THE UNIVERSITY
of EDINBURGH

Links between splicing, transcription and chromatin in *Saccharomyces cerevisiae*

Isabella Eileen Maudlin

A thesis submitted to the University of Edinburgh for the degree of Doctor of Philosophy

Supervisor: Professor Jean Beggs

September 2018

Table of contents

Acknowledgements

I would like to thank Professor Jean Beggs who provided me with enormous support and guidance throughout my PhD, who was always available for advice and who gave me the freedom to develop independent scientific thinking. It has been a great privilege and I have learned so much from Jean to take forward to the next stage of my career.

I am grateful to Dr Ema Sani and Dr Susana de Lucas for their time and patience in teaching me laboratory skills, for their guidance and help and for all the fun times we have shared from my first rotation project to the end of my PhD.

I would like to thank all members of the Beggs lab (past and present) for their help and support: Eve Hartswood, Dr Gonzalo Mendoza-Ochoa, Dr Vahid Aslandazeh, Dr Edward Wallace, Dr Shiney George, Julia Neuimeir and Charlotte Capitanchik. I would like to thank David Barrass for his patience in teaching me thiolabelling of nascent RNA and for providing help whenever I needed it. I would like to thank Vincent Géli, for initiating the collaboration between Prp22 and Set1, and for great discussions. Thanks also to Dr Alastair Kerr, Dr Gabriele Schweikert and Dr Guido Sanguinetti for helping with informatic analysis.

I would like to thank my thesis committee – Professor David Tollervey, Dr Sander Granneman and Dr Philipp Voigt were a great external source of discussion and Karen Traill for keeping me organised and helping me throughout my studies.

A special thanks to Jamie, my partner, who has been a rock for me throughout our years together and particularly during my PhD – you gave me courage when I needed it most and I will never forget your support. Special thanks also to my parents and my sisters for inspiring me and supporting me no matter what.

Last but not least, I would like to gratefully acknowledge the Wellcome Trust for funding my PhD studies.

Declaration

The work described in this thesis was undertaken between July 2015 and September 2018, in the Wellcome Centre for Cell Biology at The University of Edinburgh. The contribution of others to this work has been explicitly stated and acknowledged, otherwise this thesis is a product of my own work. No part of this dissertation has been, or will be, submitted for any degree, diploma or other professional qualification at the University of Edinburgh or any other university or similar institution.

Isabella Eileen Maudlin

September 2018

Abstract

There is increasing evidence from yeast to humans that splicing is mainly a co-transcriptional process, and it is becoming well established that splicing, transcription and chromatin are functionally coupled such that they influence one another. The present work explored the links between splicing and transcription and links between splicing and chromatin in the budding yeast *Saccharomyces cerevisiae*.

Currently, there is little mechanistic insight into the contribution of the core transcription elongation machinery to co-transcriptional spliceosome assembly and splicing. To understand how members of the core transcription elongation machinery affect splicing, I used the auxin-inducible degron (AID) system to conditionally deplete essential and non-essential transcription elongation factors and I analysed the effects on RNA polymerase II, co-transcriptional spliceosome assembly and splicing. The transcription elongation factors that I analysed are all conserved from yeast to mammals and include: Spt5, Paf1, Ctk1, Bur1 and Bur2. Most notable were the effects of depletion of the transcription elongation factor Spt5, mutations in which were known to cause splicing defects. Here, Spt5 depletion resulted in reduced recruitment of the U5 snRNP to intron-containing genes, meaning proper co-transcriptional activation of the spliceosome was inhibited, explaining how loss or mutation of Spt5 results in splicing defects. This effect was not dependent on phosphorylation of Spt5, however, the unphosphorylated form of Spt5 enhanced co-transcriptional formation of the catalytically activated spliceosome. Together, these data indicate a two-part function for Spt5 in co-transcriptional spliceosome assembly in *S. cerevisiae*. Firstly, the physical presence of Spt5 is required for proper co-transcriptional recruitment or stable association of the U5 snRNP and B complex formation. Secondly, the loss of Bur1 kinase activity and at least the unphosphorylated form of Spt5 enhances co-transcriptional formation of the catalytically activated spliceosome and splicing.

There is correlative and causative evidence that splicing affects chromatin structure and *vice versa*. Of particular interest to the present work are links between splicing and Histone 3 Lysine 4 trimethylation (H3K4me3), a chromatin mark associated with promoters of active genes. H3K4me3 has been shown to influence and be influenced

by splicing in mammalian cells. However, the molecular basis of this is unknown. To further understand the links between splicing and H3K4me3, I used the AID system to conditionally deplete essential splicing factors that act at different stages of the splicing cycle and analysed the effects on H3K4me3. Whilst depletion of splicing factors that affect the first or second catalytic step of splicing reduces H3K4me3 on intron-containing genes, notably, depletion of the late-acting factor Prp22 reduces H3K4me3 in the absence of defects in splicing catalysis, suggesting a more direct role for Prp22. Prp22 is an RNA-dependent ATPase that proofreads to product of the second step of splicing and promotes mRNA release from the post-spliceosome. Interestingly, the effect of Prp22 on H3K4me3 is dependent on its ATPase activity. Furthermore, Prp22 and Set1 were found to interact in a pull-down assay and depletion of Prp22 results in reduced recruitment of Set1 to intron-containing genes. These data show a previously unknown link between Prp22, Set1 and H3K4me3 in *S. cerevisiae*.

Collectively, these analyses provide new mechanistic insight into the links between splicing and transcription and links between splicing and chromatin in *S. cerevisiae*.

Lay summary

The DNA of most organisms is comprised of a mixture of both non-coding sequences ('introns') and coding sequences ('exons'). In order to make a protein, the DNA sequence must first be copied into RNA during a process called transcription. Before a functional protein is made, the non-coding sequences (introns) must be removed in a complex biochemical process called splicing. Splicing is conserved from simple unicellular organisms such as yeasts to more complex multicellular organisms like humans, and failures in splicing can be of great consequence to the functioning of individual cells and whole organisms. From yeast to humans, splicing occurs during transcription ('co-transcriptionally') of DNA into RNA, and interestingly there are known links between splicing and transcription such that these processes can influence one another. These links between splicing and transcription are important for splicing outcome. During my PhD I have tested the effects of loss of proteins that regulate transcription on splicing in the model yeast *Saccharomyces cerevisiae,* which has revealed novel links between core proteins involved in transcription and co-transcriptional splicing.

Interestingly, in addition to splicing and transcription influencing one another, it is known that there are links between splicing and the higher-order structure of DNA. DNA is not a naked molecule and is packaged by proteins into a higher order structure called chromatin. Unlike the DNA sequence, the way in which DNA is packed is highly changeable – and various biochemical modifications determine this, giving rise new levels of regulation of gene expression. Splicing occurs in close-proximity to chromatin, and it has been shown that splicing and chromatin affect each other, potentially serving as a splicing 'memory' in order to facilitate future rounds of splicing. During my PhD studies I have tested the effects of inhibiting splicing on chromatin structure in the model yeast *Saccharomyces cerevisiae* and found a particular stage of splicing that is important for proper levels of a particular chromatin mark.

List of abbreviations

%	percentage
℃	degrees Celsius
3'SS	3' splice site
5'SS	5' splice site
ADP	adenosine diphosphate
AID	auxin-inducible degron
ATP	adenosine triphosphate
bp	base pair
BP	branchpoint
CDK	cyclin dependent kinase
cDNA	complementary deoxyribonucleic acid
ChIP	chromatin immunoprecipitation
CTD	carboxy-terminal domain
CTR	c-terminal repeat domain
ddH_2O	double distilled H_2O
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
DNase	deoxyribonuclease
dNTP	deoxynucleotide
EDTA	ethylenediamine tetraacetic acid
Ex 2	exon 2
g	grams
g	relative centrifugal force
gDNA	genomic DNA
H3K36	histone H3 lysine 36
H3K4	histone H3 lysine 4
HA	hemagglutinin
IAA	indole-3-acetic acid
IgA	immunoglobulin A
IgG	immunoglobulin G
IP	immunoprecipitation
kDa	kilodalton
M	molar
me1	monomethylation
me2	dimethylation
me3	trimethylation
mg	milligram(s)
min	minute(s)
ml	millilitre(s)
mM	millimolar
mRNA	messenger RNA
NET	native elongating transcript
OD_{600}	optical density wavelength 600 nm
ORF	open reading frame
PCR	polymerase chain reaction

PEG	polyethylene glycol
pre-mRNA	precursor messenger RNA
qPCR	quantitative polymerase chain reaction
RNA	ribonucleic acid
RNA	ribonucleic acid
RNAi	RNA interference
RNAPI	RNA polymerase I
RNAPII	RNA polymerase II
RNAPIII	RNA polymerase III
RNase	ribonuclease
RPM	revolutions per minute
RT	reverse transcription
RT-qPCR	reverse transcription quantitative polymerase chain reaction
S2P	phosphorylated serine 2
S5P	phosphorylated serine 5
SDS	sodium dodecyl sulfate
snRNA	small nuclear ribonucleic acid
snRNP	small nuclear ribonucleic protein particle
Spt5-P	phosphorylated Spt5
ssDNA	salmon sperm deoxyribonucleic acid
TCA	trichloroacetic acid
tRNA	transfer RNA
TSS	transcription start site
U	enzyme units
V	volts
W.R.T.	with respect to
w/v	weight/volume
WT	wild-type
YMM	yeast minimal media
YPDA	yeast/peptone/dextrose/adenine
µL	microliters
µM	micromolar

Chapter 1. Introduction

1.1 Gene expression in eukaryotes

DNA encodes the information to make functional proteins, which are the building blocks of life, but does not direct protein synthesis. In a process called transcription, DNA is copied into RNA and the RNA molecule serves as a template for protein synthesis or translation. The fact that DNA makes RNA and RNA makes protein is known as the central dogma of molecular biology. Transcription in the nucleus generates precursor RNA (pre-mRNA) which is processed in various ways in order to generate a mature RNA (mRNA) that is exported to the cytoplasm for translation into a protein. RNA processing in eukaryotes is coupled to transcription. The first RNA processing event is capping at the 5' end of the pre-mRNA transcript, during which a 5' methyl cap is added to the newly made RNA and helps to distinguish protein-coding RNAs from other types of RNA, in addition to preventing premature degradation at the 5' end, regulating nuclear export and enhancing translation. Next, certain sequences are removed from the pre-mRNA.

Genes of most eukaryotes are comprised of a mixture of non-coding sequences ('introns') flanked by coding sequences ('exons'). Introns and exons are transcribed and are present in pre-mRNA, and in order to generate a functional protein, introns must be excised and exons ligated together in a process called pre-mRNA splicing. The 3' end of the pre-mRNA is processed by cleavage followed by polyadenylation, during which approximately 200 adenosine nucleotides are added to the 3' end. This gives the RNA a poly(A) tail that promotes export and translation, as well as preventing premature degradation. The pre-mRNA can then be exported from the nucleus to the cytoplasm, where it is translated into a protein by the ribosome. Eventually, the mRNA is degraded in a step-wise process: degradation is triggered by the shortening of the poly(A) tail (deadenylation) by exoribonucleases, followed by mRNA decapping and degradation in a 5' to 3' and/or 3' to 5' direction.

1.2 Transcription in eukaryotes

In eukaryotes, DNA is transcribed into RNA by three RNA polymerase (RNAP) enzymes: RNAPI, RNAPII and RNAPIII, each transcribing a specific gene type. In contrast, prokaryotes have only one RNAP enzyme. RNAPI transcribes genes encoding ribosomal RNAs (28S, 18S, and 5.8S) which collectively comprise the majority of RNA in a cell. RNAPIII transcribes genes encoding transfer RNAs (tRNAs) and other small RNAs, including the smallest ribosomal RNA. RNAPII transcribes genes encoding all proteins (mRNAs), micro RNAs, small nuclear RNAs (snRNAs) (except for U6 which is transcribed by RNAPIII) and small nucleolar RNAs (snoRNAs). All three RNAP enzymes are multi-subunit complexes, for example, RNAPII is a large (approximately 500 kDa) complex comprised of 12 distinct subunits that are conserved from yeast to humans (Rpb1-Rpb12) (reviewed in Liu *et al*., 2013).

Broadly, there are three main stages of RNAPII transcription: initiation, elongation and termination (reviewed in Shandilya *et al*., 2012). During initiation, RNAPII is recruited by general transcription factors and positioned at promoters of the gene to be transcribed and synthesises short sequences of RNA. After which, RNAPII undergoes conformational rearrangements that allow transcription elongation into the gene. At the end of the gene, RNAPII encounters a DNA sequence that is transcribed into RNA, which signals 3'end cleavage and ultimately termination of transcription. Nascent RNA transcripts synthesised by RNAPII after cleavage are degraded by a 5'-3' exonuclease (as they lack a 5' cap) which eventually triggers the release of RNAPII from the DNA template. It is important to understand the context in which RNAPII transcription occurs - in eukaryotes, DNA is packaged into nucleosomes that form a barrier to transcribing RNAPII.

1.3 Transcription through chromatin

Eukaryotic genomes are packaged into chromatin, the basic unit of which is the nucleosome. Packaging DNA into nucleosomes enables the genome to fit into the confined space of the nucleus. Nucleosomes are comprised of 147 base pairs of DNA

that are wrapped 1.65 times around histone octamers – two heterodimers of H3-H4 and two heterodimers of H2A-H2B (Luger *et al*., 1997). Histones make extensive contacts with DNA *via* positively charged amino acids – the core histones are enriched in positively charged (lysine and arginine) amino acids that bind the negatively charged DNA backbone, so that any DNA sequence can potentially be bound by histone proteins. Nucleosomes form repetitive arrays that are analogous to "beads on a string" (reviewed in Struhl and Segal, 2013). Each nucleosome within the array is separated from the next nucleosome by a sequence of unwrapped linker DNA that varies from 20 to 90 base pairs in length, depending on species and even cell type (reviewed in Szerlong and Hansen, 2011). The nucleosome arrays do not remain in this linear "beads on a string" arrangement; nucleosomes are compacted on top of one another to varying degrees of compaction (reviewed in Li and Reinberg, 2011). The highly condensed form of chromatin is known as heterochromatin, and the less condensed form is known as euchromatin. Heterochromatin is a repressive environment for transcription whereas euchromatin is a transcriptionally permissive environment as DNA that is more tightly packaged is more inaccessible by RNAPII in comparison to a more open chromatin structure.

Though necessary to compact the genome, nucleosomes present a physical barrier to RNAPII elongation. *In vivo*, promoter regions are relatively depleted of nucleosomes, and this enables recruitment of transcription factors to certain DNA sequences that facilitate recruitment of RNAPII (Yuan *et al*., 2005; reviewed in Zaret and Carroll, 2011). Once recruited, RNAPII faces a barrier of ordered nucleosomes present in gene bodies that RNAPII must overcome for productive transcription elongation. Even on a "naked" DNA template, RNAPII stalls and backtracks (Dangkulwanich *et al*., 2013), therefore transcription through the nucleosome barrier presents a challenge to RNAPII, especially as a single nucleosome is sufficient to slow RNAPII transcription and induce RNAPII pausing *in vitro* (Knezetic and Luse, 1986; Lorch *et al*., 1987; Kireeva *et al*., 2005; Hodges *et al*., 2009). In order to transcribe the nucleosome barrier, histone chaperone proteins (such as the FACT complex) associated with histones and RNAPII loosen the interactions between DNA and histones by competition and disassemble

and reassemble nucleosomes in the wake of RNAPII transcription (reviewed in Hammond *et al*., 2017).

Most genomic DNA forms nucleosomes, however there are some regions of genomes that are lower in nucleosome density compared to others, and some regions are completely nucleosome-free. Common regions of genes that are nucleosome depleted between species are enhancer, promoter and terminator elements. These nucleosome-free regions are conserved from *S. cerevisiae* to humans. In *S. cerevisiae*, a nucleosome-free region upstream of transcription start sites is flanked by two well-positioned nucleosomes -1 and +1, and towards the 3' end of genes nucleosomes become less well positioned (Mavrich *et al*., 2008; Shivaswamy *et al*., 2008). Though the presence of a highly positioned +1 nucleosome is conserved between species, the precise location of the +1 nucleosome position relative to the TSS is less well conserved between species (Mavrich *et al*., 2008).

Nucleosome positioning is at least partly governed by DNA sequence – some DNA sequences are more thermodynamically favourable for nucleosome location than others. Regions of the genome that are nucleosome-poor and regions of the genome that are nucleosome-dense can be predicted by DNA sequence alone in approximately 50% of cases (Segal *et al*., 2006). How a DNA sequence is more favourable to be wrapped around histones is determined by the flexibility of the DNA sequence. Each histone heterodimer has three DNA binding motifs spaced approximately 10 bp apart, and specific DNA dinucleotides are more favourable than others in their ability to interact with nucleosomes (Satchwell *et al*., 1986; Widom, 2001). The most favourable dinucleotides for wrapping around histones are A:T or T:A, as they are more bendy than others. Recent mapping of DNA sequences wrapped around histones in *S. cerevisiae* showed approximately 10 bp repeats of AA/AT/TA/TT dinucleotides in interaction surface of the histones which were interspersed with CC/CG/GC/GG nucleotides not present on the interaction surface (Segal *et al*., 2006; Brogaard *et al*., 2012). Long stretches of A:T or C:G base pairs are not favourable for winding around histones and stretches of A:T are common in promoter regions in *S. cerevisiae* and *C. elegans* which enable accessibility for transcription factor binding and are important

determinants of nucleosome organisation. These sequences are predictive of nucleosome occupancy in higher eukaryotes such as humans (Nelson *et al*., 1987; Kunkel and Martinson, 1981; Iyer and Struhl., 1995; Suter *et al*., 2000; Field *et al*., 2008).

Though *in vitro* and *in vivo* studies suggest DNA sequence is a determinant of nucleosome positioning, it is not the sole-determinant. For example, the +1 nucleosome observed *in vivo* is not observed *in vitro*, and this is thought to be due to interaction of transcription factors to the nucleosome depleted promoter region (Zhang *et al*., 2009). Therefore, RNAPII transcription and factors associated with RNAPII machinery can affect nucleosome organisation (reviewed in Struhl and Segal, 2013).

Though *in vitro* studies showed that RNAPII pauses on DNA templates, it was only relatively recently that this phenomenon was shown genome-wide and to correlate with nucleosome density (by using 3'end sequencing techniques that map with base pair resolution densities of RNAPII in the genome). Native elongating transcript-sequencing (NET-SEQ) takes advantage of the strong ternary RNAPII:DNA:RNA structure formed during transcription, and involves a native pull-down of engaged RNAPII followed by purification of associated nascent RNA transcripts (spliced and unspliced) and sequencing the 3' ends (Churchman and Weissman, 2011). NET-SEQ has shown that RNAPII density peaks just before nucleosomes and pauses throughout gene bodies at positions that correlate with nucleosome density in *S. cerevisiae* (Churchman and Weissman, 2011).

Using global run-on sequencing (GRO-SEQ), where a nuclear run on reaction is incubated with brominated nucleotides, and nascent RNA purified and sequenced, it was shown that RNAPII occupancy correlated with nucleosome density at the 3' ends of genes in mammalian cells (Grosso *et al*., 2012). Using precision nuclear run-on sequencing (PRO-SEQ), where a nuclear run on reaction is incubated with biotinylated nucleotides which allows purification of nascent RNA and 3'end sequencing, it was shown that RNAPII accumulates at the +1 nucleosome, and over exons in *Drosophila* (particularly the 5' end) (Kwak *et al*., 2013). Further, in *S. cerevisiae* a technique called

modification cross-linking and analysis of cDNA (mCRAC), where a tagged protein-of-interest (in this case RNAPII) is cross-linked *in vivo* to its associated RNAs which are isolated and sequenced, showed that RNAPII occupancy correlated with nucleosome density (Milligan *et al*., 2016). In mammalian cells, it was also shown using NET-SEQ that promoter-proximal pausing of RNAPII occurs just before the +1 nucleosome (Mayer *et al*., 2015). These studies correlate RNAPII occupancy with nucleosome density and show the importance of nucleosomes for RNAPII transcription.

Despite nucleosomes presenting an obstacle to RNAPII *in vitro* and the above studies demonstrating correlation of RNAPII occupancy with nucleosome density, *in vivo* transcription rates are comparable to *in vitro* transcription rates reported on 'naked' DNA templates (1-4 kb/minute) (Ardehali and Lis, 2009). This suggests that there must be mechanisms in place for RNAPII to overcome the nucleosome barrier and transcribe efficiently. *In vitro* evidence suggests that RNAPII does not separate contacts between DNA and histones but relies on other proteins to loosen these contacts in order to overcome nucleosome barriers (Hodges *et al*., 2009). In order to facilitate RNAPII transcription, chromatin structure is dynamically remodelled in various ways, including: nucleosome disassembly, nucleosome sliding, nucleosome eviction, and histone modification (reviewed in Li *et al*., 2007; Petesch and Lis, 2012). Collectively, these factors and modifications to chromatin structure affect gene expression by modulating RNAPII transcriptions and overcome the effect of DNA sequence on nucleosome occupancy.

ATP-dependent chromatin or nucleosome remodellers use ATP to translocate along DNA and can move nucleosomes by nucleosome sliding, evict entire nucleosomes, or evict/exchange histones (reviewed in Clapier *et al*., 2017). Together, these activities can therefore change the position of nucleosome, or change the composition of the nucleosome (for example, by eviction of a particular post-translationally modified histone) and in so doing make particular DNA sequences more or less accessible by the transcription machinery. There are four different families of chromatin remodeller conserved in eukaryotes: SWI/SNF, CHD, ISWI and INO80. Each family has a

conserved ATPase domain, but differ in the domains flanking it. Flanking the conserved ATPase domain, each family has 'reader' domains that bind to particular types of post-translationally modified histones. Within each family, there are several subcomplexes and each chromatin remodeller is present in complex with many other proteins (reviewed in Clapier *et al*., 2017).

Histones have N-terminal tails that protrude from the core, making them easily accessible to be post-translationally modified in a number of different ways that affect higher-order chromatin structure and therefore RNAPII transcription. Different types of protein modify and interact with histone N-terminal tails: 'writers' that modify histones (histone methyltransferases, histone acetyltransferases, kinases), 'readers' that bind to the histone modification (bromodomain, chromodomain and Tudor domain proteins) and 'erasers' (demethylases, histone deacetylases, phosphatases) that remove the modification (reviewed in Nicholson *et al*., 2015). These writers, readers and erasers can be recruited with the transcription machinery by associating with transcription factors bound to specific DNA sequences, directly by binding to histone modifications, or indirectly *via* readers bound to histone modifications (reviewed in Zhang *et al*., 2015). Most recently, it was found that some histone modifiers bind to nascent RNAPII transcripts (Battaglia *et al*., 2017; Luciano *et al*., 2017; Sayou *et al*., 2017). Histone modifications include: acetylation, phosphorylation, methylation, ubiquitination, sumoylation and ADP-ribosylation (reviewed in Bannister and Kouzarides, 2011). Some histone modifications affect chromatin structure by changing the charge of the amino acid and therefore alter contacts between histones and DNA, and other modifications rely on 'reader' proteins to mediate the effect. Importantly, histone modifications are reversible, which enables dynamic regulation of chromatin structure and gene expression. There are a vast number of histone modifications, and the combinatorial effect of these modifications gives rise to great complexity in the regulation of gene expression (reviewed in Zhang *et al*., 2015).

Modification of histones can affect chromatin compaction by changing the charge of the modified amino acid. For example, acetylation of histone lysine residues by histone acetyltransferases, and phosphorylation of serines, threonines and lysines by kinases

neutralises the positive charge of the residues modified, disrupting the electrostatic interaction between the histone and DNA and making chromatin less compact and a more favourable and accessible environment for transcription factor binding and transcription (reviewed in Bannister and Kouzarides, 2011). The less compact chromatin structure upon acetylation is highlighted by the fact that acetylated chromatin is more sensitive to DNase (Hebbes *et al*., 1994). Histone acetylation is found from yeast to humans on many different H3 and H4 lysine residues, and these modifications are enriched at promoter regions, which facilitates access by transcription factors and therefore stimulates RNAPII transcription (reviewed in Kurdistani and Grunstein, 2003; Bannister and Kouzarides, 2011). Histone phosphorylation is also conserved from yeast to humans and correlates with (but is much less common than) histone acetylation. Histone phosphorylation is associated both with activating and repressive roles in transcription regulation and chromatin compaction during mitosis (reviewed in Rossetto *et al*., 2012).

In contrast, some histone modifications are small and do not directly affect chromatin compaction but indirectly, by providing a binding site for other 'reader' proteins that modify chromatin to make it more or less compact (depending on their activity) and in this way affect transcription. An example of this is histone methylation (of lysines or arginines) by histone methyltransferases, a chromatin mark conserved from yeast to humans. There are numerous characterised histone methylation events that occur most commonly on histones H3 and H4, though is also found on histones H2A and H2B. Histone methylation is a neutral modification that does not alter the charge of the amino acid on which it is deposited but provides a binding site for other chromatin modifiers with appropriate domains that can act to make chromatin more or less compact (reviewed in Hyun *et al*., 2017). There are three possible histone methylation states conserved from yeast to humans: mono- (me1), di-(me2) and tri-methylation (me3), and the degree of methylation determines the effect of the modification, giving an added layer of complexity to this histone modification. The effects of histone methylation are context-dependent – could be activating or repressive - which, as methylation is a neutral mark, are determined by proteins that bind to the mark. For example, in higher eukaryotes chromatin remodeller Chd1 binds to trimethylated

histone H3 lysine 4 (H3K4me3) (deposited by Set1 methyltransferase from yeast to humans), and, *via* interactions with RNAPII, histone chaperone complex FACT, and DSIF, is thought to positively influence transcription elongation (Simic *et al*., 2003). In *S. cerevisiae*, Chd1 lacks the binding domain for H3K4me3 (Sims *et al*., 2005) and instead, H3K4me3 is a binding site for Yng1, which stabilises recruitment of the histone acetyltransferase NuA3, that acetylates H4K14 thereby activating transcription (Taverna *et al*., 2006). While H3K4me3 correlates with active transcription, the function of H3K4me3 in transcription is unclear, with bothww repressive and activating functions reported (Howe *et al*., 2017) and discussed in more detail in chapter 4. From yeast to humans, H3K36me3 is deposited by Set2 methyltransferase and found throughout actively transcribed protein-coding genes. H3K36me3 recruits the histone deacetylase complex Rpd3S, which prevents aberrant intragenic transcription (Carrozza *et al*., 2005).

Ubiquitination is a bulky modification of lysine residues of histones H2A and H2B, which can be mono or polyubiquitinated. Some types of ubiquitination have repressive or activating effects on transcription by recruitment of repressive or activating complexes. For example, in higher eukaryotes, H2A ubiquitination at lysine 119 recruits repressive complex PRC2 (reviewed in Weake and Workman, 2008). On the other hand, conserved from lower to higher eukaryotes, histone H2B ubiquitination on lysine 123 (H2BK123) promotes histone H3 lysine 4 di- and trimethylation (H3K4me2 and H3K4me3) by acting as a bridge for the Set1 methyltransferase. H3K4me2 and H3K4me3 are chromatin marks associated with active transcription elongation, and by helping to recruit the FACT histone chaperone complex (Xiao *et al*., 2005; Pavri *et al*., 2006; Lee *et al*., 2007; Kim *et al*., 2009; Nakanishi *et al*., 2009).

In addition to modification of histone N-terminal tails, the globular domains of histones can be post-translationally modified and new histone modifications are still being discovered, revealing further complexity of the roles of histone modifications in gene expression in eukaryotes.

1.4 Transcription initiation

RNAPII is recruited to promoters of genes by interaction with general transcription factors, which bind to a specific consensus sequence (TATA) within the promoter called the 'TATA box'. Though some eukaryotic promoters contain the consensus TATA box, genome-wide studies have shown the majority of eukaryotic promoters from yeast to humans lack this consensus sequence and rely on sequence elements downstream of promoters which are bound by other proteins that help to align RNAPII at the promoter. There are therefore two general classes of promoter in eukaryotes: TATA-less and TATA-containing (reviewed in Danino *et al*., 2015).

The complex of RNAPII, DNA and general transcription factors is called the pre-initiation complex (PIC) (reviewed in Luse, 2014). General transcription factors collectively comprise: TFIIB, TFIID, TFIIE, TFIIF and TFIIH. Firstly, the TATA box-binding protein (TBP) subunit of TFIID binds either directly to the TATA box (at TATA box promoters), or TBP-associated factors (TAF's) bind to sequences downstream of the promoter and help to align TBP at the promoter (reviewed in Shandilya and Roberts, 2012; Grünberg and Hahn, 2013). The binding of TBP induces a bend in the DNA at promoters and facilitates DNA melting. This complex is stabilised by binding of TFIIB and TFIIA. RNAPII bound to TFIIF then joins, followed by TFIIE and TFIIH which forms the PIC. TFIIH has an ATP-dependent RNA helicase subunit (Xpb) which unwinds approximately 10 base pairs of promoter DNA and enables placement of single-stranded DNA in the active site of RNAPII forming the 'transcription bubble' (reviewed in Shandilya and Roberts, 2012; Grünberg and Hahn, 2013). The initial synthesis of RNA by RNAPII is known as abortive transcription, where several short segments of RNA are transcribed and released. The PIC undergoes conformational changes – the contacts between RNAPII and general transcription factors are broken, and the kinase subunit of TFIIH phosphorylates residues of the largest subunit of RNAPII - until eventually a segment of RNA 9-11 base pairs in length is synthesised and RNAPII clears the general transcription factors and the promoter and starts transcription elongation with new proteins recruited to facilitate this (reviewed in Shandilya and Roberts, 2012;

Grünberg and Hahn, 2013). The general transcription factors are released to recruit a new RNAPII molecule for another round of transcription.

1.5 Transcription elongation

During transcription elongation, RNAPII associates with a number of different transcription elongation factors to form the transcription elongation complex that facilitates elongation and prevents RNAPII stalling or dissociation before the end of the gene. Upon promoter clearance, RNAPII in theory can begin transcribing along the gene in a 5' to 3' direction. However, in metazoans, 20-120 nucleotides downstream of the transcription start site RNAPII pauses in a phenomenon known as 'promoter-proximal pausing'. Promoter-proximal pausing is important as a quality control checkpoint for 5' capping of the nascent transcript and is also important for the regulation of developmental genes. During promoter-proximal pausing, pausing factors DSIF (Spt4/5 in *S. cerevisiae*) and NELF (unknown in *S. cerevisiae*, possibly IswI (Morillon *et al*., 2003)) cause RNAPII to pause in a stable manner, and serine 5 of the CTD is hyperphosphorylated by TFIIH. To transition RNAPII to a productive elongation state, positive transcription elongation factor (P-TEFb) (Bur1/2 and the Ctk complex (Ctk1/2/3) in *S. cerevisiae*) (Wood and Shilatifard, 2006)) phosphorylates DSIF, causing NELF to be released, and activating DSIF as a positive elongation factor. At this time, serine 5 of the CTD is dephosphorylated, and P-TEFb phosphorylates serine 2, converting RNAPII into an active elongation state (Adelman and Lis, 2012). As *S. cerevisiae* lacks NELF, it is thought that promoter-proximal pausing does not occur, however NET-SEQ studies *in S. cerevisiae* have shown that RNAPII accumulates at the 5' ends of genes near transcription start sites (Churchman and Weissman, 2011; Milligan *et al*., 2016).

After promoter clearance and release of promoter-proximal pausing, the rate of RNAPII transcription elongation is not uniform across the gene, or between genes. The RNAPII elongation rate can differ between genes up to 3-fold (Jonkers *et al*., 2014; Veloso *et al*., 2014; Fuchs *et al*., 2014). RNAPII has been shown to slow down over exons (due to increased nucleosome density) and to pause for co-transcriptional

RNA processing (5' capping, pre-mRNA splicing, cleavage and polyadenylation). Differences in RNAPII elongation rates can be explained by differences in nucleosome occupancy, histone modifications, chromatin remodellers and association of transcription elongation factors – none of which are mutually-exclusive.

Unlike RNAPI and RNAPIII, RNAPII has a large unstructured 'carboxy-terminal domain' or 'CTD' in the largest subunit (Rpb1) that protrudes from the catalytic core. The CTD is conserved from yeast to humans and comprises heptapeptide repeats with the consensus amino acid sequence YSPTSPS (reviewed in Eick and Geyer, 2013). While the CTD is conserved in presence and consensus sequence, the number of heptapeptide repeats varies seemingly by complexity of the organism – there are 26 heptapeptide repeats in *S. cerevisiae* that are almost identical, whereas the human CTD has 52 heptapeptide repeats that diverge from the consensus sequence more often (reviewed in Eick and Geyer, 2013). The CTD is a low-complexity domain that has no enzymatic activity but is likened to a 'landing pad' for the recruitment of transcription elongation factors that regulate stages of RNAPII transcription and chromatin modifiers. Additionally, the CTD is important for the recruitment of RNA processing factors, and it is thought that, as the CTD is close to the exit channel on RNAPII, the RNA processing factors can readily transfer from the CTD to nascent pre-mRNA molecule (reviewed in Eick and Geyer, 2013).

Eukaryotic life is possible without a CTD – for example, RNAPII of *Trypanosoma brucei* has no CTD, however at least parts of the CTD are essential for viability in the organisms in which it is found. Genetic experiments have shown that deletion of up to a half of the individual repeats of the CTD does not affect viability in yeast or mammals, however, deletion of half or more of the repeats impairs cell growth and viability (Bartolomei *et al*., 1988; West and Corden, 1995; Meininghaus *et al*., 2000). The 'functional unit' of the CTD was experimentally shown to consist of one heptad repeat followed by the first four residues of the next heptapeptide repeat (Stiller and Cook, 2004). At promoters, the CTD of RNAPII is unmodified (Lu *et al*., 1991), however individual residues within the heptad repeats of the CTD are subject to post-translational modification that dynamically change during the transcription cycle,

enabling diverse regulation of co-transcriptional RNA processing and chromatin modification (reviewed in Eick and Geyer, 2013). Potentially each individual residue within the heptad repeats of the CTD is subject to post-translational modifications that dynamically change during the transcription cycle, enabling diverse regulation of co-transcriptional RNA processing and chromatin modification (reviewed in Eick and Geyer, 2013). Examples of CTD post-translational modification include: phosphorylation (of serine, threonine and tyrosine), glycosylation (of serine and threonine), ubiquitination (of lysine and arginine) and methylation (of lysine and arginine). Additionally, the prolines can undergo *cis-trans* isomerisation. Considering the potential combinations of CTD post-translational modification, it was proposed that there is a CTD 'code' that marks the transcriptional progress of RNAPII, which can be read by other proteins and thereby facilitate their recruitment at the appropriate time during the transcription cycle (Buratowski, 2003).

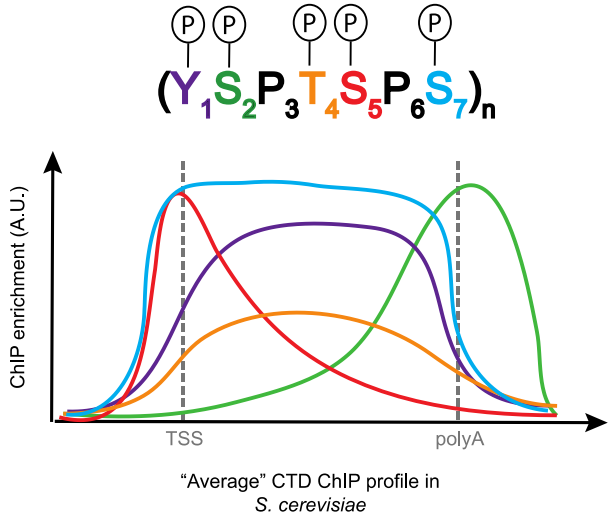
The best characterised form of CTD modification is phosphorylation. Five of the seven residues (tyrosine 1, serine 2, threonine 4, serine 5 and serine 7) of the CTD heptad repeats are reversibly phosphorylated according to the transcriptional progress of RNAPII along a gene, thereby facilitating the recruitment of RNA processing factors and chromatin modifying enzymes at the correct point in the transcription cycle (Figure 1.1A). Phosphorylation makes the CTD less compact and more extended in structure (reviewed in Meinhart *et al*., 2005). The importance of the dynamic nature of CTD phosphorylation is highlighted by genetic evidence that mutated phosphorylated resides of the CTD to alanine or glutamic acid. Using a partially truncated CTD (18 repeats), it was shown that mutation of serines 2, 5 or 7 to glutamic acid is lethal, and mutation of serines 2 and 5 to alanine is lethal in *S. cerevisiae* (West and Corden, 1995; Zhang *et al*., 2012). Mutation of serine 5 to alanine or to glutamic acid was lethal in *S. pombe* but tethering the capping enzymes to the CTD rescued this effect (Schwer and Shuman, 2011). Analysis of full-length CTD phosphorylation mutants later showed that mutation of serine 2 to alanine was not lethal but caused growth defects in both *S. cerevisiae* and *S. pombe* (Cassart *et al.,* 2012). In theory, the number of possible combinations of phosphorylated residues within a single repeat is 416 (reviewed in Eick and Geyer, 2013). Monoclonal antibodies specific to different

phosphorylated forms of the CTD were developed, which recognise different CTD phosphorylation, and have been instrumental in biochemical studies and genome-wide studies mapping the enrichment of the phosphorylation across the gene. However, results may differ depending on the antibody used, as epitope recognition by some CTD antibodies is masked by adjacent post-translationally modified residues (Table 1). Further, the use of antibodies does not allow quantitation of the combinations or degree of phosphorylation per CTD repeat. Recently, quantitative mass spectrometry of a minimal CTD in *S. cerevisiae* and humans revealed, with conservation, that most CTD repeats are unphosphorylated, multiple phosphorylation events within a single CTD repeat are rare, with only the combination of phosphorylated serine 5 followed by phosphorylated serine 2 detected (Schüller *et al*., 2016; Suh *et al*., 2016). These two mass spectrometry studies agree that tyrosine 1, threonine 4 and serine 7 phosphorylation are rare in comparison to serine 2 and serine 5 phosphorylation in yeast and humans. Together, these studies suggest that the CTD code is less complex than first thought. However, mass spectrometry may miss functionally important low-abundant or more transient populations of RNAPII enriched in these CTD phosphorylated residues. For example, serine 7 phosphorylation was shown to be important for snRNA gene expression in mammalian cells (Egloff *et al*., 2007). Furthermore, the CTD may form a complex structure where different repeats are placed in close-proximity and this may modulate protein binding. These mass spectrometry studies suggest there are factors that limit the degree of CTD phosphorylation, and this could include phosphorylated CTD residues inhibiting near-by phosphorylation events by kinases.

Though potentially limited by antibody recognition, ChIP experiments using different CTD antibodies have broadly determined the average phosphorylation profile of the CTD across protein-coding genes, and these patterns are generally conserved from yeast to humans. What is clear from these studies is that the pattern of CTD phosphorylation is intimately linked with proper recruitment of downstream proteins, such as RNA-processing factors. The most well-studied phosphorylation marks are serine 2 and serine 5 phosphorylation of the CTD, and the patterns of these marks are conserved from yeast to humans. Serine 5 is phosphorylated in yeast by Kin28 (Cdk7

in mammals) and is enriched near promoters and transcription start sites where it is important for the recruitment of mRNA capping enzymes (Cho *et al*., 1999; Schwer and Shuman, 2011), during early transcriptional elongation and is a hallmark of promoter-proximal RNAPII pausing and other types of RNAPII pausing (reviewed in Liu *et al*., 2015). In *S. cerevisiae*, serine 5 phosphorylation is thought to be important

A



B

Residue of CTD	Kinase(s)	Phosphatase(s)
Tyrosine 1	Slt2 in *S. cerevisiae* cAbl1/2 in mammals	Unknown (*S. cerevisiae* and mammals)
Serine 2	Ctk1 and Bur1 in *S. cerevisiae* Cdk12 and Cdk9 in mammals	Fcp1 and Cdc14 (*S. cerevisiae* and mammals)
Threonine 4	Unknown in *S. cerevisiae* Plk3 in mammals	Unknown (*S. cerevisiae* and mammals)
Serine 5	Kin28 in *S. cerevisiae* Cdk7 and Cdk8 in mammals	Ssu72, Rtr1 (*S. cerevisiae* and mammals)
Serine 7	Kin28 and Bur1 in *S. cerevisiae* Cdk7 and Cdk9 in mammals	Ssu72 (*S. cerevisiae* and mammals)

Figure 1.1 RNAPII CTD phosphorylation

(A) "Average" ChIP profile of CTD phosphorylation in *S. cerevisiae*. The X-axis shows the position along a gene, and the transcription start site (TSS) and polyA site are indicated.
(B) Table showing a list of the CTD kinases and phosphatases in mammals and *S. cerevisiae*. Figure adapted from Eick and Geyer, 2013.

Table 1 List of CTD antibodies and epitope recognition

Monoclonal antibody	Recognition	Epitope
4F8	Unphosphorylated CTD	$Y_1S_2P_3T_4S_5P_6S_7$
8WG16	Unphosphorylated Serine 2	$Y_1\mathbf{S_2}P_3T_4S_5P_6S_7$
1C7	Unphosphorylated Serine 5	$Y_1S_2P_3T_4\mathbf{S_5}P_6S_7$
3D12	Phosphorylated Tyrosine 1	$Y_1\mathbf{S_2}P_3\mathbf{T_4}S_5P_6S_7$ $\mathbf{Y_1}S_2P_3\mathbf{T_4}\mathbf{S_5}P_6S_7$
3E10	Phosphorylated Serine 2	$Y_1S_2P_3T_4S_5P_6\mathbf{S_7}$ $\mathbf{Y_1}\mathbf{S_2}P_3T_4\mathbf{S_5}P_6S_7$
1G7	Phosphorylated Threonine 4	$Y_1\mathbf{S_2}P_3\mathbf{T_4}\mathbf{S_5}P_6S_7$
6D7	Phosphorylated Threonine 4	$Y_1\mathbf{S_2}P_3\mathbf{T_4}\mathbf{S_5}P_6S_7$
3E8	Phosphorylated Serine 5	$\mathbf{Y_1}S_2P_3\mathbf{T_4}\mathbf{S_5}P_6S_7$
4H8	Phosphorylated Serine 5	$Y_1S_2P_3\mathbf{T_4}\mathbf{S_5}P_6\mathbf{S_7}$
4E12	Phosphorylated Serine 7	$Y_1S_2P_3\mathbf{T_4}\mathbf{S_5}P_6S_7$ $Y_1S_2P_3T_4S_5P_6\mathbf{S_7}$
H5	Phosphorylated Serine 2/5	$\mathbf{Y_1}\mathbf{S_2}P_3T_4\mathbf{S_5}P_6S_7$
H14	Phosphorylated Serine 5/2	$Y_1S_2P_3T_4\mathbf{S_5}P_6S_7$ $Y_1\mathbf{S_2}P_3T_4S_5P_6S_7$
CMA601 (Nojima *et al*., 2015)	Unphosphorylated CTD Phosphorylated Serine 2 Phosphorylated Serine 5	All isoforms
CMA602 (Nojima *et al*., 2015)	Phosphorylated Serine 2	$Y_1\mathbf{S_2}P_3T_4S_5P_6S_7$ $Y_1\mathbf{S_2}P_3T_4\mathbf{S_5}P_6S_7$ $Y_1\mathbf{S_2}P_3T_4S_5P_6\mathbf{S_7}$ $Y_1S_2P_3T_4S_5P_6\mathbf{S_7}$ $Y_1\mathbf{S_2}P_3T_4S_5P_6S_7$ $Y_1S_2P_3T_4\mathbf{S_5}P_6S_7$ $Y_1\mathbf{S_2}P_3T_4S_5P_6S_7$
CMA603 (Nojima *et al*., 2015)	Phosphorylated Serine 5	$Y_1S_2P_3T_4\mathbf{S_5}P_6S_7$ $Y_1S_2P_3T_4\mathbf{S_5}P_6S_7$ $Y_1\mathbf{S_2}P_3T_4S_5P_6S_7$ $Y_1S_2P_3T_4\mathbf{S_5}P_6\mathbf{S_7}$ $Y_1S_2P_3T_4S_5P_6\mathbf{S_7}$ $Y_1S_2P_3T_4\mathbf{S_5}P_6S_7$ $Y_1\mathbf{S_2}P_3T_4\mathbf{S_5}P_6S_7$

Residues in bold: phosphorylated. Residues in red: if this residue is phosphorylated, antibody recognition is blocked. For the Nojima antibodies, no data on adjacent residues blocking antibody recognition is available. Table derived from Eick and Geyer, 2013 and Nojima *et al*., 2015.

for full levels of serine 2 phosphorylation by recruitment of serine 2 kinases (Qiu *et al*., 2009). Serine 5 phosphorylation declines toward the 3' end of genes, and in yeast the phosphatase for serine 5 is Ssu72 (Scp1 in mammals). In contrast, serine 2 is phosphorylated by Ctk1 and Bur1 kinases in yeast (Cdk9 and Cdk12 in mammals) and is a mark of transcription elongation that is low at promoter regions and increases toward the 3' end of genes (declining after the poly(A) site) where it is important for cleavage and recruitment of 3' end processing factors from yeast to mammals (Skaar and Greenleaf, 2002; Ahn *et al*., 2004; Gu *et al*., 2013), and recruitment of pre-mRNA splicing factors and splicing in mammals (Gu *et al*., 2013). Fcp1 also dephosphorylates serine 2 of the CTD.

Unlike serine 5 and serine 2 phosphorylation, the functions of serine 7, threonine 4 and tyrosine 1 phosphorylation seem to be species-specific and their roles in gene expression less well understood (reviewed in Yurko and Manley, 2017). As with serine 5 phosphorylation, serine 7 phosphorylation (by Kin28 in yeast, Cdk7 in mammals) is highest at promoter regions but, unlike serine 5 phosphorylation, serine 7 remains phosphorylated throughout gene bodies, until the 3' end of the gene. The role of serine 7 phosphorylation in transcription remains unclear, its mutation to alanine is tolerated and only snRNA gene expression is reduced in humans (Egloff *et al*., 2007). In *S. pombe*, serine 7 phosphorylation is required for capping, however the function of serine 7 phosphorylation in *S. cerevisiae* remains unclear (St. Amour *et al*., 2012). Threonine 4 phosphorylation (kinase unknown in *S. cerevisiae*, Plk3 in mammals) is present at relatively low levels throughout gene bodies, and mutation to valine did not significantly affect transcription in yeast (Hsin *et al*., 2011; Hintermair *et al*., 2012; Rosonina *et al*., 2014). However, in chicken DT40 cells threonine 4 phosphorylation is essential and required for 3' end processing of non-polyadenylated transcripts (Hsin *et al*., 2011). This function of threonine 4 phosphorylation is not applicable to yeast, and mutation of threonine 4 to alanine is not lethal in *S. cerevisiae* (Schwer and Shuman, 2011), however mutation to glutamic acid is lethal in *S. cerevisiae* and *S. pombe* (Harlen *et al.,* 2016; Kecman *et al.,* 2018). Threonine 4 phosphorylation is important for chromatin remodelling in certain growth conditions and causes growth defects in these conditions in *S. cerevisiae* (Rosonina *et al*., 2014). Most recently,

threonine 4 phosphorylation was shown to be important for proper transcription termination and post-transcriptional splicing in *S. cerevisiae* (Stiller *et al*., 2000; Harlen *et al*., 2016). Tyrosine 1 phosphorylation (by Slt2 in *S. cerevisiae* (Yurko *et al*., 2017) and c-Abl in mammals (Baskaran *et al*., 1997)), resembles but is not identical to the pattern of serine 2 phosphorylation. Tyrosine 1 phosphorylation increases after the transcription start site and remains high throughout the gene body until just before the poly(A) site (Mayer *et al*., 2012). Tyrosine 1 phosphorylation is important for proper termination, and its mutation to alanine is lethal in yeast (Schwer and Shuman, 2011). In vertebrates, tyrosine 1 phosphorylation is essential and important for stability of the CTD (Hsin *et al*., 2014), in addition to being important in regulating antisense transcription (Descostes *et al*., 2014). In *S. cerevisiae*, the role of tyrosine 1 phosphorylation is less clear. The phosphatase for tyrosine 1 has recently been identified as Glc7 in *S. cerevisiae* and is required for proper transcription termination of non-coding RNAs (not applicable to higher eukaryotes) (Schreieck *et al*., 2014) (Figure 1.1B).

General transcription elongation factors associate with RNAPII during transcription and facilitate transcription through the chromatin barrier. In yeast, these factors have been demonstrated to be recruited to gene bodies and associate with RNAPII from promoter regions to cleavage and polyadenylation sites (Mayer *et al*., 2010). Major transcription elongation factors in yeast include Spt4/5 (DSIF in humans), Paf1 and the CTD kinases Ctk1 and Bur1 (reviewed in Shilatifard, 2004; Perales and Bentley, 2009). It is not entirely clear how these elongation factors are recruited to RNAPII during transcription, but it is known that Spt5 interacts with the body of the RNAPII elongation complex, Bur1 with the CTD, and Paf1 with Spt5 (Liu *et al*., 2009; Qiu *et al*., 2009; Qiu *et al*., 2012; Wier *et al*., 2013; Xu *et al*., 2017). Recent photoactivatable ribonucleoside-enhanced crosslinking and immunoprecipitation (PAR-CLIP) studies have shown that each of these proteins interacts with nascent RNA, and it is suggested that this facilitates the chromatin association of these factors and assembly of the transcription elongation complex (Battaglia *et al*., 2017).

Spt5 is a highly conserved and essential elongation factor from yeast to humans that forms a complex with the non-essential zinc finger protein Spt4 (reviewed in Hartzog and Fu, 2013). Spt5 is not associated with RNAPII at promoters, but tightly associates with RNAPII during transcription elongation until transcription termination and acts as a docking site for numerous protein complexes that influence RNAPII processivity, RNA processing and histone modifications (reviewed in Hartzog and Fu, 2013). It is thought that Spt5 enhances RNAPII processivity by stabilising interaction between the clamp of RNAPII and DNA (Hirtreiter *et al*., 2010; Klein *et al*., 2011; Martinez-Rucobo *et al*., 2011). Depletion of Spt5 in *S. pombe* caused genome-wide defects in transcription elongation (Shetty *et al*., 2017). In mammals, Spt5 depletion did not cause such genome-wide defects in transcription, instead it seems Spt5 is important for elongation only on long genes (Fitz *et al*., 2018). Spt5 has a conserved C-terminal region (CTR) that is comparable to the CTD of RNAPII as it is differentially phosphorylated during the course of transcription. In *S. cerevisiae*, the CTR of Spt5 contains 133 residues consisting of 15 hexapeptide repeats with the consensus S[T/A]WGG[Q/A], of which the serine and threonine residues are phosphorylated (Zhou *et al*., 2009). The CTR of Spt5 is important for RNAPII elongation and histone modification (Zhou *et al*., 2009). In *S. cerevisiae*, the Bur1/2 complex phosphorylates the CTR of Spt5 upon recruitment by the serine 5 phosphorylated CTD (Qiu *et al*., 2009). As previously discussed, in metazoans, DSIF (Spt4/5 in *S. cerevisiae*) and NELF cause RNAPII to pause in a stable manner downstream of transcription start sites during promoter-proximal pausing (reviewed in Adelman and Lis, 2012). In mammals and *S. cerevisiae* the CTR of Spt5 is non-essential, but its deletion confers sensitivity to 6-azauracil, suggesting that it promotes RNAPII elongation (Zhou *et al*., 2009).

The phosphorylation of the CTR of Spt5 by the Bur1/2 complex is important for Paf complex recruitment to elongating RNAPII (Laribee *et al*., 2005; Liu *et al*., 2009). The Paf complex is associated with RNAPII throughout actively transcribed genes where it serves as a 'platform' that co-ordinates the association of RNAPII with transcription factors and chromatin-modifying enzymes, thereby facilitating transcription elongation. The Paf complex on its own does not exhibit enzymatic activity. The Paf1

complex is composed of 5 subunits in *S. cerevisiae* (Paf1, Leo1, Ctr9, Rtf1, Cdc73), of which Paf1 and Ctr9 are the most important components for complex integrity (Betz *et al*., 2002). Rtf1, Paf1 and Ctr9 subunits are required for Rad6/Bre1-dependent histone H2BK123 mono-ubiquitination, which in turn is required for H3K4 di- and trimethylation (Wood *et al*., 2003; Krogan *et al*., 2003; Ng *et al*., 2003a; Xiao *et al*., 2005). The Paf1 complex also affects H3K36 trimethylation; deletion of Paf1 complex subunits Paf1 and Ctr9 abolishes H3K36 trimethylation in *S. cerevisiae* (Chu *et al*., 2007). The influence of Paf1 on H3K36me3 is likely indirect, since the Paf1 complex is required for full levels of serine 2 phosphorylation of the CTD of RNAPII which is required for the recruitment of Set2, the methyltransferase responsible for H3K36 trimethylation, a mark of active transcription (Mueller *et al*., 2004; Nordick *et al*., 2008). The Paf1 complex was shown to be a regulator of promoter-proximal pausing in mammalian cells in two quite conflicting studies (Chen *et al*., 2015; Yu *et al*., 2015).

