRNA carrying two magnesium ions (Mg^{2+}), helix I of the U2/U6 duplex, and the cavity of Prp8 that acts as a scaffold for the catalytic centre (Galej et al., 2013), stabilizing the catalytic RNA structures. This conformation allows the Mg^{2+} ions in the ISL (Fica et al., 2013; Hang et al., 2015; Yean et al., 2000) to catalyse the first step of splicing that produces the lariat intron-exon structure, and is observed with astonishingly high detail on the CryoEM models (Galej et al., 2014, 2016; Wan et al., 2016a; Yan et al., 2015)

C complex and the second catalytic step of splicing

The stage immediately after the first catalytic step of splicing is called C complex. Afterwards, the splice sites are repositioned to form catalytically active C complex (C* complex). Specifically, Cwc25 and Yju2 are displaced by RNA helicase Prp16 through its ATPase activity (Galej et al., 2016; Tseng et al., 2011), which leads to the repositioning of the BS away from the catalytic core to accommodate the 3'ss in its place (Fica et al., 2017; Yan et al., 2016). These

rearrangements are promoted by other splicing factors, additional to Prp16, that are specifically recruited at this stage; including Prp18 and Slu7 that help to position the 3'ss (Fica et al., 2017; Ohrt et al., 2013), which is held by loop 1 of U5 snRNA (Newman and Norman, 1992; Sontheimer and Steitz, 1993); and Prp8 (Teigelkamp et al., 1995). RNA helicase Prp22 acts later for mRNA release. Finally, the two exons are joined by an analogous mechanism as the first catalytic step of splicing.

Post-spliceosome and disassembly

Then, the spliced mRNA is released from the post-catalytic spliceosome, also called the post-spliceosome. After Prp16 has acted and left during the C to C* complex transition, RNA helicase Prp22 binds the second exon at a position around 17 nt downstream of the exon-exon junction. Then, after exon ligation, the 3' to 5 RNA helicase Prp22, disrupts the mRNA/U5 snRNA interactions, releasing the spliced mRNA from the post-spliceosome (Aronova et al., 2007; Fica et al., 2017; Schwer, 2008; Wagner et al., 1998).

Finally, Prp43, Ntr1 and Ntr2, which form an heterotrimer sometimes referred to as NTR, promote disassembly of the post-spliceosome. This heterotrimer is recruited to the post-spliceosome after Prp22 has acted and left (James et al., 2002). The RNA helicase Prp43 then releases the excised lariat from the post-spliceosome through a reaction that requires NTP (Fourmann et al., 2013; Martin et al., 2002; Tsai et al., 2005). This makes the lariat accessible for debranching by Dbr1 (Martin et al., 2002) followed by cytoplasmic 5' to 3' endonuclease degradation (Hilleren and Parker, 2003). Additional to lariat release, the NTR promotes separation of U2, U5, U6, and the nineteen complex (Tsai et al., 2005), resulting in complete disassembly of the post catalytic spliceosome.

The NTR does not act alone, as it was demonstrated that it requires U5 snRNP proteins Snu114 and Brr2 for disassembly (Small et al., 2006). Although the model was later challenged (Fourmann et al., 2013), Small and others proposed that Snu114 regulates Brr2 to unwind the U2/U6 duplex after splicing catalysis, through a mechanism analogous to U4/U6 unwinding at the B complex stage. This model proposed by Small and others, agrees with previous observations that the ATPase activity of Brr2 is particularly stimulated by U2/U6 snRNA (Xu et al., 1996). Furthermore, it was shown that NTR is bound to the U5 snRNP of the post-spliceosome through interaction of Ntr2 with Brr2 (Tsai et al., 2007). Thus, a likely scenario is that mRNA release allows recruitment of NTR to Brr2-U5 snRNP where it coordinates with Snu114-Brr2 to release the excised lariat and separate the remaining components.

It was proposed that when spliceosome assembly happens inefficiently or is interrupted, for example when the pre-mRNA substrate contains sub-optimal splice sites, the aberrant spliceosome is targeted for disassembly by NTR through an analogous mechanism as the one described above. This aspect of disassembly is explained in more detail in the "Splicing fidelity" section below.

Recycling

Additional rearrangements are required before the released U snRNPs can be recycled. Some but not all aspects of U snRNP recycling have been described. Of particular importance is the process of reannealing U6 with U4 snRNA that precedes tri-snRNP reassembly. RNA-binding protein Prp24 is essential for this process (Ghetti et al., 1995; Raghunathan and Guthrie, 1998b). After post-spliceosome disassembly, U6 snRNA binds Prp24. Then, its internal stem loop (ISL) that drives splicing catalysis is destabilized followed by U6 reannealling

with U4 snRNA, giving rise to U4/U6 snRNP. In addition, U4/U6 snRNP formation requires the association of U6 snRNA with its Lsm proteins that were released during the B to B^{act} transition (Achsel et al., 1999; Mayes et al., 1999; Ryan et al., 2002; Verdone et al., 2004). Based on the crystal structure of U6 snRNA-Prp24 (Montemayor et al., 2014) together with data of previous reports (Bae et al., 2007; Didychuk et al., 2016; Martin-tumasz et al., 2011), it was proposed that Prp24 acts like a chaperone that stabilizes the transitional conformation of U6 between its catalytic conformation and that of di-snRNP.

Afterwards, U5 binds U4/U6 snRNPs to form the tri-snRNP, through a mechanism elucidated for human cells but that could also exist in yeast (Song et al., 2010). Firstly, Prp19 binds to U4/U6 snRNPs and then ubiquitinates Prp3 (U4). This increases affinity of Prp3 for Prp8 (U5) resulting in U5 association with U4/U6 snRNPs to form the tri-snRNP. In addition, it was proposed that release of U4 during the B to Bact transition of the spliceosome is aided by Prp3 deubiquitination.

Post-catalytic U2 snRNP also undergoes structural and compositional rearrangements before it is re-used (Behrens et al., 1993; Brosi et al., 1993; van Roon et al., 2017). U2 snRNP is released from the post-spliceosome as a particle that contains Sm proteins and U2-specific factors Msl1 and Lea1, and is analogous to the 12S U2 particle of human cells. Next, this particle binds three SF3a and six SF3b proteins displaced during the Bact to B* transition, forming the active U2 snRNP that is analogous to the 17S U2 of human cells. At this point, the U2 particle is ready to join the commitment complex, leading to pre-spliceosome formation.

Co-transcriptional splicing

The idea that splicing can occur co-transcriptionally was first proposed based on the observation that spliceosomes are bound to splicing junctions of nascent RNA (Osheim et al., 1985). But it wasn't until two decades later that a series of snRNPs ChIP analyses demonstrated that step-wise spliceosome assembly occurs co-transcriptionally, in agreement with the classical model of splicing (Görnemann et al., 2005; Kotovic et al., 2003; Lacadie and Rosbash, 2005; Listerman et al., 2006). These data show co-transcriptional recruitment of U1 snRNP followed by U2 snRNP and, finally, U5 snRNP together with the NTC. The rationale behind these assays is that if splicing occurs co-transcriptionally, splicing factors can be cross-linked to DNA, as nascent pre-mRNA lies adjacent to the DNA (Wetterberg et al., 2001).

It was later demonstrated, by high density tiling microarray analysis of nascent RNA, that most introns are co-transcriptionally spliced (Carrillo Oesterreich et al., 2010). The authors observed that >80% of pre-mRNA is spliced before transcription termination, in close to half of the intron-containing genes analysed. Surprisingly, they showed that even genes with short second exons splice co-transcriptionally, and argued that this is due to Pol II pausing on terminal exons. In agreement, a parallel study demonstrated the existence of splicing-dependent Pol II pausing near the 3′ splice site (Alexander et al., 2010a). Although it is possible that Pol II pausing observed by Oesterreich and others (2010) and Alexander and others (2010) are of a different nature. Recently, it was shown by nascent RNA sequencing, that splicing can happen incredibly close to the exit channel of Pol II (Carrillo Oesterreich et al., 2016). Overall, the data support the idea that splicing is tightly coupled to transcription. Most of the first analyses of co-transcriptional splicing were done on budding yeast. However, they were soon followed by

other reports confirming that most splicing in higher eukaryotes also happens co-transcriptionally (Brugiolo et al., 2013).

Mechanisms of co-transcriptional splicing

Two non-mutually exclusive models have been proposed to explain co-transcriptional splicing. In the "recruitment model", co-transcriptional splicing is facilitated by direct recruitment of splicing factors to the transcription machinery. The first evidence in support for this was the demonstration that the absence of the carboxy-terminal domain (CTD) of the RNA Polymerase II (Pol II) can reduce splicing efficiency in mammalian cells (McCracken et al., 1997), suggesting that the CTD acts as a platform that recruits and directs RNA-processing machines. Since then, some physical links have been found between Poll II and splicing factors, including human orthologues of Mud2 and Prp19, and yeast Prp40 (U1) and Syf1 (NTC) (Chanarat et al., 2011; David et al., 2011; Morris and Greenleaf, 2000).

The kinetic model states that transcription elongation rate is modulated to allow time for splicing to occur before transcription termination, offering a "window of opportunity" for splicing factor recruitment. In this scenario, the rate of splicing competes with transcription elongation rate, and possibly also with 3′ end formation. In support for this model, it was shown that splicing is altered in response to changes of Pol II elongation rate. Of particular importance, experiments with budding yeast and mammalian cells show that a slow Pol II can favour use of upstream weak splice sites (Howe et al., 2003; Mata et al., 2003). In addition, it was observed that Pol II mutants of slow and fast transcription elongation rates affect splicing efficiency (Braberg et al., 2013) and fidelity (Aslanzadeh et al., in review).

As an extension of the kinetic model, a splicing dependent transcriptional checkpoint was proposed. This checkpoint model is based on evidence that a transcriptional elongation defect takes place, when stable binding of U2 snRNA with the BS is compromised, and that this transcription defect is mediated by transcriptional elongation factor Cus2. This suggests that there is a checkpoint that triggers Pol II pausing when pre-spliceosome formation is less efficient, as if transcription is "waiting" for splicing (Chathoth et al., 2014). So, Pol II pausing could be a mechanism by which the kinetics transcription is regulated in response to splicing.

Splicing fidelity

Accurate splicing is accomplished not only by selecting the appropriate splice sites, but also by suppressing suboptimal ones – a concept known as splicing fidelity. The kinetic model of splicing fidelity proposes that proofreading mechanisms reject suboptimal substrates due to their slow kinetic rates of catalysis [reviewed by (Semlow and Staley, 2012)]. Spliceosomal RNA helicases are central to this model.

Spliceosomal RNA helicases as fidelity factors

Most spliceosomal RNA helicases interact transiently with the spliceosome and are special in that, through their ATPase activity, they trigger the structural rearrangements that define the various phases of the spliceosome cycle (as described in the previous sections). In addition to their main role in splicing, there is evidence to suggest that these proteins proofread the first and the second step of splicing, by rejecting sub-optimal splice sites through their RNA unwinding activity.

Prp16 and the kinetic model of splicing fidelity

Prp16 was the first RNA helicase to be described as a splicing fidelity factor. An early report identified a *prp16* mutation that suppresses a splicing defect caused by a branch site mutation from TACTA**A**C to TACTA**C**C within a reporter gene (Couto et al., 1987). This work was soon followed by further characterizations of the mechanism of Prp16's activity (Burgess and Guthrie, 1993; Burgess et al., 1990). Altogether, these studies lead to the proposal of a molecular framework of splicing fidelity by the Guthrie lab. This model, which is referred to as the kinetic proofreading model, or the timer model, proposes that the action of RNA helicases can antagonize splicing catalysis, and that the rate of ATP hydrolysis by RNA helicases competes with the rate of splicing catalysis. In other words, sub-optimal substrates are rejected because they are not spliced quickly enough. For instance, if a spliceosome is assembled on a pre-mRNA containing a non-consensus branchsite sequence that does not properly base-pair with U2 snRNA, Prp16 may release Cwc25 and Yju2 before the intron is branched.

More recent studies have confirmed and extended the splicing fidelity model that was first proposed. Of particular importance, is evidence produced by the Staley lab, which showed that the Prp16-dependent discard of a pre-mRNA with mutated branchsite requires spliceosome disassembly factor Prp43 (Koodathingal et al., 2010; Mayas et al., 2010). This suggests that proofreading happens by two consecutive steps. In the first, the non-consensus splice site is rejected by an RNA helicase (*e.g.* Prp16) that would shift the structure of the catalytic centre into an unproductive state. This conformation would then trigger RNA helicase Prp43 to discard the sub-optimal substrate and disassemble the aberrant spliceosome (Figure 1.2).

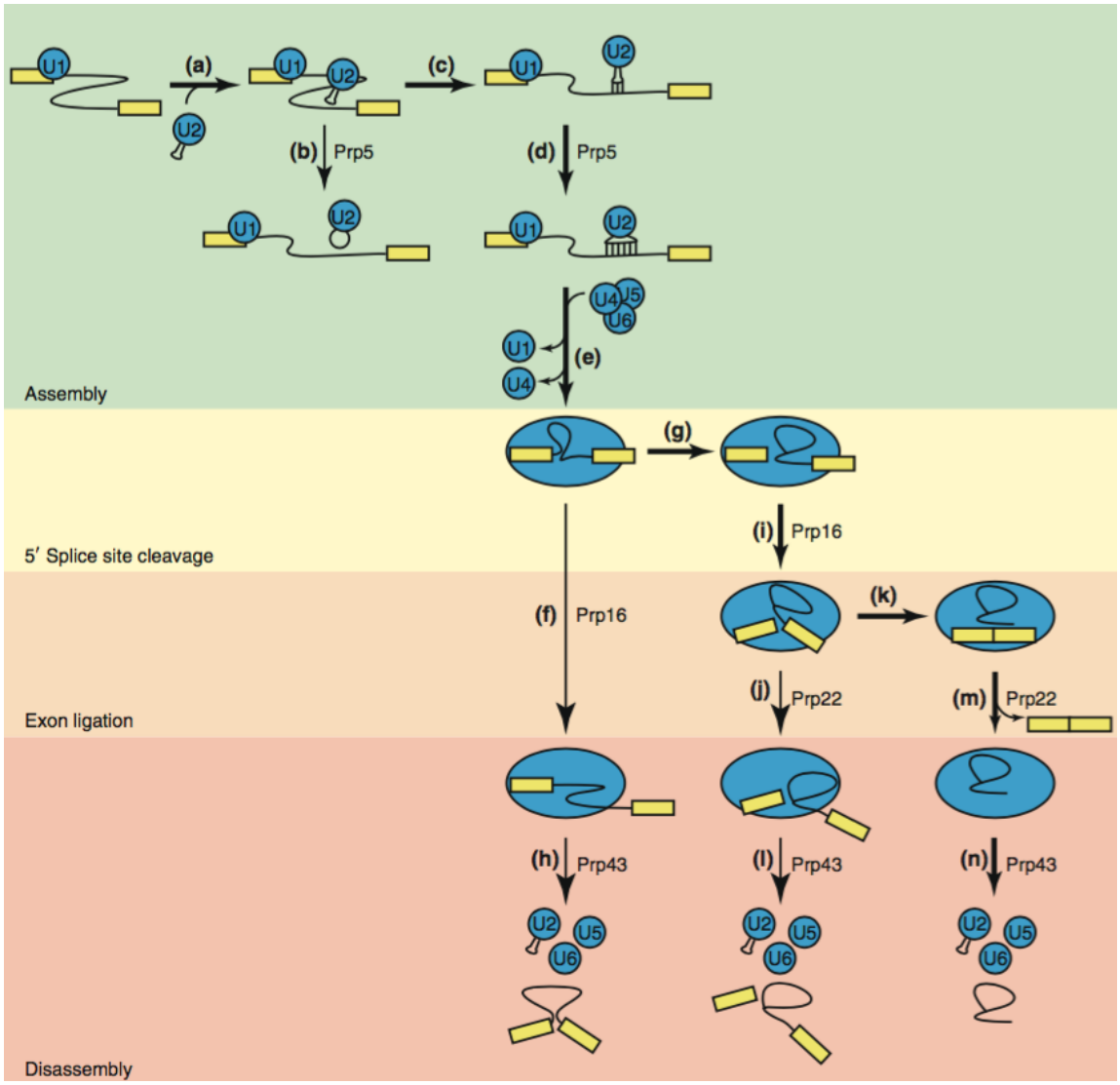


Figure 1.2. (Semlow and Staley, 2012) "Functions of known proofreading DExD/H-box ATPases in the splicing pathway".

In another report by the same lab, an *in vitro* assay was developed where the rate of splicing was reduced by introducing a mutation in the ISL of U6 snRNA that disrupts the catalytic centre. Interestingly, they observed a correlation between a reduced rate of splicing and an increased rejection of defective spliceosomes (Koodathingal et al., 2010), which required ATPase activity of Prp16 – arguably the best mechanistic evidence so far in support of the kinetic model of splicing fidelity.

Prp22's role in proofreading the 3'ss

Prp22 was the second RNA helicase to be described as a splicing fidelity factor. Although this role of Prp22 has not been as well characterized as that of Prp16, there is good evidence to support it. Of special importance is *in vitro* work from the Staley lab (Mayas et al., 2006). In wild-type conditions, a pre-mRNA containing a UAG to UAC mutation in its 3'ss undergoes the first but not the second step of splicing. However, when recombinant mutant Prp22 proteins, with impaired ATPase and/or RNA unwinding activity, were added to affinity purified spliceosomes, the UAC pre-mRNA could now undergo the second step of splicing to some extent. This suggests that the ATPase activity of Prp22 suppresses sub-optimal 3'ss, and that splicing proofreading is a general function of spliceosomal RNA helicases.

Furthermore, the authors observed that the transcript of a *ACT1-CUP1* reporter, containing a UAG to GAG mutation at its 3'ss, was spliced *in vivo* by a cold-sensitive (ATPase deficient) *prp22* strain but not by wild-type strain. Although the reporter they used is an artificial gene, this confirms that the ATPase activity of Prp22 has an effect on 3'ss selection *in vivo*. Another intriguing observation is that the *prp22* mutants also affected the second step of splicing of a substrate containing non-consensus 5'ss or branchsite sequences. This was observed by purifying arrested spliceosomes containing branched intermediates with a GUAUGU to AUAUGU mutation at the 5'ss, or a UACUA**A**C to UACUA**G**C mutation at the branchsite. Then, addition of recombinant mutant Prp22 protein, but not wild-type Prp22 protein, allowed exon ligation of some proportion of the branched substrates. This suggests that Prp22 may also assist in the proofreading of 5'ss and branchsite, although it remains to be seen whether this happens *in vivo*.

Consistent with its 3'ss proofreading role, previous evidence showed that Prp22 crosslinks around 17 nt downstream of the exon-exon junction (Schwer, 2008) and, in addition, recent Cryo-EM models have located Prp22 in the yeast (Fica et al., 2017) and human (Bertram et al., 2017) C* complex of the spliceosome (Figure 1.3). Although Prp22 was observed to interact with Prp8 away from the catalytic centre, this distance can accommodate 15-16 nt and therefore allow for a potential interaction between Prp22 and the first 15 nt of the second exon. Taken together, the data indicate Prp22 is timely and spatially positioned to translocate along the pre-mRNA in a 3' to 5' direction and, therefore, may potentially disrupt interactions that stabilize the 3'ss at the catalytic centre just before exon ligation. This is also consistent with evidence indicating Prp22 discards suboptimal substrates before but not after exon ligation (Mayas et al., 2006), and that Prp22-mediated rejection correlates with separation of the 3'ss and 5'ss (Semlow et al., 2016).

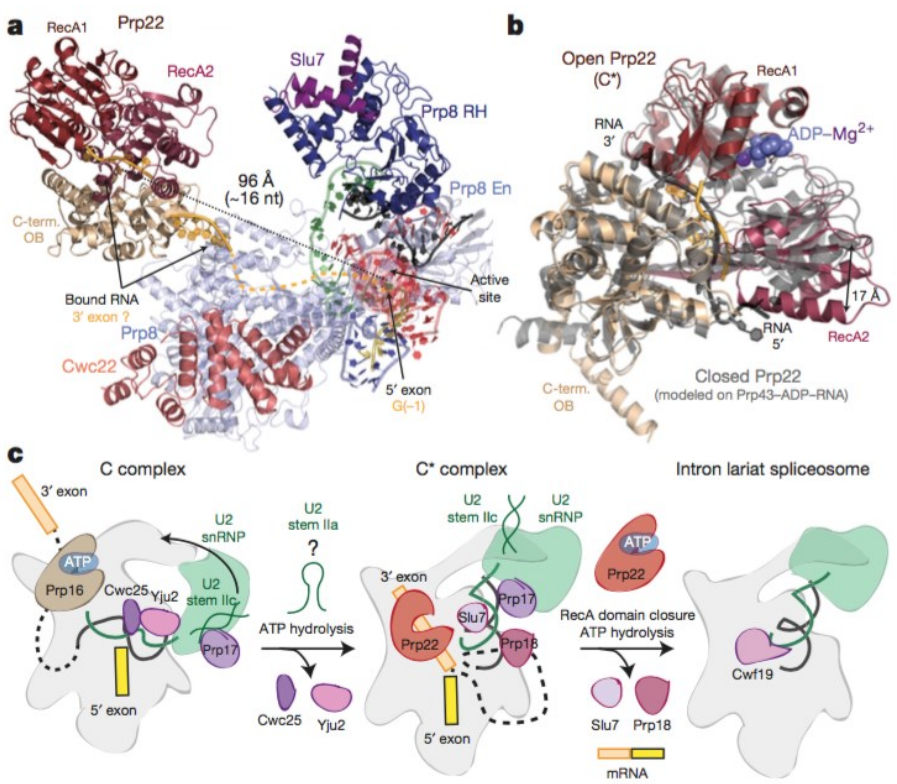


Figure 1.3. (Fica et al, 2017) "Prp22 and ATP-mediated transitions at the catalytic stage of splicing". Figures a and b are representations of the structures resolved by the authors through CryoEM, focusing on the position of Prp22 within the the C* complex. Figure c is a schematic of the model of the Prp16-mediated remodeling.

Prp16 and Prp22 may act by pulling the RNA at a distance

It was recently proposed, also by the Staley lab, that Prp16 and Prp22 influence splicing fidelity at a distance from the splice sites (Semlow et al., 2016). Evidence for this is based on *in vitro* splicing extracts with artificial pre-mRNA substrates containing 1) a chemically modified branchsite that is inefficiently branched or 2) two 3'ss separated by 3 nt, both with UAG consensus sequence. The first and the second type of substrates were used to study the role of Prp16 and Prp22, respectively. Using this assay, they first showed that the ATPase activity of these proteins is required for selection of alternative splice sites. For instance, either of the tandem 3'ss were spliced in wild-type splicing extracts, but when an ATPase-deficient mutant Prp22 protein was added to the extract, only the distal 3'ss was used.

Next, they truncated the substrates at different positions downstream of the branchsite or 3'ss splice site. With this they confirmed that a 3' tail after the splice sites is required for the function of Prp16 and Prp22, as previously suggested. Alternatively, by introducing DNA substitutions into the RNA substrate, they blocked the ability of the RNA helicases to translocate along the substrate beyond specific positions. Unexpectedly, they observed that the DNA substitutions closest to the splice sites did not inhibit the function of Prp16 or Prp22, suggesting that these proteins do not translocate through the splice sites, as was previously proposed based on crosslinking data (McPheeters and Muhlenkamp, 2003; Schwer, 2008). Overall, their data suggest 1) that Prp16 and Prp22 are not only important for splicing fidelity,

but possibly also for alternative splicing, as they allow flexibility of splice site selection, and 2) that these proteins act by pulling the RNA substrate at a distance, possibly to allow the catalytic centre to sample several potential splice sites. The latter is in agreement with the recent Cryo-EM models of the spliceosome (see above subsection).

Prp5's role in proofreading the branchsite sequence

RNA helicase Prp5 has also been associated with splicing fidelity (Xu and Query, 2007). Several *prp5* mutants were identified that, compared to wild-type *PRP5*, increase the *in vivo* splicing efficiency of a pre-mRNA containing single nucleotide substitutions at one of two positions (underlined) of the branchsite sequence TAC<u>TA</u>**A**C, which are otherwise poorly spliced with wild-type *PRP5*. They also showed that *prp5* mutants increase, although only slightly, the splicing efficiency in the presence of a mutation in the U2 snRNA that disrupts its base-pairing with the wild-type branchsite. Overall, the data suggest that Prp5 proofreads branchsite selection. However, the mechanism for this may be different to that of Prp16- or Prp22-mediated splicing proofreading, as Prp5 appears to proofread branchsite recognition by stabilizing U2 snRNA base pairing with the branchsite, instead of actively rejecting sub-optimal splice sites as is proposed for Prp16 and Prp22.

Non RNA-helicases associated with splicing fidelity

Splicing requires the correct assembly, structure and timely coordination of more than 150 protein and RNA components. For this reason, it is expected that many of these components contribute, directly or indirectly, to the selection of proper splice sites. For example, it has been observed that *isy1*Δ or *cwc21*Δ, or mutations in Prp8, Slu7, U2, U5, U6 snRNAs affect the accuracy of splice site selection (Chang and McPheeters, 2000; Frank and Guthrie, 1992;

Gautam et al., 2015; Lesser and Guthrie, 1993; Newman and Norman, 1992; Umen and Guthrie, 1996; Villa and Guthrie, 2005). However, my focus is not splice site selection *per se*, but rather the proposed model of splicing fidelity where RNA helicases reject sub-optimal splice sites. Although the process of selecting the optimal splice site and the process of rejecting sub-optimal splice sites are interdependent, they are two different aspects of splicing. While the first represents the correct functioning of the machine, the second represents quality control.

Tools to control gene expression in budding yeast

To study the *in vivo* function of an essential gene, it is very common to start by conditionally blocking its expression or activity, followed by phenotype analysis. Ideally, the conditional expression system should be fast and specific, to avoid secondary and off-target effects. The traditional conditional systems for budding yeast are temperature-sensitive mutants and metabolic regulation (*e.g. GAL* and *MET* promoters). These tools have been extremely useful in functional studies. However, they have several inconveniencies, including the secondary effects associated with temperature or media changes. Alternative tools were developed more recently, including artificial on/off promoters, anchor-away, riboswitches and degrons. In the following sections I will present an overview of these systems.

Temperature-sensitive mutants

Temperature sensitive (ts) strains typically contain a mutation in a single essential gene, which renders them able to grow at the permissive, but not restrictive temperature. Most mutant strains are generated by incubating the cells with a mutagen (e.g. EMS), followed by strain selection based on temperature phenotype and, finally, identification of the mutant gene. The

permissive temperature can range between 24-30 °C, and the restrictive temperature is around 16°C, for cold-sensitive mutants (cs), or 30-37 °C for heat-sensitive mutants (hs). It is believed that the heat-sensitivity is due to a structural instability of the mutant protein, and the cold-sensitivity to an inability to destabilize RNA:protein or RNA:RNA interactions (*e.g.* cs RNA helicases). Thus, by shifting the temperature of a ts culture one can almost instantaneously inactivate the function of protein of interest.

Temperature-sensitive mutants are the "workhorse" of yeast genetics. They are especially useful for screening libraries of ts strains against a reporter system. For example, a good proportion of splicing factors were identified by isolating strains from a ts library for their inability to splice. Also, mutants can be crossed with other mutants or with deletion strains, to identify suppressors or synthetic lethal strains that would allow one to infer functional interactions between proteins or RNA components. This can be done following a low-throughput or high-throughput approach. Over the past three-four decades, a combination of powerful yeast genetics and biochemical studies, allowed researchers to understand in great detail most of the assembly pathway of the spliceosome and the mechanisms of splicing catalysis.

Despite the extreme usefulness of ts mutants, they are far from ideal when it comes to individually studying the functions of genes. The main reason for this is that temperature shifts stress the cells, affecting splicing efficiency and other cellular processes (Bergkessel et al., 2011). Therefore, the specific effect of the mutation on splicing may be obscured by non-specific effects. In addition, many ts strains are sick even at the permissive temperature, which further decreases the contrast between the permissive and restrictive conditions.

Conditional promoters

Two types of promoters are used as molecular tools, constitutive promoters, such as P*ADH1*, P*TEF1* and P*PGK1*; and conditional promoters (CP). CPs allow control of timing and level of expression of the gene of interest (GOI). Most of these can be used to either repress or induce transcription. CP are commonly used as research tools to study gene function, and also for over-expression of heterologous proteins for industrial biotechnology.

One benefit of using a CP to inhibit expression, over other knockdown techniques, is that the sequence of the GOI need not be altered. However, by swapping the native promoter of the GOI for a CP means that, prior to repression, the expression levels of the GOI are different from wild-type. The second inconvenience is that the speed of repression is limited by the turnover rate of both the mRNA and the protein product of the GOI. That is why, it is not unusual having to wait 12-24 hours (*e.g.* after galactose to glucose shift) before complete protein depletion is observed. On the other hand, maximum protein levels are typically reached within 0.5-2 hours after inducing expression through CP.

The large variety of conditional promoters (CP) can be categorized based on the type of molecule that controls their transcription. There are natural occurring promoters that respond to cell metabolites or metals (*e.g.* galactose or Cu^{2+}), and there are modified promoters that respond to non-metabolizable drugs (*e.g.* tetracycline or B-estradiol). Each CP has a combination of characteristics that influence its overall usefulness, including but not limited to: 1.- expression level 2.- leakiness 3.- dynamic range 4.- tunability 5.- innocuousness and 6.- specificity. Although obviously not all work equally well, the question of which CP is most suitable depends on the intended use.

Native conditional promoters

GAL1, *GAL5*, *GAL7* and *GAL10* genes encode enzymes required to catabolize galactose. The expression of these, as well as many other metabolic-pathway genes, is highly regulated to manage cell resources based on the availability of nutrients. As high priority is given to glucose over other carbon and energy sources, *GAL* genes are highly expressed in the presence of galactose but strongly repressed by glucose. A glucose to galactose shift in the culture medium can result in a 1000-fold increase in transcription of galactose-controlled genes (Weinhandl et al., 2014). For these reasons, galactose-induced promoters (*e.g.* P*GAL1* and P*GAL10*) are one of the most commonly used CPs.

Key regulators of GAL promoters are transactivator Gal4 and its repressor Gal80. Gal4 activates GAL promoters by binding to their upstream activation sites (UAS) as a homodimer. Expression of Gal4 is repressed by glucose; and in absence of galactose, Gal80 forms a heterodimer with Gal4, thereby inhibiting its activity. The main disadvantage of the GAL expression system is that changing the nutritional contents of the medium can drastically affect global gene expression. For instance, it was observed that when cultures grown on galactose were subjected to pulses of glucose, the expression of about 25% of genes was changed at least 2-fold (Ronen and Botstein, 2006).

Other commonly used CPs are copper-induced and methionine-repressed promoters. When intracellular Cu^{2+} ions reach high levels, the expression of Cu^{2+} responsive genes is upregulated to reduce the excess Cu^{2+} and prevent intoxication. One of these genes is *CUP1*, which encodes a metallothionein protein that binds and sequesters Cu^{2+}. Expression of *CUP1* is activated when transcription factor Ace1 binds metal regulatory elements within P*CUP1*, in the presence of Cu^{2+}. Thus, by situating P*CUP1* directly upstream of the GOI, one can induce its expression by adding $CuSO_4$ (0.1-1mM Cu^{2+}) to the culture

medium, in a reversible manner. One benefit of using P_{CUP1} is that its induction does not require a media change, addition of $CuSO_4$ does not perturb the cells as much as carbon source shifts, and high expression levels have been observed (Da Silva and Srikrishnan, 2012). However, it has been reported to be leaky without addition of its inducer, and its dynamic range is usually low (10-25 fold) (Maya et al., 2008).

MET3 and *MET25* encode two enzymes required for methionine biosynthesis and metabolism. P_{MET3} or P_{MET25} have been used to conditionally repress expression of GOI (Mao et al., 2002). These promoters are characterized by a weak expression in the absence of Met, probably 10-fold less compared to a GAL promoter on gal+/glu- medium, and by a good repression by addition of Met (> 0.1 mM Met) to the medium, about 10-fold less expression than in Met-. Therefore, compared to GAL promoters, MET promoters can be used when wanting to avoid overexpressing the GOI (aiming at close to wild-type levels) in Met-, while avoiding the drastic nutritional changes in the growth medium when using GAL promoters.

Drug-controlled promoters

Several promoters have been engineered to respond to drugs. There are significant benefits to this. Firstly, the drugs used for this purpose are non-metabolizable, so adding small amounts of them to the culture medium can have a strong and lasting effect on GOI expression. Secondly, these drugs do not normally interfere with cell metabolism so secondary effects are minimal. Two good examples are the Tet On/Off system and the B-estradiol system.

Tet On/Off expression system

The Tet-On/Off expression system was first developed for mammalian cells (Gossen and Bujard, 1992) and adapted to yeast (Garí et al., 1997) shortly afterwards, and has worked efficiently in both of these organisms. The DNA element of the system is composed of 2 or 7 copies of the bacterial tetracycline-resistance operator (*tet*O) fused to a minimal CMV or *CYC1* promoter, which together form the tetracycline-responsive promoter P*tet*. The original protein element of the system is the bacterial tetracycline repressor (*tet*R). In the presence of tetracycline, *tet*R represses P*tet* by binding to its *tet*O element. In the yeast Tet-Off, *tet*R is fused to transcriptional repressors Ssn6 or Tur1. The Tet-On version was created by fusing *tet*R to the activation domain of virion protein 16 (VP16) of herpes simplex virus. This fusion protein is called tetracycline-controlled transactivator (*t*TA) and, in the presence of tetracycline, it activates P*tet* when bound to its *tet*O element.

Later, a mutated *tet*R was created that behaves in the opposite way to the wild-type version, as it only binds to *tet*O in the absence of tetracycline. These mutants (tetR' and *t*TA') were then used in combination with *tet*R and *t*TA, to create the dual Tet-On/Off system with improved expression control. For example, in the dual Tet-ON, when tetracycline is added to the cells, the *tet*R' is released from the P*tet* and *t*TA' binds in its place. So, the coordinated action of repressor and activator ensures both tight-control and strong expression. For instance, in the dual Tet-On/Off of yeast, the reporter protein can reach maximum levels comparable to that of the GAL system, with an impressive dynamic range of more than 1,000-fold (Bellí et al., 1998).

The general experience of the Beggs lab (personal communication from several lab members), is that the Tet-On/Off system can work well, but generating new, reliable Tet strains may be troublesome. The reason for this,

is based on the speculation that activator and repressor should be expressed to optimal levels for the system to work efficiently. If true, one should test a combination of constructs to find one that works well. For instance, the lab was able to generate a Tet-On strain only after several rounds of trial and error (Alexander et al., 2010b). An additional complication is that several rounds of transformation are required, as the system is composed of at least three separately encoded elements: promoter, activator, and repressor. This can be solved by putting all elements into a single centromeric plasmid; however, a recent attempt to accomplish this was not entirely successful.

The B-estradiol expression system

Relatively recently, a B-estradiol-inducible expression system was developed, in which an artificial transcription factor (ATF) activates an artificial promoter, only in presence of B-estradiol human hormone. Although this system is easy to use, highly efficient and very specific, at the moment it has only been tested for budding yeast.

The ATF of this system is a fusion of a DNA-binding domain, the human estrogen receptor and the VP16 transactivator. In the latest version (McIsaac et al., 2013), the DNA-binding domain is a truncated Gal4 than contains, instead of its own DNA-binding motif, a triple zinc-finger array from mouse (Z_3) or a rationally designed quadruple zinc-finger array (Z_4). The ATFs containing these elements are referred to as Z_3EV and Z_4EV. The artificial promoter is a modified P_{GAL1}, where the Gal4-binding sequence was exchanged for the corresponding zinc-finger-binding sequence; these promoters are referred to as Z_3EVpr and Z_4EVpr (in accordance with yeast gene nomenclature, in the results chapters I refer to them as P_{Z3EV} and P_{Z4EV}).

The B-estradiol expression system works in the following way. In the absence of B-estradiol, chaperone Hsp90 binds to the human estrogen receptor of the Z_nEV and stops this protein from translocating from cytoplasm to the nucleus. Conversely, in the presence of B-estradiol, Z_nEV translocates to the nucleus where it binds to the zinc-finger-binding sequence within Z_nEVpr, which results in transcriptional activation of this promoter (Figure 1.4). Using *GFP* as a reporter gene controlled by the Z_4EV, the authors observed about a 50-fold and 150-fold increase in *GFP* transcript abundance 15 and 60 minutes, respectively, after B-estradiol addition (1 uM) to the medium. Similar, but slightly lower induction levels were observed using the Z_3EV version.

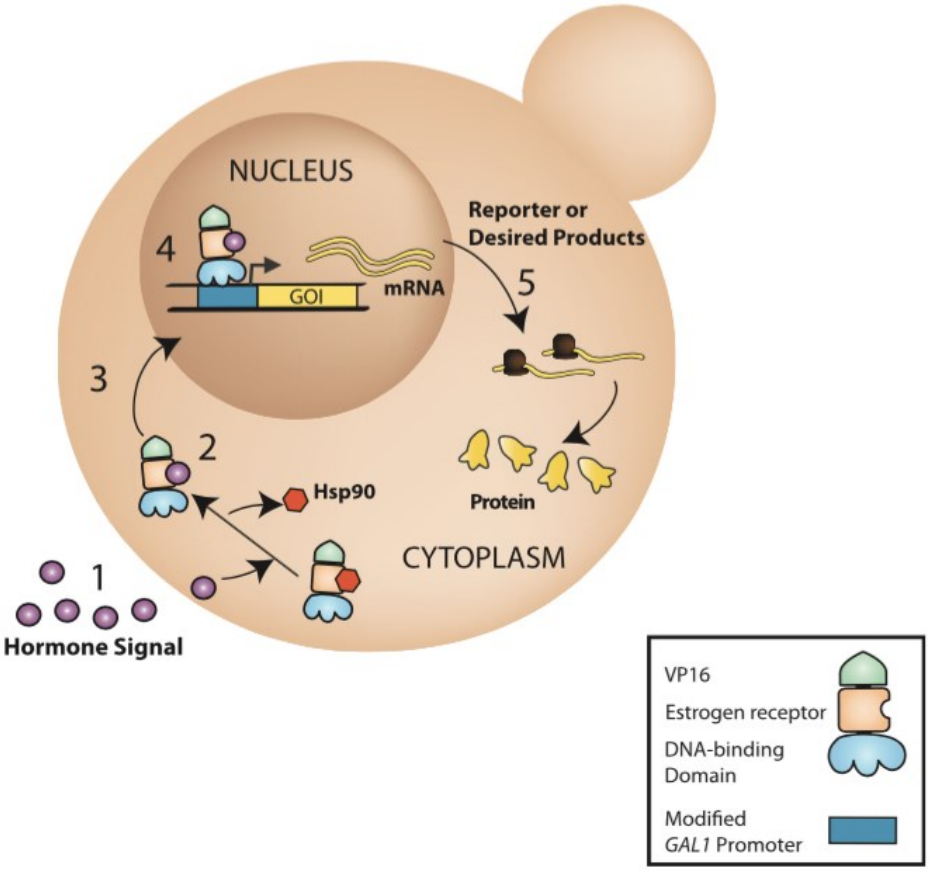


Figure 1.4. "Schematic of hormone-based gene expression system" (McIsaac et al, 2013).

Furthermore, they showed that incubating cells with B-estradiol 1) does not reduce growth rate and 2) does not affect the global transcription profile of the cells. Evidence for this is based on a microarray analysis showing that only two transcripts, from non-essential genes *RPS8A* and *YDR133C*, are 2-fold more abundant compared to the control. In contrast, a similar analysis of a previous version of the B-estradiol system, which is constituted by an ATF with full-length Gal4 and no zinc-fingers, showed that the abundance of about 400 gene transcripts is either reduced or increased by > 2-fold following B-estradiol incubation. Overall, their data indicates that induced-expression through the B-estradiol/zinc-finger system is strong and highly specific.

Riboswitches

Riboswitches are RNA-based devices that can be used to control gene expression. Typically, these RNA molecules reside within a non-coding region of an mRNA, and are composed of one or more aptamers that changes structure when bound to a small-molecule ligand. This change in structure can then regulate expression of its host mRNA through a variety of mechanisms, including modulation of transcription termination, translation inhibition, modulation of mRNA stability and alternative pre-mRNA splicing. The ligands of riboswitches can be cell metabolites (*e.g.* vitamins), ions or metals, or even antibiotics.

All known natural-occuring riboswitches exist within prokaryotes, with the one exception of the Thiamine riboswitch that regulates alternative splicing in several eukaryotic organisms (Nguyen et al., 2016a). Their function is to sense the intracellular levels of small molecules. One example, there is a riboswitch that binds to Mg^{2+} and resides within the 5' UTR of *mgtA* transcripts of *Salmonella enterica.* High levels of Mg^{2+} favour formation of a stem-loop that

promotes Rho-dependent transcriptional termination of *mgtA*, and therefore prevents transcription of the coding region of this gene, which encodes a transporter of Mg^{2+} (Hollands et al., 2012).

Since riboswitches were discovered, many of them have been engineered as tools to control trans-gene expression. Perhaps the most important advantages of riboswitches, are that they are easy to use, as they are constituted by a single element; and that they can be modified, for example to switch the specificity of the aptamer so that it binds to a different molecule. There are not many examples of engineered riboswitches for control of gene expression in yeast. However, there is one worth mentioning that worked with high efficiency. This riboswitch was generated by inserting an array of three aptamers before the start codon of the gene. When bound to their ligand tetracycline, the aptamers adopt a strong secondary structure, which is believed to prevent stable binding of ribosomes to the mRNA. Using this tool, an almost complete depletion of the target protein was achieved after 4-6 hours of tetracycline addition (Kötter et al., 2009).

Domain to induce degradation

A quick way to conditionally knockdown a gene product, is to directly degrade the protein of interest (POI). This is accomplished by fusing the POI with a domain that induces degradation, known as a degron [reviewed by (Kanemaki, 2013)]. Generally, the advantages of degrons are 1) fast depletion, 2) most genes can be targeted, 3) not necessarily species-specific and 4) reversibility. Typically, depletion of the target is achieved within 0.5-2 hours after induction. One disadvantage is that the fused degron by itself, without induction, may disrupt the POI by altering its structure or directly interfering with its function, which is a general risk of protein fusions. There are several

types of degrons, including the temperature-sensitive (ts) degron and the auxin-inducible degron. In addition, there is a different class of degrons developed for mammalian cells, which rely on conditionally exposed hydrophobic domains that are also targeted for degradation by the proteasome. However, these degrons either do not work on yeast, or have not been tested on this organism (Kanemaki, 2013).

Temperature-sensitive degron

The ts degron is comprised of a ubiquitin moiety followed by ts protein Dhfr, and is fused to the N-terminal of the POI. When this fusion protein is expressed, the ubuiquitin moiety is cleaved by a deubiquitinating enzyme, releasing the N-terminal Met. Then, by shifting the temperature to 37°C, the Dhfr unfolds exposing its N-terminal Arg and therefore, as this amino acid residue is not a Met (the N-end rule), the fusion protein is rapidly polyubiquitinated and degraded by the proteasome (Dohmen and Varshavsky, 2005). The ts degron worked well on yeast. However, because temperature shifts cause unwanted secondary effects and because the ts degron could not be implemented in mammals, this system is not widely used. Similar to the ts degron, the protease-induced protein inactivation (TIPI) degron works by exposing an N-terminal non-Met amino acid, caused by specific cleavage of the N-terminal end of the target protein by the induced-expression TEV protease (Taxis and Knop, 2012).

The auxin-inducible degron

The auxin-inducible degron (AID) was developed by transplanting the plant auxin/Tir1 pathway into other eukaryotes (Nishimura et al., 2009). The main benefits of this system are 1) the inducer does not perturb the cells and 2) it can be used with high efficiency in many eukaryotes, including budding yeast,

fission yeast, human cells, chicken cells and roundworm. For these reasons, the AID has quickly become a popular tool to study gene function in non-plant eukaryotes.

The physiological role of the plant hormone auxin, indole-3-acetic acid (IAA), is to mediate growth and development in plants. The process starts with binding of auxin to its receptor, a substrate specific F-box protein called Tir1, which is part of the plant E3 ubiquitin ligase SCF complex. Then, Auxin/Tir1 binds to its target, an IAA transcriptional repressor and, as a result, IAA is poly-ubiquitinated by the SCF complex and then degraded by the proteasome.

The AID system exploits the fact that Tir1 is specific to plants; and that the SCF complex and the ubiquitin proteasome pathway are evolutionarily conserved. Therefore, this protein depletion system can be readily transplanted to non-plant organisms. This requires heterologous expression of Tir1 and C-terminal tagging the gene of interest with the degron (*e.g. IAA17* or any of its derivatives). Then, to promote depletion of the POI, auxin is added to the culture medium (Figure 1.5). Using this system almost complete depletion of the target can be achieved within 30 minutes of auxin addition.

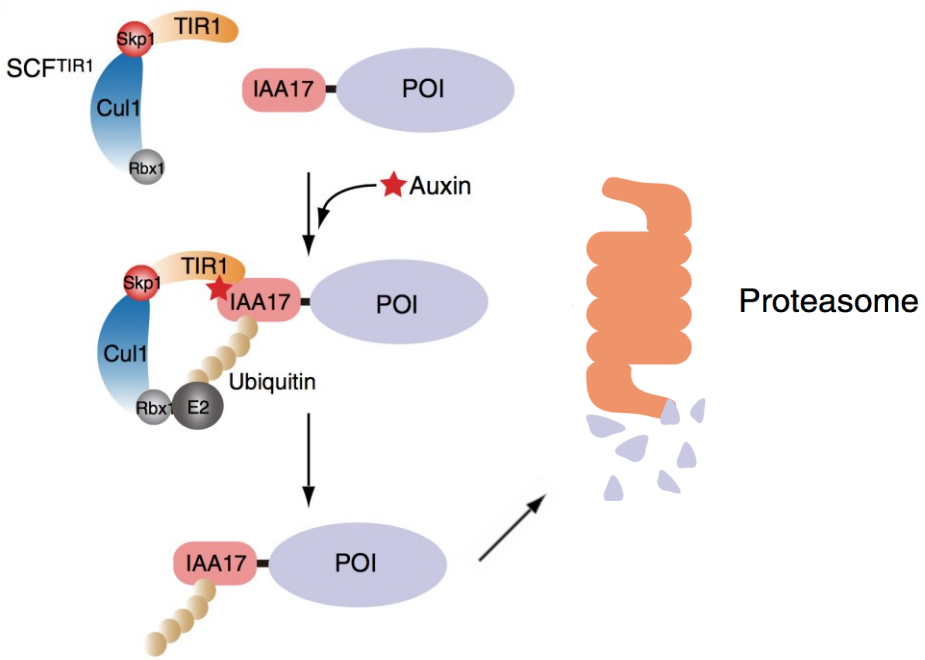


Figure 1.5. Schematic representation of the auxin-inducible system – slight modification of (Kanemaki, 2013). The protein of interest (POI) fused to the AID tag (IAA17) is targeted for degradation in the presence of the auxin plant hormone.

There are several variants of the elements that constitute the AID system. When the yeast AID system was created, Nishimura and others (2009) initially used the *Arabidopsis thaliana* Tir1 protein, but then switched to *Oryza sativa* (rice) Tir1 (*Os*Tir1), as they observed that the latter allowed higher depletion efficiency at higher temperatures (*e.g.* 30°C). OsTir1 is now the popular choice. Also, initially the expression of Tir1 was driven by a GAL promoter. Other Tir1-expressing constructs became available shortly after, for example the ones where Tir1 is constitutively expressed by the *ADH1* promoter.

The first AID tag used was the full length Iaa17 protein (25 kDa), as it is a natural target of the auxin/Tir1 pathway. As smaller tags are preferable to larger tags, the Kanemaki lab and others (Morawska and Ulrich, 2013) independently generated various truncations of Iaa17 and found one that works as well as the full version. These truncations were called miniAID (7.2 kDa) or AID* (4.8 kDa). Lastly, Naphthaleneacetic acid (NAA) is a synthetic

auxin that can also be used with the AID. Both IAA and NAA appear to work with similar efficiency, although Nishimura and others (2009) reported that there are some targets that are depleted more efficiently with one type of auxin than the other.

Anchor-away

The anchor-away technique (Haruki et al., 2008) was developed at about the same time as the AID and is also a popular choice for depleting proteins of interest in budding yeast. The strategy of this technique is to conditionally translocate a nuclear protein to the cytoplasm. The elements of the anchor-away are the human FKBP12 protein (the anchor protein) and its binding partner: the FKBP12-rapamycin-binding (FRB) domain of the human mTOR protein. The anchor-away strain was generated by fusing FKBP12 to the ribosomal protein Rpl13A, and the FRB to the protein of interest (POI).

Addition of rapamycin to the culture, allows association between Rpl13A-FKBP12, POI-FRB and rapamycin. This then leads to the sequestering of POI-FRB to the cytoplasm as it is carried away by newly assembled ribosome particles, as they are being translocated from the nucleus to the cytoplasm. In addition, the authors mutated *TOR1* (*tor1-1*), to avoid rapamycin toxicity, and deleted *FPR1* that is the homolog of human of FKBP12, to avoid binding competition between Fpr1 and FRB for rapamycin and FKBP12.

The main benefits of the anchor-away technique are its fast depletion and its high specificity. Using this system, the authors observed good depletion of all the 43 proteins they tested, within 15 minutes of rapamycin addition. Thanks to the modifications they did to the anchor-away strain, rapamycin does not prevent cell growth. However, some growth inhibition is observed on the

control strain incubated with rapamycin, on the spotting assay. When using this technique, depletion efficiency is examined by changes in growth rate and changes in cellular localization observed by fluorescent microscopy (facilitated by the use of a FRB-GFP tag).