1.6 Transcription termination

Eventually, RNAPII transcribes the 3' untranslated region (3'UTR) of protein coding genes which is a sequence that serves as a signal for binding of cleavage and polyadenylation factor and cleavage factors. After endonucleolytic cleavage of the transcript at the poly(A) site, a poly(A) tail is added by the poly(A) polymerase (Pap1). RNAPII continues transcribing after the poly(A) site, and the 5' ends of downstream transcripts produced following cleavage are targeted for 5'-3' degradation by an exonuclease (Rat1 in yeast, Xrn2 in humans) as they do not have a 5' cap. The precise mechanism of RNAPII transcription termination is not fully understood. There are two prevailing models to explain termination of RNAPII after transcript cleavage, both are applicable from yeast to humans. The allosteric model states that the loss of transcription elongation factors and conformational changes in RNAPII after the poly(A) site makes it unstable. The torpedo model states that the 5'-3' exonuclease Rat1 catches up with transcribing RNAPII and causes dissociation (reviewed in Porrua and Libri, 2015; Porrua *et al*., 2016). The proper termination of RNAPII is important to prevent transcriptional read-through into the next gene.

1.7 Pre-mRNA splicing

Pre-mRNA splicing was first discovered in 1977 in two studies that used the DNA virus human adenovirus 2 (Ad2) as a model system (Chow *et al*., 1977; Bergett *et al*., 1977). In both studies, mRNA of Ad2 was hybridised to Ad2 genomic DNA, and electron microscopy showed regions of DNA that formed loop structures that did not hybridise to mRNA. These DNA loops were introns, and this showed that genes were 'split' into exons and introns, and that introns were removed or 'spliced' out. This discovery explained the previously puzzling finding that nuclear and cytoplasmic transcripts were different in size but had the same 5' and 3' sequences. It was later shown that pre-mRNA splicing is a widespread phenomenon among eukaryotes – from yeast to humans - and almost every gene of multi-cellular organisms is spliced.

In higher eukaryotes, pre-mRNA splicing is not only essential for the production of one protein from one gene, but for the production of multiple proteins from a single gene in a process called alternative pre-mRNA splicing. For example, in humans the average number of introns per gene is 10, and at least 95% of transcripts are alternatively spliced (Pan *et al*., 2008). The importance of alternative pre-mRNA splicing is highlighted by the diversity and complexity it can generate. For example, nematodes and humans have approximately the same number of protein-coding genes in their genomes – in the human genome, there are approximately 20,000 protein-coding genes and in the *C. elegans* genome there are approximately 19,000 protein-coding genes (Ezkurdia *et al*., 2014). In *C. elegans*, approximately 4% of genes are alternatively spliced, and in contrast approximately 95% of human genes are alternatively spliced, giving approximately 200,000 expressed proteins resulting in the obvious differences in complexity between nematodes and humans (reviewed in Hodgkin, 2001; Pan *et al*., 2008). There are multiple forms of alternative splicing, including: exon skipping, intron retention, use of alternative 5' or 3' splice sites and mutual exclusivity of exons. As introns are generally removed during splicing, they may be considered as a waste of energy and relatively unimportant, however it is hypothesised that the presence of introns drives the generation of new proteins (by different combinations of splicing) and ultimately evolution.

Pre-mRNA splicing is also particularly important in disease. The correct regulation of splicing is important for the proper functioning of individual cells and whole organisms - failures may lead to the production of defective proteins and, subsequently, disease or inability to sustain life (reviewed in Wang and Cooper, 2007). Indeed, more than one third of single nucleotide polymorphisms that are disease-causing have the potential to disrupt splicing (Singh and Cooper, 2012). Further, approximately 25% of mutations in the beta-globin gene that cause the disease β-thalassemia in humans result from mutations in splice sites and subsequent defects in pre-mRNA splicing (Treisman *et al*., 1983)

The mechanism of pre-mRNA splicing is conserved from lower eukaryotes such as the budding yeast *S. cerevisiae* to higher eukaryotes such as humans. However, there are some key differences between splicing in *S. cerevisiae* to splicing in other eukaryotes, and in general *S. cerevisiae* exhibits a less complex gene architecture in terms of exon and intron densities. In fission yeast, 45% of genes contain an intron, many genes have two or more introns, and it is estimated that the transcripts of 2-3% of genes are alternatively spliced (Käufer and Potashkin, 2000; Bitton *et al*., 2015; Stepankiw *et al*., 2015). In humans, approximately 95% of transcripts are alternatively spliced, and a minority of genes are intronless. In contrast, the majority of genes in *S. cerevisiae* are intronless, with 5% (of the 6000 genes total) intron-containing. In *S. cerevisiae*, the majority of splicing is constitutive, with only few known examples of genes containing more than one intron and a few cases of alternative splicing (Davis *et al*., 2000; Juneau *et al*., 2009; Volanakis *et al*., 2013; Kawashima *et al*., 2014). Still, splicing is an extremely important and essential process in *S. cerevisiae*, as many of the 5% of genes that are spliced are essential, and the spliced transcripts are highly expressed (largely ribosomal protein transcripts) and contribute to approximately 27% of total mRNA levels (reviewed in Ares *et al*., 1999). Furthermore, on average intron-containing genes are expressed approximately 3.9 times more than intronless genes in *S. cerevisiae* (Juneau *et al*., 2006). These observations that introns stimulate transcriptional output is conserved from *S. cerevisiae* to mammals (Brinster *et al*., 1988; Mckenzie and Brennan, 1996; reviewed in Ares *et al*., 1999).

The importance of introns in *S. cerevisiae* and reason for so few introns in comparison to other species remains controversial – some propose that introns are relatively new to *S. cerevisiae*, and some propose they are in the process of being lost (reviewed in Ares *et al*., 1999; Bon *et al*., 2003). Under growth in optimal laboratory growth conditions deletion of many *S. cerevisiae* introns resulted in no substantial growth defects, however under conditions of stress, deletion of introns did result in growth defects with removal of introns decreasing RNA and protein levels (Juneau *et al*., 2006; Parenteau *et al*., 2008). While deletion of three introns did cause severe growth defects, these defects were rescued by expression of the intronless mRNA and were not due to loss of splicing (Parenteau *et al*., 2008).

Though there are key differences between *S. cerevisiae* and other eukaryotes, the fundamental biochemistry of pre-mRNA splicing is conserved from yeast to humans, and important advances in our knowledge of pre-mRNA splicing were gained from experiments using *S. cerevisiae*. For example, core components of the spliceosome (pre-mRNA processing (Prp) proteins) were first identified in temperature-sensitive mutant screens for defects in total RNA synthesis and extracts from these temperature-sensitive mutants were used to determine basic mechanisms of splicing catalysis (Lustig *et al*., 1986; Vijayraghavan *et al*., 1989; reviewed in Hossain and Johnson, 2014).

1.8 Pre-mRNA splicing is a two-step process catalysed by the spliceosome

Pre-mRNA splicing is dependent on a combination of *trans*-acting factors and short conserved sequences within introns. In total, the intron sequence is not conserved, however there are conserved sequences that define an intron, including: the 5' splice site (5'SS), the 3' splice site (3'SS) and the branch point (BP). In yeast, the consensus sequence of the 5'SS is GUAUGU, the consensus sequence of the BP is UACUAAC and the consensus sequence of the 3'SS is YAG. In contrast, splice sites of more complex organisms like humans are less well conserved, but have the consensus GU at the 5'SS and AG at the 3'SS. The BP of mammalian introns is less well conserved

(apart from the underlined adenosine), and mammalian introns contain a uridine and cytosine rich polypyrimidine tract (10 to 15 nucleotides in length) between the BP and 3'SS sequences that is thought to compensate for lack of conservation of the branchpoint sequence, and is particularly important when the BP is a long distance from the 3'SS (Reed, 1989). *S. cerevisiae* introns do not contain a polypyrimidine tract, however some introns have a uridine-rich poly-U tract upstream of the 3'SS, which enhances splice site usage and may be important for longer introns that require additional splicing factors to bring splice site together (Patterson and Guthrie, 1991; Spingola *et al*., 1999; Ma and Xia, 2011). Conserved across eukaryotes, the 5'SS GU, 3'SS AG and the underlined adenosine of the BP are essential for splicing catalysis as mutation of these residues reduces splicing efficiency or completely abolishes splicing (Langford *et al*., 1984; Chanfreau and Jacquier, 1993; Chanfreau *et al*., 1994). An exception to these conserved residues are the U12-type introns which are flanked by AT at the 5'SS and AC at the 3'SS. U12-type introns are not common and have a specialised minor U12 dependent spliceosome (Tarn and Steitz, 1996; Hall and Padgett, 1996; Burge *et al*., 1998). Consensus splice sites are important for the biochemical splicing reaction and the recruitment of splicing factors – the spliceosome must recognise these three sequences of the pre-mRNA, and therefore it is not surprising that they are conserved sequences. Pre-mRNA splicing occurs by two sequential transesterification reactions that remove the intron and join the exons (Figure 1.2A). In the first transesterification step of splicing ('branching'), the 2'OH of the branchpoint adenosine carries out a nucleophilic attack on the phosphate group of the guanosine of the 5'SS, which cuts the RNA at the 5'SS. The cut 5' end of the intron then forms a 2'-5' bond with the branchpoint adenosine, generating the intron-lariat intermediate. In the second transesterification step of splicing ('exon ligation'), the 3'OH from the 5' exon attacks the 3'SS, resulting in the ligation of the 5' and 3' exons forming a spliced transcript and release of the intron-lariat. The splicing reactions are catalysed by RNA and *in vitro* require metal ions (Mg2+), in addition to ATP (although the afore-mentioned transesterification reactions are not dependent on ATP hydrolysis) (Hardy *et al*., 1984; Hernandez and Keller, 1983; Fica *et al*., 2013). Two Mg2+ ions present in the active site of the spliceosome mediate the two transesterification reactions and splicing catalysis by RNA (Fica *et al*., 2013).

The two pre-mRNA splicing reactions are biochemically relatively simple, but are catalysed by a large (multi-megadalton) and highly dynamic RNA and protein

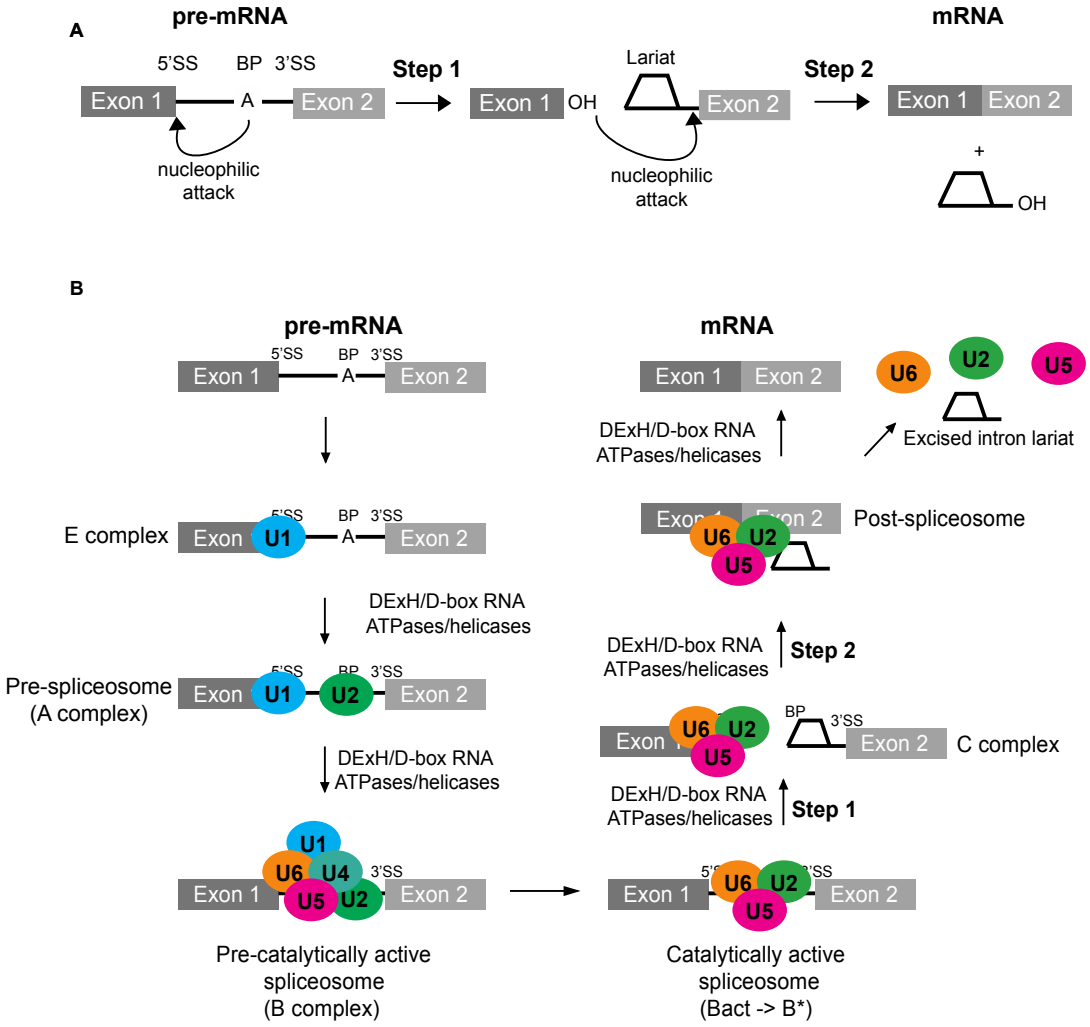


Figure 1.2. Pre-mRNA splicing cycle

(A) Splicing is a two-step transesterification reaction. In the first step, the 5'ss undergoes cleavage via a nucleophilic attack from the 2'OH group of the branchpoint (BP) adenine. The 5' splice site (SS) ligates to the BP, forming the lariat intermediate. In the second step, the 3'OH from the 5' exon attacks the 3'SS, resulting in the ligation of the 5' and 3' exons forming a mature RNA transcript (mRNA), and the release of the intron.
(B) The spliceosome cycle. The 5 snRNPs that constitute the spliceosome are thought to assemble on the pre-mRNA transcript in a step-wise manner. First, the U1 snRNP binds to the 5'SS (forming the E complex), followed by the U2 snRNP which binds to the BP (forming the A complex). The U4/U6•U5 tri-snRNP then joins forming the pre-catalytically active spliceosome (B complex). The spliceosome undergoes substantial re-arrangements between RNA and protein interactions to produce the active spliceosome (B* complex) that catalyses both steps of splicing to generate a mature RNA transcript. In addition to snRNP particles, numerous ATP-dependent DExH/D-box RNA ATPases/helicases are crucial for the fidelity of splicing and are vital in the dynamic RNA-protein rearrangements that occur in spliceosome assembly and splicing. Figure adapted from Will and Lührmann, 2011

(ribonucleoprotein) complex called the spliceosome. The spliceosome is composed of five small, nuclear RNAs (snRNAs), U1, U2, U4, U5 and U6 that specify where splicing occurs and perform catalysis. snRNAs are short RNAs that complex with at least seven other proteins to form small nuclear ribonucleoprotein particles (snRNPs) (U1, U2, U4, U5 and U6). Though the splicing reaction is RNA catalysed, there are numerous proteins that comprise the spliceosome, and these *trans*-acting factors help to position the pre-mRNA for correct splicing (reviewed in Will and Lührmann, 2011). In the spliceosome, there are approximately 90 distinct core proteins conserved from yeast to humans, although the human spliceosome has approximately twice as many distinct proteins due to the complexity of alternative splicing (Fabrizio *et al*., 2009). While the two-step splicing reaction is not ATP-dependent, the substantial structural rearrangements that occur are ATP-dependent. There are eight ATP-dependent DExD/H-box RNA helicases evolutionarily conserved in eukaryotes that are required for pre-mRNA splicing – either for structural rearrangements or spliceosome dissociation.

In vitro and *in vivo* evidence has shown that the five snRNPs assemble *de novo* in a step-wise manner on the pre-mRNA substrate to form the spliceosome, and that the spliceosome depends on the 5'SS, BP and 3'SS sequences for assembly (Figure 1.2B). Firstly, the U1 snRNP base pairs *via* the U1 snRNA to the 5'SS, and Mud2 and Msl5 (branch point binding protein) bind to the BP sequence and form a bridge with the U1 snRNP at the 5'SS. The interaction with the U1 snRNP and the 5'SS is necessary for subsequent stages of spliceosome assembly. Together, these proteins form the commitment complex, or E complex, as it commits the pre-mRNA to splicing. The U2 snRNP then binds to the BP sequence, and together, the U1 and the U2 define the intron and form the pre-spliceosome or A complex. The binding of the U2 snRNP to the BP sequence causes the BP adenosine to bulge out, making the 2'OH available for nucleophilic attack on the 5'SS during the first catalytic step of splicing. The pre-assembled 1.5 MDa tri-snRNP (U4, U5 and U6), comprised of over 30 proteins (Gottschalk *et al*., 1999), then joins the A complex to form the pre-catalytically activated spliceosome or B complex, which is the only stage of spliceosome assembly that contains all five snRNPs assembled on the pre-mRNA substrate. The U4 and U6 snRNAs within the tri-snRNP are complementary base paired to each another, and

during substantial structural re-arrangements, these base pairs are disrupted by the ATP-dependent RNA helicase activity of Brr2. The base pairs between the U1 snRNA and the 5'SS are also disrupted by the ATP-dependent RNA helicase Prp28, and in this way the U1 and U4 snRNPs leave the splicing process and the U6 snRNA base pairs with the 5'SS of the intron and the U2 snRNA. The U2:U6 duplex forms the catalytically active centre the spliceosome (Madhani *et al*., 1992; Fica *et al*., 2013).

After dissociation of the U4 snRNP, the non-snRNP nineteen complex (NTC) then associates, forming the catalytically activated spliceosome (Bact complex). The NTC comprises up to 16 proteins and promotes snRNA/pre-mRNA interactions important for the catalytic core of the spliceosome by stabilising the interaction of U5 and U6 snRNAs with the pre-mRNA substrate before the first step of splicing (Chan *et al*., 2003; Chan and Cheng, 2005). Recent cryo-electron microscopy (Cryo-EM) studies in *S. pombe* have shown that the NTC extends throughout the spliceosome (Yan *et al*., 2015).

The Bact complex is not sufficiently mature to catalyse the first step of splicing– the spliceosome is activated in an ATP-dependent manner by the RNA helicase Prp2, which releases components of the U2 snRNP SF3A and SF3B, which generates the B* complex that can catalyse the first step of splicing – branching (Warkocki *et al*., 2009; Lardelli *et al*., 2010). The release of these components of the U2 snRNP is thought to uncover the branch point adenosine to enable the first catalytic step of splicing. Before the first step of splicing, core U5 snRNP protein Prp8, which is dubbed the 'heart' of the spliceosome as it is the largest and most conserved protein present, interacts with the 5'SS and helps to position it for the first step of splicing (reviewed in Grainger and Beggs, 2005). Prp8 was shown by Cryo-EM to interact with the U2, U4, U6 snRNAs and the intron to form a cavity that generates the catalytic centre of the spliceosome (Yan *et al*., 2015).

The first step of splicing generates the C-complex, which again undergoes substantial ATP-dependent structural rearrangements in which first step splicing factors leave and the activated C* complex is formed (Tseng *et al*., 2011). Prp16, recruited by interaction with Brr2, is an ATP-dependent RNA helicase that facilitates the release

of factors required for the first step of splicing (Yju2 and Cwc25), which is thought to allow binding of second step factors Slu7 and Prp18 (van Nues and Beggs, 2001; Tseng *et al*., 2011). These second step factors and Prp8 help to place the 3'SS in the catalytic core of the spliceosome and mediate the second step of splicing by aligning and stabilising the exons in close-proximity to one another in the catalytic centre of the spliceosome *via* interaction of loop 1 of the U5 snRNA with the exons – this ultimately places the 3'SS in close-proximity to the 5'SS (Newman and Norman, 1992; Schwer and Guthrie, 1992; Teigelkamp *et al*., 1995; Crotti *et al*., 2007). The second step of splicing involves ligation of the exons and release of the intron-lariat.

After splicing catalysis is complete, the spliced mRNA is released by the action of the conserved ATP-dependent RNA helicase Prp22. The 3' to 5' RNA helicase activity of Prp22 disrupts the interactions between the U5 snRNA (and U5 snRNP (Prp8)) and the spliced RNA, thereby releasing the spliced exons from the post-spliceosome complex (Company *et al*., 1991; Wagner *et al*., 1998; Aronova *et al*., 2007; Fica *et al*., 2017). Cross-linking studies found Prp22 binds 13-33 nucleotides downstream of the 3'SS (Schwer, 2008). In recent structural Cryo-EM studies, Prp22 was found in the C* complex interacting with Prp8 which is at the core of the spliceosome (Fica *et al*., 2017). Prp22 has also been shown to have an ATP-independent role in the second step of pre-mRNA splicing (acting after Prp16) (Schwer and Cross, 1998), and is thought to facilitate 3'SS selection and perform an ATP-dependent proof-reading function before the second step – discarding suboptimal substrates (Mayas *et al*., 2006).

The ATP-dependent RNA helicase Prp43 (with Ntr1 and Ntr2) acts at the latest stage of the splicing cycle and is required to release the excised intron-lariat from the post-spliceosome complex, thereby disassembling the post-spliceosome (Martin *et al*., 2002; Tsai *et al*., 2005). In addition to disassembly of the post-spliceosome complex, Prp43 is required for release of stalled spliceosomes from earlier stages of the splicing cycle, and for pre-rRNA processing (Pandit *et al*., 2006; Leeds *et al*., 2006; Mayas *et al*., 2010). The snRNPs can then be recycled for future rounds of splicing, and the intron-lariat de-branched by the lariat debranching enzyme Dbr1 and degraded in the cytoplasm by exonucleases Xrn1 (5'-3') and Ski2 (3'-5') (reviewed in Hesselberth,

2013). The disassembly of the spliceosome is necessary for the efficiency of future rounds of splicing, so that spliceosomes do not become limiting.

It has been shown *in vitro* that pre-mRNA splicing is reversible. Using a mutant of Prp22 that failed to release the mRNA after splicing catalysis, post-catalytic spliceosomes were purified and under certain conditions pre-mRNA was the abundant species associated (not mRNA which is normally associated with post-catalytic spliceosomes). This means that both steps of splicing can be reversed *in vitro* (Tseng and Cheng, 2008). Reversibility of splicing catalysis has not been demonstrated *in vivo*, however the ordered assembly of spliceosomes was shown to be reversible by incorporating fluorescent markers in pre-mRNA and core components of the spliceosome and performing colocalization single-molecule spectroscopy and measuring association and dissociation kinetics (Hoskins *et al*., 2011).

An extremely important aspect of pre-mRNA splicing is the fidelity – as discussed earlier, errors in splicing can be detrimental to cells and so it is vital that errors are supressed. Despite the complexity and dynamics of splicing in human cells, splicing is an inherently high-fidelity reaction, with an approximate error rate as low as 1 in every 100,000 splicing events (Fox-Walsh and Hertel, 2009). The high fidelity of splicing comes from the action of ATP-dependent RNA helicases that are not only key to mediating structural rearrangements in the spliceosome but also splicing fidelity. These ATPases act to supress sub-optimal splicing events, as well as promoting optimal splicing events (reviewed in Semlow and Staley, 2012). ATPases implicated in splicing fidelity include Prp5, Prp28, Prp16, Prp22 and Prp43, and they are collectively thought to proof-read both the first and second steps of splicing catalysis; rejecting suboptimal substrates *via* their RNA unwinding activity (Burgess and Guthrie, 1993; Mayas *et al*., 2006; Xu *et al*., 2007). As the spliceosome assembles in a step-wise manner on pre-mRNA substrates, the ATPase activity of Prp5, Prp28, Prp16, Prp22 and Prp43 can reject sub-optimal splicing events throughout the splicing cycle and prevent later stages of splicing. Prp5 acts before the first step of splicing, and Prp5 mutants with no ATPase activity are able to suppress effects of BP mutations *in vivo*, and it is suggested that Prp5 proof-reads the BP sequence by competing with the interaction of the U2 snRNA and BP, disrupting these interactions if suboptimal

(Xu and Query, 2007). Prp28 also acts before the first step of splicing, and it was shown that mutants of Prp28 enhance splicing of transcripts with suboptimal 5'SS, meaning Prp28 normally functions to reject transcripts with a suboptimal 5'SS (Yang *et al*., 2013). ATPase mutants of Prp16 were shown to allow splicing of transcripts with mutant BP sequences *in vitro*, meaning that normally Prp16 regulates BP usage (Burgess and Guthrie, 1993). Further, Prp16 was shown *in vivo* to suppress cleavage of the 5'SS of sub-optimal substrates at the first step of pre-mRNA splicing catalysis *in vitro*, and pre-mRNAs, lariat intermediates and spliceosomes were disassembled by Prp43 and eventually discarded (Mayas *et al*., 2010). In the absence of the ATPase activity of Prp22, it was shown *in vitro* that mutant 3' splice sites were used more frequently, indicating Prp22 suppresses the use of aberrant splice sites prior to exon ligation (Mayas *et al*., 2006). It was proposed that these ATPases act to proof-read splicing catalysis by pulling the pre-mRNA within the spliceosome from a distance, as truncating the 3' exon downstream of the BP and 3'SS prevented the activity of Prp16 and Prp22 (Semlow *et al*., 2016). Recent cryo-EM structures support this, finding Prp22 and Prp16 at the periphery of the C* complex in humans and in yeast (Bertram *et al*., 2017; Fica *et al*., 2017).

There are two non-mutually exclusive kinetic models to explain the action of ATPases in splicing fidelity, the timer and sensor models (reviewed in Semlow and Staley, 2012). In the timer model, there is a finite amount of time for substrates to be rejected. ATPases compete with the speed of splicing catalysis – if splicing catalysis occurs faster than the ATP hydrolysis activity, the transcript will be spliced at the expense of proof-reading and *vice-versa*. In the sensor model, sub-optimal substrates are rejected at a faster rate compared to optimal substrates. This could be explained by differences in the stability of the interaction between the spliceosome and pre-mRNA substrate, or by regulation of the ATPase activity of the helicase.

Proteomics studies show that many spliceosomal proteins are post-translationally modified, including phosphorylation, ubiquitination, acetylation and glycosylation. Post-translational modification of splicing factors provide an obvious way to regulate key aspects of the splicing cycle including: structural rearrangements *via* protein-protein interactions and protein-RNA interactions, the timings of these

rearrangements, splicing fidelity and splicing factor activities (reviewed in Mckay and Johnson, 2010). The mammalian spliceosome is extensively phosphorylated and to highlight its importance, phosphorylation of proteins in the tri-snRNP is vital for tri-snRNP formation and B complex formation (Matthew *et al*., 2008). Compared with the human spliceosome, the spliceosome of *S. cerevisiae* is less extensively modified, owing to the different complexities of alternative splicing in comparison to constitutive splicing.

1.9 Links between splicing and transcription

Upon discovery of pre-mRNA splicing, it was surmised that "The three short segments forming the 5' tail of hexon mRNA are probably spliced to the body of this mRNA during post-transcriptional processing" (Berget *et al*., 1977). The suggestion that pre-mRNA splicing occurs post-transcriptionally was supported by the finding that pre-mRNA splicing could occur *in vitro* in splicing extracts and required an RNA substrate, ATP and Mg2+ (Hernandez and Keller, 1983; Padgett *et al*., 1983; Hardy *et al*., 1984). Though post-transcriptional splicing can occur, there is increasing evidence from yeast to humans that splicing occurs mainly co-transcriptionally, that is as RNAPII is still transcribing along the gene, before transcription termination (Figure 1.3) (Görnemann *et al*., 2005; Kotovic *et al*., 2003; Lacadie and Rosbash, 2005; Listerman *et al*., 2006; Carillo Oesterreich *et al*., 2010, 2016; Ameur *et al*., 2011; Khodor *et al*., 2011; Tilgner *et al*., 2012; Brugiolo *et al*., 2013; Wallace and Beggs, 2017).

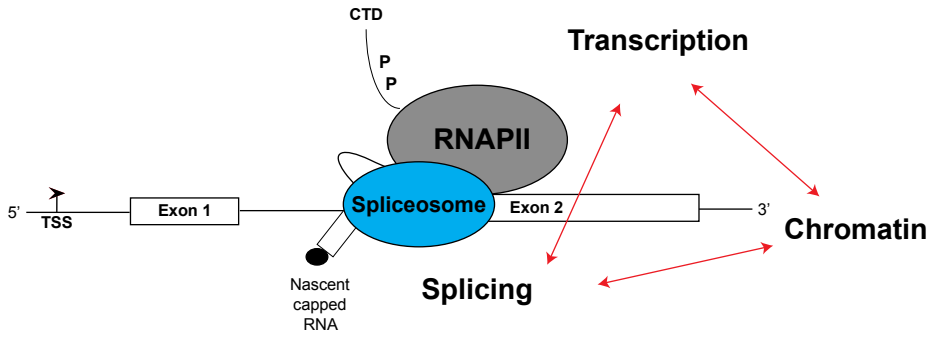


Figure 1.3. Splicing occurs mainly co-transcriptionally

During co-transcriptional splicing, the spliceosome assembles on the nascent pre-mRNA substrate. In this way, splicing, transcription and chromatin are in close-proximity and influence each other.

Co-transcriptional splicing was first observed in 1981 in electron microscopy experiments – where lariats formed on nascent pre-mRNA transcripts associated with RNAPII were visualised (Beyer *et al*., 1981). Later, co-transcriptional spliceosome assembly was visualised by EM - where spliceosomes were found associated with splice junctions of nascent pre-mRNA transcripts in *Drosophila* embryos (Osheim *et al*., 1985). Co-transcriptional spliceosome assembly in budding yeast and metazoans was elegantly demonstrated by chromatin immunoprecipitation (ChIP) experiments, which took advantage of the close proximity of the spliceosome to chromatin during co-transcriptional splicing to show the step-wise assembly of the spliceosome *in vivo*, which validated the ground-breaking *in vitro* studies that found step-wise spliceosome assembly (Kotovic *et al*., 2003; Görnemann *et al*., 2005; Lacadie *et al*., 2010; Listerman *et al*., 2006).

The extent of co-transcriptional splicing was first shown in *S. cerevisiae* by purifying chromatin-associated nascent RNA and using high-density tiling microarrays to analyse the amount of splicing that occurred before transcription was complete (Carillo Oesterreich *et al*., 2010). This indicated that more than 80% of pre-mRNA was spliced before transcription termination and that it was suggested that RNAPII pauses in terminal exons to allow sufficient time for co-transcriptional splicing to occur (splicing-dependent RNAPII pausing is discussed below). Furthermore, it was shown using a different approach in *S. cerevisiae* that the transcripts of a yeast reporter gene are co-transcriptionally spliced (Alexander *et al*., 2010a). Similar co-transcriptional splicing frequencies have since been demonstrated in higher eukaryotes (Ameur *et al*., 2011; Khodor *et al*., 2011; Tilgner *et al*., 2012; Brugiolo *et al*., 2013). The first step of splicing requires the 5'SS and BP but does not require the 3'SS. Therefore, in theory, the first step of co-transcriptional splicing could occur when the BP but not 3'SS is available upon transcription by RNAPII (Rymond and Rosbash, 1985; Liu and Cheng, 2012). Most recently single-molecule intron tracking (SMIT) experiments in *S. cerevisiae* have shown that splicing occurs in close-proximity to the exit channel of RNAPII, and it was concluded that co-transcriptional splicing catalysis occurs, on average, 26–129 nucleotides downstream of 3'SS (Carillo Oesterreich *et al*., 2016). This study suggests that as soon as the substrates are transcribed by RNAPII and

available to be bound by the spliceosome, splicing catalysis can occur, and therefore co-transcriptional splicing is an extremely well-timed and efficient process. Interestingly, high-throughput techniques have shown that not all transcripts that are spliced co-transcriptionally are spliced at the same rate – in *S. cerevisiae*, ribosomal protein transcripts are transcribed faster, more co-transcriptionally and with more fidelity than non-ribosomal protein genes. Additionally, it appears some transcripts are spliced post-transcriptionally. What governs these observations is still unknown but could be explained by a number of different factors, including: promoter elements, intron length, splice site strength, RNA secondary structure and transcription rate (Aslanzadeh *et al*., 2018; Wallace and Beggs, 2017).

By definition, co-transcriptional splicing and transcription are connected in time and space, and this raises the possibility that these processes are functionally coupled such that they influence one another. There are two non-mutually-exclusive possibilities as to how transcription and splicing are functionally coupled (i) by speed of RNAPII (termed the 'kinetic' model) and/or (ii) by interactions between RNAPII and/or transcription elongation machinery and splicing machinery (termed the 'recruitment' model) (reviewed in Perales and Bentley, 2009; Bentley, 2014). Together, the speed of RNAPII and its co-transcriptional recruitment of splicing factors are important in regulating co-transcriptional splicing.

The 'kinetic model' states that the rate of RNAPII transcription elongation affects splicing outcome by modulating the availability of splice sites and co-transcriptional spliceosome assembly (Buratowski, 2003; Bentley, 2005). There is a 'window of opportunity' for splice site recognition before the appearance of competing downstream splice sites, and a 'fast' elongation rate results in less time for splice site recognition compared to a 'slow' elongation rate. There are now multiple bodies of evidence that elongation rate of RNAPII affects splicing outcome, in particular the pattern of alternative splicing in both yeast and mammalian cells (de la Mata *et al*., 2003; Howe *et al*., 2003; Dujardin *et al*., 2014; Fong *et al*., 2014). The importance of optimal co-transcriptional spliceosome assembly for splicing outcome was highlighted by the use of fast and slow RNAPII speed mutants – in either mutant splicing fidelity

is greatly reduced and the fast mutant indicates that post-transcriptional splicing cannot compensate for lack of co-transcriptional splicing as co-transcriptional splicing and overall efficiency are greatly reduced (Braberg *et al*., 2013; Aslandazeh *et al*., 2017).

The 'recruitment model' states that physical interactions between RNAPII and/or factors associated with RNAPII facilitate co-transcriptional spliceosome assembly and splicing. Most evidence for the recruitment model surrounds the CTD of RNAPII, which is thought to function as a 'landing pad' with a 'code' that sequentially recruits RNA-processing factors during the transcription cycle (reviewed in Buratowski, 2003; Bentley, 2005). There are several links between the CTD of RNAPII and splicing. Immunoprecipitation experiments have demonstrated that the CTD of RNAPII interacts with splicing factors (Mortillaro *et al*., 1996; Yuryev *et al*., 1996; David *et al*., 2011). Genetic experiments in which the CTD was truncated or mutated at residues subject to phosphorylation resulted in reduced splicing efficiency *in vitro* and *in vivo* (McCracken *et al*., 1997; Hirose *et al*., 1999; Fong and Bentley, 2001). Furthermore, *in vitro* transcription experiments showed that the CTD stimulates pre-mRNA splicing in comparison to T7 RNA polymerase with no CTD (Hirose *et al*., 1999; Ghosh *et al*., 2000; Das *et al*., 2006). As the CTD is close to the exit-channel of RNAPII, it is proposed that this facilitates the transfer of RNA processing factors to their pre-mRNA substrates (reviewed in Neugebauer 2002; Merkhofer *et al*., 2014).

A CTD mutant, where serine 5 is mutated to alanine, results in intron retention in *S. cerevisiae* (Harlen *et al*., 2016). In mammalian cells, serine 2 phosphorylation of the CTD is required for recruitment of the U2 snRNP (shown by fluorescence *in situ* hybridization (FISH)) and proper pre-mRNA splicing (Gu *et al*., 2013). Additionally, inhibition of splicing using splicing inhibitor drugs and antisense oligonucleotides against snRNAs was shown to reduce the levels of serine 2 phosphorylation of the CTD in human cells (Koga *et al*., 2015), indicating a reciprocal relationship.

Serine 2 and serine 5 phosphorylation of the CTD of RNAPII are also linked to splicing-dependent RNAPII pausing, whereby RNAPII pauses near splice sites. Using an inducible reporter gene, it was shown in *S. cerevisiae* that RNAPII accumulates

periodically (the pause comes and goes – representing waves of transcription) in the vicinity of the 3'SS, at which point the CTD of RNAPII becomes transiently hyperphosphorylated at serine 5 (a mark not normally found in gene bodies and a hallmark of paused RNAPII) followed by hyperphosphorylation at serine 2 (a mark of productive transcription elongation). Through genetic suppression of a splicing mutation, this pausing was demonstrated to be splicing-dependent (Alexander *et al*., 2010b). Further, if pre-spliceosome formation is blocked, RNAPII hyperphosphorylated at serine 5 of the CTD accumulates over introns (Chathoth *et al*., 2014). It was proposed that this represents a splicing-dependent transcriptional checkpoint that is dependent on pre-spliceosome formation, where a signal causes RNAPII to pause to allow sufficient time for spliceosome assembly and quality control to occur, thereby enhancing both splicing efficiency and fidelity.

ChIP-sequencing showed that RNAPII was enriched in exons over introns in human cells, and so did PRO-SEQ in *Drosophila* (Brodsky *et al*., 2005; Schwartz *et al*., 2009; Kwak *et al*., 2013). NET-SEQ has revealed genome-wide RNAPII pausing near splice sites in both mammalian cells and *S. cerevisiae*, providing strong evidence that RNAPII pausing near splice sites is highly conserved from lower to higher eukaryotes (Mayer *et al*., 2015; Nojima *et al*., 2015; Harlen *et al*., 2016). Using mammalian NET-SEQ, it was shown that RNAPII has higher density over the 5' and 3' ends of constitutive exons, indicating RNAPII pausing at these boundaries, and that on exons that are skipped RNAPII is lower in occupancy compared to exons retained in alternative splicing (Mayer *et al*., 2015). Nojima *et al*., (2015) found by mammalian NET-SEQ, using antibodies that detect serine 5 phosphorylation of the CTD, that RNAPII was associated with splicing intermediates and enriched over spliced exons. Importantly, this enrichment was demonstrated to be splicing-dependent by use of splicing inhibitor drugs. In *S. cerevisiae*, Harlen *et al*., (2016) also showed by NET-SEQ that RNAPII pauses over intron-exon junctions. Harlen *et al*., (2016) also demonstrated using ChIP-nexus (He *et al*., 2015) that serine 5 phosphorylated RNAPII peaks downstream of the 3'SS and is associated with recruitment of the U1 snRNP, and by mass spectrometry that the splicing factors (U1 – NTC) are associated with serine 5 phosphorylated RNAPII. In mCRAC, tagged RNAPII is cross-linked *in vivo*

to its associated RNAs, followed by a second purification step using antibodies against different phosphorylated states of the CTD of RNAPII. Associated RNAs are then isolated, fragmented and sequenced (Milligan *et al*., 2016). Using mCRAC, Milligan *et al*., (2016) showed differences in CTD phosphorylation state between intronless and intron-containing genes, and that these were associated with splice sites. Overall, these studies demonstrate genome-wide causative and correlative links between different phosphorylated states of the CTD, RNAPII transcription and splicing. Many of these observations may be explained by higher nucleosome densities in exons compared to introns, and different histone modifications (as previously discussed).

Recently, it was shown in *S. cerevisiae* that ubiquitination of a particular lysine residue in the catalytic subunit of RNAPII is linked to splicing-dependent RNAPII pausing and co-transcriptional splicing (Milligan *et al*., 2017). Deletion of Bre5, which is part of a complex that de-ubiquitinates the RNAPII, led to RNAPII elongation defects and reduced splicing efficiency on a reporter gene. Mutation of the ubiquitinated residue of RNAPII to a ubiquitin-resistant residue led to defects in co-transcriptional pre-mRNA splicing genome-wide. A model was proposed whereby RNAPII is ubiquitinated in response to splicing in which RNAPII pauses, splicing occurs, and then de-ubiquitination occurs and RNAPII is released to continue elongation (Milligan *et al*., 2017).

In addition to RNAPII and its CTD affecting co-transcriptional spliceosome assembly and co-transcriptional splicing, it is possible that core members of the transcription elongation machinery play an important role in co-transcriptional splicing, and indeed there are some links between the transcription elongation machinery and splicing that are discussed in detail in Chapter 3.

1.10 Links between splicing and chromatin

By definition, splicing that occurs co-transcriptionally occurs not only in close-proximity to the transcription machinery but also to the chromatin template beneath (Figure 1.3). There is mounting causative and correlative evidence that chromatin can

influence splicing and *vice-versa* from *S. cerevisiae* to mammals. It is thought that chromatin can affect splicing in two non-mutually exclusive ways (i) by modulating the elongation rate of RNAPII (splicing is extremely sensitive to transcription speed, and nucleosomes act as a barrier to RNAPII transcription) and (ii) by interacting with splicing factors, providing a platform for their co-transcriptional recruitment (reviewed in Haque and Oberdoerffer, 2014; reviewed in Fiszbein *et al*., 2017).

Most evidence suggests that nucleosome occupancy and histone modification can influence splicing outcome. There is correlative evidence that nucleosome positioning and histone modifications are differentially enriched on exons compared with introns. Genome-wide studies from worms to humans found that nucleosome occupancy is significantly higher and more ordered in exons compared to introns, and in constitutively spliced exons compared to alternatively spliced exons (Schwartz *et al*., 2009; Spies *et al*., 2009; Amit *et al*., 2012; Tilgner *et al*., 2012). It was also found that nucleosome occupancy inversely correlates with the strength of the splice site, so that exons flanked by weak splice sites are enriched in nucleosomes in worms and humans (Tilgner *et al*., 2009, Spies *et al*., 2009). In addition, genome-wide mapping has found histone modifications enriched in exons relative to introns from worms to humans, including H3K36me3, H3K4me3 and H2BK123ub, even when considering differential nucleosome density in exons and introns (Andersson *et al*., 2009; Kolasinska-Zwierz *et al*., 2009; Schwartz *et al*., 2009; Spies *et al*., 2009; Dhami *et al*., 2010; Huff *et al*., 2010; Shieh *et al*., 2011). In the case of H3K36me3, this mark is enriched in constitutively spliced exons compared to alternatively spliced exons (Andersson *et al*., 2009; Kolasinska-Zwierz *et al*., 2009), and this mark is lower on intronless genes compared to intron-containing genes (Kim *et al*., 2011). Together, nucleosome positioning and chromatin modifications present within exons can present a barrier to RNAPII that slows RNAPII and consequently this is supposed to affect pre-mRNA splicing by modulating splice site availability and/or by modulating co-transcriptional recruitment of splicing factors (Haque and Oberdoerffer, 2014). Evidence for this includes the fact that RNAPII was shown to accumulate over spliced exons and its elongation rate was shown to slow over exons (Brodsky *et al*., 2005; Chodavarapu *et al*., 2010; Wilhelm *et al*., 2011; Kwak *et al*., 2013; Veloso *et al*., 2014;

Mayer *et al*., 2015). Given the average length of an exon in mammals (145 bp) is similar to the amount of DNA win a nucleosome, it was suggested that this facilitates exon definition (Schwartz *et al*., 2009; Tilgner *et al*., 2009; Carrillo Oesterreich *et al*., 2011).

Though these genome-wide studies suggest links between splicing and chromatin, they are correlative. The first causative evidence that chromatin affects splicing came from the finding that treatment of mammalian cells with histone deacetylase inhibitor, trichostatin A, resulted in inhibition of fibronectin EDI exon inclusion (Nogués *et al*., 2002). It was later shown that the same inhibition of histone deacetylase activity causes exon skipping of the neural cell adhesion molecule (Schor *et al*., 2009). Further evidence that suggested chromatin affects splicing came from a study in mammalian cells that used a series of alternatively spliced reporter genes that were identical in gene sequence but differed in the response of their promoters to different steroid hormones (some were responsive, some were not). This study found that patterns of alternative splicing were altered in the reporter genes with steroid-responsive promoters, and not with the non-responsive promoters. Importantly, RNAPII occupancy was unaffected, indicating that the changes observed in alternative splicing were due to differences in factors recruited to promoters (chromatin modifiers) (Aubroeuf *et al*., 2002). In *S. cerevisiae*, the histone acetyltransferase of the SAGA complex, Gcn5, was shown to genetically interact with components of the U2 snRNP (Msl1 and Lea1) and specifically affect their co-transcriptional recruitment to genes and pre-mRNA splicing (Gunderson and Johnson, 2009). The same group later showed that mutation of the histone H3 residues modified by Gcn5 results in pre-mRNA accumulation, and deletion of histone deacetylases results in changes in co-transcriptional spliceosome assembly and splicing defects, meaning that the level of histone acetylation is important for splicing outcome (Gunderson *et al*., 2011). In *S. cerevisiae*, deletion of Set2, which methylates H3K36, reduced co-transcriptional recruitment of the U2 and U5 snRNPs, and caused pre-mRNA splicing defects (Sorenson *et al*., 2016). These effects on co-transcriptional spliceosome assembly and splicing were independent of RNAPII occupancy, suggesting that chromatin structure can facilitate recruitment of splicing factors co-transcriptionally.

There is mounting evidence that splicing factors interact with chromatin modifiers and thereby affect splicing outcome, in addition to evidence that the presence of chromatin modifiers can affect splicing outcome. For example, SWI/SNF chromatin remodellers were also shown to interact with the U2 and U5 snRNPs and affect alternative splicing in *Drosophila* and humans (Batsché *et al*., 2006; Tyagi *et al*., 2009). In mammalian cells, regulators of alternative splicing PTB and hnRNP proteins associate with histone H3 lysine 9 trimethylation (H3K9me3) (Vermeulen *et al*., 2010). H3K36me3 was also shown to recruit PTB and affect alternative splicing *via* MRG15 which binds to H3K36me3 in mammalian cells (Luco *et al*., 2010). In mammalian cells, H3K4me3 was also found interact with the U2 snRNP and to be important for pre-mRNA splicing outcome (Sims *et al*., 2007; Luco *et al*., 2010; Vermeulen *et al*., 2010) and the links between H3K4me3 and splicing are discussed in detail in Chapter 4. Further, it was shown that deletion of Bre1, the factor responsible for histone H2B monoubiquitination resulted in reduced pre-mRNA splicing efficiency of a subset of genes (~20) in *S. cerevisiae* (Moehle *et al*., 2012). Furthermore, histone variant H2A.Z was shown have genetic interactions with the U2 snRNP, and to be required for efficient co-transcriptional spliceosome rearrangements and pre-mRNA splicing of a subset of genes by reducing RNAPII elongation (Neves *et al*., 2017).

As well as chromatin affecting both constitutive and alternative splicing, there is growing evidence in mammalian systems that splicing can in turn affect chromatin. For example, it was shown using a β-globin reporter gene that mutation of the 3'SS or treatment with splicing inhibitor drug spliceostatin A resulted in changes to H3K36me3 distribution without affecting transcription (Kim *et al*., 2011). Additionally, inhibition of splicing (using spliceostatin A or meayamycin splicing inhibitor drugs) was shown to impair recruitment of Set2 and reduce the level of H3K36me3 without affecting RNAPII occupancy (de Almeida *et al*., 2011). It was further shown using a reporter gene that changes in alternative splicing patterns by altering splice site strengths, and treatment with splicing inhibitor drug meayamycin alters nucleosome (H3) positioning (Keren-Shaul *et al*., 2013). Further, binding of Hu proteins (which regulate exon inclusion or exclusion) to nascent pre-mRNA inhibits the activity of histone deacetylase 2 which increases exon exclusion possibly by

increasing RNAPII elongation rate (*via* level of acetylation) (Zhou *et al*., 2011). It was also demonstrated in mammalian cells that splicing affects H34K4me3 (Bieberstein *et al*., 2012), and the links between H3K4me3 and splicing are discussed in detail in Chapter 4. Though these studies collectively clearly demonstrate that splicing affects chromatin modifications, it is not yet clear mechanistically how for the particular chromatin modifications studied. It is possible that splicing factors acting co-transcriptionally modulate the association of chromatin modifiers with RNAPII during transcription.

It is hypothesised that there is a 'chromatin code' for splicing that is established during the first rounds of co-transcriptional splicing that is maintained in subsequent rounds, and this chromatin code serves to influence processes such as transcription elongation and splicing factor recruitment, which ultimately combine to influence splicing outcome (Kim *et al*., 2011). Though the general conservation of the links between splicing and chromatin in lower and higher eukaryotes would suggest these links are important, the extent of conservation of particular phenomena is not yet clear, and some might be species-specific.

1.11 *S. cerevisiae* as a model organism to study pre-mRNA splicing, and the links between splicing and transcription and links between splicing and chromatin

The budding yeast *S. cerevisiae* is a simple eukaryote that shares many things in common with more complex organisms and is one of the most widely used model organisms to study conserved biological processes – from DNA replication to pre-mRNA splicing. Whilst conserved from yeast to humans, biological processes are much less complex in *S. cerevisiae*, making it a simpler system to manipulate and learn from. There are many advantages to using *S. cerevisiae* as a model organism, including well-established methods of genetic manipulation, short generation time, and the relative affordability of maintenance.

The core splicing machinery is conserved from *S. cerevisiae* to humans - approximately 90 distinct splicing factors are found in both species, although the

human spliceosome has approximately twice as many distinct proteins due to the complexity of alternative splicing (Fabrizio *et al*., 2009). The conservation of splicing machinery between yeast and humans was most recently highlighted by structural studies of the activated spliceosome (C-complex) – the core structure was found to be conserved from yeast to humans, with the major difference being 11 additional proteins in the human C-complex (Wan *et al*., 2016; Galej *et al*., 2016; *et al*., 2018). Further, the finding that splicing is mainly co-transcriptional is conserved from yeast to humans, and many of the links between splicing and transcription and splicing and chromatin are conserved from yeast to humans – indeed most of the core transcription machinery and chromatin modifications are conserved. For example, the rate of RNAPII affecting pre-mRNA splicing outcome, the phenomenon of splicing-dependent RNAPII pausing, and splicing affecting transcription and chromatin modifications and *vice-versa* have all been demonstrated in human cells and *S. cerevisiae*. Due to the conservation of spliceosome structure and links between splicing and transcription and splicing and chromatin, *S. cerevisiae* is an ideal model to study some (but not all) aspects of splicing. The major caveat to working with *S. cerevisiae* is the lack of alternative splicing.

In contrast to mammalian cells, where splicing inhibitor drugs (such as spliceostatin) are available for *in vivo* use, to date, there are no functional splicing inhibitor drugs that work *in vivo* in *S. cerevisiae*. Drugs that are available to inhibit pre-mRNA splicing in mammalian cells are not able to penetrate the thick cell wall of *S. cerevisiae* and splicing studies in *S. cerevisiae* have traditionally relied on the use of *in vitro* techniques, mutants and temperature-sensitive mutants as many splicing factors are essential and therefore classical gene deletion is not an option. To block specific stages of pre-mRNA splicing *in vivo* in *S. cerevisiae*, this study made use of the auxin-inducible degron (AID) system, which is a powerful approach derived from plants to deplete proteins-of-interest in a reversible manner in response to the plant hormone auxin (Figure 1.4) (Nishimura *et al*., 2009). The system is now established from *S. cerevisiae* to mammalian cells. By targeting proteins for degradation, the AID system bypasses traditional problems with genetic techniques such as protein stability and gene deletion lethality (Nishimura *et al*., 2009, Natsume *et al*., 2016).

Auxin is a plant hormone that regulates gene expression by interacting with the F-box transport inhibitor response I (Tir1) protein, which results in the interaction of the E3 ubiquitin ligase SCF (SkpI, Cullin and F- box protein)-TIR1 complex with the AUX/IAA transcriptional repressors. Upon this interaction, the E2 ubiquitin-conjugating enzyme is recruited that polyubiquitylates the AUX/IAA proteins, resulting in subsequent degradation by the proteasome. The SCF pathway is conserved from lower to higher eukaryotes, however TIR1 and AUX/IAA proteins are only present in plant species. The AID system takes advantage of this conservation, such that expression of TIRI and a protein-of-interest fused to an AUX/IAA recognition motif in another eukaryotic organism can result in the depletion of the protein-of-interest in an auxin-dependent manner.

The application of the AID system is not limited to splicing factors. The present work also made use of the AID system to conditionally deplete non-essential and essential transcription elongation factors. As with splicing factors, most studies have made use of mutants, and previous adaptation or compensatory mutations may have occurred in these strains, and therefore phenotypes observed may be indirect. The AID system therefore provides a novel approach to study *in vivo* direct consequences of loss of essential and non-essential proteins.

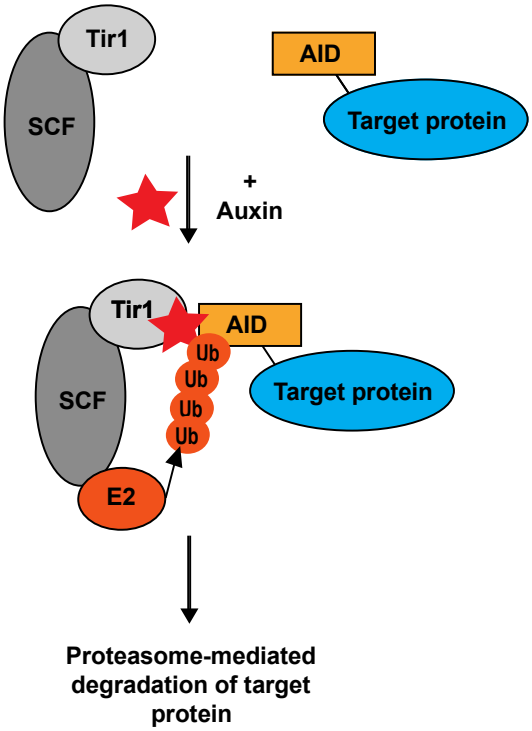


Figure 1.4 The AID system

In the present work, the AID system was used to conditionally deplete proteins-of-interest. The AID-degron system is derived from plants, in which auxin is a member of a family of hormones that regulate gene expression by interacting with the F-box transport inhibitor response I (TIR1) protein, which results in the interaction of the E3 ubiquitin ligase SCF (SkpI, Cullin and F-box protein)-TIR1 with the AUX/IAA transcriptional repressors. Upon this interaction, the E2 ubiquitin-conjugating enzyme is recruited that polyubiquitylates (ub) the AUX/IAA proteins, resulting in subsequent degradation by the proteasome. The SCF pathway is conserved from lower to higher eukaryotes, however TIR1 and AUX/IAA are only present in plant species. The AID-degron system takes advantage of this conservation, such that expression of TIRI and a protein-of-interest fused to an AUX/IAA recognition motif (AID) in another eukaryotic organism can result in the depletion of the protein-of-interest in an auxin- dependent manner. Figure adapted from Nishimura *et al*., 2009.

1.12 Aims of the present work

The present work used *S. cerevisiae* as a model organism to study the links between splicing and transcription and links between splicing and chromatin. There were two major aims of the present work:

- To identify novel links between splicing and transcription by using the AID system to conditionally deplete essential transcription elongation factors and analysing any effects on co-transcriptional spliceosome assembly and splicing (presented in Chapter 3)

- To decipher which stage of the splicing cycle, if any, affects the chromatin mark H3K4me3 in *S. cerevisiae* by using the AID system to conditionally deplete essential splicing factors (presented in Chapter 4)

Chapter 2. Materials and methods

2.1 General yeast maintenance

Yeast cultures were maintained in Yeast Peptone Dextrose Adenine (YPDA) (1% yeast extract, 2% Bacto-peptone, 2% glucose, 0.003% adenine sulphate) or Yeast Minimal Media (YMM) (0.67% Yeast Nitrogen Base without amino acids (Formedium #CYN0401), 2% glucose) supplemented with Kaiser synthetic complete drop out powder (Formedium) according to the manufacturer's instructions.

For experiments, single colonies were picked from YPDA/YMM selective agar plates no older than 2 weeks. Cells were grown at 30℃, 180 rpm in the appropriate selective media.

Glycerol stocks were prepared by adding a 1:1 volume of 30% (w/v) sterile glycerol to an overnight culture. Glycerol stocks were stored at -80℃.

A list of strains used in this study can be found in Appendix 1.

2.2 Time course experiments

For all auxin depletion time courses, cells were diluted from an overnight culture to an OD_{600} of 0.2 in YPDA and grown to an OD_{600} of 0.7. At an OD_{600} of 0.7, cells were treated with 0.75 mM Indole-3-acetic acid (IAA) (auxin) (Acros organics #122160100) dissolved in DMSO for the appropriate amount of time.

To induce TIR1 using the β-estradiol system, cells were diluted from an overnight culture to an OD_{600} of 0.2 in -LEU YMM media. At an OD_{600} of 0.7, cells were treated with 10 µM β-estradiol (dissolved in 100% ethanol) (Sigma-Aldrich #E8875) and 0.75 mM IAA for 40 minutes.

To induce wild-type and mutant Prp22 expression using the β-estradiol system and deplete AID-tagged endogenous Prp22 at the same time, the appropriate volume of cells were diluted from an overnight culture to an OD_{600} of 0.2 in -LEU YMM media. At an OD_{600} of 0.6, cells were treated with 10 µM β-estradiol (dissolved in 100%

ethanol) for 30 minutes, followed by 0.75 mM IAA for 60 minutes.

Cells were harvested for chromatin immunoprecipitation (ChIP), RNA extraction and protein extraction as described below. Each time course was performed in biological triplicate.

2.3 PCR

To tag endogenous genes of interest, the Phusion high fidelity enzyme (NEB #MO530L) was used according to manufacturer's instructions in a final volume of 100 µl. PCR products were ethanol precipitated in 2.5 volumes of 100% ethanol at -20°C for at least 30 minutes. DNA was pelleted by centrifugation at 20,000 x *g* for 20 minutes and washed twice in 70% ethanol. The supernatant was discarded and the pellet air dried and dissolved in 50 µl ddH_2O. A list of primers and plasmids used in this study can be found in Appendix 2 and Appendix 3, respectively.

2.4 Yeast transformation

Yeast were transformed using the TRAFO LiAc/PEG method (Gietz and Schiestl, 2007). 50 ml cells were grown to an OD_{600} of 0.7 in the appropriate selective media. Cells were harvested by centrifugation at 1000 x *g* for 2 minutes, washed twice in ddH_2O, once with 100 mM LiAc and resuspended in 400 µl 100 mM LiAc. 100 µl of cells were used per transformation, and the following transformation mix was added: 240 µl 50% PEG4000, 36 µl 1 M LiAc, 50 µl 10 mg/ml ssDNA (Roche #11467140001) (boiled before use) and 10 µg of PCR product or 100 ng plasmid DNA. A negative control with no PCR/plasmid DNA was always included. The mix was incubated at 42°C for 40 minutes and spun briefly at 1000 x *g*. The cells were resuspended in 500 µl media and incubated at 30°C rotating for 2 hours to recover. After the recovery period, cells were harvested by centrifugation at 1000 x *g* for 2 minutes and re-suspended in 100 µl ddH_2O. Cells were plated on the appropriate selective media and incubated at 30°C for 2-3 days, or until single colonies were visible. If overgrown, cells were replica plated on appropriate selective media and incubated at 30°C for 1-2 days.

2.5 DNA preparation and PCR

To check for positive yeast transformants, genomic DNA was extracted from an overnight culture of a single colony grown in 3 ml selective media. Cells were harvested by centrifugation at 1000 x *g* for 2 minutes. The pellet was re-suspended in 200 µl ddH_2O, 200 µl phenol:chloroform (1:1) pH 5.2, 200 µl DNA extraction buffer (2% Triton-x100, 1% SDS, 100 mM NaCl, 10 mM Tris-HCl pH 8.0, 1 mM EDTA pH 8.0) and 200 µl 0.5 mm zirconia beads (Thistle scientific #11079105z). The mixture was vortexed for 2 minutes followed by centrifugation at 18,000 x *g* for 5 minutes. The aqueous phase was transferred to 2.5 volume 100% ethanol and DNA precipitated at -20℃ for 30 minutes. The sample was centrifuged at 18,000 x *g* for 10 minutes, washed twice in 70% ethanol and the DNA pellet re-suspended in 200 µl ddH_2O.

Each strain was verified by PCR using TAQ DNA polymerase (Invitrogen #10342020) according to the manufacturer's instructions using oligonucleotides listed in Appendix 2.

2.6 Bacteria growth

The *Escherichia coli* (*E. coli*) strain NEB-5-alpha (NEB #C2987H), was used to propagate plasmid DNA. Standard techniques were used for growing and transforming *E. coli.*

2.7 Plasmids and cloning

Plasmid DNA was extracted using the Qiagen QIAprep Spin Miniprep Kit (#27104) according to the manufacturer's instructions. For restriction enzyme analysis, NEB restriction enzymes were used according to the manufacturer's instructions. For ligation reactions, T4 DNA ligase (NEB #MO202) was used according to the recommended protocol. For gel extractions, the Qiagen Gel Extraction Kit (#28704) and protocol was used to purify DNA. All plasmids used in this study are listed in Appendix 3.

Gibson assembly was used to generate the plasmids pZL-PRP22WT-V5, pZL-PRP22T757A-V5 and pZL-PRP22I764A-V5. The backbone of these plasmids is

derived from pZTRL and the inserts derived from p360-PRP22-V5, p360-T757A-V5 and p360-I764A-V5. 5 µg of pZTRL was digested with SfiI (NEB), the DNA fragments separated by agarose gel electrophoresis and the backbone segment extracted and purified using the Qiagen Gel Extraction Kit (#28704) according to the manufacturer's instructions. The insert segments of Prp22 were amplified by PCR using NEB Phusion enzyme using primers listed in Appendix 2 according to manufacturer's instructions. 5 µg of the insert PCR products were digested with SfiI and cleaned-up using the MinElute PCR clean-up kit (Qiagen #28006). The NEB Gibson assembly kit (#E2611) was used to generate pZL-PRP22WT-V5, pZL-PRP22T757A-V5 and pZL-PRP22I764A-V5 according to the manufacturer's instructions.

2.8 DNA sequencing

DNA was sequenced using the Sanger protocol at the Sequencing Facility at the University of Dundee (https://www.dnaseq.co.uk).

2.9 Protein sample preparation and western blotting

10 ml of cells at an OD_{600} of 0.8 were fixed in 5 ml of 100% methanol on dry ice and harvested by centrifugation at 1000 x *g*. Protein samples were prepared using a modified NaOH lysis and trichloroacetic acid (TCA) precipitation protocol (Volland *et al*., 1994). Cells were re-suspended in 830 µl ddH_2O and 217 µl of 1 M NaOH, mixed and incubated on ice for 10 minutes. 58 µl 100% TCA was added and samples mixed and incubated on ice for 15 minutes. Samples were centrifuged at 20,000 x *g* for 5 minutes, the supernatant discarded, and the pellet resuspended in 77.8 µl of dissociation buffer (0.1M Tris-HCl pH6.8, 4mM EDTA pH8, 4% (w/v) SDS, 2% (v/v) 2-mercaptoethanol, Orange G) and 35 µl 1 M Tris base. Protein samples were incubated at 96℃ for 10 minutes, centrifuged at 18,000 x *g* for 4 minutes and the supernatant retained for western blotting. Protein concentration was measured using the Bradford assay (Bio-rad #5000006).

For western blotting, 25 µg of protein were run on a NuPAGE 4-12% gel Bis-Tris (Invitrogen #NP0323BOX) at 180 V in 1 X MOPS-SDS buffer (Invitrogen #1862491).

Proteins were transferred to a Bio-rad nitrocellulose membrane (0.2 µm, #LC2009) using a semi-wet transfer unit (Bio-rad) at 100 V for 1 hour at 4℃ in Tris-Glycine transfer buffer (200 mM Tris, 1.5 M glycine) with 10% methanol. The membrane was blocked in PBS and 5% (w/v) milk powder for 1 hour with gentle rocking. Primary antibodies were diluted in PBS with 0.2% TWEEN-20 and 5% milk powder and incubated for 1 hour at room temperature or overnight. The membrane was washed 3 times for minutes each in PBS and 0.2% TWEEN-20. The membrane was incubated with the appropriate secondary antibodies diluted in PBS, 5% (w/v) milk powder and 0.2% TWEEN-20 for 1 hour, rocking at room temperature. The membrane was washed 3 times for 10 minutes each in PBS and 0.2% TWEEN-20, followed by a final wash in 1 X PBS.

The Odyssey infrared imaging system (LI-COR Bioscience) was used to visualise proteins of interest after transfer. To quantitate, the median method of the Odyssey software was used. Unless otherwise stated, data were normalised against the 3-Phosphoglycerate Kinase (Pgk1) loading control. Student's t-tests were performed to determine the P-values for differences observed between wild-type and depletion/induction conditions. Primary and secondary antibodies used are listed in Tables 2 and 3, respectively.

Table 2. Primary antibodies used for western blotting in this study

Antibody	Concentration	Conditions
Rabbit anti-Prp40 (rabbit 11 Eurogentec 2014)	1:1000	Overnight 4℃
Rabbit anti-Prp8 (R1703 Final bleed Boon peptide 5/046)	1:1000	Overnight 4℃
Rat anti-FLAG (Agilent #200474)	1:1000	Overnight 4℃ or RT 1 hour
Mouse anti-histone H3 (Diagenode #C15210011)	1:1000	Overnight 4℃ or RT 1 hour
Rabbit anti-histone H3K4me3 (Active motif #39159)	1:1000	Overnight 4℃ or RT 1 hour
Rabbit anti-histone H3K4me2 (Active motif #39141)	1:1000	Overnight 4℃ or RT 1 hour
Rabbit anti-histone H3K4me1 (Active motif #39297)	1:1000	Overnight 4℃ or RT 1 hour

Mouse anti-Rpb1 (Diagenode #C15100055-100)	1:1000	Overnight 4℃ or RT 1 hour
Mouse anti-Rpb3 (Biolegend #665004)	1:1000	Overnight 4℃ or RT 1 hour
Rat anti-S2P (Chromotek #3E10)	1:500	Overnight 4℃
Rat anti-S5P (Chromotek #3E8)	1:500	Overnight 4℃
Mouse anti-Pgk1 (Abcam #22C5D8)	1:5000	Overnight 4℃ or RT 1 hour
Mouse anti-c-MYC (Santa Cruz #SC-40x)	1:1000	Overnight 4℃ or RT 1 hour
Mouse anti-V5 (Invitrogen #MA5-15253)	1:1000	Overnight 4℃ or RT 1 hour
Mouse anti-HA (Santa Cruz #F-7)	1:1000	Overnight 4℃ or RT 1 hour
Rabbit anti-Spt5-P (#1761, kind gift from Steven Hahn) (Liu *et al*., 2009)	1:1000	Overnight 4℃
Rabbit anti-Spt5 (kind gift from Grant Hartzog) (Hartzog *et al*., 1998)	1:1000	Overnight 4℃

Table 3. Secondary antibodies used for western blotting in this study

Antibody	Concentration	Conditions
Goat anti-mouse IRDye680RD (LI-COR #926-68070)	1:10,000	RT 1 hour
Goat anti-rabbit IRDye680RD (LI-COR #926-32223)	1:10,000	RT 1 hour
Goat anti-rat IRDye800RD (LI-COR #926-32219)	1:10,000	RT 1 hour
Goat anti-rabbit IRDye800RD (LI-COR #925-32211)	1:10,000	RT 1 hour

2.10 Growth analysis

An overnight culture from a single colony was diluted to an OD_{600} of 0.1 in 200 µl of the appropriate media in a sterile flat-bottomed 96 well plate. A Sunrise Absorbance Microplate Reader (Tecan Trading AG) was used to grow the cells at 30℃ 180 rpm. OD_{600} measurements were taken every 15 minutes for 18 hours. To test the auxin sensitivity of the AID-tagged strains, the cells were grown in the presence of 0.75 mM IAA (dissolved in DMSO). The growth rate of each strain was measured in biological triplicate.

2.11 Total RNA preparation and RT-qPCR

10 ml of cells at an OD_{600} of 0.8 were fixed in 5 ml methanol on dry ice and harvested by centrifugation at 1000 x *g*. RNA was extracted using a modified GTC:Phenol method from Tollervey and Mattaj (1987). Cell pellets were re-suspended in 400 µl AE buffer (50mM NaOAc pH 5.3, 10mM EDTA, 0.5% SDS), 600 µl of phenol citrate pH 4.3 and 300 µl zirconia beads. Cells were lysed in using a Mini-Beadbeater-24 (BioSpec Products) at 2000 rpm twice for 2 minutes followed by 2 minutes on ice. The sample was placed on dry-ice for 2 minutes. Cells were centrifuged at 20,000 x *g* for 5 minutes and the upper phase was transferred to a new tube containing 500 µl phenol:chloroform (5:1) pH 4.3 (ThermoFisher Scientific, AM9720). The sample was vortexed and centrifuged at 20,000 x *g* for 5 minutes. The upper phase was transferred to a new tube containing 500 µl chloroform, vortexed and centrifuged at 18,000 x *g* for 5 minutes. The upper phase containing RNA was precipitated in 1/3rd volume 10 M LiCl for 30 minutes at -20℃. The sample was centrifuged at 18,000 x *g* for 5 minutes at 4℃ and washed twice in 70% EtOH. The pellet was re-suspended in 50 µl ddH_2O, and RNA concentration measured using a Nanodrop.

For reverse transcription, 5 µg of RNA were diluted in ddH_2O to a total volume of 8 µl. Residual DNA was removed by DNase treatment for 15 minutes at 37℃ (1 µl 10x DNase I buffer, 0.9 µl DNase I enzyme (Promega #M6101) and 0.1 µl RNase inhibitor (Promega #N2511)). The DNase I was inactivated at 75℃ for 10 minutes. 2.5 µl of a mix containing the desired reverse primers (3 µM each) was added to the DNase-treated RNA and incubated at 72℃ for 5 minutes. The sample was split into two – one for the reverse transcription and a negative control with no reverse transcriptase added. 5 µl of a reverse transcriptase master-mix was added to each tube: 2 µl 5 x reverse transcription buffer, 0.75 µl 10 mM dNTPs (Promega #U1515), 0.25 µl RNase inhibitor, 1.7 µl ddH_2O and 0.3 µl reverse transcriptase (Roche #3531287001) (or water in the negative control). cDNA synthesis was performed at 55℃ for at least 2 hours, and the cDNA diluted to a final volume of 200 µl in ddH_2O.

For qPCR analysis, 2.5 µl cDNA was mixed with 2 µl Brilliant III SYBR green qPCR mix (Agilent #60088251) and 0.5 µl primer mix (3 µM each). Each qPCR reaction was

run in technical triplicate.

The following cycling parameters were used: 5 minutes at 95℃, 40 cycles of 10 seconds at 94℃, 10 seconds at 60℃ and 15 seconds at 72℃. The absolute quantification/second derivative maximum analysis in the Light Cycler 480 software 1.5.0 (Roche) was used to quantitate the signal and data were normalised to the intronless gene *ALG9* and/or time zero/wild-type conditions, or to *SCR1* and/or time zero/wild-type conditions. A list of primers used for reverse transcription and RT-qPCR can be found in Appendix 4. Student's t-tests were performed to determine the P-values for differences observed between wild-type and depletion/induction conditions.

2.12 Chromatin immunoprecipitation (ChIP)

50 ml of culture at an OD_{600} of 0.8 was cross-linked linked in 1% (w/v) formaldehyde for 10 minutes shaking at room temperature. The reaction was stopped by incubating the cells for 5 minutes with 2.5 ml of 2.5 M glycine. Cells were harvested by centrifugation at 1000 x *g* and washed twice in ice-cold 1 X PBS. Cell pellets were re-suspended in 350 µl FA1 buffer (50 mM HEPES-KOH pH 7.5, 140 mM NaCl, 1mM EDTA pH 8.0, 1% Triton X-100, 0.1% sodium deoxycholate, one complete EDTA-free proteinase inhibitor tablet (Roche #11836145001), four PhosSTOP tablets (Sigma Aldrich #000000004906845001) and 350 µl zirconia beads. The cells were disrupted using the Mini-Beadbeater-24 (BioSpec Products) twice at 2000 rpm for 2 minutes with 2 minutes on ice in between. The sample was separated from the beads by centrifugation at 1000 x *g* for 2 minutes. The sample was centrifuged at 20,000 x *g* for 15 minutes at 4℃. The pellet was re-suspended in 600 µl FA1 buffer, and the sample sonicated using a New Twin Biorupt sonicator (Diagenode) for 15 cycles 30 seconds on and 30 seconds off. The sample was centrifuged at 20,000 x *g* for 30 minutes at 4℃ and the supernatant containing solubilised chromatin was retained. Protein concentration was determined using the Bradford assay.

For immunoprecipitation, the appropriate amount of chromatin was incubated in 20 µl Protein A/G Dynabeads (Life Technologies #10001D/10003D) conjugated to antibody on a rotating wheel overnight at 4℃. A list of the antibodies and amount of chromatin

used per IP can be found in Table 4.

The beads were washed 3 times in FA1 buffer, twice in FA2 buffer (50 mM HEPES-KOH pH 7.5, 0.5 M NaCl, 1mM EDTA pH 8.0, 1% Triton X-100, 0.1% sodium deoxycholate), twice in FA3 buffer (10 mM Tris-HCl pH 8.0, 250 mM LiCl, 1mM EDTA pH 8.0, 0.5% NP-40, 0.5% Na deoxycholate) and once in Tris-EDTA pH 8.0 0.05% TWEEN-20. Crosslinking was reversed with 150 µl elution buffer (50 mM Tris-Hcl, 10 mM EDTA, 1% SDS) and 3 µl Proteinase K (25 mg/ml) and incubated at 42℃ for 2 hours and 65℃ overnight, shaking. An input sample equal to 10% of the protein that was used for the immunoprecipitation was prepared and crosslinking was reversed as above. The QIAGEN mini column clean-up kit was used to purify DNA according to the manufacturer's instructions, and DNA eluted in 400 µl of 10 mM Tris pH 8.0. Samples were analysed by qPCR as described above using primers listed in Appendix 5. The positions of amplicons used for qPCR for each gene are shown to scale in appendix 6. The ChIP data was using the relative threshold cycle (Ct) values for each sample. ChIP data are presented as percentage of input normalised to the first amplicon of each gene, to Rpb1, to histone H3 or to another negative control region specified in the figure legend. Student's t-tests were performed to determine the P-values for differences observed between wild-type and depletion/induction conditions.

Table 4. List of antibodies and conditions used for ChIP in this study

Chromatin (µg)	Antibody and concentration	Dynabeads A/G
250	Rabbit anti-Prp40 (rabbit 11 Eurogentec 2014); 5 µl/IP	Protein A
250	Rabbit anti-Prp8 (R1703 Final bleed Boon peptide 5/046); 5 µl/IP	Protein A
500	Rabbit anti-HA (Abcam #AB9110); 5 µl/IP	Protein A
250	Mouse anti-FLAG (Sigma M2 #F1804); 5 µl/IP	Protein G
250	Mouse anti-c-MYC (Santa Cruz #SC-40x); 5 µl/IP	Protein G
250	Rabbit anti-V5 (Abcam #AB9116); 5 µl/IP	Protein A
125	Mouse anti-histone H3 (Diagenode #C15210011); 1.5 µl/IP	Protein G
125	Rabbit anti-histone H3K4me3 (Active motif #39159); 3 µl/IP	Protein A
125	Rabbit anti-histone H3K4me2 (Active motif #39141); 3 µl/IP	Protein A
125	Rabbit anti-histone H3K4me1 (Active motif #39297); 3 µl/IP	Protein A
125	Mouse anti-Rpb1 (Diagenode #C15100055-100); 1.5 µl/IP	Protein G

125	Mouse anti-Rpb3 (Biolegend #665004); 2 µl/IP	Protein G
125	Rat anti-S2P (Chromotek #3E10); 10 µl/IP	Protein G
125	Rat anti-S5P (Chromotek #3E8); 10 µl/IP	Protein G

2.13 Thiolabelling of nascent RNA

The protocol used was developed by David Barrass (Barrass *et al*., 2015). Cells containing the plasmid pRS426–FUI1 that encodes the uracil permease Fui1 were grown in -URA YMM to an OD_{600} of 0.8. To label the nascent RNA, 100 µM 4-thiouracil (Acros organics # 359930010) was added to the culture and 30 ml of culture snap frozen by immersion into 20 ml methanol on dry ice at the following time points post 4-thiouracil addition: 1 minutes, 2 minutes, 4 minutes and 8 minutes. Cells were harvested by centrifugation at 1000 x *g* and washed once in ice-cold ddH_2O.

RNA was extracted as described above and re-suspended in a final volume of 90 µl TE pH 7.0 in a 0.2 ml tube. RNA was biotinylated at 65℃ for 15 minutes by the addition of 1/10th volume HPDP-biotin solution (4mM in dimethyl formamide). RNA was purified from excess HPDP-biotin solution ThermoFisher Scientific #21341), in a 0.5ml Zeba desalting column (ThermoFisher Scientific #89882), using TE buffer and following the manufacturer's instructions. RNA was precipitated at -20℃for 30 minutes by the addition of 1/3rd volume 10 M LiCl. The sample was spun at 18,000 x *g* for 5 minutes at 4 ℃, and pellet washed twice in 80% EtOH rotating at RT. The RNA was pelleted at 20,000 x *g* for 5 minutes at 4℃ and re-suspended in 200 µl ddH_2O at 65 ℃. RNA concentration was measured using a Nanodrop, and the concentration across samples equalised to final volume of 250 µl ddH_2O. Bead buffer (final concentration 1 mM Tris-HCl pH7.0, 20 mM NaCl, 2.5 mM $MgCl_2$, 100 mM NaPi pH 6.8, 0.1% SDS) was added to the RNA and the samples were well mixed, with care being taken to avoid precipitation by adding the NaPi and SDS separately from the other components, all at room temperature.

Magnetic streptavidin beads (NEB #S1420S) were prepared by adding 50 µl beads to a 1.5 ml low-binding tube and washing them in 200 µl of bead buffer. The beads were blocked in 200 µl bead buffer supplemented with 2.5 µl of 5 mg/mL tRNA and 10 µl

of 20 mg/mL glycogen for 20 minutes in a rotating wheel at RT, followed by a final wash in 200 µl bead buffer. The RNA was added to the beads and incubated on a rotating wheel for 30 minutes at RT. The beads were washed four times in 200 µl of bead buffer and eluted in 40 µl of 0.7 M 2-mercaptoethanol (freshly prepared). RNA was precipitated at -20°C for 1 hour by adding 2.5 volumes 100% EtOH, $1/10^{th}$ volume 3 M NaOAc pH 5.3 and 10 µg glycogen (Roche # 10901393001). RNA pellets were washed twice in 70% ethanol and re-suspended in 10 µl TE pH 7.0. RNA quality was analysed using an Agilent 2100 bioanalyser RNA Nano-Chip according to the manufacturer's instructions. Reverse transcription followed by qPCR was performed as described in section 2.11.

2.14 Co-immunoprecipitation

250 ml of culture at an OD_{600} of 0.8 was harvested by centrifugation at 1000 x *g* and washed twice in ice-cold 1 X PBS. The cell pellet was re-suspended in 900 µl lysis buffer (50 mM Tris-HCl pH 7.5, 2 mM Mg_2Cl_2, 150 mM NaCl, 0.2% NP-40 and one complete EDTA-free proteinase inhibitor tablet (Roche #11836145001) and 400 µl zirconia beads. Cells were lysed using a Mini-Beadbeater-24 (BioSpec Products) twice at 2000 rpm for 2 minutes followed by 2 minutes on ice. The sample was centrifuged at 1000 x *g* for 2 minutes, the supernatant was collected and additionally centrifuged at 20,000 x *g* for 30 minutes at 4°C and used for immunoprecipitation.

The concentration of protein was measured using the Bradford assay, and 1 mg of protein used per IP. Prior to immunoprecipitation, extract was pre-cleared by adding ½ volume unconjugated Protein A/G Dynabeads (Life Technologies ##10001D/10003D). 50 µl Protein A/G Dynabeads conjugated to antibody were incubated with the appropriate volume of extract on a rotating wheel overnight at 4°C. The next day, beads were washed 8 times in lysis buffer (non-bound fraction kept for analysis). 20 µl of loading buffer was added to the beads, input and non-bound samples, which were boiled for 10 minutes before loading on a NuPAGE 4-12% Bis-Tris gel Bis-Tris (Invitrogen) and western blotting was performed as described above. A list of antibodies used for co-immunoprecipitation can be found in table 5, and antibodies used for western blotting can be found in Tables 2 and 3.

Table 5. List of antibodies and conditions used in this study for co - immunoprecipitation in this study

Amount of extract (µg)	Antibody and concentration	Dynabeads A/G
1000	Mouse anti-FLAG (Sigma M2 #F1804); 10 µl/IP	Protein G
1000	Mouse anti-c-KMYC (Santa Cruz #SC-40x); 10 µl/IP	Protein G
1000	Rabbit anti-Prp8 (R1703 Final bleed Boon peptide 5/046) 10 µl/IP	Protein A

2.15 Native elongating transcript (NET) purification and RT-qPCR

The NET protocol used was a modified version of the protocol developed by Churchman and Weissman (2001). One litre of cells at an OD_{600} of 0.8 were harvested by centrifugation at 4℃ for 2 minutes, washed in ice-cold PBS and the pellet snap frozen in liquid nitrogen. Cells were subjected to cryogenic lysis using a mixer mill (SPEX 6780) for 5 rounds at 10 cps, for 2 minutes with 2 minutes cooling in between.

For each immunoprecipitation of TAP-tagged Rpb3, 1 g of ground cells was re-suspended in 5 ml ice-cold lysis buffer (20 mM HEPES pH 7.4, 110 mM KOAc, 0.5% Triton X-100, 0.1% TWEEN-20, 10 mM $MnCl_2$, one complete EDTA-free proteinase inhibitor tablet (Roche #11836145001), 50 U/ml SUPERas.In RNase inhibitor (ThermoFisher Scientific #AM2694)) and 500 µl RNase-free DNase I (Promega #M6101) and incubated at 4℃ for 20 minutes. The lysate was centrifuged at 20,000 x *g* at 4 ℃ for 10 minutes. An input sample (20 µl) was collected at this point for western blotting, and for a total RNA reference (100 µl). To immunoprecipitate Rpb3-TAP, the supernatant was added to 500 µl of Protein A sepherose beads and incubated at 4℃ for 1.5 hours on a rotating wheel. The beads were centrifuged at 3000 x *g* and a non-bound fraction was collected for future western blotting.

Beads were washed 4 times in 10 ml ice-cold wash buffer (20 mM HEPES pH 7.4, 110 mM KOAc, 0.5% Triton X-100, 0.1% TWEEN-20, 10 mM $MnCl_2$, one complete EDTA-free proteinase inhibitor tablet (Roche #118361ss45001), 50 U/ml SUPERas.In RNase inhibitor (ThermoFisher Scientific #AM2694), 1 mM EDTA) rotating at 4℃

for 2 minutes followed by centrifugation at 1000 x *g* at 4℃. The beads were re-suspended in 300 µl wash buffer and transferred to a 1.5 ml tube. A fraction of the beads was collected at this point for future Western blotting. Rpb3-TAP was eluted from the beads in 300 µl TEV cleavage buffer (1X AcTEV buffer, 50 U/ml SUPERas.In RNase inhibitor (ThermoFisher Scientific #AM2694), 1 mM DTT, 5 µl AcTEV protease (ThermoFisher Scientific #12575-015)) at 4℃ on a rotating wheel overnight. A fraction of the beads (20 µl) was collected at this point for future western blotting. The beads were passed through a Pierce spin column (ThermoFisher Scientific #69725) and the elution collected (a fraction was taken for future Western blotting). The beads left in the column were re-suspended in 300 µl of ddH_2O and the column centrifuged at 1000 x *g*. The elutions were combined and RNA extracted using the Qiagen miRNeasy Mini Kit (#217004) and eluted in 30 µl of RNase-free water according to the manufacturer's instructions. Reverse transcription followed by qPCR was performed as described in section 2.11.

Chapter 3. Links between splicing and transcription

3.1 Background

Major elongation factors in yeast include Spt5, Paf1 and the CTD kinases Ctk1 and Bur1 (reviewed in Shilatifard, 2004; Perales and Bentley, 2009). In addition to the CTD of RNAPII, it is possible that elongation factors that dynamically associate with and/or modify RNAPII to facilitate transcription through the nucleosome barrier influence co-transcriptional splicing (Figure 3.1). This could occur by elongation factors facilitating the initial recruitment of splicing factors to newly synthesised pre-mRNA substrates, by providing a platform to stabilise their association once recruited, or by modulating the speed of RNAPII, none of which are mutually-exclusive.

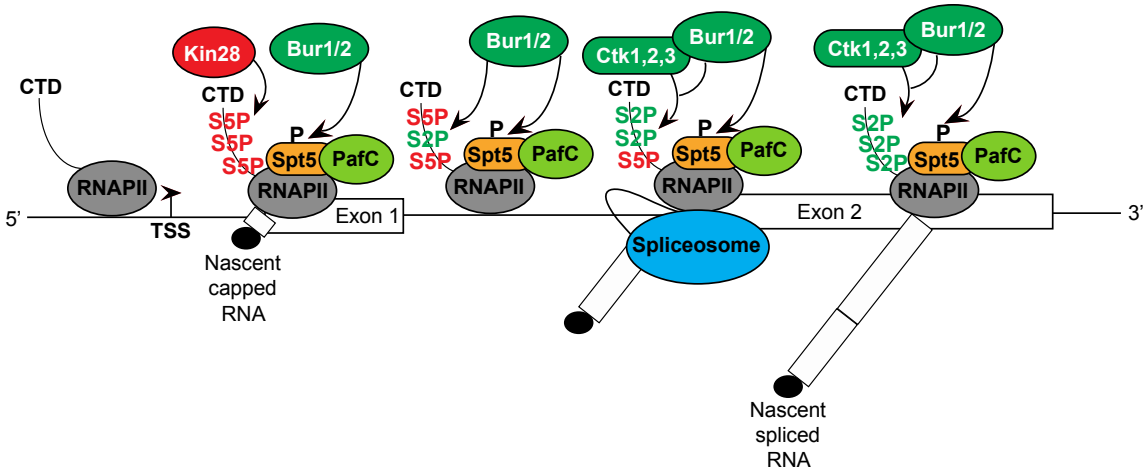


Figure 3.1. Splicing occurs in close-proximity to the transcription elongation machinery

Splicing occurs in close-proximity to the transcription elongation machinery. RNAPII is recruited to promoters in an unphosphorylated state and without Spt5. The CTD of RNAPII is then hyperphosphorylated at serine 5 at the 5' end of genes by the kinase Kin28 (budding yeast). The Bur1/2 serine 2 CTD kinase complex is recruited by serine 5 phosphorylation of the CTD at the 5'end of genes where it augments the function of major serine 2 kinase Ctk1 and phosphorylates serine 2 of the CTD towards the 5' end of genes. In addition to phosphorylating serine 2 of the CTD of RNAPII, Bur1/2 phosphorylates Spt5 transcription elongation factor. Phosphorylated Spt5 recruits the Paf1 complex (PafC) which facilitates RNAPII transcription elongation through chromatin. Ctk1 acts toward the 3'end of genes to phosphorylate serine 2 of the CTD of RNAPII. The spliceosome is recruited to pre-mRNA substrates co-transcriptionally and is by definition in close-proximity to RNAPII and factors associated with RNAPII. Additionally, the spliceosome is in close-proximity to chromatin. Figure is not to scale.

Currently, there is little functional or mechanistic insight into the contribution of the core transcription elongation machinery to co-transcriptional spliceosome assembly and splicing (reviewed in Neugebauer, 2002; Merkhofer *et al*., 2014). There is some evidence that core members of the transcription elongation complex affect splicing outcome. For example, mutations in Spt4 and Spt5 result in splicing defects in *S. cerevisiae* (Lindstrom *et al*., 2003; Burckin *et al*., 2005; Xiao *et al*., 2005), and depletion of Spt4 in mammalian cells results in changes to alternative splicing patterns (Liu *et al*., 2012). Furthermore, depletion of Spt5 in mammalian cells causes pre-mRNA accumulation on some genes (Diamant *et al*., 2012). Additionally, it was shown in yeast that Spt5 exhibits intron bias, and that Spt5 co-immunoprecipitates with Prp40, a core member of the U1 snRNP (Moore *et al*., 2006; Shetty *et al*., 2017). Most recently, depletion of Spt5 in *S. pombe* was shown by RNA sequencing to cause pre-mRNA accumulation (Shetty *et al*., 2017). Furthermore, in mammalian cells it was proposed that splicing influences the elongation rate of RNAPII *via* P-TEFb (Fong and Zhou, 2001; Lin *et al*., 2008). In mammalian cells, P-TEFb was shown to interact with splicing factor SKIP (Prp45 in *S. cerevisiae*) (Bres *et al*., 2005). More recently, in *Arabidopsis*, ELF7 (homolog of Paf1) of the Paf1 complex has been shown to interact with splicing factor SKIP (Cao *et al*., 2015; Li *et al*., 2016; Yang *et al*., 2016). Recently, elongation factors Bur1, Bur2, Spt5 and Ctk1 were found to crosslink to introns in *S. cerevisiae*, suggesting that these factors may be recruited by binding to pre-mRNA (Battaglia *et al*., 2017).

While these studies demonstrate that elongation factors are important for splicing outcome, there is no clear mechanistic insight into how this might be, and effects on co-transcriptional spliceosome assembly or co-transcriptional splicing have not been investigated. Most studies have made use of mutant or deletion systems that may have adapted over time, and therefore there may be gaps in our knowledge as to how co-transcriptional spliceosome assembly is finely tuned. Furthermore, these studies have mainly analysed the effects on splicing in total RNA (not co-transcriptional splicing) and have not analysed the effects on co-transcriptional spliceosome assembly. This is surprising as any effects on splicing are likely due to changes in a particular stage of co-transcriptional spliceosome assembly. Currently, a systematic characterisation of

the physical contribution of the core transcription elongation machinery to co-transcriptional spliceosome assembly and co-transcriptional splicing is lacking. The present work describes the use of the AID system to conditionally deplete essential and non-essential components of the transcription elongation machinery in *S. cerevisiae* in order to determine their role in co-transcriptional spliceosome assembly and splicing. The AID system has the benefit of enabling the direct determination of the physical contribution of these factors, without prior adaptation or secondary effects likely present in deletion/mutant systems. In the present work, Spt5, the Bur1/2 complex, Ctk1 and Paf1 were each conditionally depleted using the AID system and subsequent effects on RNAPII occupancy, co-transcriptional spliceosome assembly, splicing and co-transcriptional splicing were determined.

Effects on RNAPII occupancy and co-transcriptional spliceosome assembly were analysed by ChIP followed by qPCR across intron-containing genes. This was particularly important, as both the kinetic and recruitment models rely on it. For example, should depletion of a particular elongation factor reduce RNAPII occupancy, by proxy this would indirectly reduce co-transcriptional spliceosome assembly due to there being fewer pre-mRNA transcripts to bind to or due to loss of interactions between splicing machinery and RNAPII, and the conclusion that that elongation factor is required for co-transcriptional spliceosome assembly could be misinterpreted.

As splicing factors assemble on pre-mRNA during co-transcriptional splicing, their close proximity to chromatin enables them to be crosslinked to DNA and analysed by ChIP. In this way, the co-transcriptional recruitment of splicing factors and spliceosome assembly can be monitored *in vivo* (Kotovic *et al*., 2003; Görnemann *et al*., 2005; Tardiff and Rosbash, 2006). In the present work, ChIP using antibodies against core members of the spliceosome was used to test if depletion of elongation factors by the AID system affects co-transcriptional spliceosome assembly. Effects on pre-mRNA splicing and transcription were determined using RT-qPCR on total (steady-state) RNA and, if interesting effects on co-transcriptional spliceosome assembly were observed by ChIP after depletion of a particular elongation factor,

effects on co-transcriptional splicing and transcription were determined by NET-RT-qPCR Churchmann and Weissman, 2001; Aslanzadeh *et al*., 2017).

To enable depletion, each protein-of-interest was tagged at the C-terminus with the AID* cassette comprised of the AUX/IAA (AID*) recognition motif for auxin-mediated depletion and a 6 X FLAG tag for immunodetection (Morawska and Ulrich, 2013) (Figure 3.2). AID-mediated depletion of a protein-of-interest requires expression of Tir1, and tagging was performed in a W303 parental strain with *Oryza sativa TIR1* (*OsTIR1*) either constitutively expressed under a weak ADH1 promoter (P*ADH1*-396) (integrated at the *his3-11,15* locus) (Santangelo and Tornow, 1990; Mendoza-Ochoa *et al*., 2018), or conditionally expressed from a β-estradiol-inducible promoter present on a centromeric plasmid (McIsaac *et al*., 2014; Mendoza-Ochoa *et al*., 2018). *OsTIR1* expression is repressed in the absence of β-estradiol, and induced by the addition of β-estradiol, thereby enabling auxin-mediated degradation of an AID*-tagged target protein.

In the present work, correct integration of the AID* cassette at each locus was confirmed by PCR, expression was confirmed by western blotting (data not shown). Additionally, effects of the AID* tag on cell growth tested (with and without auxin treatment).

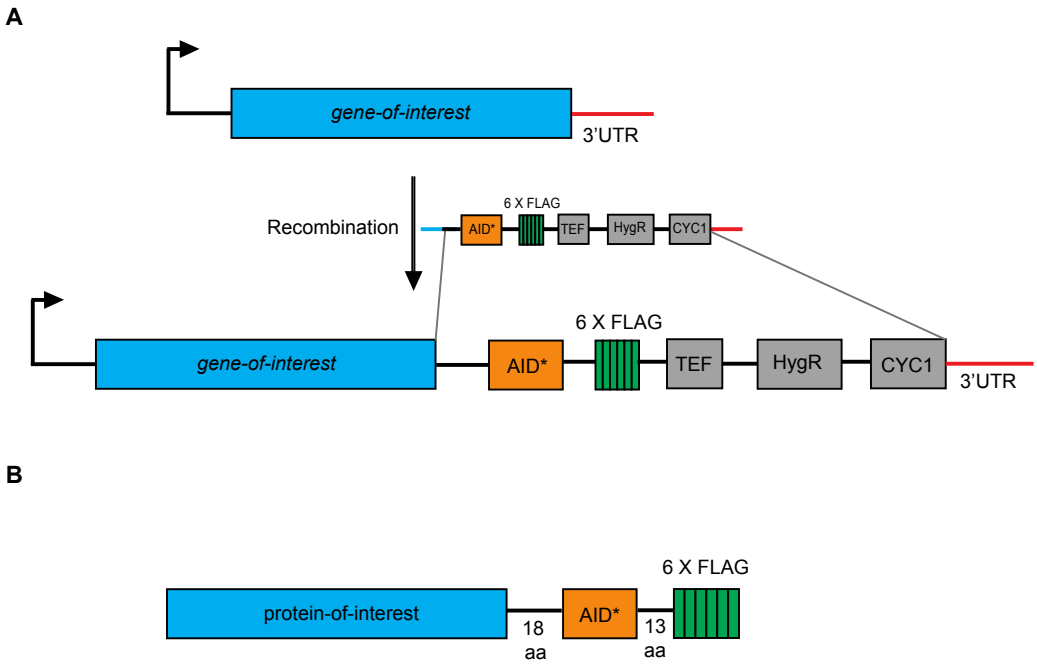


Figure 3.2. Endogenously tagging proteins-of-interest to enable depletion using the AID system

(A) In this study, endogenous genes were modified with an AID* (AUX/IAA recognition motif) sequence at the C-terminus in W303 yeast strains expressing *OsTIR*. PCR was used to amplify the AID* cassette, and each PCR product contained sequence homology within the C-terminus of each gene (blue) and the 3'UTR (red) such that, upon transformation, recombination would occur at the regions indicated by the grey lines. In addition to the AID* sequence, the cassette contains a 6XFLAG tag for immunodetection and hygromycin resistance (HygR) gene for selection of positive transformants (driven by a TEF promoter (P_{TEF})). At the 3' end, a CYC1 terminator (T_{CYC1}) is present.
(B) The tagging of each gene with the AID*-6FLAG cassette results in the production of a fusion protein with the AID* to allow auxin-mediated depletion of each protein and a 6 X FLAG tag for immunodetection. In addition, 2 linker regions of 18 amino acids (aa) and 13 aa flanked the AID* sequence (Morawska and Ulrich, 2013).

3.2 Use of the AID system to conditionally deplete Spt5

To determine whether the physical presence of Spt5 affects co-transcriptional spliceosome assembly *in vivo* in *S. cerevisiae*, Spt5 was conditionally depleted using the AID system. Spt5 was C-terminally AID*-6FLAG tagged in a strain that allowed conditional induction of *OsTIR1* using the β-estradiol system (described in section 3.1) (McIsaac *et al*., 2013; Mendoza-Ochoa *et al*., 2018). Experiments were performed with and without 40 minutes of β-estradiol and auxin treatment.

The effect of the AID* tag on Spt5 function and auxin-sensitivity was assessed by growth analysis of the parental strain (W303) with no AID* tag, and the Spt5-AID* strain with or without auxin and β-estradiol treatment (Figure 3.3A). In the absence of auxin, the Spt5-AID* strain grew comparably to the parental strain, showing that the C-terminal AID* tag does not affect Spt5 function. After auxin and β-estradiol addition to deplete Spt5, the Spt5-AID* strain grew more slowly than the parental strain – which was expected as Spt5 is essential in *S. cerevisiae*.

Western blotting showed that depletion of Spt5 for 40 minutes resulted on average in a significant reduction to 40% of the undepleted amount of Spt5 (student's t-test, $P<0.05$) (Figure 3.3B). It is estimated that Spt5 is present at approximately 4632 molecules/cell (Kulak *et al*., 2014), therefore after 40 minutes of depletion there are approximately 1853 molecules/cell of Spt5 left.

ChIP-qPCR analyses showed that, in addition to being depleted in whole cell extracts, Spt5-AID* was significantly depleted across each the intron-containing genes analysed (Figure 3.3C). On average, relative to conditions prior to depletion, there was 66%, 65% and 61% of Spt5 left across each of *ACT1*, *RPS13* and *ECM33* (student's t-test, $P<0.05$).

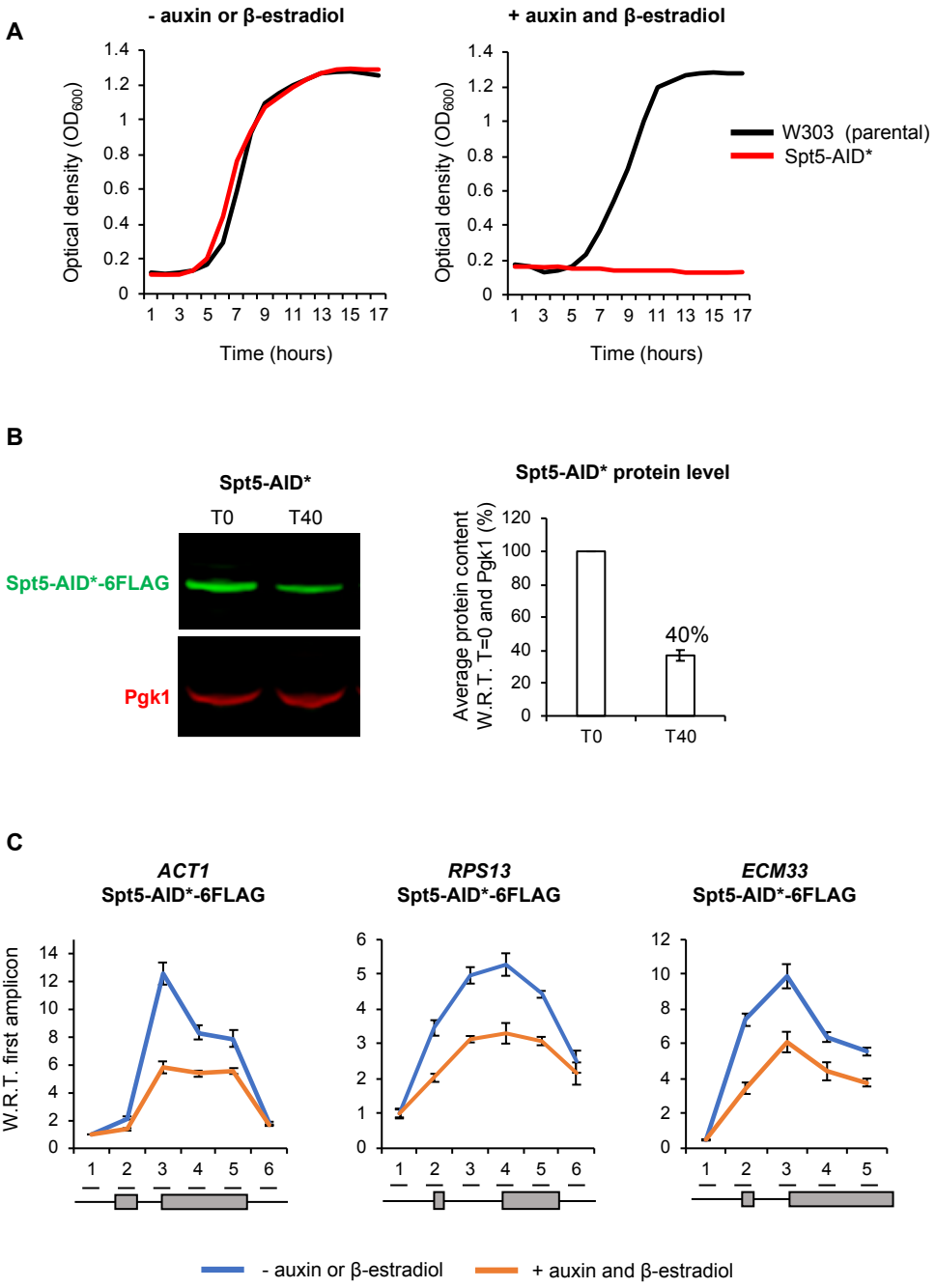


Figure 3.3. Use of the AID system to conditionally deplete Spt5

(A) Growth of the control parental strain (W303; no AID*-tagged proteins; black) and Spt5-AID* tagged strain (red) was measured over a period of 17 hours without (-) auxin or β-estradiol or with (+) auxin and β-estradiol treatment. Mean of three biological replicates.

(B) Western blot probed with anti-FLAG and anti-Pgk1 as a loading control. Samples were taken before (T0) and 40 minutes (T40) after addition of auxin and β-estradiol. Spt5-AID* depletion was quantified and shown as the percentage mean of 3 biological replicates after 40 minutes of auxin and β-estradiol addition with respect to (W.R.T) time zero and normalised to the Pgk1 signal. Error bars = standard error of the mean.

(C) Anti-FLAG ChIP followed by qPCR analysis of the intron-containing genes *ACT1*, *RPS13*, *ECM33* without (-) auxin and β-estradiol (blue) or with (+) 40 minutes of auxin and β-estradiol

(orange) addition to deplete Spt5-AID*-6FLAG. The X-axis shows the positions of amplicons used for ChIP qPCR analysis – the exons of each gene are in grey. The data are presented as the mean percentage of input with respect to (W.R.T.) the first amplicon of each gene. Mean of at least three biological replicates. Error bars = standard error of the mean.

3.3 Characterisation of the effect of Spt5 depletion on RNAPII occupancy, serine 2 and serine 5 phosphorylation of the CTD of RNAPII

To determine whether the depletion of Spt5 in these conditions affects the occupancy or phosphorylation status of RNAPII, ChIP was performed using antibodies against RNAPII (Rpb1), serine 2 (S2P) or serine 5 (S5P) CTD phosphorylation states followed by qPCR across intron-containing genes *ACT1*, *RPS13* and *ECM33*. Depletion of Spt5 for 40 minutes did not significantly affect RNAPII occupancy across *ACT1* or *ECM33* (student's t-test, $P>0.05$), however on *RPS13* there was a small but significant reduction in RNAPII occupancy on average to 81% and 84% at amplicons 2 and 3 (exon 1 and intron) (student's t-test, $P<0.05$) relative to conditions without depletion (Figure 3.4A).

ChIP-qPCR using an antibody against serine 2 (S2P) phosphorylation of the CTD showed the correct profile of serine 2 phosphorylation – an increase in this mark was observed toward the 3' end of the genes relative to RNAPII occupancy. Depletion of Spt5 resulted in increased S2P phosphorylation across the intron-containing genes tested, though only significantly at amplicon 6 of *ACT1* (exon 2), amplicon 2 of *RPS13* (exon 1) and amplicons 3 to 5 of *ECM33* (3'SS-exon 2) (student's t-test, $P<0.05$) (Figure 3.4B).