The aims of this study

In the past years, many novel tools to control gene expression have been created. One of those is the auxin-inducible system. This system offers several advantages over previous methods (*e.g.* ts mutants and *GAL* promoters), including high specificity and speed of depletion. **My preliminary objective was** to use the AID system to study several interesting splicing factors including Prp45 and its interaction partners Prp46 and Prp22.

However, I observed that the targeted depletion of these proteins causes unexpected secondary effects on early stages of spliceosome assembly. The hypothesis to explain this was that these secondary effects were not caused by the AID system by itself but were caused by the accumulation of stuck spliceosomes that were preventing the recycling of splicing components – a phenomenon that has not been well described *in vivo*. During the first part of my PhD, **my objective was to test this hypothesis.**

After more than a year of using the AID system, the lab and I observed that AID strains expressing different levels of Tir1 have different depletion efficiencies. The hypothesis was that an optimal expression level of the Tir1 is critical for the efficient depletion of the target protein. **My second objective was** to test this hypothesis while creating a new version of the AID system for yeast with a more efficient depletion, by allowing a better control of the level of Tir1.

During the final stage of my PhD, I used this new version of the AID system as a novel way of swapping expression of wild-type Prp22 for mutants of Prp22, in combination with a genome wide analysis of splicing. **The objective of this approach was to study the role of Prp22 in splicing fidelity**. Although the models that explain splicing fidelity are well supported by *in vitro* evidence, this phenomenon has rarely been studied globally *in vivo*. Therefore, this novel approach offered the opportunity to discover new insights into the process of splicing fidelity.

Chapter 2: Materials and Methods

Sources of reagents

Common chemicals were purchased through Sigma-Aldrich, unless stated otherwise. Speciality reagents, commercial kits, or chemicals particularly important for the methods described in this thesis, are listed in the tables below. % units refer to weight/volume when the chemical was in powder form and volume/volume when the chemical was in liquid form.

Table 2.1. Reagents

Name	Company	Use	Cat. No.
20 mg/mL glycogen	Roche	4tU labelling	10901393001
4tU	Acros Organics	4tU labelling	359930010
Biotin - EZlink HPDP	Thermo Scientific	4tU labelling	21341
magnetic beads - Streptavidin	New England Biolabs	4tU labelling	S1420S
RNase inhibitor: Superasin	Ambion	4tU, RT and RIP	AM2696
IAA	Acros Organics	AID	122160100
2 mg/mL BSA	Sigma-Aldrich	Bradford	1076192
5 x Bradford solution	Bio-rad	Bradford	5000006
magnetic beads - Protein A	Life Technologies	ChIP	10001D
magnetic beads - Protein G	Life Technologies	ChIP	10003D
Phenol/Chloroform(1:1) pH 5.2	Fisher	DNA extraction	BP17531
Monarch DNA Gel	New England Biolabs	gel extraction	T1020S
QIAquick PCR	Qiagen	DNA purification	28106
MinElute PCR	Qiagen	DNA purification	28006
Gibson Assembly Master Mix	New England Biolabs	Gibson Assembly	E2611S
QIAprep Miniprep	Qiagen	Miniprep	27106
Zirconia beads 0.5 mm	Thistle Scientific	multiple	11079105z
B-estradiol	Sigma-Aldrich	multiple	E8875
cOmplete proteinase inhibitor	Roche	multiple	11836145001
DNA Polymerase - Phusion	New Englang Biolabs	PCR	M0535L
dNTP mix 10 mM each	Promega	PCR and RT	U1515
2x SYBR green III master mix	Agilent	qPCR	60088251
Phenol pH 4.3	Sigma-Aldrich	RNA extraction	P46082
Phenol/Chloroform (5:1) pH 4.3	Fisher	RNA extraction	BP17541
Reverse Transcriptase	Roche	RT	3531287001
DNAase I - RQ1	Promega	RT	M6101
10 mg/mL fragmented SSDNA	Roche	transformation	11467140001
E. coli competents: NEB5-a	New England Biolabs	transformation	C2987H
NuPAGE minigels	Invitrogen	Western blot	NP0323BOX
MOPS SDS Running buffer	Invitrogen	Western blot	1862491
Odyssey MWM	Li-cor	Western blot	928-40000
Immobilon-FL PVDF	Millipore	Western blot	IPFL00010

Table 2.2. Antibodies

Antigen	Clone	Organism	Company	Use	Dilution	Cat. No.
FLAG		Rat	Agilent	WB	1:1000	200474
Gapdh		Mouse	Abcam	WB	1:500	Ab9485
Pgk1		Mouse	Abcam	WB	1:5000	Ab113687
HA	12CA5	Mouse	Roche	WB	1:500	11583816001
Myc	9E10	Mouse	Millipore	WB	1:500	05-419
V5		Mouse	Invitrogen	WB	1:2500	MA5-15253
V5		Rabbit	Abcam	WB	1:2500	Ab15828
Rat IgG		Goat	Li-cor	WB	1:5000	925-32219
Mouse IgG		Goat	Li-cor	WB	1:5000	925-68020
Mouse IgG		Goat	Li-cor	WB	1:5000	925-32210
HA	F-7	Mouse	Santa Cruz	ChIP		sc-7392 X
Myc	9E10	Mouse	Santa Cruz	ChIP		sc-40 X
FLAG	M2	Mouse	Sigma-Aldrich	ChIP		F1804

Growth media and buffers

The ingredients for the growth media where purchased from Formedium. Growth media was prepared by the local media services in the WCCB using standard procedures, and stored at room temperature for up to one month. Rich media (YPDA, LB and SOC) were sterilized by autoclave at 121°C for 15 minutes. Minimal media was filtered sterilized through 0.2 uM filter units. Solid media contained 2% bacto-agar. Media used are listed in the table below.

Table 2.3. Media and commonly used buffers

Name	Component	Concentration
YPDA	Yeast extract	1%
	Bacto-peptone	2%
	Glucos	2%
	Adenine Sulfate	0.003%
YMM	Yeast Nitrogen Base w/o AA	0.67%
	Glucose	2%
	Synthetic complete drop-out mixture	according to instructions
LB	Bacto-tryptone	1%
	Yeast extract	0.5%
	NaCl	0.5%
SOC	Bacto-tryptone	2%

	Yeast extract	0.5%
	NaCl	0.06%
	KCl	0.02%
	MgCl2	0.1%
	MgSO4	0.12%
	Glucose	0.4%
PBS	NaCl	140 mM
	KCl	3 mM
	Na2HPO4	8 mM
	H2PO4	1.4 mM
	MgCl2	20 mM
TAE	Tris base	2 M
	Acetic acid	5.71%
	EDTA	50 mM
TE	Tris base	2 M
	Acetic acid	5.71%
TBS	Tris-HCl	10 mM
	NaCl	150 mL
Tris-glycine transfer buffer	Tris base	20 mM
	Glycine	150 mM

Table 2.4. Special buffers

Name	**Component**	**Volume**
DNA extraction	Triton-x100	4 mL
	10% SDS	20 mL
	4 M NaCl	5 mL
	1 M Tris pH 8	2 mL
	0.5 M EDTA pH 8	0.4 mL
	ddH20	to 200 mL
RNA extraction buffer	3 M NaOAc pH 5.3	3.3 mL
	0.5 M EDTA, pH 8	4 mL
	ddH20	to 200 mL
4xLaemmli	1 M Tris, pH 6.8	5 mL
	0.5 M EDTA, pH 8	0.4
	10% SDS	20 mL
	Glycerol	10 mL
	Orange G	a pinch
	ddH20	to 50 mL
4xLaemmli/4MTris buffer	4xLaemmli	21 mL
	1 M Tris-base	9 mL
	2-mercaptoethanol*	14 uL / 1 mL
	ddH20	to 30 mL
IPP150	1 M Tris pH 8	1 mL
	5M NaCl	3 mL
	10% NP-40	1 mL
	1M MgCl2	150 uL
	cOmplete Prot.In.	1 tab/50 mL
	ddH20	to 100 mL
1M NaPi pH 6.8	1M NaH2PO4	50 mL
	1M Na2HPO4	50 mL
	ddH20	to 100 mL
10xNaTMg	1M Tris Cl pH7.0	1 mL
	5M NaCl	4 mL

	1M MgCl2	2.5 mL
	ddH2O	to 10 mL
NaSTPMg	10xNaTMg	5 mL
	1M NaPi pH 6.8	5 mL
	10% SDS	0.5 mL
	ddH2O	to 50 mL
FA1	HEPES KOH 0.5 M pH 7.5	25 mL
	NaCl 4M	8.75 mL
	EDTA 0.5M pH8	0.5 mL
	TRITON	2.5 mL
	Na deoxycholate 10%	2.5 mL
	cOmplete proteinase inhibitor	1 tab/50 mL
	ddH20	to 250 mL
FA2	HEPES KOH 0.5 M pH 7.5	25 mL
	NaCl 4M	31.25 mL
	EDTA 0.5M pH8	0.5 mL
	TRITON	2.5 mL
	Na deoxycholate 10%	2.5 mL
	ddH20	to 250 mL
FA3	Tris-HCl pH8 1M	2.5 mL
	LiCl 4M	15.6 mL
	EDTA 0.5M pH8	0.5 mL
	NP-40	1.25 mL
	Na deoxycholate 10%	12.5 mL
	ddH20	to 250 mL
ChIP elution buffer	5 ml TRIS-HCl pH 7.5 1M	5 mL
	1.75 ml EDTA 0.5mM	1.75 mL
	SDS 10%	5 mL
	ddH20	to 50 mL

Oligonucleotides

Table 2.5. oligonucleotides for tagging and checking insert

Name	Sequence
Rrp44_S3	TTACGTTTTCGATAAGGTCGAAGTTCAAGTTAGGTCGGTGATGGATCCAATTACTAGCAA GCGTAAGGCAGAATTATTGTTAAAAcgtacgctgcaggtcgac
Rrp44_S2	TACATTGACTACTTTACGATGTGTTTTATATATGAGTTATGAATTCCTTTTCGTTTTTATAT CCTGATACTGAAGCATCTTCCATatcgatgaattcgagctcg
Rrp44_CT_F	ccggtagggcaagcatagaa
Rrp44_CT_R	ggataacatgaaagcgcgca
Prp22_S3	GACTAAGCTCAATAAGGCAGTCAAGGGAAAGGGCATTAGGTATCAAGAGGcgtacgctgcagg tcgac
Prp22_S2	ATATAGGTCTATAAAACTCGATAATTATAATGCATAAAAAGCTAACAATGatcgatgaattcgagct cg
S-22 (F)	tggagatgctgaaagccagtctagg
S-22 (R)	tttgtaactaaacgagcaatggcagc
Lea1_F2	CTTCTTTAGAAGAGATTGCCAGGCTGGAAAAACTACTCTCTGGTGGTGTTcggatccccgggtt aattaa
Lea1_S2	TTTATAATTCTTTTTTTTTAAGTCATTGAACAGTCGCACTAACCAAAAGAatcgatgaattcgagct cg
S_Lea1_CT (F)	CGATGGAGATCATGAATTTGG
S_Lea1_CT (R)	GAGCACCATTATTTGTTTTCGTT

YHC1_S3	TCCGACGGCATACGGAAACCGTCGAGTGCCAACGGATATAAAAGGAGGCGGTATGGAAATcgtacgctgcaggtcgac
YHC1_S2	ATTGTTTTTTTTGGCAAAATATAGTCTACATATAATATTTAATAATCTAATCTGATAGCGatcgatgaattcgagctcg
S_YHC1_CT (F)	gcaaaccacgatcaaggagt
S_YHC1_CT (R)	tctcatcataccattttcattcg
Prp45 (S3)	TTCAGTTTACTAAAGCTGAATCCGATGATAAATCGGATAACTATGGCGCCcgtacgctgcaggtcgac
Prp45 (S2)	ATACAACTCAAGCACAAGAATGCTTTGTTTTCCTAGTGCTCATCCTGGGCatcgatgaattcgagctcg
S_Prp45_CT (F)	TTATGACAACCCACTGTTCGTCC
S_Prp45_CT (R)	CATCACACCTCAGCGATAATGAC
Prp46 (S3)	AGTCAGAACCGGGGCTAGCGTGGAACCCCAACTTAAGCGCCAAAAGATTTcgtacgctgcaggtcgac
Prp46 (S2)	TCATTGTATATACACGTATACAGGGTACGTACTTTTTCCATCTACTCCCAatcgatgaattcgagctcg
S_Prp46_CT (F)	GTGAACGGAGTGTTCTTTGTAGCAC
S_Prp46_CT (R)	TAAAGTTACCGAGCGCTATTGCGGC
Syf1_S3	AATCAACCTCTTCATATTCGATTAATCCAGATGAAATAGAACTAGATATTcgtacgctgcaggtcgac
Syf1_S2	AAAAAGAACTTATGGTTTCGAAAATGATGCATGATTTTACATAGCTTATAatcgatgaattcgagctcg
S-Syf1 (F)	ATGCTCCCAACTCGCATACT
S-Syf1 (R)	CACAACCCAAACATCCACAG

Table 2.6. oligonucleotides for RT-qPCR

Name	**Sequence**
ACT_B_F	AGGGGCTTGAAATTTGGAAAAA
ACT_L_F	AGGGGCTTGAAATTTGGAAAAA
ACT_B_R	GCAACAAAAAGAATGAAGCAATCG
ACT_L_R	GCAAGCGCTAGAACATACATAGTACA
ACT_3_F	TTGCTTCATTCTTTTTGTTGCT
ACT_3_R	GCAAAACCGGCTTTACACAT
ACT_m_R	GCAAAACCGGCTTTACACAT
ACT_m_F	TCGAAAATTTACTGAATTAACAATGGA
ACT_E_F	GCTGCTTTGGTTATTGATAACGGTTC
ACT_E_R	GATGGGAAGACAGCACGAGGAG
RPL28_L_F	gagcgcaattatgaaaaagagttacca
RPL28_Lb_R	ttccaaatggaactacatacAtagtaaaacag
RPL28_V_R	ttggttctttcattccctcttcca
RPL28_Vm_F	TCCAGATTCACTAAGACTAGAAAGCACAGA
RPL28_m_R	TGACCACCGGCCATACCTCT
RPL28_3_F	tttttgtacagCCGGTAAAGGTCGT
RPL28_3_R	GATGTTGACCACCGGCCATAC
RPL28_E_F	AGAGGTATGGCCGGTGGTCA
RPL28_E_R	CAGAAATGAGCTTGTTGCTTGTGG
RPL39_5_F	aacacagatagatcaacATGGCTgtatgt
RPL39_5_R	ggtggtaaggtcatttagatggatgtg
RPL39_3_F	cgtatgtgcacgatatgtttcccttt
RPL39_3_R	GTGGCAATGGTCTGTTTTGCTTC
RPL39_m_R	GTGGCAATGGTCTGTTTTGCTTC
RPL39_m_F	agatcaacATGGCTGCTCAAAAGTC
RPL39_E_F	AGCAAAACAGACCATTGCCACA
RPL39_E_R	TGTTCATCTTGGTTCTTCTCCAGTTTC
RPS13_V_F	TCGTATGCACAGTGCCgtatgtt
RPS13_V_R	gtcatctgatttagcgaactattcaatgc

RPS13_3_F	tccaattccactaaatattactttaaacagGGTA
RPS13_3_R	CTTGAACCAAGCTGGAGCATTTCT
RPS13_m_F	TCGTATGCACAGTGCCGGTAA
RPS13_m_R	AGGACAACTTGAACCAAGCTGGAG
RPS13_E_F	CTAGAAATGCTCCAGCTTGGTTCAA
RPS13_E_R	TCAAACCCTTTCTCGCGTACTTG
ALG9_F	TAAGCTGGCATGTGCTGCATTC
ALG9_R	TTTGCATGATTCGGTTGATTGG
U1_ F	TGATTTTTGGGGCCCTTTGTT
U1_ R	TCTTTTGAAAGGCCCCAGCTC
U2_F	TTATGTCCAACGCGGGATTTG
U2_R	CCGCACTAGCACCCCATACC
U4_F	TGAGGATTCGTCCGAGATTGTGT
U4_R	AAGCGAACACCGAATTGACCA
U5s_F	AATGGCGGAGGGAGGTCAAC
U5s_R	CAACACCCGGATGGTTCTGG
U6_F	CGCGAAGTAACCCTTCGTGGA
U6_R	GTTCATCCTTATGCAGGGGAACTG
ECM33_V_F	AGTGCCTCCGCTCTAGCTGgt
ECM33_V_R	cgagatttgtgaggaaagaggcaaa
ECM33_3_F	tctaagcagCTAACTCAACTACTTCTATTCCATC
ECM33_3_R	TTTTGTCCAAATCAGCTTGAGCAGT
ECM33_m_F	GCCTCCGCTCTAGCTGctaactc
ECM33_m_R	ttgagcagtagcagtggcagaagt
ECM33_E_F	CTTCTGCCACTGCTACTGCTCAAG
ECM33_E_R	AGCAGCGGAACCCAAGTCAC
SCR1-F	tctctgtctggtgcggcaag
SCR1-R	tcacgggtcacctttgctga

Oligonucleotides in **bold** where added to the reverse mix

Table 2.7. oligonucleotides for ChIP

Name	**Sequence**
I ACT1 -71 F	TACATCAGCTTTTAGATTTTTCACGCTTACTGCTT
I ACT1 +34 R	GATGGTGCAAGCGCTAGAACATACCAGAAT
II ACT1 +368 F	TGTACTAACATCGATTGCTTCATTCTTTTTGTTGC
II ACT1 +413 R	GACGATAGATGGGAAGACAGCACGAGGA
III ACT1 +561 F	ATCTGGCATCATACCTTCTACAACGA
III ACT1 +653 R	GTTTGATTTAGGGTTCATTGGAGCTT
IV ACT1 +1119 F	TCTGCCGGTATTGACCAAACTACTTA
IV ACT1 +1210 R	CCGGACATAACGATGTTACCGTATAA
V ACT1 +1595 F	ATGTGTTTTGTCTCTCCCTTTTCTACGAAAATTTC
V ACT1 +1734 R	TGATCATATGATACACGGTCCAATGGATAAACAT
I ECM33 -592_F	GCAGTATCATCCTTCACGACCC
I ECM33 -510_R	GCGTCTTTCCCGTTTTTGC
II ECM33 +9_F	CAAGAACGCTTTGACTGCTACTG
II ECM33 +145 _R	GAAGAGGACCACGAATCTACTCG
III ECM33 +430_F	ACTTCTGCCACTGCTACTGCTC
III ECM33 +562_R	AGGAACCATCAATCTCTTGGATAC
IV ECM33 +1073_F	TTGGTCAATCTTTGTCTATCGTCTC
IV ECM33 +1173_R	TGTGTTGTTAGCAATGATGAAACC
V ECM33 +1531_F	TCTAAGAAGTCTAAGGGTGCTGCTC
V ECM33 +1582_R	TGAATGAAGTGGCTGGAACAAG

I RPS13_Pm_F	agtcgtgattgaattaacaatttctttctca
I RPS13_Pm_R	GCACTGTGCATACGACCCATttt
II RPS13_I_F	gctgggtgattccaatttcttttaca
II RPS13_I_R	cataaaggcggctagccatcag
III RPS13_3_F	tccaattccactaaatattactttaaacagGGTA
III RPS13_3_R	CTTGAACCAAGCTGGAGCATTTCT
IV RPS13_Eb_F	TTCACAGATTGGCCAGATACTACAGAAC
IV RPS13_Eb_R	TTGACCAAAGCGGAGGCAGT
V RPS13 +1288_F	GGACCGTTCTCAGAAACATTCCA
V RPS13 +1400_R	CCAACTTGACTTGTCATGCTTGTGT
ALG9_F	TAAGCTGGCATGTGCTGCATTC
ALG9_R	TTTGCATGATTCGGTTGATTGG

Plasmids

To generate plasmids pURA3-AID*-6FLAG (Figure S2.1), pURA3-AID*-9myc and pURA3-AID*-6HA, HPH marker in pHyg-AID plasmids was replaced by *Kluyveromyces lactis* URA3 marker from pVAS-osTIR, using AscI and SacI restriction sites. This special URA3 marker, called URA3-looper, is flanked by a 142 nt repeat sequence from its 5'UTR which allows to pop-out the marker by 5-FOA selection. The series of plasmids containing the AID* tag were kindly provided by the Ulrich lab (Morawska and Ulrich, 2013) through the request of Jane Reid.

The pBest-TIR1-LEU2 centromeric plasmid (Figure 4.6), which contains all elements for the controlled expression of OsTIR1, was constructed by Gibson Assembly (NEB master mix), following the instructions of the manufacturer. Gibson Assembly is a technique that does not require restriction enzyme ligation, where up to five PCR products with complementary ends are cloned into the desired plasmid in a single reaction. pBest-TIR1-LEU2 was generated from the following parts: (1) pRS415 was digested with SpeI/AscI (2) P*Z4EV* from pMN10 (3) Os*TIR1*-V5-T*TEF1* from pVAS-osTIR and (4) P*ACT1*-Z4EV from YMN3 genomic DNA. pMN10 was kindly provided by the McIsaac lab (McIsaac et al., 2013) through the request of Vahid Aslanzadeh.

Plasmids were constructed containing wild-type PRP22 and *prp22* mutants, C-terminally tagged with V5. This was performed by Gibson Assembly using the following parts: (1) MscI/SalI digested p360-PRP22 (Figure S2.2) (2) C-terminal end of *PRP22* from p360-PRP22-URA3, p358-T757A-TRP1 or p358-I764A-TRP1 (2) 3'UTR of *PRP22*, plus V5 tag as part of a primer, from p360-PRP22. Then, these plasmids were further modified to substitute the native promoter for the Z4EV promoter of the B-estradiol expression system. This was performed by Gibson Assembly using the following parts: (1) Eco53KI/ClaI digested p360-PRP22-V5 (2) P$_{Z4EV}$ from pMN10 (3) 5' end of *PRP22* from p360-PRP22-V5. This resulted in plasmids: p360-Z4p-PRP22-V5, p360-Z4p-T757A-V5 and p360-Z4p-I764A-V5 (Figure S5.1). Plasmids p360-PRP22-URA3, p358-T757A-TRP1 and p358-I764A-TRP1 were kindly provided by Beate Schwer (Schneider et al., 2004).

The plasmids were cloned on *E. coli* NEB-5-alpha strains (NEB), and verified by sanger sequencing through the DNA Sequencing and Services of MRC/PPU of the University of Dundee.

Table 2.8. Plasmids

Name	Description	Source
pMK200	To integrate P*ADH1*-689-Os*TIR1* into ura3 locus	YGRC Osaka University
pHyg-AID*-6FLAG	To C-terminally tag a gene of interest	Ulrich lab
pHyg-AID*-9myc	To C-terminally tag a gene of interest	Ulrich lab
pHyg-AID*-6HA	To C-terminally tag a gene of interest	Ulrich lab
pVAS-osTIR1	To integrate P*ADH1*-397-Os*TIR1* into ura3 locus	Vahid Aslanzadeh
pURA3-AID*-6FLAG	URA3-looper version of pHyg-AID*-6FLAG	This study
pURA3-AID*-9myc	URA3-looper version of pHyg-AID*-9myc	This study
pURA3-AID*-6HA	URA3-looper version of pHyg-AID*-6HA	This study
pBest-TIR1-LEU2	For the B-estradiol overexpression of OsTir1 (CEN)	This study
pRS425-FUI1-LEU2	For overexpression of yeast uracil permease FUI1 (CEN)	David Barrass
p360-PRP22-URA3	For expression of wild-type Prp22 under control of its native promoter (CEN)	Beate Schwer

p358-T757A-TRP1	For expression of mutant Prp22 protein T757A under control of its native promoter (CEN)	Beate Schwer
p358-I764A-TRP1	For expression of mutant Prp22 protein I764A under control of its native promoter (CEN)	Beate Schwer
p360-PRP22-V5	C-terminal V5 tagged version of p360-PRP22 (CEN)	This study
p360-T757A-V5	C-terminal V5 tagged version of p358-T757A (CEN)	This study
p360-I764A-V5	C-terminal V5 tagged version of p358-I764A (CEN)	This study
p360-Z4p-PRP22-V5	For the B-estradiol overexpression of wild-type Prp22 (CEN)	This study
p360-Z4p-T757A-V5	For the B-estradiol overexpression of mutant Prp22 T757A (CEN)	This study
p360-Z4p-I764A-V5	For the B-estradiol overexpression of mutant Prp22 I764A (CEN)	This study

Strains

YGM1 contains the P*ADH1-689*-Os*TIR1* cassette and was created by integrating a StuI-linearized pMK200 into the *ura3-1* locus of W303. YBRT1 strain contains the P*ADH1-396*-Os*TIR1* cassette and was kindly provided by Barbara Terlow (this lab). YZTR41 strain was created by integrating a KanMX-P*Z4EV* PCR amplicon using pMN10 as template, directly upstream of *APE2* start site, followed by *URA3* pop-in/pop-out substitution of *APE2* coding region for Os*TIR1*-V5. AID-tagged strains were created by transforming YGM1 or YZTR41 with PCR products of the pHyg-AID*- or pURA3-AID*- cassettes, respectively. *UPF1* gene was deleted on some genes using the URA3-looper marker. pMK200 plasmid was created by the Kanemaki lab and obtained through the Yeast Genetic Resource Centre, Osaka University. pYMN3 plasmid was kindly provided by the MIsaac lab through the request of Vahid Aslanzadeh.

Table 2.9. Strains

Name	Genotype	Source
W303	MATα *ade2-1 ura3-1 his3-11,15 trp1-1 leu2-3,112 can1-100*	
YGM1	MATα *ade2-1 his3-11,15 trp1-1 leu2-3,112 can1-100 ura3-1*::*URA3*-P*ADH1*-689-Os*TIR1*	This study
YBRT	MATα *ade2-1 ura3-1 trp1-1 leu2-3,112 can1-100 his3-11,15:*:P*ADH1*-397-Os*TIR1*	B. Terlow
YMN3	MATα *ura3*Δ *leu2*Δ0::P*ACT1*-Z4EV-NatMX	S. McIsaac
YZTR41	MATα *ura3*Δ *leu2*Δ0::P*ACT1*-Z4EV-NatMX *ape2*::PZ4EV-Os*TIR1*-V5	This study
Prp22-AID*(YGM1)	MATα *ade2-1 his3-11,15 trp1-1 leu2-3,112 can1-100 ura3-1*::*URA3*-P*ADH1*-689-Os*TIR1 PRP22::PRP22*-AID*-6FLAG-HygMX	This study
Prp45-AID*(YGM1)	MATα *ade2-1 his3-11,15 trp1-1 leu2-3,112 can1-100 ura3-1*::*URA3*-P*ADH1*-689-Os*TIR1 PRP45::PRP45*-AID*-6FLAG-HygMX	This study
Prp46-AID*(YGM1)	MATα *ade2-1 his3-11,15 trp1-1 leu2-3,112 can1-100 ura3-1*::*URA3*-P*ADH1*-689-Os*TIR1 PRP46::PRP46*-AID*-6FLAG-HygMX	This study
Syf1-AID*(YGM1)	MATα *ade2-1 his3-11,15 trp1-1 leu2-3,112 can1-100 ura3-1*::*URA3*-P*ADH1*-689-Os*TIR1 SYF1::SYF1*-AID*-6FLAG-HygMX	This study
Prp16-AID*(YGM1)	MATα *ade2-1 his3-11,15 trp1-1 leu2-3,112 can1-100 ura3-1*::*URA3*-P*ADH1*-689-Os*TIR1 PRP16::PRP16*-AID*-6FLAG-HygMX	B. Terlow
Prp4-AID*(YBRT)	MATα *ade2-1 ura3-1 trp1-1 leu2-3,112 can1-100 his3-11,15:*:P*ADH1*-397-Os*TIR1 PRP4::PRP4-AID*-6FLAG-HygMX*	I. Maudlin
Prp22-AID*(YBRT)	MATα *ade2-1 ura3-1 trp1-1 leu2-3,112 can1-100 his3-11,15:*:P*ADH1*-397-Os*TIR1 PRP22::PRP22-AID*-6FLAG-HygMX*	E. Sani
Yhc1-AID*(YZTR41)	MATα *ura3*Δ *leu2*Δ0::P*ACT1*-Z4EV-NatMX *ape2*::PZ4EV-Os*TIR1*-V5 *YHC1::YHC1*-AID*-6FLAG	This study
Prp22-AID*(YZTR41)	MATα *ura3*Δ *leu2*Δ0::P*ACT1*-Z4EV-NatMX *ape2*::PZ4EV-Os*TIR1*-V5 *PRP22::PRP22*-AID*-6FLAG	This study
Rrp44-AID*(YZTR41)	MATα *ura3*Δ *leu2*Δ0::P*ACT1*-Z4EV-NatMX *ape2*::PZ4EV-Os*TIR1*-V5 *RRP44::RRP44*-AID*-6FLAG	This study
YZTR41-du	MATα *ura3*Δ *leu2*Δ0::P*ACT1*-Z4EV-NatMX *ape2*::PZ4EV-Os*TIR1*-V5 upf1Δ	This study
Prp22-AID*(YZTR41)-du	MATα *ura3*Δ *leu2*Δ0::P*ACT1*-Z4EV-NatMX *ape2*::PZ4EV-Os*TIR1*-V5 *PRP22::PRP22*-AID*-6FLAG upf1Δ	This study
Prp22-AID*-du-Z22CEN	MATα *ura3*Δ *leu2*Δ0::P*ACT1*-Z4EV-NatMX *ape2*::PZ4EV-Os*TIR1*-V5 *PRP22::PRP22*-AID*-6FLAG upf1Δ p360-Z4p-PRP22-V5	This study
Prp22-AID*-du-Z57CEN	MATα *ura3*Δ *leu2*Δ0::P*ACT1*-Z4EV-NatMX *ape2*::PZ4EV-Os*TIR1*-V5 *PRP22::PRP22*-AID*-6FLAG upf1Δ p360-Z4p-T757A-V5	This study
Prp22-AID*-du-Z64CEN	MATα *ura3*Δ *leu2*Δ0::P*ACT1*-Z4EV-NatMX *ape2*::PZ4EV-Os*TIR1*-V5 *PRP22::PRP22*-AID*-6FLAG upf1Δ p360-Z4p-I764A-V5	This study

General manipulation and preservation of yeast cultures

Yeast liquid cultures were grown in flasks at 30°C and shaken at 180 rpm. Except during transformation, only yeast that contain a plasmid were grown on the appropriate selective media, otherwise they were grown on YPDA without antibiotics. Log-phase cultures were prepared by diluting an overnight culture to OD_{600} of 0.2, and then allowing it grow for at least two doublings. Cells were harvested by centrifugation at 3,000 x g for 2 minutes in a swinging-bucket centrifuge, or in a fixed-rotor centrifuge for cultures smaller than 2 mL, unless otherwise stated. Isolated colonies of yeast strains were kept at 4°C in agar plates for up to two weeks. For long-term storage, a saturated culture was diluted 1:1 with 50% glycerol and stored at - 80°C. All cultures were manipulated using standard aseptic technique.

Yeast transformation

Yeast were transformed using the TRAFO LiAc/PEG method (Gietz and Schiestl, 2007). To prepare the competent cells, 50 mL of log-phase culture at OD_{600} of 1, was washed twice with 10 mL of ddH_2O, and then once with 1 mL of 0.1 M LiAc. The pellet was resuspended in 100 uL of 0.1 M LiAc, and placed left on ice. An aliquot of 10 mg/mL SSDNA carrier was boiled for 5 minutes and placed on ice. The transformation mix was freshly prepared by thoroughly mixing 240 uL of PEG 4000 50%, 36 uL of 1 M LiAc, 20 uL of freshly boiled SSDNA carrier and 25 uL of transformation DNA (20 ug of purified PCR product or 100 ng of plasmid)

To transform, 40 uL of the competent cells were added to the transformation mix, and thoroughly vortexed for approximately 1 minute. The tube was incubated at 30°C for 30 minutes, and then at 42°C for 40 minutes in a water bath. The tube was spun at 3,000 x g for 5 seconds, and the supernatant discarded using a micropipette. Then, the pellet was resuspended in 300 uL of

PBS, spread on YPDA agar using glass beads and incubated at 30°C overnight. The next day, the cells were replica plated on selective media and incubated for 2-4 days until colonies were visible. Then, to colony purify, single colonies were re streaked on selective media. Positive transformants where verified by PCR and, in the case of gene tagging, by western blotting.

PCR and agarose gel electrophoresis

End-point PCR was performed using DNA polymerase Phusion HF Hot-Start, following the instructions of the manufacturer. To improve PCR efficiency, DMSO was routinely added to the PCR mix at a final concentration of 2.5%. The thermocycler program used is stated in the table below. PCR products were separated on 1.5-2% agarose TAE gels and SYBR safe, and the bands visualized under UV light using a Syngene gel doc system.

End-point PCR program

Step 1:	98°C	00:30
Step 2:	98°C	00:10
Step 3:	60°C	00:20
Step 4:	72°C	1 min/kb
Step 5:	repeat steps 2-4 x 35	
Step 6:	72°C	10:00

DNA extraction

DNA template for PCR was extracted from yeast cells with phenol/chloroform. Using a 1 mL micro pipette tip, a patch of cells was taken from an agar plate, and resuspended in a microfuge tube containing 300 uL of DNA extraction buffer, 300 uL of phenol/chloroform (1:1) pH 5.2, and 300 uL of zirconia beads. The tube was shaken for 2 minutes in a Mini-Beadbeater-24 (BioSpec Products) at 2,000 rpm, and then spun at maximum speed for 5

minutes. 200 uL of the aqueous phase was take and placed in a new tube containing 200 uL of isopropanol and 20 uL of NaAc buffer. The contents were mixed by inverting the tube three times and then the tube was placed at -80°C for 5 minutes. Then, tube was spun at maximum speed for 5 minutes and the supernatant was carefully discarded by aspiration. The pellet was washed once with 70% ethanol and, finally, the pellet was resuspended in 100 uL of ddH_2O. DNA concentration and purity was analysed using the Nanodrop.

Growth analysis

Growth rate was analysed through a spotting assay or through a growth curve analysis. For a growth curve analysis, log phase cultures were re diluted to OD_{600} of 0.1 and then OD_{600} measurements were taken every following hour for up to 6 hours, using 1-mL cuvettes and a standard spectrophotometer. For a spotting assay, log phase cultures were diluted to OD_{600} of 0.3 and then, three serial dilutions of 10-fold were made on a flat-bottom 96-well plate, with a total volume of 250 uL per well. These serially diluted cultures were then spotted on agar plates using a hedgehog. YPDA control plates were included at the start and end of the spotting procedure. Plates were incubated at 30°C for 1-2 days before taking images using a Syngene gel doc system.

Auxin-inducible depletion

AID depletion was performed though the following procedure. A 2.5 mL solution stocks of 1.5 M IAA was prepared by dissolving 440 mg of IAA into 2 mL of DMSO, and stored at -20°C for up to 2 months. Then, no more than one hour before the log phase culture is ready, the IAA stock solution was pre-diluted 1:500 in 25% ethanol. Finally, to deplete the target, the pre-diluted IAA stock is added at 1:500 to the log phase culture. The pre-dilution step is not

strictly necessary but, by doing this, the IAA is dissolved more efficiently in the medium. Auxin can be added to a final concentration of up to 0.75 mM

B-estradiol induction

For the B-estradiol system, a stock solution of 10 mM B-estradiol (in ethanol) was prepared and stored at -20°C for up to 2 months. To induce expression, B-estradiol is added to a log phase culture at a final concentration of 1-10 uM. Just as with the IAA stock, a pre-dilution of the B-estradiol stock on 25% ethanol was sometimes done, to facilitate dispensing it into a small culture.

Western blotting

Protein extraction

First, total proteins were extracted following the NaOH/TCA precipitation method with a few modifications. 5 mL of a log phase culture were taken and added to a 15-mL falcon tube containing 5 mL of methanol chilled on dry ice. The contents were mixed by inverting the tube twice. Cells were then harvested by centrifugation, washed with 1 mL of cold PBS and placed in a 1.5 mL safelock tube. The pellet was resuspended in 540 mL of 0.2 M NaOH and incubated on ice for 10 minutes. 30 uL of 100% TCA was added to the pellet. The tube was vortexed briefly and incubated on ice for 5 minutes. The cells were spun at 3,000 x g for 30 seconds, the supernatant was eliminated by aspiration, and the tube was spun again and aspirated to eliminate the dregs. 60 uL of 4xLaemmli/4MTris was added to the pellet, and the tube was placed on a thermomixer at 1400 rpm and 65°C for 5 minutes to resuspend the pellet. Then, the samples were boiled for 5 minutes and then spun at maximum speed for 5 minutes. The supernatant was placed in a new tube and stored at -20°C and the pellet was discarded.

Bradford quantification

Total proteins were quantified, using the Bradford Bio-rad reagent in 1 mL cuvettes, following the instructions of the manufacturer. In order to keep the SDS concentration below the threshold of compatibility, the Laemmli protein samples were first diluted 1:20 in water, and then 6 uL of this was mixed with 1 mL of 1 x Bradford solution into a cuvette. A blank solution was prepared by diluting 4xLaemmli/4MTris 1:20 in water. The standards were generated by mixing 1, 2, 3, 4, 5 and 6 uL of 1 mg/mL of BSA, with 6 uL of blank solution and 1 mL of 1 x Bradford solution, into a cuvette. The contents of the cuvettes were mixed with a micropipette and then measured at 595 nM with a spectrophotometer. Using the OD readings, a standard curve was generated in Excel and the protein concentration was estimated for each sample using the standard curve.

SDS-PAGE and electro-transfer

SDS-PAG was performed using the Invitrogen NuPAGE electrophoresis system, with mini Bis-Tris gels of 10-15 wells and 4-15% gradient acrylamide. Each well was loaded with 10-30 ug of protein extract. 2 uL of Odyssey molecular weight marker was added to one well. The proteins were separated by running the gel at 120 V until the dye reached the bottom of the gel. Then, the gel was taken and assembled into the Mini Trans-Blot Cell (BioRad), following the instructions of the manual. Protein was transferred to a low-fluorescent PVDF membrane (Millipore) using Tris-Glycine transfer buffer, at 100 V for 1 hour.

Immunodetection

The transferred proteins were immunodetected by fluorescence using the Li-cor system through the following procedure. The membrane containing the transferred proteins was (1) incubated for 30 minutes in 5% skimmed milk PBS on gentle shaking (2) incubated for 1 hour with the primary antibody in 5% skimmed milk PBS and 0.2% Tween-20 with gentle shaking (3) washed for 10 minutes with 0.1% Tween-20 PBS three times (4) incubated for 1 hour with the secondary antibody in 5% skimmed milk PBS 0.2% Tween-20 and 0.01% SDS with gentle shaking and (5) washed for 10 minutes with 0.1% Tween-20 PBS three times. Finally, the membrane was scanned in the Li-cor Odyssey scanner and the image was analysed using the Odyssey Image Studio software.

RNA extraction

RNA was extracted from yeast following a modified method of phenol/chloroform extraction. First, 5 mL of a log phase culture were added to a 15-mL falcon tube containing 5 mL of methanol chilled on dry ice, and mixed. Cell pellet was harvested by centrifugation, washed once with 1 mL of cold PBS and placed in a 2-mL screw cap tube. 300 uL of zirconia beads, 700 uL of RNA extraction buffer, and 700 uL of phenol pH 4.3 were added to the cell pellet. To lyse the cells, the tube was shaken for 2 minutes at 2,000 rpm in a Mini-Beadbeater-24 (BioSpec Products), placed on ice for 2 minutes, and this cycle repeated twice. The tube was placed on dry-ice for 5 minutes and then spun at maximum speed for 8 minutes. The aqueous phase was taken and put into a new tube containing 800 uL of phenol/chloroform (5:1) pH 4.3, vigorously vortexed for 30 seconds, and spun for 3 minutes at maximum speed. The aqueous phase was taken, put into a new tube containing 700 uL of chloroform, vigorously vortexed for 5 seconds and spun for 2 minutes at

maximum speed. To precipitate, the aqueous phase was taken and put into a new tube containing 300 uL of 10 M LiCl, mixed by inversion, and placed at -20°C for 2 hours. The pellet was recovered by spinning at maximum speed for 5 minutes at 4°C, and washed by adding 1 mL of 70% ethanol and placing in a rotating wheel for 10 minutes. 15 uL of ddH20 were added to the pellet, and incubated at 65°C for no more than 5 minutes, until the pellet is dissolved. The dissolved RNA was analysed by Nanodrop to determine concentration and purity, and then stored at -80°C.

Reverse transcription

To obtain cDNA, RNA extracted as described above were reverse transcribed through the following procedure. RNA concentration was equalized across the multiple samples. First, the reverse primer mix was prepared by mixing each desired reverse primer to a final concentration of 3 uM per primer. The RT mix was prepared by mixing 2 uL of 5 x RT buffer, 0.75 uL of 10 mM dNTP mix, 0.1 RNase inhibitor, 1.75 uL of ddH_2O and 0.4 uL of reverse transcriptase. A No-RT mix was prepared in the same way but without enzyme. DNase I master mix was prepared by mixing 1 uL of DNase I 10 x buffer, 0.9 uL of DNase I, 0.1 uL of RNase inhibitor and 0.5 uL ddH_2O. Then, for each sample, 8 uL of RNA extract with at least 700 ng/uL of RNA and 2.5 uL of DNase I master mix were added to the well of a 96 well plates for PCR. The plate was mixed and placed in a thermocycler at 37°C for 20 minutes, and then at 75°C for 10 minutes. 2.5 uL of the reverse primer mix was added to each well and mixed. The plate was heated at 72°C for 3 minutes and then immediately placed on ice. The RNA samples were split in two wells and 5 uL of RT mix, or 5 uL No-RT mix, was added to each well. The plate was spun, mixed and heated to 55°C for one hour. Finally, 190 uL of ddH_2O was added to each well to make up to 200 uL and the plate was stored at -20°C.

Quantitative PCR

DNA samples retrieved by Reverse Transcription, Chromatin Immunoprecipitation or RNA Immunoprecipiation were analysed by quantitative PCR (qPCR) using the SYBR green method and the 384-well Lightcycler 480 II (Roche) equipment. Each qPCR reaction contained 2 uL of 2xSYBR Green III Master Mix (Agilent), 0.25 uL of 3 uM forward primer, 0.25 uL of 3 uM reverse primer, and 2.5 uL of DNA samples. The qPCR program used is listed in the table below.

qPCR program

Step 1:	94°C	02:00
Step 2:	94°C	00:10
Step 3:	60°C	00:10
Step 4:	72°C	00:15
Step 5:	repeat steps 2-4 x 35	
Step 6:	95°C	00:10
Step 7:	60°C	00:10
Step 8:	ramp to 94°C	00:30

To analyse the data, threshold cycle (Ct) values were retrieve and the relative abundance of the amplicon was calculated by dividing 2^{-Ct} of the query amplicon against 2^{-Ct} of the control amplicon. Unless otherwise stated, the control amplicon was for *ALG9* in the case of RT-qPCR analysis, for *U2* in the case of RIP analysis, or the amplicon of the lowest value within each gene in the case of ChIP analysis. For the ChIP analysis, the relative abundance of the amplicon in the immunoprecipitate was divided against that of the 10% input control.

Chromatin immunoprecipitation

Chromatin immunoprecipitation was performed through the following procedure, optimized by Ross Alexander and Ema Sani, with minor modifications.

Cross-linking and cell disruption

50 mL of log phase culture were placed in a 50 mL falcon tube containing 1.37 mL of 37% formaldehyde, and shaken gently for 10 minutes. 2.5 mL of 2.5 M glycine was added to the tube and shaken gently for 5 minutes. The cells were harvested by centrifugation. The cell pellet was washed with ice-cold PBS, once with 50 mL and then again with 1 mL, and transferred to a 2-mL screw cap tube. 300 uL of zirconia beads and 350 uL of FA1 buffer where added to the tube. The cells were disrupted by shaking three times at 2,000 rpm in a Mini-Beadbeater-24 (BioSpec Products) for 2 minutes, with 2 minutes on ice in between. To separate the lysate from the zirconia beads, the tube was perforated in the bottom with a flame-heated needle, and placed inside a 5-mL syringe (without its plunger) that is inside a 15-mL falcon tube. This assembled device was spun for 1 minute at 1000 x g and the lysate was transferred to a new 1.5-mL tube.

Chromatin extraction and Immunoprecipitation

The lysate was spun at maximum speed for 15 minutes at 4°C. The supernatant was discarded. The pellet was resuspended in 300 uL of FA1 buffer, transferred to a Diogenode tube, and sonicated for 10 cyles at 30 seconds ON and 30 seconds on at 4°C. The tube was spun for 5 minutes at maximum speed at 4°C, and the supernatant was placed in a new tube and quantified by Bradford (as described above). To immunoprecipitate, 20 uL of

dynabeads (coupled to protein A or G) and 5 ug of antibody was mixed with 500 ug of chromatin extract and incubated overnight at 4°C in a rotating wheel. Then, the beads were washed three times with FA1, three times with FA2, three times with FA3 and once with 0.05% Tween-20/TBS buffer. The washed beads were transferred to a new tube. To elute the DNA and reverse cross-link, 150 uL of ChIP elution buffer and 75 ug of Proteinase K were added to the washed beads and to 50 ug of chromatin extract as the 10% input control. The tubes were incubated for 2 hours at 42°C, and 4 hours at 65°C, while shaking at 1400 rpm. The supernatant was recovered and its DNA was purified using the Qiagen MinElute PCR kit. The purified DNA was taken to a final volume of 400 uL with ddH_2O, then analysed by qPCR.

RNA immunoprecipitation

RNA immunoprecipitation was performed as described previously (Tardiff et al., 2006), with a few modifications. All steps during and after cell harvesting were performed at 4 °C. 50 mL of log phase culture was harvested by centrifugation and the pellet was washed once with 10 mL TBS, once with 1 mL IPP150 buffer, and transferred to a 2-mL screw cap tube. 1.5 mL of IPP150, 40 units of RNaseIN and 500 uL of zirconia beads were added to the tube. The cells were disrupted by shaking three times at 2,000 rpm in a Mini-Beadbeater-24 (BioSpec Products) beads beater for 1 minutes, with 1 minutes on ice in between. The lysate was separated from the zirconia beads as described above (ChIP protocol), spun at maximum speed for 5 minutes, and the supernatant was transferred to a new tube. For immunoprecipitation, 20 uL of dynabeads (coupled to protein A or G) and 5 ug of antibody was mixed with the soluble lysate and incubated for 2 hours in a rotating wheel. The beads were then washed five times with 1 mL of IPP150 (+ 40U/mL of RNaseIN) on a rotating wheel for 5 minutes each, and then placed on a new tube. RNA was extracted

and purified following the phenol/chloroform method described above, by adding phenol and RNA extraction buffer directly to the washed beads. The purified RNA was taken to a final volume of 400 uL with ddH_2O and analysed by qPCR.

4-thio-uracil labelling of newly synthesized RNA

Strains analysed by this method (4-thio-uracil (4tU) labelling newly synthesized RNA (nsRNA)), were previously transformed with pFui-LEU plasmid for overexpression of the uracil permease under its own promoter in a 2 uM plasmid. After RNA extraction, all following steps were carried out using RNase free reagents and materials. The procedure was developed and optimized by David Barrass and is described below.

Thiolation

A log phase culture was grown in –ura –leu (Keiser) drop out YMM. 4 thio uracil was added to the culture to at a final concentration of 10 uM, mixed vigorously and left for 1 minute in normal growth conditions. 30 mL of this culture were added to a 50-mL falcon containing 20 mL of methanol previously chilled on dry-ice. The cell pellet was harvested by centrifugation, washed with 1 mL of cold PBS and transferred to a 2-mL screw cap tube. RNA was extracted following the procedure described above, and the purified RNA pellet was resuspended in 90 uL of TE pH 7 buffer.

Biotinylation

The resuspended RNA was transferred to a 8-strip PCR tube and heated to 65°C for 15 seconds. 10 uL of HPDP biotin solution (4mM in dimethyl formamide) was added, followed by incubation at 65°C for 15 minutes. The

biotinylated RNA was purified using 0.5ml Zeba column, using TE buffer and following the manufacturer's instructions. To precipitate, 1/10 volume of NaOAc buffer and 2.5 volumes of ethanol was added, mixed, placed at -20°C for 30 minutes, and spun for 5 minutes at maximum speed at 4°C. The supernatant was discarded and the pellet was washed with 80% ethanol, once on a rotating wheel for 1 hour and a second time overnight at 4°C to wash out any remaining unincorporated biotin.

Purification of snRNA

The washed pellet of biotinylated RNA was dissolved in 242 uL of ddH_2O at 65°C, and 5 uL of dissolved RNA were stored separately for concentration determination by nanodrop and later analysis of total RNA. The RNA concentration was equalized across the multiple samples. 30 uL of 10xNaTMg buffer, 30 uL of 1 M NaPi buffer and 3 uL of 10% SDS were added to the RNA, and mixed. To prepare the magnetic streptavidin beads, 50 uL of the slurry were added to a low retention 1.5-mL tube and placed on a magnetic rack; washed with 200 uL of NaSTPMg buffer; blocked with 200 uL of NaSTPMg buffer, 2.5 uL of 5 mg/mL tRNA and 10 uL of 20 mg/mL glycogen for 20 minutes in a rotating wheel; and washed again with 200 uL of NaSTPMg. The prepared beads were incubated with the RNA samples in a rotating wheel for 30 minutes, washed four times with 200 uL of NaSTPMg, and eluted twice with 50 uL of freshly prepared 0.7 M b-mercaptoethanol. To precipitate the RNA, 10 uL of 3 M NaOAc pH 5.3 and 2.5 uL of 20 mg/mL glycogen were added to the 100 uL elute, mixed and incubated for 1 hour at -20°C. The pellet was spun for 5 minutes at maximum speed, washed with 70% ethanol, and resuspended in 10 uL of TE at room temperature. The quality of the RNA was assessed using the Bioanalyzer, following the instructions of the manufacturer.

RNA next-generation sequencing

Samples, library preparation and sequencing

Total RNA was extracted and purified as described above. A total of eight samples were processed, including the wild-type strain, Prp22 depletion, and two mutants of Prp22, in biological duplicates. The quality of the RNA was assessed by Agilent Bioanalyzer 2100 RNA Nano-Chip, following the instructions of the manufacturer.

RNA samples were shipped in dry-ice to BGI Genomics (Hong Kong) for library preparation and strand specific 150 bp paired-end deep sequencing by HiSeq Illumina 4000. Total RNA was treated with DNase I and then with Ribo-Zero Gold rRNA Removal kit for yeast. Retrieved RNA was randomly fragmented. First-strand cDNA was synthesized with random primers using Invitrogen SuperScript II First-strand Synthesis SuperMix. The product was purified with Agencourt RNA Clean XP Beads, followed by synthesis of the second strand. The ends of the cDNA library were repaired and then index adapters were ligated. To enrich the cDNA fragments, several rounds of PCR amplification were performed. The PCR products are purified with Ampure XP Beads. Then, the eight libraries were loaded in two wells (1 biological replicate per well) and Illumina sequenced.

Genome mapping

After sequencing, the raw reads were filtered by BGI Genomics. Data filtering included removal of adapter sequences, contamination and low-quality reads from raw reads. The clean reads in FASTQ format were shipped to WTCCB in an external hard drive. A report was generated and sent that includes the quality control analysis (QC) and read statistics results (table below).

Table 2.10. Read statistics results

No.	Sample Name	Read length(nt)	Clean Reads	Clean bases	Q20(%)	GC(%)
1	57-1	150	210,586,726	31,588,008,900	99.06;97.14	42.3
2	57-2	150	242,973,566	36,446,034,900	99.02;97.03	42.33
3	64-1	150	196,993,362	29,549,004,300	98.44;96.74	42.26
4	64-2	150	203,629,174	30,544,376,100	99.06;96.76	42.4
5	WT-1	150	210,963,050	31,644,457,500	98.99;96.89	42.25
6	WT-2	150	192,272,846	28,840,926,900	99.04;96.73	42.5
7	d-1	150	204,400,598	30,660,089,700	98.92;96.93	42.4
8	d-2	150	169,650,916	25,447,637,400	98.30;97.16	42.59

Further processing and analysis of the reads was performed through the Linux-based server (bifx-cli) of the Bioinformatics Core Facility at the WTCCB (University of Edinburgh) under the username of s1325554, unless otherwise stated. To confirm the quality of the reads, a QC analysis was done using the FastQC tool (Andrews, 2010).

Genome mapping was performed by aligning the clean reads to the genome sequence of *Saccharomyces cerevisiae*, using the STAR mapper program (Dobin and Gingeras, 2016). First, a genome index was generated using the yeast reference genome version sacCer3 and the corresponding GTF file (version R64-1-1.88) downloaded from ENSEMBL. Then, the clean reads were mapped to the genome using the genome index and the spliced-junction output files (SJ.out.tab) were further analysed.

Code to generate a yeast genome index for STAR mapping:

```
STAR --runThreadN 60 --runMode genomeGenerate –-genomeDir /path-to-genome-directory/ --genomeFastaFiles /path-to-genome-Fasta-Files/Saccharomyces_cerevisiae.R64-1-1.dna.toplevel.fa --sjdbGTFfile /path-to-genome-GTF-File/Saccharomyces_cerevisiae.R64-1-1.88.gtf --sjdbOverhang 50
```

Code to map reads to reference genome:

```
STAR --runThreadN 60 --genomeDir /path-to-genome-directory/ --readFilesIn /path-to-reads-first-strand/cleanreads_1.fq/ path-to-reads-second-strand/cleanreads_2.fq
```

Analysis of mRNA junctions

First, common gene names and 5'ss-3'ss motifs were annotated to the splice junction table (SJ.out.tab) through a program created by Sander Granneman called "STARtab2GTF.py". Then, non-annotated splice junctions were located within this table and categorized based on the type and direction of the splice site shift (*e.g.* upstream 5'ss, downstream 5'ss, upstream 3'ss… etc), using a program created by Vahid Aslanzadeh called "STARtab2GTF2Juncanot.py". A non-annotated splice junction is defined as a junction within an intron-containing gene that is not annotated in the GTF file. Data in this table was then used to measure relative mRNA abundance and the frequency of splicing errors, using a python-based notebook created by Vahid Aslanzadeh called "Vahid-altevents.ipynb" as a basis. This python notebook was modified and adapted according to the characteristics of my data. For each sample, number of reads was normalized relative to the total number of uniquely mapped reads. The equation drawn below was used to calculate the Splicing Error Frequency (SEF) score, which represent non-annotated junction reads proportional to annotated junction reads, within each gene. Prior to SEF score calculation, spliced junction events that do not contain at least 1 read in each of two replicates from one experiment, were filtered out; this means that not every non-annotated spliced junction event contains SEF score, or SEF ratio values, on all four strains.

$$SEF = \frac{\text{number of reads that align to the non-annotated splice junction}}{\text{number of reads that align to the annotated splice junction}}$$

Fisher's exact test was used to estimate statistical significance between the SEF values of mutants compared to wild type. For this test, the average counts (reads) from biological replicates, of both annotated and non-annotated, were used to calculate the p-value of mutant compared to wild-type (table below) and test whether the ratios between strains are imbalanced.

Example of 2x2 contingency table for Fisher's exact test (giving p-value < 0.05)

*LSM7*_down3ss_7nt	**Prp22 depletion**	**Wild-type**
Non-annotated	(37+42)/2	(13+15)/2
Annotated	(148+289)/2	(562+577)/2

Cells containing average reads of two biological replicates (real sequencing values)

Data visualization was performed using Seaborn and Matplotlib visualization libraries through python programming language.

Chapter 3: Studying the spliceosome cycle with the AID system

Abstract

Spliceosome assembly and splicing catalysis have been extensively studied, yet there are still many gaps in our understanding of the final steps of the spliceosome cycle: disassembly and recycling. One aspect of this that is not well understood is what happens to the cycle when assembly goes wrong? Specifically, how does the cell deal with aberrant spliceosomes that lack essential protein components? To answer this question, with an *in vivo* approach I rapidly knocked down several splicing factors and measured changes in abundance of intermediate complexes of the spliceosomes and in co-transcriptional assembly of the spliceosome. I show that in the absence of Prp4, Prp16 or Prp22, intermediate complexes of the spliceosome accumulate and co-transcriptional formation of the pre-spliceosome is reduced (Figure 3.1). This suggests that, at least in some cases, aberrant spliceosomes that lack essential splicing factors are not targeted for disassembly and recycled, contrary to what was previously proposed. Furthermore, a kinetic analysis of Prp22 depletion suggests that spliceosome recycling is possibly a rate-limiting step for splicing and critical in budding yeast, as a recycling defect has an immediate negative impact on new rounds of splicing. The work presented here warns about potential systematic secondary effects when disturbing single components of spliceosomes.

Introduction

Splicing has been extensively studied, in most cases with biochemical and genetic approaches. As a result, we now have a good mechanistic understanding of this process, with many molecular interactions confirmed and extended through high resolution cryo-EM modelling (Fica et al., 2017; Galej et al., 2016; Hang et al., 2015; Nguyen et al., 2015; Wan et al., 2016b; Yan et al., 2015). Nevertheless, some gaps remain. For example, few reports have analysed spliceosome assembly *in vivo,* and even fewer have analysed disassembly and recycling *in vivo*.

In normal conditions, after splicing completion spliced mRNA is released from the spliceosome by the RNA-helicase Prp22 (Schwer, 2008; Wagner et al., 1998). Then, the post-spliceosome complex is disassembled by Prp43, Ntr1 and Ntr2 (Arenas and Abelson, 1997; Martin et al., 2002; Tsai et al., 2005), which allows the free snRNPs to be recycled into new rounds of splicing. One aspect that is not well understood is the link between quality control and spliceosome disassembly. Are aberrant spliceosomes detected and disassembled, perhaps to prevent errors in splicing?

In vitro-based studies suggest that substrates with suboptimal splice sites are discarded from the spliceosome by RNA helicases Prp16, Prp22 and/or Prp43, reviewed by (Koodathingal and Staley, 2013). Thus, it was proposed that these proteins have the additional role of proofreading splicing. This model aims to explain how introns are selected and what happens when the spliceosome encounters aberrant pre-mRNAs. However, it does not explain what happens to spliceosomes with mutant or missing protein components. Examples of cases like these include experiments where spliceosome components are knocked down, and in splicing factor mutant strains, possibly including those

associated with human diseases. Are these aberrant spliceosomes targeted for disassembly and recycled, just as spliceosomes with aberrant pre-mRNAs are proposed to be discarded and disassembled? The answer to this question is important in two ways. Firstly, it is important to accurately interpret knockdown experiments of spliceosome components. Secondly, it will improve our understanding of how cells coordinate the assembly and disassembly cycle of the spliceosome and other large molecular particles.

A common way to study the role of splicing factors *in vivo* is through RNAi or protein depletion knockdown techniques, transcriptional repression or conditional mutant strains. In some situations, the knockdown of one element may cause indirect disturbances in the whole system, making the results difficult to interpret. Separating specific from indirect effects is critical for most functional analyses. In the case of splicing, because the spliceosome exists in a highly dynamic cycle of assembly and disassembly, where elements seldom exist in isolation, it is not unlikely that perturbing one element will affect other parts the cycle, which can lead to misleading interpretations. This is one reason why it is important to understand how aberrant complexes are dealt with by the cell. When knocking down splicing factors, if aberrant complexes are not disassembled and recycled, it could lead to a recycling defect with systematic effects on the assembly pathway making it difficult to separate the direct from indirect effects.

To improve our understanding of spliceosome disassembly and recycling *in vivo*, my objectives were to deplete splicing factors required at different stages of assembly and ask 1) do arrested complexes accumulate? 2) Is spliceosome assembly affected? And, if this is the case 3) how quickly? To this end, I used the auxin-inducible degron (AID) system (Nishimura et al., 2009) for a quick depletion of splicing factors, to allow analysis before secondary effects appear;

RNA immunoprecipitation (RIP) that indicates if intermediate complexes have accumulated; and chromatin immunoprecipitation (ChIP) as a proxy for spliceosome assembly.

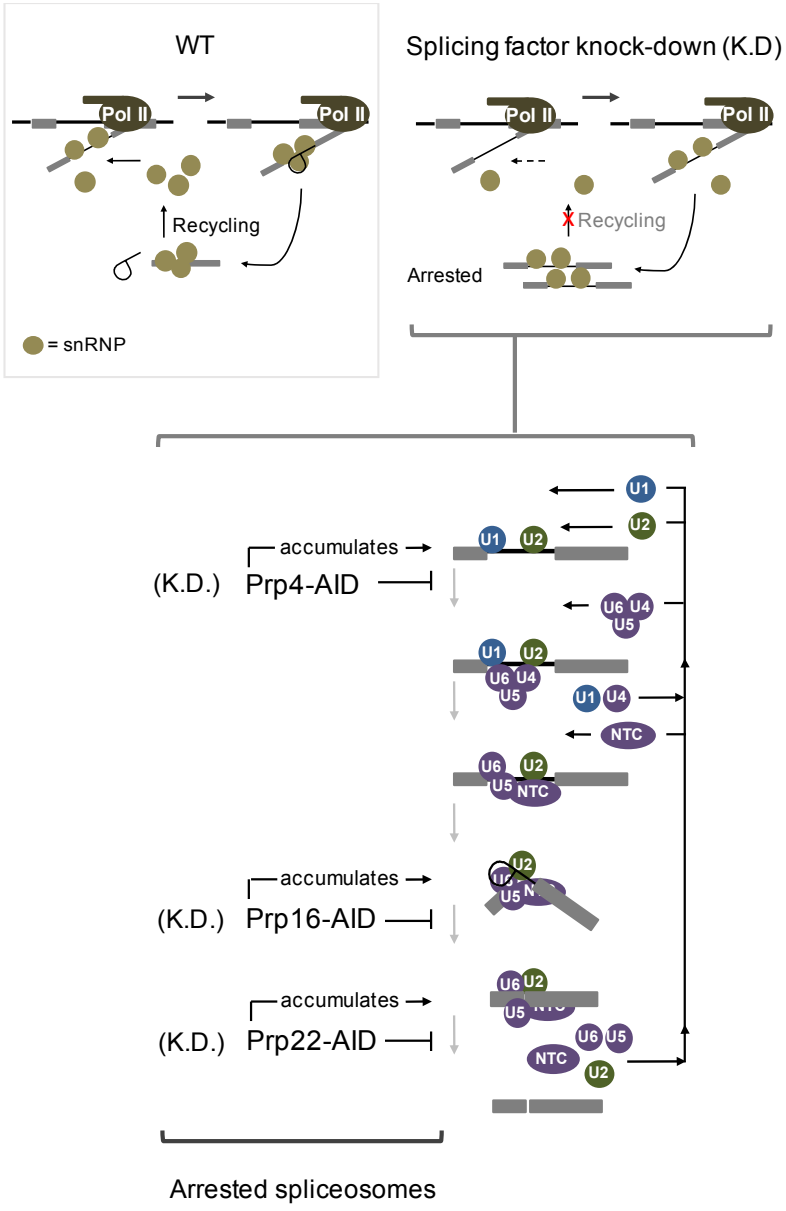
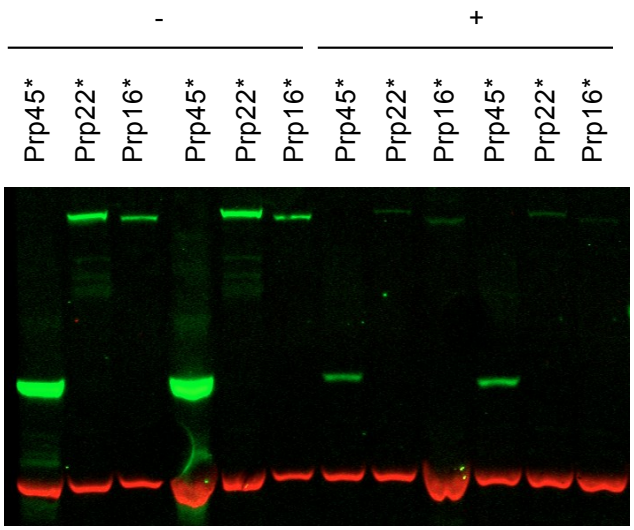
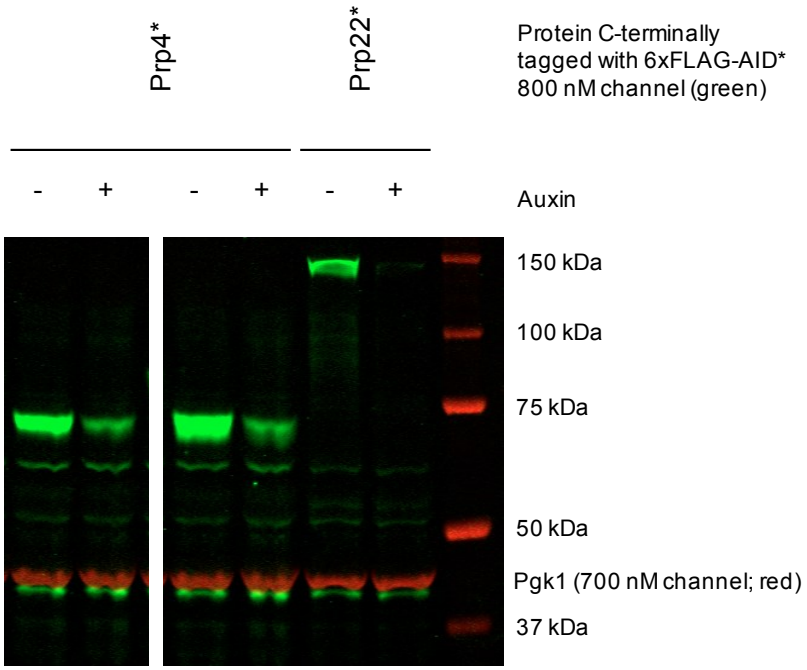


Figure 3.1. Schematic representation of the model of the arrested spliceosomes. On normal conditions (WT) spliceosome components (e.g. snRNPs), are co-transcriptionally recruited to the newly synthesized pre-mRNA transcripts. After splicing completion, spliceosome components are release and are free to bind new transcripts. When splicing factors are knocked-down (K.D.) (e.g. protein depletion), arrested spliceosomes can accumulate and, as a result, inhibit the recycling process. The schematic bellow shows which complex intermediate accumulates when Prp4, Prp16 or Prp22 are depleted.

A


-
+
Auxin
Prp45*
Prp22*
Prp16*
Prp45*
Prp22*
Prp16*
Prp45*
Prp22*
Prp16*
Prp45*
Prp22*
Prp16*
Protein C-terminally
tagged with 6xFLAG-AID*
(800 nM channel; green)
Pgk1
(700 nM channel; red)

B


Prp4*
Prp22*
Protein C-terminally
tagged with 6xFLAG-AID*
800 nM channel (green)
- + - + - +
Auxin
150 kDa
100 kDa
75 kDa
50 kDa
Pgk1 (700 nM channel; red)
37 kDa

C

Protein	Size (kDa)	Protein levels after 30 min. of auxin addition
Prp45*	58	6% ± 0.4
Prp22*	145	6% ± 0.5
Prp16*	137	7% ± 0.2
Prp4*	68	19% ± 0.2
Pgk1	45	

Figure 3.2. AID-tagged Prp4, Prp16, Prp22 and Prp45 were depleted to low levels by 30 minutes of auxin induction. Auxin was added to cultures of four different strains, each one with an AID*-6FLAG C-terminal tag on **(A)** PRP45, PRP22, PRP16 or **(B)** PRP4 genes. Samples were analyzed before (-) and after 30 minutes (+) of auxin (IAA) addition. The abundance of the AID target was measured with anti-FLAG LICOR-based western blotting. **(C)** Protein levels after auxin addition were calculated as the ratio of + auxin/- auxin (x100) of anti-FLAG signal, normalized to anti-Pgk1 signal as a loading control. Three biological replicates where analyzed (only two of them are shown in panels A-B). Standard error of the mean of the three biological replicates was calculated (±).

Results

Depletion of essential splicing factors can cause an accumulation of intermediate complexes

I analysed the AID depletion of splicing factors Prp4, Prp16, Prp22 and Prp45 (Figure 3.2). The main reason I focused on these four proteins is that they act at different stages of assembly and, therefore, cumulatively would provide a wider perspective of the effects of the absence of essential splicing factors. In the AID system, addition of auxin to the medium allows the heterologous plant Tir1 protein to direct the AID-tagged target for proteasome degradation. There are many variants of the AID tag. I used the AID*-6FLAG (Morawska and Ulrich, 2013). AID-tagged strains Prp16, Prp22 and Prp45 derive from YGM1 and Prp4 derives from YBRT1. The differences between YGM1 and YBRT1 are explained in more detail in Chapter 4. Based on the time course

depletions described in Chapter 4, most analyses presented in this chapter were at 30 minutes after auxin addition. At that time, pre-mRNA abundance has stabilized at its highest level.

Using RIP to measure changes in the balance of intermediate complexes of the spliceosome

To test whether intermediate complexes of the spliceosome accumulate when Prp4, Prp16, Prp22 or Prp45 is depleted, I performed RIP. This was done by pulling down Lea1 (U2 snRNP) and then measuring assosociated snRNAs (released from the beads) by RT-qPCR. In normal conditions, a RIP assay results in U2 interacting with U1 and U4 at low levels, but more with U5 and U6 (Figure 3.3A). The changes in the level of association between U2 and the other snRNPs, can show if complexes are accumulating and if so, helps to indicate which intermediate complex accumulates, if the composition is relatively homogeneous.

Depletion of Prp4 increases association of U2 with U1 and decreases association of U2 with U5 and U6, relative to undepleted

Prp4 is a component of the U4 snRNP and is required for tri-snRNP recruitment (Lygerou et al., 1999), which allows the pre-spliceosome to transition to Complex B. As expected, when Prp4 is depleted I observe that association of Lea1 with U1 is increased, and its association with U4, U5 and U6 snRNA decreases, relative to undepleted culture (- auxin) (Figure 3.3B). This suggests that without the recruitment of the tri-snRNP, spliceosome assembly is stuck at the pre-spliceosome stage, which contains U1 and U2 but not U4, U5 and U6 snRNP.

Depletion of Prp16 reduces association of U2 with U4 and increases association of U2 with U5 and U6, relative to undepleted

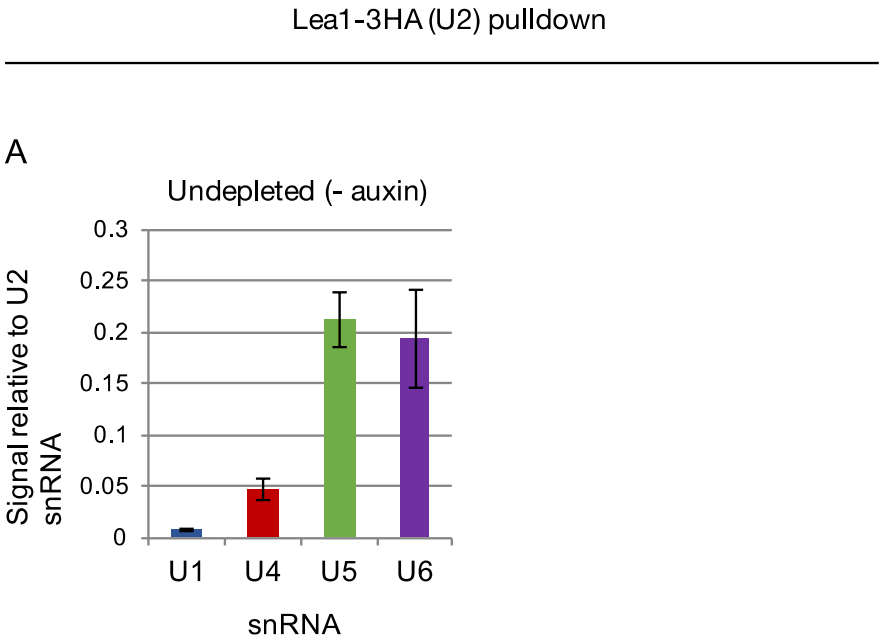
After the first catalytic step of splicing, the spliceosome undergoes structural rearrangements that set the stage for the second catalytic splicing reaction. These intermediate complexes, between the first and second step, are called Complexes C and C* (catalytically activated). Prp16 is recruited at this stage and plays a role in the second step of splicing. Interestingly, when Prp16 is depleted, I observe that association of Lea1 with U1 and U4 is decreased and its association with U5 and U6 snRNAs is increased relative to undepleted (Figure 3.3B), which agrees with an accumulation of Complex C that contains U2, U5 and U6 but not U1 or U4 snRNAs (Table 3.1).

Depletion of Prp22 increases association of U2 with U5 and U6, relative to undepleted

Prp22's main role is the release of the mRNA after splicing completion, with a proposed additional (ATP-independent) role in the second step of splicing for some genes (Schwer and Gross, 1998). This splicing factor is recruited to complex C* (Fica et al., 2017), which contains U2, U5 and U6 snRNP. When depleting Prp22 I observed that association of Lea1 with U5 and U6 snRNA is increased relative to undepleted (Figure 3.3B), almost to the same extent as seen with depletion of Prp16. This suggests that, like depletion of Prp4 and Prp16, depletion of Prp22 results in accumulation of an arrested complex. The lack of decrease in U2's association with U4 snRNA, in contrast to depletion of Prp16, could reflect an abnormality of the post-spliceosome that has accumulated.

RIP analysis of Prp45 depletion

The nineteen complex, and possibly also Prp45 (Makarov et al., 2002), is reported to join complex B, stabilizing complex B* when U1 and U4 snRNPs leave (Tarn et al., 1993). When depleting Prp45, I observe smaller changes in the association of U2 with the other snRNAs, compared to the other strains. Although this RIP profile does not clearly point to the accumulation of any specific intermediate complex (Table 3.1), it may reflect a mild accumulation of Complex B – the stage where Prp45 is predicted to join.



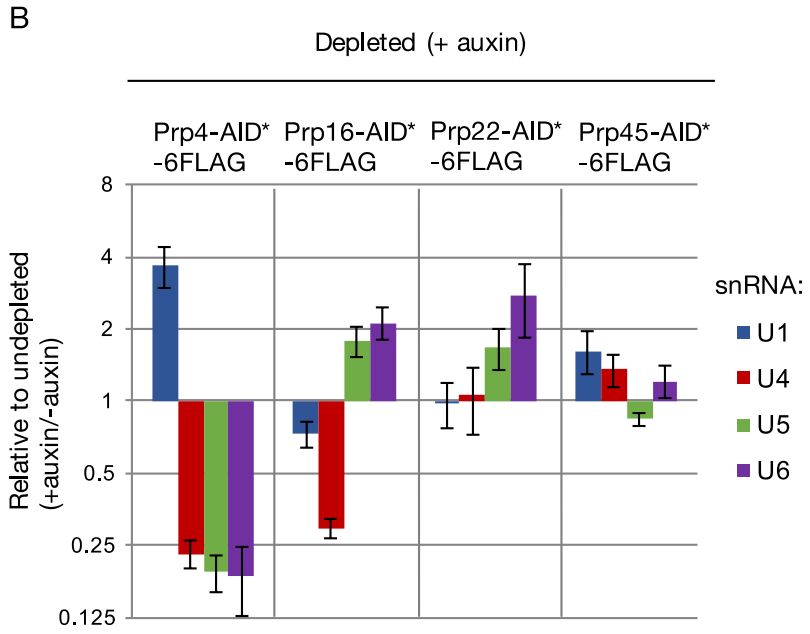


Figure 3.3. Intermediate complexes of the spliceosome accumulate in the absence of splicing factors.
(A) RIP assay where Lea1-3HA was pulled down and from the beads U snRNAs were measured by RT-qPCR. Data is presented as signal relative to U2 snRNA, to correct for differences in pull-down efficiencies.
(B) RIP assays as in (A), before and after depletion of Prp4-AID*-6FLAG, Prp16-AID*-6FLAG, Prp22-AID*-6FLAG or Prp45-AID*-6FLAG. Data are presented as changes relative to undepleted cultures (+auxin/-auxin). Error bars denote standard error of three biological replicates.

Table 3.1. Predicted change in Lea1 (U2) interaction with U1, U4, U5 and U5 snRNAs (by RIP) when intermediate complexes accumulate

Arrested:	U1	U4	U5	U5
Pre-spliceosome	↑	↓	↓	↓
Complex B	↑	↑	↑	↑
Complex Bact	↓	↓	↑	↑
Complex B*	↓	↓	↑	↑
Complex C	↓	↓	↑	↑
Post-spliceosome	↓	↓	↑	↑

The conclusion from the RIP data

Overall, I conclude that intermediate complexes of the spliceosome accumulate in the absence of splicing factors Prp4, Prp16 or Prp22, which suggests that proper disassembly does not happen when essential splicing factors are missing. The following question is whether the depletions also affect new rounds of spliceosome assembly. In other words, does the disassembly defect cause a recycling defect?

Pre-spliceosome formation is inhibited when depleting Prp4, Prp12, Prp22 or Prp45

ChIP as a measure of spliceosome assembly

Since it was demonstrated by ChIP that splicing happens co-transcriptionally (Görnemann et al., 2005; Kotovic et al., 2003; Tardiff and Rosbash, 2006), this technique has been used many times to analyse snRNP recruitment *in vivo*. The concept of co-transcriptionality dictates that spliceosome components can only be ChIPed to the DNA while they are assembled on newly synthesized pre-mRNAs. Therefore, this technique can be used as a proxy for newly assembled spliceosomes. From now onwards, frequently I will use the term spliceosome assembly or pre-spliceosome formation instead of co-transcriptional spliceosome recruitment.

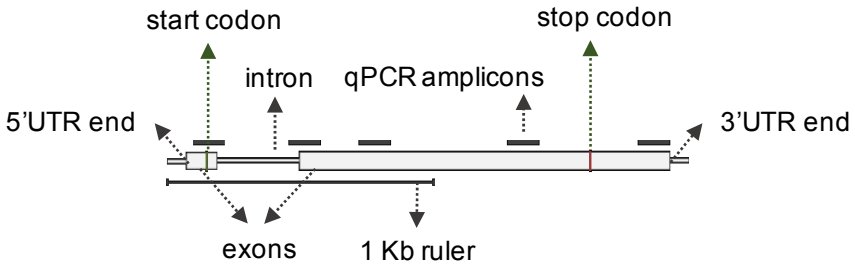
A characteristic of this technique is that different splicing factors may produce different gene occupancy profiles. The earlier the splicing factor is recruited to the transcript, the more 5′ the ChIP signal occurs on the gene. Hence, splicing factors crosslink to DNA sites that indirectly correlate with residence within the spliceosome assembly cycle rather than binding site on the RNA.

As a measurement of assembly of the spliceosome, I performed ChIP on core splicing factors Prp40 (U1 snRNP) and Lea1-3HA (U2 snRNP), and qPCR on intron-containing genes *ACT1*, *ECM33* and *RPS13*. U1 and U2 snRNPs are the main components of the pre-spliceosome, one of the earliest intermediate complexes at the start of the spliceosome cycle. Thus, this combined ChIP analysis will indicate whether early spliceosome assembly is affected when depleting Prp4, Prp16, Prp22 and Prp45.

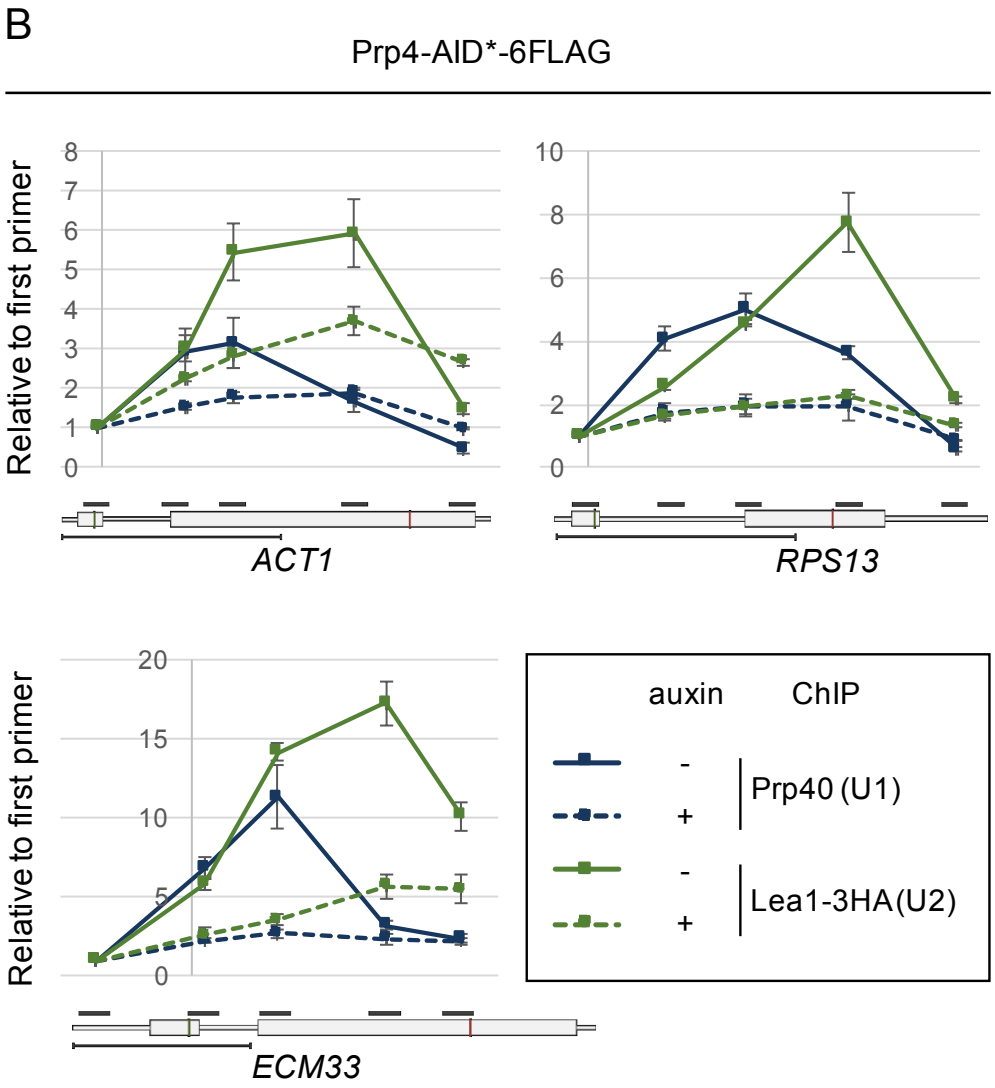
Depletion of Prp4 reduces ChIP occupancy of U1 and U2

On normal condition (no auxin), the ChIP profiles I observe are as expected, with Prp40 (U1) signal peaking near the 3'ss and Lea1 (U2) around the middle of the second exon (Figure 3.4B). The drop in U1 ChIP signal after the 3'ss reflects its displacement as the spliceosome cycle progresses. The profile is different across the different genes. This is due to the influence of several factors including the structure of the gene and the position of the amplicons along the gene. This is one reason why more than one gene was analysed. Interestingly, after depleting Prp4 I observe a reduced ChIP signal of Prp40 (U1) and Lea1 (U2), on the three genes analysed (Figure 3.4B), indicating that formation of the pre-spliceosome is reduced when recruitment of the tri-snRNP is blocked.

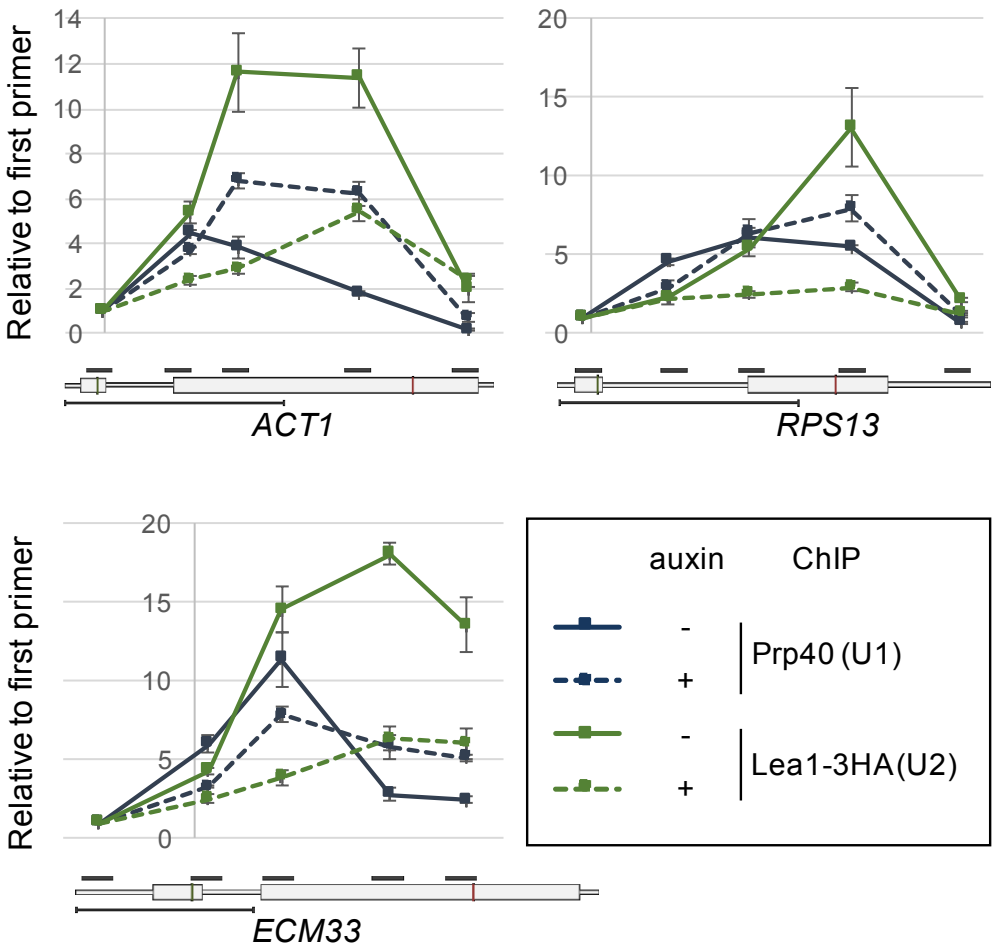
A


start codon
stop codon
intron
qPCR amplicons
5'UTR end
3'UTR end
exons
1 Kb ruler

B

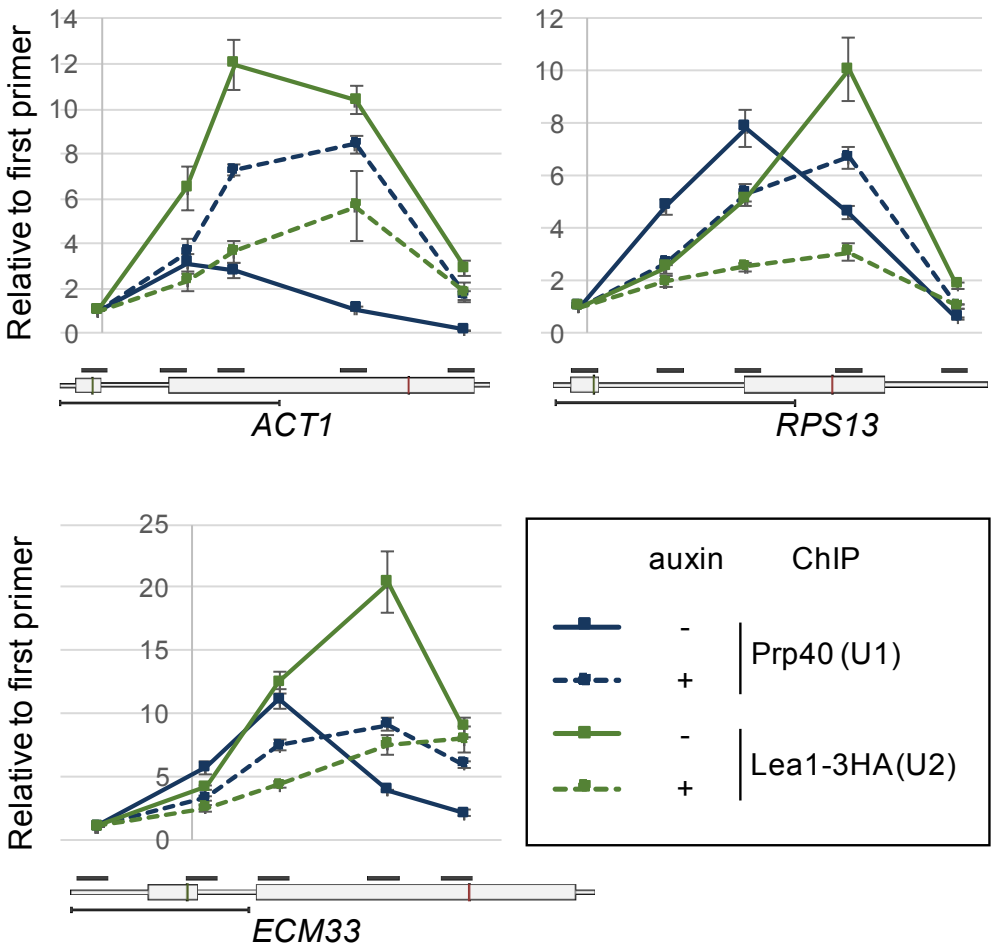

Prp4-AID*-6FLAG
Relative to first primer
ACT1
RPS13
ECM33
auxin
ChIP
-
+
Prp40 (U1)
Lea1-3HA(U2)

C


Prp16-AID*-6FLAG
Relative to first primer
14
12
10
8
6
4
2
0
ACT1
20
15
10
5
0
RPS13
Relative to first primer
20
15
10
5
0
ECM33
auxin
ChIP
-
+
Prp40 (U1)
-
+
Lea1-3HA(U2)

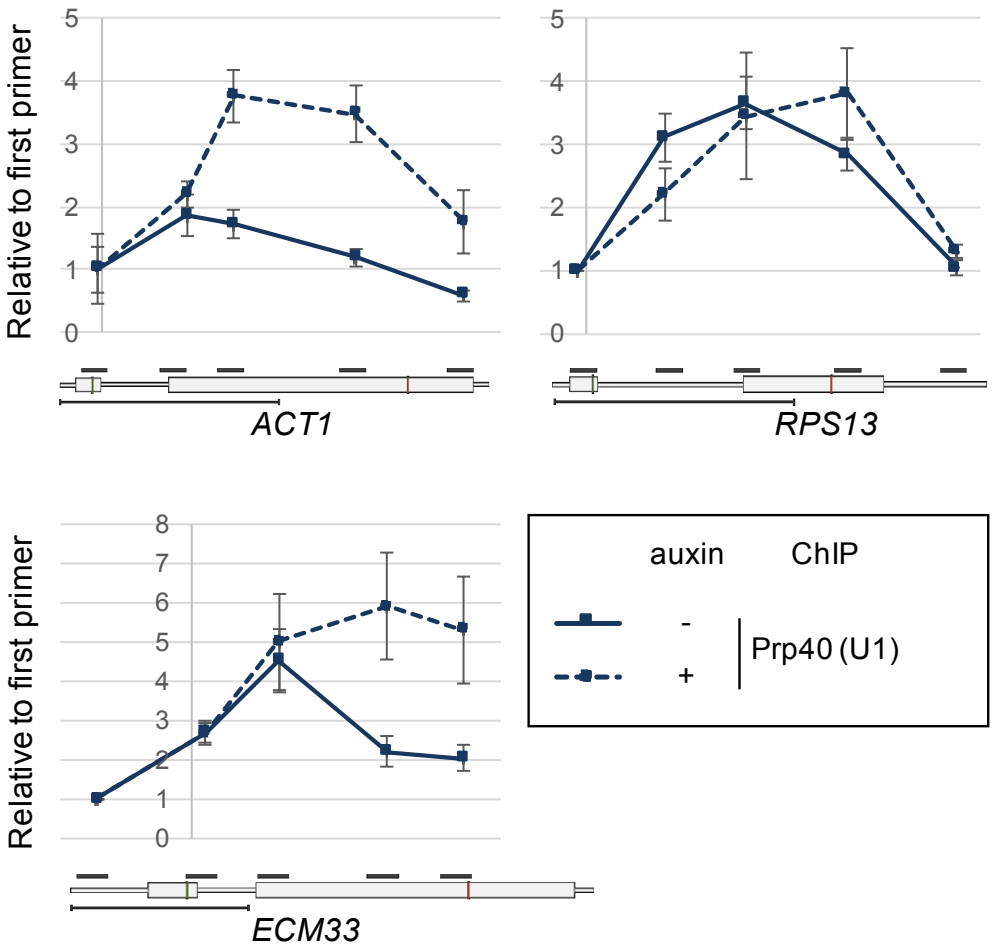
D

Prp22-AID*-6FLAG



E

Prp45-AID*-6FLAG



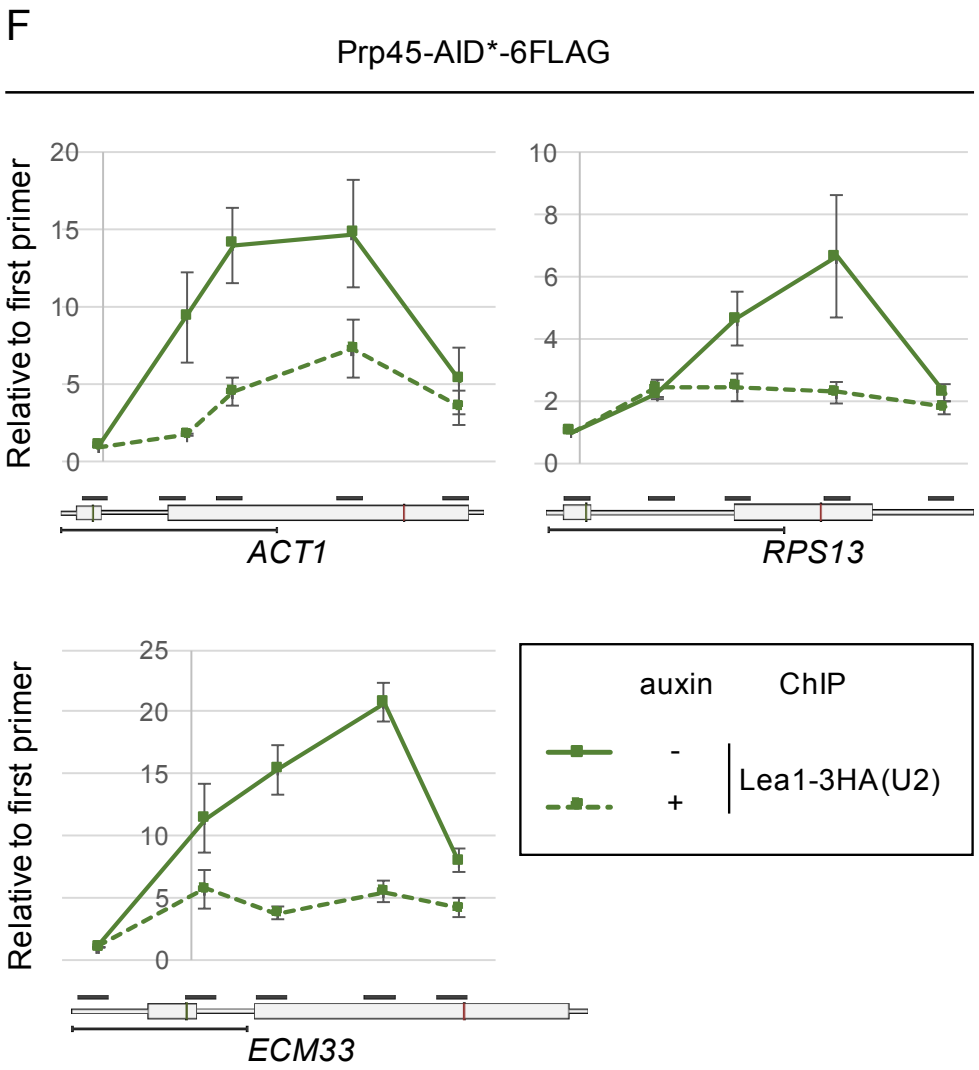


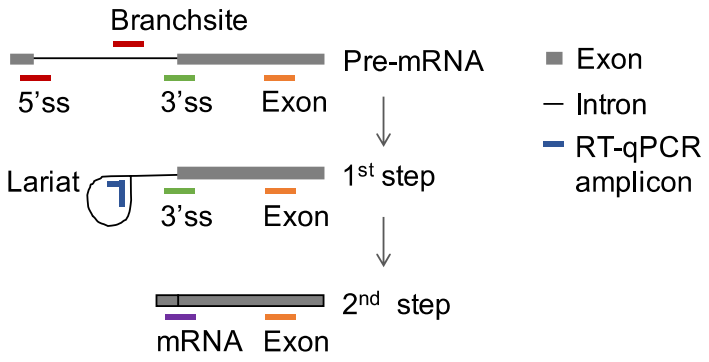
Figure 3.4. Co-transcriptional assembly of the pre-spliceosome is reduced in the absence of splicing factors Prp4, Prp16, Prp22 or Prp45. ChIP of Prp40 (blue lines) and Lea1-3HA (green lines) from the U1 and U2 snRNPs, respectively, was performed before and after auxin-induced depletion. qPCR was done with primer for *ACT1, ECM33* and *RPS13*. Data is presented as ChIP signal relative to the first amplicon of the gene (see Materials and Methods). The x-axis contains amplicon location within the structure of the gene. Solid lines represent data of undepleted (- auxin) and dashed lines of depleted (+ auxin). Error bars denote standard error of three biological replicates. U1 and U2 ChIP when depleting Prp45-AID*-6FLAG are plotted separately (E-F) since they are at different scales due to experimental variations. **(A)** Labels of the gene diagrams in the x-axis of (B-E) **(B)** Depletion of Prp4-AID*-6FLAG **(C)** Depletion of Prp16-AID*-6FLAG **(D)** Depletion of Prp22-AID*-6FLAG **(E)** Depletion of Prp45-AID*-6FLAG (U1 ChIP) **(F)** Depletion of Prp45-AID*-6FLAG (U2 ChIP)

Depletion of Prp16, Prp22 or Prp45 increases ChIP occupancy of U1, and reduces occupancy of U2

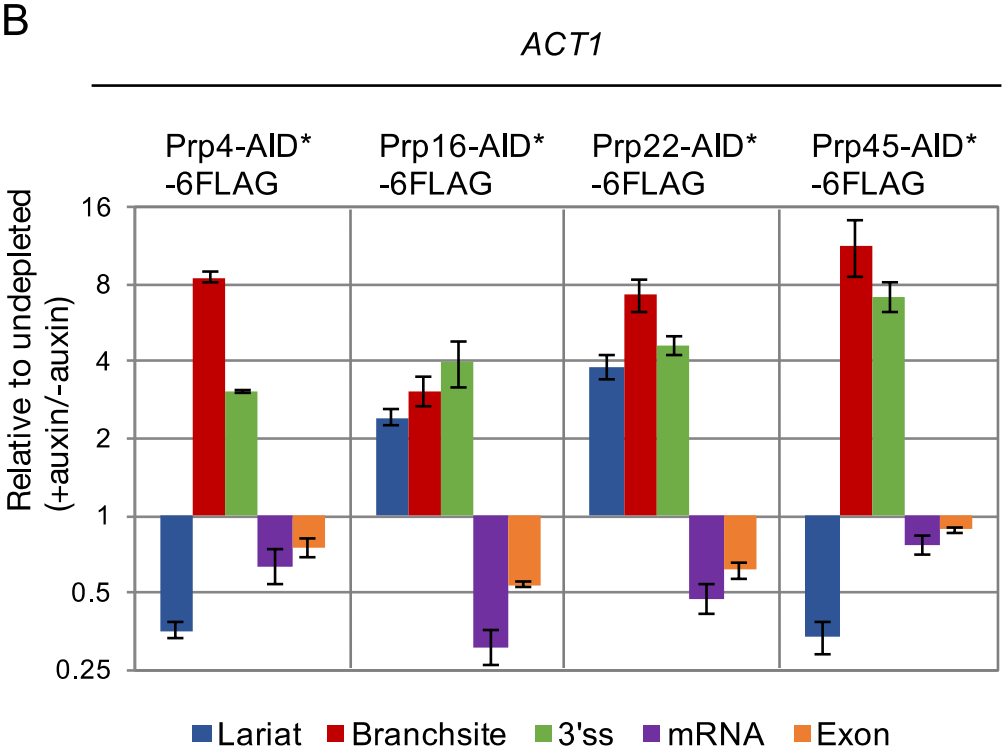
In the depletion of late-acting splicing factors Prp16 or Prp22, or NTC-associated protein Prp45, I observed a reduced ChIP signal of Lea1 (U2), with Prp40 ChIP signal persisting towards the 3' end of the gene (Figure 3.4C-F). The change in Prp40 ChIP profile seems to be gene dependent, as Prp40 (U1) is pronouncedly increased on *ACT1* but only slightly on *ECM33* or *RPS13*. The increase in Prp40 ChIP signal is likely a consequence of the absence of U2 snRNP. Without the progression of assembly, U1 snRNP is not displaced and would crosslink stronger towards the end of the gene compared to the untreated control. As in the Prp4 depletion, these results indicate that pre-spliceosome formation is reduced when depleting Prp16, Prp22 or Prp45.