ChIP-qPCR using an antibody against serine 5 phosphorylation of the CTD (S5P) showed the correct profile – a decline in this mark toward the 3'end of the gene relative to RNAPII occupancy. Depletion of Spt5 did not significantly alter the amount of S5P at the 5'end to the middle of the genes, however, at the very 3'end of the gene there was a significant increase in S5P on *RPS13* on average to 172% at amplicon 6 (exon 2) and on *ECM33* on average to 124% at amplicon 5 (exon 2) (student's t-test, $P<0.05$) relative to conditions without depletion (Figure 3.4C).

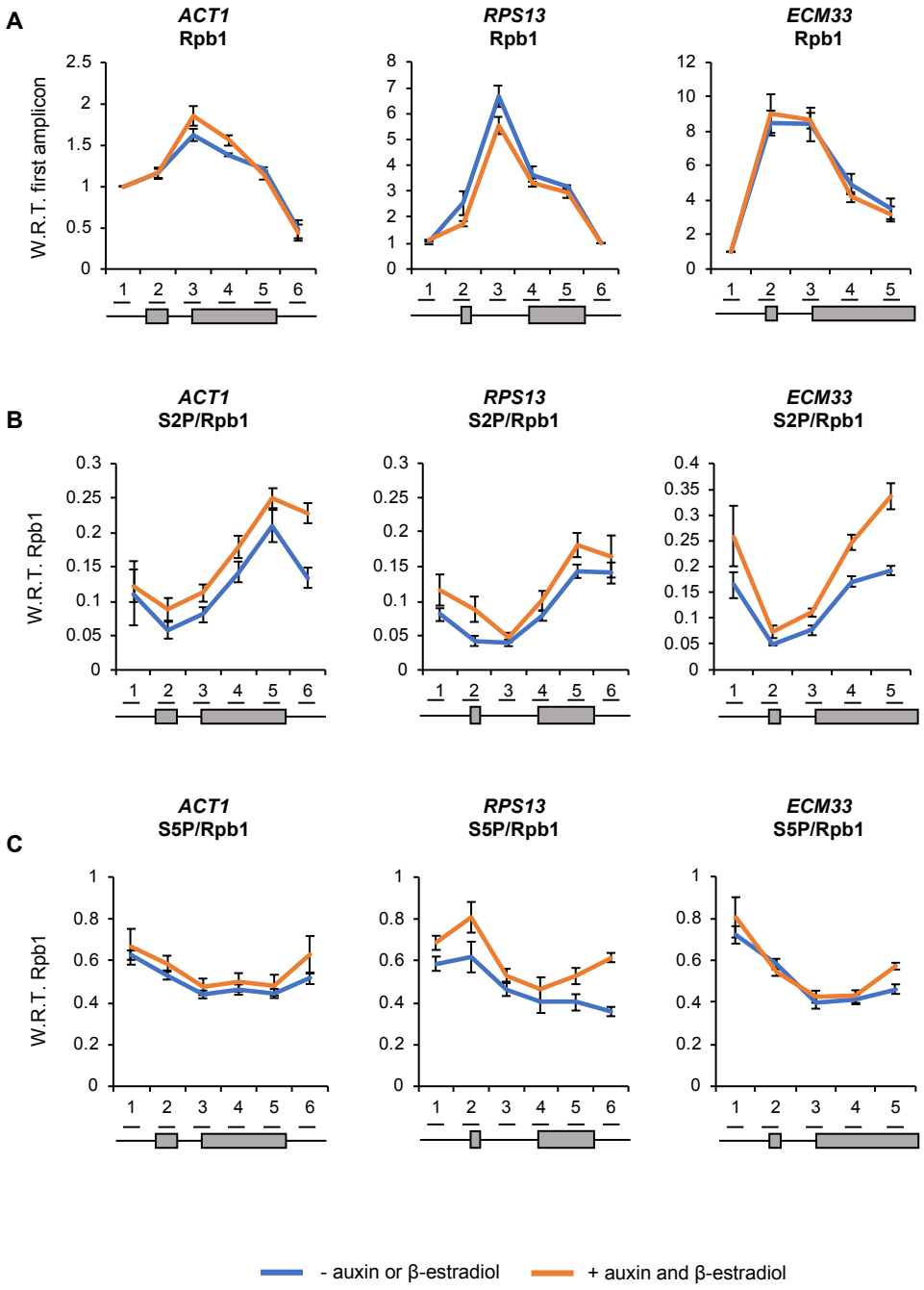


Figure 3.4. Characterisation of the effect of Spt5 depletion on RNAPII occupancy, serine 2 and serine 5 phosphorylation of the CTD of RNAPII

(A) Anti-Rpb1 (RNAPII), (B) anti-serine 2 phosphorylation (S2P) and (C) anti-serine 5 phosphorylation (S5P) ChIP and qPCR across intron-containing genes *ACT1*, *RPS13* and *ECM33* without auxin or β-estradiol (blue) and with 40 minutes of auxin and β-estradiol addition to deplete Spt5-AID* (orange). The X-axis shows the positions of amplicons used for ChIP qPCR analysis – the exons are in grey. The Rpb1 ChIP data are presented as the mean percentage of input with respect to (W.R.T.) the first amplicon of each gene. The S2P and S5P ChIP data are relative to Rpb1 occupancy. Mean of at least 3 biological replicates. Error bars = standard error of the mean.

3.4 Depletion of Spt5 reduces the co-transcriptional recruitment of the U5 snRNP without affecting earlier stages of co-transcriptional spliceosome assembly

To determine whether the physical presence of Spt5 affects co-transcriptional spliceosome assembly, ChIP-qPCR was performed across intron-containing genes *ACT1*, *RPS13* and *ECM33* with and without depletion of Spt5 for 40 minutes. Antibodies were used that detect core members of the spliceosome: the U1 snRNP (Prp40), U2 snRNP (Lea1-3HA) and U5 snRNP (Prp8), which allowed a detailed picture of which stage, if any, of co-transcriptional spliceosome assembly may be affected by depletion of Spt5.

In conditions without auxin or β-estradiol, the ChIP profile of U1, U2 and U5 was as expected – the U1 and U2 snRNPs peak near the 3'SS, and the U5 snRNP peaks downstream (toward the 3'end of the gene). ChIP-qPCR showed that depletion of Spt5 did not significantly or consistently affect U1 or U2 snRNP occupancies where they peak (student's t-test, $P>0.05$) (Figures 3.5A and 3.5B), relative to conditions without depletion. In contrast, depletion of Spt5 resulted in a significant reduction in U5 snRNP occupancy to on average 43% where the U5 snRNP peaks on *ACT1* (amplicon 5, exon 2) (student's t-test, $P<0.05$), 42% where the U5 snRNP peaks on *RPS13* (amplicon 5, exon 2) (student's t-test, $P<0.05$) and 39% and 41% where the U5 snRNP peaks on *ECM33* (amplicons 4 and 5, exon 2) (student's t-test, $P<0.05$) (Figure 3.5C), relative to conditions without depletion. As recruitment of the U1 and U2 snRNP's was unaffected, this suggests that the physical presence of Spt5 is not required for pre-spliceosome assembly but is required for either the co-transcriptional recruitment or stable association of the U5 snRNP to form a pre-catalytically activated spliceosome (B complex).

The loss of U5 snRNP occupancy upon Spt5 depletion could be explained by a reduction in the total cellular level of the U5 snRNP. To test this, western blotting was performed with an antibody against the U5 snRNP (Prp8) using extracts with and without 40 minutes of Spt5 depletion. The average cellular level of Prp8 after depletion of Spt5 was 113% relative to conditions without depletion, and this difference was not statistically significant (student's t-test, $P>0.05$) (Figure 3.5D).

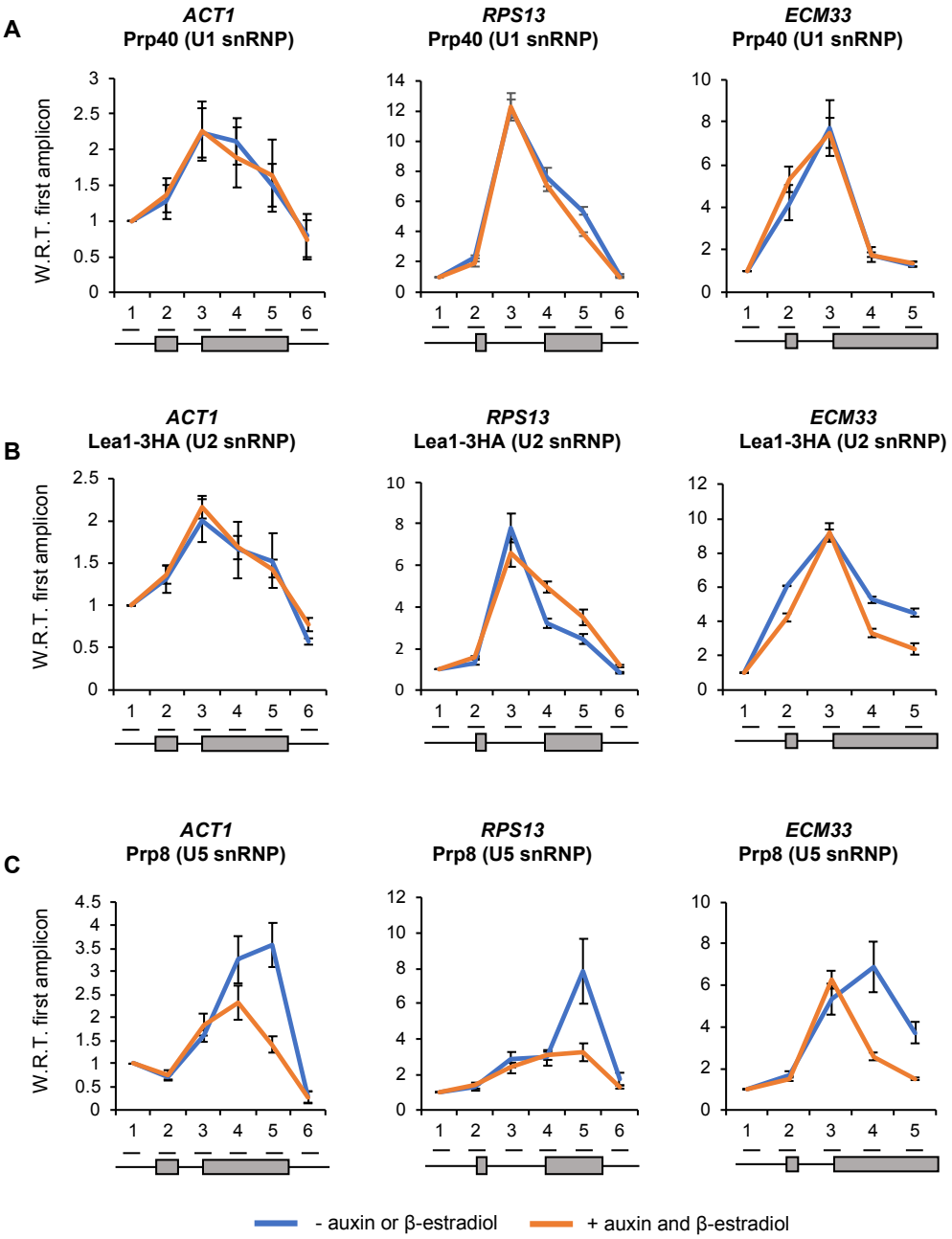


Figure 3.5. Depletion of Spt5 reduces the co-transcriptional recruitment of the U5 snRNP without affecting earlier stages of co-transcriptional spliceosome assembly

(A) Anti-Prp40 (U1 snRNP), (B) Anti-Lea1-HA (U2 snRNP) and (C) Anti-Prp8 (U5 snRNP) ChIP and qPCR across intron-containing genes *ACT1*, *RPS13* and *ECM33* without auxin or β-estradiol (blue) and with 40 minutes of auxin and β-estradiol addition to deplete Spt5-AID* (orange). The X-axis shows the positions of amplicons used for ChIP qPCR analysis – the exons are in grey. The ChIP data are presented as the mean percentage of input with respect to (W.R.T.) the first amplicon of each gene. Mean of at least 3 biological replicates. Error bars = standard error of the mean.

D

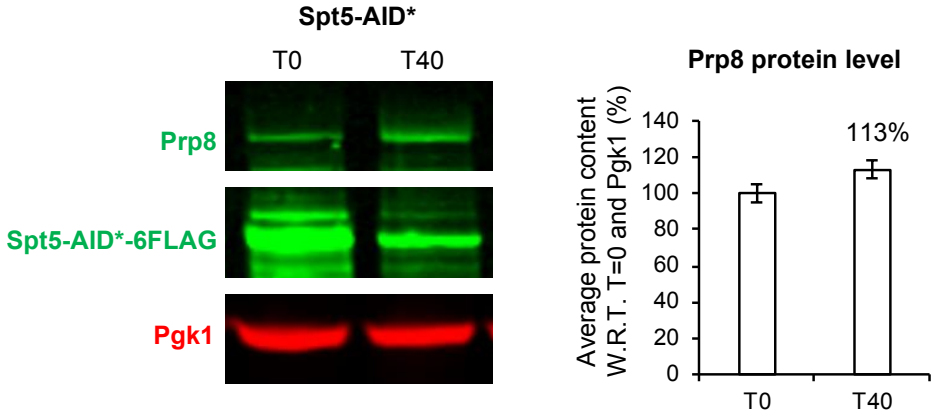


Figure 3.5 (continued). Depletion of Spt5 reduces the co-transcriptional recruitment of the U5 snRNP without affecting earlier stages of co-transcriptional spliceosome assembly

(D) Western blot probed with anti-Prp8 (U5 snRNP), anti-FLAG and anti-PGK1 as a loading control. Samples were taken before (T0) and 40 minutes (T40) after addition of auxin and β-estradiol. Prp8 was quantified and shown as the percentage mean of 3 biological replicates after 40 minutes of auxin and β-estradiol addition with respect to (W.R.T) time zero and normalised to the PGK1 signal. Error bars = standard error of the mean.

3.5 Depletion of Spt5 affects pre-mRNA splicing

Having observed an effect on the co-transcriptional recruitment of the U5 snRNP upon Spt5 depletion, it was next determined whether splicing catalysis was affected in these conditions. Splicing RT-qPCR was performed for transcripts of the intron-containing genes analysed by ChIP (*ACT1*, *RPS13* and *ECM33*). Primers were used that detected pre-mRNA (3'SS or the BP), lariat (excised intron lariat or lariat-exon 2) and mRNA.

An increase in 3'SS and BP signals are indicative of pre-mRNA accumulation and a first step splicing defect. Accumulation of the 3'SS and lariat only is indicative of a second step splicing defect (lariat-exon 2). Accumulation of the 3'SS, BP and lariat is indicative of a mixture of a first and second step splicing defect. Accumulation of lariat only (no pre-mRNA accumulation) is indicative of the excised intron-lariat that may be still in the post-spliceosome complex. Lariat primers were only available for *ACT1*,

as it is a particularly difficult species to design primers for. All primers used were designed by David Barrass (Beggs lab). This RT-qPCR assay was used to determine which stage of splicing catalysis, if any, was affected by depletion of Spt5 for 40 minutes (Figure 3.6A).

RT-qPCR on total (steady state) RNA showed that depletion of Spt5 resulted in a significant accumulation of pre-mRNA species of *ACT1* (BP and 3'SS), and a significant reduction in lariat, indicating a first step splicing defect (student's t-test, $P<0.05$) (Figure 3.6B). *ECM33* also significantly accumulated pre-mRNA (5'SS) upon Spt5 depletion, showing a defect in splicing catalysis (student's t-test, $P<0.05$) (Figure 3.6B). For *RPS13*, significant increases in pre-mRNA species (5'SS and 3'SS) were detected after depletion of Spt5, showing a defect in splicing catalysis (student's t-test, $P<0.05$) (Figure 3.6B). No significant changes were detected in exon 2 levels for any of the genes analysed, indicating no effect on transcription upon Spt5 depletion (Figure 3.6B).

Overall, the splicing RT-qPCR data fit with the ChIP analysis (Figure 3.5) that Spt5 depletion affects co-transcriptional recruitment of the U5 snRNP and B complex formation, which would be expected to cause a defect in the first step of pre-mRNA splicing.

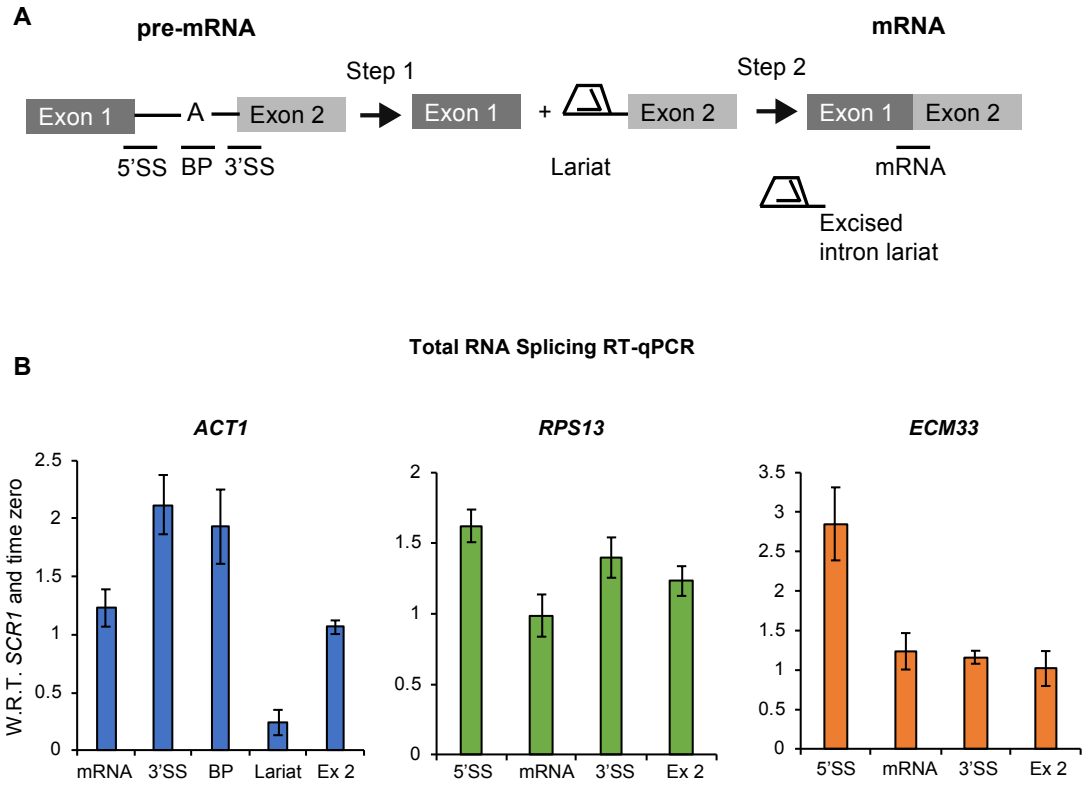


Figure 3.6. Depletion of Spt5 affects pre-mRNA splicing

(A) Splicing RT-qPCR cartoon showing the primers used for the splicing RT-qPCR assay for an average gene. Primers used detected pre-mRNA (5'SS or BP and 3'SS), lariat (excised intron or intron-exon 2), exon 2 (ex 2) and mRNA.
(B) RT-qPCR analysis of the intron-containing genes *ACT1*, *RPS13* and *ECM33* after depletion of Spt5-AID*. Normalised to the *SCR1* PolIII transcript and time zero (without auxin and β-estradiol addition). Mean of at least 3 biological replicates. Error bars = standard error of the mean.

3.6 Spt5 interacts with core members of the spliceosome

It was possible that the effect of Spt5 depletion on co-transcriptional recruitment of the U5 snRNP was due to a physical interaction between Spt5 and the spliceosome in wild-type conditions. To test this, co-immunoprecipitation experiments were performed in which Spt5-AID*-6FLAG was pulled down using a FLAG antibody, followed by Western blotting using antibodies against core members of the spliceosome: the U1 snRNP (Prp40), U2 snRNP (Lea1-3HA) and U5 snRNP (Prp8). Interaction between Spt5-AID*-6FLAG and the U5 snRNP was detected by western blotting (Figures 3.7A and 3.7B). This interaction was RNA-independent, as the

addition of RNase did not affect the co-immunoprecipitation (Figures 3.7A and 3.7B). No pull down of Prp8 was detected using a control strain with untagged Spt5, meaning the co-immunoprecipitation was specific. No interaction was detected between Spt5-AID*-6FLAG and the U1 snRNP (Figure 3.7B) or the U2 snRNP proteins (Figure 3.7C).

It was possible that the interactions detected between Prp8 and Spt5 was indirect, *via* RNAPII. To test this, co-immunoprecipitation experiments were performed in which the U1 snRNP (Prp40), U2 snRNP (Lea1-3HA) and U5 snRNP (Prp8) were each pulled down, and western blotting performed using a FLAG antibody to detect Spt5-AID*-6FLAG, and an antibody against RNAPII (Rpb1). Under these conditions, the U1, U2 and U5 snRNPs each pulled-down Spt5-AID*-6FLAG (Figure 3.7D) but with different efficiencies – the U2 and U5 snRNPs pulled down more Spt5-AID*-6FLAG than the U1 snRNP. These differences are not explained by differences in protein abundance, since the predicted number of molecules/cell of each Prp40, Lea1 and Prp8 were similar (372, 203 and 225 respectively (Kulak *et al*., 2014)) and each was well immunoprecipitated (Figures 3.7D-3.7F). In theory, this could be due to the different antibodies used or the different snRNP affinities for Spt5.

An interaction between the U2 snRNP and Rpb1 was detected (Figure 3.7E). No interaction between the U1 snRNP or U5 snRNP with Rpb1 was detected (Figure 3.7E). This indicates that the co-immunoprecipitation between the U2 snRNP and Spt5 that was detected could be indirect, *via* RNAPII. As the U1 snRNP and U5 snRNP co-immunoprecipitate Spt5 without RNAPII, these interactions may be more specific. Overall, the most consistent (reciprocal) interaction was detected between the U5 snRNP and Spt5, though I cannot rule out interactions with other splicing factors that may be more transient.

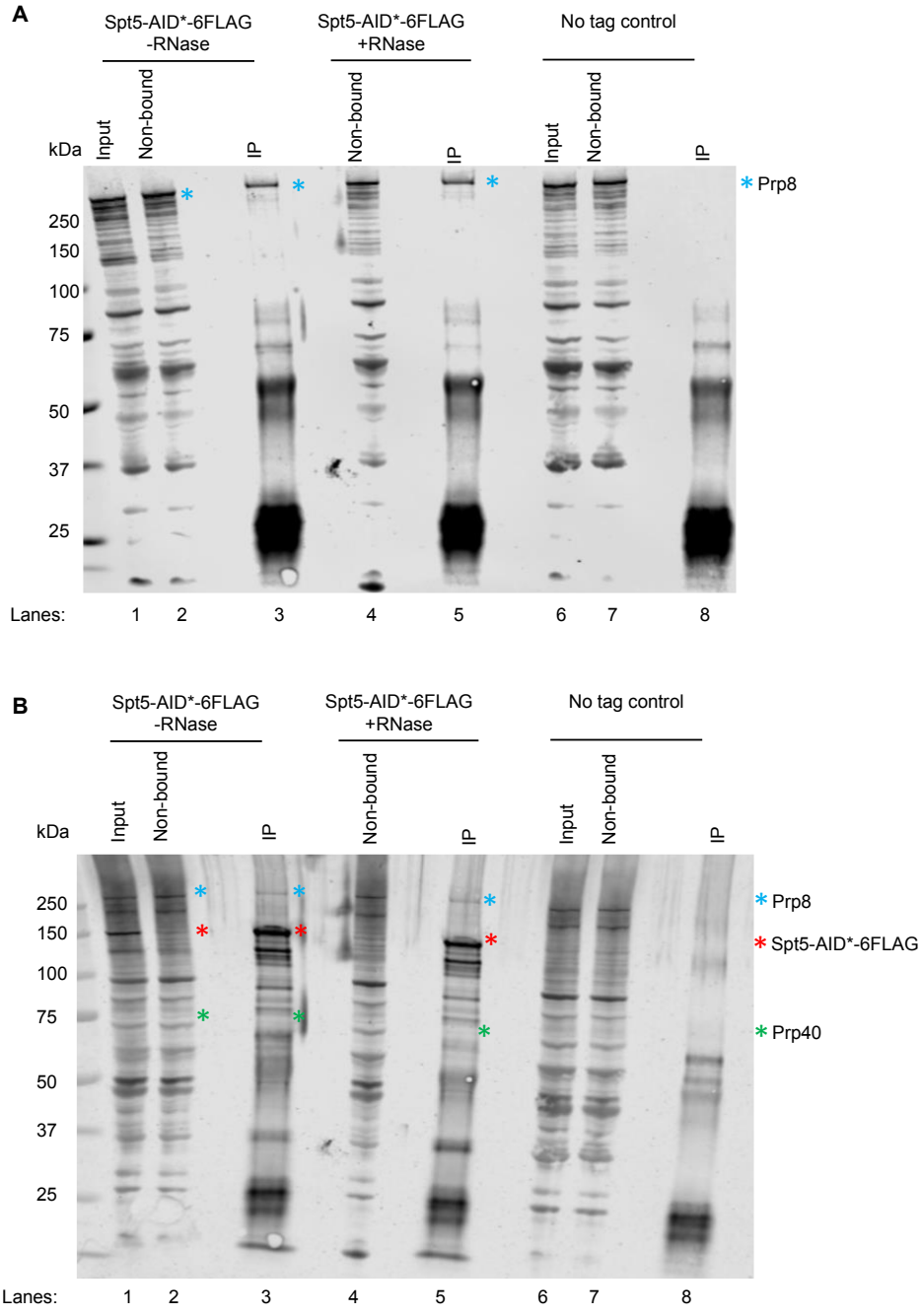


Figure 3.7. Spt5 interacts with core members of the spliceosome

(A) Western blots from a co-immunoprecipitation experiment in which Spt5-AID*-6FLAG was pulled down using a FLAG antibody with (lane 3) or without RNase (lane 5). Additionally, a strain with no flag-tagged Spt5 was used as a negative control (lane 8). Input (lanes 1, 6), non-bound (lanes 2, 4, 7) and immunoprecipitation (IP) (lanes 3, 5, 8) samples were loaded. The blot was probed with an antibody against Prp8 (U5 snRNP). Prp8 is indicated by the blue asterisk.
(B) The blot in (A) was re-probed with a FLAG antibody to show that Spt5-AID*-6FLAG was successfully pulled-down in this assay. Spt5-AID*-6FLAG is indicated by the red asterisk. The blot was re-probed with an antibody against Prp40 (U1 snRNP), indicated by the green asterisk.

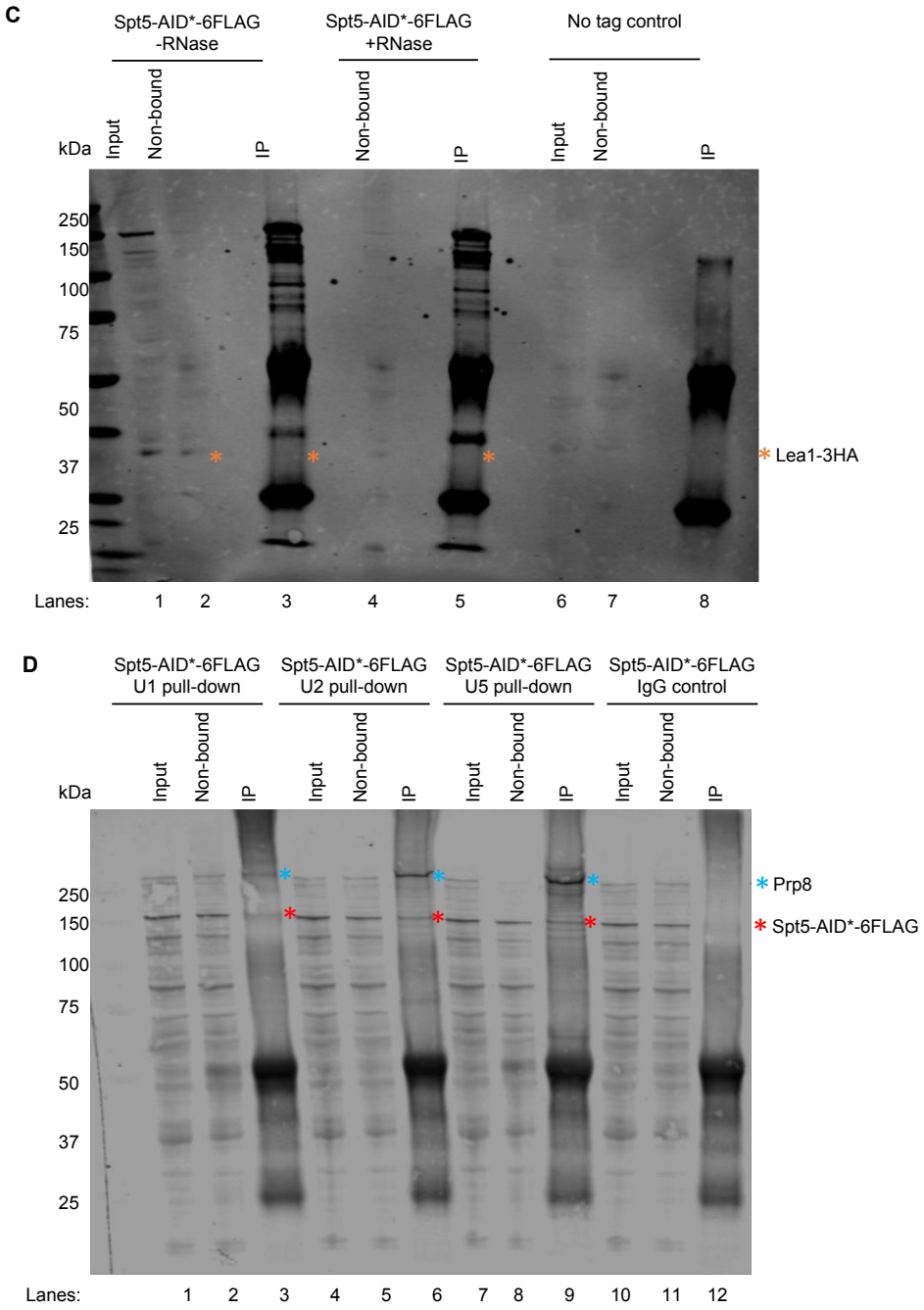


Figure 3.7 (continued). Spt5 interacts with core members of the spliceosome

(C) The blot was also re-probed with an antibody against HA (Lea1; U2 snRNP), indicated by the orange asterisk.
(D) Western blot from co-immunoprecipitation experiments in which the U1 snRNP (Prp40) (lane 3), U2 snRNP (Lea1-HA) (lane 6) and U5 snRNP (Prp8) (lane 9) were each pulled down using the indicated antibodies in the Spt5-AID*-6FLAG strain. Additionally, a negative rabbit IgG control was included in which rabbit IgG was used for a pull-down in the Spt5-AID*-6FLAG strain (lane 12). Input (lanes 1, 4, 7, 10) non-bound (lanes 2, 5, 8, 11) and immunoprecipitation (IP) (lanes 3, 6, 9, 12) samples were loaded. The blot was probed with antibodies against Prp8 (U5 snRNP) and FLAG (Spt5-AID*-6FLAG). Prp8 is indicated by the blue asterisk. Spt5-AID*-FLAG is indicated by the red asterisk.

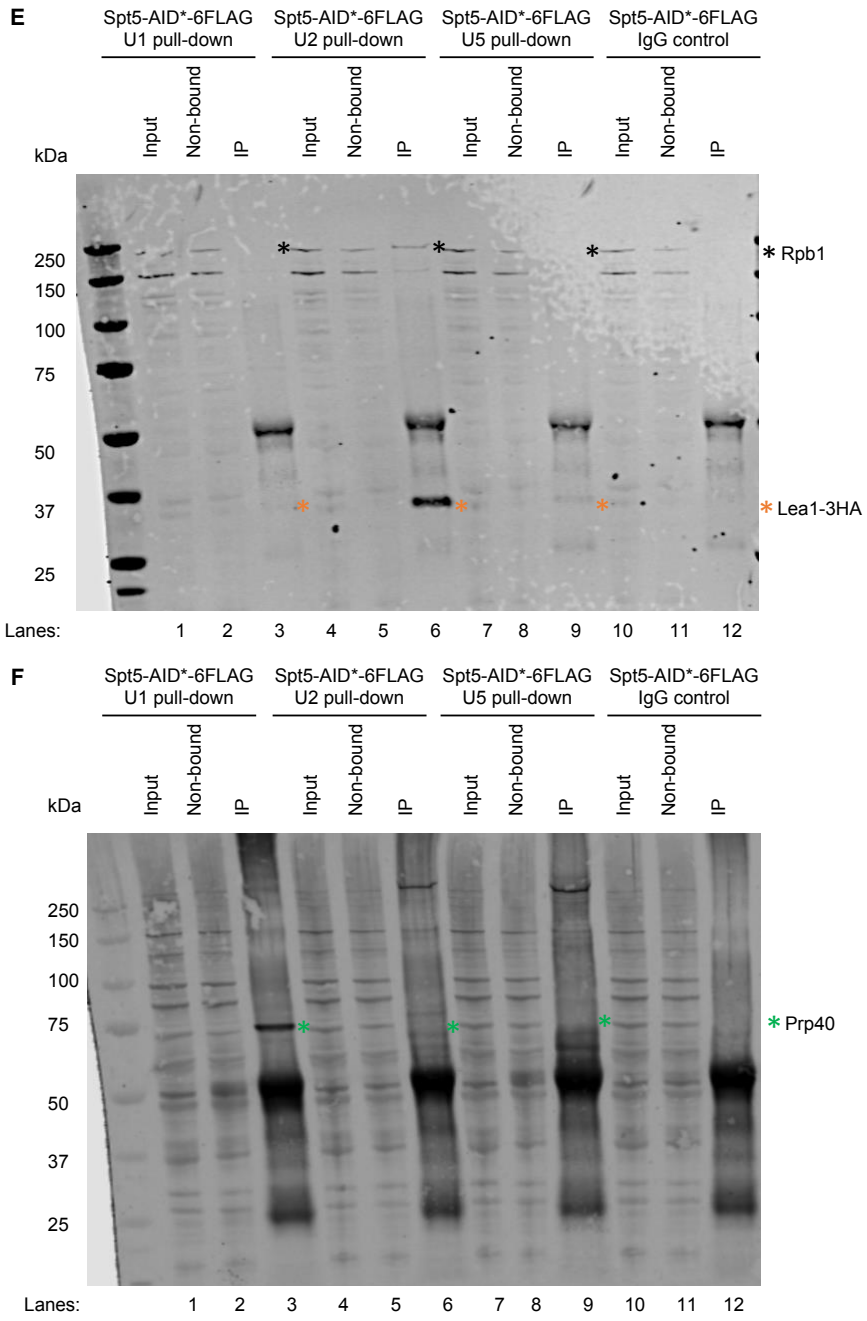


Figure 3.7 (continued). Spt5 interacts with core members of the spliceosome

(E) The blot in (D) was washed and re-probed with antibodies against Rpb1 (RNAPII) and HA (Lea1). Rpb1 is indicated by the black asterisk. Lea1-HA is indicated by the orange asterisk.
(F) The blot in (D) was re-probed with an antibody against Prp40 (U1 snRNP). Prp40 is indicated by the green Asterisk.
Predicted molecular weights (kDa): Spt5-AID*-6FLAG = 129 kDa, Prp8 = 280 kDa, Prp40 = 69 kDa.

3.7 Use of the AID system to conditionally deplete the Bur1/2 complex

The CTR of Spt5 is phosphorylated by the Bur1/2 complex in *S. cerevisiae* (Liu *et al*., 2009). Bur1 is the essential cyclin-dependent kinase and Bur2 is its non-essential cyclin by homology to other cyclins - although its expression does not cycle (Yao *et al*., 2000). To determine whether the effect of loss of Spt5 is equivalent to loss of its phosphorylation, and whether Bur1/2 play a role in co-transcriptional spliceosome assembly, Bur1 and Bur2 were each C-terminally AID*-tagged in a strain that constitutively expressed Tir1 that allowed their conditional depletion in the presence of auxin (described in section 3.1). Both Bur1 and Bur2 were analysed to allow one to distinguish specifically whether it is the physical presence of Bur1 or its kinase activity that is important for any observations made. Experiments were performed with or without 30 minutes of auxin treatment.

The effect of the tag on function of Bur1 and Bur2, and auxin-sensitivity, was assessed by growth analyses of the parental strain (W303-OsTIR) with no AID* tag and the Bur1-AID* and Bur2-AID* strains with or without auxin addition (Figure 3.8A). In the absence of auxin, the Bur1-AID* and Bur2-AID* strains grew comparably to the parental strain, indicating that the AID* tag does not affect Bur1 or Bur2 function. After auxin addition, the Bur1-AID* and Bur2-AID* strains each grew slightly more slowly than the parental strain. The effect on growth observed upon Bur1 depletion was less severe than expected, as Bur1 is essential.

Western blotting showed that 30 minutes of auxin treatment significantly reduced the level of Bur1 and Bur2 on average to 12% each relative to conditions prior to auxin treatment (student's t-test, $P<0.05$) (Figure 3.8B). It is estimated that Bur1 is present at approximately 1056 molecules/cell, and Bur2 at approximately 682 molecules/cell, giving approximately 127 molecules/cell of Bur1 and 82 molecules/cell of Bur2 remaining after auxin treatment (Kulak *et al*., 2014).

ChIP-qPCR analyses showed that, in addition to being depleted in whole cell extracts, Bur1 and Bur2 were both well depleted across the intron-containing genes *ACT1*,

RPS13 and *ECM33* in response to auxin (Figure 3.8C). Bur1 was significantly depleted on average to 66% across *ACT1,* 67% across *RPS13* and 46% across *ECM33* (student's t-test, $P<0.05$). Bur2 was significantly depleted on average to 50% across each of *ACT1* and *RPS13* and 58% across *ECM33*, relative to conditions prior to depletion (student's t-test, $P<0.05$).

To test whether depletion of Bur2 affects either the protein level or the recruitment of Bur1 to genes, Bur1 was C-terminally V5 epitope tagged in the Bur2-AID* strain. Western blotting showed that depletion of Bur2 did not significantly affect the total cellular level of Bur1 (101% relative to conditions prior to Bur2 depletion) (student's t-test, $P>0.05$) (Figure 3.8D). ChIP-qPCR showed that depletion of Bur2 did not significantly change Bur1-V5 occupancy across intron containing genes *ACT1*, *RPS13* and *ECM33* (student's t-test, $P>0.05$), relative to conditions prior to depletion (Figure 3.8E). Depletion of Bur2 was therefore a suitable way to distinguish whether the physical presence of Bur1 or its kinase activity was responsible for any changes observed in co-transcriptional spliceosome assembly and splicing.

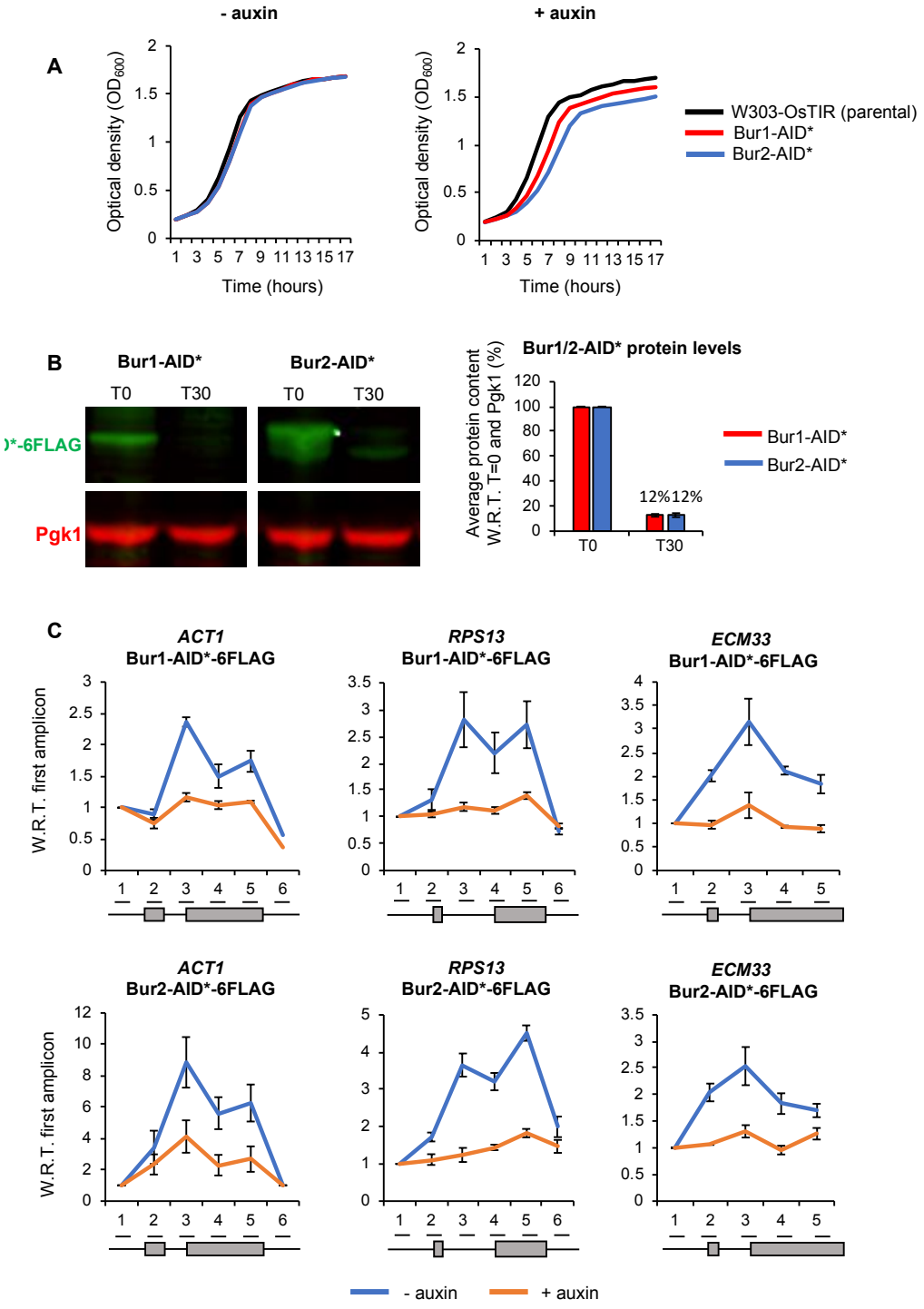


Figure 3.8. Use of the AID system to conditionally the Bur1/2 complex

(A) Growth (OD_{600}) of the control parental strain (W303-OsTIR; no AID*-tagged proteins; black), Bur1-AID*-tagged strain (red) and Bur2-AID*-tagged (blue) strain was measured over a period of 17 hours without auxin (-) or with auxin (+). Mean of three biological replicates.
(B) Western blot probed with anti-FLAG and anti-Pgk1 as a loading control. Samples were taken before (T0) and 30 minutes (T30) after addition of auxin. The depletion of Bur1-AID* (red) and Bur2-AID* (blue) was quantified and shown as the percentage mean of 3 biological

replicates after 30 minutes of addition with respect to (W.R.T) T0 and normalised to the PG1 signal. Error bars = standard error of the mean.
(C) Anti-FLAG ChIP followed by qPCR analysis of the intron-containing genes: *ACT1*, *RPS13*, *ECM33* 0 minutes (- auxin; blue) and 30 minutes (+ auxin; orange) after auxin addition to deplete Bur1-AID* or Bur2-AID*. The X-axis shows the positions of amplicons used for ChIP qPCR analysis – the exons are in grey. The data are presented as the mean percentage of input with respect to (W.R.T.) the first amplicon of each gene. Mean of at least three biological replicates. Error bars = standard error of the mean.

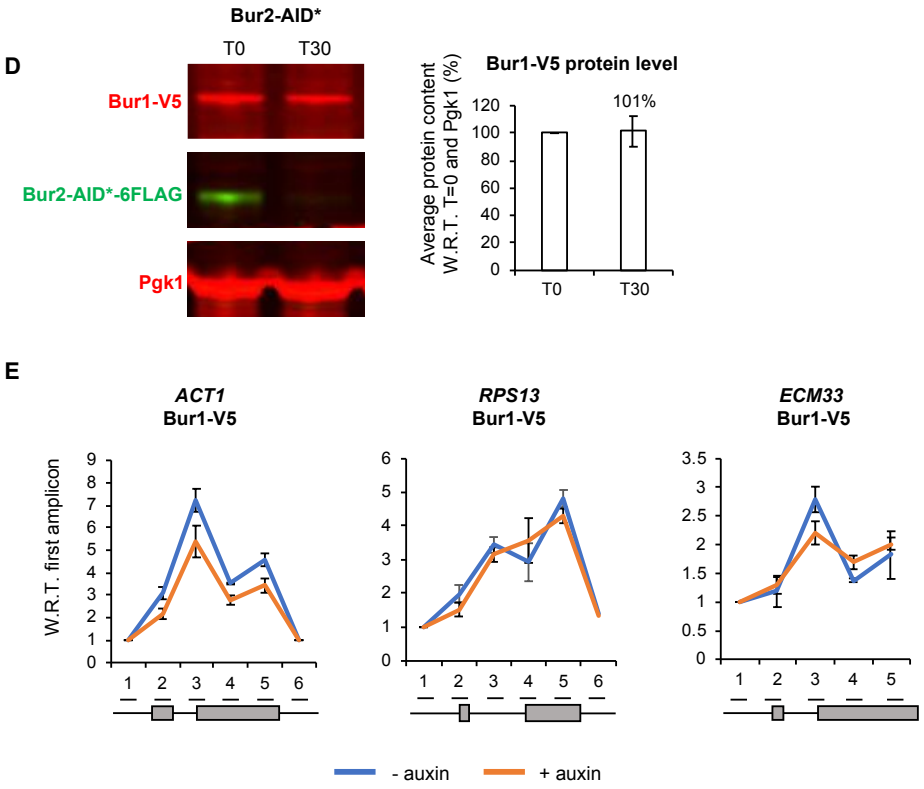


Figure 3.8 (continued). Use of the AID system to conditionally the Bur1/2 complex

(D) Western blot probed with anti-V5 and anti-Pgk1 as a loading control. Samples were taken before (T0) and 30 minutes (T30) after addition of auxin to deplete Bur2-AID* in a strain in which Bur1 was V5 tagged. The amount of Bur1-V5 was quantified and shown as the percentage mean of 3 biological replicates at T0 with respect to (W.R.T) T0 and normalised to the Pgk1 signal. Error bars = standard error of the mean.
(E) Anti-V5 ChIP followed by qPCR analysis of the intron-containing genes: *ACT1*, *RPS13*, *ECM33* at T0 (- auxin; blue) and after 30 minutes (+ auxin; orange) of auxin treatment to deplete Bur2-AID*. The X-axis shows the positions of amplicons used for ChIP qPCR analysis – the exons are in grey. The data are presented as the mean percentage of input with respect to (W.R.T.) the first amplicon of each gene. Mean of at least three biological replicates. Error bars = standard error of the mean.

3.8 Depletion of Bur1 and Bur2 reduces Spt5 phosphorylation, without reducing serine 2 or serine 5 phosphorylation of the CTD of RNAPII

To determine whether depletion of Bur1 or Bur2 for 30 minutes was an appropriate amount of time to reduce Spt5 phosphorylation, western blotting was performed using an antibody specific to the phosphorylated form of Spt5 (gift from Prof. Steven Hahn (Liu *et al*., 2009)) and an antibody specific to the central region of Spt5 (kind gift from Prof. Grant Hartzog (Hartzog *et al*., 1998)). Additionally, western blotting was performed using antibodies against serine 2 (S2P) or serine 5 (S5P) phosphorylated forms of the CTD of RNAPII to determine any effects on these CTD phosphorylation states upon Bur1 or Bur2 depletion. The quantitation of the S2P or S5P phosphorylated forms of the CTD were normalised to Rpb3, a subunit of RNAPII, to account for potential differences in the amount of RNAPII upon depletion of Bur1 or Bur2.

Western blotting showed that depletion of Bur1 or Bur2 for 30 minutes resulted in a significant reduction in Spt5 phosphorylation on average to 18% and 5% respectively, relative to conditions prior to auxin addition and to total Spt5 (student's t-test, $P<0.05$) (Figure 3.9A). Western blotting showed that depletion of Bur1 or Bur2 for 30 minutes resulted in serine 2 phosphorylation levels increasing to, on average, 120% and 117% respectively, relative to conditions prior to auxin addition and to Rpb3 (Figure 3.9B). Only the increase after depletion of Bur1 was significant (student's t-test, $P<0.05$) (Figure 3.9B). Serine 5 phosphorylation levels increased significantly to, on average, to 141% and 123% after depletion of Bur1 and Bur2 respectively relative to conditions prior to auxin addition and to Rpb3 (student's t-test, $P<0.05$) (Figure 3.9C). Depletion of Bur1 and Bur2 for 30 minutes was therefore a suitable amount of time to reduce Spt5 phosphorylation.

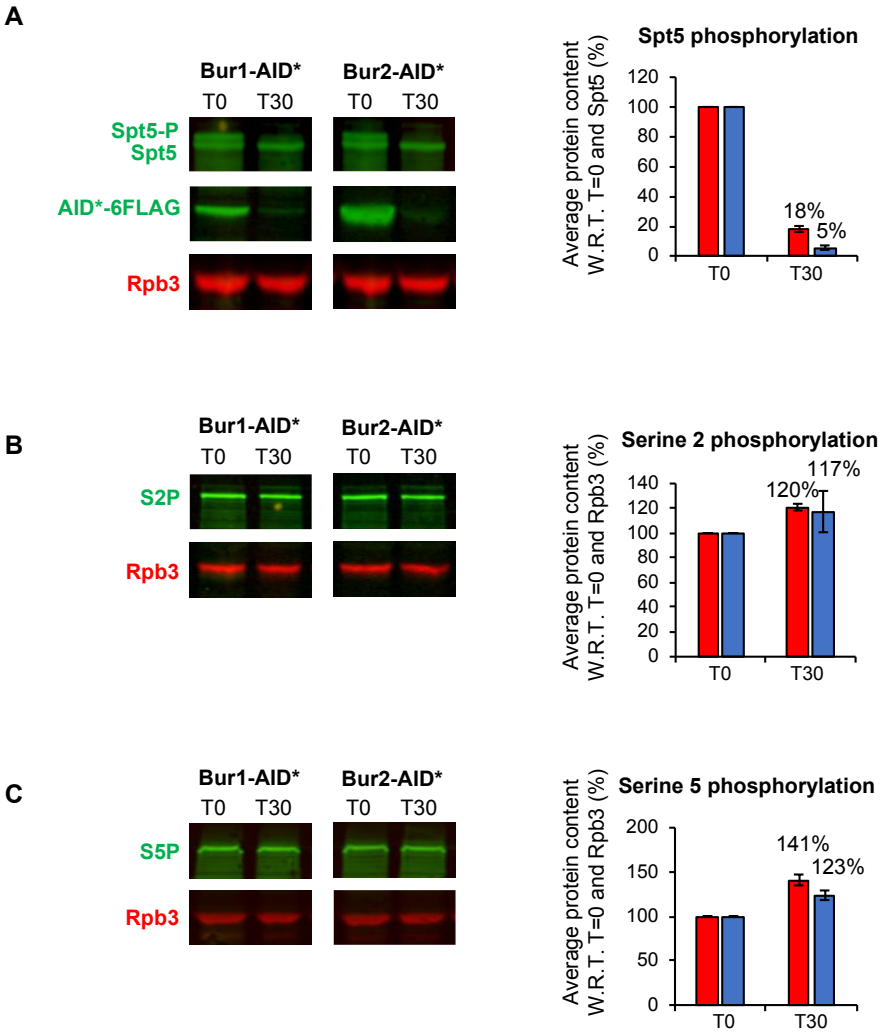


Figure 3.9. Depletion of Bur1 and Bur2 reduces Spt5 phosphorylation, without reducing serine 2 or serine 5 phosphorylation of the CTD of RNAPII

(A) Western blot probed with antibodies against phosphorylated Spt5 (anti-Spt5-P), anti-Spt5, anti-FLAG and anti-Rpb3 (RNAPII; as a loading control). Samples were taken before (T0) and 30 minutes (T30) after addition of auxin to deplete Bur1-AID* or Bur2-AID*. Spt5 phosphorylation levels were quantified after depletion of Bur1-AID* (red) or Bur2-AID* (blue) and shown as the percentage mean of 3 biological replicates after 30 minutes of auxin treatment with respect to (W.R.T) time zero and normalised to Spt5 signal.
(B) Western blot probed with anti-serine 2 phosphorylation (S2P) and anti-Rpb3 (RNAPII) as a loading control. Samples were taken before (T0) and 30 minutes (T30) after addition of auxin to deplete Bur1-AID* or Bur2-AID*. Serine 2 phosphorylation levels were quantified after depletion of Bur1-AID* (red) and Bur2-AID* (blue) and shown as the percentage mean of 3 biological replicates after 30 minutes of auxin addition with respect to (W.R.T) time zero and normalised to Rpb3 signal.
(C) Western blot probed with anti-serine 5 phosphorylation (S5P) and anti-Rpb3 (RNAPII) as a loading control. Samples were taken before (T0) and 30 minutes (T30) after addition of auxin to deplete Bur1-AID* or Bur2-AID*. Serine 5 phosphorylation levels were quantified after depletion of Bur1-AID* (red) and Bur2-AID* (blue) and shown as the percentage mean of 3 biological replicates after 30 minutes of auxin addition with respect to (W.R.T) time zero and normalised to Rpb3 signal.
All data (A-C): mean of three biological replicates, error bars = standard error of the mean.

3.9 Characterisation of the effects of depletion of Bur1/2 on RNAPII occupancy, serine 2 and serine 5 phosphorylation of the CTD of RNAPII

To determine whether the depletion of Bur1 or Bur2 in these conditions affects the occupancy or phosphorylation status of RNAPII, ChIP was performed using antibodies against total RNAPII (Rpb1), phosphorylated serine 2 (S2P) or phosphorylated serine 5 (S5P) of the CTD followed by qPCR across intron-containing genes *ACT1*, *RPS13* and *ECM33*. Depletion of Bur1 for 30 minutes did not significantly affect RNAPII occupancy across *ACT1*, *RPS13* or *ECM33* (student's t-test, $P>0.05$), relative to conditions prior to depletion (Figure 3.10A). Depletion of Bur2 did not significantly affect RNAPII occupancy across *ACT1* (student's t-test, $P>0.05$), however there was a small but significant reduction in RNAPII occupancy after depletion of Bur2 on *RPS13* on average to 63% and 68% at amplicons 2 and 3 (exon 1 and intron) and on *ECM33* on average to 78% and 90% at amplicons 2 and 3 (exon 1 and exon 2) (student's t-test, $P<0.05$), relative to conditions prior to depletion (Figure 3.10D).

Neither depletion of Bur1 nor depletion of Bur2 significantly affected serine 2 (S2P) or serine 5 (S5P) phosphorylation of the CTD of RNAPII across the intron-containing genes *ACT1*, *RPS13* or *ECM33* (student's t-test, $P>0.05$), relative to conditions prior to depletion (Figures 3.10B, 3.10C, 3.10E, 3.10F). The lack of reduction in serine 2 phosphorylation of the CTD of RNAPII observed by western blotting (Figure 3.9) and ChIP upon depletion of Bur1 and Bur2 was surprising, as deletion of Bur2 and inhibition of Bur1 activity were reported to reduce the level of serine 2 phosphorylation by western blotting and by ChIP (Liu *et al*., 2009; Qiu *et al*., 2009; Dronamraju and Strahl, 2014).

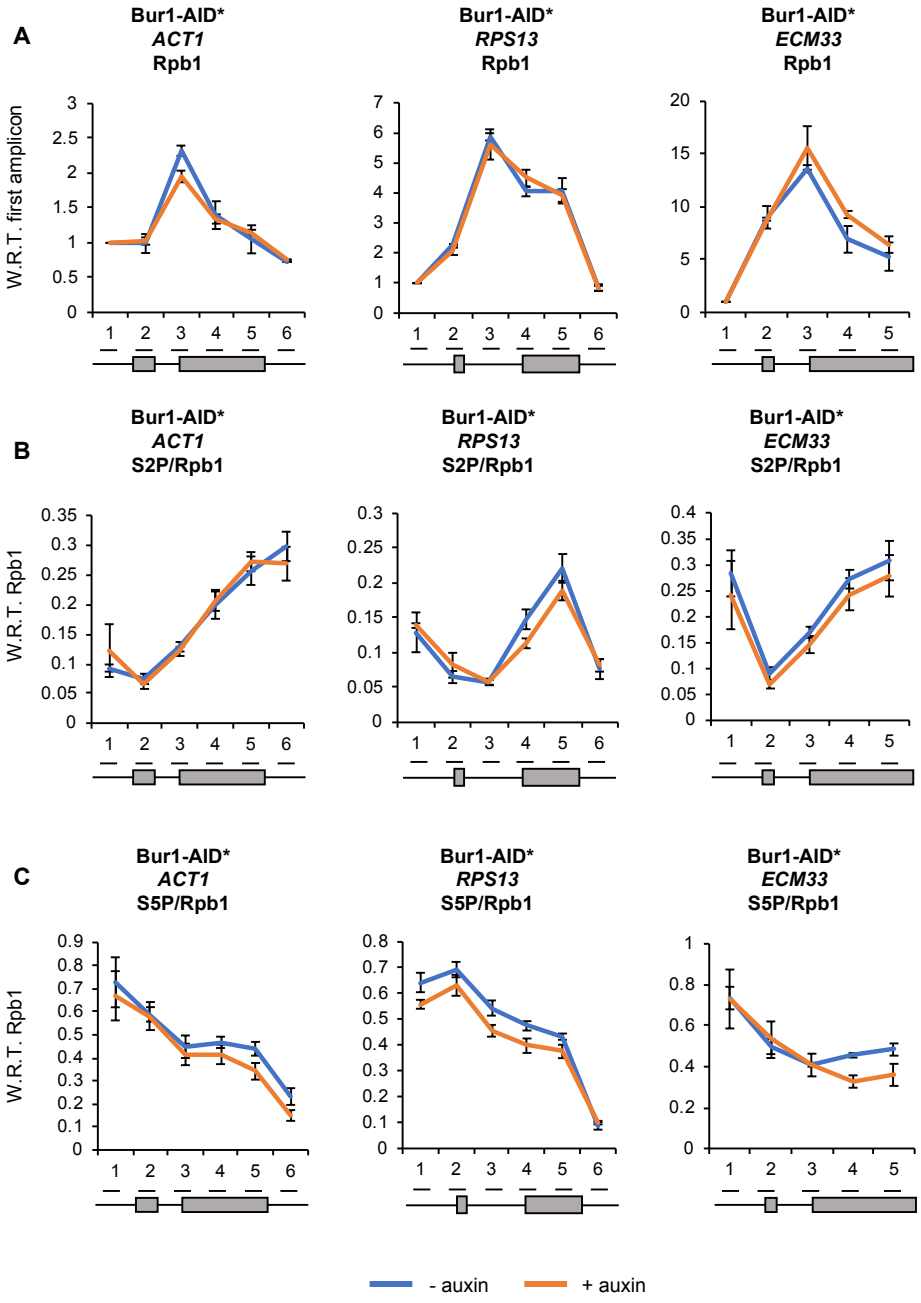


Figure 3.10. Characterisation of the effects of depletion of Bur1/2 on RNAPII occupancy, serine 2 and serine 5 phosphorylation of the CTD of RNAPII

(A, D) Anti-Rpb1 (RNAPII), (B, E) anti-serine 2 phosphorylation (S2P) and (C, F) anti-serine 5 phosphorylation (S5P) ChIP and qPCR across intron-containing genes *ACT1*, *RPS13* and *ECM33* without auxin (T0; blue) and after 30 minutes of auxin treatment to deplete Bur1-AID* (T30; orange; A, B, C) and Bur2-AID* (T30; orange; D,E,F). The X-axis shows the positions of amplicons used for ChIP qPCR analysis – the exons are in grey. The Rpb1 ChIP data are presented as the mean percentage of input with respect to (W.R.T.) the first amplicon of each gene. The S2P and S5P ChIP data are relative to Rpb1 occupancy.
All data (A-F): mean of at least 3 biological replicates, error bars = standard error of the mean.

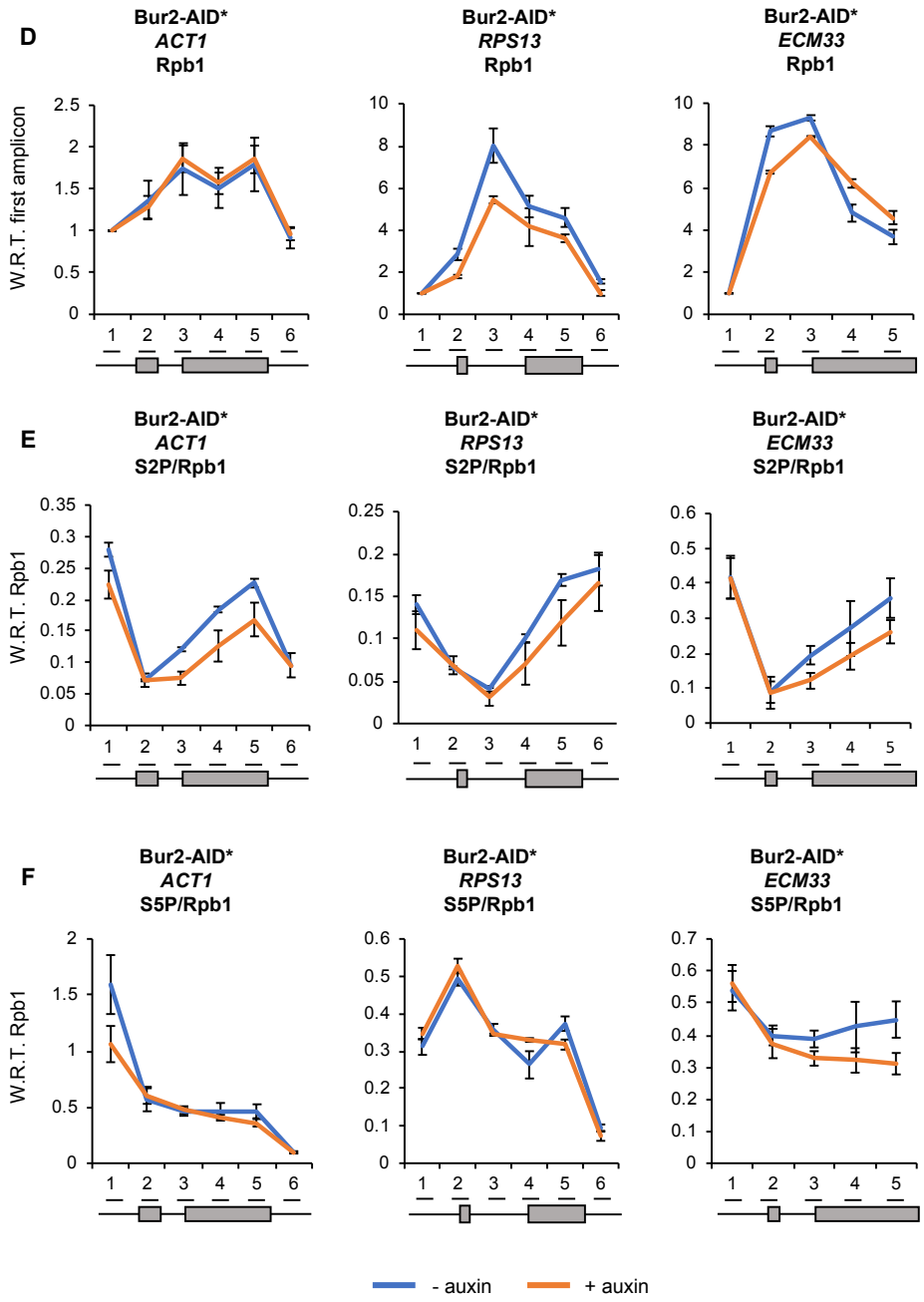


Figure 3.10 (continued). Characterisation of the effects of depletion of Bur1/2 on RNAPII occupancy, serine 2 and serine 5 phosphorylation of the CTD of RNAPII

(A, D) Anti-Rpb1 (RNAPII), (B, E) anti-serine 2 phosphorylation (S2P) and (C, F) anti-serine 5 phosphorylation (S5P) ChIP and qPCR across intron-containing genes *ACT1*, *RPS13* and *ECM33* without auxin (T0; blue) and after 30 minutes of auxin treatment to deplete Bur1-AID* (T30; orange; A, B, C) and Bur2-AID* (T30; orange; D,E,F). The X-axis shows the positions of amplicons used for ChIP qPCR analysis – the exons are in grey. The Rpb1 ChIP data are presented as the mean percentage of input with respect to (W.R.T.) the first amplicon of each gene. The S2P and S5P ChIP data are relative to Rpb1 occupancy.
All data (A-F): mean of at least 3 biological replicates, error bars = standard error of the mean.

3.10 Depletion of Bur1 causes co-transcriptional accumulation of the U2 and U5 snRNPs

To determine whether depletion of Bur1 or Bur2 in these conditions affects co-transcriptional spliceosome assembly, ChIP-qPCR was performed across intron-containing genes *ACT1*, *RPS13* and *ECM33* with or without depletion of Bur1 or Bur2 for 30 minutes. Antibodies were used that detect core members of the spliceosome: the U1 snRNP (Prp40), U2 snRNP (Lea1-3HA) and U5 snRNP (Prp8), which allowed a detailed picture of which stage, if any, of co-transcriptional may be affected by loss of Bur1 or Bur2.

ChIP-qPCR analysis showed that depletion of Bur1 did not significantly affect U1 snRNP occupancy across the intron-containing genes tested (student's t-test, $P>0.05$) (Figure 3.11A). In contrast, the U2 snRNP significantly accumulated on average to 190% and 124% on *ACT1* (amplicons 3 and 4, 3'SS and exon 2) and 198% on *ECM33* (amplicon 3, 3'SS), relative to conditions prior to depletion (student's t-test, $P<0.05$) (Figure 3.11B). Depletion of Bur1 also caused the U5 snRNP to significantly accumulate on average to 184% on *ACT1* (amplicon 4, exon 2), 186% and 174% on *RPS13* (amplicons 4 and 5, 3'SS and exon 2) and 190% on *ECM33* (amplicon 4, exon 2) (student's t-test, $P<0.05$), relative to conditions prior to depletion (Figure 3.11C).

Depletion of Bur2 did not significantly affect U1 snRNP occupancy across *ACT1* or *ECM33*, however across *RPS13* U1 snRNP occupancy was significantly reduced on average to 74.8%, relative to conditions prior to depletion (student's t-test, $P<0.05$) (Figure 3.11D). Depletion of Bur2 resulted in a significant accumulation of the U2 snRNP on *ACT1* on average to 130% (amplicon 3, 3'SS) and on *ECM33* to 130% (amplicon 3, 3'SS) (student's t-test, $P<0.05$), relative to conditions prior to depletion (Figure 3.11E). In contrast, the U2 snRNP was significantly reduced on average to 81.4% across *RPS13* upon Bur2 depletion (student's t-test, $P<0.05$), relative to conditions prior to depletion (Figure 3.11E). Depletion of Bur2 caused the U5 snRNP to significantly accumulate on average to 243% on *ACT1* (amplicon 4, exon 2) and 146% on *ECM33* (amplicon 4, exon 2), relative to conditions prior to depletion (student's t-test, $P<0.05$) (Figure 3.11F). Depletion of Bur2 did not significantly affect

the occupancy of the U5 snRNP across *RPS13*, relative to conditions prior to depletion (student's t-test, P>0.05) (Figure 3.11F).

In summary, a similar phenotype of U2 and U5 snRNP accumulation was observed at least on *ACT1* and *ECM33* upon Bur1 and Bur2 depletion. Not only did the U2 and U5 snRNPs accumulate on these genes, but they peaked earlier, indicating co-transcriptional spliceosome assembly is faster in the absence of Bur1. As the U2 and U5 snRNPs accumulated on *ACT1* and *ECM33* and not *RPS13* upon Bur2 depletion, this means the effect of Bur1/2 depletion on co-transcriptional spliceosome assembly is gene-specific. Further, as the effect of Bur2 depletion appears less severe than the effect of Bur1 depletion, it may that Bur1 has some residual kinase activity without Bur2 or have Bur2-independent substrates that are important for this phenotype. There is precedent for this as Bur1 is essential whereas Bur2 is non-essential (Yao *et al*., 2000).

The effects on co-transcriptional spliceosome assembly observed after depletion of Bur1 and Bur2 are very different to the effects of Spt5 depletion (Figure 3.5). Therefore, the loss of phosphorylation of the CTR of Spt5 is not responsible for the loss of U5 snRNP recruitment observed upon Spt5 depletion.

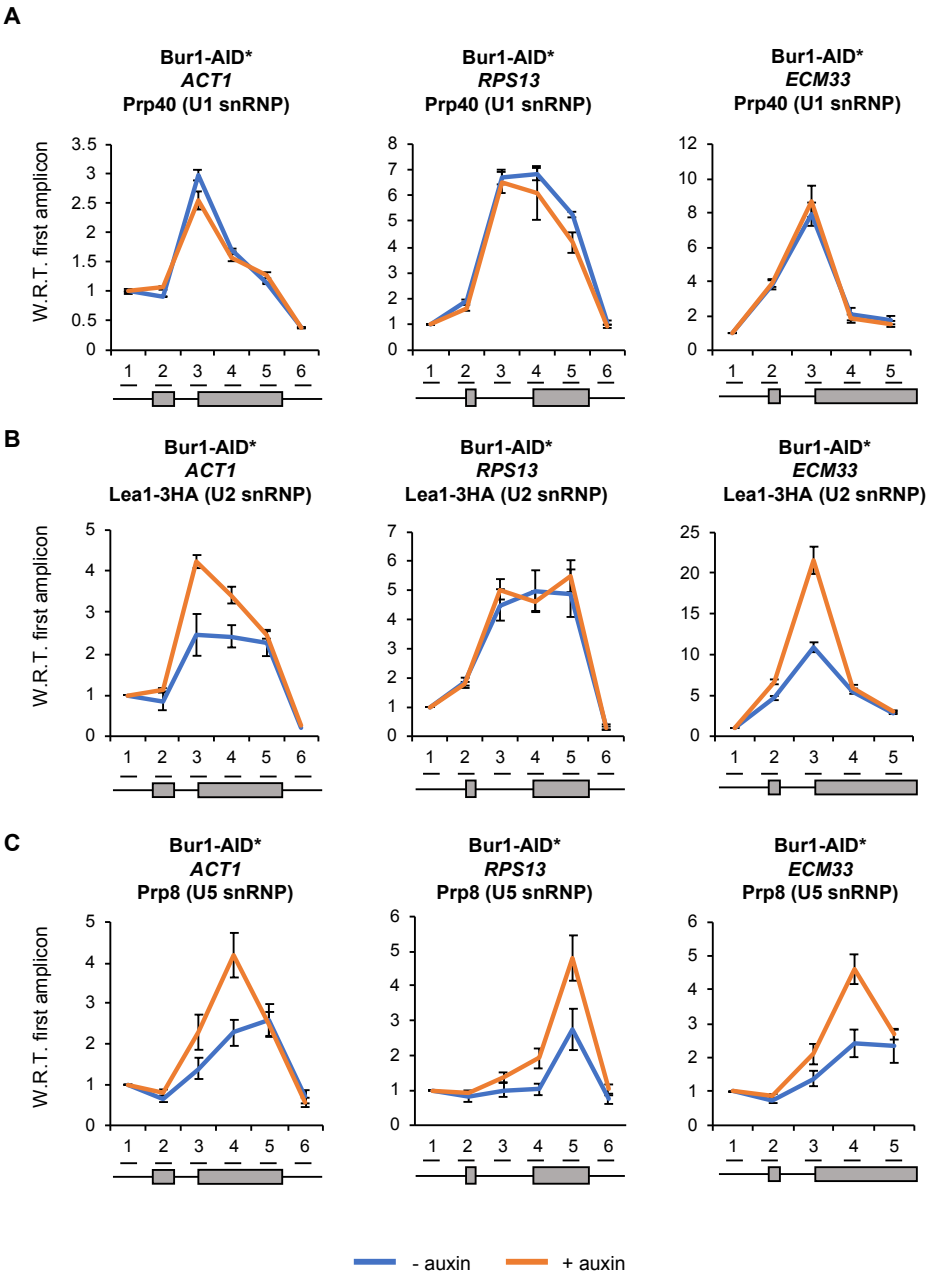


Figure 3.11. Depletion of Bur1 causes co-transcriptional accumulation of the U2 and U5 snRNPs

(A, D) Anti-Prp40 (U1 snRNP) (B, E) Anti-Lea1-HA (U2 snRNP) and (C, F) Anti-Prp8 (U5 snRNP) ChIP followed by qPCR analysis of the intron-containing genes: *ACT1*, *RPS13*, *ECM33* 0 minutes (no auxin; blue) and after 30 minutes (orange) of auxin addition to deplete Bur1-AID* (A-C) and Bur2-AID* (D-F). The X-axis shows the positions of amplicons used for ChIP qPCR analysis – the exons are in grey. The data are presented as the mean percentage of input with respect to (W.R.T.) the first amplicon of each gene.
All data (A-F): mean of at least 3 biological replicates, error bars = standard error of the mean.

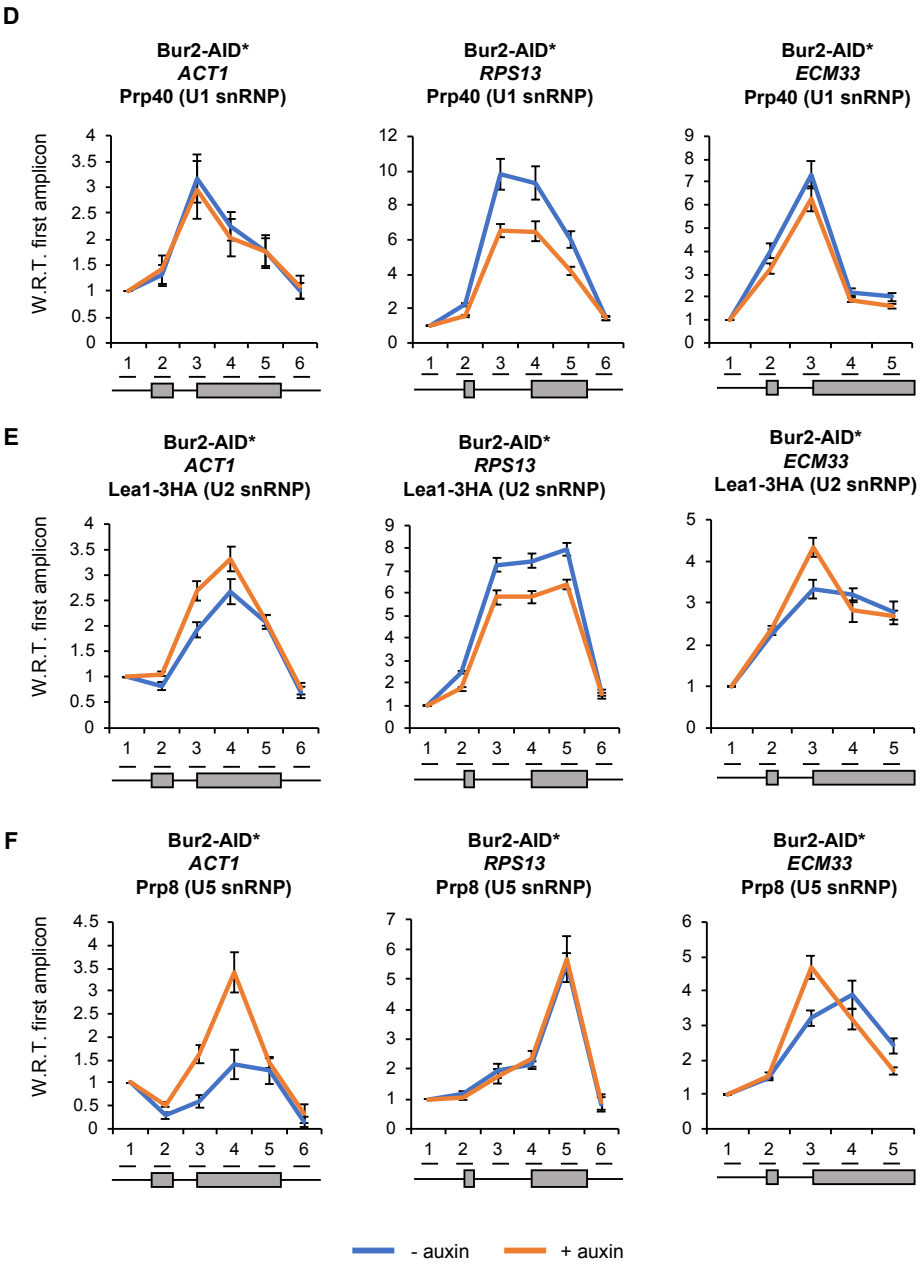


Figure 3.11 (continued). Depletion of Bur1 causes co-transcriptional accumulation of the U2 and U5 snRNPs

(A, D) Anti-Prp40 (U1 snRNP) (B, E) Anti-Lea1-HA (U2 snRNP) and (C, F) Anti-Prp8 (U5 snRNP) ChIP followed by qPCR analysis of the intron-containing genes: *ACT1*, *RPS13*, *ECM33* 0 minutes (no auxin; blue) and after 30 minutes (orange) of auxin addition to deplete Bur1-AID* (A-C) and Bur2-AID* (D-F). The X-axis shows the positions of amplicons used for ChIP qPCR analysis – the exons are in grey. The data are presented as the mean percentage of input with respect to (W.R.T.) the first amplicon of each gene.
All data (A-F): mean of at least 3 biological replicates, error bars = standard error of the mean.

3.11 Depletion of Bur1 increases co-transcriptional spicing efficiency

The accumulation of the U2 and U5 snRNP's at *ACT1* and *ECM33* after depletion of Bur1 or Bur2 could be explained in three ways. Firstly, as U1 snRNP occupancy is normal upon Bur1/2 depletion, that there is a block to the progression of co-transcriptional spliceosome assembly at the formation of the catalytically activated spliceosome (Bact complex) onwards and that complexes get stuck at the gene. Secondly, Bact formation and later steps in the splicing cycle might occur faster and more efficiently in the absence of Bur1. Thirdly, RNAPII elongates more slowly (could be pausing more frequently or for longer periods of time) and so the snRNPs are detected better by ChIP. If there was a block to progression of co-transcriptional spliceosome assembly, defects in splicing catalysis would be predicted and if the Bact complex forms more efficiently, perhaps splicing catalysis would be more efficient. To distinguish between these two possibilities, RT-qPCR was performed on total (steady-state) RNA (as described in section 3.5) upon depletion of Bur1 or Bur2. RT-qPCR showed that Bur1 depletion significantly reduced the 3'SS and BP pre-mRNA species of *ACT1*, in addition to the lariat (student's t-test, $P<0.05$), whilst the mRNA and exon 2 levels did not significantly change (student's t-test, $P>0.05$) (Figure 3.12A). For *ECM33,* Bur1 depletion significantly reduced the 3'SS (student's t-test, $P<0.05$) and did not change the levels of the 5'SS, mRNA or exon 2 (student's t-test, $P>0.05$) (Figure 3.12A). No significant change in pre-mRNA species were detected for *RPS13* upon Bur1 depletion (student's t-test, $P>0.05$), however the mRNA and exon 2 levels were significantly reduced ($P<0.05$) (Figure 3.12A). In contrast, no significant reduction in pre-mRNA levels were observed for *ACT1*, *RPS13* or *ECM33* upon Bur2 depletion (student's t-test, $P>0.05$) (Figure 3.12B). The only significant difference upon Bur2 depletion was an increase in exon 2 levels of *ACT1* and the 5'SS of *ECM33* (student's t-test, $P<0.05$) (Figure 3.12B) indicating increased levels of gene transcription.

Next, the effect of Bur1 depletion on co-transcriptional splicing was tested by NET-RT-qPCR using the same primers described previously (section 3.5). NET-RT-qPCR showed a significant reduction in 3'SS, BP and lariat species associated with RNAPII

co-transcriptionally upon Bur1-AID* depletion for *ACT1* (student's t-test, P<0.05), whilst the exon 2 and mRNA signals did not significantly change (student's t-test, P>0.05) (Figure 3.12C). For *ECM33,* Bur1-AID* depletion significantly reduced the amount of 3'SS associated with RNAPII (student's t-test, P<0.05), significantly increase amount of mRNA associated with RNAPII (student's t-test, P<0.05) and did not significantly change the levels of the 5'SS, mRNA or exon 2 associated with RNAPII (student's t-test, P>0.05) (Figure 3.12C). No significant changes were detected in pre-mRNA, mRNA or exon 2 levels of *RPS13* associated with RNAPII upon Bur1-AID* depletion (student's t-test, P>0.05) (Figure 3.12C). As exon 2 levels did not significantly change, Bur1 depletion does not cause a transcription defect on *ACT1*, *ECM33* or *RPS13*.

Together, the total RNA RT-qPCR and NET-RT-qPCR experiments agree. A reduction in pre-mRNA (BP) and the products of the first step of splicing (3'SS and lariat) was observed for *ACT1*, without significant changes to mRNA or exon 2 levels, suggesting pre-mRNA is spliced faster. For *ECM33*, the 5'SS, exon 2 or mRNA species do not significantly change, however the 3'SS is significantly reduced, suggesting that the second step of pre-mRNA splicing may be faster. This indicates that different steps of splicing are more efficient upon Bur1 depletion for these genes. This effect is gene-specific, as it was not observed for *RPS13*, which fits with the ChIP-qPCR data where the U2 snRNP did not accumulate on *RPS13* (Figure 3.11). This is consistent with the idea that Bact complex formation and progression past it in the spliceosome cycle is more efficient, rather than a block to the progression of co-transcriptional spliceosome assembly. As the exon 2 levels associated with RNAPII did not significantly change (by NET-RT-qPCR), this would suggest that RNAPII does not pause more upon Bur1 depletion. The splicing RT-qPCR analysis of total RNA upon Bur1 and Bur2 depletion did not agree, with only depletion of Bur1 increasing co-transcriptional splicing efficiency. Again, these differences between Bur1 depletion and Bur2 depletion can be explained by the fact that when Bur2 is depleted, the co-transcriptional recruitment of Bur1 was unaffected, and Bur1 may have some residual kinase activity without Bur2 or have Bur2-independent substrates.

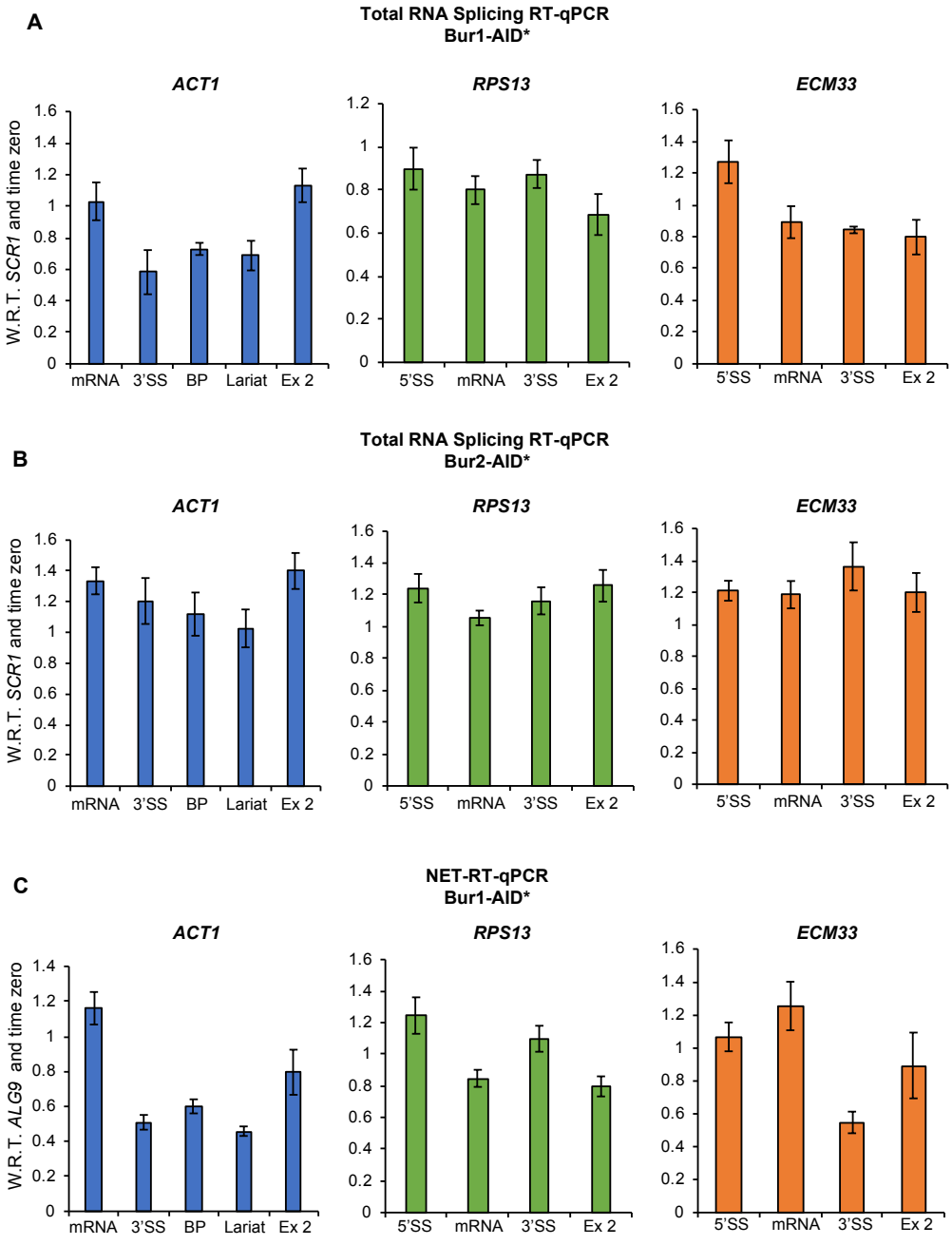


Figure 3.12. Depletion of Bur1 increases co-transcriptional pre-mRNA spicing efficiency

RT-qPCR analysis of total RNA from the intron-containing genes *ACT1*, *RPS13* and *ECM33* *after* depletion of Bur1-AID* (A) or Bur2-AID* (B). Normalised to the *SCR1* PolIII transcript and time zero (no auxin).
(C) NET-RT-qPCR analysis of the intron-containing genes *ACT1*, *RPS13* and *ECM33* *after* depletion of Bur1-AID*. Normalised to the intronless *ALG9* PolII transcript and time zero (before auxin addition).
Primers used detected pre-mRNA (5'SS or BP and 3'SS), lariat (excised intron or intron-exon 2), exon 2 (ex 2) and mRNA (see figure 3.6A for cartoon).
All data (A-C): mean of at least 3 biological replicates, error bars = standard error of the mean.

3.12 Loss of Spt5 phosphorylation resembles depletion of Bur1 and Bur2

Bur1 phosphorylates the first serine of the hexapeptide repeats of the CTR of Spt5 (Liu *et al*., 2009). To test whether the effect of Bur1-AID* and Bur2-AID* depletion is due to the loss of the phosphorylation to the CTR of Spt5 (and not another Bur1 substrate), Spt5 phosphomimetic and phosphomutants were used (these strains were a kind gift from Prof. Alan Hinnebusch (Liu *et al*., 2009; Qiu *et al*., 2012)). In the phosphomimetic Spt5, the first serine of each hexapeptide was mutated to glutamic acid (E). In the phosphomutant of Spt5, the first serine of each hexapeptide (S[T/A]WGG[Q/A]) was mutated to alanine (A) (Liu *et al*., 2009; Qiu *et al*., 2012).

ChIP using antibodies that detect core members of the spliceosome: U1 snRNP (Prp40), U2 snRNP (Lea1-V5) and U5 snRNP (Prp8) was performed in strains containing either wild-type Spt5 (WT), phosphomutant Spt5 (S->A) or phosphomimetic Spt5 (S->E). Additionally, ChIP was performed using an antibody against RNAPII (Rpb1) to test for any differences in RNAPII occupancy between the wild-type and mutant strains.

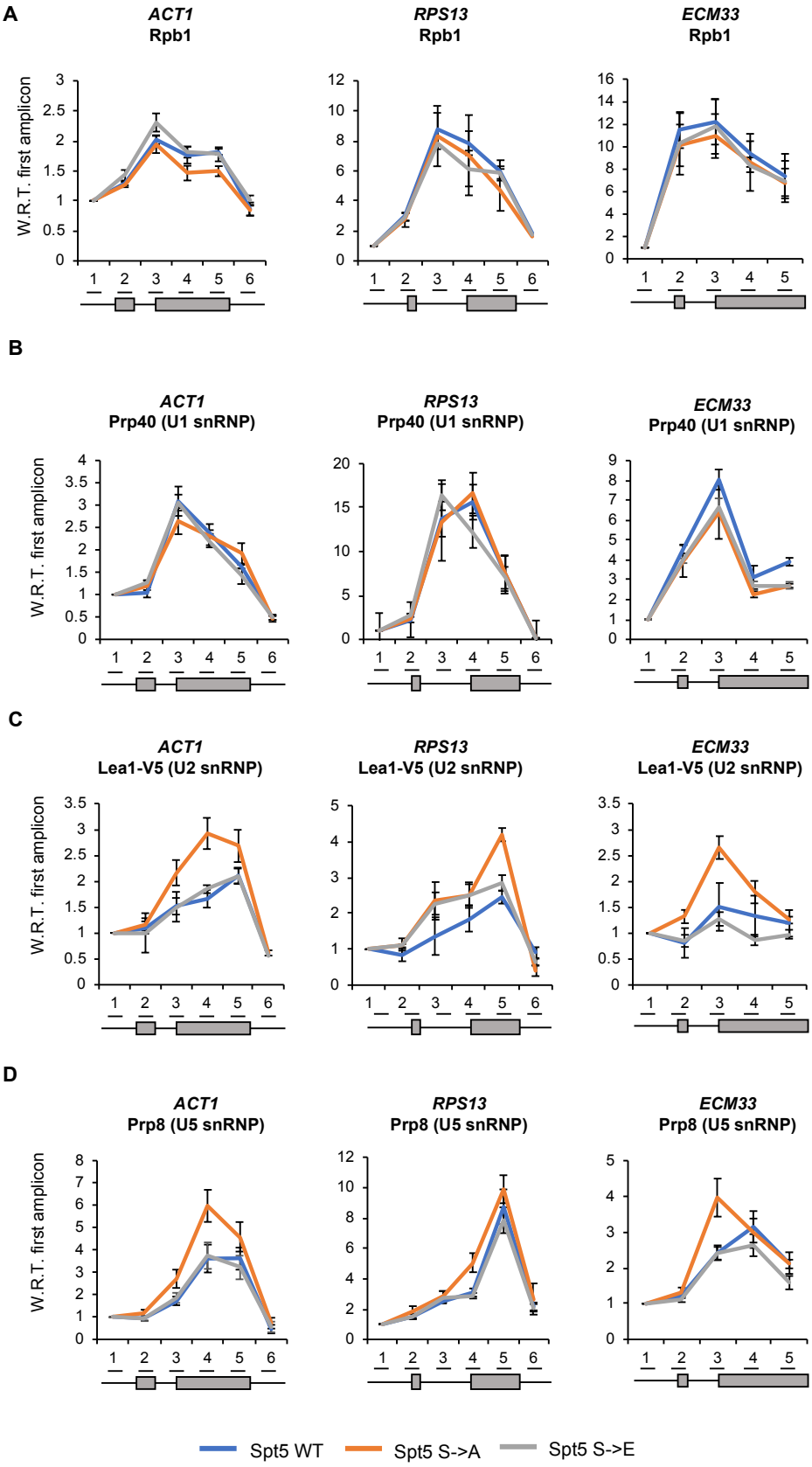
ChIP-qPCR analysis across intron-containing genes *ACT1*, *RPS13* and *ECM33* showed no significant change in RNAPII or U1 snRNP occupancy between WT, S->A or S->E Spt5 (student's t-test, $P>0.05$) (Figures 3.13A and 3.13B). The U2 snRNP significantly accumulated on *ACT1* to 176% at amplicon 4 (exon 2), *ECM33* to 176% at amplicon set 3 (3'SS) and to 170% at amplicon 5 (exon 2) on *RPS13*, relative to WT (student's t-test, $P>0.05$) (Figure 3.13C).

The U5 snRNP showed a significant increase in occupancy on *ACT1* in the Spt5 phosphomutant (S->A) on average relative to WT to 157%, 165% and 127% at amplicons 3, 4 and 5 (3'SS-exon 2) (student's t-test, $P<0.05$), respectively, whilst the Spt5 phosphomimetic (S->E) did not significantly change relative to WT (student's t-test, $P>0.05$) (Figure 3.13D). The U5 snRNP also increased significantly in occupancy in the Spt5 phosphomutant (S->A) on *RPS13* at amplicon 4 (3'SS) on average to 164% relative to WT (student's t-test, $P<0.05$), with no significant changes observed with

the phosphomimetic (S->E) relative to WT (student's t-test, P>0.05) (Figure 3.13D). For *ECM33,* the U5 snRNP also increased significantly in occupancy in the Spt5 phosphomutant (S->A) at amplicon 3 (3'SS) on average by 163% relative to WT (student's t-test, P<0.05), with no significant changes observed with the phosphomimetic (S->E) relative to WT (student's t-test, P>0.05) (Figure 3.13D).

Next, the effect of Spt5 phosphomimetic/phosphomutant mutations on splicing was tested by RT-qPCR on total (steady-state) RNA (as described in section 3.5) (Figure 3.13E). RT-qPCR showed no significant changes between phosphomimetic or phosphomutant Spt5 relative to WT Spt5. Both Spt5 mutants showed significant increases in mRNA levels for *ACT1* and *RPS13* (student's t-test, P<0.05) (not *ECM33*). There were no consistent changes in pre-mRNA species. No significant changes were observed in pre-mRNA species for each mutant relative to WT Spt5 (student's t-test, P>0.05), there was a significant reduction in the 5'SS species of *RPS13* for both mutants relative to WT Spt5 (student's t-test, P<0.05), and there was a significant increase in the 5'SS species of *ECM33* for both mutants relative to WT Spt5 (student's t-test, P<0.05)

The accumulation of the U5 snRNP in the Spt5 S->A phosphomutant relative to wild-type conditions resembles depletion of Bur1 and Bur2, which fits nicely, as Bur1 and Bur2 depletion abolish Spt5 phosphorylation (Figures 3.9 and 3.11). The effects of the Spt5 phosphomutant on co-transcriptional spliceosome assembly are milder than depletion of Bur1, which is analogous to the effects of depletion of Bur2 on co-transcriptional spliceosome assembly being milder to depletion of Bur1. No consistent changes in pre-mRNA species were observed by RT-qPCR for the Spt5 phosphomimetic/phosphomutant strains relative to the WT Spt5 strain, and relative to each other no significant differences – again, analogous to depletion of Bur2. The ChIP data further agrees with the idea that the loss of phosphorylation of the CTR of Spt5 is not responsible for the loss of U5 snRNP recruitment observed upon Spt5 depletion (Figure 3.5).


A
ACT1
Rpb1
RPS13
Rpb1
ECM33
Rpb1
W.R.T. first amplicon
B
ACT1
Prp40 (U1 snRNP)
RPS13
Prp40 (U1 snRNP)
ECM33
Prp40 (U1 snRNP)
C
ACT1
Lea1-V5 (U2 snRNP)
RPS13
Lea1-V5 (U2 snRNP)
ECM33
Lea1-V5 (U2 snRNP)
D
ACT1
Prp8 (U5 snRNP)
RPS13
Prp8 (U5 snRNP)
ECM33
Prp8 (U5 snRNP)
Spt5 WT
Spt5 S->A
Spt5 S->E

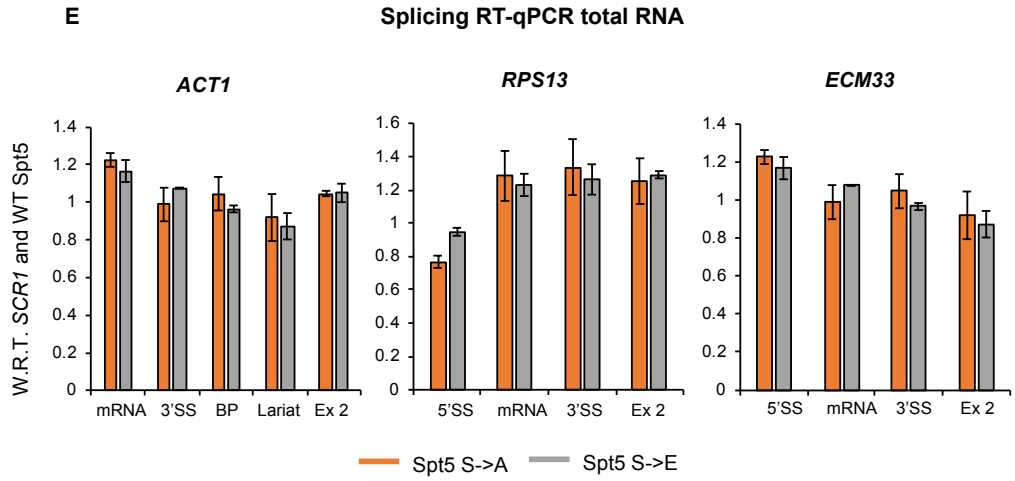


Figure 3.13. Loss of Spt5 phosphorylation resembles depletion of Bur1 and Bur2

(A) Anti-Rpb1 (RNAPII), (B) Anti-Prp40 (U1 snRNP), (C) Anti-Lea1-HA (U2 snRNP) and (D) Anti-Prp8 (U5 snRNP) ChIP followed by qPCR analysis of the intron-containing genes: *ACT1*, *RPS13*, *ECM33* in strains expressing either wild-type (WT, blue) Spt5, Spt5 phosphomutant (S->A, orange) or Spt5 phosphomimetic (S->E, grey). The X-axis shows the positions of amplicons used for ChIP qPCR analysis – the exons are in grey. The data are represented as the mean percentage of input with respect to (W.R.T.) the first amplicon of each gene.
(E) Splicing RT-qPCR analysis of the intron-containing genes *ACT1*, *RPS13* and *ECM33* in strains expressing either Spt5 phosphomutant (S->A, orange) or Spt5 phosphomimetic (S->E, grey). Normalised to the *SCR1* PolIII transcript and WT Spt5. Primers used detected pre-mRNA (5'SS or BP and 3'SS), lariat (excised intron or intron-exon 2), exon 2 (ex 2) and mRNA (see figure 3.6A for cartoon). Mean of at least 3 biological replicates. Error bars = standard error of the mean.
All data (A-E): mean of at least 3 biological replicates, error bars = standard error of the mean.

3.13 Use of the AID system to conditionally deplete Ctk1

In *S. cerevisiae*, Ctk1 is the major yet non-essential serine 2 CTD kinase of RNAPII (Cho *et al*., 2001). In the present work, Ctk1 was C-terminally AID*-tagged to enable conditional depletion using the AID system in order to determine whether serine 2 phosphorylation of the CTD plays an important role in co-transcriptional spliceosome assembly and splicing *in vivo* in *S. cerevisiae* (described in section 3.1). Additionally, as Bur1 is proposed to augment the phosphorylation of Ctk1 and phosphorylate serine 2 of the CTD at the 5' end of genes (Qiu *et al*., 2009), depletion of Ctk1 served as a control for the effects on co-transcriptional spliceosome assembly and splicing observed upon Bur1 depletion.

The effect of the AID* tag on Ctk1 function and auxin-sensitivity were assessed by growth analyses of the parental strain (W303-OsTIR) with no AID* tag and the Ctk1-AID* strain with or without auxin treatment (Figure 3.14A). In the absence of auxin, the Ctk1-AID* strain grew comparably to the parental strain, indicating that the AID* tag does not affect Ctk1 function. After auxin addition, the Ctk1-AID* strain grew more slowly than the parental strain. Though Ctk1 is non-essential, its deletion has multiple phenotypes, including decreased anaerobic growth (Samanfar *et al*., 2013).

Western blotting showed that depletion of Ctk1-AID* for 30 minutes resulted on average in a significant reduction to 12% of Ctk1 left relative to cells with no depletion (student's t-test, $P<0.05$) (Figure 3.14B). It is estimated that Ctk1 is present at approximately 471 molecules/cell (Kulak *et al*., 2014), and therefore there may be approximately 57 molecules/cell of Ctk1 left after 30 minutes of auxin treatment.

ChIP-qPCR analyses showed that, in addition to being depleted in whole cell extracts, Ctk1-AID* was significantly depleted on average to 60% across *ACT1*, 68% across *RPS13* and 59% across *ECM33* after 30 minutes of auxin treatment relative to conditions prior to auxin addition (student's t-test, $P<0.05$) (Figure 3.14C).

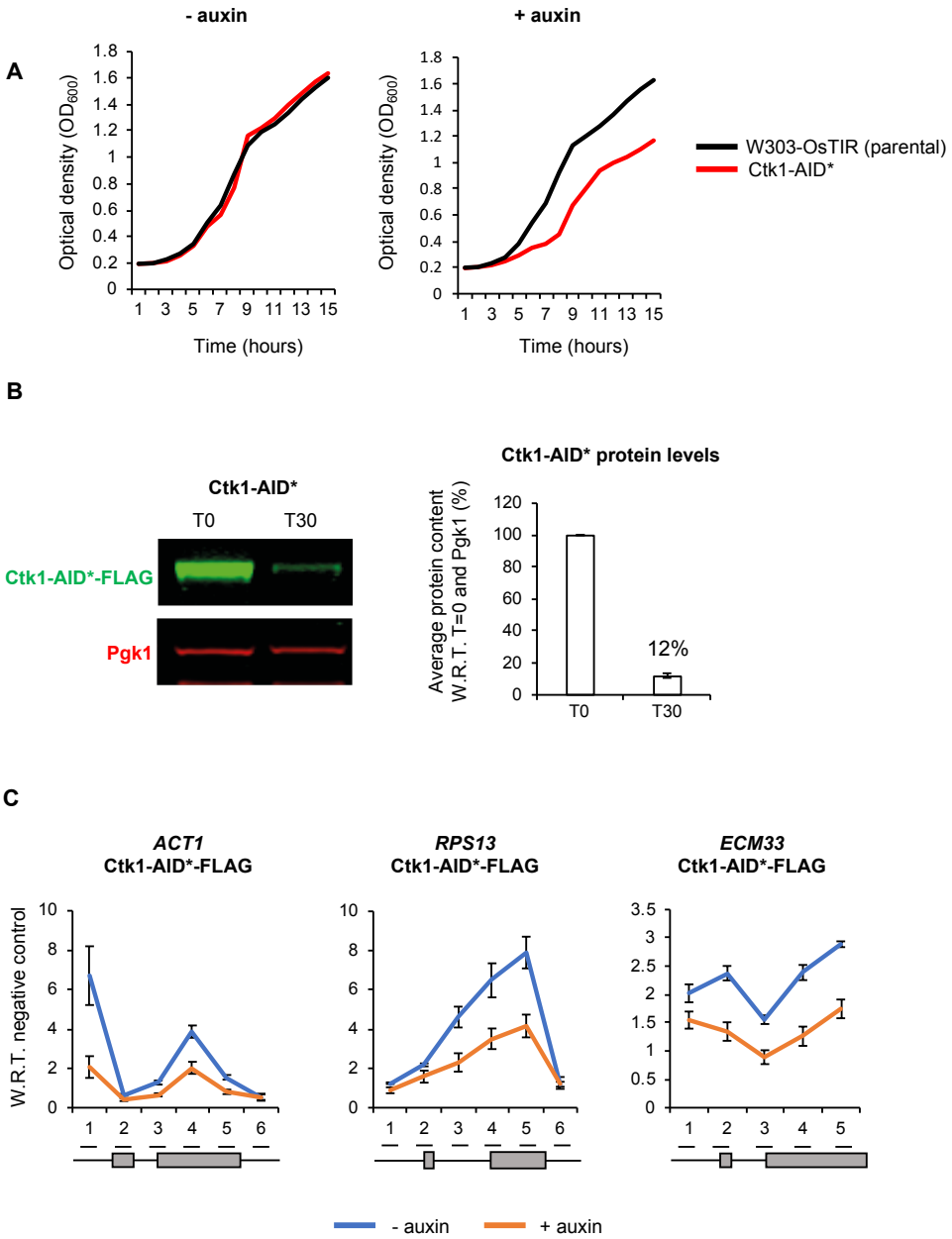


Figure 3.14. Use of the AID system to conditionally deplete Ctk1

(A) Growth of the control parental strain (W303-OsTIR; no AID*-tagged proteins; black) and Ctk1-AID* tagged strain (red) was measured over a period of 15 hours without auxin (-) or with auxin (+). Mean of three biological replicates.
(B) Western blot probed with anti-FLAG and anti-Pgk1 as a loading control. Samples were taken before (T0) and 30 minutes (T30) after addition of auxin. Ctk1-AID* depletion was quantified and shown as the percentage mean of 3 biological replicates after 30 minutes of auxin addition with respect to (W.R.T) time zero and normalised to the Pgk1 signal. Error bars = standard error of the mean.
(C) Anti-FLAG ChIP followed by qPCR analysis of the intron-containing genes: *ACT1*, *RPS13*, *ECM33* 0 minutes (no auxin; blue) and after 30 minutes (+ auxin; orange) of auxin addition to deplete Ctk1-AID*. The X-axis shows the positions of amplicons used for ChIP qPCR analysis – the exons are in grey. The data are presented as the mean percentage of input normalised to a negative control region. Mean of at least three biological replicates. Error bars = standard error of the mean.