Together, the RIP and ChIP analyses suggest that, when Prp4, Prp16, Prp22, or Prp45 is missing, accumulation of arrested complexes sequesters splicing machinery and prevents formation of new pre-spliceosomes. The following question is whether splicing catalysis is also affected.

A


Branchsite
Pre-mRNA
5'ss
3'ss
Exon
Lariat
1st step
3'ss
Exon
2nd step
mRNA
Exon
Exon
Intron
RT-qPCR amplicon

B


ACT1
Prp4-AID* -6FLAG
Prp16-AID* -6FLAG
Prp22-AID* -6FLAG
Prp45-AID* -6FLAG
Relative to undepleted (+auxin/-auxin)
16
8
4
2
1
0.5
0.25
Lariat
Branchsite
3'ss
mRNA
Exon

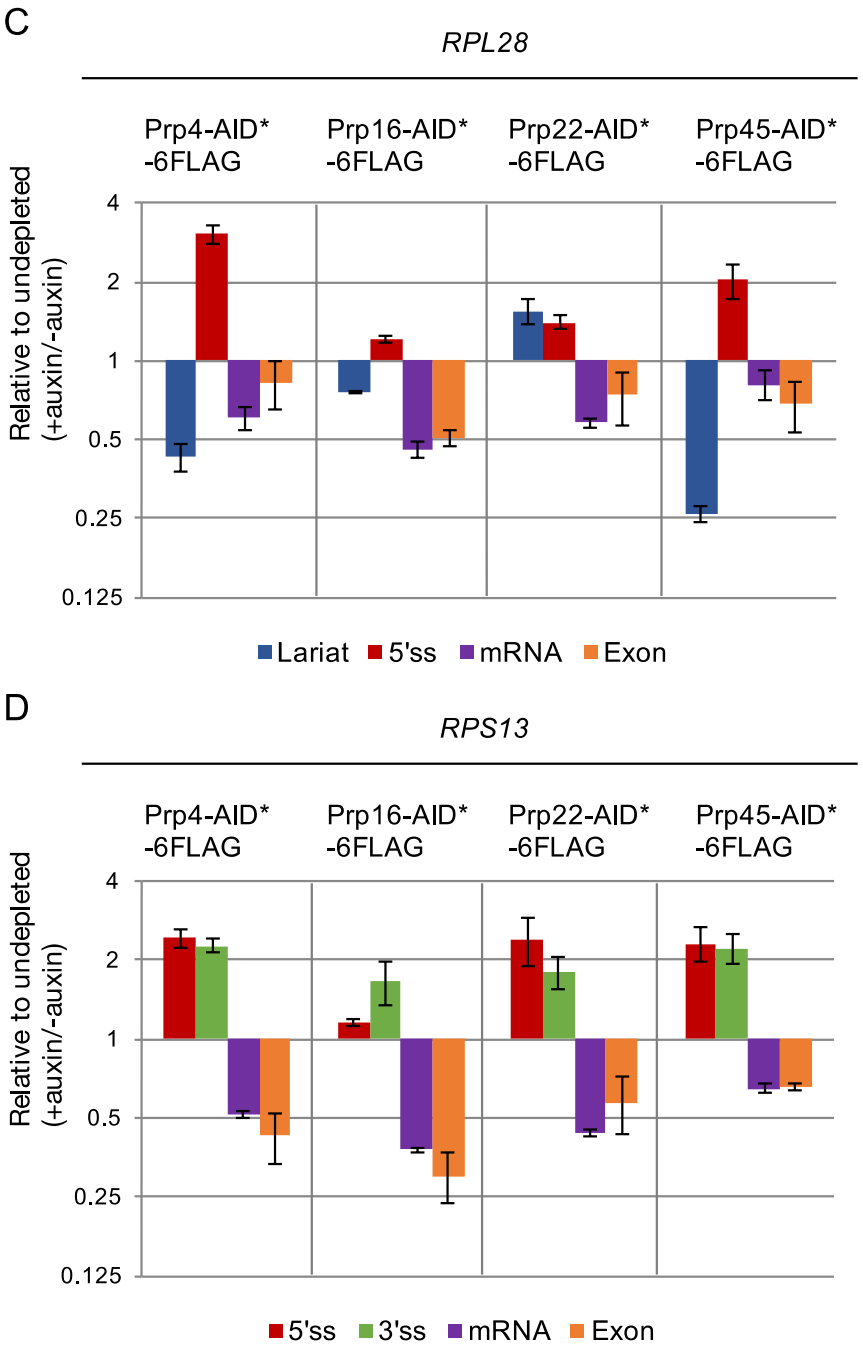


Figure 3.5. Depletion of Prp16 or Prp22 causes a first step of splicing defect. **(A)** Diagram of the principle behind using RT-qPCR to measure the efficiency of the first and second steps of splicing. Samples were analyzed before and after auxin-induced depletion of Prp4-AID*-6FLAG, Prp16-AID*-6FLAG, Prp22-AID*-6FLAG or Prp45-AID*-6FLAG on **(B)** *ACT1*, **(C)** *RPL28* and **(D)** *RPS13*. Signal was normalized against *ALG9* and data is presented as changes relative to undepleted (+auxin/-auxin). Error bars denote standard error of three biological replicates.

Measuring splicing efficiency *in vivo*

Splicing efficiency can be measured by RT-qPCR

David Barrass (this lab) designed qPCR primers that specifically detect 5′ splice site (5′ss) or branch site (BS), 3′ splice site (3′ss), lariat, mRNA and exon (Figure 3.5A) of intron-containing genes (Alexander et al., 2010b), including *ACT1*, *RPL28* and *RPS13*. An increase in 5′ss or BS and 3′ss abundance is interpreted as a defect in the first step of splicing; an increase in lariat and 3′ splice site abundance is interpreted as a defect in the second step of splicing; and a defect in mRNA release would likely result in protection of the excised lariat that is still in the post-spliceosome complex, from debranching enzyme Dbr1 (Martin et al., 2002), so in this case we will see an increase in lariat but not BS or 3′ss. Therefore, with this assay, one can measure the relative efficiency of splicing and distinguish between a first and second step defect of the two catalytic steps of splicing.

Late-acting splicing factors, Prp16 and Prp22, are required for the first catalytic step of splicing in vivo

After depleting Prp16 and Prp22 I observed an increase in lariat and 3′ss abundance of *ACT1* (Figure 3.5B), which is indicative of a defect in the second step. It is important to remember that although Prp22's main role is mRNA release, it has been proposed to have an additional role in the second step of splicing (Schwer and Gross, 1998) for some genes, including *ACT1* – the gene that was analysed here. As predicted by the hypothesis, depletion of Prp16 or Prp22 results in a defect of the first step as well as the second step of splicing, as seen by an increase in BS signal; in contrast, depletion of Prp4 or Prp45 results in a first but not a second step defect (Figure 3.5B). This is an interesting observation by itself, because there is no evidence that Prp16 or Prp22 is

recruited to the spliceosome before the first step of splicing. The RT-qPCR analysis of two other genes, *RPL28* and *RPS13* (Figure 3.5B), and an exon normalization instead of an *ALG9* normalization (Figure S3.1), overall show similar results as for *ACT1.* Overall, the results suggest that the accumulation of arrested complexes, caused by depletion of Prp16 or Prp22, also affects splicing catalysis, not only the co-transcriptional assembly of the spliceosome.

Kinetic analysis of Prp22 depletion

Finally, I asked how soon the recycling defect takes place after depletion of Prp22. One reason why addressing this question is important, is that it will show if it is possible to separate the direct from indirect effects of the induced depletion – something that is desirable in almost any knockdown experiment. Secondly, the result would indicate how critical is recycling for new rounds of splicing. By depleting Prp22, I take the direct effect as the reduced efficiency of the second step of splicing, which eventually feeds-back and indirectly affects the first step of splicing.

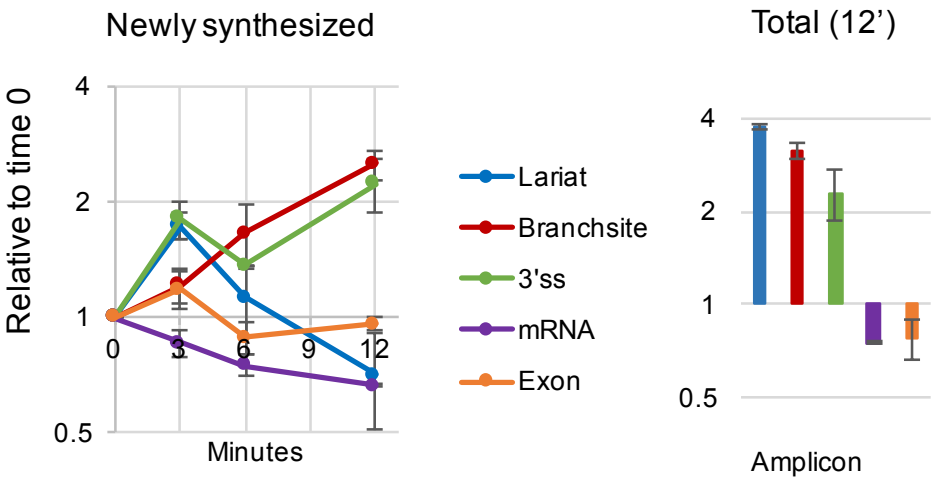
Depletion of Prp22 quickly causes a defect in the first step of splicing

To this end, I used the 4-thiouracil (4tU) technique developed by David Barrass, which allows the measurement of splicing efficiency of 4tU-labelled newly synthesised RNA (nsRNA) (Barrass et al., 2015). With this technique, I performed a time course analysis, of 0, 3, 6 and 12 minutes after auxin-induced depletion of Prp22, with a 1 minute 4tU labelling of each sample. nsRNA abundance was measured by RT-qPCR with primers for *ACT1*, *ECM33*; ribosomal protein genes *RPL28*, *RPS13* and *RPL39*; and control primers for *ALG9* (intron-less gene), *U6* and *SCR1* (a very abundant Pol III transcript) for normalization.

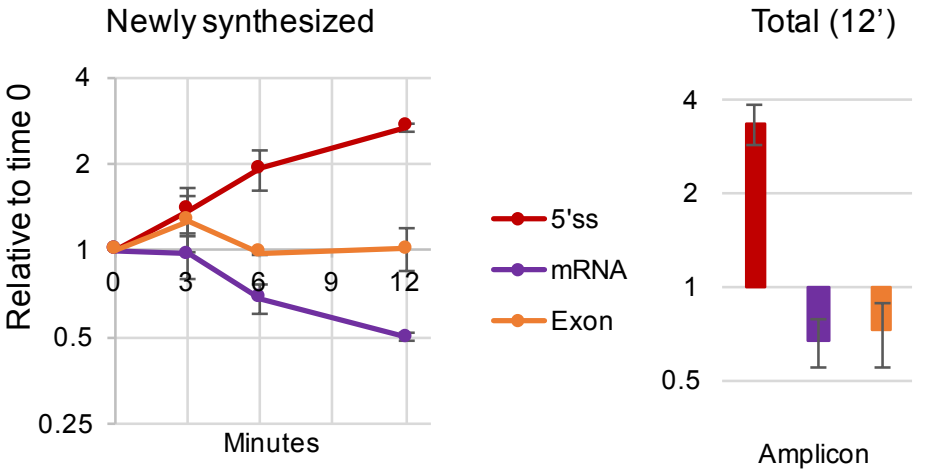
After depleting Prp22, on ACT1 I observed an apparent switch from a second step defect (lariat and 3'ss increase) to a first step defect (lariat decrease and branchsite increase) of splicing of newly synthesized *ACT1* (Figure 3.6A). Interestingly, the second step defect (minute 3) of splicing is quickly followed by the first step defect (minute 6), with barely enough time to distinguish the two states. This kinetic analysis on *ACT1* agrees with the hypothesis that the defect in the second step of splicing is a primary consequence of depleting Prp22, and the defect in the first step of splicing of *ACT1* (and possibly other genes too) is a secondary effect.

The first step of splicing of other genes is affected earlier (Figure 3.6B_E). A first step defect (5'ss increase) with *ECM33* is observed already at 3 minutes, slightly earlier than with *ACT1*. Pre-mRNA abundance (5'ss and/or 3'ss) of RPGs *RPS13*, *RPL39* and *RPL28* also increases but to a lesser extent (around 1.5-fold). However, lariat abundance of *RPL28* is more affected (6-fold decrease at minute 12) than *RPL28* pre-mRNA. In contrast to *ACT1*, the abundance of newly synthesized lariat of *RPL28* does not increase before it drops, which could suggest that Prp22 is not required for the second step of splicing of this transcript or that the first step of splicing appeared so quickly that the phenotype associated with the second step of splicing was unresolved on this transcript. I conclude that depletion of Prp22 quickly causes a defect in the first step of splicing.

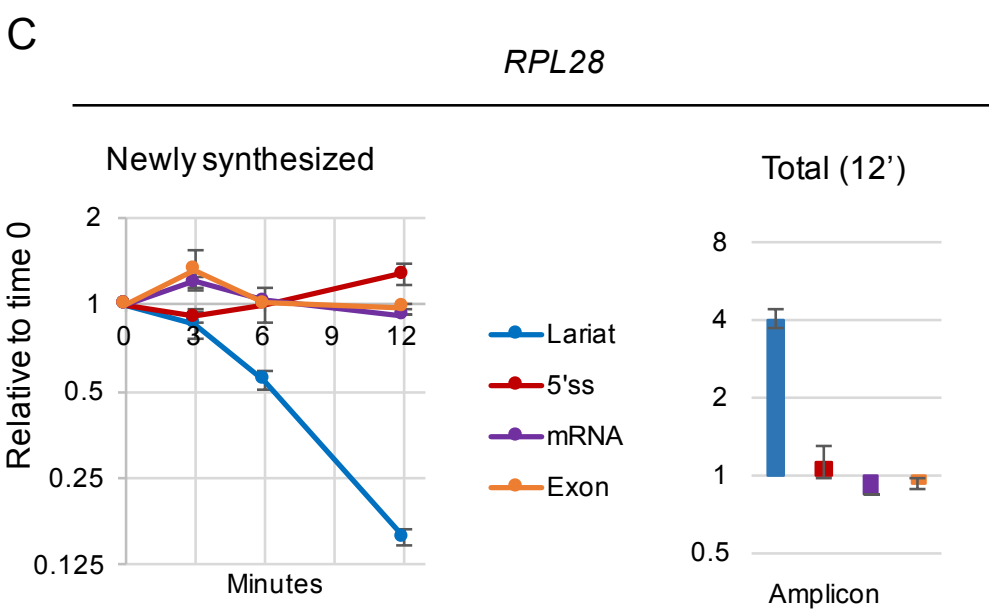
A


ACT1
Newly synthesized
Total (12')
Relative to time 0
4
2
1
0.5
0
3
6
9
12
Minutes
Lariat
Branchsite
3'ss
mRNA
Exon
Amplicon

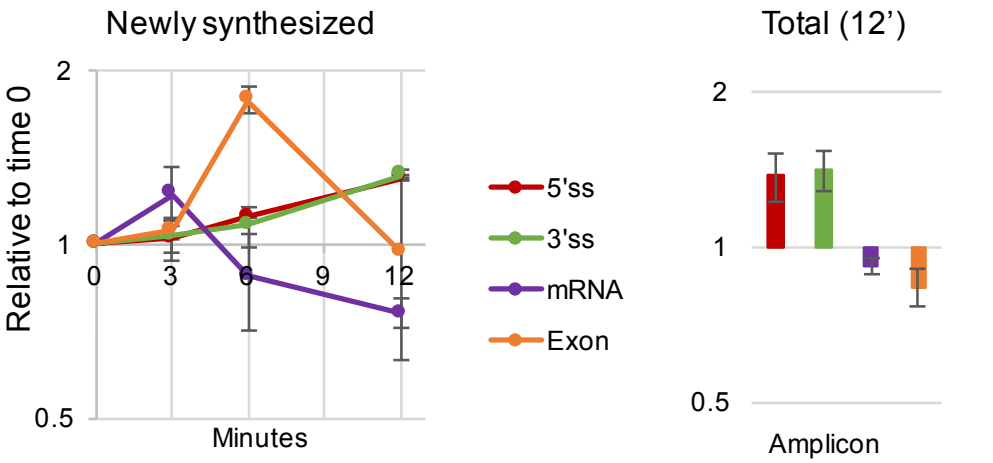
B


ECM33
Newly synthesized
Total (12')
Relative to time 0
4
2
1
0.5
0.25
0
3
6
9
12
Minutes
5'ss
mRNA
Exon
Amplicon

C


RPL28
Newly synthesized
Relative to time 0
2
1
0.5
0.25
0.125
0
3
6
9
12
Minutes
Lariat
5'ss
mRNA
Exon
Total (12')
8
4
2
1
0.5
Amplicon

D


RPS13
Newly synthesized
Relative to time 0
2
1
0.5
0
3
6
9
12
Minutes
5'ss
3'ss
mRNA
Exon
Total (12')
2
1
0.5
Amplicon

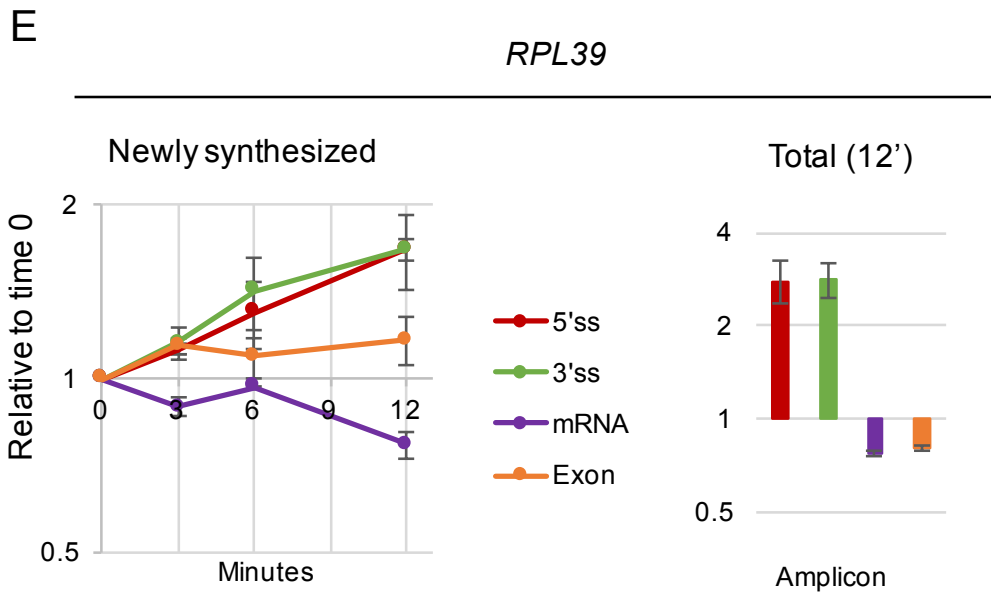
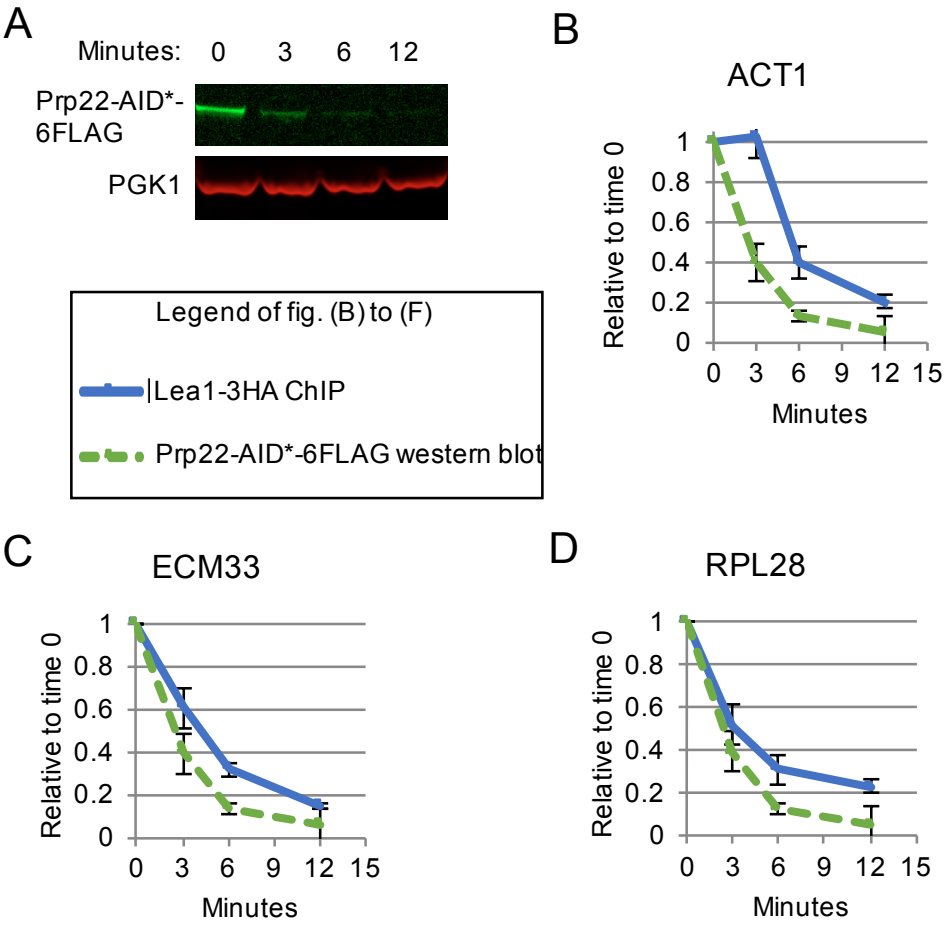


Figure 3.6. Depletion of Prp22 quickly causes a defect in the first step of splicing. 4tU-labeled newly synthesized RNA was measured, from time-course samples of auxin-induced depletion of Prp22-AID*-6FLAG. RT-qPCR was done with primers for **(A)** *ACT1*, **(B)** *ECM33*, **(C)** *RPL28*, **(D)** *RPS13* and **(E)** *RPL39*. Total RNA (unlabeled) was also analyzed from 0 and 12 minute samples. Signal was normalized against *SCR1* and data is presented as relative to time 0. X-axis of the time course plots (left) represents minutes after addition of auxin. Error bars denote standard error of four biological replicates.

pre-spliceosome formation is quickly reduced

To complement the 4tU kinetics result, I performed an additional Prp22 depletion time course, with ChIP of Lea1 (U2) and western blot of Prp22 (Figure 3.7). As anticipated, on *ACT1*, Lea1 ChIP signal drops at the same time (6 minutes after auxin) as the second step defect switches to a first step defect, as previously shown by the 4tU RT-qPCR analysis. On *ECM33*, Lea1 ChIP signal is already reduced 3 minutes after auxin addition, which also correlates with the 4tU RT-qPCR assay that shows an increase in *ECM33* pre-mRNA already at minute 3. As for *ECM33*, Lea1 ChIP signal on the RPGs drops very quickly after adding auxin, almost as quickly as Prp22 is being depleted. Prp22 ChIP analysis on *ACT1* and *ECM33* confirms that after auxin addition, Prp22 that is engaged in co-transcriptional splicing, drops as quickly as total protein

levels of Prp22 (Figure S3.3). Taking together the 4tU and ChIP kinetics analyses, I conclude that the absence of Prp22 leads to a quick decrease in pre-spliceosome formation.



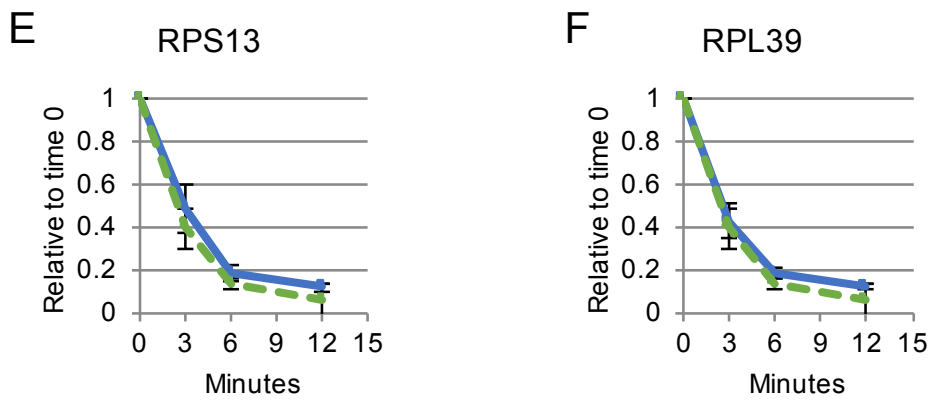


Figure 3.7. Co-transcriptional recruitment of U2 snRNP is quickly inhibited after Prp22 depletion.
(A) Anti-FLAG and anti-Pgk1 western blot after auxin addition.
(B-F) Lea1-3HA ChIP-qPCR of *ACT1*, *ECM33*, *RPL28*, *RPS13* and *RPL39* after auxin addition. ChIP data (blue solid line) is presented as fold over background (*ALG9*), and relative to time 0. Quantification of (A), as Pgk1 normalized and relative to time 0, is plotted in (B-F) as dashed green lines. X-axis represents time (minutes) after addition of auxin. Error bars denote standard error of four biological replicates.

Discussion

Depletion of essential splicing factors can cause an accumulation of intermediate complexes

The RIP data suggest that the absence of essential splicing factors can cause arrested spliceosomes to accumulate. This means that *in vivo*, these types of aberrant complexes that contain normal substrates but lack an essential splicing factor, are not properly disassembled. This is apparently in contrast with the proposal that defective spliceosomes are targeted for disassembly, mainly based on evidence that mutations in disassembly factors Spp382 (NTR1) and Prp43 suppress the growth phenotype of *prp38-1* and *prp8-1* (Konarska and Query, 2005). Taking together both analyses, it seems plausible that when the splicing defect is strong, as in depletion of essential splicing factors, if disassembly happens, it is at a rate lower than the rate of accumulation of arrested spliceosomes. After all, the absence of an essential

splicing factor is not physiologically normal, therefore, it is not difficult to believe that the cell is not fully equipped to deal with this situation.

The RIP data have some gaps, in part due to the limitations of the technique. Firstly, it must be pointed out that even though the RIP technique produces quantitative data, it does not provide accurate quantitative information about the level of accumulation of an intermediate complex, with respect to wild type. The RIP analysis can only qualitatively tell whether an intermediate complex is accumulating and can help to identify which intermediate complex accumulates. To give an example, in the case of depletion of Prp16 that we predict to accumulate complex C, association of U2 with U5 and U6 is increased about 2-fold relative to undepleted Figure 3.3). This does not mean that there is a two-fold increase in complex C abundance. In only 1 out of 8 intermediate complexes of the spliceosomes, U2 is not in complex with U5 and U6. Therefore, a hypothetical large increase in the abundance of complex C (U2/U5/U6), relative to other complexes, may only mildly increase association of U2 with U5 and U6.

Unlike the 4-fold reduction in association of U2 with U4 when Prp16 is depleted, the association of U2 with U1 and U4 is not reduced when Prp22 is depleted (Figure 3.3), which is different than what is predicted by the accumulation of a post-spliceosome (Table 3.1). It is believed that after the dissociation of the post-spliceosome, which contains mRNA/U2/U5/U6, the tri snRNP (U5/U6-U4) is reformed. However, I speculate that depletion of Prp22 can result in an abnormal post-spliceosome complex where U4 snRNP is prematurely bound to U6 snRNP to some extent. Alternatively, if Brr2 is sequestered, perhaps Complex B forms and dissociates, giving a normal level of U1, U2, U4, U5 and U6, in addition to post-spliceosome accumulation. Either of these possible cases could explain why by depleting Prp22, I do not

observe a reduced association between Lea1 and U4 snRNA relative to undepleted.

Alternatively, it could be that a variety of spliceosome intermediate complexes accumulate by depleting Prp22, in a substrate-specific manner, in such a way that some reduction in the U2/U4 association cancels off when looking at global averages. It may also be that this strain is already starting to develop a splicing defect (explained in Chapter 4), before auxin induction, which could explain why no change is observed in U4 after auxin addition. However, when comparing the relative level of association of U2 with the other snRNAs (Figure S3.2), before the addition of auxin, I do not observe that in this strain the U4 signal ratio is low enough to explain the lack of reduction in U4, in the depleted condition.

The RIP data of Prp45 depletion may reflect a mild accumulation of Complex B (that contains all snRNPs) or no accumulation at all (Figure 3.3). Cryo-EM structural data has revealed that Prp45 adopts an extended and unstructured form within the spliceosome, where it interacts with many proteins (Yan et al., 2015). Given its extended and embedded nature, it is likely that the depletion of this protein will cause the spliceosome to lose structural stability and fall apart. Considering this, it is possible that Prp45 depletion does not result in the accumulation of a stable and intact intermediate complex, which would explain the RIP result.

Spliceosomes can be stalled *in vitro*

In vitro systems are often used to stall the spliceosome at specific stages to individually analyse, by mass-spectrometry for example, the different intermediate complexes of the spliceosome. This is done by using a combination of strategies including, pre-mRNA substrates with mutated

splice sites, altering the levels of ATP and/or using temperature-sensitive mutant proteins (Fabrizio et al., 2009; Ohrt et al., 2013; Warkocki et al., 2009). However, this does not automatically imply that arrested spliceosomes can also accumulate *in vivo*. For example, it is possible that known or unknown disassembly factors that are not essential for splicing catalysis *per se*, may be missing in *in vitro* extracts, allowing the spliceosome to be stalled *in vitro* but not *in vivo*. In this sense, the observations described in this chapter that intermediate complexes of the spliceosome can accumulate *in vivo* by depleting essential splicing factors suggest that this strategy can be used, as an alternative to *in vitro* systems, to enrich and then purify specific intermediate complexes.

Depletion of Prp4, Prp16, Prp22 or Prp45 inhibit pre-spliceosome formation

The Prp40 (U1) and Lea1 (U2) ChIP assays show that depletion of Prp4, Prp16, Prp22 or Prp45 inhibits co-transcriptional pre-spliceosome formation (Figure 3.4), which suggests that knockdown of essential splicing factors, including late-acting splicing factors, can inhibit early spliceosome assembly. Due to the nature of the ChIP technique, it could be asked whether depletion affects only co-transcriptional assembly, allowing splicing to continue post-transcriptionally. However, this does not seem to be the case as the RT-qPCR analysis of *ACT1* splicing, which does not discriminate between co- or post-transcriptional splicing, shows that depletion of late-acting Prp16 or Prp22 causes a first step splicing defect (Figure 3.5). This implies that ChIP data do not represent an uncoupling of co- and post- transcriptional assembly, but rather a general inhibition of spliceosome assembly.

The arrested spliceosome model: depletion of a splicing factor can cause a recycling defect that feeds back to pre-spliceosome formation

The RIP and ChIP data support the arrested spliceosome model

Together, the RIP and ChIP analyses suggest that, when depleting Prp4, Prp16, Prp22 or Prp45, new pre-spliceosomes are not formed because U1 and/or U2 are sequestered in arrested spliceosomes. In the Prp4 depletion, the RIP data suggests that U1 and U2 snRNP are sequestered in arrested complexes, while the ChIP data show that co-transcriptional recruitment of U1 and U2 snRNP is reduced. With Prp16 or Prp22 depletion, the RIP data suggest that U2, U5 and U6 are sequestered in arrested complexes, while the ChIP data show that recruitment of U2 (but not U1) snRNP is reduced. The good agreement between the RIP and ChIP data argues that the arrested spliceosome model is the most likely explanation for the observations.

Arguments against a direct role of Prp4, Prp16 or Prp22 in pre-spliceosome formation

An alternative explanation for some of the observations, is that Prp4, Prp16 or Prp22 might play a direct role in pre-spliceosome formation. Although this "direct-role" hypothesis is tempting, there are good arguments against it that are well supported by published proteomic and structural data. These published data support the involvement of Prp4 after, but not before, pre-spliceosome formation. Prp4 is a U6 snRNP component (Banroques and Abelson, 1989) that is recruited to the spliceosome, as part of the pre-assembled tri-snRNP particle (Stevens and Abelson, 1999) that joins the spliceosome after binding of U1 and U2 snRNPs to the pre-mRNA to form Complex B. Soon after its recruitment, Prp4 and other proteins are released during the transition of B to Bact (Fabrizio et al., 2009). Furthermore, *In vitro* immunoprecipitation assays demonstrated that Prp4 is required for

recruitment of the tri-snRNP to the pre-mRNA (Banroques and Abelson, 1989). Recently, a CryoEM yeast tri-snRNP model confirmed these previous observations by showing Prp4 extensively interacting with other tri-snRNP proteins such as Snu13, Prp31, Prp3 and Prp6 (Nguyen et al., 2015). Therefore, it seems that Prp4, similar to other proteins that also containing a WD domain, acts as a protein platform to mediate protein-protein interactions that stabilize the structure of tri-snRNP (Ayadi et al., 1998). In addition, it is possibly that, together with other tri-snRNP proteins also released during the B to Bact transition, Prp4 helps to conceal the catalytic cavity of Prp8 before the spliceosome is catalytically activated (Nguyen et al., 2015). As an interaction between U1 components and Prp4 have not been observed, nor the presence of Prp4 has been detected within the pre-spliceosome in proteomic analysis of the pre-spliceosome (Behzadnia et al., 2007), overall there is good evidence supporting a role of Prp4 after, but not before, pre-spliceosome formation. This agrees with the model that depletion of Prp4 inhibits pre-spliceosome formation as an indirect effect caused a recycling defect.

The roles of Prp16 and Prp22 in splicing have been well studied through biochemical and genetic approaches. These studies demonstrated that Prp16 and Prp22 act at late stages of spliceosome assembly, to promote the second catalytic step of splicing and the release of the spliced mRNA, respectively (Burgess and Guthrie, 1993; Cordin and Beggs, 2013; Schwer, 2008; Schwer and Guthrie, 1991; Semlow et al., 2016; Tseng et al., 2011; Wagner et al., 1998). It has been observed *in vitro*, however, that Prp16 can promote the first catalytic step of splicing, but only for transcripts that contain a mutated non-canonical branschsite sequence (Tseng et al., 2011). As virtually all yeast intron-containing genes contain consensus branchsite sequences, Prp16 is not expected to be required directly for the first step of splicing *in vivo*. Furthermore, the transient interaction of Prp16 only with Complex C, and

Prp22 with Complex C* and the post-spliceosome, but not with earlier intermediate complexes, has been confirmed by proteomic analysis (Fabrizio et al., 2009; Lardelli et al., 2010) and recent CryoEM modelling (Fica et al., 2017; Galej et al., 2016). Overall, there is good evidence supporting a role of these proteins after, but not before pre-spliceosome formation or the first catalytic step of splicing, which agrees with the model that depletion of Prp16 or Prp22 inhibit pre-spliceosome formation as an indirect effect caused by a recycling defect.

How to explain the combined RIP and ChIP data of Prp45 depletion

The observation that depletion of Prp45 also causes reduced pre-spliceosome formation (Figure 3.4E) is somewhat surprising, as the corresponding RIP analysis (Figure 3.3) pointed to a mild accumulation of Complex B or no accumulation at all. If Complex B were accumulating to significant levels, ChIP of U1 would be reduced as in Prp4 depletion (Figure 3.4B), but this is not what I observed. On the other hand, the ChIP data by itself could suggest that Prp45 is directly required for the recruitment of the U2 snRNP, one step earlier than anticipated for a NTC-associated protein. However, this does not seem to be the case, as I don't observe reduced association between Lea1 (U2) and all the other snRNAs by RIP.

As discussed with the U2/U4 RIP data of Prp22 depletion, it could be that depletion of Prp45 causes accumulation of a mixture of intermediate complexes in such a way that the Lea1 (U2) RIP analysis will give the appearance of no change – that is, a global mixture of intermediate complexes that individually produce opposite predicted profiles (Table 3.1), such as pre-spliceosome and Complex Bact. This could be the case if Prp45 is required at different stages of spliceosome assembly, in a substrate-specific manner (see

NTC in Introduction). Although, the data are not sufficient to provide a satisfactory explanation for what happens when Prp45 is depleted, it is likely that the loss of U2 snRNP recruitment is caused by a recycling defect, as with the other depletions.

ChIP signal of Prp40 is low before adding auxin, which is why in this case only the U1 and U2 ChIP plots are presented separately. ChIP of Prp40 was first performed on the Prp45 AID-tagged strain and was later improved before analysing the other AID-tagged strains. For this reason, it is likely that low Prp40 signal is due to a suboptimal ChIP assay. However, I cannot rule out that the co-transcriptional recruitment of U1 snRNP is already affected before adding auxin, due to a possible disruption of Prp45's activity by the C-terminal tag. If true, it would be interesting because it might suggest that the C-terminal of Prp45 is important for stable binding of U1 to the pre-mRNA.

Depletion of non-RNA helicase proteins can also lead to a recycling defect

It can be argued that the lack of proper disassembly is due to the absence of RNA helicases (Prp16 or Prp22), that have been shown to be required for the discard and disassembly (proofreading) of spliceosomes with aberrant pre-mRNAs (Koodathingal and Staley, 2013). However, this argument becomes weak with my observation that depletion of Prp4, which is not an RNA helicase, also results in the accumulation of an arrested spliceosome; and that depletion of Prp4 or Prp45, unexpectedly reduces pre-spliceosome assembly (Figure 3.4). Depleting a different tri-snRNP protein, Prp3 (Figure S3.1A), produces a very similar RIP result as depleting Prp4, which suggests any defect that prevents tri-snRNP recruitment may cause accumulation of an arrested pre-spliceosome. Additionally, the proposed role of RNA helicases in proofreading was almost exclusively analysed with aberrant pre-mRNA, not

with the absence of essential splicing factors as I have done. Therefore, it is entirely possible that the splicing proofreading machinery is not prepared to deal with aberrant spliceosomes that lack an essential protein. They can only target normal spliceosomes with aberrant pre-mRNA (*e.g.* a pre-mRNA with non-canonical splice sites). Therefore, the Prp4, Prp45 and Prp3 depletion results suggest that knockdown of proteins that are not RNA helicases, may also lead to a recycling defect.

Arrested spliceosomes leading to a recycling defect has been observed before

Tardiff and Rosbash (2006) showed that metabolic depletion of U5 snRNA leads to the accumulation of arrested pre-spliceosomes and to the reduced co-transcriptional recruitment of U1 and U2 snRNPs. It is worth pointing out that they were more focused on demonstrating the *in vivo* step-wise recruitment of snRNPs than in showing how a global splicing inhibition can lead to a defect in the recycling of spliceosome components. In a way, my depletion of Prp4 (Figure 3.4B) is like Tardiff and Rosbash's U5 snRNA depletion as both approaches will block assembly and recruitment of the tri-snRNP. However, one criticism of their work is that their GAL depletion was left for 16 hours, therefore their observations may be a secondary effect of a prolonged splicing defect. For example, it may be that formation of new pre-spliceosomes was prevented due a reduced expression of splicing proteins that are encoded by intron-containing genes (like Mud1 and Smd2), as depletion of U5 snRNA would stop splicing globally.

The work presented in this thesis confirms Tardiff and Rosbash's observations, and extends it by showing that knockdown of late-acting splicing factors Prp16 or Prp22, or NTC-associated Prp45, also leads to a recycling defect. This means that other complexes, besides the pre-spliceosome, such as complex C and the post-spliceosome, are also susceptible

to a recycling defect, which lead me to reason that a recycling defect is a common result of knocking down splicing factors. Additionally, I showed that spliceosome assembly can be inhibited almost immediately (3-6 minutes) after knocking down a splicing component, which is much earlier than the previous observation of 16 hours of U5 snRNA depletion by Tardiff and Rosbash.

Prp16 and Prp22 are indirectly required for the first step of splicing

The apparent requirement of Prp16 or Prp22 for the first step of splicing is an old observation. I have reproduced this observation with depletion of Prp16 and Prp22 combined with an RT-qPCR analysis of splicing efficiency (Figure 3.5). In 1991, Company and others observed more *ACT1* pre-mRNA in *in vitro* splicing reactions when a temperature-sensitive *prp22* strain was heat-inactivated (Company et al., 1991). To explain their result, they speculated that defective spliceosomes are not recycled into new rounds of splicing, thus accumulating pre-mRNA – just like our hypothesis. A similar observation by the same lab (John Abelson lab), but *in vivo*, was presented earlier but was not properly discussed (Vijayraghavan et al., 1989). After Company and others (1991), the laboratories of Christine Guthrie and John Abelson also observed *in vivo* pre-mRNA accumulation; of *RP51A*, *U3* snRNA and *ACT1* with *prp22* cold-sensitive mutants, as well as with *prp16 cs* mutants (Noble and Guthrie, 1996); and of *ACT1* with heat-sensitive mutant *prp22-1* (Wagner et al., 1998). To the best of my knowledge, evidence that supports the hypothesis that blocking splicing can produce a feed-back effect due to a recycling defect was not produced until 2006 (Tardiff and Rosbash, 2006). This issue is now resolved here with evidence that depletion of Prp16 and Prp22 leads to a recycling defect, which explains why the first step of splicing is inhibited.

A recycling defect appears quickly after depleting Prp22

The U2 ChIP and 4tU kinetic analysis showed that the co-transcriptional recruitment of U2 snRNP and the first step of splicing are inhibited almost immediately after Prp22 depletion (Figure 3.6), which suggests that blocking spliceosome disassembly can quickly cause a recycling defect. This was surprising, as our expectation was that the recycling defect would appear late after the arrested complexes accumulate gradually over time. To the best of my knowledge, it is still not accurately known what proportion of the splicing machinery is in its free form compared to how much is within the spliceosome. So, this observation is interesting because it could suggest that the pool of free spliceosome components is small and that recycling is a rate-limiting step for splicing.

Another observation is that the 4tU and ChIP data of RPGs do not entirely agree between them, in contrast with the non-RPGs *ACT1* and *ECM33*. U2 ChIP signal is dramatically reduced (from minute 3) on the RPGs, *RPL28*, *RPS13* and *RPL39* (Figure 3.7D-F); even though splicing catalysis of these genes appear to be only mildly affected when looking at the relative change in 5'ss and 3'ss signal (Figure 3.6C-E). I speculate on two possible explanations for this. It could be that pre-mRNA degradation is relatively higher for RPGs, so these pre-mRNAs do not accumulate. In support of this, it has been shown that the abundance of pre-mRNA of RPGs is reduced in a variety of stress conditions, for example heat shock, glucose depletion or osmotic stress; probably due to stress-response mechanisms that down regulate their expression (Pleiss et al., 2007). Thus, it may be that the splicing defect is also triggering a stress response that contributes to less production of pre-mRNA of RPGs, through the down regulation of transcription or increased pre-mRNA degradation. Alternatively, it could be that the co-transcriptional splicing of RPGs is more affected than their post-transcriptional splicing. In

this way, the differences between the two experiments would be due to the different nature of the techniques. 4tU RT-qPCR measures total splicing catalysis and ChIP only measures co-transcriptional spliceosome assembly.

The data agree with the "hungry spliceosome model"

Interestingly, the kinetic results are compatible with the "hungry spliceosome model", which says that pre-mRNAs compete for a limited pool of spliceosome components (Munding et al., 2013). This proposal is mainly based on evidence that blocking transcription of RPGs leads to an increase in splicing efficiency of other transcripts that are otherwise poorly spliced. So, if pre-mRNAs are in excess and spliceosome assembly is a rate-limiting step for splicing, as their data suggest, it would predict that inhibiting spliceosome recycling should immediately affect new rounds of splicing, which is what I observe. One novel implication of this model, as pointed out by the authors, is that availability of spliceosome components relative to pre-mRNA abundance is important for gene expression regulation. In support for this idea, an mRNA-seq and mass spectrometry analysis of murine bone marrow cells discovered that for some genes, intron retention correlated with downregulation of core splicing factors, suggesting that this novel aspect of splicing regulation might exist in higher eukaryotes too (Wong et al., 2013).

Can some observations be explained by an uncontrolled depletion?

After the experiments reported in this chapter were completed, it was discovered that the Prp22 AID-tagged (YGM1 background) strain I used has leaky depletion. By leaky depletion I mean that Prp22 is depleted to a certain extent, probably to 60% of wt levels, before the addition of auxin. This is because the YGM1 background strain constitutively expresses OsTIR1 to very high levels, which results in very fast but uncontrolled depletion (explained

in more detail in Chapter 4). The main potential problem when starting off with less target protein, is that the changes measured may be underestimated, as almost all comparisons were done against the no auxin culture of the same strain. However, in this case it appears that this is not a problem as, in almost all my data, I observe a significant change when comparing induced with un-induced cultures. In addition, Prp22 is an essential protein and no growth defect is observed in this AID strain without auxin (Chapter 3). However, I do observe a small splicing defect with no auxin. There is about 50% more lariat of *ACT1* in the Prp22 strain compared to the parental untagged strain (Figure S4.2), which suggests that the phenotype after auxin-induction is greater than it appears when compared to no auxin, and that Prp22 is reduced to the minimum required for normal growth prior to auxin addition. Therefore, I conclude that the leakiness effect does not change the interpretation of the results or the conclusions.

In relation to the literature

An example where the arrested spliceosome model can provide an alternative interpretation to published data

It was reported that depletion of NTC components causes accumulation of the free form of U4 snRNA, and concluded that this complex has an indirect role in U4/U6 snRNP biogenesis (Chen et al., 2006). As they also observe a decrease in total levels of U6 snRNA, they propose that U6 becomes unstable and is therefore discarded when spliceosome assembly does not progress in the absence of NTC. However, I suggest that, considering the arrested spliceosome model, there could an alternative interpretation. If absence of the NTC results in the accumulation of an aberrant intermediate complex that sequesters U6, perhaps bound to the U2 snRNA, U4/U6 snRNP recycling might also be reduced – a possibility that was not discussed in the article.

Assuming proteins are important for the stability of the U2/U6 duplex, it is likely that this possible increase in U2/U6 association was missed in their results since their total RNA extraction method involved breaking cells with phenol/chloroform that would have taken apart RNA-protein complexes. So, ruling out this possibility would require a native pulldown of U2 snRNP to know if its association with U6 snRNA is increased in the depletion.

Is this model valid for higher eukaryotes?

Is there at least some indication that the arrested spliceosome model is important for higher eukaryotes as it is for budding yeast? The answer is not obvious because it is not known whether higher eukaryotes possess more sophisticated quality control mechanisms that will target and disassemble aberrant complexes. Also, it is still not known how critical is spliceosome recycling in these organisms, it could be that spliceosome components are not limiting, as has been proposed for yeast (Munding et al., 2013). If spliceosome components are in excess relative to pre-mRNA abundance, it could be that a defect in recycling might not have a big impact on new rounds of splicing. To try to answer the question, within the scarce literature on the homologues of Prp16 and Prp22, I investigated if accumulation of pre-mRNA in the knockdown of these late-acting splicing factors has been observed.

Interestingly, I found a report where a homozygous lethal mutation of the zebrafish orthologue of Prp22 (called Dhx8) was analysed (English et al., 2012). Based on a microarray analysis, they selected 19 genes that are less expressed in the mutant and involved in hematopoiesis that is the focus of the article. Then they measured splicing of these genes using RT-PCR probes flanking the intron. Interestingly, in 13 out of 19 genes they observe more unspliced pre-mRNA compared to wild type. If the zebrafish Dhx8 has the

same function as the yeast Prp22 (release of spliced mRNA from the post-spliceosome after splicing completion), this suggests that absence of Dhx8 may be causing a feedback on the earliest stages of splicing; which means that the arrested spliceosome model is possibly also valid in higher eukaryotes. However, I do not know how many genes are less expressed in the *dhx8* mutant, as the microarray data web link found in the supplementary material was outdated. If only a small proportion of transcripts are less expressed relative to wild type, it could be that the mutation is not highly penetrant and only a few genes are particularly sensitive to defects in spliceosome recycling.

Is there agreement between the model and data of RNAi-knockdown of splicing factors?

There is a recent report of particular interest where all core and regulatory components of the spliceosome were systematically knocked down with RNAi in HeLa cells and the 35 alternative splicing (AS) events were measured with high-throughput capillary electrophoresis (Papasaikas et al., 2015). Each knockout condition was profiled based on the magnitude and direction of the changes in AS. Surprisingly, they observe that knockdown of a significant fraction of core splicing factors, produces changes in AS, instead of generally reducing splicing efficiency as one would anticipate. Next, by modelling the AS profile data they generated a network based on similarities between all possible pairs of knockdown conditions.

Within the network, a big and dense cluster can be observed that contains most core splicing factors and most of the functional associations of the whole network. In contrast, splicing regulators such as SR and hnRNP proteins are located at the periphery of the network with few functional associations with other individual factors. Not surprisingly, this means that knockdown of core splicing factors tends to have more similar effects between them than

knockdown of regulatory splicing proteins, which produce distinct effects. If it is true that knocking down essential splicing factors often results in a feed-back effect on the earliest stages of assembly as I propose, it is possible that this phenomenon is contributing to the high degree of similarities that the authors observe among core splicing factors.

On the other hand, the big and dense cluster is not entirely homogeneous. Within this cluster there are sub-clusters, which tend to contain splicing factors that physically interact or share functions, thus they estimate that about half of all associations can be explained by previous knowledge. This means that, despite the overall similarities, knocking-down different core splicing factors can produce different effects on AS. Based on this surprising observation, the authors propose that core splicing factors have regulatory potential. However, at least a fraction of such differences may be due to differences in the efficiency of the RNAi-knockdown, as it is known that RNAi is generally inefficient and no evidence was presented of knockdown efficiency in this report. Colleagues and I have some evidence to suggest that a recycling defect, of the type I have observed, requires a complete knockdown of the splicing factor target. Therefore, it could be that if the RNAi-knockdown were near 100% efficient, then the differences they observed would be smaller.

A recycling defect may require a complete knockdown

When Isabella Maudlin and Ema Sani, members of this lab (Beggs lab), analysed the splicing defect of the depletion of Prp22 AID-tagged in a different background strain (YBRT1 instead of YGM1) by RT-qPCR, they were surprised to observe what appeared to be a disassembly defect but not a first step defect with *ACT1*. The data are not shown here because they do not belong to me. However, I've taken the liberty to discuss it because it is related

to my work and it addresses an important point. At first, the result appeared to disagree with the arrested spliceosome model. If disassembly is blocked then why is there no recycling defect that inhibits the first step of splicing? The simplest explanation for this is that the YBRT1 strain, although is not leaky does not deplete as efficiently as YGM1 (see Chapter 3). After 30 minutes of auxin-induction, Prp22 (YBRT1) was only depleted down to 40% of normal levels (data of Isabella Moudlin). Therefore, it seems that splicing is not globally affected enough to cause a recycling defect. Satisfyingly, when auxin induction was left for more than 1 hour, a first step defect was now observed in this strain too. Overall, this suggests that only a strong splicing defect will cause a recycling defect. Perhaps that is one reason why disease-related mutations of core splicing factors do not affect splicing systematically, since generally in these cases the splicing defect is not highly penetrant.

Concluding remarks and perspectives

The results described here send a reminder that perturbing one element of a dynamic complex, such as the spliceosome, often cause unwanted systematic effects on the assembly cycle. As evidence for this, I have shown that the biggest effects seen with my depletions are the consequence of secondary effects that are not directly related to the precise function of the depleted factors. In addition, I have shown that the secondary effect appears almost immediately after depletion, suggesting that it is technically difficult to completely separate it from the direct effect. The take-home message is, to avoid misinterpreting data derived from perturbing spliceosome components, one should think about recycling inhibition as a possible source of secondary effects.

The Prp22 depletion kinetic analyses, supports the idea that spliceosome components are limiting and suggests that recycling is a rate-limiting step for

splicing. Together with the "hungry spliceosome model" (Munding et al., 2013), this supports the need to better understand the relationship between the availability of spliceosome components and alternative splicing. One way to start answering this question would be to reduce the expression level of recycling factors and see if this correlates with changes in alternative splicing. If true, this would suggest that for the cell it is important to maintain an adequate balance between the abundance of the spliceosome components and the abundance of pre-mRNA substrates, as changes in these ratios could potentially perturb gene expression.

Acknowledgements

I would first like to thank Jane Reid for her help in my initial lab training and her contribution to the planning of my experiments and discussion of my results. I must also thank David Barrass and Ema Sani for teaching me the 4tU labelling and ChIP techniques and their feedback on my results. I also thank former honours student Barbara Terlow for sharing her Os*TIR1*-expressing strain YBRT1 and her AID-tagged *PRP16* (in YGM1), and PhD student Isabella Maudlin for sharing her AID-tagged *PRP4* (in YBRT1). All people mentioned above are former or current members of the Beggs lab.

Chapter 4: An improved AID system for budding yeast

Abstract

In the AID system, fusion of the gene of interest with an AID tag, allows the plant Tir1 protein to target this fusion protein for degradation when auxin is added to the medium. I observe that high expression levels of Tir1 lead to a fast depletion of the target protein, but also to an uncontrolled depletion without auxin addition, suggesting that an optimal expression level of Tir1 is critical. To further develop this technique, I created a new version of the AID system where Tir1 expression is rapidly induced by B-estradiol prior to auxin addition. I show that this gratuitous B-estradiol AID provides a fast depletion of the target protein, with low uninduced depletion. An AID system with these characteristics for budding yeast was lacking until now.

Chapter Introduction

Probably the most widely used strategy to study the functions of essential genes *in vivo*, is to analyze the effect of blocking their expression and/or activity with conditional systems (reviewed in Chapter 1). Ideally, the conditional system should be fast and specific, to avoid secondary and off-target effects. The traditional conditional systems for budding yeast are temperature-sensitive mutants and metabolic regulation (*e.g. GAL* and *MET* promoters). These tools have been very useful in functional studies. However, they have disadvantages, including: 1) changes in temperature or growth medium can cause other, unwanted effects on cell metabolism, for example on transcription and splicing (Bergkessel et al., 2011; Kresnowati et al., 2006; Ronen and Botstein, 2006); and 2) metabolic regulation is often slow-acting, as

both the mRNA transcript and the encoded protein must be turned-over before the phenotype is manifested.

The auxin-inducible degron (AID) is a recent technique that has become increasingly popular mainly because it allows fast depletion of the target protein, the inducer is a small molecule auxin that does not perturb the cells, and it is easy to implement. Setting up the AID system involves 1) the heterologous expression of Tir1, a plant auxin-binding receptor that is part of the conserved E3 ubiquitin ligase complex and 2) fusion of the gene of interest with an AID tag (plant derived) that allows Tir1 to direct this protein for polyubiquitination that targets it for degradation by the proteasome (Nishimura et al., 2009). The AID system is described in more detail in Chapter 1.

In the original AID system for budding yeast, *Oryza sativa* (rice) *TIR1* (Os*TIR1*) expression was driven by a GAL promoter. To avoid perturbations to cell metabolisms associated with the GAL system, we initially used other strains where Os*TIR1* is under control of constitutive promoters of different strength, with mixed levels of success. I observed that depletion rate, and also uncontrolled depletion of the target protein in the absence of auxin, correlates with a high expression level of OsTir1, which agrees with previous observations in *Arabidopsis thaliana* (Gray et al., 1999; Dos Santos Maraschin et al., 2009). This indicates that constitutive expression of *TIR1* in yeast is not ideal either.

Using the novel B-estradiol expression system where transcription of a gene of interest is induced in a highly specific manner with no off-target effects, I generated a new version of the AID system for budding yeast, called B-estradiol AID, where expression of OsTir1 is induced shortly before adding

auxin. I show that this system does not deplete the target in the absence of auxin and can still quickly deplete the target protein after less than 30 minutes of auxin addition. An abundant target protein was also quickly depleted, which suggests that target abundance is not an obstacle. Furthermore, I constructed a centromere plasmid that contains all the elements for the B-estradiol-mediated expression of OsTir1, and which can be introduced into a variety of budding yeast strains with a one-step transformation.

Results

YGM1 strain allows fast AID depletion

Selecting the AID tools to use

My first aim in the lab was to generate strains for auxin-induced depletion of the proteins of interest. Instead of the original AID tools for yeast, GAL-OsTir1 and IAA17 (as the AID tag) (Nishimura et al., 2009), I used newer ones that we believed would work better, including: 1) a codon optimized Os*TIR1* for improved yeast expression, controlled by a constitutive promoter, to avoid the problems associated with the GAL expression system and 2) an AID*-6FLAG to easily detect the target using a commercially available antibody, and to avoid a big C-terminal tag that might affect the activity of the target protein. The AID* ($IAA17_{71\text{-}114}$) (Morawska and Ulrich, 2013) is a small derivative of the original $IAA17_{1\text{-}229}$.

Generating the YGM1 strain and testing its depletion efficiency

To generate the OsTir1-expressing strain, I integrated a StuI-digested pMK200 (Kanemaki lab) plasmid into the *ura3-1* locus of strain W303. pMK200 contains a yeast codon-optimized Os*TIR1* under the control of a strong 689 bp *ADH1* constitutive promoter (P*ADH1-689*) (Figure S4.1). This yeast strain was called YGM1.

I C-terminally tagged splicing factors Prp45, Prp46, Prp22 and Syf1 with pHyg-AID*-6FLAG on the genome of YGM1. These proteins were selected because of the initial interest in studying two candidates coupling factors of splicing with transcription, Prp45 and Syf1, and the interaction partners of Prp45: Prp46 and Prp22. Afterwards, I measured their growth rate, without and with addition of IAA (auxin) at 0.75 mM. This allowed me to confirm that OsTir1 expression, auxin and the AID tags by themselves do not inhibit growth (Figure 4.1). The exception is the AID-tagged Syf1 strain that grew slower than wild-type in the absence of auxin. Also, I observed that auxin addition reduces growth rate of all four AID tagged strains, which was the first indication that this depletion system works, as all target proteins are essential for growth.

To test depletion efficiency, I measured the levels of the target proteins before and after auxin addition. First, I performed a time course western blot analysis of AID-tagged Prp46 and Prp22. With this, I observed that target protein levels dropped to around 5% within 10 minutes after auxin addition. Afterwards, I observed that AID-tagged Prp45 and Syf1 are also efficiently depleted, reaching 5% of starting levels by 30 minutes of auxin addition. The AID-tagged Syf1 strain was not further analyzed due to its growth deficiency. The results indicate that the YGM1 strain allows fast depletion.

Through a time-course RT-qPCR analysis of splicing intermediates of *ACT1* transcripts, I measured changes in splicing efficiency when Prp45, Prp46 and Prp22 were depleted (Figure 4.2). With this I observed that depletion of Prp45 results in more branch site and 3'splice site, and less lariat, mRNA and exon2 signal of *ACT1* transcripts. This is indicative of a defect in the first step of splicing, as expected considering that Prp45 joins the spliceosome before the first step of splicing (Fabrizio et al., 2009; Ohi et al., 2002). Depletion of Prp46 shows a similar result as Prp45 depletion, except that the splicing defect seems milder. Depletion of Prp22 results in more branch site, 3'splice site and lariat, and less mRNA and exon of *ACT1* transcripts, which may suggest that absence of Prp22 reduced the efficiency of both the first and the second step of splicing (Chapter 3). Accumulation of splicing intermediates appears to stabilize between 20 and 30 minutes after auxin addition. The results confirm that auxin-induced depletion of splicing factors Prp45, Prp46 and Prp22, quickly leads to a splicing defect. With two additional experiments, I attempted to verify that the strains do not have a splicing defect without the addition of auxin; however, because the data generated cannot be directly compared to that of a mutant strain, I could not reach a conclusion (Figure S4.2).

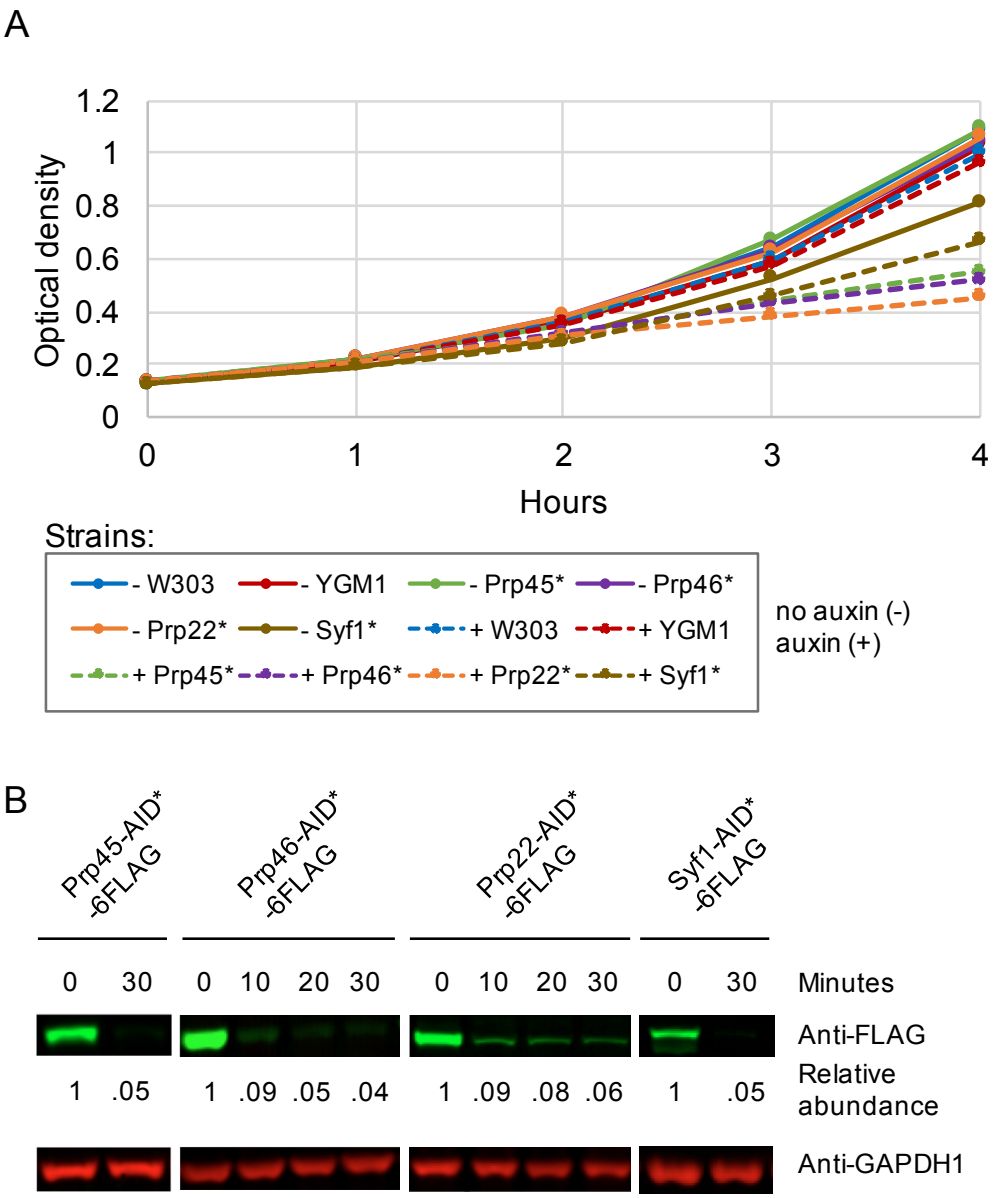


Figure 4.1. Prp45, Prp46, Prp22 or Syf1 were depleted to low levels, using the auxin-inducible degron system.
(A) Growth rate analysis of the AID-tagged strains, including the untagged parental strain YGM1 and wild-type W303, with (+) or without (-) auxin. * = AID*-6FLAG tag.
(B) Abundance of AID*-6FLAG tagged proteins was measured at 0, 10, 20 or 30 minutes after auxin addition, through anti-FLAG LICOR-based western blotting. Relative abundance was calculated as the ratio of + auxin/- auxin anti-FLAG signal, normalized to anti-Gapdh signal as a loading control.

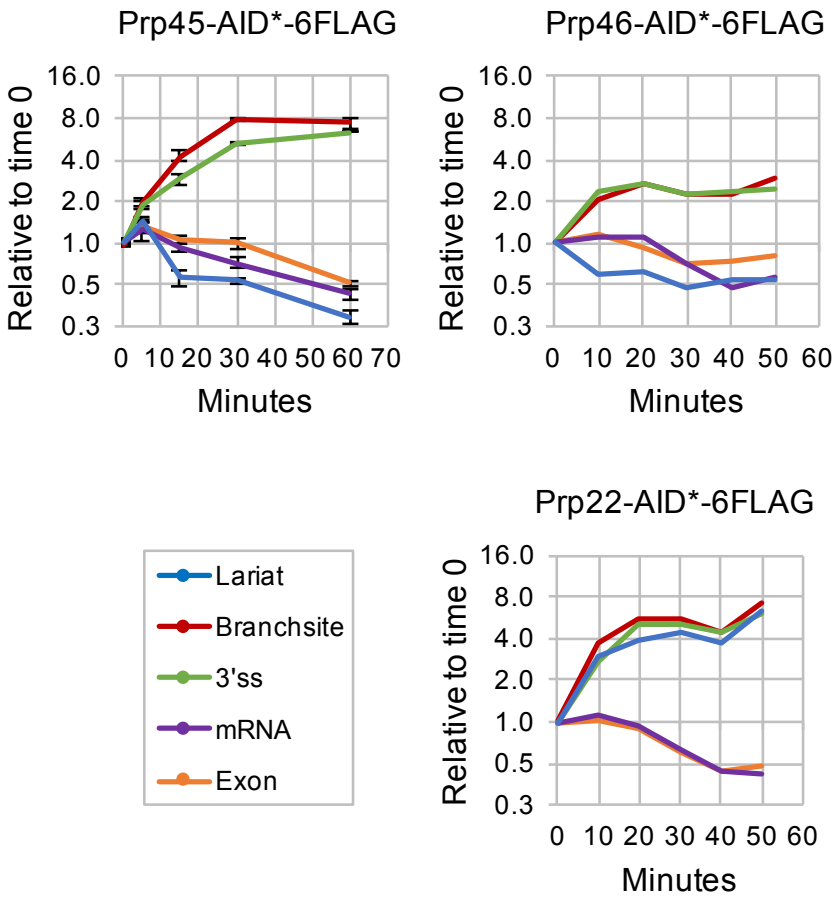


Figure 4.2. Splicing efficiency analysis by RT-qPCR of *ACT1*, after auxin-induced depletion of Prp45, Prp46 and Prp22. Time course analysis of auxin addition. Data were normalized to *ALG9*. Error bars for the Prp45-AID*6FLAG strain denote standard error of two biological replicates. Data of only one trial is presented for Prp46-AID*6FLAG and Prp22-AID*6FLAG strains.

YGM1 has low success rate

Many additional genes were later AID-tagged in YGM1 by Barbara Terlow and myself. Only 11 out of 21 of these strains did not grow on auxin-containing media, even though all tagged genes are essential for growth (Table 4.1). In contrast, all genes AID-tagged in a different background strain created by Barbara called YBRT1, where OsTir1 is expressed by a weak 397 bp *ADH1* promoter ($P_{ADH1\text{-}397}$), resulted in strains that responded to auxin but the products were depleted less efficiently. To explain the behaviour of the AID-

tagged strains, I speculated that the high strength of P*ADH1-689* controlling OsTir1 expression in YGM1 may be lethal when combined with the AID-tagging of some genes, which may force cells to adapt by inactivating the AID pathway.

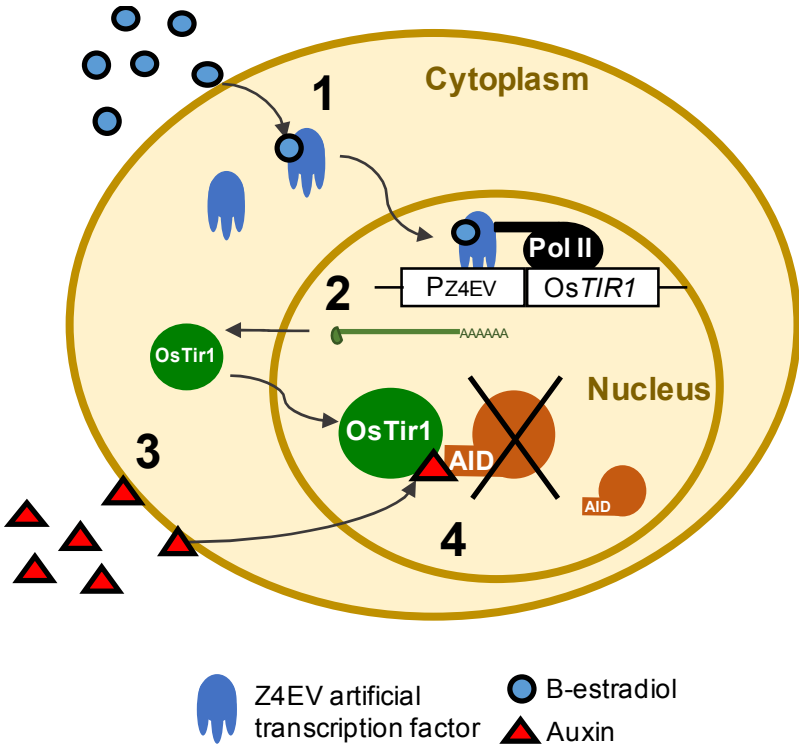
The B-estradiol AID allows fast and tightly-controlled depletion

To investigate the relationship between Tir1 expression level and a possible uncontrolled depletion of the target protein, and at the same improve the current AID system for yeast, I constructed a new strain (YZTR41), where OsTir1 is only expressed when adding B-estradiol to the medium (Figure 4.3). This variation of the AID system was named B-estradiol AID. The B-estradiol expression system (McIsaac et al., 2013), is composed of an artificial promoter (P*Z4EV*) that is specifically activated by its Z4EV artificial transcription factor (ATF) bound to B-estradiol. It should be noted that I added an N-terminal nuclear localization signal (NLS) to this Os*TIR1* gene, although it is not expected it to be important, as it was previously shown that the addition of an NLS to Os*TIR1* does not improve depletion efficiency (Tanaka et al., 2015).

Table 4.1. AID*-6FLAG tagged strains derived from YGM1

	Tagged gene	Growth *	Complex **	Helicase	Response to auxin***	Source ****
1	*PRP45*	✔	NTC	No	✔	GM
2	*PRP46*	✔	NTC	No	✔	GM
3	*PRP22*	✔	Complex C	Yes	✔	GM
4	*SYF1*	✗	NTC	No	✔	GM
5	*PRP5*	✗	Pre-spliceosome	Yes	✔	BT
6	*PRP16*	✔	Complex C	Yes	✔	BT
7	*PRP28*	✔	Comple B	Yes	✔	BT
8	*PRP3*	✗	Tri-snRNP	No	✔	BT
9	*PRP17*	✔	Complex C	No	✔	BT
10	*PRP39*	?	U1	No	✗	BT
11	*PRP9*	?	U2	No	✗	BT
12	*RRP6*	?	Exosome	No	✗	BT
13	*RRP44*	?	Exosome	No	✗	BT
14	*UPF1*	?	NMD	No	✗	BT
	PRP3	✗	Tri-snRNP	No	✔	GM
	PRP39	?	U1	No	✗	GM
15	*PRP2*	?	Complex Bact	Yes	✔	GM
16	*YHC1*	?	U1	No	✗	GM
17	*PRP4*	?	Tri-snRNP	No	✗	GM
18	*PRP6*	?	Tri-snRNP	No	✗	GM
19	*PRP40*	?	U1	No	✗	GM
20	*KIN28*	?		No	✗	GM
21	*PRP19*	?	NTC	No	✗	GM

* Some grew normally (✔), other did not (✗). Growth rate was not analyzed on strains that did not respond to auxin (?)
** Some are part of the nine teen complex (NTC) or U1, U2 or tri- snRNPs, others are RNA helicases recruited to the pre-spliceosome, complex B, Bact or C
*** Growth was inhibited (✔) or unaffected (✗) after auxin addition
**** Strains were made by Barbara Terlow (BT) or myself (GM)

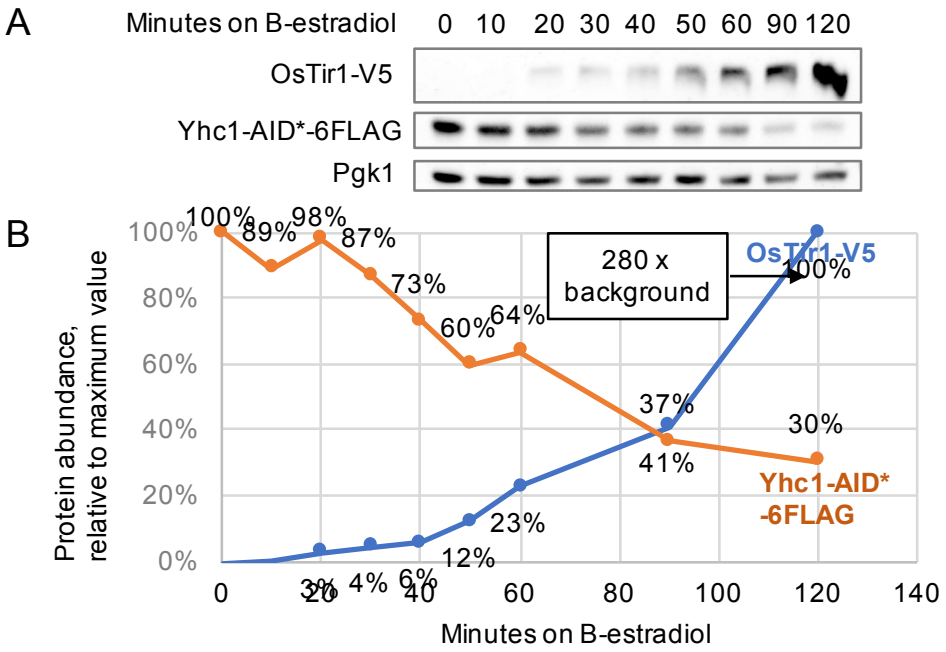


B-estradiol AID strain YZTR41: *ura3*Δ, *leu2*Δ0::P*ACT1*-Z4EV-NatMX, *ape2*Δ::KanMX- PZ4EV-Os*TIR1*-V5, GOI-AID*-6FLAG-HygMX

Figure 4.3. Illustration of the principle behind the B-estradiol auxin-inducible degron (B-estradiol AID). Key to this system, is that rice Tir1 (OsTir1) is only expressed when B-estradiol (McIsaac et al., 2013) is added to the media. Therefore, depletion of the AID target requires the addition of two inducers, B-estradiol and auxin.
(1) The presence of B-estradiol allows the Z4EV artificial transcription factor to be translocated to the nucleus, where it specifically binds the Z4EV artificial promoter (PZ4EV). **(2)** This activates transcription of the downstream gene Os*TIR1*, a component of the E3 ubiquitin ligase complex (SCR). **(3)** When auxin is present, it binds OsTir1 and this allows the polyubiquitination and **(4)** subsequent degradation of the AID-fused gene of interest (GOI).

OsTIR1 expressed to high levels causes uncontrolled depletion

To test the system, first I AID-tagged Yhc1 (U1) splicing factor in the B-estradiol AID strain (YZTR41). Then I performed a western blot time course after B-estradiol addition (Figure 4.4) without auxin. OsTir1 is first detected 20 minutes after B-estradiol addition and its signal continuously increases, reaching levels of 280 times above background at 120 minutes, which demonstrates that the product is quickly and strongly expressed after induction. As predicted, AID-tagged Yhc1 levels inversely correlated with OsTir1 levels. Ych1 signal drops to 30% of its initial value 120 minutes after B-estradiol addition. Interestingly, I do not observe Yhc1 levels significantly reduced (below 90%) during the first 20 minutes, the time during which OsTir1 is not detected. This demonstrates that, even in the absence of auxin, the AID target is depleted when Tir1 is expressed to high levels.



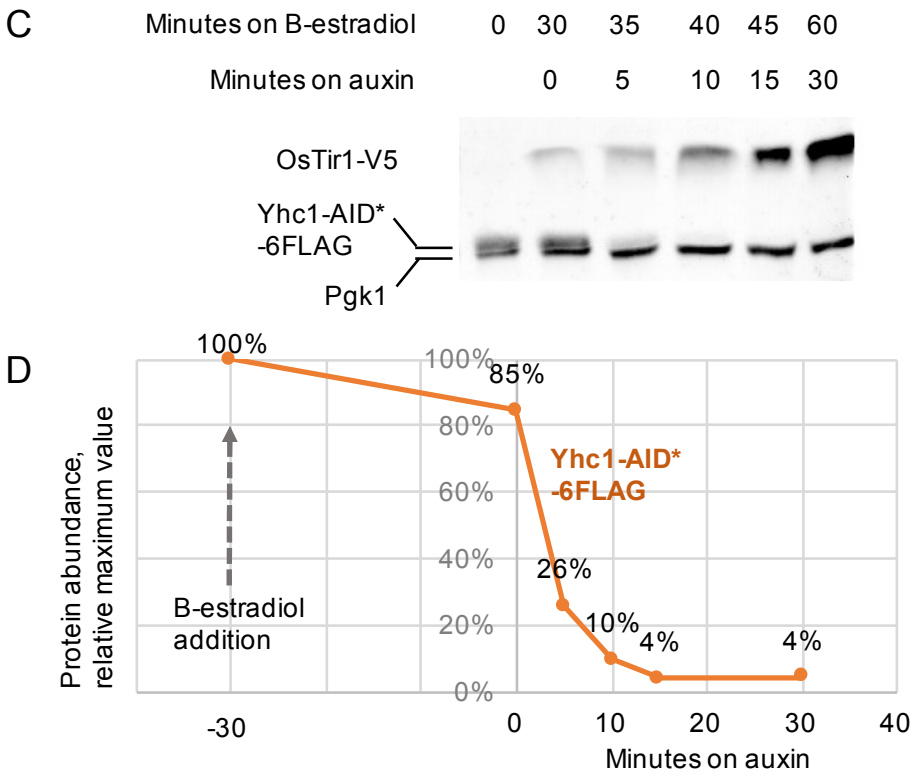


Figure 4.4. When Tir1 is expressed to high levels, the AID-tagged protein is depleted even in the absence of auxin.
(A) A time course western blot analysis of B-estradiol induction of OsTIR1-V5. OsTir1-V5 and Yhc1-AID*-6FLAG were probed with anti-V5 and anti-FLAG antibodies using LICOR system. Protein signal was normalized against anti-Pgk1.
(B) Quantitation of (A). Data are presented as relative to the highest value.
(C) A time course western blot analysis where auxin was added to the cultures 30 minutes after addition of B-estradiol.
(D) Quantitation of (C). Data are presented as relative to the highest value.

Adding auxin at the right time allows a fast and controlled depletion

To test the auxin-induced depletion efficiency of the system, I performed an additional time course analysis where auxin was added after B-estradiol. To achieve the fastest depletion possible, the idea is to give enough time for OsTIR1 levels to build up and then add auxin, before the target protein level starts to drop. Based on the previous results (Figure 4.5), I decided to add auxin 30 minutes after B-estradiol. After just 10 minutes, Yhc1 reached 10% of

initial level (time 0 of B-estradiol addition). However, at the time auxin was added, Yhc1 was at 85% of initial level, which means that auxin should be added slightly sooner, ideally before target levels drop below 90%.

Afterwards, two additional proteins were AID-tagged, Prp22 and exosome component Rrp44. As with AID-tagged Yhc1, two sequential time course analyses were performed, the first with B-estradiol-only (to optimize timing) and the second with addition of auxin after B-estradiol (to test depletion efficiency). Based on the results of the first time-course (Figure 4.5), I decided to add auxin 20 and 40 minutes after B-estradiol to Prp22 and Rrp44 AID-tagged cultures, respectively. At the time of auxin addition, Prp22 and Rrp44 levels are 95% and 88% of initial values respectively, almost within the acceptable range. After 15 minutes of auxin addition, Prp22 and Rrp44 levels quickly dropped to 11% and 4% respectively. Together, the results demonstrate that by optimizing the time of auxin addition relative to B-estradiol addition, one can achieve a fast and controlled depletion of the target protein.

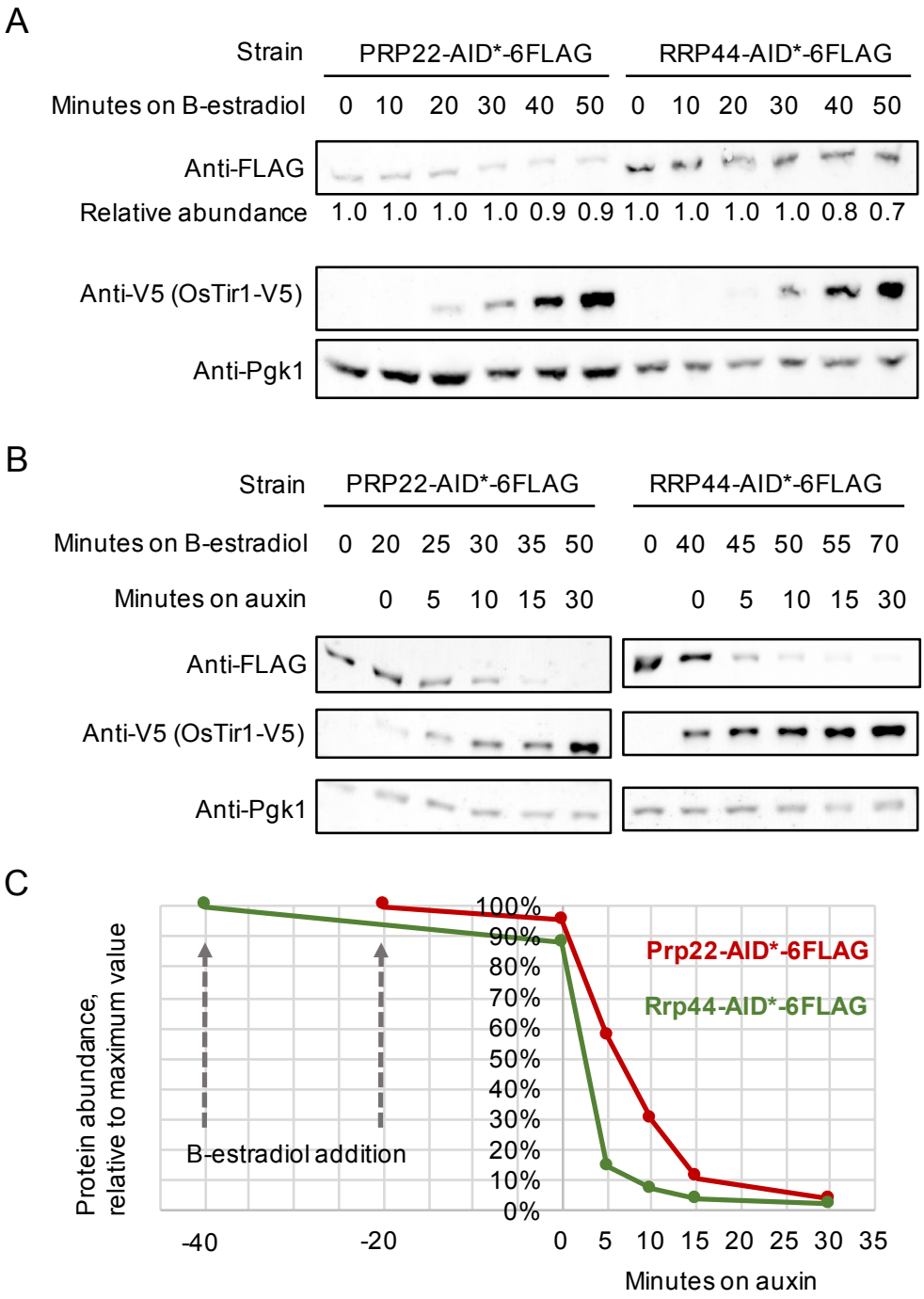
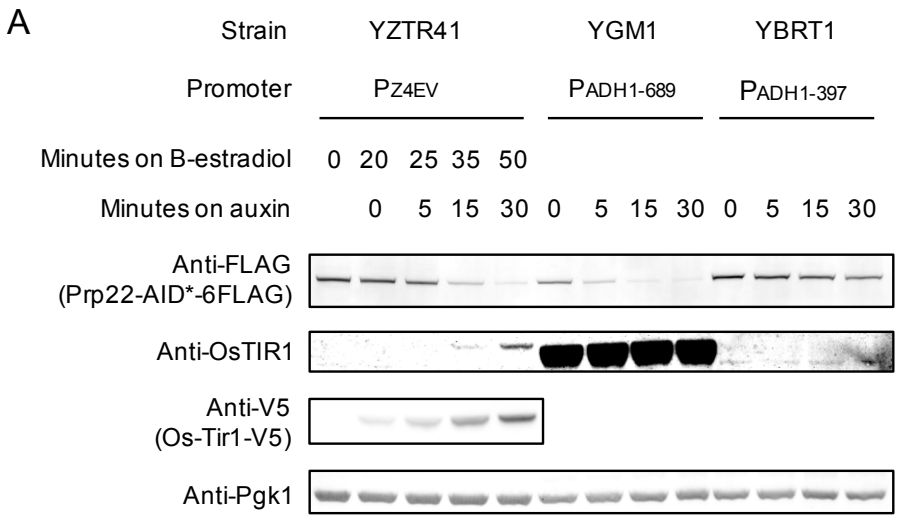


Figure 4.5. Prp22 and Rrp44 are quickly depleted by the B-estradiol AID.
(A) Western blot analysis on Prp22 and Rrp44 AID*-6FLAG tagged Z4TR41 strains after B-estradiol addition. Data are presented as anti-FLAG signal relative to time 0 and Pgk1-normalized (relative abundance).
(B) As in (A) but auxin addition 20 and 40 minutes after B-estradiol addition, on cultures of the Prp22-AID*-6FLAG and Rrp44-AID*-6FLAG strains, respectively.
(C) Quantitation of (B). Data are presented as Pgk1-normalized and relative to the maximum value.

A quick depletion requires high expression levels of OsTIR1

Finally, I compared the B-estradiol AID strain YZTR41, with the previous AID strains than constitutively express OsTir1, YGM1 and YBRT, through a time course depletion of AID-tagged Prp22 (Figure 4.7). The first observation to note is that OsTir1 is dramatically more abundant in YGM1 than in the other strains; it was not even detected in YBRT1. Secondly, without auxin Prp22 abundance is at similar levels in YBRT1 and YZTR41 but about 40% lower in YGM1. This shows that an uncontrolled depletion is also observed on AID strains that grow and splice well, like this AID-tagged Prp22 YGM1 strain. Thirdly, by 30 minutes of auxin addition, Prp22 levels dropped to 4% and 11% in YGM1 and YZTR41, but only to 75% in YBRT1, relative to normal levels. Together, these observations suggest that a quick depletion requires high levels of Tir1, and that a controlled and efficient depletion of the target protein is difficult to achieve through a constitutive expression of Tir1.



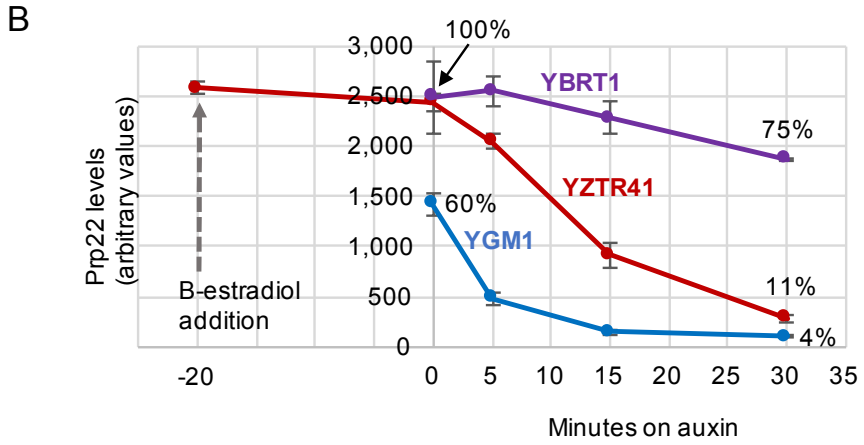


Figure 4.7. B-estradiol AID (YZTR41 strain) allows fast and controlled depletion of the target protein. Depletion efficiency and starting values of Prp22-AID*-6FLAG were compared between B-est AID strain YZTR41 (P*Z4EV*-Os*TIR1*), YGM1 (P*ADH1*-689-Os*TIR1*) and YBRT1 (P*ADH1*-397-Os*TIR1*).
(A) A time course western blot analysis of auxin induction. B-estradiol was added 20 minutes before auxin on cultures of the B-est AID strain, YZTR41.
(B) Quantitation of (A). Data are presented as raw values/1000. Error bars denote standard error of two biological replicates.

A new plasmid containing the B-estradiol TIR1 components

The current B-estradiol AID strain has two inconveniences. The first is that transferring all components to a different background strain would require at least two transformation steps of homologous recombination, which is inefficient. The second is that it may not be possible to make further modifications to the current strain, as most selection markers are already used, only ura and leu are available. To solve this, I cloned the P*ACT1*-Z4EV ATF and PZ4EV-Os*TIR1*-V5 constructs, into a plasmid containing a centromere self-replicating origin (CEN6/ARS) and a *LEU2* marker (Figure 4.6). This plasmid will allow the quick incorporation of the B-estradiol AID components into almost any budding yeast strain. Unlike YZTR41, in this plasmid Os*TIR1* lacks NLS at its N-terminus (see above). Due to lack of time, I could not test this plasmid. Fortunately, Isabella Maudlin (Beggs lab) has successfully used it to deplete some of her proteins of interest. In addition, she is planning to make

new versions of it with different markers, increasing even further the versatility of the B-estradiol AID system.

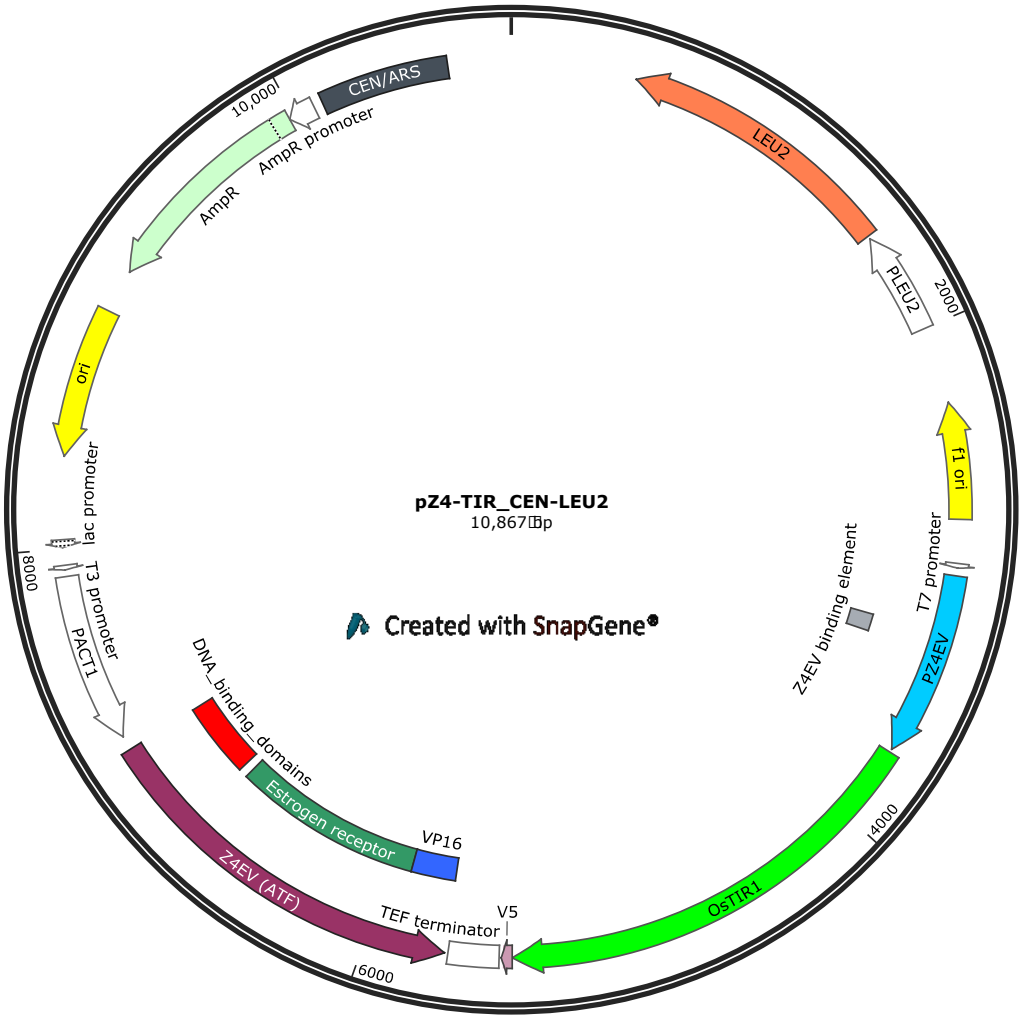


Figure 4.6. Map of plasmid pZ4-TIR1-CEN-LEU. It can be used for the B-estradiol induced expression of OsTir1 – in a similar way as the previous B-estradiol AID strain YZTR41 (Figure 4.3).

High auxin concentrations inhibit growth on minimal media

Members of the Tollervey lab observed that auxin (IAA) at 750 uM, the concentration that we routinely used, inhibits growth on CSM recipe of yeast minimal media (YMM). Thus, before performing experiments that required growing cells on YMM, I attempted to reproduce their result and to titrate auxin down to non-detrimental levels.

With a spotting assay, I evaluated the growth of W303, YGM1, PRP22-AID*-6FLAG (YGM1) and PRP9-AID*-6FLAG (YBRT1) strains on Kaiser or CSM recipes of YMM. AID-tagged strains were included in the analysis as a negative control for growth. I observed that non AID-tagged strains grew less well on Kaiser YMM with 750 uM of auxin (IAA) (Figure 4.8A). In contrast, no growth was observed on CSM YMM with the same high concentration of auxin (Figure 4.8B). This suggests that the apparent toxic effect of auxin is related to nutrient content in the media, as the CSM mix has fewer and less nutrients than Kaiser. I then performed an additional spotting assay with a range of auxin concentrations (7.5, 30, 75 and 750 uM) on 2x Kaiser YMM (Figure 4.8C). I used 2x instead of 1x Kaiser with the hope that an increase in the amount of nutrients would decrease even further the inhibitory effect on growth. With this, I observed no difference in growth between no auxin and 75 uM or lower of auxin. A growth rate analysis of YGM1 and PRP22-AID*-6FLAG strains on 2x Kaiser YMM broth confirmed that addition of 75 uM by itself does not inhibit growth (Figure 4.8D).