3.14 Depletion of Ctk1 reduces serine 2 phosphorylation of the CTD of RNAPII

To determine whether depletion of Ctk1 for 30 minutes was a sufficient amount of time to reduce serine 2 phosphorylation of the CTD of RNAPII, western blotting was performed using antibodies against the serine 2 (S2P) or serine 5 (S5P) phosphorylated forms of the CTD of RNAPII. The quantitation of the S2P or S5P phosphorylated forms of the CTD were normalised to Rpb3, a subunit of RNAPII, to account for potential differences in the amount of RNAPII upon depletion of Ctk1. Additionally, western blotting was performed using antibodies specific to the phosphorylated and unphosphorylated forms of Spt5 (see section 3.8 for details).

Western blotting showed that depletion of Ctk1 for 30 minutes resulted in no significant reduction in Spt5 phosphorylation relative to conditions prior to auxin addition and to total Spt5 (student's t-test, $P>0.05$) (Figure 3.15A). Depletion of Ctk1 for 30 minutes resulted in a significant reduction in the level of serine 2 phosphorylation on average to 17% relative to conditions prior to auxin addition and to Rpb3 (student's t-test, $P<0.05$) (Figure 3.15B). Serine 5 phosphorylation increased significantly on average to 146% after depletion of Ctk1, relative to conditions prior to auxin addition and to Rpb3 (student's t-test, $P<0.05$) (Figure 3.15C).

To determine whether Ctk1 depletion resulted in changes in RNAPII occupancy, serine 2 (S2P), and serine 5 (S5P) phosphorylation of the CTD of RNAPII on a gene-by-gene basis, ChIP-qPCR was performed across intron-containing genes *ACT1*, *RPS13* and *ECM33* using antibodies against RNAPII (Rpb1), S2P and S5P CTD phosphorylation. ChIP-qPCR analyses showed that depletion of Ctk1 for 30 minutes did not significantly affect RNAPII occupancy across *ACT1*, *RPS13* or *ECM33* relative to conditions prior to auxin addition (student's t-test, $P>0.05$) (Figure 3.15D). Depletion of Ctk1 in these conditions significantly reduced serine 2 phosphorylation of the CTD relative to RNAPII occupancy. On *ACT1*, serine 2 phosphorylation was significantly reduced on average to 65%, 28%, 21% and 16% at amplicons 3 to 6 (3'SS-exon2) respectively, relative to conditions prior to auxin addition (student's t-test, $P<0.05$) (Figure 3.15E). On *RPS13*, serine 2 phosphorylation was significantly

reduced on average to 50%, 40%, 26% and 42% at amplicons 3 to 6 (3'SS-exon 2) respectively, relative to conditions prior to auxin addition (student's t-test, $P<0.05$) (Figure 3.15E). On *ECM33*, serine 2 phosphorylation was significantly reduced on average to 69%, 57%, 41%, 25% and 34% at amplicons 1 to 5 respectively, relative to conditions prior to auxin addition (student's t-test, $P<0.05$) (Figure 3.15E).

Serine 5 phosphorylation did not significantly change across *RPS13* relative to RNAPII occupancy (student's t-test, $P>0.05$) (Figure 3.15F). However, serine 5 phosphorylation was significantly increased on average to 122% at amplicon 5 (exon 2) for *ACT1* (student's t-test, $P<0.05$), and on average to 123% at amplicons 3-5 (3'SS-exon 2) of *ECM33* (student's t-test, $P<0.05$), relative to conditions prior to auxin addition and to RNAPII occupancy (Figure 3.15F).

Ctk1 depletion for 30 minutes was therefore an effective way to reduce serine 2 phosphorylation of the CTD.

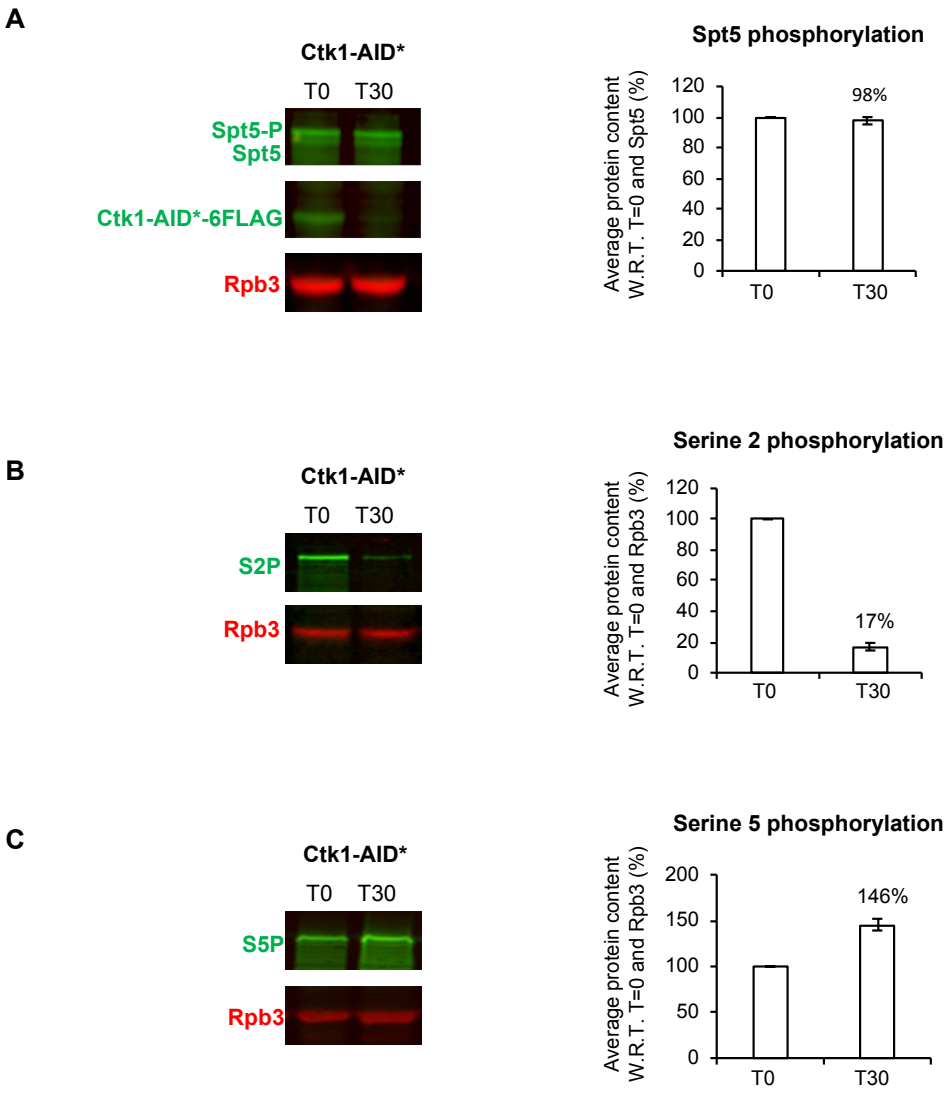


Figure 3.15. Depletion of Ctk1 reduces serine 2 phosphorylation of the CTD of RNAPII

(A) Western blot probed with anti-Spt5 phosphorylation (Spt5-P), anti-Spt5, anti-FLAG and anti-Rpb3 (RNAPII) as a loading control. Samples were taken before (T0) and 30 minutes (T30) after addition of auxin to deplete Ctk1-AID*. Spt5 phosphorylation levels were quantified after depletion of Ctk1-AID* and shown as the percentage mean of 3 biological replicates after 30 minutes of auxin treatment with respect to (W.R.T) time zero and normalised to Spt5 signal.
(B) Western blot probed with anti-serine 2 phosphorylation (S2P) and anti-Rpb3 (RNAPII) as a loading control. Samples were taken before (T0) and 30 minutes (T30) after addition of auxin to deplete Ctk1-AID*. Serine 2 phosphorylation levels were quantified after depletion of Ctk1-AID* and shown as the percentage mean of 3 biological replicates after 30 minutes of auxin addition with respect to (W.R.T) time zero and normalised to Rpb3 signal.
(C) Western blot probed with anti-serine 5 phosphorylation (S5P) and anti-Rpb3 (RNAPII) as a loading control. Samples were taken before (T0) and 30 minutes (T30) after addition of auxin to deplete Ctk1-AID*. Serine 5 phosphorylation levels were quantified after depletion of Ctk1-AID* and shown as the percentage mean of 3 biological replicates after 30 minutes of auxin addition with respect to (W.R.T) time zero and normalised to Rpb3 signal.

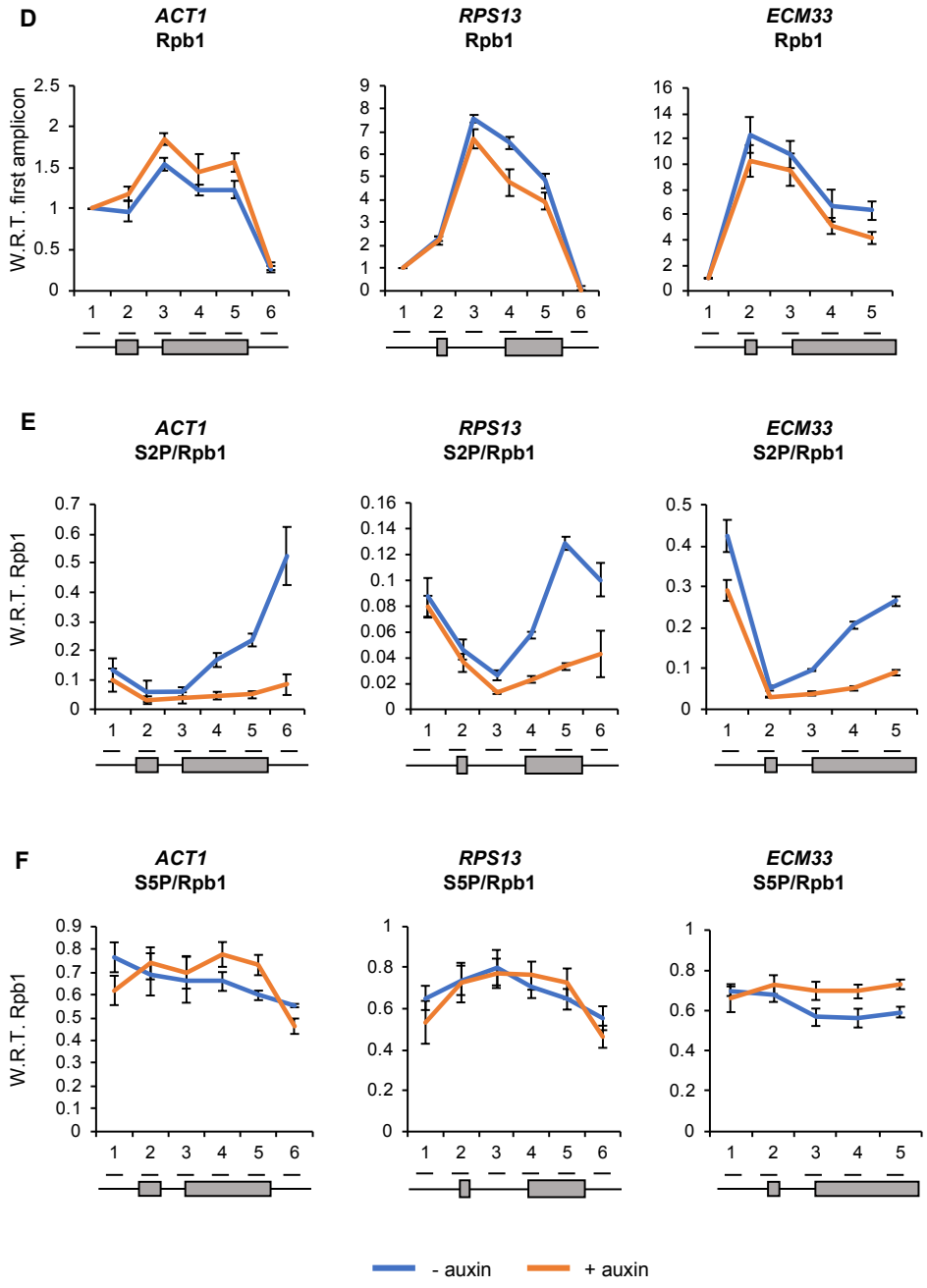


Figure 3.15 (continued). Depletion of Ctk1 reduces serine 2 phosphorylation of the CTD of RNAPII

(D) Anti-Rpb1 (RNAPII), (E) anti-serine 2 phosphorylation (S2P) and (F) anti-serine 5 phosphorylation (S5P) ChIP and qPCR across intron-containing genes *ACT1*, *RPS13* and *ECM33* without auxin (T0; blue) and with 30 minutes of auxin addition to deplete Ctk1-AID* (T30; orange). The X-axis shows the positions of amplicons used for ChIP qPCR analysis – the exons are in grey. The Rpb1 ChIP data are presented as the mean percentage of input with respect to (W.R.T.) the first amplicon of each gene. The S2P and S5P ChIP data are relative to Rpb1 occupancy.
All data (A-F): mean of at least three biological replicates, error bars = standard error of the mean.

3.15 Depletion of Ctk1 does not affect co-transcriptional spliceosome assembly in *S. cerevisiae*

To test whether loss of serine 2 phosphorylation affected co-transcriptional spliceosome assembly, ChIP-qPCR was performed across intron-containing genes (*ACT1*, *RPS13* and *ECM33*) with and without depletion of Ctk1 for 30 minutes. Antibodies were used that detect core members of the spliceosome: the U1 snRNP (Prp40), U2 snRNP (Lea1-13MYC) and U5 snRNP (Prp8). ChIP-qPCR analyses showed no significant change in the occupancy of U1, U2 or U5 snRNPs upon Ctk1 depletion for 30 minutes, relative to conditions prior to auxin addition (student's t-test, $P>0.05$) (Figures 3.16A-3.16C).

Therefore, in these conditions, *in vivo* loss of Ctk1 and serine 2 phosphorylation does not affect co-transcriptional spliceosome assembly, which is in contrast to findings in mammalian cells that a mutation of serine 2 of the CTD to alanine reduced recruitment of the U2 snRNP to transcription sites (assayed by FISH), and caused defects in pre-mRNA splicing (Gu *et al*., 2013).

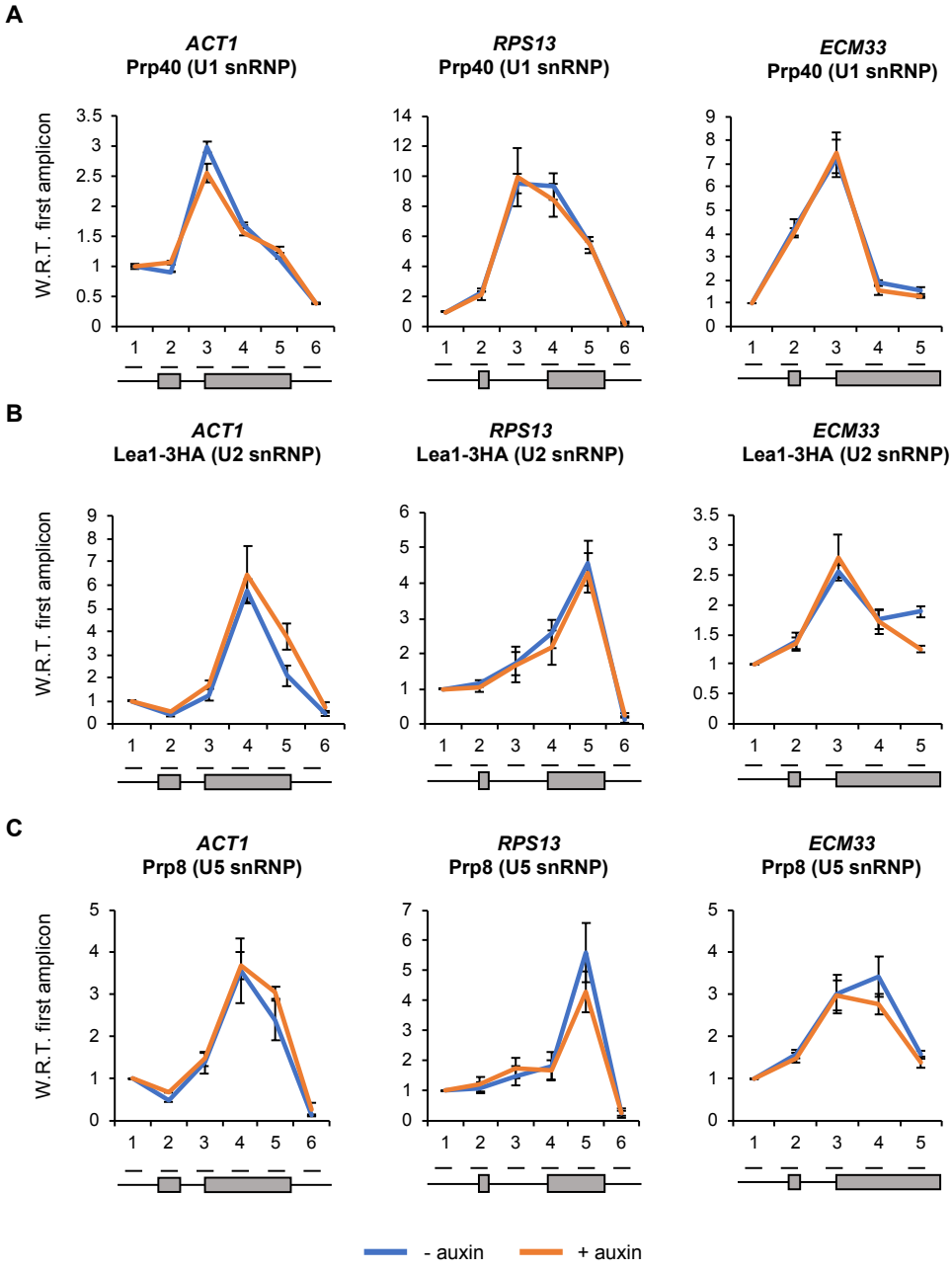


Figure 3.16. Depletion of Ctk1 does not affect co-transcriptional spliceosome assembly in *S. cerevisiae*

(A) Anti-Prp40 (U1 snRNP), (B) anti-Lea1-MYC (U2 snRNP) and (C) anti-Prp8 (U5 snRNP) ChIP followed by qPCR analysis of the intron-containing genes: *ACT1*, *RPS13*, *ECM33* 0 minutes (no auxin; blue) and after 30 minutes (orange) of auxin addition to deplete Ctk1-AID*. The X-axis shows the positions of amplicons used for ChIP qPCR analysis – the exons are in grey. The data are presented as the mean percentage of input with respect to (W.R.T.) the first amplicon of each gene. Mean of at least three biological replicates. Error bars = standard error of the mean.

3.16 Depletion of Ctk1 does not affect pre-mRNA splicing in *S. cerevisiae*

To test whether depletion of Ctk1 affected pre-mRNA splicing, RT-qPCR was performed on total (steady-state) RNA, as described in section 3.5. No significant accumulation of pre-mRNA species upon depletion of Ctk1 was observed for *ACT1*, *RPS13* and *ECM33* (student's t-test, P>0.05) (Figure 3.17). Additionally, there was a significant reduction in pre-mRNA of *RPS13* (5'SS and 3'SS) and *ECM33* (5'SS), in addition to significant reductions in mRNA and exon 2 levels for both genes (student's t-test, P<0.05) (Figure 3.17). This was not observed for *ACT1*. This indicates that on *RPS13* and *ECM33* there are transcription defects.

Overall, depletion of Ctk1 does not result in pre-mRNA splicing defects, which fits with the ChIP-qPCR data which showed no effect on co-transcriptional spliceosome assembly upon Ctk1 depletion (Figure 3.16).

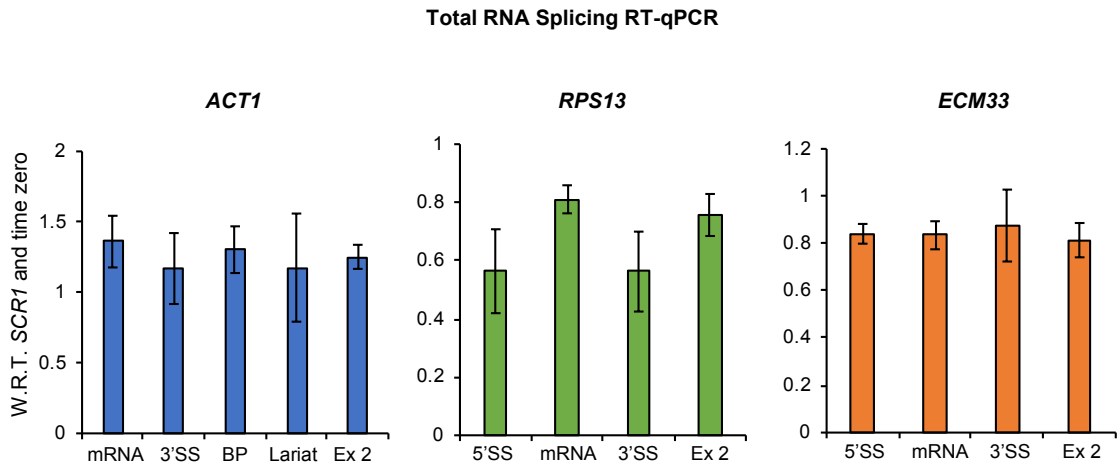


Figure 3.17. Depletion of Ctk1 does not affect pre-mRNA splicing in *S. cerevisiae*

RT-qPCR analysis of total RNA from the intron-containing genes *ACT1*, *RPS13* and *ECM33* after 30 minutes of depletion of Ctk1-AID* (A). Normalised to the *SCR1* PolIII transcript and time zero (no auxin). Primers used detected pre-mRNA (5'SS or BP and 3'SS), lariat (excised intron or intron-exon 2), exon 2 (ex 2) and mRNA (see figure 3.6A for cartoon). Mean of at least 3 biological replicates. Error bars = standard error of the mean.

3.17 Use of the AID system to conditionally deplete Paf1

Spt5 functions to promote transcription elongation through recruitment of downstream factors that modulate transcription and chromatin state. It has been demonstrated using Spt5 phosphomutants/phosphomimetics that the phosphorylation of the CTR of Spt5 by the Bur1/2 complex is important for Paf1 complex recruitment to elongating RNAPII (Laribee *et al*., 2005; Liu *et al*., 2009). To test whether the effects on co-transcriptional spliceosome assembly observed upon Spt5 and Bur1/2 depletion is due to loss of the physical presence of Paf1, non-essential Paf1 was C-terminally AID*-tagged and depleted using the AID system (described in section 3.1). Experiments were performed with and without 30 minutes of auxin treatment.

Both the effect of the tag on Paf1 function and auxin-sensitivity were assessed by growth analyses of the parental strain (W303-OsTIR) with no AID* tag and the Paf1-AID* strain with or without auxin (Figure 3.18A). In the absence of auxin, the Paf1-AID* strain grew comparably to the parental strain, and with auxin the Paf1-AID* grew more slowly than the parental strain. Though Paf1 is non-essential, its deletion has multiple phenotypes, including decreased anaerobic growth (Samanfar *et al*., 2013).

Western blotting showed that depletion of Paf1 for 30 minutes resulted in a significant reduction of Ctk1 on average to 8% relative to cells with no depletion (student's t-test, $P<0.05$) (Figure 3.18B). Paf1 is present at approximately 2542 molecules/cell (Kulak *et al*., 2014), and therefore after 30 minutes of depletion there may be approximately 203 molecules/cell of Paf1 remaining.

ChIP-qPCR analyses showed that, in addition to being depleted in whole cell extracts, Paf1 was significantly depleted relative to conditions prior to auxin addition on average to 28% across *ACT1*, 60% across *RPS13* and 55% across *ECM33* after 30 minutes of auxin treatment (student's t-test, $P<0.05$) (Figure 3.18C).

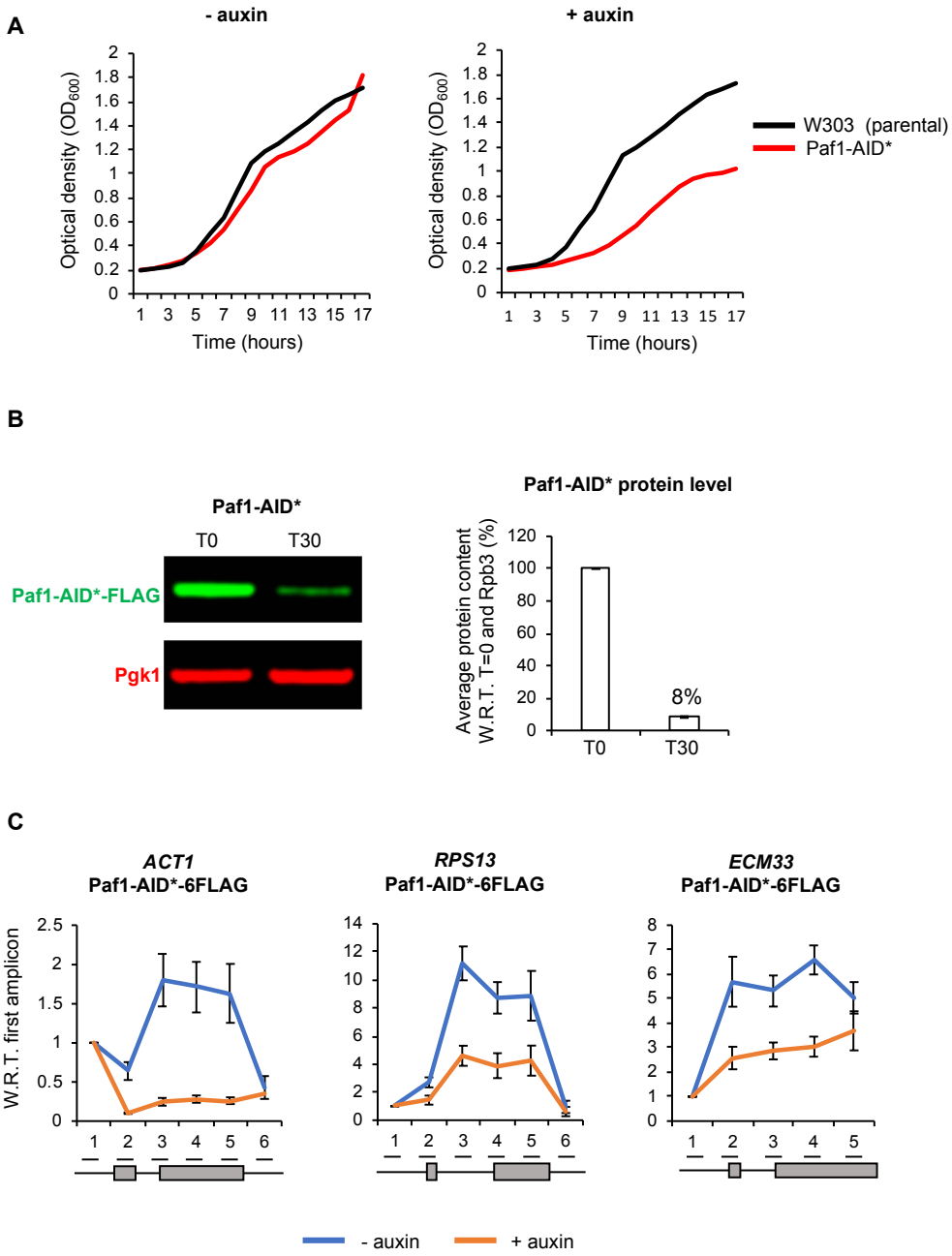


Figure 3.18. Use of the AID system to conditionally deplete Paf1

(A) Growth of the control parental strain (W303-OsTIR; no AID*-tagged protein; black) and Paf1-AID* tagged strain (red) was measured over a period of 17 hours without auxin (-) or with auxin (+). Mean of three biological replicates.
(B) Western blot probed with anti-FLAG and anti-Pgk1 as a loading control. Samples were taken before (T0) and 30 minutes (T30) after addition of auxin. Paf1-AID* depletion was quantified and shown as the percentage mean of 3 biological replicates after 30 minutes of auxin addition with respect to (W.R.T) time zero and normalised to the Pgk1 signal. Error bars = standard error of the mean.
(C) Anti-FLAG ChIP followed by qPCR analysis of the intron-containing genes: *ACT1*, *RPS13*, *ECM33* 0 minutes (no auxin; blue) and 30 minutes (+ auxin; orange) after auxin addition to deplete Paf1-AID. The X-axis shows the positions of amplicons used for ChIP qPCR analysis – the exons are in grey. The data are presented as the mean percentage of input with respect to (W.R.T.) the first amplicon of each gene. Mean of at least three biological replicates. Error bars = standard error of the mean.

3.18 Depletion of Paf1 causes accumulation of RNAPII in gene bodies

To determine whether Paf1 depletion changed RNAPII occupancy, serine 2, or serine 5 phosphorylation of the CTD of RNAPII on a gene-by-gene basis, ChIP-qPCR was performed across intron-containing genes *ACT1*, *RPS13* and *ECM33* using antibodies against RNAPII (Rpb1), serine 2 (S2P) or serine 5 (S5P) CTD phosphorylation.

ChIP-qPCR analyses showed that depletion of Paf1 for 30 minutes significantly increased RNAPII occupancy across *ACT1*, *RPS13* or *ECM33* (Figure 3.19A). For *ACT1*, RNAPII occupancy was significantly increased on average to 139% and 156% at amplicons 4 and 5 (exon 2), relative to conditions prior to auxin addition (student's t-test, $P<0.05$). For *RPS13* RNAPII occupancy was significantly increased on average to 230% and 200% at amplicons 3 and 4 (intron, 3'SS), relative to conditions prior to auxin addition (student's t-test, $P<0.05$). For *ECM33* RNAPII occupancy was significantly increased on average to 191% at amplicon 3 (3'SS), relative to conditions prior to auxin addition (student's t-test, $P<0.05$).

Depletion of Paf1 in these conditions significantly increased serine 2 phosphorylation of the CTD relative to RNAPII occupancy on *ACT1*, *RPS13* and *ECM33* (Figure 3.19B). On *ACT1*, serine 2 phosphorylation was significantly increased on average to 156% and 395% at amplicons 5 and 6 (exon 2 and 3' end), relative to conditions prior to auxin addition (student's t-test, $P<0.05$). On *RPS13*, serine 2 phosphorylation was significantly increased on average to 147% at amplicon set 6 (3'end), relative to conditions prior to auxin addition (student's t-test, $P<0.05$). On *ECM33*, serine 2 phosphorylation was significantly increased on average to 176%, 143%, 151% and 155% at amplicons 2-5 (exon 1-exon 2) respectively, relative to conditions prior to auxin addition (student's t-test, $P<0.05$).

Serine 5 phosphorylation of the CTD was also significantly increased upon depletion of Paf1 relative to RNAPII occupancy on *ACT1*, *RPS13* and *ECM33* (Figure 3.19B). On *ACT1*, serine 5 phosphorylation was significantly increased on average to 255%, 232%, 355%, 210% and 306% at amplicons 1, 2, 3, 4 and 6 respectively, relative to

conditions prior to auxin addition (student's t-test, P<0.05). On *RPS13*, serine 5 phosphorylation was significantly increased on average to 240% and 352% at amplicons 1 and 2 (5'end and exon1), relative to conditions prior to auxin addition (student's t-test, P<0.05). On *ECM33*, serine 5 phosphorylation was significantly increased on average to 184%, 150% and 147% at amplicons sets 1-3 (5'end-exon 2), relative to conditions prior to auxin addition (student's t-test, P<0.05).

Overall, Paf1 depletion resulted in significant accumulation of RNAPII on *ACT1*, *ECM33* and *RPS13* from the intron or 3'SS to exon 2 within the gene body. Serine 5 and serine 2 phosphorylation of the CTD were also increased at particular locations along the gene. The RNAPII that accumulated in the gene bodies was only hyperphosphorylated at serine 5 on *ACT1* and *ECM33*, and at serine 2 on *ECM33*. This accumulation of RNAPII in gene bodies is similar to the increase in RNAPII occupancy observed by Chen *et al* (2015) upon Paf1 knockdown in mammalian cells. The increase in serine 2 phosphorylation observed by ChIP-qPCR was surprising as previous ChIP-qPCR studies showed that deletion of Paf1 reduced serine 2 phosphorylation in *S. cerevisiae* (Mueller *et al*., 2004; Nordick *et al*., 2008). However, Chen et al (2015) also observe an increase in serine 2 phosphorylation after Paf1 knockdown in mammalian cells (Chen *et al*., 2015).

To test whether loss of the physical presence of Paf1 affected co-transcriptional spliceosome assembly, ChIP-qPCR was performed across intron-containing genes (*ACT1*, *RPS13* and *ECM33*) with and without depletion of Ctk1 for 30 minutes. Antibodies were used that detect core members of the spliceosome: the U1 snRNP (Prp40), U2 snRNP (Lea1-3HA) and U5 snRNP (Prp8). ChIP-qPCR analyses showed no significant change in the occupancy of U1, U2 or U5 upon Paf1 depletion for 30 minutes on *ACT1* or *ECM33*, relative to conditions prior to auxin addition (student's t-test, P>0.05) (Figures 3.19D-3.19F). For *RPS13*, depletion of Paf1 significantly increased U1 snRNP occupancy to 300%, 257% and 207% at amplicons 3-5 (intron-exon2) (student's t-test, P<0.05), without significantly affecting U2 or U5 snRNP occupancy (student's t-test, P>0.05), relative to conditions prior to auxin addition (Figures 3.19D-3.19F). In agreement, ChIP-qPCR analyses of co-transcriptional

spliceosome assembly in a Paf1 deletion strain show no significant effect on co-transcriptional spliceosome assembly across the same intron-containing genes (data not shown).

In summary, gene-specific effects were observed for *RPS13*, where the U1 snRNP significantly accumulated. However, as no consistent effects on co-transcriptional spliceosome assembly were observed upon Paf1 depletion, the effects on co-transcriptional spliceosome assembly observed upon Spt5 depletion, Bur1 or Bur2 depletion and the phosphomutant of Spt5 are not due to loss of Paf1 recruitment.

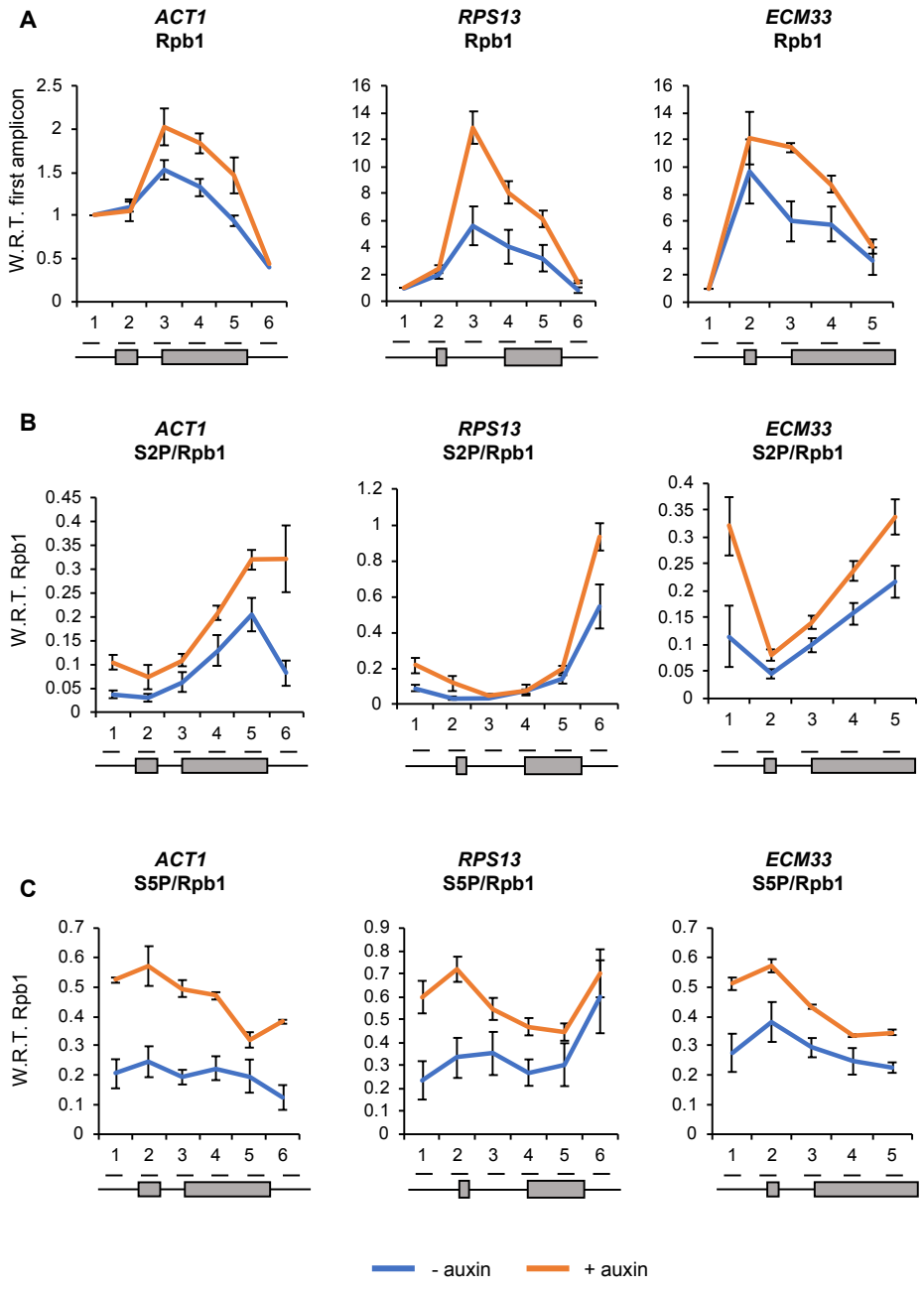


Figure 3.19. Depletion of Paf1 causes accumulation of RNAPII in gene bodies

(A) Anti-Rpb1 (RNAPII), (B) anti-serine 2 phosphorylation (S2P), (C) anti-serine 5 phosphorylation (S5P), (D) anti-Prp40 (U1 snRNP), (E) anti-Lea1-HA and (F) anti-Prp8 (U5 snRNP) ChIP and qPCR across intron-containing genes *ACT1*, *RPS13* and *ECM33* without auxin (blue) and with 30 minutes of auxin addition to deplete Paf1-AID* (orange). The X-axis shows the positions of amplicons used for ChIP qPCR analysis – the exons are in grey. The Rpb1, Prp40, Lea1-HA and Prp8 ChIP data are presented as the mean percentage of input with respect to (W.R.T.) the first amplicon of each gene. The S2P and S5P ChIP data are relative to Rpb1 occupancy. All data (A-F): mean of at least 3 biological replicates, error bars = standard error of the mean.

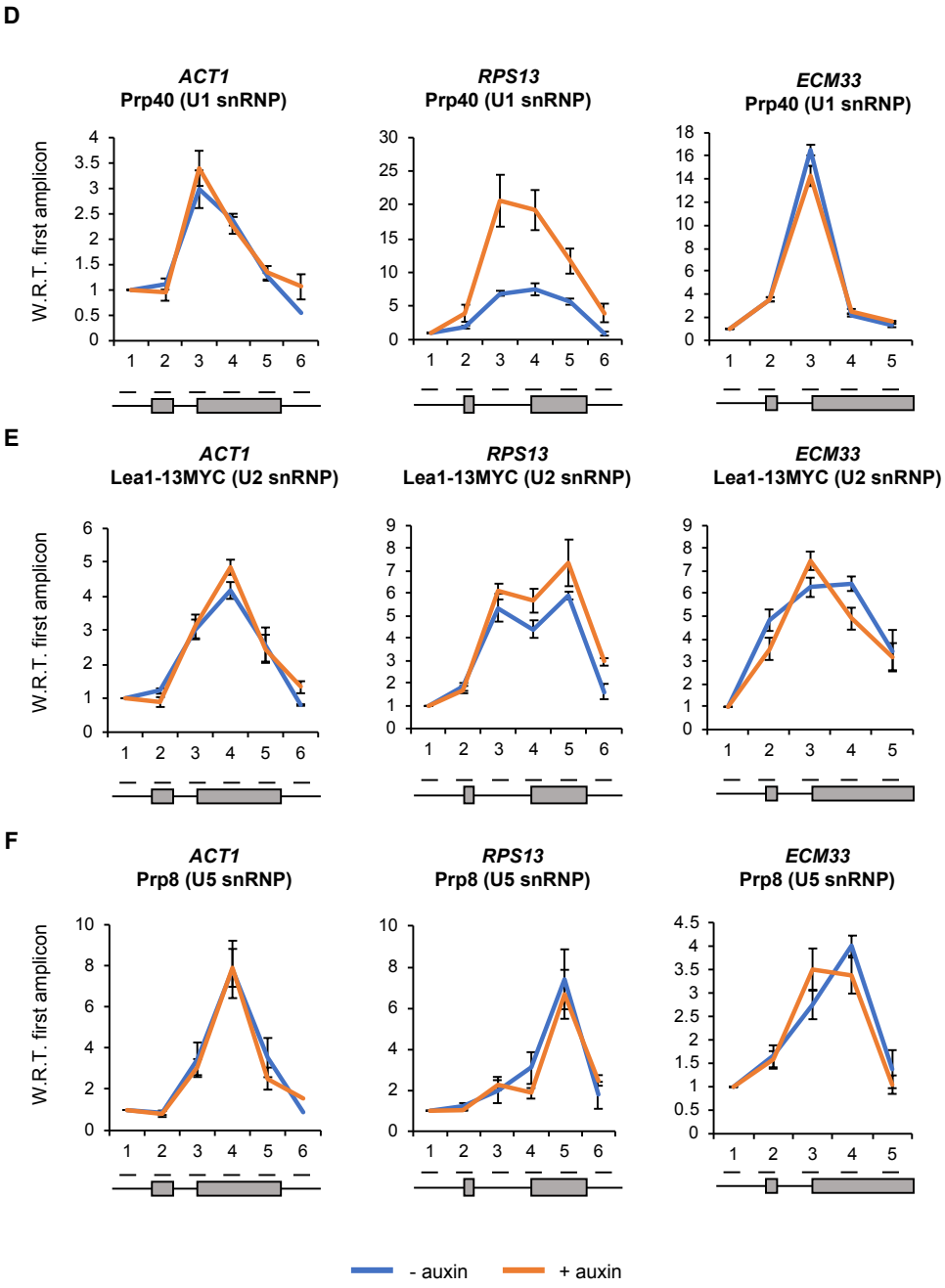


Figure 3.19 (continued). Depletion of Paf1 causes accumulation of RNAPII in gene bodies

(A) Anti-Rpb1 (RNAPII), (B) anti-serine 2 phosphorylation (S2P), (C) anti-serine 5 phosphorylation (S5P), (D) anti-Prp40 (U1 snRNP), (E) anti-Lea1-HA and (F) anti-Prp8 (U5 snRNP) ChIP and qPCR across intron-containing genes *ACT1*, *RPS13* and *ECM33* without auxin (blue) and with 30 minutes of auxin addition to deplete Paf1-AID* (orange). The X-axis shows the positions of amplicons used for ChIP qPCR analysis – the exons are in grey. The Rpb1, Prp40, Lea1-HA and Prp8 ChIP data are presented as the mean percentage of input with respect to (W.R.T.) the first amplicon of each gene. The S2P and S5P ChIP data are relative to Rpb1 occupancy. All data (A-F): mean of at least 3 biological replicates, error bars = standard error of the mean.

3.19 Depletion of Paf1 does not affect pre-mRNA splicing

RT-qPCR was performed on total (steady-state) RNA to determine whether Paf1-AID* depletion affected pre-mRNA splicing, as described in section 3.5. No significant changes in the levels of pre-mRNA species, mRNA species or exon 2 levels were observed upon depletion of Paf1-AID* for *ACT1*, *RPS13* and *ECM33* relative to conditions prior to auxin addition (student's t-test, P>0.05) (Figure 3.20).

The splicing RT-qPCR analysis showing no defects in pre-mRNA splicing agrees with the ChIP-qPCR data that showed no consistent changes in co-transcriptional spliceosome assembly upon Paf1-AID* depletion (Figure 3.19).

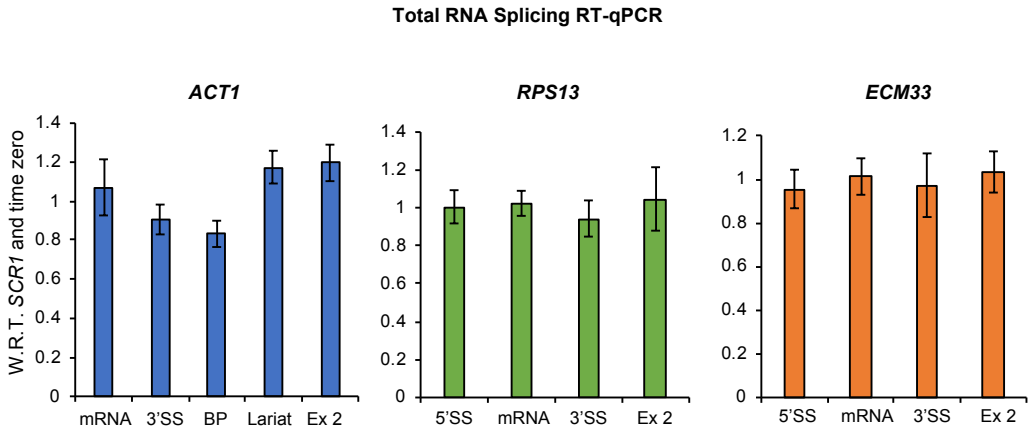


Figure 3.20. Depletion of Paf1 does not affect pre-mRNA splicing

RT-qPCR analysis of total RNA from the intron-containing genes *ACT1*, *RPS13* and *ECM33* after 30 minutes of depletion of Paf1-AID* (A). Normalised to the *SCR1* PolIII transcript and time zero (no auxin). Primers used detected pre-mRNA (5'SS or BP and 3'SS), lariat (excised intron or intron-exon 2), exon 2 (ex 2) and mRNA (see figure 3.6A for cartoon). Mean of at least 3 biological replicates. Error bars = standard error of the mean.

3.20 Discussion

Despite mounting evidence for functional links between transcription and splicing, the majority of this evidence is not mechanistic. A 'recruitment' model for co-transcriptional spliceosome assembly has been suggested, whereby interactions between RNAPII and/or transcription elongation factors and splicing factors during transcription facilitate co-transcriptional spliceosome assembly and therefore affect splicing outcome (reviewed in Buratowski, 2003; Bentley, 2005). Most evidence for the recruitment model is assumptive, and most studies do not directly demonstrate that RNAPII, the CTD of RNAPII (and its phosphorylated states) or transcription elongation factors are required for co-transcriptional spliceosome assembly. Therefore, the recruitment model of co-transcriptional splicing currently lacks mechanistic evidence. By using the AID system to conditionally deplete essential and non-essential transcription elongation factors (Spt5, Bur1, Bur2, Ctk1 and Paf1), the experiments described here aimed to provide mechanistic insights into the links between the transcription elongation machinery and splicing in *S. cerevisiae*. Previous studies have made use of mutants or deletions and in these strains or systems adaptation or compensatory mutations may have occurred, and therefore previous observations of the functional roles of these proteins may have been secondary to their primary role. For this reason, the AID system provided a novel approach not only to assess effects of the core transcription elongation factors on co-transcriptional spliceosome assembly and splicing, but also effects on RNAPII occupancy and phosphorylation status of the CTD of RNAPII.

Here, evidence is provided for a two-part function for Spt5 in co-transcriptional spliceosome assembly. ChIP, as a well-established assay for co-transcriptional spliceosome assembly, showed that whilst the U1 and U2 snRNPs were recruited normally upon Spt5 depletion, the recruitment of the U5 snRNP was significantly reduced (Figure 3.5). As the U1 and U2 snRNPs exhibit normal co-transcriptional recruitment, at least the pre-spliceosome or A complex can form in the absence of Spt5. This supports step-wise spliceosome assembly. In the absence the U5 snRNP, one might expect the pre-spliceosome to be unstable. Here, the A complex probably

forms and dissociates in an equilibrium such that no changes were observed by ChIP upon Spt5 depletion compared to undepleted. There is precedent for this given that individual stages of step-wise spliceosome assembly were shown by single molecule imaging to be reversible in *S. cerevisiae* (Hoskins *et al*., 2011). It was also shown *in vivo* that in the absence of the U5 snRNP pre-spliceosomes can form (Tardiff and Rosbash, 2006). However, Price *et al* (2014) found that the U5 snRNP promotes spliceosome maturation events that occur before tri-snRNP recruitment.

As the U5 snRNP is an integral component of the tri-snRNP, the loss of co-transcriptional recruitment of the U5 snRNP upon Spt5 depletion observed in the present work indicates that B complex cannot form efficiently or stably in the absence of Spt5. These ChIP experiments cannot distinguish between Spt5 affecting the initial recruitment of the U5 snRNP, or Spt5 affecting the stable association of the U5 snRNP once recruited. In the absence of stable association of the tri-snRNP and B complex formation, the spliceosome cannot be activated, and no splicing catalysis can occur (at least co-transcriptionally). Splicing defects would be predicted in these conditions, and indeed RT-qPCR analyses showed that depletion of Spt5 resulted in pre-mRNA accumulation on the same genes tested by ChIP (Figure 3.6). This is consistent with previous studies which found that mutations in Spt5 caused splicing defects in *S. cerevisiae*, and that depletion of Spt5 results in pre-mRNA accumulation in *S. pombe* (Lindstrom *et al*., 2003; Burckin *et al*., 2005; Shetty *et al*., 2017). While these splicing defects may be modest, it is possible that post-transcriptional splicing could partially compensate for lack of co-transcriptional splicing.

By nature, co-transcriptional splicing is highly sensitive to changes in RNAPII elongation rate and a possible explanation as to how depletion of Spt5 reduces U5 snRNP occupancy is that it affects RNAPII elongation. The prevailing view is that Spt5 reduces pausing or premature termination of RNAPII during transcription elongation and thereby acts as a positive transcription elongation factor through the chromatin barrier (Morillon *et al*., 2003; Mason and Struhl, 2005; Quan and Hartzog, 2010; Diamant *et al*., 2016; Shetty *et al*., 2017). Loss of Spt5 would therefore be predicted to cause a reduction in RNAPII elongation rate, and reduced RNAPII

elongation rate correlates with reduced RNAPII occupancy along genes (Mason and Struhl 2005; Malik *et al*., 2017). Loss of RNAPII occupancy would be predicted in turn to affect co-transcriptional spliceosome assembly and splicing by affecting pre-mRNA substrate availability. I discard this as an explanation for reduced U5 snRNP recruitment for two reasons. Firstly, no consistent changes in RNAPII occupancy were observed in the present work upon Spt5 depletion (Figure 3.4) and there was no effect on RNA production (as determined by exon 2 levels) (Figure 3.6) for the genes analysed, indicating no transcription defect upon Spt5 depletion. This contrasts with a recent study in which Spt5 was depleted for 4.5 hours using the AID system in *S. pombe*, where Spt5 depletion caused reduced RNAPII occupancy 3' ends of genes (indicating RNAPII struggles transcribing past a barrier) and reduced RNA synthesis genome-wide in *S. pombe* upon Spt5 depletion (Shetty *et al*., 2017). This long depletion reduced the cellular level Spt5 to 13% and may have had consequences for RNAPII occupancy (Shetty *et al*., 2017). Secondly, in the present work the loss of splicing factor recruitment upon Spt5 depletion was specific to the U5 snRNP; no loss of earlier-acting splicing factors was observed upon Spt5 depletion and one would predict that changes to RNAPII elongation rate and loss of RNAPII occupancy would affect all stages of co-transcriptional spliceosome assembly. In particular, one might expect the U1 snRNP to be particularly sensitive to changes in RNAPII elongation rate as U1 snRNP occupancy correlates with level of transcription (Harlen *et al*., 2016; Wallace and Beggs, 2017). A further possibility is that the elongation rate of RNAPII is faster upon Spt5 depletion. However, there is no evidence to suggest that loss of Spt5 increases elongation rate of RNAPII – *in vitro* and *in vivo* studies show that loss of Spt5 increases RNAPII pausing, so one would predict elongation rate to be slower in the absence of Spt5 (Morillon *et al*., 2003; Mason and Struhl, 2005; Quan and Hartzog, 2010; Diamant *et al*., 2016; Shetty *et al*., 2017).