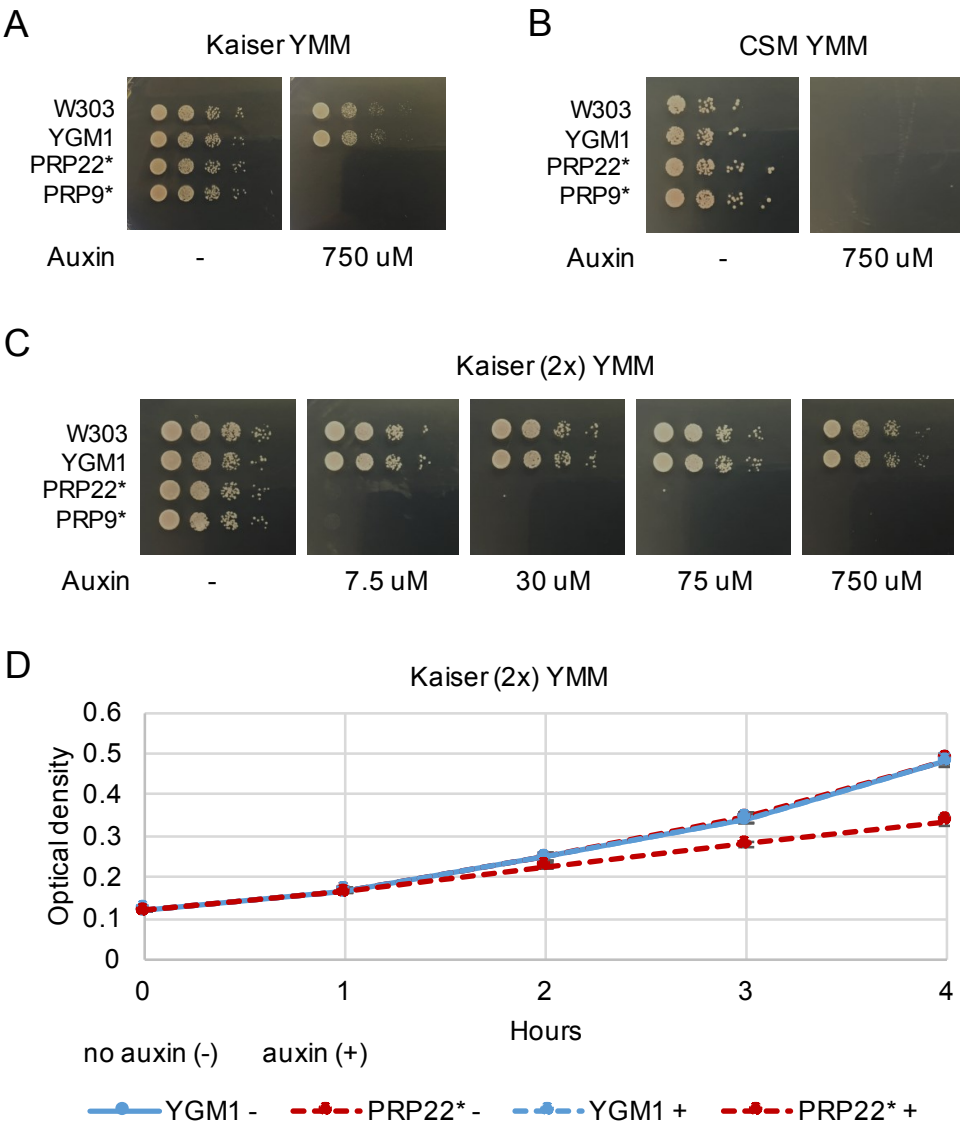


Figure 4.8. A low auxin concentration (75 uM) does not inhibit grown on Kaiser (2x) YMM
(A) Cells of W303, YGM1, PRP22-AID*-6FLAG (YGM1) and PRP9-AID*-6FLAG (YBRT1) strains, were spotted on agar containing Kaiser YMM without (-) or with 750 uM of auxin (IAA)
(B) As (A) but on CSM YMM
(C) As (A) but on Kaiser (2x) YMM without (-) or with 7.5 - 750 uM of auxin (IAA)
Assays from different panels were produced on different days
(D) Growth rate of YGM1 and PRP22-AID*-6FLAG (YGM1) on Kaiser (2x) YMM were measured without (-) or with 75 uM of auxin (+)

Finally, I investigated whether depletion is as efficient when adding low amounts of auxin, as when adding the usual high amounts. In cultures grown on 2x Kaiser YMM, I measured Prp22 levels at time 0, 6 and 12 minutes after

auxin addition, and observed that Prp22 levels are similar when adding 30, 75 or 750 uM of auxin. I therefore conclude that addition of 75 uM of auxin (IAA) leads to the efficient depletion of AID-tagged Prp22 (Figure 4.9). Overall, the results suggest that it is important to induce with only the minimal amount of auxin to avoid possible secondary effects associated with the growth inhibitory effect of auxin at high concentrations on YMM.

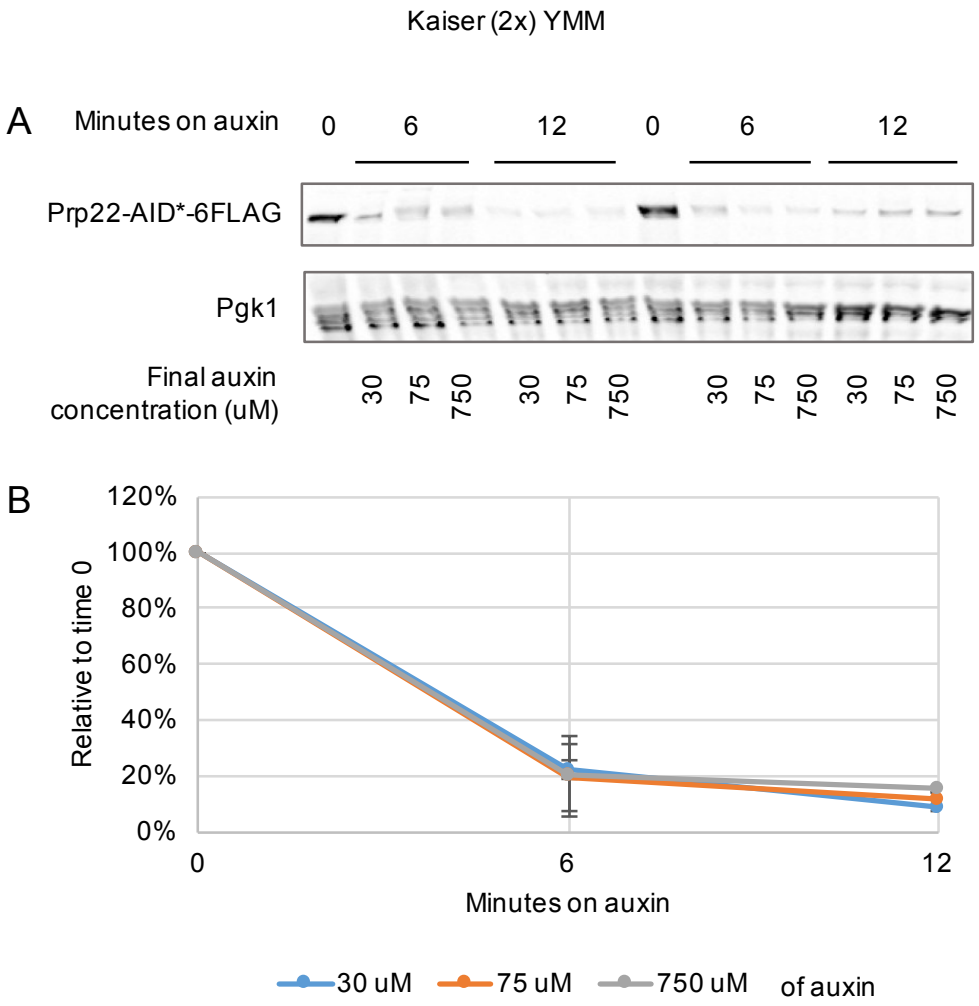


Figure 4.9. Addition of 750, 75, or 30 uM of auxin lead to similar depletion efficiencies of Prp22 (2x Kaiser YMM)
(A) Prp22-AID*-6FLAG (YGM1) abundance at 0, 6 and 12 minutes after adding 30, 75 or 750 uM of auxin (IAA)
(B) Quantitation of (A). Data are presented as Pgk1-normalized, relative to time 0. Error bars denote standard error of two biological replicates.

Discussion

High TIR1 abundance leads to fast but uncontrolled depletion

It was previously suggested that Tir1 is a rate-limiting factor of auxin response in *Arabidopsis thaliana*. This was based on observations that over-expression of Tir1 leads to an auxin-response phenotype (Gray et al., 1999) and depletion of Tir1 substrates (Aux/IAA proteins) (Dos Santos Maraschin et al., 2009), even without exogenous auxin. It is likely that the endogenous plant auxin is mediating the Tir1 induced auxin-response the authors observed.

Here I present evidence that confirms that, in the transgenic plant-derived AID system of budding yeast, Tir1 is also a rate-limiting factor, and that the target protein may be depleted without auxin addition when Tir1 is expressed to high levels, as seen in plants. This does not imply that Tir1 can bind its target in the absence of auxin. It could be that uncontrolled depletion is mediated by residual auxin in the growth media and/or endogenous auxin (IAA), as evidence suggests that budding yeast can synthesize IAA (Rao et al., 2010). Therefore, it is possible that using minimal media instead of rich media can help to reduce uncontrolled depletion, as the former does not contain yeast extract.

Why does YGM1 have a low success rate?

Strains generated by AID tagging essential genes on YGM1 fell into one of three categories according to their growth phenotype (Table 4.1). In the first category, strains grew normally without auxin and slowly after auxin addition (for example, AID-tagged Prp45, Prp16 and Prp22) which is the expected phenotype. In the second, growth was normal even after auxin addition (for example AID-tagged Prp39 and Prp9), and in the third, growth was slow

without auxin and slower after auxin addition (for example, AID-tagged Syf1 and Prp3).

The recent evidence linking OsTIR1 expression levels with uncontrolled depletion (Figure 4.4), can now explain why many of these AID-tagged strains do not respond to auxin (second category). It seems that the process of AID tagging certain genes is lethal and thus creates a strong selective pressure for a compensatory mutation or, more likely, the ejection of the pMK200 from the genome, which would break the AID pathway and uncontrolled depletion and therefore allow the cells to grow. In agreement with this, we generally observed that a low transformation efficiency (data not shown) coincided with a lack of growth inhibition when adding auxin.

As for the third category, it seems that in some cases the degree of un-induced depletion may not be high enough to kill the cells, but enough to negatively affect growth rate. This could be the case of AID-tagged Syf1 (Figure 4.1). Before auxin addition, AID-tagged Syf1 strain grows slower than wild-type. Although it appears that AID-tagged Syf1 is less abundant than AID-tagged Prp45, Prp16 and Prp22 proteins, this does not necessarily suggest that the levels of AID-tagged Syf1 are already reduced even without auxin addition. One reason for this is that these AID-tagged proteins were probed and scanned separately, on different days. Therefore, one cannot reliably compare protein levels across the images, due to technical variations of the assay and differences in the brightness adjustments of the images.

Overall, the data suggests that the degree of un-induced depletion is protein dependent, and/or that cells can grow well even with low amount of some essential proteins. It could be that even the AID-tagged strains (YGM1-derived) that grow as wild-type, and that respond to auxin addition (first

category), may partially deplete the target even before auxin addition. An example of this is AID-tagged Prp22 (YGM1). Considering this, it may be useful to investigate if levels of AID-tagged Prp45, Prp46 or Syf1 are already reduced before adding auxin, as the case of AID-tagged Prp22. This would require generating a new set of strains where the proteins of interest are tagged with 6FLAG, lacking the AID* domain, or AID*-FLAG tagging into the genome of a strain that does not express OsTir1. However, I decided not to pursue this, because it did not seem to be a high priority experiment that might affect my main conclusions.

The B-estradiol AID system

An ideal induction system should be gratuitous, tightly controlled and fast. The original yeast AID based on GAL-OsTIR1 (Nishimura et al., 2009) presumably does not have the problem of uncontrolled depletion without auxin, as OsTir1 expression is only induced before auxin addition. However, this system is not ideal because it perturbs cell metabolism due to the shifts in medium nutrients (from glucose to galactose). Here I have shown that the constitutive expression of Tir1 is not ideal either, because a strong promoter may lead to uncontrolled depletion (*e.g.* the YGM1 strain) and a weak promoter may deplete inefficiently (*e.g.* the YBRT1 strain) (Figure 4.7). In principle, one may find a constitutive promoter of intermediate strength that expresses OsTir1 at optimal levels for a certain target. However, it is likely that the optimal OsTir1 level depends on the abundance of the target protein, as my observations suggest (Figure 4.5). Furthermore, observations by Isabella Maudlin suggest that even the very low expression levels of OsTir1 in YBRT1 leads to uncontrolled depletion of some target protein (personal communication).

I showed that the gratuitous B-estradiol expression of OsTir1 solves the issues of the previous AID systems, as it allows for a fast and controlled depletion of the target protein (Figure 4.7). An additional advantage of this system is that the time between B-estradiol and auxin addition can be tuned according to the abundance of the target protein. More abundant targets may require a longer time of B-estradiol pre-incubation, which will lead to higher OsTir1 abundance and therefore higher depletion efficiency. This feature is not only useful for abundant targets but also to specifically modulate depletion speed. In most cases the fastest depletion is better, but there are situations where a gradual depletion of the target is desirable, as it has been suggested that different degrees of depletion can produce different outcomes (see discussions in Chapter 3).

Why does auxin at high concentrations inhibit growth?

There are at least two possible reasons why cells grow slowly on YMM containing high concentrations of auxin (IAA). An interesting and recent report, shows that auxin induces adhesion and filamentation in budding yeast – morphological changes associated with plant infection, which suggests that yeast can perceive and respond to auxin (Prusty et al., 2004). In relation to this, growing evidence indicates that the plant/yeast phytohormone signalling is bidirectional, as it has been shown that yeast can synthesize a variety of plant hormones including auxin, reviewed by (Chanclud and Morel, 2016; Fu et al., 2015). Prusty and others (2004) also observed that auxin at high concentrations inhibits cell growth, and that the expression profile of cells treated with sub inhibitory concentrations of auxin (100 uM) is distinct to that of cells exposed to diverse environmental stress conditions. So, based on the literature it seems that rather than auxin being toxic to yeast, the reduced growth is part of the

auxin-induced response of yeast while infecting plants, as discussed by the authors.

Alternatively, it is possible that auxin at high concentrations interferes with yeast metabolism. Auxin is structurally similar to its precursor amino acid, tryptophan. Based on this, it may be that an excess of auxin competes with the biochemical reactions of tryptophan (e.g. misincorporation into proteins). The inverse correlation between the richness of the medium and the degree of growth inhibition at high auxin concentrations agrees with this; growth was more inhibited in CSM YMM (50 mg/L Tryptophan) than in Kaiser YMM (76 mg/L Tryptophan) medium (Figure 4.8); in contrast, no growth inhibition was observed in YPDA. Although this idea is speculative, it would not be inconceivable as many non-protein amino acid analogues are toxic to cells.

Acknowledgements

I would like to thank Jane Reid for introducing the AID system to the lab and her suggestions on how to improve it. I also thank David Barrass, Barbara Terlow, Vahid Aslanzadeh, Ema Sani, Isabella Maudlin, Susana De Lucas, Charlotte, and Edward Wallace, for contributing with reagents and/or to the collective knowledge of the AID system in the lab. I must specifically note that YBRT1 strain was generated by Barbara T., Jane R. and David B., Prp22 was AID-tagged on YBRT1 by Ema S. and pVAS-osTIR1 plasmid (containing the Os*TIR1*-V5 construct) was generated by Vahid A. I also thank Masato Kanemaki for his valuable comments and for kindly providing reagents of the AID system, including the anti-OsTir1 antibody. Finally, I thank R. Scott McIsaac for kindly providing reagents of the B-estradiol expression system, through the request of Vahid A., including YMN3 yeast strain and pMN10 plasmid.

Chapter 5: Global analysis of Prp22's role in splicing fidelity

Abstract

Prp22 is a spliceosomal RNA helicase that contributes to splicing fidelity by proofreading exon ligation. To better understand the function of Prp22 in splicing fidelity, I globally measured splicing errors in the absence of Prp22 or its helicase activity, using the B-estradiol auxin-inducible degron (AID), combined with high depth RNA-sequencing. Through this analysis, I generated preliminary data to suggest that highly expressed genes compete better for splicing factors, when they become limiting due to a splicing defect. Secondly, I show that absence of Prp22 or its helicase activity leads to proportionately more errors in 3'ss selection, which demonstrates that Prp22 is required for proper 3'ss selection of transcripts of endogenous genes *in vivo*. In summary, this chapter presents promising preliminary data that will be analysed in more depth on ongoing studies. The expectation is that the complete work will help to improve our understanding of Prp22-mediated 3'ss selection.

Introduction

Splicing fidelity is important, as incorrect splicing of a gene transcript can lead to its reduced expression, or to its translation into an aberrant protein that may be toxic to the cell. It was proposed that splicing fidelity is influenced by spliceosomal RNA helicases, including Prp16, Prp22, Prp43 and Prp5, as a secondary role mechanistically associated to their main role in triggering structural transitions throughout the spliceosome cycle. The current model suggests that these RNA helicases promote splicing fidelity by rejecting suboptimal splice sites through their RNA unwinding activity. For instance,

Prp22's main role is to release the spliced mRNA from the post catalytic spliceosome and, in addition, evidence suggests that it proofreads exon ligation by rejecting sub-optimal 3' splice sites through the same RNA unwinding mechanism that releases spliced mRNA.

Although our understanding of splicing proofreading has advanced significantly, there are many important mechanistic details that are waiting to be uncovered. One interesting question that has not been fully answered is, how does the spliceosome distinguish between the canonical splice site and a potentially competing sub-optimal splice site?

As previous reports are mainly based on *in vitro* systems containing artificial pre-mRNAs with non-consensus splice sites or chemically modified snRNAs, a global *in vivo* analysis has the potential to provide new insights. To try to fill this gap, I performed an RNA-sequencing analysis to measure the frequency of splicing errors in the absence of Prp22 or its helicase activity. The main goals were to 1) confirm the requirement of Prp22 for proper 3’ splice site selection and 2) to identify possible substrates subject to Prp22-dependant rejection. In wild-type I detected around 180 aberrant 3’ splice sites (consistent with previous reports), and observed that about 20% of these are significantly more frequent in *prp22* mutants compared to wild-type. The data will be analysed in more depth in ongoing studies to try to better understand how Prp22 distinguishes optimal from suboptimal substrates.

Results

Swapping expression of wild-type Prp22 for mutants of Prp22

The experimental approach was to analyse three conditions: 1) depletion of Prp22, 2) depletion of Prp22 with the simultaneous induced expression of ATPase-deficient Prp22 mutant T757A or 3) simultaneous induction of ATPase-hyperactive Prp22 mutant I7564. All strains are *upf1*Δ and therefore, deficient in non-sense mediated decay (NMD). For simplicity, I will often refer to these strains (conditions) as Δ22/-, Δ22/T757A and Δ22/I764A, respectively. Both mutants lack RNA unwinding activity (Schneider et al., 2004), which was observed to be important for rejection of non-consensus 3'ss (Mayas et al., 2006). I included these mutants in this study to 1) determine whether the *in vivo* role of Prp22 in splicing fidelity is related to its RNA unwinding activity and 2) to avoid a possible inhibition of the second catalytic step of splicing, as it has been proposed that Prp22 has an ATPase-independent role in the second catalytic step of splicing on some gene transcripts (Schwer and Gross, 1998).

The strains Δ22/T757A and Δ22/I764A that express Prp22 mutant proteins were generated by transforming a B-estradiol AID strain (YZTR41) that is Prp22 AID-tagged (Δ22/-), with a centromeric plasmid (Figure S5.2) containing the *T757A* or *I764A* mutant alleles under the control of the B-estradiol-regulated artificial promoter PZ4EV (Figure 5.1). I tested the three different conditions by performing a western blot time-course analysis. At time 0, B-estradiol and auxin were added simultaneously to achieve a gradual depletion, as it is likely that this would allow more time for the spliceosome to make fidelity mistakes before the first step of splicing is inhibited as a consequence of other splicing factors becoming limited.

Through this assay, I verified that wild-type Prp22 was efficiently depleted in the three strains, and at a slower rate than in the previous experiment, where I added auxin 20 minutes after B-estradiol (Chapter 4 Figure 4.7). In this new experiment, 15, 30 and 45 minutes after addition of the inducers, Prp22 levels were reduced to about 80%, 30% and 10% of normal levels, respectively. Also, I observed that the Prp22 T757A and I764A mutant proteins were produced quickly, reaching detectable levels at 15 minutes and about 6 or 10 times more than wild-type levels at 45 minutes.

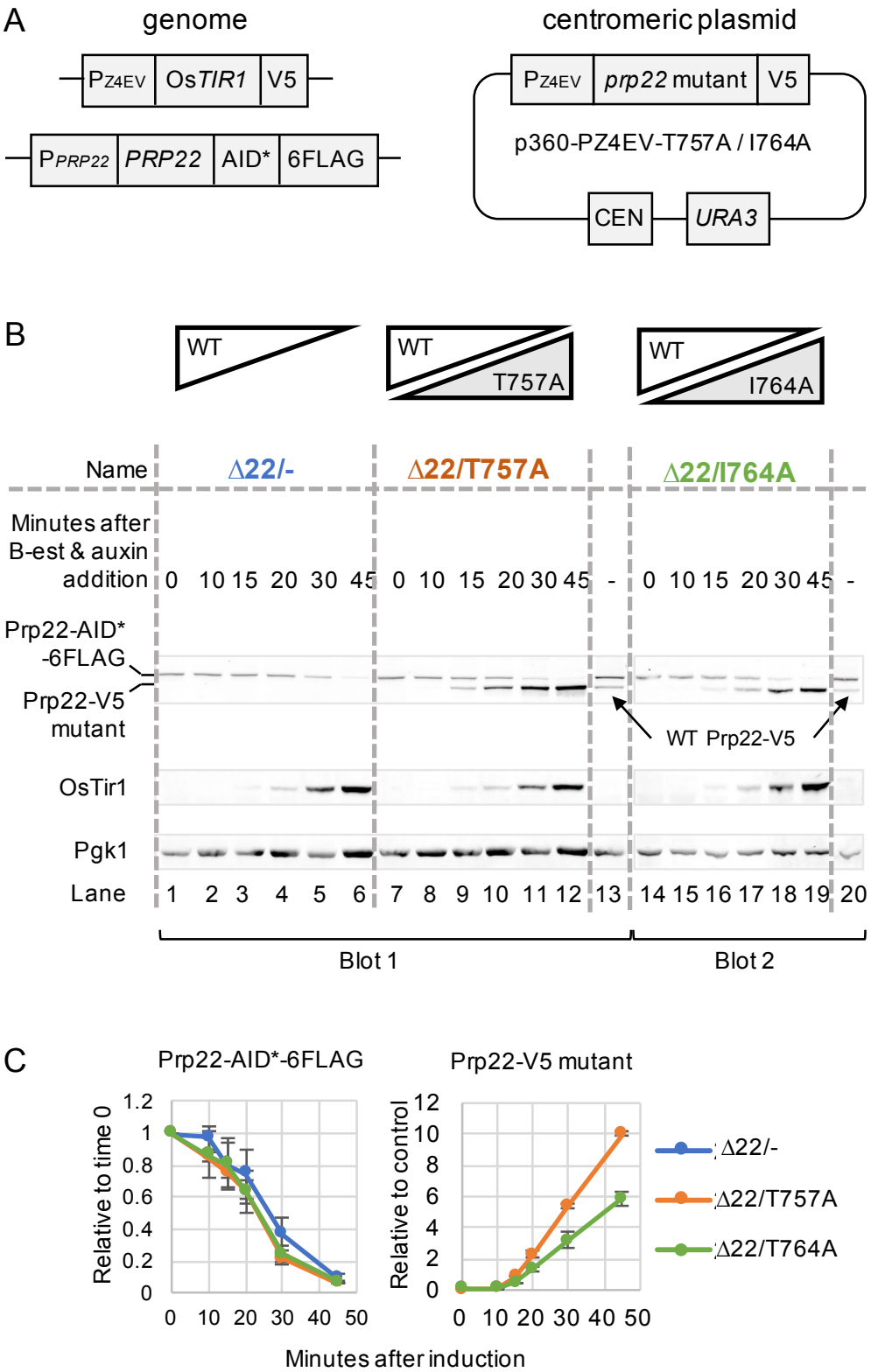

A
genome
centromeric plasmid
PZ4EV
OsTIR1
V5
PPRP22
PRP22
AID*
6FLAG
PZ4EV
prp22 mutant
V5
p360-PZ4EV-T757A / I764A
CEN
URA3
B
WT
WT
T757A
WT
I764A
Name
Δ22/-
Δ22/T757A
Δ22/I764A
Minutes after B-est & auxin addition
0 10 15 20 30 45 0 10 15 20 30 45 - 0 10 15 20 30 45 -
Prp22-AID* -6FLAG
Prp22-V5 mutant
WT Prp22-V5
OsTir1
Pgk1
Lane
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
Blot 1
Blot 2
C
Prp22-AID*-6FLAG
Prp22-V5 mutant
Relative to time 0
Relative to control
Δ22/-
Δ22/T757A
Δ22/T764A
Minutes after induction

Figure 5.1. The B-estradiol AID system is used to simultaneously induce gradual depletion of WT Prp22, and express Prp22 helicase-deficient mutants.
A) Illustration highlighting the key elements of the system. The simultaneous addition of B-estradiol and auxin leads to expression of OsTIR1 and Prp22 mutants through the B-estradiol promoter (PZ4EV), and depletion of AID-*6FLAG tagged Prp22 (WT). ATPase-deficient mutant T757A and ATPase-hyperactive mutant I764A are both deficient in RNA unwinding. They are reported to allow the second step of splicing *in vitro* but not mRNA release from the post-spliceosome (Schneider et al., 2004).
B) Time course analysis of simultaneous addition of B-estradiol and auxin to cultures of AID*-6FLAG tagged Prp22 containing plasmids p360-mock, p360-PZ4EV-*prp22*-T757A or p360-PZ4EV-*prp22*-I764A; hereafter these strains will be referred to as Δ22/-, Δ22/T757A and Δ22/I764A, respectively. Extracts from cells containing plasmid p360-PRP22-V5 was added as a control (lanes 13 and 20).
C) Quantitation of (B). Data is presented as Pgk1-normalized and relative to time 0 or to the control of lanes 13 and 20 in (B). Error bars denote standard deviation of two biological replicates.

Prp22 mutants T757A and I764A inhibit growth

I then investigated whether cell growth is affected when mutant proteins are expressed in combination with the wild-type. To this end, I measured the growth rate of strains lacking an AID tag on wild-type Prp22, and containing the plasmids for the B-estradiol-induction of the mutants. I observed that over-expression of either of the mutants inhibited growth to about the same low level as when depleting wild-type Prp22 (Figure 5.2), which indicates that these mutants are semi-dominant or dominant negative but not recessive, in apparent contradiction to published data (Schneider et al., 2004).

Loss of Prp22 RNA unwinding activity causes a first step of splicing defect

I then analysed the splicing phenotype of Δ22/-, Δ22/T757A and Δ22/I764A strains 30 and 45 minutes after B-estradiol and auxin addition, by RT-qPCR of *ACT1, ECM33, RPL28, RPS13* and *RPL39* transcripts. I observed more lariat and pre-mRNA (5'ss, branchsite and 3'ss amplicons) compared to wild-type (Figure 5.3), as I previously observed with depletion of Prp22 in the YGM1 strain (Chapter 3 Figures 3.5). As one mutant retains a high level of ATPase activity but both mutants lack RNA unwinding activity, which suggests that

loss of Prp22's RNA unwinding activity, is sufficient to cause a recycling defect leading to a block in the first step of splicing. Generally, I observe similar qualitative results across the three strains. However, the ones expressing a mutant protein accumulate more lariat and pre-mRNA than the depletion only, especially at 30 minutes. This can be explained by the apparent dominant nature of the mutants, combined with the observation that, after B-estradiol and auxin addition, the protein levels of the mutants start to increase before wild-type levels start to drop (Figure 5.1). This difference in the degree (or timing) of the splicing defect should be considered for interpreting the following results.

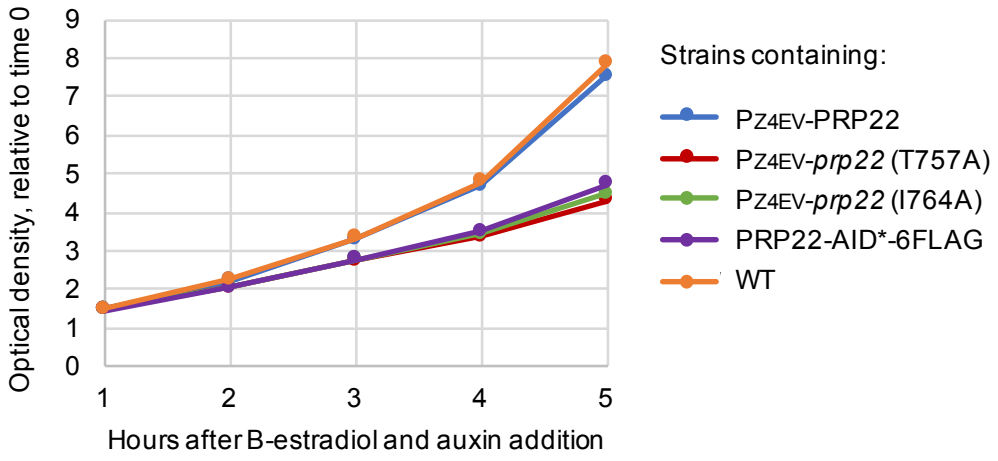


Figure 5.2. Prp22 mutant proteins, T757A and I764A, inhibit growth even without wild-type depletion. B-estradiol and auxin were added at time 0 and optical density was measured every following hour on strains with induced over-expression of Prp22 wild-type (blue), T757A (red) or I764A (green). An AID-tagged Prp22 (purple) and a BY4741 wild-type strain (orange) were also included in the analysis as positive and negative controls of growth inhibition.

In summary, these results demonstrate that 1) the WT-for-mutant swap technique works well and 2) that the Δ22/-, Δ22/T757A and Δ22/I764A strains produced the expected splicing phenotype when induced; which leads to the second part of this chapter where I describe the results of the splicing fidelity analysis.

RNA-sequencing analysis to study splicing fidelity

For RNA-sequencing, I used biological duplicate cultures treated for 45 minutes with B-estradiol and auxin, which were previously analyzed by Western blot and RT-qPCR (Figures 5.1 and 5.3). This includes Δ22/T757A, Δ22/I764A, Δ22/-, and WT that lacks the AID-tag and contains a *URA3*-containing mock plasmid. Ribosomal RNA depletion, library preparation (no poly-A selection) and paired-end Illumina sequencing was performed by BGI Genomics (Hong-Kong). An average of 85 million uniquely mapped reads were generated per sample, out of which 0.9-1.9 million are mRNA junction reads (Table 5.1). Samples from mutants contain about half as many junction reads as WT, due to their splicing defect.

Table 5.1. Total number of mapped reads per sample

Sample (replicate)	Mapped reads*	Junctions**
Δ22/I757A (1)	89,855,130	903,293
Δ22/I757A (2)	104,000,443	1,033,597
Δ22/T764A (1)	84,386,081	910,649
Δ22/T764A (2)	86,571,595	889,714
Δ22/- (1)	87,701,109	1,358,104
Δ22/- (2)	71,259,795	1,025,710
WT (1)	88,473,836	1,904,651
WT (2)	79,995,411	1,667,470

* Number of uniquely mapped reads
** Number of uniquely mapped mRNA junctions

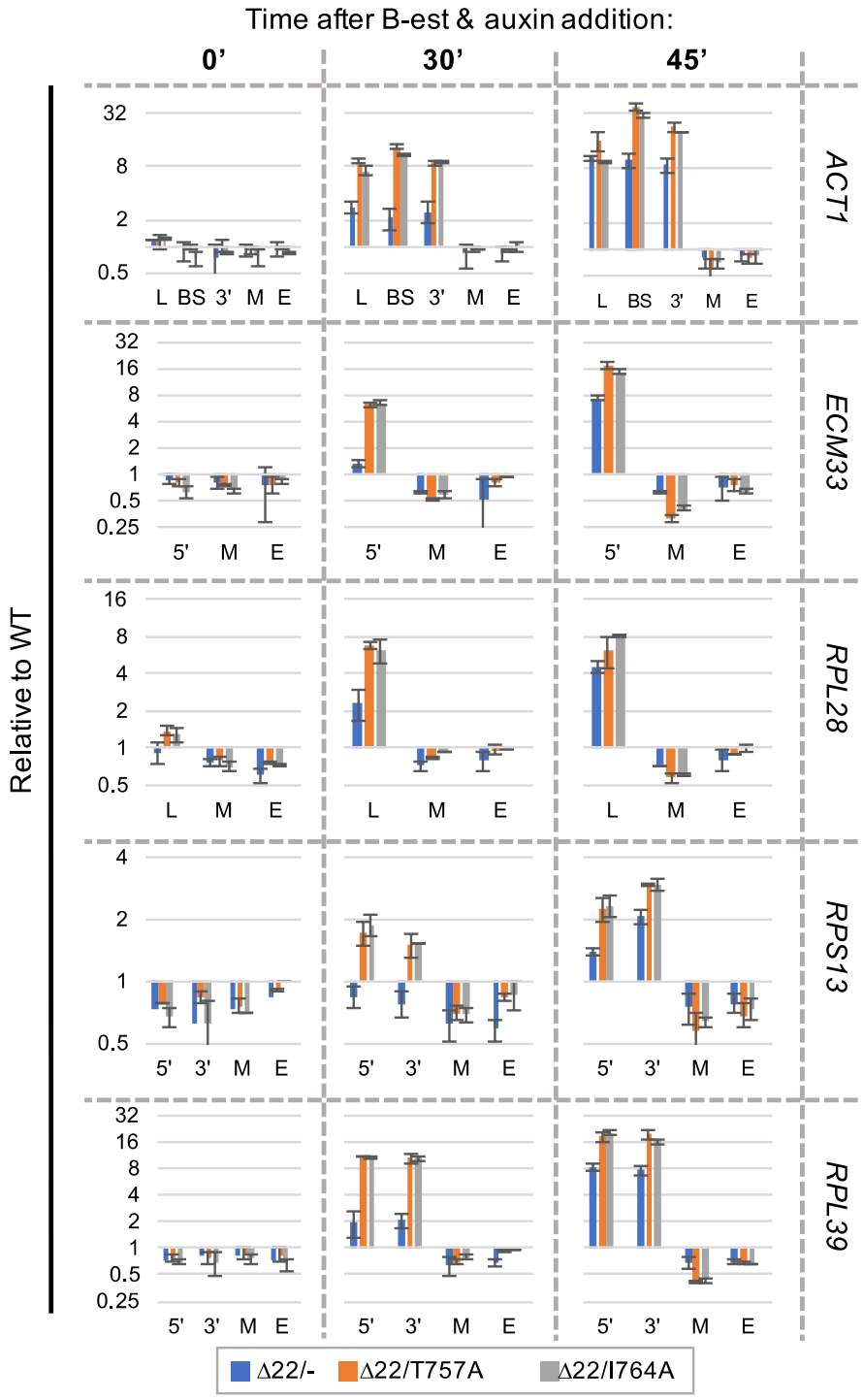


Figure 5.3. Induced expression of Prp22 helicase-deficient mutants leads to reduced splicing efficiency. Cell samples were prepared as in Figure 5.1. Splicing efficiency was analyzed at time 0, 30 and 45 minutes after induction, through RT-qPCR with amplicons to lariat (L), 5' splice site (5), branch site (BS), 3' splice site (3), mRNA (M) and/or second exon (E) of *ACT1, ECM33, RPL28, RPS13* and *RPL39* transcripts. The control strain (WT) contains plasmid p360-mock and is not Prp22 AID*-6FLAG tagged. Data are presented as relative to WT. Error bars denote standard error of two biological replicates.

Abundant transcripts appear to be less affected by the splicing defect

To measure the splicing defect per gene, I divided mRNA junction reads of mutants by WT, which I then used to generate a clustered heatmap (Figure 5.4A). Here, I observed a lower mutant/WT junction ratio for most intron-containing genes, confirming that depletion of Prp22 or absence of its RNA unwinding activity globally inhibits splicing. This splicing inhibition is likely due to reduced spliceosome assembly and not splicing catalysis, as it was previously demonstrated that these RNA helicase mutants affect spliced mRNA release from the post-spliceosome, but not the second step of splicing (Schneider et al., 2004).

Also, I observe that Ribosomal Protein Gene (RPG) transcripts are proportionally less affected by the splicing defect. To investigate whether this is due to their higher expression level compared to non-RPGs, I produced a correlation plot between number of junction reads in WT against the mutant/WT junction ratio (Figure 5.4B), as an attempt to compare "expression level" vs. "splicing defect". Here, I observe an inverse correlation between these two features, within both RPGs and non-RPGs. Therefore, it seems that intron-containing gene transcripts that are very abundant (with high number of junction reads in WT), such as most RPGs, tend to be proportionately less affected by a splicing defect (high mutant/WT ratio of junction reads), and vice versa. This complements previous evidence that RPGs are spliced particularly fast (Barrass et al., 2015), co-transcriptionally and efficiently (Wallace and Beggs, 2016) and that they compete better for spliceosome components (Munding et al., 2013).

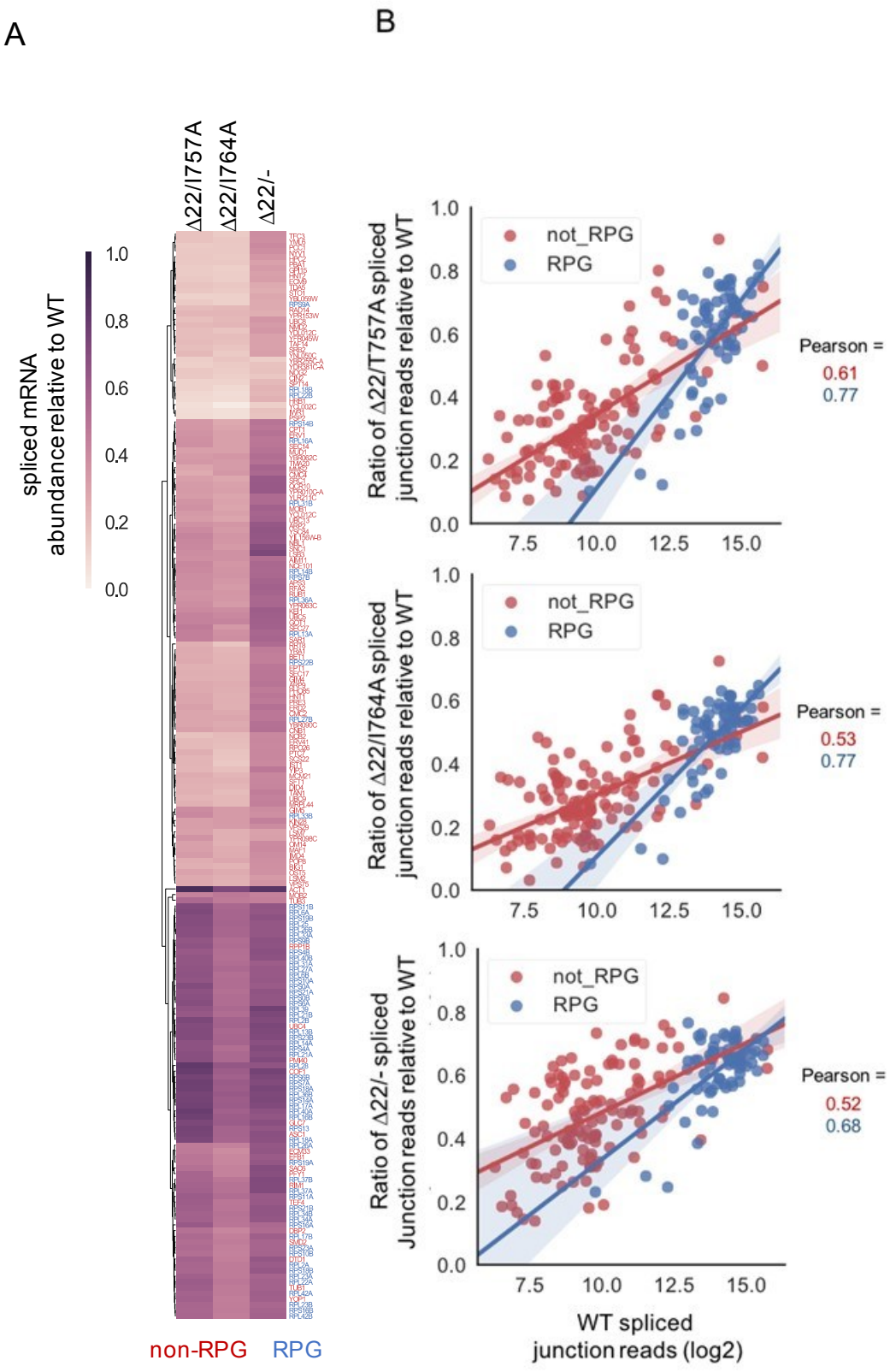
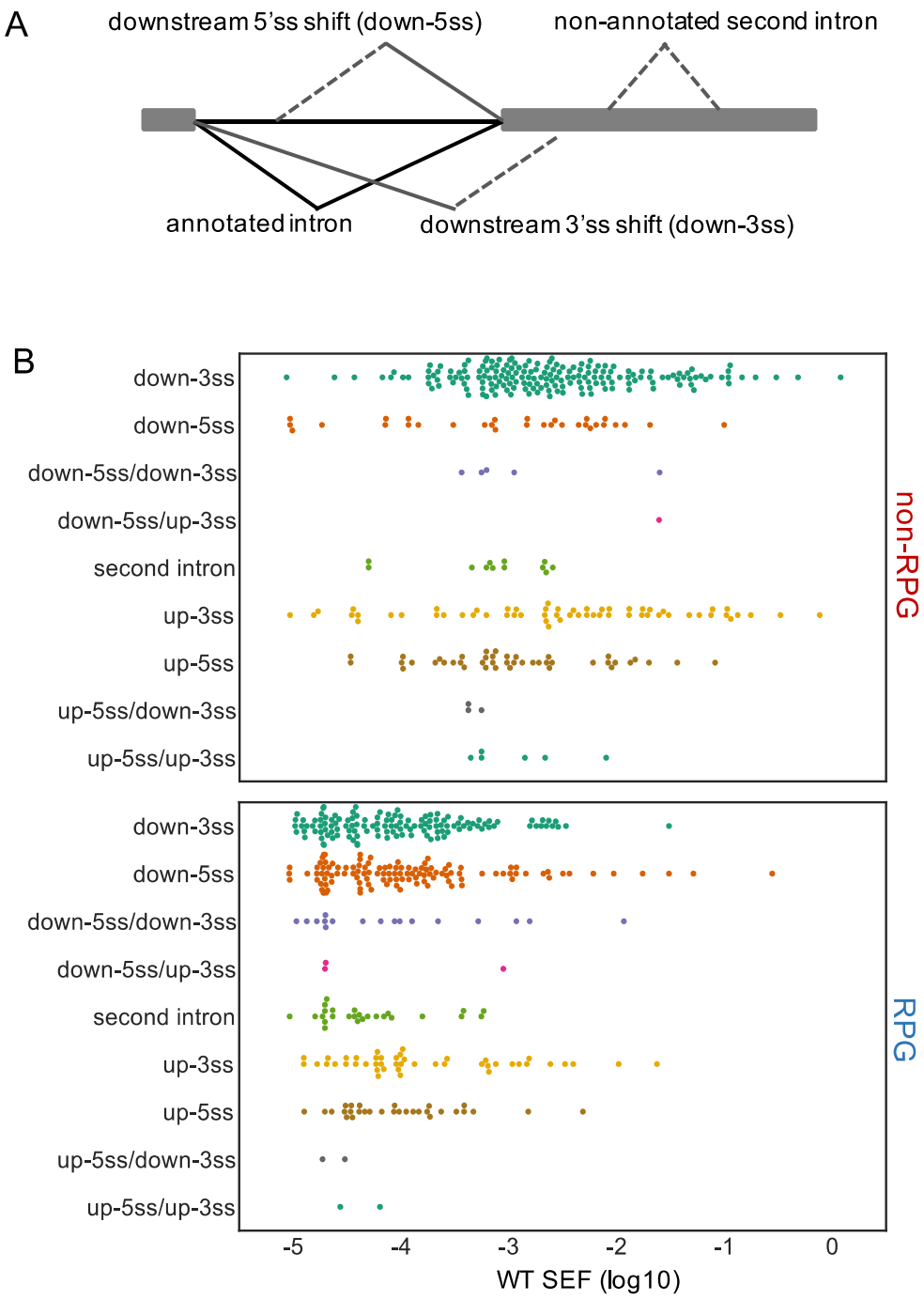

A
Δ22/I757A
Δ22/I764A
Δ22/-
spliced mRNA abundance relative to WT
1.0
0.8
0.6
0.4
0.2
0.0
non-RPG
RPG
B
not_RPG
RPG
Ratio of Δ22/T757A spliced junction reads relative to WT
Pearson =
0.61
0.77
Ratio of Δ22/I764A spliced junction reads relative to WT
Pearson =
0.53
0.77
Ratio of Δ22/- spliced Junction reads relative to WT
Pearson =
0.52
0.68
7.5
10.0
12.5
15.0
WT spliced junction reads (log2)

Figure 5.4. Spliced mRNA is less abundant compared to WT, and genes that are less expressed are more affected by this splicing defect. Illumina RNA sequencing analysis was performed from cultures treated with B-estradiol and auxin for 45 minute as described previously, which includes depletion of Prp22 (Δ22/-), depletion of Prp22 with simultaneous expression of T757A (Δ22/T757A) or I764A (Δ22/I764A) Prp22 mutants, and WT. mRNA junction reads of mutants were divided by those of WT. Data represent the average of two biological duplicates.
A) Hierarchical clustering of ratios of spliced mRNA reads relative to WT. Names of non-ribosomal protein genes (non-RPG) were labeled in red and ribosomal protein genes (RPG) in blue.
B) Correlation analysis between ratios of mRNA reads relative to WT (y-axis), and mRNA reads in WT (x-axis). Pearson correlation coefficient was calculated separately for RPGs and non-RPGs.

The normal frequency of splicing errors

To measure the errors in splicing, I first discarded the non-annotated splicing events that do not have at least 1 read on each two replicates from one experiment. Then, for each non-annotated junction, I obtained the splicing error frequency (SEF) score, which is the proportion of reads of the non-annotated junction with respect to the number of reads of the annotated junction. For wild-type only, I then plotted the SEF of non-RPGs and RPGs separately, according to the type of splice site shift (downstream 5'ss, downstream 3'ss… etc) (Figure 5.5). This allowed me to confirm, as very recently shown by a similar splicing fidelity analysis (Aslanzadeh et al., 2017, in review), that normally splicing errors: 1) are mostly rare – only 77 out of 308 non-annotated junctions have a SEF > 0.01, 2) are rarer for RPG than non-RPG transcripts and 3) happen more frequently for 3'ss than 5'ss.



$$\text{Splicing Error Frequency} = \frac{\text{non-annotated junction reads}}{\text{annotated junction reads}}$$

Figure 5.5. The normal frequency of non-annotated spliced mRNAs is dependent on the gene category (RPG) and type of splice site shift. Here, a splicing error is defined as a downstream or upstream shift relative to the annotated 5'ss and/or 3'ss **(A)**. The splicing error frequency (SEF) for each detected non-annotated splicing event (a total of 308) in WT, was calculated as the read counts of the non-annotated junction divided by the read counts of the annotated junction within the same gene. Data were plotted separately for non-RPG and RPG transcripts **(B).** Data represent the average of two biological duplicates.

Depletion of Prp22 or its RNA-unwinding activity alters the frequency of aberrant junctions

I then asked whether the frequency of splicing errors is increased in mutant relative to wild-type. First, as a general overview, I generated a correlation plot of SEF scores, Δ22/T757A, Δ22/I764A or Δ22/- against WT. Here, I observe that the frequency of about 1/5 of all aberrant splicing events, is significantly different (p-value < 0.05) in mutant compared to wild-type. Surprisingly, the change happens in both directions (for different transcripts there can be more or fewer aberrant events in mutant than in WT), which indicates that absence of Prp22 does not cause an indiscriminate increase in splicing errors. Furthermore, I observe changes in aberrant junctions that have low frequency error and also in the ones that have high frequency error (Figure 5.6), which indicates that the effect is not biased by how rare the aberrant junction is in WT, or whether it is a non-RP or RP transcript.

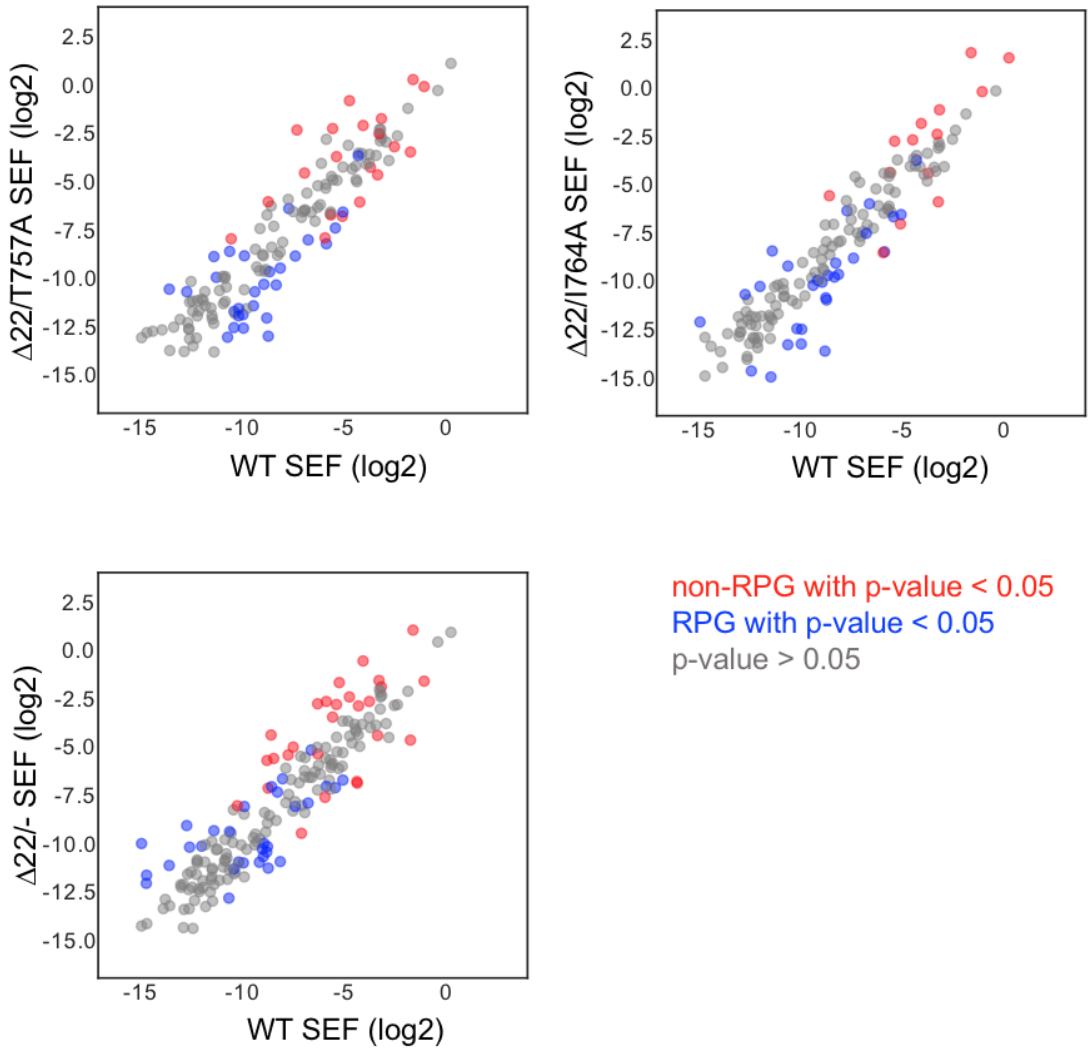


Figure 5.6. Absence of Prp22 or its RNA unwinding activity changes the frequency of splicing errors of many non-annotated junctions. SEF scores of mRNA junctions in Δ22/- , Δ22/T757A and Δ22/I764A, are plotted against WT. Splicing events that are significantly different from WT (p-value < 0.05) are shown in red (non-RPG) or blue (RPG); the ones with p-values > 0.05 are shown in grey. Data represents the average of biological duplicates.

Aberrant 3'ss are more frequent in depletion over wild-type

To analyse the changes in splicing error frequency in more detail, I then plotted the ratio of SEF scores of mutant over wild-type, according to the type of the splice site shift (Figure 5.7). I only included the non-annotated junctions that are significantly different (p-value < 0.05) from wild-type, which includes

about 60 splicing events. Within this set, most events correspond to downstream shifts of the 3'ss (blue) or 5'ss (orange), or an upstream shift of the 3'ss (green), relative to the annotated splice sites; other types of splice site shifts were very rare or absent, and were therefore not included in this plot.

Interestingly, I observe that many non-annotated splicing events with a 3'ss shift are more frequent in mutant than in wild-type (Figure 5.7A). In contrast, virtually none of the 5'ss shifts show this effect, which confirms that Prp22 is important for proper selection of 3'ss in particular. This holds true for Prp22 depletion (Δ22/-) and for the induced expression of helicase-deficient mutants (Δ22/T757A, Δ22/I764A); although the effect is more pronounced in Δ22/-.

Next, I plotted the data separately for non-RPGs and RPGs (Figure 5.7B). Although, I do observe that virtually all 5'ss shifts that are affected in the mutant belong to RPGs, this likely not unusual considering that this bias is also observed in wild-type (Figure 5.5). There may be a tendency of aberrant upstream 3'ss, but not downstream 3'ss, of non-RPG to be more affected in the mutant, although I do not have a reasonable explanation for this. I only present the data for non-RPGs and RPGs separately for Δ22/-, as I observed similar results with the other two mutants (data not shown).

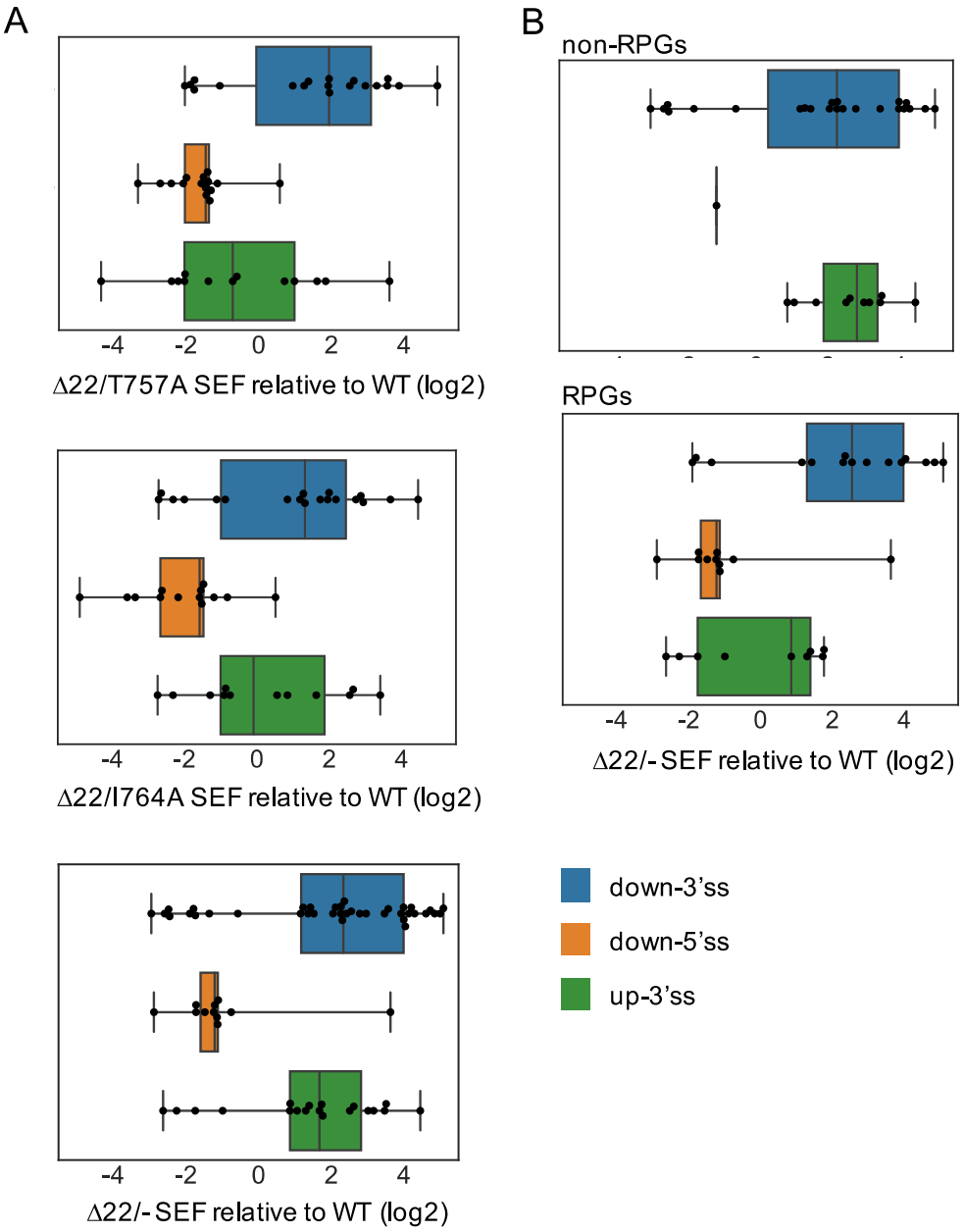
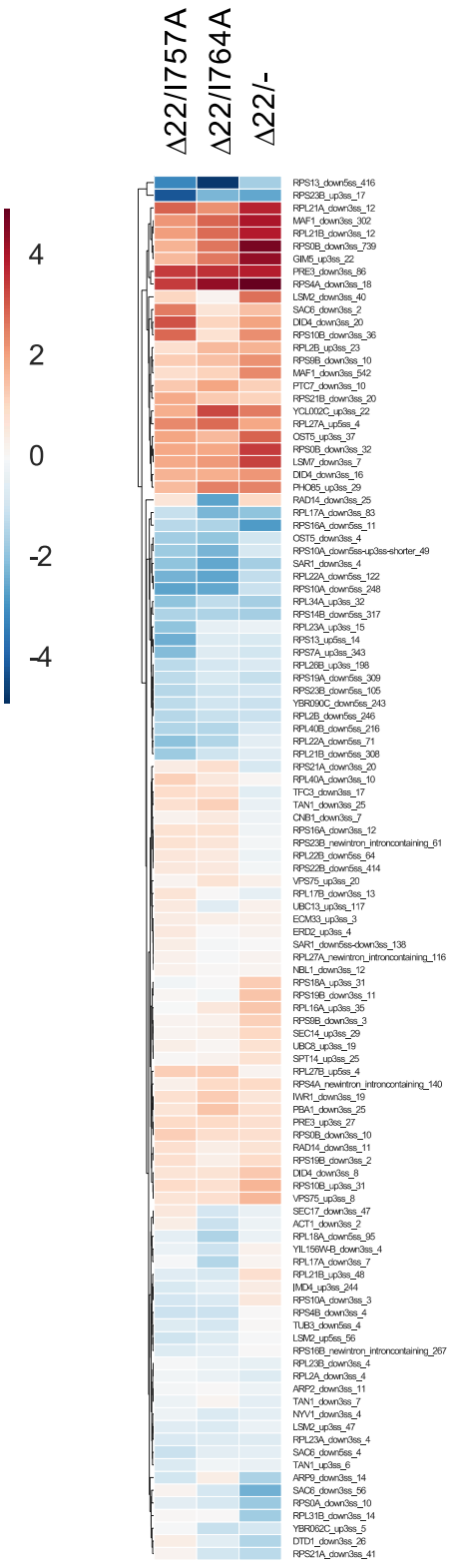


Figure 5.7. The frequency of many non-annotated 3'ss is increased relative to wild-type. The SEF score relative to WT (as in Figure 5.6) was calculated separately for mRNA junctions with a downstream (blue) or upstream shift (green) in 3'ss selection, and downstream shift in 5'ss selection (orange), for Δ22/T757A, Δ22/I764A and Δ22/- **(A).** Non-RPGs (red) and RPGs (blue) from Δ22/- of (A), were plotted separately **(B)**.

I then compared the mutant/WT SEF ratio across strains using a clustered heatmap (Figure 5.8). From this analysis, it seems that the effect is mostly similar between the three strains, although the depletion looks slightly different than for the two helicase-deficient mutants. Compared with the helicase-deficient mutants, the depletion appears shifted towards positive log2 values, as observed in the previous plot (Figure 5.6). These differences may be due to differences in the strength (or timing) of the splicing defect (Figures 5.1 and 5.3), which may indicate that a prolonged splicing defect, like the case of Δ22/T757A and Δ22/I764A, reduces the chances of detecting the rare aberrant junctions, maybe due to RNA turnover.

Overall, the results demonstrate that Prp22 is required for suppression of many non-annotated 3'ss *in vivo*. The observation that, in this respect, there are more similarities than differences between depletion and mutants (Figure 5.7) confirms that it is the RNA unwinding activity, and not the physical presence of the Prp22 protein by itself, that is important for its role in splicing proofreading.



SEF relative to WT

Figure 5.8. Absence of Prp22 protein (Δ22/-) or its RNA unwinding activity (Δ22/T757A or Δ22/I764A) results in a similar splicing error frequency profile. A clustered heat map was generated with the SEF scores relative to WT, of the non-annotated splicing events that were detected in all conditions (Δ22/T757A, Δ22/I764A, Δ22/- and WT).

Top 10 hits

I then selected 10 aberrant junctions, to show that there are at least a few aberrantly spliced transcripts that are notably affected by the absence of wild-type Prp22, both in relative and absolute numbers. As the Prp22 depletion strain (Δ22/-) generally showed the largest difference with respect to wild-type, from this data set I selected 10 aberrant junctions, that have a SEF score > 0.05 and a fold over wild-type SEF > 2 (and p-value < 0.05) (Figure 5.9) – read counts shown in Figure S5.3.

One of the most dramatic cases is a junction from *LSM7* transcript, where an alternative 3'ss was used 7 nt downstream of the annotated 3'ss. In wild-type conditions, this aberrantly spliced junction represents about 6% of the total population of *LSM7* spliced mRNAs but, when Prp22 is depleted, this number increases to about 67%. Another notable example is an aberrantly spliced mRNA of *YCL002C*, with a 3'ss shift 22 nt upstream. When Prp22 is depleted the frequency of this splicing event changes from about 33% to 203%, relative to wild-type.

In all top 10 aberrantly spliced transcripts, as in most non-annotated spliced transcripts, frameshifts in the coding sequences introduce stop codons, and they would therefore be targeted for NMD. This supports the idea that both the splicing machinery and RNA degradation proteins, contribute to the quality control of spliced mRNAs, as part of a tier-two strategy (Egecioglu and Chanfreau, 2011). Thus, aberrantly spliced mRNAs that have leaked through

the splicing proofreading mechanism are targeted for degradation in the cytoplasm, ensuring that only accurately spliced transcripts are translated.

Gene	splice site	nt shift	SEF as % WT	SEF as % Δ22/T757A	SEF as % Δ22/I764A	SEF as % Δ22/-	SEF fold over WT (0 5 10 15 20 25)
GIM5	up-3ss	22	**1**	19	4	13*	
RAD14	up-3ss	13	**3**	10	7	31*	
LSM7	down-3ss	7	**6**	23*	28*	67*	
HNT2	up-3ss	26	**2**	14	3	16*	
YCL002C	up-3ss	22	**33**	118*	350*	203*	
PHO85	up-3ss	29	**2**	7*	15*	14*	
IWR1	down-3ss	32	**4**	56*	43	18*	
DID4	down-3ss	20	**2**	20*	5	9*	
VPS75	up-3ss	8	**10**	17	19	33*	
PTC7	down-3ss	10	**11**	29*	45*	26*	

Figure 5.9. Top 10 non-annotated spliced junctions. 10 non-annotated junctions were selected for having a high frequency (SEF > 5%) and the largest increase in frequency relative to WT [(Δ22 SEF) / (WT SEF)]. * SEF values with statistical significant difference against WT (p-value > 0.05)

Discussion

How to study the role of Prp22 in splicing fidelity *in vivo*?

It is not trivial to study the role of Prp22 in splicing fidelity *in vivo*. Firstly, its role in fidelity is associated with its RNA unwinding activity, which is essential for cell viability. Therefore, it may not be not possible to generate a mutant allele of Prp22 that has lost its function in splicing fidelity while still conserving normal function of releasing the spliced mRNA from the post-

spliceosome – its canonical role in the spliceosome cycle. So, at the moment one cannot perturb its role in fidelity in a very specific manner. Furthermore, in Chapter 3 I showed that a quick depletion of Prp22 quickly inhibits the first catalytic step of splicing – a secondary effect that appears only a few minutes after the primary effect (Chapter 3 Figure 3.6). In other words, shortly after depleting Prp22, there is less production of newly spliced mRNAs, which includes the aberrantly spliced mRNAs I aimed to measure. This could mean that aberrantly spliced mRNAs may be difficult to detect under these conditions.

Fortunately, my data from Chapter 3 suggest that depletion of Prp22 accumulates arrested complexes. These complexes likely contain spliced mRNAs, including aberrantly spliced mRNAs. Furthermore, data generated by Isabella Maudlin shows that a mild depletion of Prp22 does not globally inhibit the first step of splicing (data not shown), which suggests that a gradual depletion of Prp22 may allow more time for the spliceosome to make mistakes, before the first step of splicing is inhibited due to accumulation of arrested complexes. Considering all arguments, there was good reason to believe that if Prp22 is required for *in vivo* splicing fidelity, its depletion will lead to more errors in splicing that are detectable by RNA sequencing. However, for the reasons mentioned above, it is likely that the measured effect is an underestimate of the full contribution of Prp22 to splicing fidelity.

Swapping expression of wild-type for mutant

By depleting wild-type Prp22 while inducing expression of a mutant Prp22, I demonstrated that the B-estradiol AID system can facilitate the study of mutant alleles *in vivo*. Alternatively, an experiment like this could have been done through a combination of inducible and repressible promoters (*e.g.* Gal

and Met promoters). However, most metabolic regulation systems require nutrient shifts that result in pleiotropic effects and, in addition, may not be as fast-acting as the B-estradiol AID.

This swap technique can be especially useful for studying recessive or semi-dominant mutants that may not show the full phenotype unless the wild-type is absent. In the case of recessive mutants, it may be sufficient to express the mutant alleles under their native promoter instead of through an inducible promoter.

Prp22 mutants behave as dominant negative

As published evidence suggests that the Prp22 mutants T757A and I764A are recessive, based on 5-FOA plasmid shuffling (Schneider et al., 2004), initially I had these alleles under control of their native promoter (p360-PRP22-V5 plasmids). However, I observed that the expression level of these mutant proteins is about 10 times less than wild-type (data not shown). Based on this, I then constructed new plasmids by substituting the native promoter for the PZ4EV B-estradiol promoter, and with these I performed a growth rate analysis that suggests that the mutants are indeed not recessive (Figure 5.2). To explain these apparently contradictory observations, I speculate that the attempt to simultaneously express Prp22 wild-type and dominant-negative mutants T757A and I764A (*e.g.* through plasmid shuffling), may result in strong selective pressure to reduce expression of these toxic proteins through a compensatory mutation, hence their low protein level.

RNA unwinding activity of Prp22 is indirectly required for the first step of splicing *in vivo*

In the previous chapter I showed that depletion of Prp22 protein indirectly leads to a defect in the first step of splicing. Here, I complement those previous results by showing, through an RT-qPCR analysis, that the induced expression of Prp22 helicase-deficient mutants also negatively affects the first step of splicing (Figure 5.3). Although there is no direct evidence that these mutants physically associate with the spliceosome, there are good reasons to believe so, including: 1) although lacking RNA unwinding activity it was shown, through an RNA binding assay, that these mutants can bind to an RNA substrate (Schneider et al., 2004), 2) it was shown that a Prp22 cold-sensitive mutant (ATPase-deficient at restrictive temperature) physically associates with a stuck post-spliceosome (Lardelli et al., 2010), and 3) the apparent dominant nature of the T757A and I764A mutants implies that they compete with the WT Prp22 for binding to the spliceosome. Overall, the data suggest that loss of Prp22 RNA unwinding activity without loss of the protein *per se*, is sufficient to inhibit spliceosome disassembly and recycling *in vivo*.

Abundant gene transcripts appear to be are less affected by the splicing defect

I have observed that for highly abundant transcripts, when splicing is inhibited, very often the number of mRNA junctions is proportionally less reduced in mutants compared to WT (Figure 5.4A). This is interesting because it may suggest that intron-containing transcripts that are highly expressed, tend to compete better for splicing machinery that has become limiting – in this case due to absence of Prp22 function that inhibits recycling of splicing components. This supports the proposition that, at least in budding yeast, pre-mRNA substrates compete for splicing components, a concept known as the

"hungry spliceosome" (Munding et al., 2013). It also supports the idea that intron-containing RPG transcripts, which are highly abundant, are spliced faster (Barrass et al., 2015) more co-transcriptionally (Carrillo Oesterreich et al., 2016; Wallace and Beggs, 2016) and more faithfully (Aslanzadeh et al., 2017, in review). It is likely that these associations between expression level and many aspects of splicing efficiency are the reflection of natural selection favouring efficient mRNA processing of highly transcribed genes.

Alternatively, there could be a different non-mutually exclusive explanation for this observation. Theoretically, the mutant/WT ratio of mRNA junctions I used as a measure of the splicing defect, depends on splicing efficiency, and to some extent also on the turn-over rate of spliced mRNAs. In other words, some spliced mRNAs that were already there before the splicing defect will be present at the time cell samples were taken. So, if spliced mRNA transcripts that are more abundant tend to be turned over at a slower rate, then it could be that the apparent inverse correlation between "level of gene expression" and "degree of splicing defect" is actually due to differences in mRNA stability rather than differences in their ability to compete for limiting splicing components. So, to discard this possibility, later on it might be useful to do a similar analysis as Figure 5.4B but using read counts within the intron (un-spliced pre-mRNAs) instead of mRNA junction reads. Either way, there is no easy way to measure splicing efficiency without the influence of RNA turnover rate, except probably by isolating newly synthesized RNA through a fast labelling technique (Barrass et al., 2015). Therefore, although my observations fit well with previously proposed models, they should be interpreted with caution.

Prp22 is important for suppression of suboptimal 3'ss

Through a genome-wide analysis of splicing errors, I showed that the frequency of many non-annotated 3'ss is increased in the absence of Prp22 or its helicase activity (Figure 5.7). This is the first evidence that Prp22 is required for proper 3'ss selection with endogenous gene transcripts *in vivo,* as most previous studies linking RNA helicases with splicing fidelity were based on *in vitro* assays. Furthermore, I showed a few examples where absence of Prp22 leads to a substantial increase in non-productive splicing (Figure 5.9). Although indirect, this is the first evidence to indicate that the role of Prp22 in splicing fidelity is physiologically important.

As only about a fifth of all aberrant 3'ss are more frequent in mutant relative to wild-type, it may be that Prp22 only proofreads a particular set of suboptimal substrates. As mentioned above, because the fidelity function of Prp22 is linked to its essential role in splicing, it is likely that the measured splicing error is an underestimate with respect to the full contribution of Prp22 to splicing fidelity. Therefore, it could also be that the effect of many aberrant 3'ss is masked by the splicing defect.

It was unexpected that, compared to wild-type, many of the aberrant 5'ss have a lower SEF (Figure 5.7) in the mutants, suggesting that splicing fidelity is higher in these cases. I can only speculate that this is due to the reduced pool of splicing components caused by depletion of Prp22. If the availability of the splicing machinery is reduced, then optimal splice sites may compete even better for splicing, thus suboptimal substrates will be spliced less frequently. This can also explain the aberrant 3'ss events with a lower SEF than wild-type.

Errors in 3'ss selection are common

I have confirmed that 3'ss selection in wild-type is particularly prone to errors (Figure 5.5). This is interesting, but probably not surprising given that the consensus 3' motif is short ($_{A}$T/CAG-), compared to the 5' motif ($_{AAG}$-GTATGT$_{T}$) or the branchsite ($_{T}$ACTAAC$_{A}$). In other words, because there is less information to find the appropriate 3'ss, the spliceosome may very often use alternative AG- sequences near the branchsite. This is one reason why 3'ss proofreading may be especially important. On the other hand, typically the first T/CAG- immediately after the branchsite is selected for splicing, so branchsite location constrains 3'ss selection. This explains why there are many more aberrant downstream 3'ss than upstream 3'ss even in wild-type, as the distance between the branchsite and canonical 3'ss is limited. In this sense, it would be interesting to investigate whether Prp22 contributes to the preferential selection of T/CAG- 3'ss that are closer to the branchsite.

To try to understand what type of sub-optimal substrates are targeted by Prp22, I manually inspected the most significant aberrant introns listed in Figure 5.9. Some of these contain upstream 3'ss with non-consensus motif (e.g. AAG-), including *GIM5, YCL002C, RAD14, HNT2, PHO85* transcripts (Figure S5.2). This may suggest that Prp22 specifically inhibits non-consensus 3'ss (NAG-) located between the branchsite and the canonical 3'ss. On the other hand, there are other aberrant introns that contain a consensus 3'ss downstream of the canonical 3'ss, including *LSM7, IWR1* and *PTC7* transcripts. This may suggest that, in addition, Prp22 inhibits downstream 3'ss that may compete with the annotated 3'ss. So, it seems that Prp22 influences 3'ss selection based on motif sequence and also on branch site proximity. Although this is a very preliminary observation based on manual inspection, it does suggest that it is worth performing a systematic analysis as part of ongoing studies.

In search for non GT/AG introns

Canonical 5'ss (-GT) and 3'ss (AG-) motifs represent the vast majority of both canonical and aberrant introns (Kawashima et al., 2014). For this reason, all the analyses described in this chapter, only include aberrant junctions derived from spliced GT/AG introns. As it was previously shown that helicase-deficient Prp22 mutants, but not wild-type Prp22, allow *in vitro* splicing of a non AG- 3'ss (TAC-) (Mayas et al., 2006), I became interested in searching for this type of aberrant introns within my sequencing data. However, I later discovered that STAR mapper does not annotate strandness to the junction output file (SJ.out.tab) for those introns with motifs different than GT/AG, GC/AG and AT/AC. Therefore, they are filtered out in the downstream data analysis.

So, as a crude attempt to search for potentially interesting "non-canonical" introns, I manually inspected the SJ.out.tab files that contain read counts linked to intron coordinates but no gene name annotations. However, I did not find splicing events that seemed promising. The few sequence discontinuities with 1) motifs different than GT/AG and 2) different number of reads in mutants compared to wild-type, do not appear to originate from splicing events, as they flank sequences of < 50 or > 1000 bp, or belong to intron-less protein coding genes or non-coding RNAs. So, it is possible that Prp22 is not particularly required for rejection of aberrant introns containing non GT/AG motifs as was previously observed *in vitro*. However, to rule this out would require a more thorough analysis.

Ongoing studies

One reason why this study is valuable, is that the resulting dataset may allow us to better understand what triggers the rejection of aberrant 3'ss. This will

involve investigating whether the set of aberrantly spliced introns that become more frequent in mutant relative to wild-type, have unusual features. For example, one could look for associations between changes in the frequency of splicing errors and specific 3'ss sequences, secondary structure strength within different regions of the pre-mRNA, distance to the branchsite or canonical 3'ss, among others.

Finally, in the first part of the RNA-seq analysis I attempted to quantify the splicing defect, and to compare it across genes with different expression level. Through this analysis, I produced preliminary evidence of a correlation between expression level and the strength of the splicing defect, which may confirm that highly expressed genes tend to compete better for a limited pool of spliceosomes. However, mRNA stability may influence the mRNA junction ratio of mutant/wild-type (which was my measure of the splicing defect). Therefore, it is worth to complement this previously observed correlation using intron-read density instead of mRNA junction reads. This is an additional analysis that should be done for publication.

Acknowledgements

I would like to acknowledge Beate Schwer's generosity in providing plasmids p360-PRP22-URA3, p358-T757A-TRP1 and p358-I764A-TRP1. The RNA-seq experiment would not have been possible without the advice and guidance of Vahid Aslanzadeh, especially regarding NGS data analysis. I thank him for that, and also for sharing his python program "STARtab2GTF2Juncanot.py" and his python-based notebook "Vahid-altevents.ipynb". Finally, I thank Sander Granneman for sharing his easy-to-use python program "STARtab2GTF.py".

Closing Remarks

Molecular tools to conditional repression gene expression are very useful in *in vivo* functional studies of cell biology. The ideal purpose of these tools is to specifically disrupt the element of interest without causing unwanted secondary effect. The AID is one of few systems that has the potential of accomplishing this. However, others and I have observed that previous version of this system for budding yeast had one of three possible issues: they stressed the cells, had leaky depletion, or depleted slowly. In this regard, one of my most important contribution was the development of the B-estradiol AID. This system does not appear to have any of the issues of the previous systems, and has the added benefit of tuneable depletion rate to accommodate depletion of low or highly abundant target proteins.

However, one limitation of the B-estradiol AID is that it requires the addition of two effectors, and this could increase the chances of causing pleiotropic effects. A global analysis has shown that the transcriptome profile of yeast is virtually unchanged by the addition of B-estradiol to the medium (McIsaac et al., 2013). On the other hand, although addition of auxin does not affect cell growth on normal condition, a good transcriptomic analysis of yeast exposed to auxin under diverse conditions is still lacking. Others and I have observed that exposing yeast to high concentrations of auxin can be detrimental to growth, under conditions of nutrient deprivation (Figure 4.8) or under stress (Prusty et al., 2004). Furthermore, there is evidence that yeast can perceive and respond to auxin as part of a plant-infection pathway (Prusty et al., 2004). Therefore, for future studies it would be both useful and interesting to perform a systematic genome wide analysis of yeast exposed to auxin under different conditions.

Prp22 is an interesting protein because, as other spliceosomal RNA helicases, it directly triggers key structural changes within the spliceosome and, in addition, it was proposed to proofread the second step of splicing catalysis. As most previous studies of Prp22 were based on *in vitro* assays, I became interested in studying the role of this protein *in vivo*. My first observation was that the absence of this protein inhibits the first stages of spliceosome assembly and splicing catalysis. This was initially surprising, as this protein acts at the final stages of splicing. Although many year ago it was reported that disruption of late-acting splicing factors, such as Prp22, inhibits the first step of splicing (Company et al., 1991; Noble and Guthrie, 1996), the explanation given for why this happens has not been well supported by evidence.

Given the gap in the literature, my first objective was to demonstrate why Prp22 appears to be required for the earliest stages of *in vivo* splicing. I later observed that depletion of this protein, or others, correlates with accumulation intermediate complexes of the spliceosome and to a very quick reduction of co-transcriptional assembly of the sequestered splicing components. This strongly indicates that Prp22 is only indirectly required for the first step of splicing, as its absence leads to a recycling defect and, therefore, to reduced formation of new spliceosomes. This finding is important in two ways. Firstly, it helped to better understand the role of Prp22 in splicing *in vivo* and, secondly, it showed that perturbing one element of the spliceosome can cause unwanted systematic effects on the assembly cycle.

Finally, I used the improved AID system to study the proposed role of Prp22 in splicing fidelity *in vivo*, by combining a novel depletion approach with an RNA sequencing analysis. With this, I observed that absence of Prp22 or its ATPase activity correlates with a higher frequency of aberrantly spliced 3'ss compared to wild-type. This result is the first demonstration that Prp22 is

important for proper splice site selection on endogenous genes *in vivo*, which strongly indicates that splicing proofreading is a physiologically relevant process. The challenge of these experiments derived from the fact that, as I previously observed, depletion of this protein indirectly inhibits the first steps of splicing. This meant that the type of aberrantly spliced mRNA transcripts that I aimed to detect, were rare. However, because the transcripts were sequencing with very high-depth, it was possible to measure these rare events, and to compare them with wild-type conditions. The plan for the near future, is to attempt to find common features within the genes that become more affected, to try to better understand the process by which Prp22 rejects suboptimal 3'ss.

In summary, I believe that the molecular tools I generated, together with my observation that strong and quick secondary effects may arise when perturbing essential proteins, will be useful for future functional studies.

Supplementary information

Figure S2.1 Plasmid pURA3-AID*-6FLAG

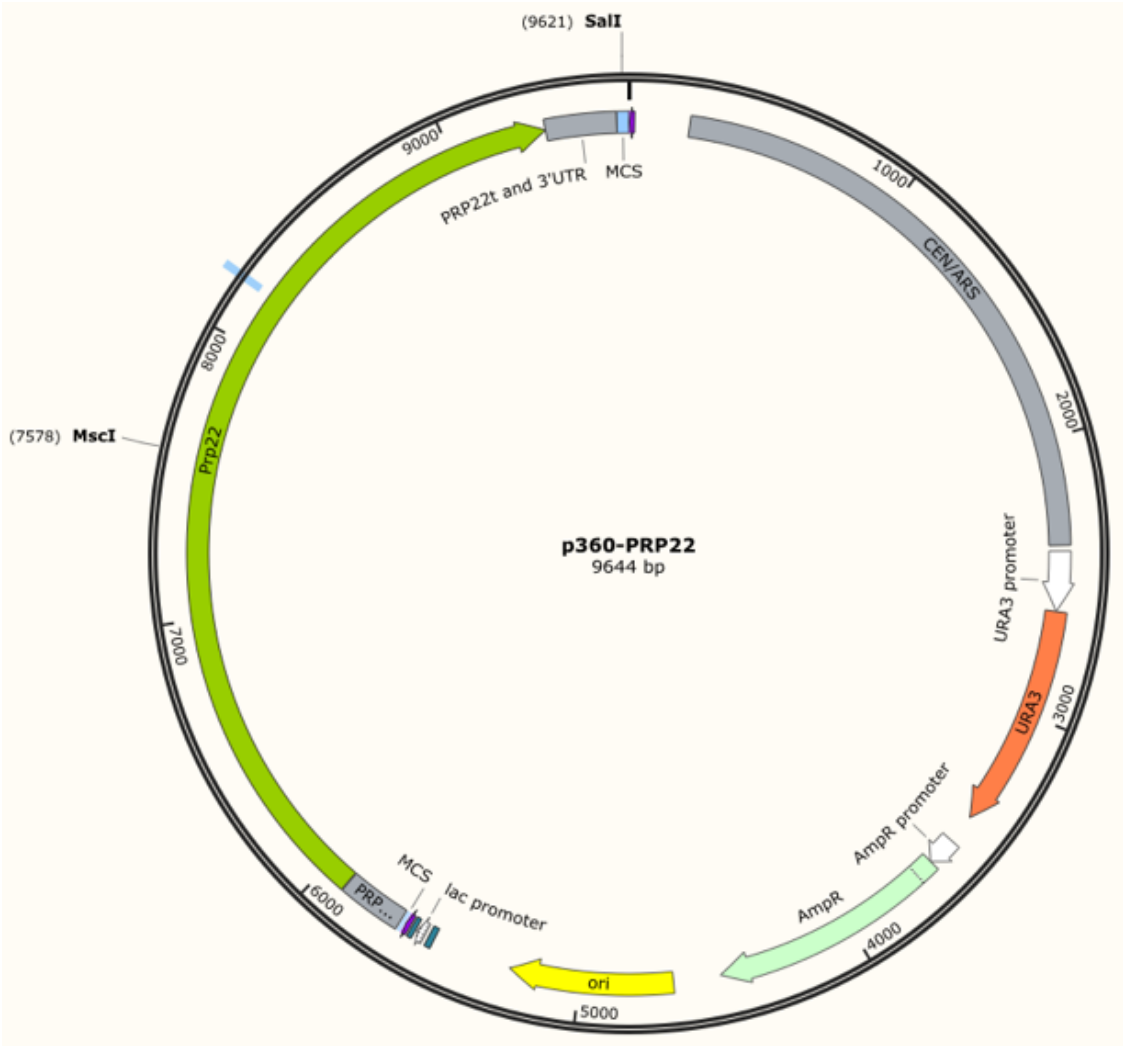
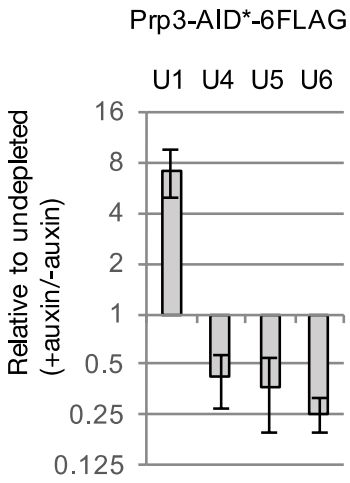
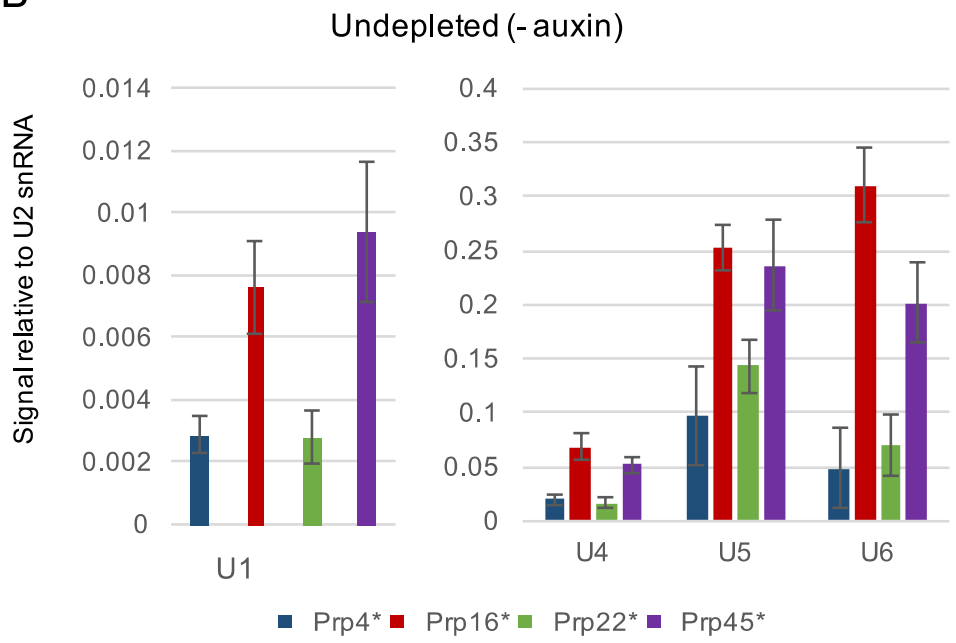


Figure S2.2 Plasmid p360-PRP22

A


Prp3-AID*-6FLAG
U1 U4 U5 U6
Relative to undepleted (+auxin/-auxin)
16
8
4
2
1
0.5
0.25
0.125

B


Undepleted (- auxin)
Signal relative to U2 snRNA
0.014
0.012
0.01
0.008
0.006
0.004
0.002
0
U1
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
U4
U5
U6
Prp4*
Prp16*
Prp22*
Prp45*

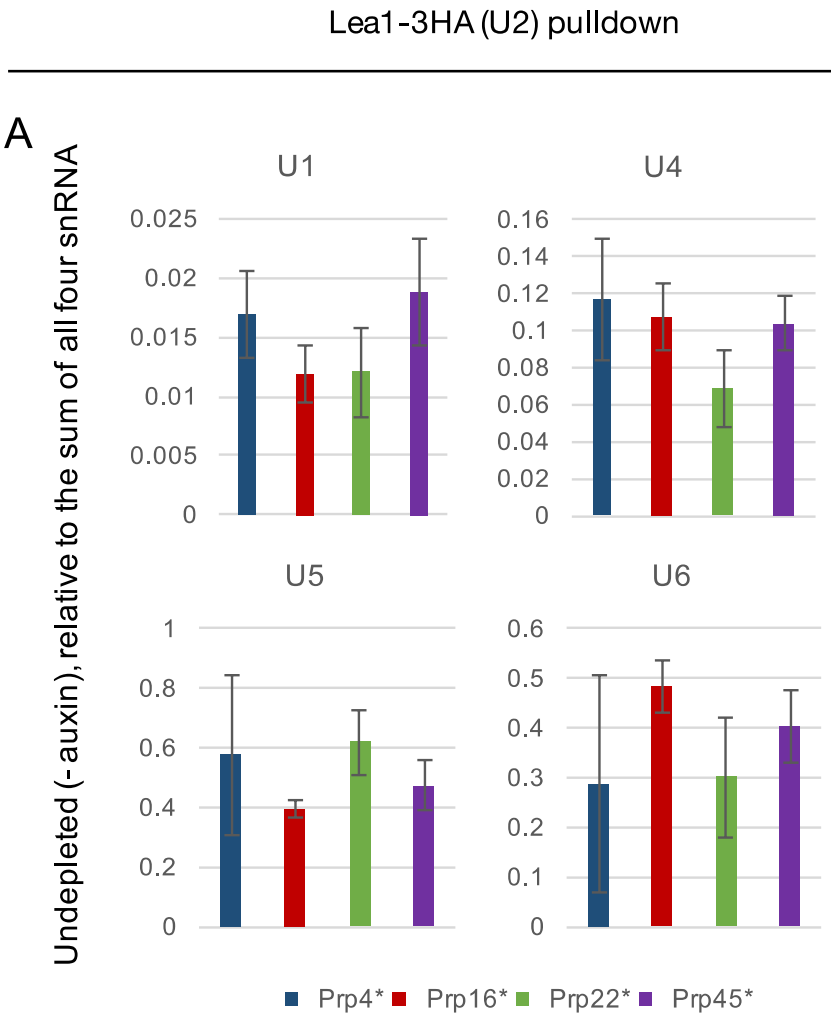
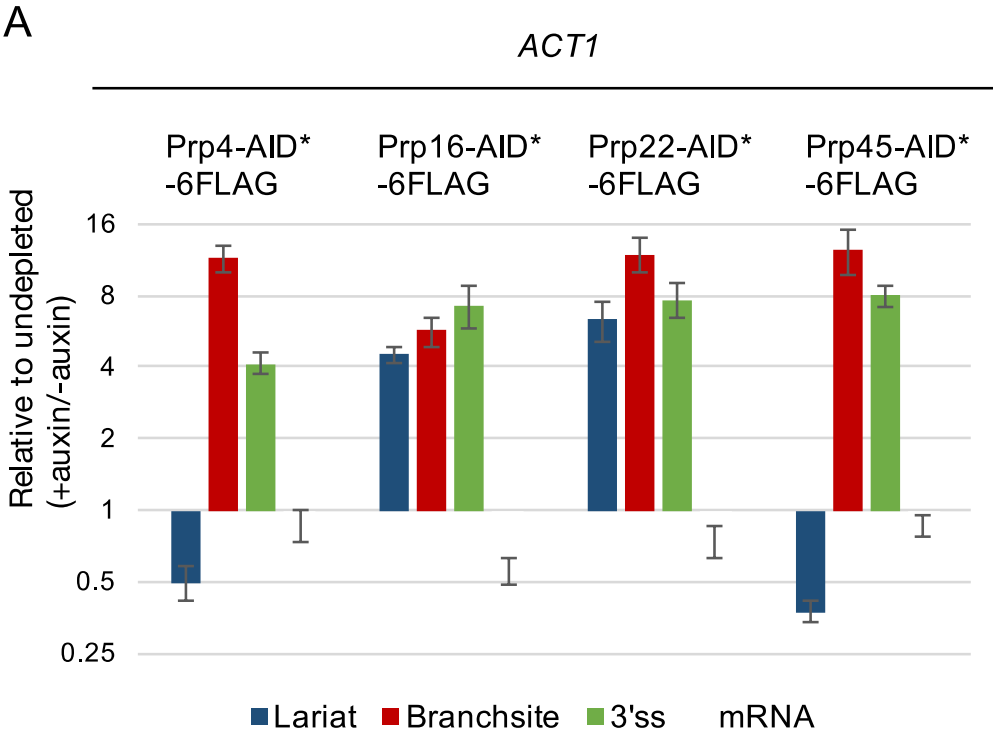
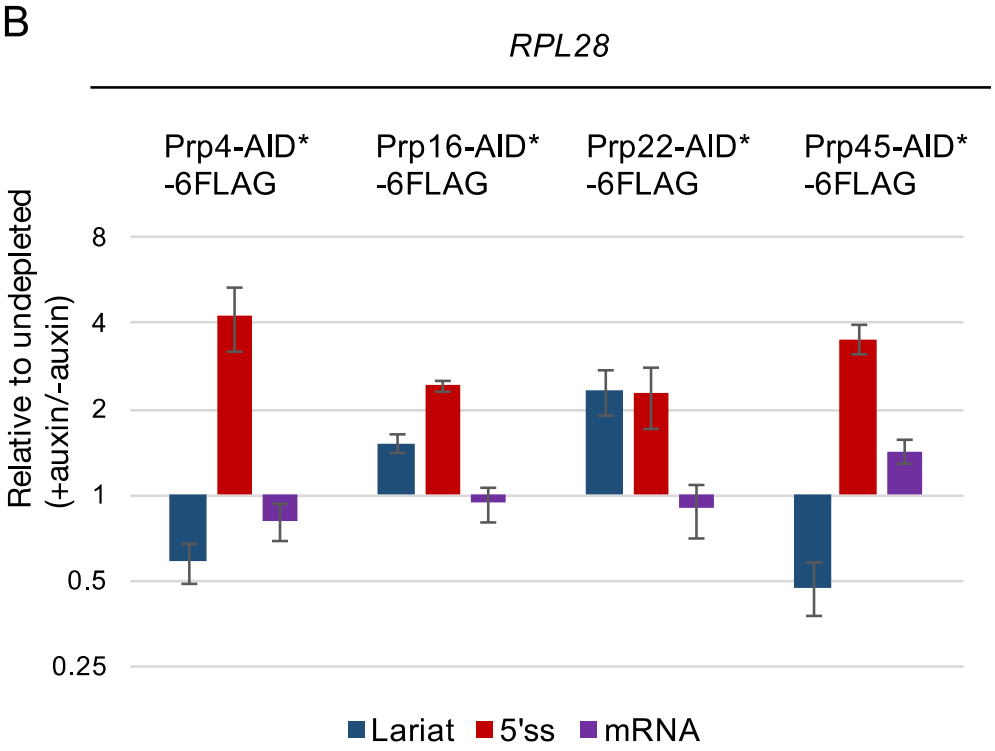


Figure S3.1. Additional analysis of the RIP data relate to Figure 3.3 Data was normalized to U2 snRNA signal to correct for differences in pull-down efficiencies. Error bars denote standard error of three biological replicates.
(A) When Prp3-AID*-6FLAG (tri-snRNP) is depleted, interaction of Lea1-3HA (U2 snRNP) with U1 is increased, and with U4, U5 and U6 snRNAs is decreased, like in the Prp4-AID*-6FLAG depletion.
(B) Starting values (before auxin-induced depletion) of the -AID*-6FLAG strain, demonstrating the necessity of normalized the data and presenting it as the ratio of depleted/undepleted, since experiments done at different days can produce values of different scales.
(C) Same data as in (B) but presented as relative to the sum of U1, U4 U5 and U6 (U2 normalized), to correct for differences in scale. This allows one to compare, between the different strains, the level of interaction of U2 with the other snRNAs before the addition of auxin.

A


ACT1
Prp4-AID*
-6FLAG
Prp16-AID*
-6FLAG
Prp22-AID*
-6FLAG
Prp45-AID*
-6FLAG
Relative to undepleted
(+auxin/-auxin)
16
8
4
2
1
0.5
0.25
Lariat
Branchsite
3'ss
mRNA

B


RPL28
Prp4-AID*
-6FLAG
Prp16-AID*
-6FLAG
Prp22-AID*
-6FLAG
Prp45-AID*
-6FLAG
Relative to undepleted
(+auxin/-auxin)
8
4
2
1
0.5
0.25
Lariat
5'ss
mRNA

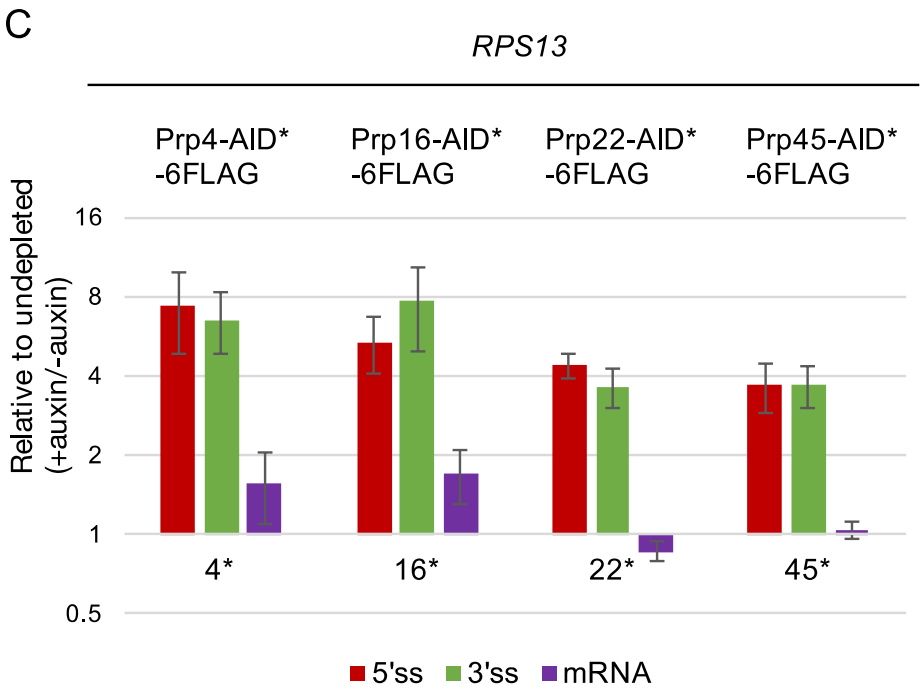


Figure S3.2. Data as in Figure 3.5 presented as signal relative to exon 2 instead of normalized to ALG9. This normalization corrects for potential differences in expression rate (after auxin-induced depletion) and is another way at looking at splicing efficiency.

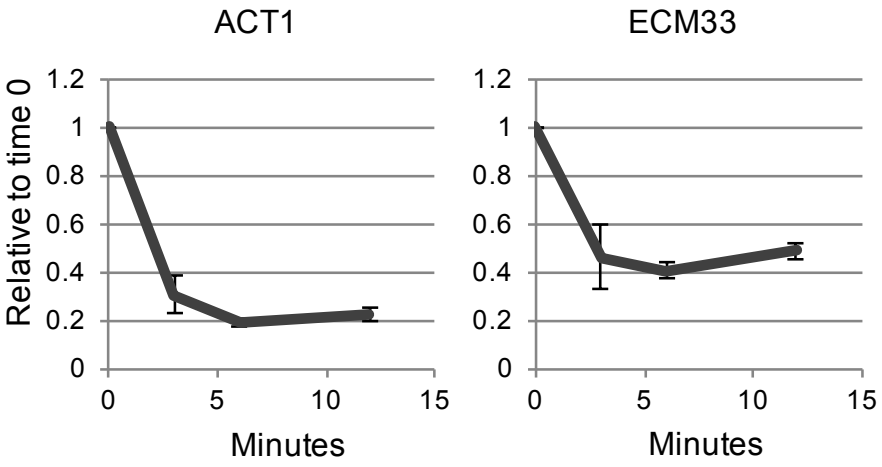


Figure S3.3. Anti-FLAG ChIP of Prp22-AID*-6FLAG depletion, related to Figure 3.7. This confirms that after induced depletion, Prp22-AID*-6FLAG engaged in co-transcriptional splicing, drops as quickly as total levels of Prp22-AID*-6FLAG (as shown in **Figure 3.6.** by western blotting). Only *ACT1* and *ECM33* are shown because, with the primers available at the time of the experiment, ChIP of Prp22-AID*-6FLAG to *RPL28*, *RPS13* or *RPL39* was unsuccessful (data not shown).