Co-immunoprecipitation experiments showed a reciprocal interaction between Spt5 and Prp8 (U5 snRNP) that is not *via* RNAPII and not RNA-dependent (Figure 3.7). It was difficult to detect the U1 or U2 snRNPs upon pulldown of Spt5, but pulldown of the U1 snRNP and U2 snRNP detected Spt5. However, the U2 snRNP pulled down RNAPII too, indicating that this interaction could be indirect, *via* RNAPII. The finding

that, in these conditions, U2 interacts with RNAPII and not with U1 or U5 was in itself interesting. One study found that Prp40 (U1 snRNP) binds to the phosphorylated CTD of RNAPII (Morris and Greenleaf, 2000; Gasch *et al*., 2006), and another by mass spectrometry that the splicing factors (U1 – NTC) are associated with the serine 5 phosphorylated CTD of RNAPII (Harlen *et al*., 2016). Another study found that splicing factor U2AF65 interacts directly with the serine 2 phosphorylated CTD of RNAPII (David *et al*., 2011; Gu *et al*., 2013). U2AF facilitates stable binding of the U2 snRNP to pre-mRNA, and therefore the finding in the present work that U2 interacts with RNAPII fits with these previous observations. Though I cannot rule out interactions between Spt5 and other splicing factors, overall the strongest interaction detected between Spt5 and a splicing factor in the present work was that between Spt5 and the U5 snRNP. An interaction between Prp40 and Spt5 was detected previously, and therefore the co-immunoprecipitation assay in the present work may have missed some interactions (Moore *et al*., 2006). Mass spectrometry would give a clearer indication of the breadth of Spt5-splicing factor interactions. Taken together, these data provide mechanistic insight into how Spt5 could affect splicing - by modulating co-transcriptional recruitment of the U5 snRNP to form the B complex, possibly by direct interaction, at least with the U5 snRNP.

It was possible that the reduction in U5 snRNP recruitment caused by Spt5 depletion was due to loss of Spt5 CTR phosphorylation. The CTR of Spt5 is phosphorylated during transcription by the Bur1/2 complex in *S. cerevisiae* (Liu *et al*., 2009). In the present work, depletion of Bur1 and Bur2 significantly reduced Spt5 phosphorylation, as expected, however, no significant reduction in serine 2 phosphorylation was observed by western blotting (Figure 3.9) or by ChIP (Figure 3.10) upon depletion of Bur1 or Bur2, meaning in these conditions the kinase activity of Bur1 is not required for serine 2 phosphorylation of the CTD. This was unexpected, as deletion of Bur2 and inhibition of Bur1 activity were reported to reduce the level of serine 2 phosphorylation by western blotting and at the 5' end of the *ARG1* gene by ChIP (Liu *et al*., 2009; Qiu *et al*., 2009; Dronamraju and Strahl, 2014). While the loss of serine 2 phosphorylation in whole cell extracts in these studies is clear, the ChIP-qPCR showing loss of serine 2 phosphorylation at the 5' end of genes was performed on one

gene, the intronless gene *ARG1*, which is of course not representative of all genes. Differences in observations made could also be due to the difference in using a conditional system in comparison to a deletion system, or due to differences in antibodies used for detection of the phosphorylated CTD.

Depletion of Bur1 and Bur2 did not result in loss of U5 snRNP recruitment to intron-containing genes, and therefore, the effect observed upon Spt5 depletion was not simply due to loss of phosphorylated Spt5. This is consistent with the fact that the CTR of Spt5 is non-essential in *S. cerevisiae* (Yao *et al*., 2000; Ding *et al*., 2010), and agrees with the findings in *S. pombe* where deletion of the CTR of Spt5 did not cause significant splicing defects in comparison to depletion of Spt5 (Shetty *et al*., 2017). In contrast, Bur1 and Bur2 depletion resulted in a significant accumulation of the U2 and U5 snRNPs on *ACT1* and *ECM33* (Figure 3.11). Further, U2 and U5 snRNPs seemed to peak earlier along the genes, indicating that co-transcriptional spliceosome assembly is faster in the absence of Bur1. As U1 snRNP occupancy did not change, that means that U1 associates and dissociates normally and at least the Bact complex (and complex preceding it) forms normally upon Bur1 and Bur2 depletion.

The observed accumulation of the U2 and U5 snRNPs could potentially be explained by changes in RNAPII elongation rate upon depletion of Bur1 or Bur2 - RNAPII could be pausing more frequently (or for longer periods of time) and accumulating at particular locations and, if this delays progression through the spliceosome cycle, it might explain better detection of the snRNPs by ChIP. However, depletion of Bur1 did not significantly affect RNAPII occupancy, CTD phosphorylation or RNA production (exon 2 levels of nascent RNA and steady-state RNA) for the genes tested (Figures 3.10 and 3.12). Additionally, as discussed below, Paf1 depletion increases RNAPII occupancy on the intron-containing genes tested without causing accumulation of U2 or U5 snRNPs. Further, the occupancy of the U1 snRNP was unaffected in these conditions, and if the elongation rate of RNAPII was affected one would expect changes in occupancy of all the snRNPs, particularly the U1 snRNP (as discussed previously). Though mutations of Bur1 confer sensitivity to 6-azauracil and loss of RNAPII occupancy, which suggests a role for Bur1 in transcription elongation

and processivity (Keogh *et al*., 2003; Murray *et al*., 2001), other studies have not found such a role (Liu *et al.,* 2009). The finding in the present work that depletion of Bur1 does not significantly affect RNAPII occupancy agrees with the finding that inhibition of Bur1 kinase activity does not affect RNAPII occupancy or transcription (Liu *et al*., 2009), and that deletion of Bur2 has no effect on RNAPII occupancy in *S. cerevisiae* (Qiu *et al*., 2009). However, it was recently shown in fission yeast using PRO-SEQ that inhibition of Bur1 kinase activity reduced RNAPII elongation rate (Booth *et al*., 2018). In a similar way, PRO-SEQ could be used to test effects Bur1 depletion on RNAPII elongation rate.

If not due to changes in RNAPII elongation rate, the increased ChIP signal for the U2 or U5 snRNPs upon Bur1 depletion could be explained in two ways. Firstly, as U1 snRNP occupancy is normal upon Bur1/2 depletion, there could be a block to the progression of co-transcriptional spliceosome assembly at some stage past the formation of the catalytically activated spliceosome (Bact complex) (i.e. subsequent to dissociation of U1 snRNP, which would otherwise accumulate) and that Bact or subsequent complexes get stuck at the gene. Alternatively, Bact complex formation and later steps in the splicing cycle might occur more faster and more efficiently in the absence of Bur1. If there was a block to co-transcriptional spliceosome progression, this would be predicted to prevent spliceosome recycling, which would lead to defects in co-transcriptional spliceosome assembly (as snRNPs would be limiting) and defects in splicing catalysis. In contrast, if the Bact complex forms more efficiently, one might predict splicing catalysis would be more efficient, as a result of more co-transcriptional splicing. Loss of Bur1 was found to enhance co-transcriptional splicing of *ACT1* and *ECM33* by NET-RT-qPCR, and because of this we favour the idea that loss of Bur1 enhances co-transcriptional assembly of spliceosomes and that splicing occurs more efficiently in the absence of Bur1 (Figure 3.12). Further, as the U2 and U5 snRNPs peak earlier by ChIP upon Bur1 depletion, this indicates that co-transcriptional spliceosome assembly occurs not only more efficiently but faster, as well as showing that there are no spliceosome recycling defects. We also rule out the possibility of a post-splicing complex accumulating at the gene, as lariat does not accumulate (should

be detected by NET-RT-qPCR) and there were no defects observed in spliceosome recycling or splicing catalysis upon Bur1 depletion (Figure 3.12).

The gene-specific effects of U2 and U5 snRNP accumulation could be explained by differences between non-ribosomal protein genes and ribosomal protein genes. *ACT1* and *ECM33* are both non-ribosomal, whereas *RPS13* is ribosomal. Ribosomal protein genes are spliced quickly and co-transcriptionally in comparison to non-ribosomal genes in *S. cerevisiae*, and perhaps as *RPS13* is already spliced quickly and co-transcriptionally the any effects are not apparent (Wallace and Beggs, 2017).

As a kinase, Bur1 could potentially phosphorylate multiple targets, and it is possible that phosphorylation of proteins other than Spt5 was responsible for the effects on co-transcriptional spliceosome assembly observed upon Bur1 and Bur2 depletion (Figure 3.11). In contrast to depletion of Bur1, no significant changes in pre-mRNA splicing were observed upon depletion of Bur2, which fits with the less severe effects on co-transcriptional spliceosome assembly observed upon Bur2 depletion (Figures 3.11 and 3.12). The less severe effects of Bur2 depletion could be explained by the fact that when Bur2 is depleted the level and co-transcriptional recruitment of Bur1 were unaffected, and Bur1 may have some residual kinase activity without Bur2 or have Bur2-independent substrates that are important for this phenotype. It has been suggested that Bur1 has targets independent of Bur2, or is regulated by other cyclins, as Bur1 is essential whereas Bur2 is non-essential (Yao *et al*., 2000). The finding here that loss of Bur1 kinase activity does not cause splicing defects but enhances co-transcriptional splicing fits with the fact that Bur2 and the CTR are not essential in *S. cerevisiae* (Yao *et al*., 2000; Ding *et al*., 2010). The finding in the present work that a phosphomutant of Spt5 (equivalent to depletion of Bur1/2) showed the same accumulation of the U2 and U5 snRNPs (without significant changes in RNAPII occupancy) as a depletion of Bur1/2 cemented the idea that the phosphorylation status of Spt5 is important for this phenomenon, though the phenotype is milder than depletion of Bur1 and is more similar to depletion of Bur2 (Figures 3.13 and 3.11). Further, comparison of the WT and phosphomutant Spt5 showed no consistent changes in pre-mRNA splicing, which is also analogous to depletion of Bur2 (Figures

3.12 and 3.13). This milder phosphomutant phenotype could be explained by the fact that expression of the Spt5 WT/phosphomutant/phosphomimetic proteins was not inducible, and therefore the strains may have adapted so that changes to co-transcriptional spliceosome assembly are harder to observe. Spt5 phosphorylation by Bur1 does appear to be Bur2-dependent (Figure 3.9A), and therefore another (Bur2-independent) target of Bur1 may be involved that is important for the effects on co-transcriptional spliceosome assembly observed upon Bur1 depletion. At this stage, it cannot be concluded that it is the phosphorylation of Spt5 by Bur1 alone that is responsible for the accumulation of U2 and U5 snRNPs. Bur2-independent targets of Bur1 could be assayed by comparative mass spectrometry after depletion of either Bur1 or Bur2.

Spt5 phosphorylation is required for recruitment of the Paf complex, which facilitates transcription elongation by RNAPII (Laribee *et al*., 2005; Liu *et al*., 2009). It is possible that the effects observed upon depletion of Bur1/2 and with the Spt5 phosphomutant were due to loss of Paf complex recruitment. However, depletion of Paf1, the core component of the Paf complex, did not cause accumulation of the U2 and U5 snRNPs, meaning that this phenotype cannot be explained by loss of Paf1 recruitment (Figure 3.19). Furthermore, no significant changes in pre-mRNA splicing were observed (Figure 3.20). Interestingly, Paf1 depletion did result in RNAPII accumulation and an increase in serine 2 and serine 5 phosphorylation of the CTD of RNAPII (Figure 3.19). This accumulation of RNAPII in gene bodies is similar to the increase in RNAPII occupancy observed by Chen *et al*., (2015) upon Paf1 knockdown in mammalian cells. The increase in serine 2 phosphorylation observed by ChIP-qPCR was unexpected as previous ChIP-qPCR studies showed that deletion of Paf1 reduced serine 2 phosphorylation in *S. cerevisiae* (Mueller *et al*., 2004; Nordick *et al*., 2008). However, Chen et al (2015) also observe an increase in serine 2 phosphorylation after Paf1 knockdown in mammalian cells (Chen *et al*., 2015). In *S. cerevisiae*, NET-SEQ showed that deletion of Paf1 resulted in a shift of the 5' peak of RNAPII toward the 3' end of genes (Fischl *et al*., 2017). These data also act as a control for the Bur1 depletion data – depletion of Paf1 increases RNAPII occupancy without causing co-transcriptional accumulation of the U2 and U5 snRNPs.

How might loss of Bur1 kinase activity and Spt5 phosphorylation enhance co-transcriptional formation of spliceosomes? In mammals, phosphorylation of the CTR of Spt5 has been shown to convert Spt5 from an inhibitor of transcription elongation to a promoter of transcription elongation. Though this two-part function for Spt5 has not been demonstrated in *S. cerevisiae* (Spt5 is thought to only act only as a positive regulator), an intriguing possibility is that loss of Spt5 phosphorylation causes transient pausing of RNAPII and this transient pause is sufficient to promote co-transcriptional splicing. Indeed, RNAPII has been demonstrated to transiently pause in a splicing-dependent manner on an inducible reporter gene in *S. cerevisiae*, the mechanism for which is currently unknown (Alexander *et al*., 2010b). Further, RNAPII pausing was also demonstrated at intron-exon junctions in *S. cerevisiae* (Harlen *et al*., 2016). As pausing is highly transient, it would be difficult to detect using the methods described in the present work and use of an inducible reporter gene where transcription is synchronised or more high-throughput methods such as NET-SEQ would be more informative. As previously discussed, the changes in pausing would still have to be transient enough not to be detected in the present work (no changes in RNAPII occupancy observed). A second non-mutually-exclusive model is that at least the unphosphorylated form of Spt5 and Bur1 kinase activity facilitates structural rearrangements or dynamics in the spliceosome, potentially *via* interaction with splicing factors acting at the B to Bact transition, such as Brr2 (Maeder *et al*., 2009). This could be analogous to the interaction of Spt5 with the capping enzymes - the unphosphorylated CTR of Spt5 facilitates recruitment of capping enzymes, whilst phosphorylation by Bur1 releases the capping enzymes once capping is complete and releases paused RNAPII (Wen and Shatkin, 1999; Lidschreiber *et al*., 2013; Doamekpor *et al*., 2014, 2015). Both models for the role of unphosphorylated form of Spt5 in splicing would rely on exchange of phosphorylated Spt5, *de novo* recruitment of unphosphorylated Spt5 or the inhibition of Bur1 kinase activity and activity of a phosphatase. In *S. pombe*, Dis2 (PP1 in mammals, Glc7 in *S. cerevisiae*) was recently shown to be the phosphatase for Spt5 that functions in the absence of Bur1 kinase activity (Parua *et al*., 2018).

Proteins in the spliceosome could be regulated by Bur1 phosphorylation. There are multiple reversible phosphorylation events during spliceosome assembly that are thought to be important for structural rearrangements within the spliceosome (reviewed in Will and Luhrmann, 2011). In the present work, dephosphorylation of proteins acting at the B to Bact transition by loss of Bur1 could promote the B to Bact transition. Dephosphorylation of U1 and SR proteins is required for the first step of splicing in mammals (Tazi *et al*., 1993; Cao *et al*., 1997). In mammals, SF3b1 has been shown to be phosphorylated by kinase DYRK1A, which is important for splicing catalysis (Wang *et al*., 1998; de Graaf *et al*., 2006). Cyclin-dependent kinases have recently been shown to regulate the association of SF3b1 with chromatin (Murthy *et al*., 2018). However, many of the phosphorylated domains discussed above are missing in *S. cerevisiae,* and the kinases and phosphatases absent so the importance of phosphorylation of the spliceosome in *S. cerevisiae* is less clear than in mammals (reviewed in Will and Luhrmann, 2011).

Together, experiments described here suggest a two-part function for Spt5 in co-transcriptional spliceosome assembly in *S. cerevisiae* (Figure 3.21). Firstly, its physical presence is required for proper recruitment or stable association of the U5 snRNP and formation of a pre-catalytically activated spliceosome (B complex) (possibly by direct/indirect interaction between Spt5 and at least the U5 snRNP). This phenomenon is not dependent on downstream recruitment of Paf1 or the phosphorylation status of Spt5. Secondly, the loss of Bur1 kinase activity and at least the unphosphorylated CTR of Spt5 promotes co-transcriptional formation of a

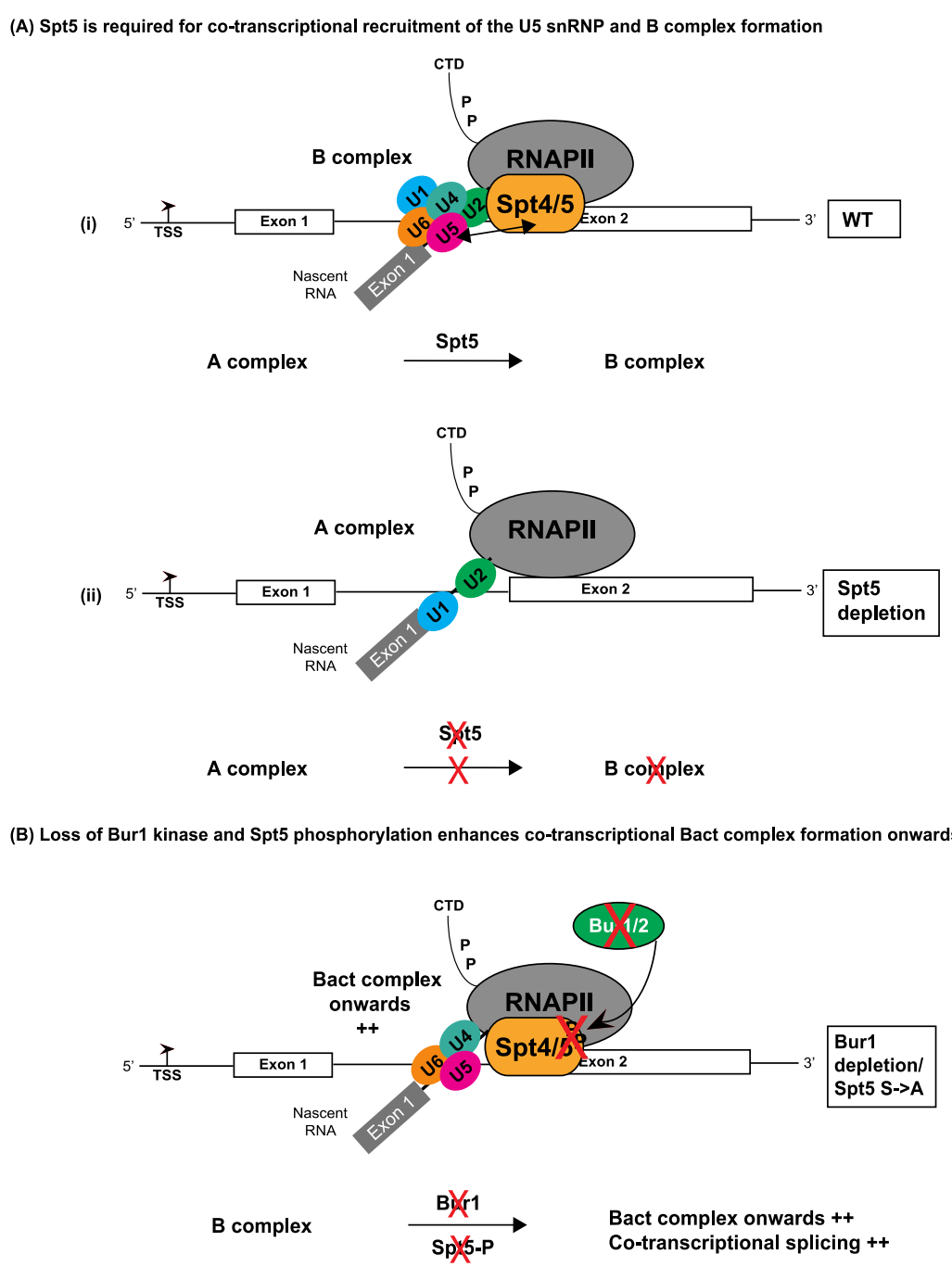


Figure 3.21. A two-part role for Spt5 in co-transcriptional spliceosome assembly in *S. cerevisiae*

(A) (i) In wild-type (WT) conditions (without Spt5 depletion), Spt5 facilitates proper recruitment or stable association of the U5 snRNP and therefore formation of a pre-catalytically activated spliceosome (B complex). This may be mediated, either directly or indirectly, by interaction between Spt5 and the U5 snRNP. (ii) Upon Spt5 depletion (indicated by the red cross), the U5 snRNP is not properly recruited or stably associated, and the B complex cannot form (indicated by the red cross), leading to defects in splicing catalysis.
(B) The loss of Bur1 kinase activity (indicated by the red cross) and at least the subsequent unphosphorylated CTR (indicated by the red cross) of Spt5 directly or indirectly promotes (++) co-transcriptional assembly of spliceosomes at the stage of Bact formation or later, and in this way promotes co-transcriptional splicing.

catalytically activated spliceosome (Bact complex onwards) and promotes co-transcriptional splicing in an unknown mechanism.

Whether this model would be applicable for higher eukaryotes remains to be seen. It is certainly possible, as Spt5 is highly conserved from yeast to humans (DSIF in humans), as is phosphorylation of its CTR by Bur1/2 (P-TEFb) and splicing occurs co-transcriptionally. There is evidence that Spt5 affects alternative splicing and constitutive splicing in mammalian cells (Diamant *et al*., 2012; Liu *et al*., 2012). Furthermore, there is some evidence that P-TEFb affects alternative splicing patterns in mammalian cells, by causing inclusion of an alternative exon (Barboric *et al.,* 2009). However, this is *via* phosphorylation of the CTD of RNAPII which facilitates recruitment of SR splicing factor SF2/ASF, which is important for splice site recognition (Barboric *et al*., 2009). In mammalian cells, DSIF is converted to a positive elongation factor by interaction with elongation factor Tat-SF1 (Chen *et al.,* 2009). Tat-SF1 and yeast homolog Cus2 interact with splicing factors, and it has been suggested that the phosphorylation of DSIF by P-TEFb and recruitment of Tat-SF1 could facilitate co-transcriptional splicing (Yan *et al.,* 1998; Fong and Zhou, 2001; Lenasi and Barboric, 2010). In mammalian cells, there is strong evidence that P-TEFb phosphorylates serine 2 of the CTD of RNAPII, and therefore this model is not applicable to *S. cerevisiae* (Wood and Shilatifard, 2006; Bartkowiak *et al*., 2011). While Spt5 and its phosphorylation by P-TEFb are conserved, the functions are likely divergent – this is exemplified by the fact that phosphorylation of the DSIF in mammalian systems is known to be important for release of RNAPII promoter-proximal pausing, a function that has not been demonstrated in *S. cerevisiae* (Adelman and Lis, 2012). Furthermore, there are key differences in splicing and gene architecture between yeast and humans that complicate matters. The splicing machinery is expanded in human cells to accommodate alternative splicing - 95% of transcripts are alternatively spliced and only 3% of genes are intronless (Kornblihtt *et al*., 2013). This contrasts with *S. cerevisiae* where 5% of genes contain a single intron and there is very little alternative splicing (Qin *et al*., 2016; Hossain *et al*., 2016; Juneau *et al*., 2009).

Depletion of Ctk1 was an effective way to reduce serine 2 phosphorylation at the 3' ends of genes, without affecting RNAPII occupancy (Figure 3.15). The large decrease in serine 2 phosphorylation of the CTD upon Ctk1 depletion fits with the idea that Ctk1 is the major serine 2 CTD kinase in *S. cerevisiae* (Cho *et al*., 2001). In contrast, Bur1 depletion did reduce serine 2 phosphorylation of the CTD but significantly reduced Spt5 phosphorylation (in contrast, Ctk1 depletion did not significantly affect Spt5 phosphorylation) (Figures 3.9, 3.10 and 3.15). These data fit with the idea that the activity of P-TEFb is split between Bur1 and Ctk1 in *S. cerevisiae* (Keogh *et al*., 2003; Wood and Shilatifard, 2006). It was surprising that loss of serine 2 phosphorylation by Ctk1 depletion in these conditions did not affect co-transcriptional spliceosome assembly or splicing (Figures 3.16 and 3.17). The finding that loss of serine 2 phosphorylation does not affect splicing agrees with a study in *S. cerevisiae* that expressed truncated CTD mutants (8 repeats) in which the serine 2 residues were mutated to alanine (Harlen *et al*., 2016). These results contrast with findings in mammalian cells that a mutation of serine 2 of the CTD to alanine reduced recruitment of the U2 snRNA to transcription sites (assayed by FISH), and caused defects in pre-mRNA splicing (Gu *et al*., 2013). Further, it was shown that splicing factor U2AF65 interacts directly with the serine 2 phosphorylated CTD of RNAPII, suggesting a reciprocal relationship between serine 2 phosphorylation and splicing in mammalian cells (David *et al*., 2011). The observed differences may be explained by differences in the complexity of splicing between *S. cerevisiae* and mammals. The Ctk1 depletion act as a useful control for the Bur1 depletion experiments and shows that the effects observed upon Bur1 depletion are not likely to be due to loss of serine 2 phosphorylation.

Chapter 4. Links between splicing and chromatin

4.1 Background

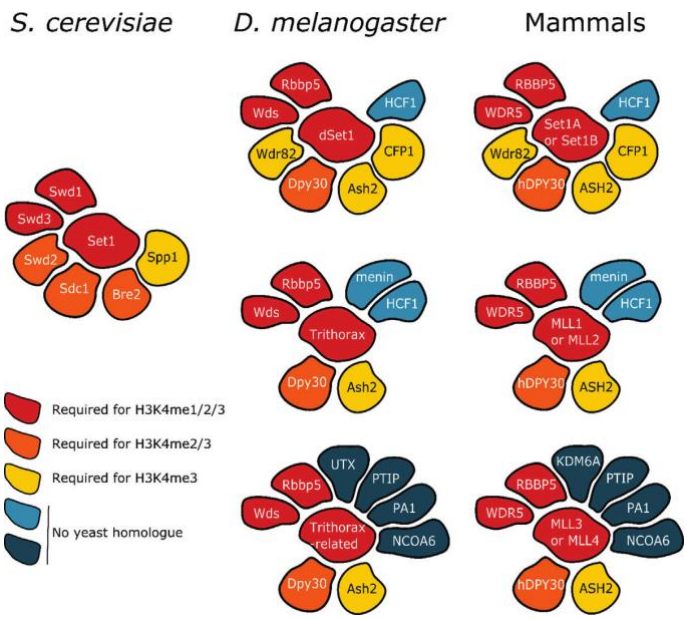
Of particular interest to the present work were the links between splicing and H3K4me3. There are several pieces of evidence that H3K4me3 affects splicing and *vice-versa*. In mammalian cells, RNAi knockdown of Ash2 (a subunit of the Set1 complex specific for H3K4me3) and RNAi knockdown of Chd1 (a reader of H3K4me3 and chromatin remodeller) both result in reduction of the association of the U2 snRNP component SF3a with chromatin and reduced pre-mRNA splicing efficiency (Sims *et al*., 2007). This showed the first functional link between H3K4me3 and pre-mRNA splicing. It was later shown that Ash2 overexpression affects patterns of alternative splicing in mammalian cells (Luco *et al*., 2010)). The authors also found that H3K4me3 is able to enrich for core members of the U2 snRNP, and that Chd1 physically interacts with SF3a, providing a model whereby Chd1 binding to H3K4me3 recruits the U2 snRNP and influences pre-mRNA splicing (Sims *et al*., 2007). The H3K4me3 reader Sgf29 has also been demonstrated to interact with U2 snRNP components (Vermeulen *et al*., 2010).

Subsequently, examination of genome-wide ChIP-sequencing data sets in mammalian cells showed that the peak of H3K4me3 was enriched at the first 5'SS of intron-containing genes but not at downstream splice sites, even when proximal to the transcription start site (Bieberstein *et al*., 2012). When the 5'SS was near a promoter, a single peak of H3K4me3 was observed but when the first exon was larger than 500 nucleotides, two peaks of H3K4me3 were observed. Further, a reporter gene with a 3'SS mutation that blocks pre-mRNA splicing, or an intronless version, exhibit reduced levels of H3K4me3 relative to the wild-type intron-containing reporter (Bieberstein *et al*., 2012). In addition, the use of splicing inhibitor drug spliceostatin reduced H3K4me3 levels (Bieberstein *et al*., 2012). In contrast, deletion of Set1 in *S. cerevisiae* did not affect co-transcriptional spliceosome assembly and microarray experiments showed reduced pre-mRNA splicing of only a subset of genes (Hérissant *et al*., 2014). Deletion of Chd1 in *S. cerevisiae* resulted in the strongest changes to

H3K4me3 on intron-containing genes, and lead to an increased pre-mRNA splicing efficiency (Lee *et al*., 2017).

There are three possible histone methylation states of H3K4: mono- (me1), di-(me2) and tri-methylation (me3). The patterns of H3K4me3, H3K4me2 and H3K4me1 form a gradient, with the peak of H3K4me2 on a nucleosome downstream of H3K4me3, and the peak of H3K4me1 on a nucleosome downstream of H3K4me2 on actively transcribed genes (Howe *et al*., 2017). These different methylation states are thought to be brought about by a "time model", which correlates the amount of time Set1 spends at a particular genomic location to a higher level of H3K4 methylation (Ng *et al*., 2003b; Soares *et al*., 2017). The histone methyltransferases responsible for H3K4 methylation are conserved from yeast to humans, and each contain the Set1 methyltransferase, which contains within it a conserved SET histone methyltransferase domain. Metazoans have multiple complexes responsible for H3K4 methylation while in *S. cerevisiae* there is a single histone methyltransferase called Set1. Set1 is present in a complex called COMPASS (complex associated with Set1), and in *S. cerevisiae*, the COMPASS complex is composed of Set1 and seven other structural proteins: Bre2, Swd1, Spp1, Swd2, Swd3 and Sdc1 and Shg1. Set1, Swd1 and Swd3 are required for all three H3K4 methylation states, Sdc1, Swd2 and Bre2 are required for H3K4me2 and H3K4me3, and Spp1 is specifically required for H3K4me3 *in vivo* (Cheng *et al*., 2004; Dehé *et al*., 2006; Lee *et al*., 2007; Schneider *et al*., 2005) (Figure 4.1). In mammals, there are at least six histone methyltransferase complexes responsible for H3K4me3 (SET1A, SET1B, MLL1, MLL2, MLL3 and MLL4). The mammalian Set1-containing complexes are most similar to the *S. cerevisiae* Set1 complex (reviewed in Shilatifard, 2012) (Figure 4.1). In mammals, the six Set1-containing complexes are not functionally redundant, as deletion of individual complex components is embryonic lethal (Glaser *et al*., 2006; Lee *et al*., 2008).

Set1 is recruited during transcription to the CTD of transcribing RNAPII *via* the Paf complex, which interacts with the serine 5 phosphorylated CTD of RNAPII (Krogan *et al*., 2003; Ng *et al*., 2003b). Serine 5 phosphorylation is enriched at the 5'end of genes and could help to explain the enrichment of H3K4me3 there. It was also shown



(Howe *et al*., 2017)

Figure 4.1. The Set1 complex in *S. cerevisiae, D. melanogaster* and mammals

Set1 complexes in *S. cerevisiae*, *D. melanogaster* and mammals. Subunits required for H3K4me3 are shown in yellow, subunits required for H3K4me2/3 are shown in orange, subunits required for H3K4me1/2/3 are shown in red. Species-specific subunits are shown in blue. Figure taken from (Howe *et al*., 2017).

that the Swd2 subunit of the Set1 complex interacts with histone H2B ubiquitination (Lee *et al*., 2007). However, recent studies showed that the distribution of Set1 does not correlate with serine 5 phosphorylation (Sayou *et al*., 2017), and it was proposed that there must be alternative ways by which Set1 is recruited to chromatin. Set1 has two RNA recognition motifs (Schlichter and Cairns, 2005), and was recently shown by several groups to preferentially bind the 5'ends of nascent RNAPII transcripts, which has been proposed to explain the enrichment of H3K4me3 at the 5' ends of genes – due to increased Set1 occupancy *via* RNA binding (Battaglia *et al*., 2017; Luciano *et al*., 2017; Sayou *et al*., 2017). Despite recent advances, the precise mechanism of Set1 recruitment is not yet fully characterised (Luciano *et al*., 2017).

Genome-wide ChIP studies have shown a conserved peak of H3K4me3 from yeast to humans at the transcription start sites of actively transcribed genes, the level of which positively correlates with level of transcription (Benayoun *et al*., 2014; Chen *et al*.,

2015; Howe *et al*., 2017). As H3K4me3 correlates with active transcription, it is generally thought of as an activating chromatin mark that stimulates RNAPII transcription. However, knockdown of H3K4me3 in *S. cerevisiae* and mammalian systems shows little effect on global steady-state gene expression, arguing against a global role of H3K4me3 as a transcription activator (Margaritis *et al*., 2012; Howe *et al*., 2017). In fact, several studies suggest a role for Set1 in gene silencing in yeast (Nislow *et al*., 1997; Briggs *et al*., 2001; Margaritis *et al*., 2012; Castelnuovo *et al*., 2014). As knockdown of Set1 reduces all H3K4 methylation states, a strategy used to specifically study the function of H3K4me3 is to specifically deplete Spp1 (Cfp1 in mammals), which reduces H3K4me3 without affecting H3K4me1 or H3K4me2. In *S. cerevisiae*, neither Set1 nor Spp1 is an essential gene, and global loss of H3K4me3 does not affect viability and does not result in significant changes to gene expression (Briggs *et al*., 2001; Miller *et al*., 2001). In mammalian cells, loss of H3K4me3 prevents proper embryogenesis, although cultured embryonic stem cells are viable (Carlone and Skalnik, 2001; Carlone *et al*., 2005). Depletion of Cfp1 (Spp1) to reduce H3K4me3 in mammalian cells did not result in significant changes to gene expression (Clouaire *et al*., 2012). Therefore, the precise role for H3K4me3 in transcription from yeast to mammals currently remains unclear (Howe *et al*., 2017). The lack of effects on transcription upon loss of H3K4me3 suggests that low levels of H3K4me3 are sufficient for any effect on transcription it might have. In *S. cerevisiae*, H3K4me3 (and H3K4me2) was shown to be a highly stable chromatin mark, remaining at the *GAL10* gene for 5 hours after switching off expression of *GAL10*, which reduced transcription and Set1 recruitment within 1 hour (Ng *et al*., 2003b). This means that either the demethylase responsible for removal of H3K4 methylation is not very active, or that loss of H3K4 methylation normally relies on dilution of histones during DNA replication (Ng *et al*., 2003b; Radman-Livaja *et al*., 2010). It also infers that H3K4 methylation is not an inherited mark through cell division, and its placement is *de novo* with each round of transcription (Ng *et al*., 2003b).

The particular stage of the pre-mRNA splicing cycle that might affect H3K4me3 is unknown, and the precise molecular mechanisms involved remain unclear. Further, it is not known whether the link between H3K4me3 and splicing is conserved from

mammals to *S. cerevisiae*. The present work describes the application of the AID system to conditionally deplete essential splicing factors in *S. cerevisiae* to block different stages of spliceosome assembly and pre-mRNA splicing in order to determine any effects on H3K4me3 *in vivo*. Splicing factors were selected at different stages throughout the splicing cycle, in an attempt to pin-point which particular stage of splicing is important for any changes in H3K4me3 observed.

In the present work, the AID system was used to conditionally deplete the following essential splicing factors: Prp39, Prp9, Slu7 and Prp22, which are all conserved from yeast to humans (Figure 4.2). Prp39, Prp9, Slu7 and Prp22 were each conditionally depleted using the AID system. Effects on pre-mRNA splicing were determined by RT-qPCR and subsequent *in vivo* effects on H3K4me3 determined by ChIP-qPCR. Prp39 is a core member of the U1 snRNP that is essential for splicing, functioning at the earliest stage of splicing (5'SS recognition) and required for the commitment of a pre-mRNA to the splicing pathway (Lockhart and Rymond, 1994; Gottschalk *et al*., 1998). Prp9 is a core member of the U2 snRNP and is essential in forming the pre-spliceosome and for pre-mRNA splicing (Hodges and Beggs, 1994; Wiest *et al*., 1996). Prp22 is an essential DExH-box ATP-dependent 3'-5' RNA helicase that disrupts the contacts between the U5 snRNP and mRNA after exon ligation to release the spliced mRNA (Company *et al*., 1991; Schwer and Cross, 1998; Wagner *et al*., 1998). Prp22 has also been shown to be required for the second catalytic step (Schwer and Cross, 1998), and has been proposed to facilitate 3'SS selection and perform a proof-reading function prior to the second step – discarding suboptimal substrates (Mayas *et al*., 2006).

At this point, I would like to gratefully acknowledge Dr. Emanuela Sani (Beggs lab) who generated and characterised the Slu7-AID* and Prp22-AID* strains used in this study, and who provided the H3K36me3 data. I would also like to gratefully acknowledge David Barrass (Beggs lab) who helped me perform 4-thiouracil labelling and purification of newly synthesised RNA.

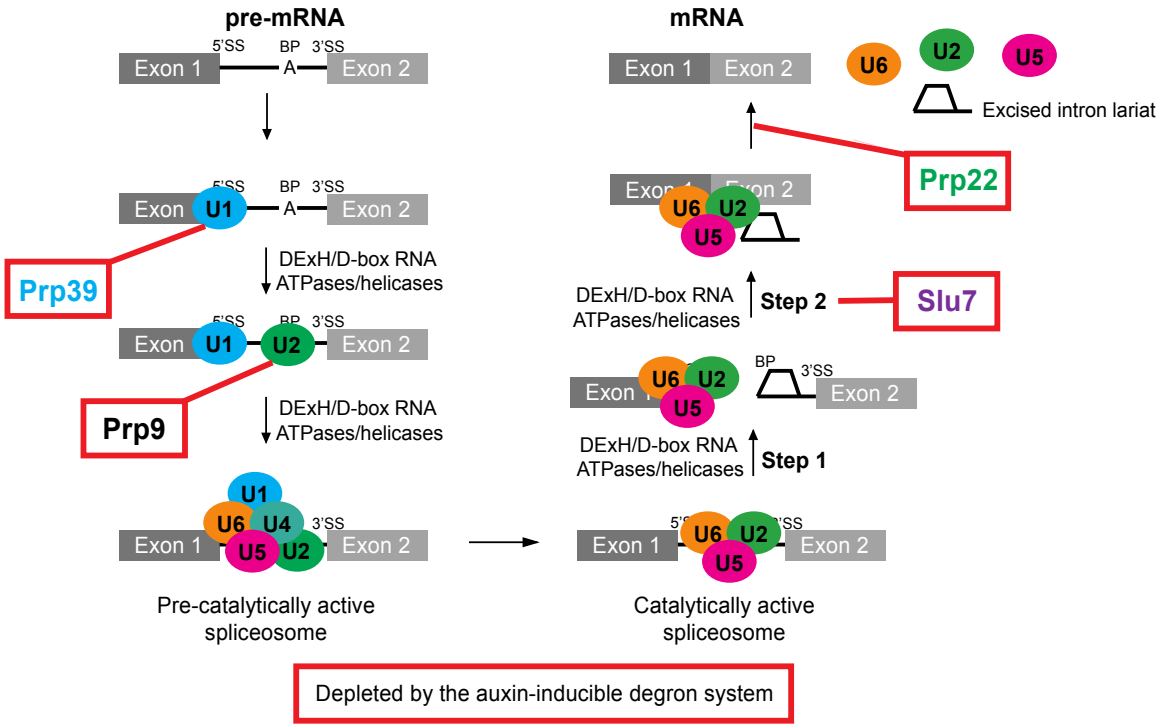


Figure 4.2. Choosing splicing factors throughout the splicing cycle to address which stage of splicing, if any, affects H3K4me3

Classical splicing cycle showing where the essential splicing factors of interest to this work act in the splicing cycle: Prp39 (blue, U1 snRNP), Prp9 (black, U2 snRNP), Slu7 (purple) and Prp22 (green). Figure adapted from Will and Lührmann (2011).

4.2 Use of the AID system to conditionally deplete essential splicing factors

To enable analysis of the effects of depletion of essential splicing factors on H3K4me3 *in vivo*, Prp39, Prp9, Slu7 and Prp22 were each C-terminally AID*-tagged in a strain that constitutively expressed Tir1, enabling their conditional depletion in the presence of auxin by the AID system. Experiments were performed with or without 30 minutes of auxin treatment.

Firstly, the effect of the AID* tag on function of each splicing factor and auxin-sensitivity was assessed by comparing the growth of the parental strain (W303-OsTIR) that has no AID* tag with the Prp39-AID* strain, the Prp9-AID* strain, the Slu7-AID* strain and the Prp22-AID* strain with or without auxin treatment (Figure 4.3A). In the absence of auxin, the AID*-tagged strains grew comparably to the parental strain, showing that the C-terminal AID* tag does not affect their function. After auxin addition to deplete the AID*-tagged proteins, each AID*-tagged strain grew more slowly and the parental strain was unaffected. The growth defect upon depletion of the AID*-tagged proteins was expected as these proteins are essential in *S. cerevisiae*.

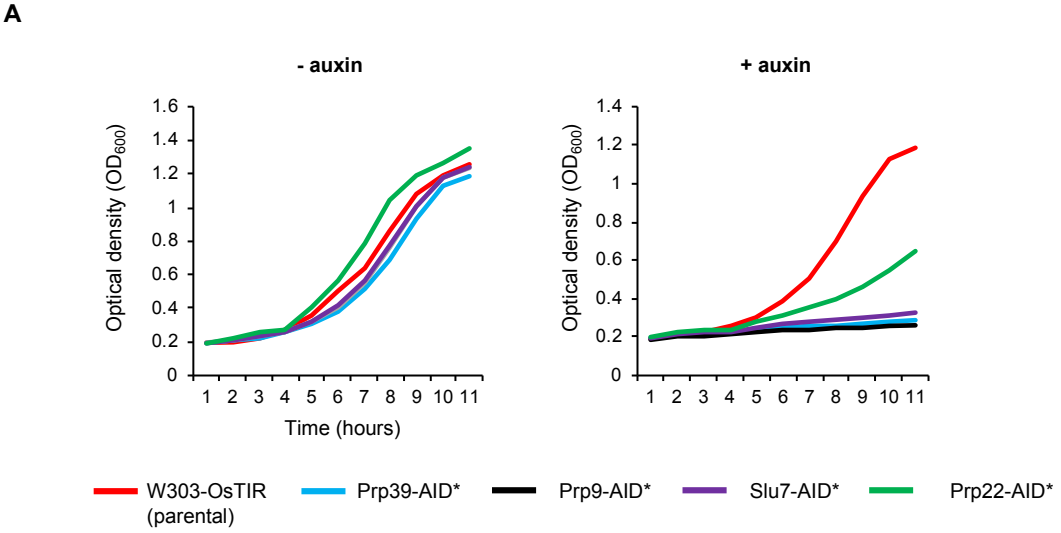


Figure 4.3. Use of the AID system to conditionally deplete essential splicing factors

(A) Growth of the control parental strain (W303-OsTIR; no AID*-tagged proteins; red), Prp39-AID* strain (red), Prp9-AID* strain (black), Slu7-AID* strain (purple) and Prp22-AID* strain (green) was measured over a period of 11 hours without (-) auxin or with (+) auxin treatment. Mean of three biological replicates.

Western blotting showed that 30 minutes of auxin treatment significantly reduced the level of each of Prp39, Prp9, Slu7 and Prp22 on average to 16%, 16%, 19% and 45% relative to conditions without auxin (student's t-test, $P<0.05$) (Figure 4.3B). It is estimated that Prp39 is present at approximately 177 molecules/cell, Prp9 at approximately 291 molecules/cell, Slu7 at approximately 425 molecules/cell and Prp22 at approximately 232 molecules/cell. Therefore, after auxin treatment for 30 minutes, there were approximately 28 molecules/cell of Prp39, 47 molecules/cell of Prp9, 81 molecules/cell of Slu7 and 104 molecules/cell of Prp22 remaining (Kulak *et al*., 2014). It was interesting that Prp39, Prp9 and Slu7 were all better depleted than Prp22 in these experimental conditions. This is not explained by differences in protein copy number, as the estimates of protein copy number for Prp39, Prp9, Slu7 and Prp22 are all very similar, and is probably explained by differing accessibility of the C-terminal AID* tag for degradation by the AID system (Kulak *et al*., 2014).

To test whether the chromatin-associated pool of each splicing factor was depleted, ChIP was performed using a FLAG antibody followed by qPCR analyses across intron-containing genes *ACT1* and *ECM33* (Figure 4.3C). Firstly, the ChIP profile across the intron-containing genes *ACT1* and *ECM33* was correct – the earlier acting splicing factors Prp39 and Prp9 peak earlier along the genes (towards the 5'end) than later acting factors Slu7 and Prp22 (peak toward the 3'end, at the end of exon 2). ChIP-qPCR showed that, in addition to being depleted in whole cell extracts, each of the essential splicing factors was well depleted from chromatin (Figure 4.3C). Prp39 was significantly depleted on average to 44% at amplicons 3 and 4 (3'SS and exon 2) on *ACT1*, and 52% at amplicons 2 and 3 (exon 1 and 3'SS) on *ECM33* (student's t-test, $P<0.05$), relative to conditions without auxin (Figure 4.3C). Prp9 was significantly depleted on average to 51% at amplicons 3, 4 and 5 (3'SS-exon 2) on *ACT1*, and 47% at amplicons 3, 4 and 5 (3'SS-exon 2) on *ECM33* (student's t-test, $P<0.05$), relative to conditions without auxin (Figure 4.3C). Slu7 was significantly depleted on average to 54% at amplicon 5 (exon 2) on *ACT1*, and 70% at amplicons 4 and 5 (exon 2) on *ECM33* (student's t-test, $P<0.05$), relative to conditions without auxin (Figure 4.3C). Prp22 was significantly depleted on average to 54% at amplicon 5 (exon 2) on *ACT1*, and 37% at amplicons 3 and 4 (3'SS and exon 2) on *ECM33* (student's t-test, $P<0.05$),

relative to conditions without auxin (Figure 4.3C). The finding that each factor was depleted co-transcriptionally was important, as it is the chromatin-associated pool that would most likely affect H3K4me3.

Next, a splicing RT-qPCR assay was used to determine whether depletion of Prp39, Prp9, Slu7 and Prp22 in these conditions resulted in the expected splicing defects (Figures 4.3D and 4.3E). Prp39 and Prp9 are first step splicing factors, and their depletion would be expected to result in accumulation of pre-mRNA (5'SS, 3'SS and BP) species, and reduction in lariat (product of first step), which is exactly what was observed for the intron-containing genes *ACT1* and *RPL28* (Figure 4.3E). Depletion of Prp39 and Prp9 also caused pre-mRNA accumulation of intron-containing genes *RPS13* (5'SS and 3'SS) and *ECM33* (5'SS), where no lariat primer sets were available (Figure 4.3E. Depletion of Slu7 resulted in the accumulation of 3'SS, BP and lariat species of *ACT1*, 5'SS and lariat of *RPL28*, 5'SS and 3'SS of *RPS13* and 5'SS of *ECM33*. The accumulation of lariat in addition to pre-mRNA (5'SS/BP) species indicates a mixture of a first and second step splicing defect upon Slu7 depletion (Figure 4.3E). Depletion of each of Prp39, Prp9 and Slu7 caused a significant reduction in mRNA and exon 2 levels for all genes tested. In contrast, Prp22 depletion resulted in no pre-mRNA accumulation on all genes tested, however for *ACT1* and *RPL28,* the lariat species significantly accumulated (Figure 4.3E). As no significant accumulation of the 3'SS *ACT1* was observed in these conditions, this means that the lariat accumulating must be excised intron-lariat, and under these conditions, depletion of Prp22 does not cause defects in splicing catalysis, but causes defects in spliceosome disassembly after the second step. No significant decrease in mRNA or exon 2 levels were observed for any of the genes tested upon Prp22 depletion in these conditions, giving further evidence for no defects in splicing catalysis in these conditions.

In summary, the AID system worked well to deplete the essential splicing factors Prp39, Prp9, Slu7 and Prp22 from whole cell extracts and chromatin and in addition caused the expected defects in splicing catalysis. The loss of exon 2 and mRNA species upon depletion of Prp39, Prp9 and Slu7 could be explained by degradation of spliced RNA. As pre-mRNA species accumulate, it is not indicative of a transcription defect.

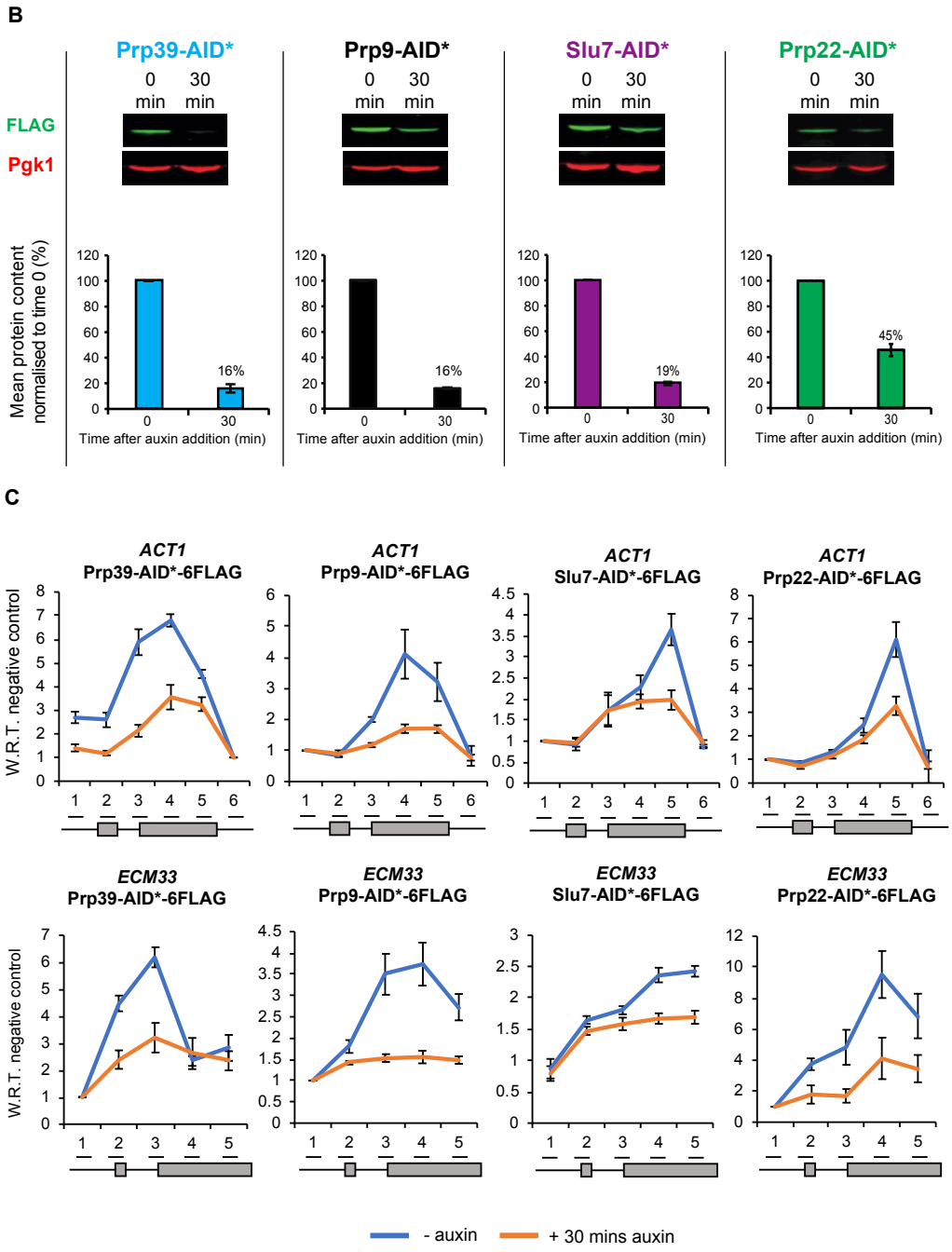


Figure 4.3 (continued). Use of the AID system to conditionally deplete essential splicing factors

(B) Western blot probed with anti-FLAG and anti-Pgk1 as a loading control. Samples were taken before (0) and 30 minutes (30) after addition of auxin. Prp39-AID*, Prp9-AID*, Slu7-AID* and Prp22-AID* depletion was quantified and shown as the percentage mean of 3 biological replicates after 30 minutes of auxin addition with respect to (W.R.T) time zero and normalised to the Pgk1 signal. Error bars = standard error of the mean.
(C) Anti-FLAG ChIP followed by qPCR analysis of the intron-containing genes *ACT1* and *ECM33* without (-) auxin (blue) or with (+) 30 minutes of auxin (orange) addition to deplete Prp39-AID*, Prp9-AID*, Slu7-AID* and Prp22-AID*. The X-axis shows the positions of amplicons used for ChIP qPCR analysis – the exons are of each gene in grey. The data is presented as the mean percentage of input normalised to a negative control amplicon. Mean of at least three biological replicates. Error bars = standard error of the mean.