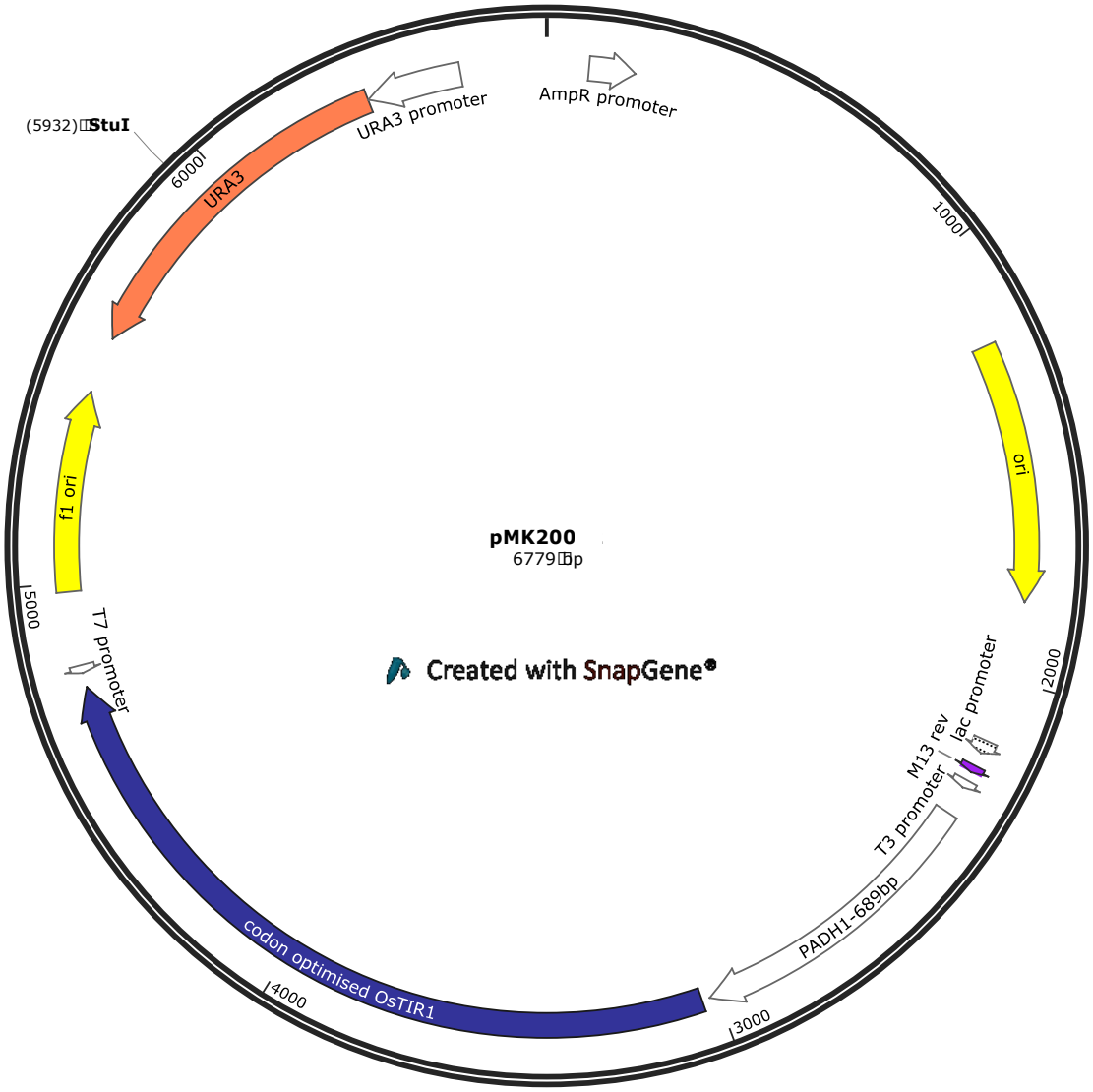


Figure S4.1. Map of plasmid pMK200. StuI linearize pMK200 (Kanemaki lab) was inserted into *ura3-1* of W303 to produce strain YGM1.

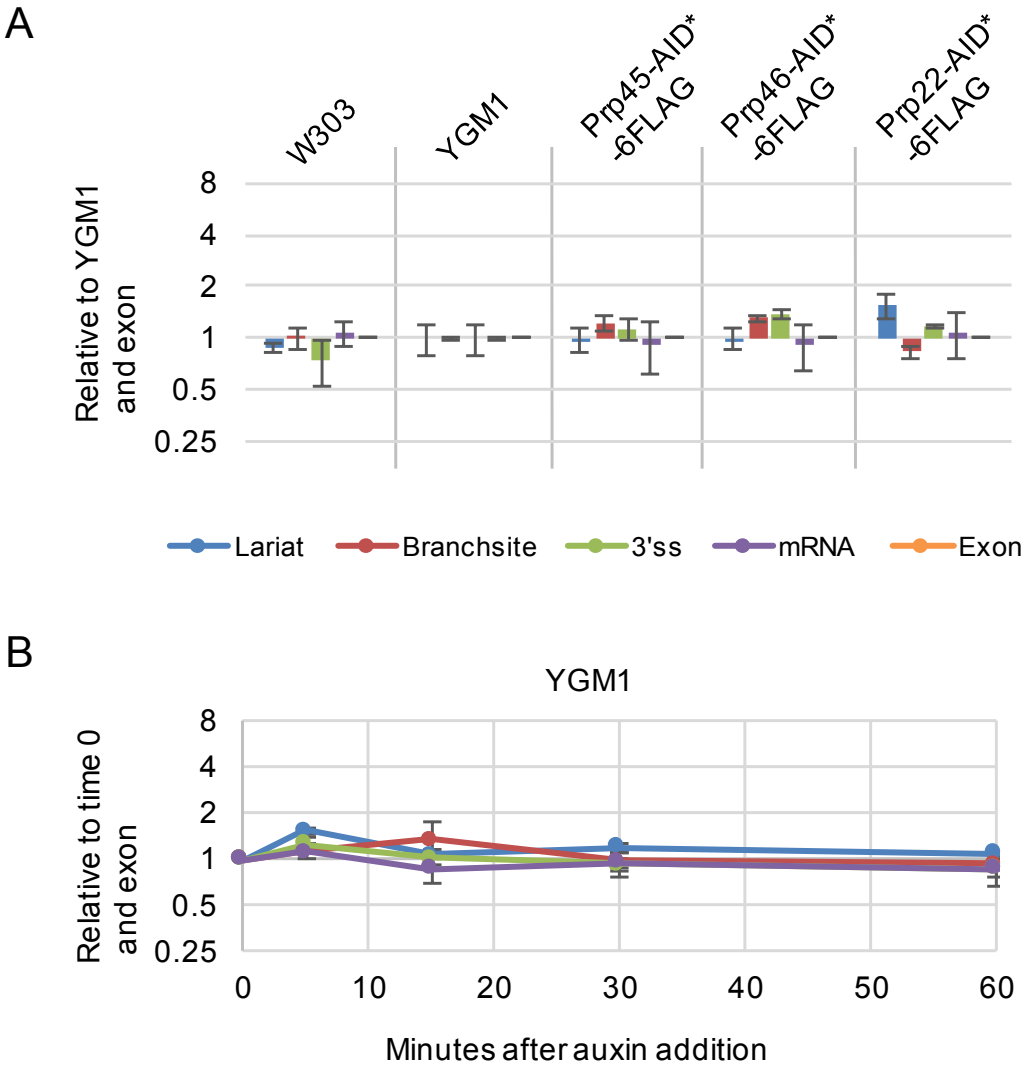


Figure S4.2. Abundance of splicing intermediates of *ACT1* transcript of YGM1 strain and its derivatives. Related to Figure 4.2.
(A) Wild-type strain W303 and AID-tagged strains compared with their untagged parental strain YGM1. Data are presented as relative to YGM1 and normalized to exon.
(B) Effect of auxin addition on YGM1 strain. Data are presented as relative to time 0 and normalized to exon.
Error bars denote standard error of three biological replicates.

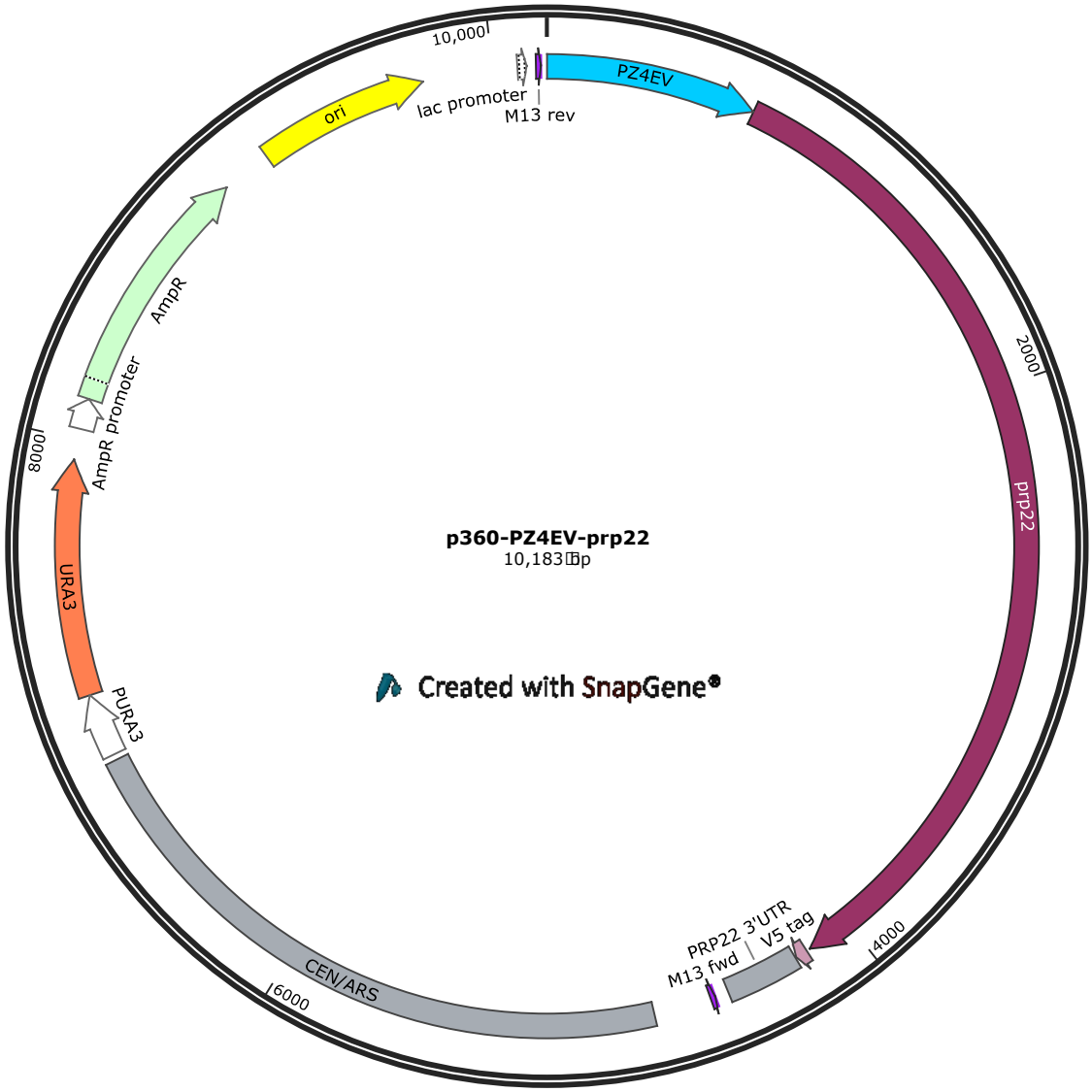


Figure S5.1. Map of plasmids p360-PZ4EV-prp22 (-WT, -T757A or -I764A). Series of centromeric plasmids for the B-estradiol induced expression of Prp22 wild-type, or Prp22 mutants T757A or I764A.

Figure S5.2. Related to Figure 5.9. Alt intron = non-annotated intron, BS = branchsite, blue or red boxes contain the 3'ss motif.

Non-annotated splicing event			Reads in non-annotated spliced junction								Reads in annotated spliced junction							
Gene	**Event**	**Shift**	**57_1**	**57_2**	**64_1**	**64_2**	**d_1**	**d_2**	**wt_1**	**wt_2**	**57_1**	**57_2**	**64_1**	**64_2**	**d_1**	**d_2**	**wt_1**	**wt_2**
GIM5	up-3ss	22	5	1	7	4	24	19	0	5	165	131	150	135	228	120	429	412
RAD14	up-3ss	13	9	0	6	3	17	28	9	6	45	57	61	62	93	66	298	253
LSM7	down-3ss	7	42	75	77	51	227	172	55	56	239	268	249	213	339	259	954	881
HNT2	up-3ss	26	5	1	2	0	14	9	7	1	21	36	33	30	78	70	236	199
YCL002C	up-3ss	22	12	17	21	24	61	59	73	66	19	11	8	7	37	26	228	196
PHO85	up-3ss	29	7	14	23	20	37	42	13	15	136	144	149	148	289	277	562	577
IWR1	down-3ss	32	3	7	6	0	6	3	4	5	10	10	8	11	27	23	136	108
DID4	down-3ss	20	34	30	7	9	37	22	15	16	146	172	157	186	375	278	808	652
VPS75	up-3ss	8	15	44	23	38	100	72	82	56	143	188	175	158	272	245	628	723
PTC7	down-3ss	10	21	49	48	37	64	57	61	57	121	121	108	83	258	205	528	520

57 = T757A mutant, 64 = I764A mutant, d = Prp22 depletion, wt = wild-type

Figure S5.3. Reads of non-annotated and annotated spliced junctions of "Top-10" hits (related to Figure 5.9), of the two biological replicates of each experiment

Bibliography

Abovich, N., and Rosbash, M. (1997). Cross-intron bridging interactions in the yeast commitment complex are conserved in mammals. Cell *89*, 403–412.

Achsel, T., Brahms, H., Kastner, B., Bachi, A., Wilm, M., and Lührmann, R. (1999). A doughnut-shaped heteromer of human Sm-like proteins binds to the 3'-end of U6 snRNA, thereby facilitating U4/U6 duplex formation in vitro. EMBO J. *18*, 5789–5802.

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

Alexander, R., Barrass, J., Dichtl, B., Kos, M., Obtulowicz, T., Marie-cecile, R., Koper, M., Karkusiewicz, I., Mariconti, L., Tollervey, D., et al. (2010b). RiboSys, a high-resolution, quantitative approach to measure the in vivo kinetics of pre-mRNA splicing and 3′-end processing in Saccharomyces cerevisiae. RNA *16*, 2570–2580.

Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc.

Arenas, J.E., and Abelson, J.N. (1997). Prp43: An RNA helicase-like factor involved in spliceosome disassembly. Biochemistry *94*, 11798–11802.

Ares, M., Grate, L., Pauling, M.H., Ares, M., Grate, L., and Pauling, M.H. (1999). A handful of intron-containing genes produces the lion's share of yeast mRNA. RNA *5*, 1138–1139.

Aronova, A., Bacíková, D., Crotti, L.B., Horowitz, D.S., and Schwer, B. (2007). Functional interactions between Prp8, Prp18, Slu7, and U5 snRNA during the second step of pre-mRNA splicing. RNA *13*, 1437–1444.

Ayadi, L., Callebaut, I., Saguez, C., Villa, T., Mornon, J.-P., and Banroques, J. (1998). Functional and structural characterization of the Prp3 binding domain of the yeast Prp4 splicing factor. J. Mol. Biol. *284*, 673–687.

Bae, E., Reiter, N.J., Bingman, C.A., Kwan, S.S., Lee, D., Jr, G.N.P., Butcher, S.E., and Brow, D.A. (2007). Structure and Interactions of the First Three RNA Recognition Motifs of Splicing Factor Prp24. *367*, 1447–1458.

Banroques, J., and Abelson, J.N. (1989). PRP4: a protein of the yeast U4/U6 small nuclear ribonucleoprotein particle. Mol. Cell. Biol. *9*, 3710–3719.

Barrass, J.D., Reid, J.E.A., Huang, Y., Hector, R.D., Sanguinetti, G., Beggs, J.D., and Granneman, S. (2015). Transcriptome-wide RNA processing kinetics revealed using extremely short 4tU labeling. Genome Biol *16*, 282.

Behrens, S.E., Tyc, K., Kastner, B., Reichelt, J., and Lührmann, R. (1993). Small nuclear ribonucleoprotein (RNP) U2 contains numerous additional proteins and has a bipartite RNP structure under splicing conditions. Mol. Cell. Biol. *13*, 307–319.

Behzadnia, N., Golas, M.M., Hartmuth, K., Sander, B., Kastner, B., Deckert, J., Dube, P., Will, C.L., Urlaub, H., Stark, H., et al. (2007). Composition and three-dimensional EM structure of double affinity-purified, human prespliceosomal A complexes. EMBO J. *26*, 1737–1748.

Bellí, G., Garí, E., Piedrafita, L., Aldea, M., and Herrero, E. (1998). An activator/repressor dual system allows tight tetracycline-regulated gene expression in budding yeast. Nucleic Acids Res. *26*, 942–947.

Bergkessel, M., Whitworth, G.B., and Guthrie, C. (2011). Diverse environmental stresses elicit distinct responses at the level of pre-mRNA processing in yeast. RNA *17*, 1461–1478.

Bertram, K., Agafonov, D.E., Liu, W.-T., Dybkov, O., Will, C.L., Hartmuth, K., Urlaub, H., Kastner, B., Stark, H., and Lührmann, R. (2017). Cryo-EM structure of a human spliceosome activated for step 2 of splicing. Nature *542*, 318–323.

Braberg, H., Jin, H., Moehle, E.A., Chan, Y.A., Wang, S., Shales, M., Benschop, J.J., Morris, J.H., Qiu, C., Hu, F., et al. (2013). From structure to systems: high-resolution, quantitative genetic analysis of RNA polymerase II. Cell *154*, 775–788.

Brosi, R., Gröning, K., Behrens, S.E., Lührmann, R., and Krämer, A. (1993). Interaction of mammalian splicing factor SF3a with U2 snRNP and relation of its 60-kD subunit to yeast PRP9. Science *262*, 102–105.

Brugiolo, M., Herzel, L., and Neugebauer, K.M. (2013). Counting on co-transcriptional splicing. F1000Prime Rep. *5*, 9.

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

Burgess, S., Couto, J.R., and Guthrie, C. (1990). A putative ATP binding protein influences the fidelity of branchpoint recognition in yeast splicing. Cell *60*, 705–717.

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

Carrillo Oesterreich, F., Herzel, L., Straube, K., Hujer, K., Howard, J., and Neugebauer, K.M. (2016). Splicing of Nascent RNA Coincides with Intron Exit from RNA Polymerase II. Cell *165*, 372–381.

Chan, S.P., and Cheng, S.C. (2005). The Prp19-associated complex is required for specifying interactions of U5 and U6 with pre-mRNA during spliceosome activation. J. Biol. Chem. *280*, 31190–31199.

Chan, S.P., Kao, D.I., Tsai, W.Y., and Cheng, S.C. (2003). The Prp19p-associated complex in spliceosome activation. Science *302*, 279–282.

Chanarat, S., Seizl, M., and Strässer, K. (2011). The Prp19 complex is a novel transcription elongation factor required for TREX occupancy at transcribed genes. Genes Dev. *25*, 1147–1158.

Chanclud, E., and Morel, J. (2016). Review Plant hormones : a fungal point of view. 1–9.

Chang, J.S., and McPheeters, D.S. (2000). Identification of a U2/U6 helix Ia mutant that influences 3' splice site selection during nuclear pre-mRNA splicing. RNA *6*, 1120–1130.

Chathoth, K.T., Barrass, J.D., Webb, S., and Beggs, J.D. (2014). A splicing-dependent transcriptional checkpoint associated with prespliceosome formation. Mol. Cell *53*, 779–790.

Chen, C.-H., Kao, D.-I., Chan, S.-P., Kao, T.-C., Lin, J.-Y., and Cheng, S.-C. (2006). Functional links between the Prp19-associated complex, U4/U6 biogenesis, and

spliceosome recycling. RNA *12*, 765–774.

Chen, J.Y.F., Stands, L., Staley, J.P., Jackups, R.R., Latus, L.J., and Chang, T.H. (2001). Specific alterations of U1-C protein or U1 Small nuclear RNA can eliminate the requirement of Prp28p, an essential DEAD Box splicing factor. Mol. Cell *7*, 227–232.

Chiang, T.-W., and Cheng, S.-C. (2013). A weak spliceosome-binding domain of Yju2 functions in the first step and bypasses Prp16 in the second step of splicing. Mol. Cell. Biol. *33*, 1746–1755.

Chorev, M., and Carmel, L. (2012). The function of introns. Front. Genet. *3*, 1–15.

Chung, S., Mclean, M.R., and Rymond, B.C. (1999). Yeast ortholog of the Drosophila crooked neck protein promotes spliceosome assembly through stable U4/U6.U5 snRNP addition. RNA *5*, 1042–1054.

Company, M., Arenas, J., and Abelson, J.N. (1991). Requirement of the RNA helicase-like protein PRP22 for release of messenger RNA from spliceosomes. Nature *349*, 487–493.

Cordin, O., and Beggs, J.D. (2013). RNA helicases in splicing. RNA Biol. *10*, 83–95.

Couto, J.R., Tamm, J., Parker, R., and Guthrie, C. (1987). A trans-acting suppressor restores splicing of a yeast intron with a branch point mutation. Genes Dev. *1*, 445–455.

David, C.J., Boyne, A.R., Millhouse, S.R., and Manley, J.L. (2011). The RNA polymerase II C-terminal domain promotes splicing activation through recruitment of a U2AF65-Prp19 complex. Genes Dev. *25*, 972–983.

Didychuk, A.L., Montemayor, E.J., Brow, D.A., and Butcher, S.E. (2016). Structural requirements for protein-catalyzed annealing of U4 and U6 RNAs during di-snRNP assembly. Nucleic Acids Res. *44*, 1398–1410.

Dobin, A., and Gingeras, T. (2016). Mapping RNA-seq Reads with STAR. Curr. Protoc. Bioinform. 51:11.14.1-11.14.19. doi: 10.1002/0471250953.bi1114s51.

Dohmen, R.J., and Varshavsky, A. (2005). Heat-Inducible Degron and the Making of Conditional Mutants. In Methods in Enzymology, pp. 799–822.

Egecioglu, D.E., and Chanfreau, G. (2011). Proofreading and spellchecking: a two-tier strategy for pre-mRNA splicing quality control. RNA *17*, 383–389.

English, M.A., Lei, L., Blake, T., Wincovitch, S.M., Sood, R., Azuma, M., Hickstein, D., and Liu, P.P. (2012). Incomplete splicing, cell division defects, and hematopoietic blockage in dhx8 mutant zebrafish. Dev. Dyn. *241*, 879–889.

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

Fica, S.M., Tuttle, N., Novak, T., Li, N.-S., Lu, J., Koodathingal, P., Dai, Q., Staley, J.P., and Piccirilli, J.A. (2013). RNA catalyses nuclear pre-mRNA splicing. Nature *503*, 229–234.

Fica, S.M., Oubridge, C., Galej, W.P., Wilkinson, M.E., Bai, X.-C., Newman, A.J., and Nagai, K. (2017). Structure of a spliceosome remodelled for exon ligation. Nature *542*, 377–380.

Fourmann, J.B., Schmitzová, J., Christian, H., Urlaub, H., Ficner, R., Boon, K.L., Fabrizio,

P., and Lührmann, R. (2013). Dissection of the factor requirements for spliceosome disassembly and the elucidation of its dissociation products using a purified splicing system. Genes Dev. *27*, 413–428.

Frank, D., and Guthrie, C. (1992). An essential splicing factor, SLU7, mediates 3' splice site choice in yeast. Genes Dev. *6*, 2112–2124.

Fu, S.-F., Wei, J.-Y., Chen, H.-W., Liu, Y.-Y., Lu, H.-Y., and Chou, J.-Y. (2015). Indole-3-acetic acid: A widespread physiological code in interactions of fungi with other organisms. Plant Signal. Behav. *10*, e1048052.

Galej, W.P., Oubridge, C., Newman, A.J., and Nagai, K. (2013). Crystal structure of Prp8 reveals active site cavity of the spliceosome. Nature *493*, 638–643.

Galej, W.P., Nguyen, T.H.D., Newman, A.J., and Nagai, K. (2014). Structural studies of the spliceosome: Zooming into the heart of the machine. Curr. Opin. Struct. Biol. *25*, 57–66.

Galej, W.P., Wilkinson, M.E., Fica, S.M., Oubridge, C., Newman, A.J., and Nagai, K. (2016). Cryo-EM structure of the spliceosome immediately after branching. Nature *537*, 197–201.

Garí, E., Piedrafita, L., Aldea, M., and Herrero, E. (1997). A Set of Vectors with a Tetracycline-Regulatable Promoter System for Modulated Gene Expression inSaccharomyces cerevisiae. Yeast *13*, 837–848.

Gautam, A., Grainger, R.J., Vilardell, J., Barrass, J.D., and Beggs, J.D. (2015). Cwc21p promotes the second step conformation of the spliceosome and modulates 3′ splice site selection. Nucleic Acids Res. *43*, 3309–3317.

Ghetti, A., Company, M., and Abelson, J.N. (1995). Specifity of Prp24 binding to RNA: a role for Prp24 in the dynamic interaction of U4 and U6 snRNAs. RNA *1*, 132–145.

Gietz, R.D., and Schiestl, R.H. (2007). High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. Nat. Protoc. *2*, 31–34.

Görnemann, J., Kotovic, K.M., Hujer, K., and Neugebauer, K.M. (2005). Cotranscriptional spliceosome assembly occurs in a stepwise fashion and requires the cap binding complex. Mol. Cell *19*, 53–63.

Gossen, M., and Bujard, H. (1992). Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. Proc. Natl. Acad. Sci. U. S. A. *89*, 5547–5551.

Grainger, R.J., and Beggs, J.D. (2005). Prp8 protein: At the heart of the spliceosome. RNA *11*, 533–557.

Gray, W.M., Pozo, J.C., Walker, L., Hobbie, L., Risseeuw, E., Banks, T., Crosby, W.L., Yang, M., Ma, H., and Estelle, M. (1999). Identification of an SCF ubiquitin–ligase complex required for auxin response in Arabidopsis thaliana William. Genes Dev *53*, 1678–1691.

Hang, J., Wan, R., Yan, C., and Shi, Y. (2015). Structural basis of pre-mRNA splicing. Science *349*, 1191–1198.

Haruki, H., Nishikawa, J., Laemmli, U.K., Young, R.A., Sali, A., Blobel, G., Frank, J., Hager, G., Bustin, M., and Misteli, T. (2008). The anchor-away technique: rapid, conditional establishment of yeast mutant phenotypes. Mol. Cell *31*, 925–932.

Hilleren, P.J., and Parker, R. (2003). Cytoplasmic Degradation of Splice-Defective Pre-mRNAs and Intermediates. Mol. Cell *12*, 1453–1465.

Hollands, K., Proshkin, S., Sklyarova, S., Epshtein, V., Mironov, A., Nudler, E., and Groisman, E.A. (2012). Riboswitch control of Rho-dependent transcription termination. Proc. Natl. Acad. Sci. U. S. A. *109*, 5376–5381.

Hossain, M.A., Claggett, J.M., Edwards, S.R., Shi, A., Pennebaker, S.L., Cheng, M.Y., Hasty, J., and Johnson, T.L. (2016). Posttranscriptional Regulation of Gcr1 Expression and Activity Is Crucial for Metabolic Adjustment in Response to Glucose Availability. Mol. Cell *62*, 346–358.

Howe, K.J., Kane, C.M., and Ares, M. (2003). Perturbation of transcription elongation influences the fidelity of internal exon inclusion in Saccharomyces cerevisiae. RNA 993–1006.

James, S.-A., Turner, W., and Schwer, B. (2002). How Slu7 and Prp18 cooperate in the second step of yeast pre-mRNA splicing. RNA *8*, 1068–1077.

Juneau, K., Palm, C., Miranda, M., and Davis, R.W. (2007). High-density yeast-tiling array reveals previously undiscovered introns and extensive regulation of meiotic splicing. Proc. Natl. Acad. Sci. *104*, 1522–1527.

Juneau, K., Nislow, C., and Davis, R.W. (2009). Alternative Splicing of PTC7 in Saccharomyces cerevisiae Determines Protein Localization. Genetics *183*, 185–194.

Kanemaki, M.T. (2013). Frontiers of protein expression control with conditional degrons. Pflugers Arch. *465*, 419–425.

Kawashima, T., Douglass, S., Gabunilas, J., Pellegrini, M., and Chanfreau, G.F. (2014). Widespread Use of Non-productive Alternative Splice Sites in Saccharomyces cerevisiae. PLoS Genet. *10*, e1004249.

Konarska, M.M., and Query, C.C. (2005). Insights into the mechanisms of splicing: more lessons from the ribosome. Genes Dev. *19*, 2255–2260.

Koodathingal, P., and Staley, J.P. (2013). Splicing fidelity. RNA Biol. *10*, 1073–1079.

Koodathingal, P., Novak, T., Piccirilli, J.A., and Staley, J.P. (2010). The DEAH box ATPases Prp16 and Prp43 cooperate to proofread 5' splice site cleavage during Pre-mRNA splicing. Mol. Cell *39*, 385–395.

Koonin, E. V (2006). The origin of introns and their role in eukaryogenesis: a compromise solution to the introns-early versus introns-late debate? Biol. Direct *1*, 22.

Kotovic, K., Lockshon, D., Boric, L., and Neugebauer, K.M. (2003). Cotranscriptional Recruitment of the U1 snRNP to Intron-Containing Genes in Yeast. Mol. Cell. Biol. *23*, 5768–5779.

Kötter, P., Weigand, J.E., Meyer, B., Entian, K.-D., and Suess, B. (2009). A fast and efficient translational control system for conditional expression of yeast genes. Nucleic Acids Res. *37*, e120.

Kresnowati, M.T.A.P., Van Winden, W.A., Almering, M.J.H., Ten Pierick, A., Ras, C., Knijnenburg, T.A., Daran-Lapujade, P., Pronk, J.T., Heijnen, J.J., and Daran, J.M. (2006). When transcriptome meets metabolome: fast cellular responses of yeast to sudden relief of glucose limitation. Mol. Syst. Biol. *2*, 49.

Lacadie, S.A., and Rosbash, M. (2005). Cotranscriptional spliceosome assembly dynamics and the role of U1 snRNA:5'ss base pairing in yeast. Mol. Cell *19*, 65–75.

Laggerbauer, B., Achsel, T., and Lührmann, R. (1998). The human U5-200kD DEXH-box protein unwinds U4/U6 RNA duplices in vitro. Proc. Natl. Acad. Sci. U. S. A. *95*, 4188–4192.

Lardelli, R.M., Thompson, J.X., Yates, J.R., and Stevens, S.W. (2010). Release of SF3 from the intron branchpoint activates the first step of pre-mRNA splicing. RNA *16*, 516–528.

Lesser, C.F., and Guthrie, C. (1993). Mutations in U6 snRNA that alter splice site specificity: implications for the active site. Science *262*, 1982–1988.

Listerman, I., Sapra, A.K., and Neugebauer, K.M. (2006). Cotranscriptional coupling of splicing factor recruitment and precursor messenger RNA splicing in mammalian cells. Nat. Struct. Mol. Biol. *13*, 815–822.

Lygerou, Z., Christophides, G., and Séraphin, B. (1999). A novel genetic screen for snRNP assembly factors in yeast identifies a conserved protein, Sad1p, also required for pre-mRNA splicing. Mol. Cell. Biol. *19*, 2008–2020.

Makarov, E.M., Makarova, O. V, Urlaub, H., Gentzel, M., Will, C.L., Wilm, M., and Lührmann, R. (2002). Small nuclear ribonucleoprotein remodeling during catalytic activation of the spliceosome. Science *298*, 2205–2208.

Mao, X., Hu, Y., Liang, C., and Lu, C. (2002). MET3 promoter: A tightly regulated promoter and its application in construction of conditional lethal strain. Curr. Microbiol. *45*, 37–40.

Martin, A., Schneider, S., and Schwer, B. (2002). Prp43 is an essential RNA-dependent ATPase required for release of lariat-intron from the spliceosome. J. Biol. Chem. *277*, 17743–17750.

Martin-tumasz, S., Richie, A.C., Ii, L.J.C., Brow, D.A., and Butcher, S.E. (2011). A novel occluded RNA recognition motif in Prp24 unwinds the U6 RNA internal stem loop. *39*, 7837–7847.

Mata, M. De, Alonso, C.R., Fededa, J.P., Pelisch, F., Cramer, P., Bentley, D., and Kornblihtt, A.R. (2003). A slow RNA Polymerase II affects alternative splicing in vivo. Mol. Cell *12*, 525–532.

Matera, A.G., and Wang, Z. (2014). A day in the life of the spliceosome. Nat. Rev. Mol. Cell Biol. *15*, 108–121.

Maya, D., Quintero, M.J., de la Cruz Muñoz-Centeno, M., and Chávez, S. (2008). Systems for applied gene control in Saccharomyces cerevisiae. Biotechnol. Lett. *30*, 979–987.

Mayas, R.M., Maita, H., and Staley, J.P. (2006). Exon ligation is proofread by the DExD/H-box ATPase Prp22p. Nat. Struct. Mol. Biol. *13*, 482–490.

Mayas, R.M., Maita, H., Semlow, D.R., and Staley, J.P. (2010). Spliceosome discards intermediates via the DEAH box ATPase Prp43p. Proc. Natl. Acad. Sci. U. S. A. *107*, 10020–10025.

Mayes, A.E., Verdone, L., Legrain, P., and Beggs, J.D. (1999). Characterization of Sm-like proteins in yeast and their association with U6 snRNA. EMBO J. *18*, 4321–4331.

McCracken, S., Fong, N., and Yankulov, K. (1997). The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. Nature *385*, 357–361.

McIsaac, R.S., Oakes, B.L., Wang, X., Dummit, K. a, Botstein, D., and Noyes, M.B. (2013). Synthetic gene expression perturbation systems with rapid, tunable, single-gene specificity in yeast. Nucleic Acids Res. *41*, e57.

McPheeters, D.S., and Muhlenkamp, P. (2003). Spatial Organization of Protein-RNA Interactions in the Branch Site-3 Splice Site Region during pre-mRNA Splicing in Yeast. Mol. Cell. Biol. *23*, 4174–4186.

Montemayor, E.J., Curran, E.C., Hong Hong, L., Andrews, K.L., Treba, C.N., Butcher, S.E., and Brow, D.A. (2014). Core structure of the U6 small nuclear ribonucleoprotein at 1.7-Å resolution. Nat. Struct. Mol. Biol. *21*, 544–551.

Morawska, M., and Ulrich, H.D. (2013). An expanded tool kit for the auxin-inducible degron system in budding yeast. Yeast *30*, 341–351.

Morris, D.P., and Greenleaf, A.L. (2000). The splicing factor, Prp40, binds the phosphorylated carboxyl-terminal domain of RNA polymerase II. J. Biol. Chem. *275*, 39935–39943.

Munding, E.M., Shiue, L., Katzman, S., Donohue, J., and Ares, M. (2013). Competition between Pre-mRNAs for the splicing machinery drives global regulation of splicing. Mol. Cell *51*, 338–348.

Newman, A.J., and Norman, C. (1992). U5 snRNA interacts with exon sequences at 5' and 3' splice sites. Cell *68*, 743–754.

Nguyen, G.T.D.T., Scaife, M.A., Helliwell, K.E., and Smith, A.G. (2016a). Role of riboswitches in gene regulation and their potential for algal biotechnology. J. Phycol. *52*, 320–328.

Nguyen, T.H.D., Galej, W.P., Bai, X., Savva, C.G., Newman, A.J., Scheres, S.H.W., and Nagai, K. (2015). The architecture of the spliceosomal U4/U6.U5 tri-snRNP. Nature *523*, 47–52.

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

Nishimura, K., Fukagawa, T., Takisawa, H., Kakimoto, T., and Kanemaki, M. (2009). An auxin-based degron system for the rapid depletion of proteins in nonplant cells. Nat. Methods *6*, 917–922.

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

Ohi, M.D., Link, a. J., Ren, L., Jennings, J.L., McDonald, W.H., and Gould, K.L. (2002). Proteomics Analysis Reveals Stable Multiprotein Complexes in Both Fission and Budding Yeasts Containing Myb-Related Cdc5p/Cef1p, Novel Pre-mRNA Splicing Factors, and snRNAs. Mol. Cell. Biol. *22*, 2011–2024.

Ohrt, T., Odenwälder, P., Dannenberg, J., Prior, M., Warkocki, Z., Schmitzová, J., Karaduman, R., Gregor, I., Enderlein, J., Fabrizio, P., et al. (2013). Molecular dissection of step 2 catalysis of yeast pre-mRNA splicing investigated in a purified system. RNA *19*, 902–915.

Osheim, Y.N., Miller, O.L., and Beyer, A.L. (1985). RNP particles at splice junction

sequences on Drosophila chorion transcripts. Cell *43*, 143–151.

Pan, Q., Shai, O., Lee, L.J., Frey, B.J., and Blencowe, B.J. (2008). Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. Nat. Genet. *40*, 1413–1415.

Papasaikas, P., Tejedor, J.R., Vigevani, L., and Valcarcel, J. (2015). Functional splicing network reveals extensive regulatory potential of the core spliceosomal machinery. Mol. Cell *57*, 7–22.

Parenteau, J., Durand, M., Véronneau, S., Lacombe, A.-A., Morin, G., Guérin, V., Cecez, B., Gervais-Bird, J., Koh, C.-S., Brunelle, D., et al. (2008). Deletion of many yeast introns reveals a minority of genes that require splicing for function. Mol. Biol. Cell *19*, 1932–1941.

Parenteau, J., Durand, M., Morin, G., Gagnon, J., Lucier, J.F., Wellinger, R.J., Chabot, B., and Elela, S.A. (2011). Introns within ribosomal protein genes regulate the production and function of yeast ribosomes. Cell *147*, 320–331.

Pleiss, J.A., Whitworth, G.B., Bergkessel, M., and Guthrie, C. (2007). Rapid, Transcript-Specific Changes in Splicing in Response to Environmental Stress. Mol. Cell *27*, 928–937.

Prusty, R., Grisafi, P., and Fink, G.R. (2004). The plant hormone indoleacetic acid induces invasive growth in Saccharomyces cerevisiae. Proc. Natl. Acad. Sci. *101*, 4153–4157.

Raghunathan, P.L., and Guthrie, C. (1998a). RNA unwinding in U4/U6 snRNPs requires ATP hydrolysis and the DEIH-box splicing factor Brr2. Curr. Biol. *8*, 847–855.

Raghunathan, P.L., and Guthrie, C. (1998b). A spliceosomal recycling factor that reanneals U4 and U6 small nuclear ribonucleoprotein particles. Science *279*, 857–860.

Rao, R.P., Hunter, A., Kashpur, O., and Normanly, J. (2010). Aberrant synthesis of indole-3-acetic acid in Saccharomyces cerevisiae triggers morphogenic transition, a virulence trait of pathogenic fungi. Genetics *185*, 211–220.

Ronen, M., and Botstein, D. (2006). Transcriptional response of steady-state yeast cultures to transient perturbations in carbon source. Proc. Natl. Acad. Sci. U. S. A. *103*, 389–394.

van Roon, A.-M.A.M., Oubridge, C., Obayashi, E., Sposito, B., Newman, A.J., Seraphin, B., and Nagai, K. (2017). Crystal structure of U2 snRNP SF3b components: Hsh49p in complex with Cus1p binding domain. RNA *23*, 968–981.

Ryan, D.E., Stevens, S.W., and Abelson, J. (2002). The 5’ and 3’ domains of yeast U6 snRNA: Lsm proteins facilitate binding of Prp24 protein to the U6 telestem region. RNA *8*, 1011–1033.

Dos Santos Maraschin, F., Memelink, J., and Offringa, R. (2009). Auxin-induced, SCFTIR1-mediated poly-ubiquitination marks AUX/IAA proteins for degradation. Plant J. *59*, 100–109.

Schneider, C., Agafonov, D.E., Schmitzová, J., Hartmuth, K., Fabrizio, P., and Lührmann, R. (2015). Dynamic Contacts of U2, RES, Cwc25, Prp8 and Prp45 Proteins with the Pre-mRNA Branch-Site and 3’ Splice Site during Catalytic Activation and Step 1 Catalysis in Yeast Spliceosomes. PLOS Genet. *11*, e1005539.

Schneider, S., Campodonico, E., and Schwer, B. (2004). Motifs IV and V in the DEAH box splicing factor Prp22 are important for RNA unwinding, and helicase-defective Prp22 mutants are suppressed by Prp8. J. Biol. Chem. *279*, 8617–8626.

Schwer, B. (2008). A conformational rearrangement in the spliceosome sets the stage for Prp22-dependent mRNA release. Mol. Cell *30*, 743–754.

Schwer, B., and Gross, C.H. (1998). Prp22, a DExH-box RNA helicase, plays two distinct roles in yeast pre-mRNA splicing. EMBO J. *17*, 2086–2094.

Schwer, B., and Guthrie, C. (1991). PRP16 is an RNA-dependent ATPase that interacts transiently with the spliceosome. Nature *349*, 494–499.

Semlow, D.R., and Staley, J.P. (2012). Staying on message: Ensuring fidelity in pre-mRNA splicing. Trends Biochem. Sci. *37*, 263–273.

Semlow, D.R., Blanco, M.R., Walter, N.G., and Staley, J.P. (2016). Spliceosomal DEAH-Box ATPases Remodel Pre-mRNA to Activate Alternative Splice Sites. Cell *164*, 985–998.

Shcherbakova, I., Hoskins, A. a., Friedman, L.J., Serebrov, V., Corrêa, I.R., Xu, M.Q., Gelles, J., and Moore, M.J. (2013). Alternative Spliceosome Assembly Pathways Revealed by Single-Molecule Fluorescence Microscopy. Cell Rep. *5*, 151–165.

Da Silva, N.A., and Srikrishnan, S. (2012). Introduction and expression of genes for metabolic engineering applications in Saccharomyces cerevisiae. FEMS Yeast Res. *12*, 197–214.

Small, E.C., Leggett, S.R., Winans, A. a., and Staley, J.P. (2006). The EF-G-like GTPase Snu114p Regulates Spliceosome Dynamics Mediated by Brr2p, a DExD/H Box ATPase. Mol. Cell *23*, 389–399.

Song, E.J., Werner, S.L., Neubauer, J., Stegmeier, F., Aspden, J., Rio, D., Harper, J.W., Elledge, S.J., Kirschner, M.W., and Rape, M. (2010). The Prp19 complex and the Usp4Sart3 deubiquitinating enzyme control reversible ubiquitination at the spliceosome. Genes Dev. *24*, 1434–1447.

Sontheimer, E.J., and Steitz, J.A. (1993). The U5 and U6 small nuclear RNAs as active site components of the spliceosome. Science *262*, 1989–1996.

Staley, J.P., and Guthrie, C. (1999). An RNA Switch at the 5' Splice Site Requires ATP and the DEAD Box Protein Prp28p. Mol. Cell *3*, 55–64.

Stevens, S.W., and Abelson, J. (1999). Purification of the yeast U4/U6.U5 small nuclear ribonucleoprotein particle and identification of its proteins. Proc. Natl. Acad. Sci. U. S. A. *96*, 7226–7231.

Tanaka, S., Miyazawa-Onami, M., Lida, T., and Araki, H. (2015). iAID: an improved auxin-inducible degron system for the construction of a "tight" conditional mutant in the budding yeast Saccharomyces cerevisiae. Yeast *32*, 567–581.

Tardiff, D., and Rosbash, M. (2006). Arrested yeast splicing complexes indicate stepwise snRNP recruitment during in vivo spliceosome assembly. RNA *12*, 968–979.

Tardiff, D., Lacadie, S., and Rosbash, M. (2006). A genome-wide analysis indicates that yeast pre-mRNA splicing is predominantly posttranscriptional. Mol. Cell *24*, 917–929.

Tarn, W.Y., Lee, K.R., and Cheng, S.C. (1993). Yeast precursor mRNA processing protein PRP19 associates with the spliceosome concomitant with or just after dissociation of U4 small nuclear RNA. Proc. Natl. Acad. Sci. U. S. A. *90*, 10821–10825.

Tarn, W.Y., Hsu, C.H., Huang, K.T., Chen, H.R., Kao, H.Y., Lee, K.R., and Cheng, S.C.

(1994). Functional association of essential splicing factor(s) with PRP19 in a protein complex. EMBO J. *13*, 2421–2431.

Taxis, C., and Knop, M. (2012). TIPI: TEV Protease-Mediated Induction of Protein Instability. In Methods in Molecular Biology (Clifton, N.J.), pp. 611–626.

Teigelkamp, S., Newman, A.J., and Beggs, J.D. (1995). Extensive interactions of PRP8 protein with the 5' and 3' splice sites during splicing suggest a role in stabilization of exon alignment by U5 snRNA. EMBO J. *14*, 2602–2612.

Tsai, R.-T., Tseng, C.-K., Lee, P.-J., Chen, H.-C., Fu, R.-H., Chang, K.-J., Yeh, F.-L., and Cheng, S.-C. (2007). Dynamic Interactions of Ntr1-Ntr2 with Prp43 and with U5 Govern the Recruitment of Prp43 To Mediate Spliceosome Disassembly. Mol. Cell. Biol. *27*, 8027–8037.

Tsai, R.T., Fu, R.H., Yeh, F.L., Tseng, C.K., Lin, Y.C., Huang, Y.H., and Cheng, S.C. (2005). Spliceosome disassembly catalyzed by Prp43 and its associated components Ntr1 and Ntr2. Genes Dev. *19*, 2991–3003.

Tseng, C.-K., Liu, H.-L., and Cheng, S.-C. (2011). DEAH-box ATPase Prp16 has dual roles in remodeling of the spliceosome in catalytic steps. RNA *17*, 145–154.

Umen, J.G., and Guthrie, C. (1996). Mutagenesis of the yeast gene PRP8 reveals domains governing the specificity and fidelity of 3' splice site selection. Genetics *143*, 723–739.

Verdone, L., Galardi, S., Page, D., and Beggs, J.D. (2004). Lsm Proteins Promote Regeneration of Pre-mRNA Splicing Activity. Curr. Biol. *14*, 1487–1491.

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

Villa, T., and Guthrie, C. (2005). The Isy1p component of the NineTeen Complex interacts with the ATPase Prp16p to regulate the fidelity of pre-mRNA splicing. Genes Dev. *19*, 1894–1904.

Wagner, J.D., Jankowsky, E., Company, M., Pyle, A.M., and Abelson, J.N. (1998). The DEAH-box protein PRP22 is an ATPase that mediates ATP-dependent mRNA release from the spliceosome and unwinds RNA duplexes. EMBO J. *17*, 2926–2937.

Wallace, E.W.J., and Beggs, J.D. (2016). Extremely fast and incredibly close: co-transcriptional splicing in budding yeast. RNA *23*, 601–610.

Wan, R., Yan, C., Bai, R., Huang, G., and Shi, Y. (2016a). Structure of a yeast catalytic step I spliceosome at 3.4 Å resolution. Science *353*, 895–904.

Wan, R., Yan, C., Bai, R., Wang, L., Huang, M., Wong, C.C.L., and Shi, Y. (2016b). The 3.8 Å structure of the U4/U6.U5 tri-snRNP: Insights into spliceosome assembly and catalysis. Science. *351*, 466–475.

Wang, E.T., Sandberg, R., Luo, S., Khrebtukova, I., Zhang, L., Mayr, C., Kingsmore, S.F., Schroth, G.P., and Burge, C.B. (2008). Alternative isoform regulation in human tissue transcriptomes. Nature *456*, 470–476.

Warkocki, Z., Odenwälder, P., Schmitzová, J., Platzmann, F., Stark, H., Urlaub, H., Ficner, R., Fabrizio, P., and Lührmann, R. (2009). Reconstitution of both steps of Saccharomyces cerevisiae splicing with purified spliceosomal components. Nat. Struct. Mol. Biol. *16*, 1237–1243.

Weinhandl, K., Winkler, M., Glieder, A., and Camattari, A. (2014). Carbon source dependent promoters in yeasts. Microb. Cell Fact. *13*, 1–17.

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

Will, C., and Lührmann, R. Spliceosome structure and function. Cold Spring Harb Perspect Biol doi: 10.1101/cshperspect.a003707.

Wong, J.J.L., Ritchie, W., Ebner, O.A., Selbach, M., Wong, J.W.H., Huang, Y., Gao, D., Pinello, N., Gonzalez, M., Baidya, K., et al. (2013). Orchestrated intron retention regulates normal granulocyte differentiation. Cell *154*, 583–595.

Xu, Y.Z., and Query, C.C. (2007). Competition between the ATPase Prp5 and Branch Region-U2 snRNA Pairing Modulates the Fidelity of Spliceosome Assembly. Mol. Cell *28*, 838–849.

Xu, D., Nouraini, S., Field, D., Tand, S.-J., and Friesen, J.D. (1996). An RNA-dependent ATPase associated with U2/U6 snRNAs in pre-mRNA splicing.pdf. Nature *381*, 709–713.

Yan, C., Hang, J., Wan, R., Huang, M., Wong, C.C.L., and Shi, Y. (2015). Structure of a yeast spliceosome at 3.6-angstrom resolution. Science *349*, 1182–1191.

Yan, C., Wan, R., Bai, R., Huang, G., and Shi, Y. (2016). Structure of a yeast step II catalytically activated spliceosome. Science. 10.1126/science.aak9979.

Yean, S.L., Wuenschell, G., Termini, J., and Lin, R.J. (2000). Metal-ion coordination by U6 small nuclear RNA contributes to catalysis in the spliceosome. Nature *408*, 881–884.