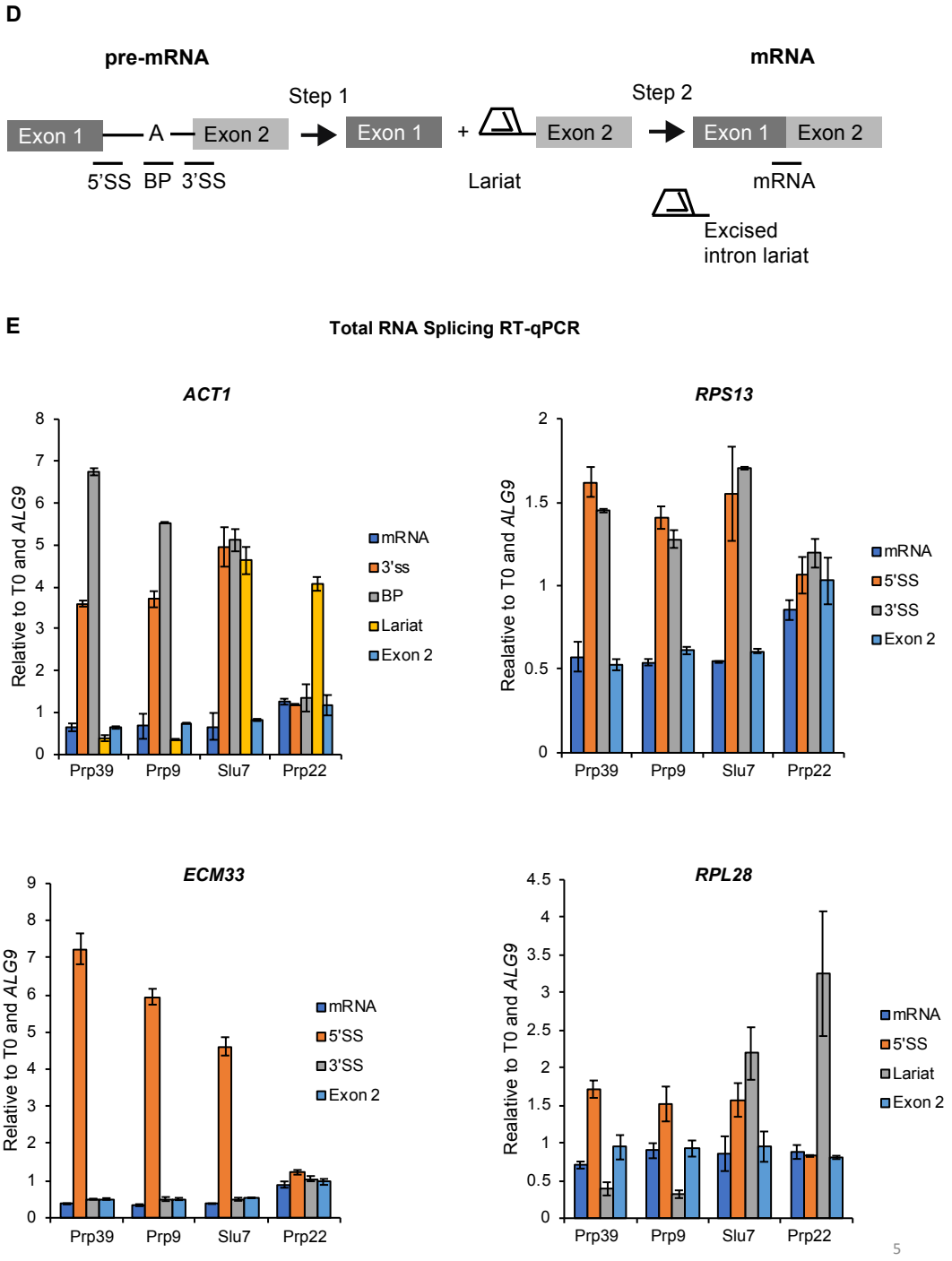


Figure 4.3 (continued). Use of the AID system to conditionally deplete essential splicing factors

(D) Splicing RT-qPCR cartoon showing the primers used for the splicing RT-qPCR assay for an average gene. Primers used detected pre-mRNA (5'SS or BP and 3'SS), lariat (excised intron or intron-exon 2), exon 2 (ex 2) and mRNA.

(E) RT-qPCR analysis of total RNA from the intron-containing genes *ACT1*, *RPS13*, *ECM33* and *RPL28* after 30 minutes of depletion of Prp39-AID*, Prp9-AID*, Slu7-AID* and Prp22-AID*. Normalised to *ALG9* intronless transcript and time zero (without auxin addition). Mean of at least 3 biological replicates. Error bars = standard error of the mean.

4.3 Depletion of essential splicing factors that affect the first, or first and second, step of splicing reduces the level of H3K4me3

To determine whether the level of h3K4me3 is altered by depletion of splicing factors that affect the first (Prp39, Prp9), or first and second (Slu7) steps in these, ChIP was performed using antibodies against H3K4me3 and histone H3 followed by qPCR across intron-containing genes *ACT1*, *RPS13* and *ECM33*. H3K4me3 peaks towards the 5' end of actively transcribed genes and declines toward the 3'end, and H3K4me3 across *ACT1*, *RPS13* and *ECM33* showed this correct pattern. Depletion of Prp39 on average significantly reduced the level of H3K4me3 to 59% at amplicon 2 (exon 1) of *ECM33* (student's t-test, $P<0.05$), 66%, 68% and 71% at amplicons 1-3 (5'-3'SS) of *ACT1* (student's t-test, $P<0.05$), and to 54% at amplicon 2 (exon 1) of *RPS13* (student's t-test, $P<0.05$), relative to conditions without depletion (Figure 4.4A). Depletion of Prp9 on average significantly reduced the level of H3K4me3 to 67% and 75% at amplicons 1 and 2 (5'-exon 1) of *ECM33* (student's t-test, $P<0.05$), 75% at amplicon 2 (exon 1) of *ACT1* ($P<0.05$), and to 74% at amplicon 2 (exon 1) of *RPS13* (student's t-test, $P<0.05$), relative to conditions without depletion (Figure 4.4B). Depletion of Slu7 on average significantly reduced the level of H3K4me3 to 76% and 75% at amplicons 2 and 3 (exon 1 and 3'SS) of *ECM33* (student's t-test, $P<0.05$), 82%, 78% and 80% at amplicons 1-3 (5'-3'SS) of *ACT1* ($P<0.05$), and to 71% at amplicon 2 (exon 1) of *RPS13* (student's t-test, $P<0.05$), relative to conditions without depletion (Figure 4.4C).

Overall, under these conditions depletion of splicing factors that affect the first (Prp39, Prp9), or first and second (Slu7) steps reduces the level of H3K4me3, which is consistent with previous findings in mammalian cells of a reduction in H3K4me3 upon inhibition of splicing (Bieberstein *et al*., 2012).

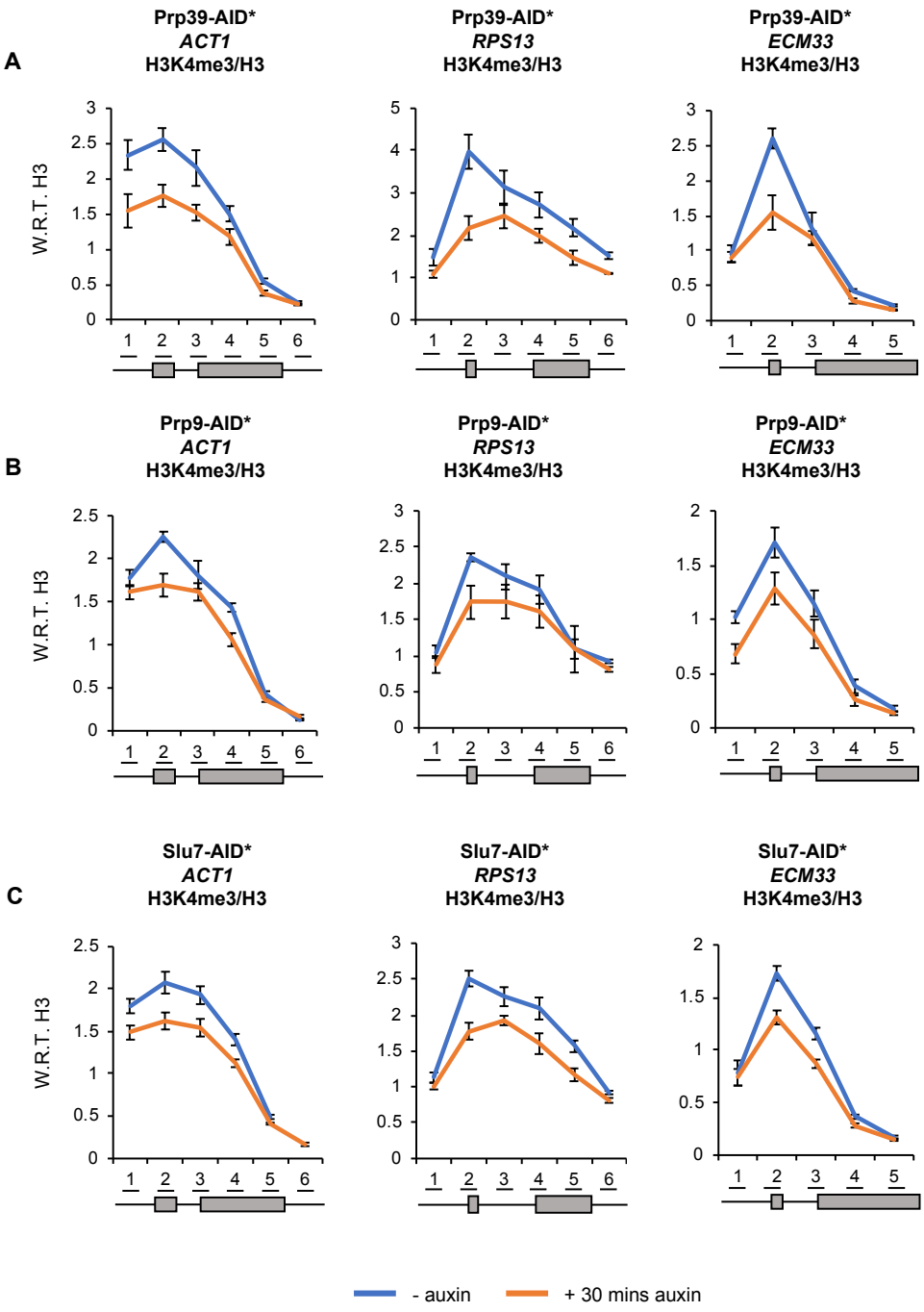


Figure 4.4. Depletion of essential splicing factors that affect the first, or first and second, step of splicing reduces the level of H3K4me3

Anti-H3K4me3 ChIP followed by qPCR analysis of the intron-containing genes *ACT1, RPS13* and *ECM33* without (-) auxin (blue) or with (+) 30 minutes of auxin (orange) addition to deplete (A) Prp39-AID*, (B) Prp9-AID* and (C) Slu7-AID*. The X-axis shows the positions of amplicons used for ChIP qPCR analysis – the exons are in grey. The data is presented as the mean percentage of input with respect to (W.R.T.) total histone H3. Mean of at least three biological replicates. Error bars = standard error of the mean.

4.4 Prp22 is the latest-acting factor whose depletion causes a reduction in H3K4me3 without causing defects in splicing catalysis

It was predicted from previous studies in mammalian cells that depletion of splicing factors that affect the first or second step of splicing would result in loss of H3K4me3. We next tested whether depletion of late-acting Prp22 for 30 minutes, which did not cause defects in splicing catalysis, reduced H3K4me3 levels. ChIP was performed using antibodies against H3K4me3 and histone H3 followed by qPCR across intron-containing genes *ACT1*, *RPS13* and *ECM33*. Depletion of Prp22 on average significantly reduced levels of H3K4me3 to 60% and 66% at amplicons 1 and 2 (5' and exon 1) of *ECM33* (student's t-test, $P<0.05$), 65%, 64% and 70% at amplicons 1-3 (5'-3'SS) of *ACT1* (student's t-test, $P<0.05$), and to 69% and 73% at amplicons 2 and 3 (exon 1 and intron) of *RPS13* (student's t-test, $P<0.05$), relative to conditions without depletion (Figure 4.5A).

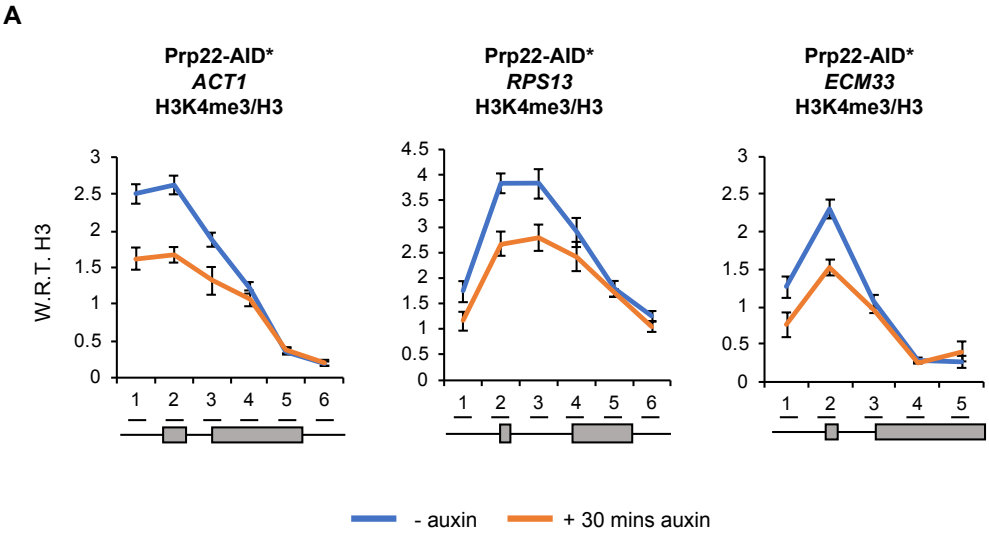


Figure 4.5. Prp22 is the latest-acting factor whose depletion causes a reduction in H3K4me3 without causing defects in splicing catalysis

(A) Anti-H3K4me3 ChIP followed by qPCR analysis of the intron-containing genes *ACT1, RPS13* and *ECM33* without (-) auxin (blue) or with (+) 30 minutes of auxin (orange) addition to deplete Prp22-AID*. The X-axis shows the positions of amplicons used for ChIP qPCR analysis – the exons are in grey. The data is presented as the mean percentage of input with respect to (W.R.T.) total histone H3. Mean of at least three biological replicates. Error bars = standard error of the mean.

The finding that Prp22 depletion caused a reduction in H3K4me3 without causing defects in splicing catalysis (after 30 minutes of depletion) was unexpected. To confirm this, a more thorough time course of Prp22 depletion was performed with 15, 30, 45, 60 and 120 minutes of auxin treatment. Western blotting showed that Prp22 was on average significantly depleted to 66%, 35%, 26%, 20% and 20% after 15, 30, 45, 60 and 120 minutes of auxin treatment, respectively (student's t-test, $P<0.05$) (Figure 4.5B). Splicing RT-qPCR showed that Prp22 depletion resulted in no significant accumulation of pre-mRNA species (3'SS and BP) (student's t-test, $P>0.05$) of *ACT1*, but a significant accumulation of lariat after 15 and 30 minutes of auxin treatment (student's t-test, $P<0.05$) (Figure 4.5C). After 45, 60 and 120 minutes of auxin treatment, pre-mRNA (3'SS and BP) and lariat species of *ACT1* significantly accumulated (student's t-test, $P<0.05$), indicating that there was a first and second step splicing defect at these later time points upon further Prp22 depletion (Figure 4.5C).

Further, 4-thiouracil labelling and purification of newly synthesised RNA was performed (Barrass *et al*., 2015) (data generated in collaboration with David Barrass, whom I gratefully acknowledge). Prp22 was depleted on average to 40%, 20% and 9% relative to wild-type levels, and each Prp22-depleted culture was subjected to 4-thiouracil labelling for 1, 2, 4, 6 and 8 minutes and newly synthesised RNA was purified, in addition to total RNA (Figure 4.5D). 4-thiouracil labelling agreed with earlier experiments – when there was on average 40% of Prp22 left, an accumulation of lariat was observed for *ACT1*, with little effect on pre-mRNA species (Figure 4.5E). When Prp22 was further depleted on average to 20% and 9%, the *ACT1* lariat further accumulated with the 3'SS also increasing – indicating a defect in the second step of splicing (Figure 4.5E). The 4-thiouracil labelling further shows that the primary consequence of Prp22 depletion is the accumulation of the excised intron-lariat, followed by a second step defect in pre-mRNA splicing upon further depletion.

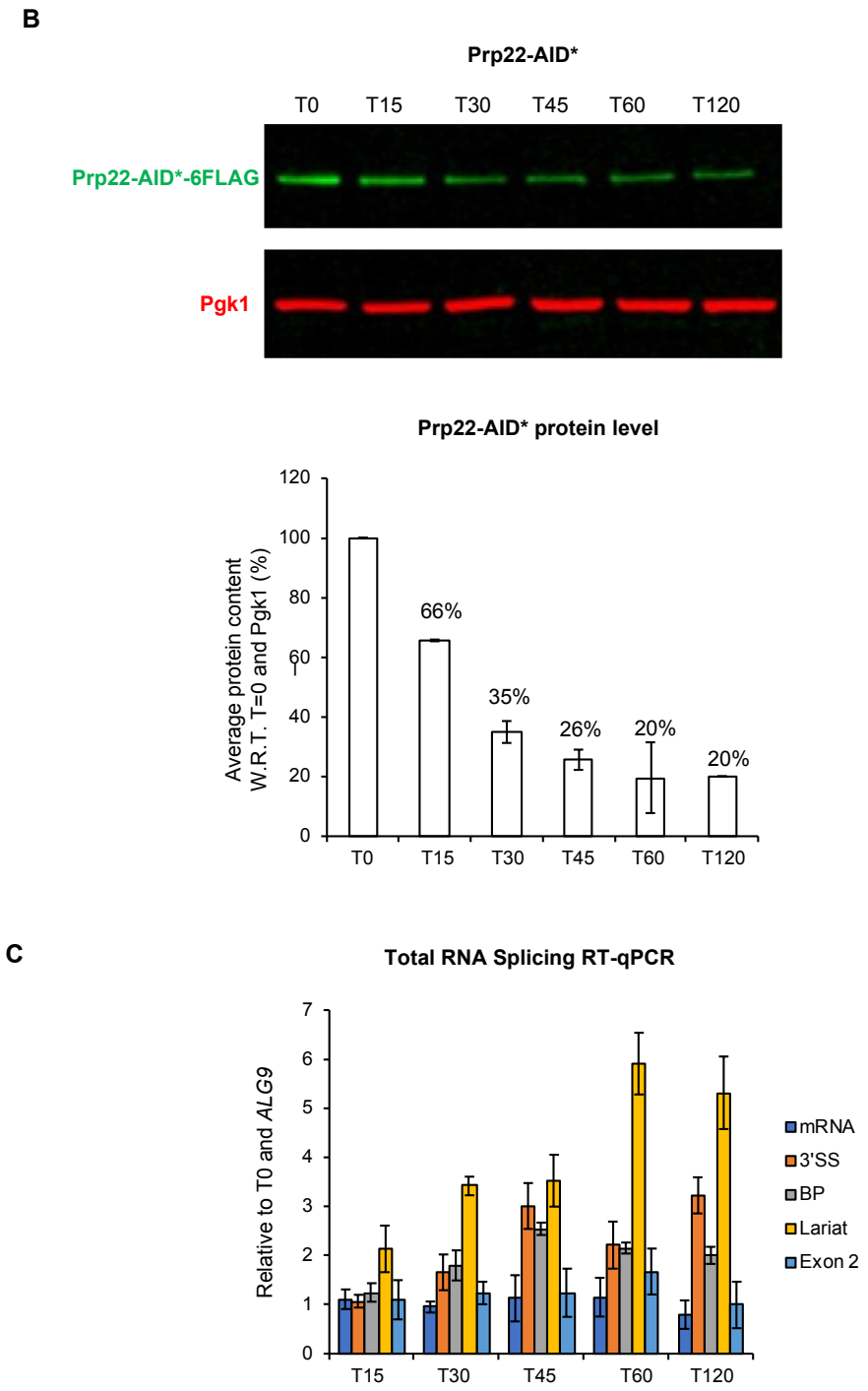


Figure 4.5 (continued). Prp22 is the latest-acting factor whose depletion causes a reduction in H3K4me3 without causing defects in splicing catalysis

(B) Western blot probed with anti-FLAG and anti-Pgk1 as a loading control. Samples were taken before (T0) and 15 (T15), 30 (T30), 45 (T45), 60 (T60) and 120 (T120) minutes after addition of auxin. Prp22-AID* depletion was quantified and shown as the percentage mean of 3 biological replicates after 30 minutes of auxin addition with respect to (W.R.T) time zero and normalised to the Pgk1 signal. Error bars = standard error of the mean.
(C) RT-qPCR analysis of total RNA from the intron-containing gene *ACT1* after 15, 30, 45, 60 and 120 minutes of depletion of Prp22-AID*. Normalised to *ALG9* transcript and time zero (without auxin addition). Mean of at least 3 biological replicates. Error bars = standard error of the mean.

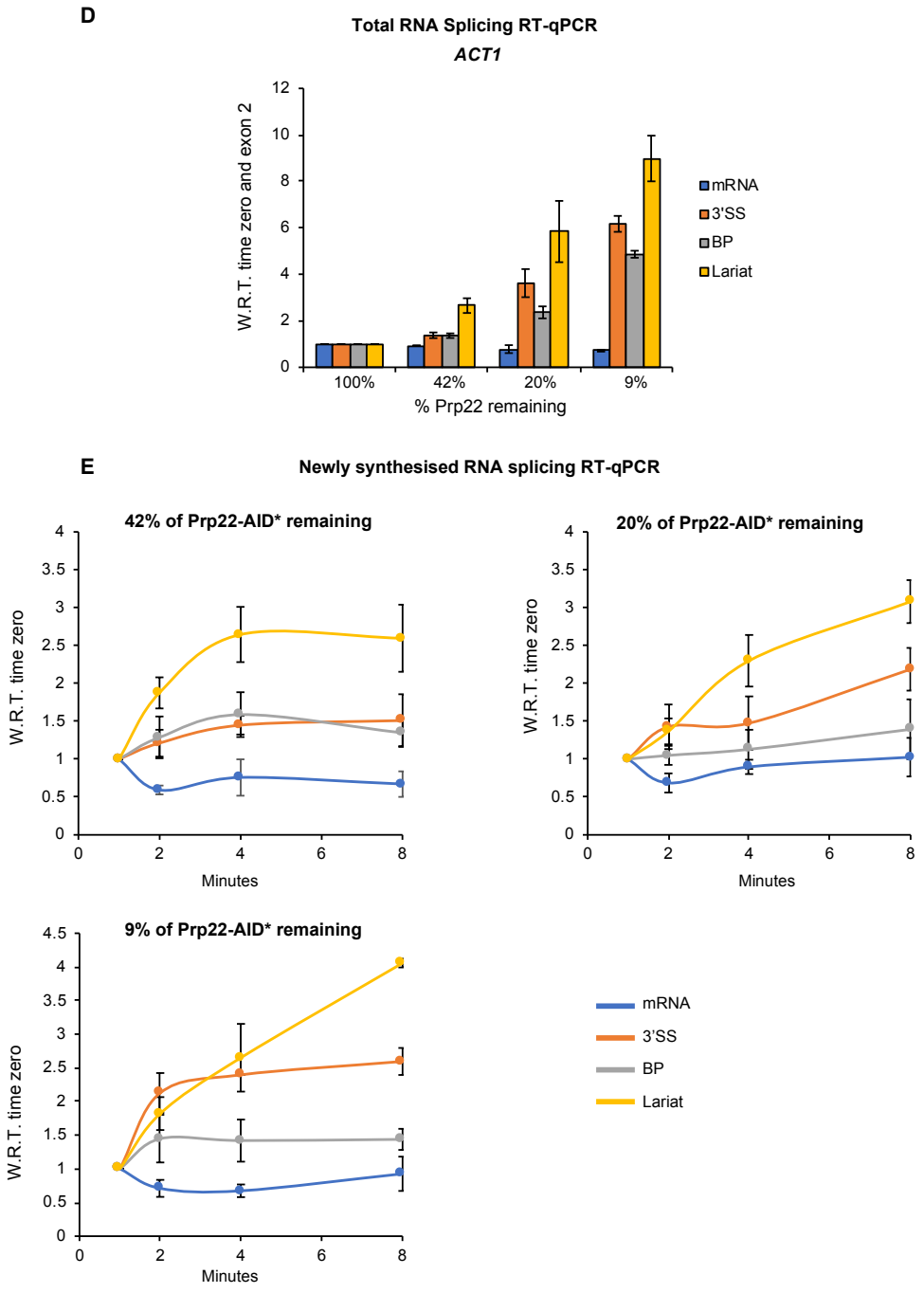


Figure 4.5 (continued). Prp22 is the latest-acting factor whose depletion causes a reduction in H3K4me3 without causing defects in splicing catalysis

(D) RT-qPCR analysis of total RNA from the intron-containing gene *ACT1* with on average 40%, 20% and 9% of Prp22 remaining. Normalised to exon 2 transcript and time zero (without auxin addition). Mean of at least 3 biological replicates. Error bars = standard error of the mean.
(E) RT-qPCR analysis of newly synthesised RNA 4-thiouracil labelled for 1, 2, 4, 6 and 8 minutes were analysed with on average 40%, 20% and 9% of Prp22 remaining. The X-axis in the lower panels shows the 4-thiouracil labelling time points (1, 2, 4, 6 and 8 minutes). All values are normalised to the 1 minute labelling value and time zero (without auxin addition). Mean of 2 biological replicates. Error bars = standard error of the mean

ChIP was performed using antibodies against H3K4me3 and histone H3 followed by qPCR across intron-containing genes *ACT1*, *RPS13* and *ECM33* before and after depletion of Prp22 for 30 minutes and 60 minutes (Figure 4.5F). Depletion of Prp22 on average significantly reduced level of H3K4me3 on *ACT1* at amplicons 1 to 3 (5'-3'SS) to 65%, 64% and 70% after 30 minutes of depletion, and 63%, 65% and 78% after 60 minutes of depletion (student's t-test, $P<0.05$), relative to conditions without depletion (Figure 4.5F). For *RPS13*, the level of H3K4me3 was significantly reduced at amplicons 2 and 3 (exon 1 and intron) to 69% and 73% after 30 minutes of depletion and on average to 61% across *RPS13* (all amplicons) after 60 minutes of depletion (student's t-test, $P<0.05$), relative to conditions without depletion (Figure 4.5F). For *ECM33*, the level of H3K4me3 was significantly reduced at amplicons 1 and 2 (5' and exon 1) to 60% and 66% after 30 minutes of depletion, and on average to 51% across *ECM33* (all amplicons) after 60 minutes of depletion (student's t-test, $P<0.05$), relative to conditions without depletion (Figure 4.5F). Therefore, no further Prp22 depletion was observed for longer incubation with auxin.

Next, we asked whether other chromatin marks were affected by loss of Prp22 depletion in these conditions. One chromatin mark of interest was H3K36me3, which is enriched over exons of actively expressed genes (these data were generated in collaboration with Dr Emanuela Sani, whom I gratefully acknowledge). Following depletion of Prp22 for 30 minutes and 60 minutes, ChIP was performed using antibodies against H3K36me3 and histone H3, followed by qPCR across intron-containing genes *ACT1*, *RPS13* and *ECM33* (Figure 4.5G). The pattern of H3K36me3 was correct– an increase in this mark towards the 3' end of the genes tested. After 30 minutes of depletion of Prp22, no significant changes in H3K36me3 were detected (student's t-test, $P>0.05$), relative to conditions without depletion (Figure 4.5G). However, after 60 minutes of Prp22 depletion H3K36me3 was significantly reduced at positions along the genes tested. For *ACT1*, the level of H3K36me3 was significantly reduced at amplicon 5 (exon 2) to 58% after 60 minutes of depletion (student's t-test, $P<0.05$), relative to conditions without depletion (Figure 4.5G). For *RPS13*, the level of H3K36me3 was significantly reduced at amplicons 3 to 5 (intron-exon2) to 62%, 68% and 60% after 60 minutes of depletion (student's t-test, $P<0.05$),

relative to conditions without depletion (Figure 4.5G). For *ECM33*, the level of H3K36me3 was significantly reduced at amplicons 3 to 5 (intron-exon2) to 56%, 56% and 54% after 60 minutes of depletion (student's t-test, $P<0.05$), relative to conditions without depletion (Figure 4.5G).

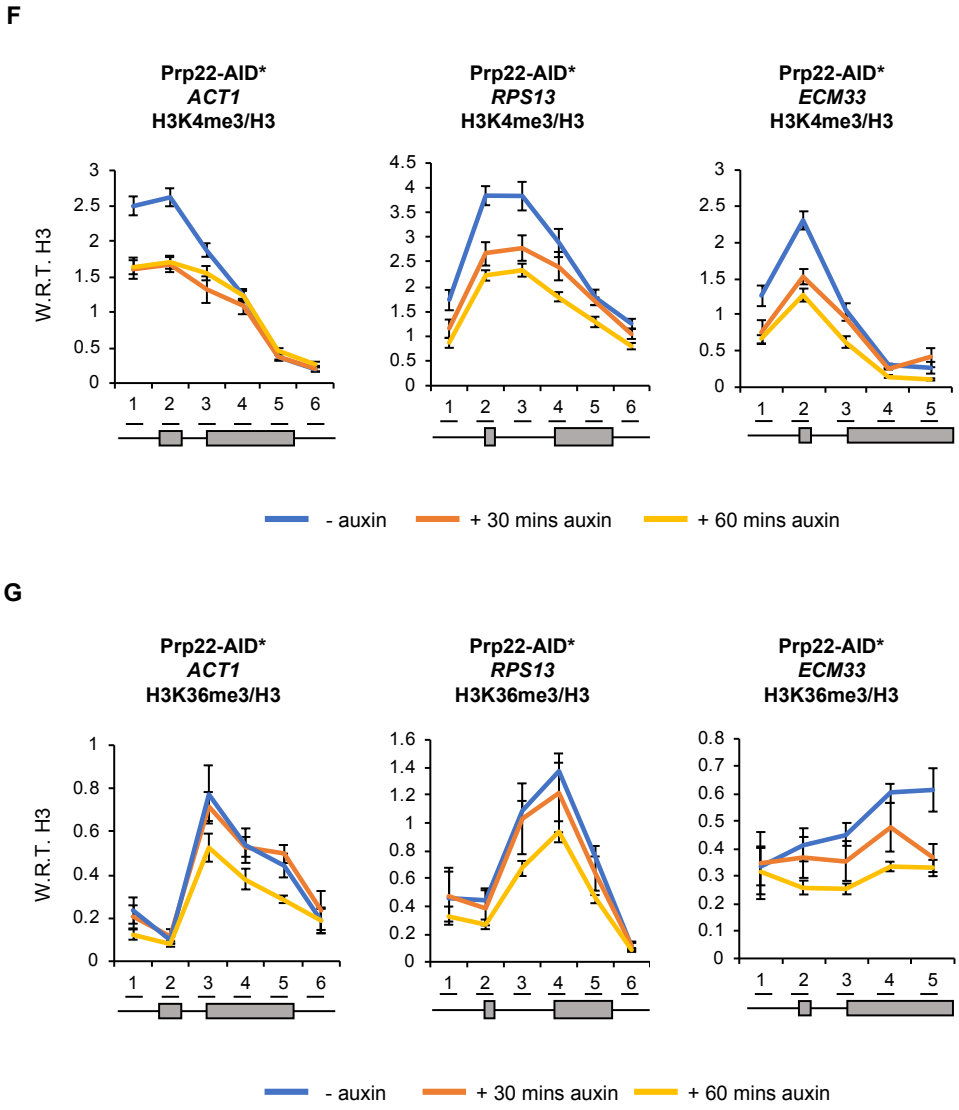


Figure 4.5 (continued). Prp22 is the latest-acting factor whose depletion causes a reduction in H3K4me3 without causing defects in splicing catalysis

(F) Anti-H3K4me3 and (G) Anti-H3K36me3 (from Dr. Emanuela Sani) ChIP followed by qPCR analysis of the intron-containing genes *ACT1, RPS13* and *ECM33* without (-) auxin (blue) or after 30 minutes (orange) and 60 minutes (yellow) of auxin addition to deplete Prp22-AID*. The X-axis shows the positions of amplicons used for ChIP qPCR analysis – the exons are in grey. The data is presented as the mean percentage of input with respect to (W.R.T.) total histone H3. Mean of at least three biological replicates. Error bars = standard error of the mean.

To test whether H3K4me1 and H3K4me2 were affected by Prp22 depletion, Prp22 was depleted for 30 minutes and 60 minutes, then ChIP was performed using antibodies against H3K4me2, H3K4me1 and histone H3 was performed followed by qPCR across intron-containing genes *ACT1*, *RPS13* and *ECM33* (Figures 4.5H and 4.5I). H3K4me2 peaks after H3K4me3, and H3K4me1 peaks after H3K4me2 on *ACT1*, *RPS13* and *ECM33* – indicating the expected methylation patterns. After 30 minutes of depletion of Prp22, H3K4me1 was significantly reduced on average to 35% at amplicon 4 (exon 2) of *ACT1*, 24% and 28% at amplicons 2 and 3 (exon 1 and 3'SS) of *ECM33*, respectively, and to 19% at amplicon 5 (exon 2) of *RPS13*, relative to conditions without depletion (student's t-test, $P<0.05$) (Figure 4.5H). After 60 minutes of depletion of Prp22, H3K4me1 was significantly reduced on average to 47% and 40% at amplicons 4 and 5 (exon 2) of *ACT1*, respectively, 46% across amplicons 2 to 5 (exon 1-exon2) of *ECM33*, respectively, and to 51% and 44% at amplicons 4 and 5 (exon 2) of *RPS13*, respectively, relative to conditions without depletion (student's t-test, $P<0.05$) (Figure 4.5H). After 30 minutes of depletion of Prp22, H3K4me2 did not significantly change across *ACT1*, *ECM33* or *RPS13* (student's t-test, $P>0.05$), but was significantly reduced after 60 minutes of depletion to 55% at amplicon 5 of ACT1 (exon 2), to 53% at amplicons 1 and 2 (5' and exon 1) of *ECM33*, and to 46% and 50% at amplicons 4 and 5 (exon 2) of *RPS13*, respectively, relative to conditions without depletion (student's t-test, $P<0.05$) (Figure 4.5I).

To test whether the effect of Prp22 depletion on H3K4me3 affected intronless genes, ChIP was performed using antibodies against H3K4me2 and histone H3 followed by qPCR across intronless genes *ALG9* and *FMP27* after depletion of Prp22 for 30 minutes and 60 minutes (Figure 4.5J). No significant difference in H3K4me3 levels was detected across the intronless genes detected after 30 and 60 minutes of Prp22 depletion (student's t-test, $P>0.05$) (Figure 4.5J).

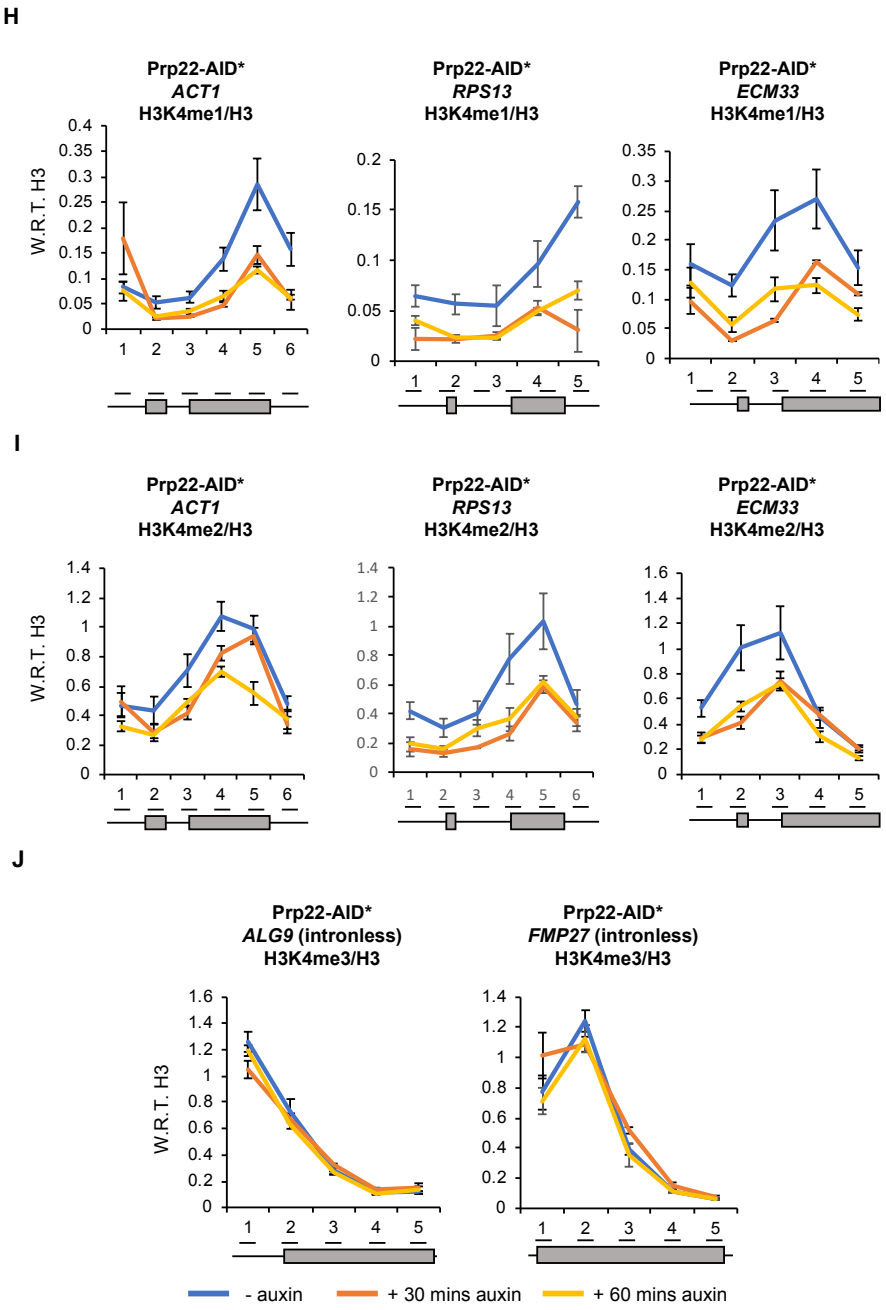


Figure 4.5 (continued). Prp22 is the latest-acting factor whose depletion causes a reduction in H3K4me3 without causing defects in splicing catalysis

(H) Anti-H3K4me1 and (I) Anti-H3K4me2 ChIP followed by qPCR analysis of the intron-containing genes *ACT1, RPS13* and *ECM33* without (-) auxin (blue) or after 30 minutes (orange) and 60 minutes (yellow) of auxin addition to deplete Prp22-AID*. The X-axis shows the positions of amplicons used for ChIP qPCR analysis – the exons are in grey. The data is presented as the mean percentage of input with respect to (W.R.T.) total histone H3. Mean of at least three biological replicates. Error bars = standard error of the mean.
(J) Anti-H3K4me3 ChIP followed by qPCR analysis of the intronless genes *ALG9* and *FMP27* without (-) auxin (blue) or after 30 minutes (orange) and 60 minutes (yellow) of auxin addition to deplete Prp22-AID*. The X-axis shows the positions of amplicons used for ChIP qPCR analysis – the exons are in grey. The data is presented as the mean percentage of input with respect to (W.R.T.) total histone H3. Mean of at least three biological replicates. Error bars = standard error of the mean.

In summary, the primary consequence of Prp22 depletion appears to be accumulation of the excised intron-lariat, followed by a second step defect in pre-mRNA splicing after further depletion. Depletion of Prp22 for 30 minutes resulted in a reduction in the level of H3K4me3 on three intron-containing genes tested, with no defects in splicing catalysis observed. Depletion of Prp22 for 60 minutes did not further reduce the level of H3K4me3 relative to 30 minutes of depletion. Depletion of Prp22 did not reduce the level of H3K4me3 on the intronless genes tested, meaning the effect may be specific to intron-containing genes. After 30 minutes of Prp22 depletion, H3K4me1 was reduced at some positions along the genes tested, and this was increased after 60 minutes of depletion of Prp22. In contrast, H3K4me2 was not significantly reduced after 30 minutes of Prp22 depletion but was significantly reduced after 60 minutes of depletion. After 30 minutes of Prp22 depletion, no significant reduction in H3K36me3 was detected – only after 60 minutes was there a significant reduction in this mark.

4.5 Depletion of Prp39, Prp9 and Slu7 reduces co-transcriptional recruitment of Prp22 to genes

To test whether depletion of Prp39, Prp9, or Slu7 reduces the recruitment of Prp22 to intron-containing genes, Prp22 was C-terminally 13MYC epitope tagged in each of the Prp39-AID*, Prp9-AID* and Slu7-AID* strains. The recruitment of Prp22 to the intron-containing genes after depletion of these essential splicing factors was tested by ChIP-qPCR across *ACT1*, *RPS13* and *ECM33*. Depletion of Prp39 significantly reduced Prp22 recruitment on average to 67% at amplicon 5 (exon 2) on *ACT1*, 67% at amplicon 5 (exon 2) on *RPS13* and 68% at amplicon 5 (exon 2) on *ECM33* (student's t-test, $P<0.05$) (Figure 4.6A). Depletion of Prp9 significantly reduced Prp22 recruitment on average to 44% at amplicon 5 (exon 2) on *ACT1*, 72% at amplicon 5 (exon 2) on *RPS13* and 75% at amplicon 5 (exon 2) on *ECM33* (student's t-test, $P<0.05$) (Figure 4.6B). Depletion of Prp9 significantly reduced Prp22 recruitment on average to 53% at amplicon 5 (exon 2) on *ACT1*, 46% at amplicon 5 (exon 2) on *RPS13* and 73% at amplicon 5 (exon 2) on *ECM33* (student's t-test, $P<0.05$) (Figure 4.6C).

As the spliceosome is known to assemble in a step-wise manner on pre-mRNA substrates, this finding that depletion of earlier acting splicing factors reduces co-transcriptional recruitment of Prp22 to genes was not unexpected. As Prp22 was the latest-acting factor analysed that caused a reduction in H3K4me3, and as depletion of earlier splicing factors reduced co-transcriptional recruitment of Prp22 to genes, we decided to focus our efforts on the links between H3K4me3 and Prp22.

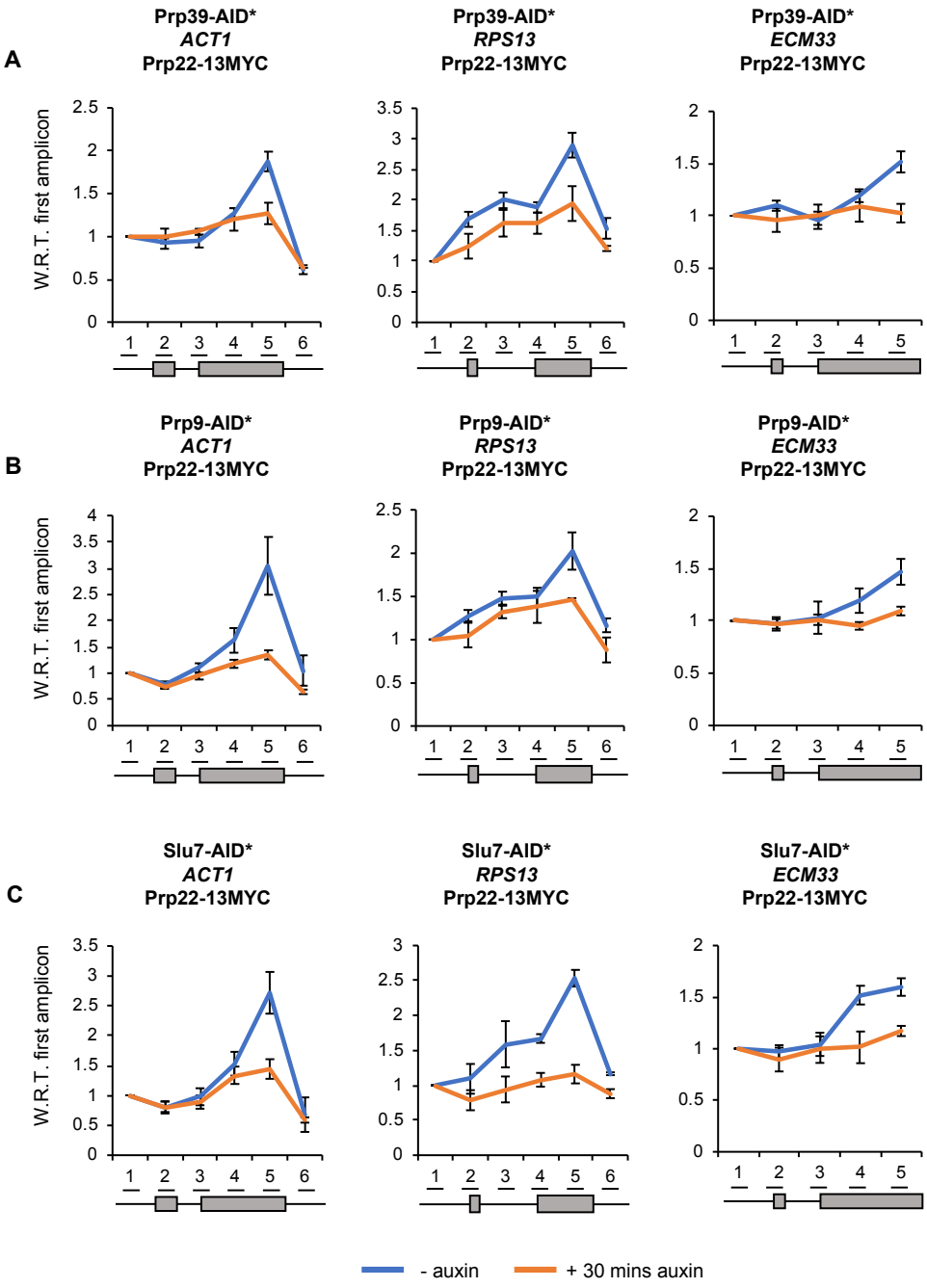


Figure 4.6. Depletion of Prp39, Prp9 and Slu7 reduces co-transcriptional recruitment of Prp22 to genes

Anti-MYC (Prp22-13MYC) ChIP followed by qPCR analysis of the intron-containing genes *ACT1, RPS13* and *ECM33* without (-) auxin (blue) or with (+) 30 minutes of auxin (orange) addition to deplete (A) Prp39-AID*, (B) Prp9-AID* and (C) Slu7-AID*. The X-axis shows the positions of amplicons used for ChIP qPCR analysis – the exons are of each gene in grey. The data is presented as the mean percentage of input with respect to (W.R.T.) the first amplicon of each gene. Mean of at least three biological replicates. Error bars = standard error of the mean.

4.6 Depletion of Prp22 reduces Set1 recruitment

In *S. cerevisiae*, a single histone methyltransferase, Set1, is required for H3K4me3. To test whether the reduced H3K4me3 upon Prp22 depletion was due to loss of Set1 recruitment, Set1 was N-terminally 9MYC epitope tagged in the Prp22-AID* strain (Dehé *et al*., 2006). Prp22 was depleted using the AID system, and recruitment of 9MYC tagged Set1 tested by ChIP-qPCR across the intron-containing genes *ACT1*, *RPS13* and *ECM33* after depletion of Prp22 for 30 minutes and 60 minutes.

Depletion of Prp22 significantly reduced Set1 recruitment across *ACT1* on average to 47% (amplicons 1-4; 5' – exon 2) and 58% (amplicons 1-5; 5'– exon 2) after depletion of Prp22 for 30 minutes and 60 minutes, respectively (student's t-test, $P<0.05$), relative to conditions without depletion (Figure 4.7A). Across *RPS13,* depletion of Prp22 significantly reduced Set1 recruitment on average to 47% (amplicons 1-6; 5'-3') and 51% (amplicons 2-5; exon 1 – exon 2) after depletion of Prp22 for 30 minutes and 60 minutes, respectively (student's t-test, $P<0.05$), relative to conditions without depletion (Figure 4.7A). Across *ECM33,* depletion of Prp22 significantly reduced Set1 recruitment on average to 47% (amplicons 1-4; 5'-exon 2) and 51% (amplicons 1-4; 5'-exon 2) after depletion of Prp22 for 30 minutes and 60 minutes, respectively (student's t-test, $P<0.05$), relative to conditions without depletion (Figure 4.7A).

As Set1 is recruited to transcribing RNAPII, it is possible that loss of Set1 recruitment upon depletion of Prp22 is caused by reduced RNAPII occupancy. To test whether Prp22 depletion resulted in loss of RNAPII occupancy, ChIP was performed using an antibody against RNAPII (Rpb1) followed by qPCR across *ACT1*, *RPS13* and *ECM33* depletion (Figure 4.7B). After depletion of Prp22 for 30 or 60 minutes, no significant change in RNAPII occupancy was observed across *ACT1 or ECM33* (student's t-test, $P>0.05$), relative to conditions without depletion, although after 60 minutes of auxin treatment RNAPII occupancy was reduced to 65% at amplicon 3 of *RPS13* (student's t-test, $P<0.05$) (Figure 4.7B).

More specifically, Set1 is recruited to RNAPII that is phosphorylated at serine 5 on the CTD, and therefore another possible explanation for loss of Set1 recruitment is by loss of phosphorylation of serine 5 in the CTD of RNAPII. ChIP was performed using an antibody against phosphorylated serine 5 (S5P) in the CTD followed by qPCR, as before. After 30 and 60 minutes of auxin treatment to deplete Prp22, no significant change in serine 5 phosphorylation was observed across *ACT1*, *RPS13* and *ECM33* (student's t-test, $P>0.05$), relative to conditions without depletion (Figure 4.7C).

The loss of Set1 occupancy upon Prp22 depletion could be explained by a reduction in the total cellular level of Set1. To test this, western blotting was performed with an antibody against the MYC epitope tag (to detect 9MYC-Set1) using extracts with and without 60 minutes of Prp22 depletion. The average cellular level of Set1 after depletion of Prp22 was 87% relative to conditions without depletion, but this difference was not statistically significant (student's t-test, $P>0.05$) (Figure 4.7D).

In summary, depletion of Prp22 for 30 and 60 minutes reduced Set1 recruitment at three intron-containing genes tested. This reduction in Set1 recruitment is apparently not due to reduced occupancy of RNAPII or of CTD S5P on these genes. Furthermore, the loss of Set1 recruitment was not due to a reduction in the total cellular level of Set1.

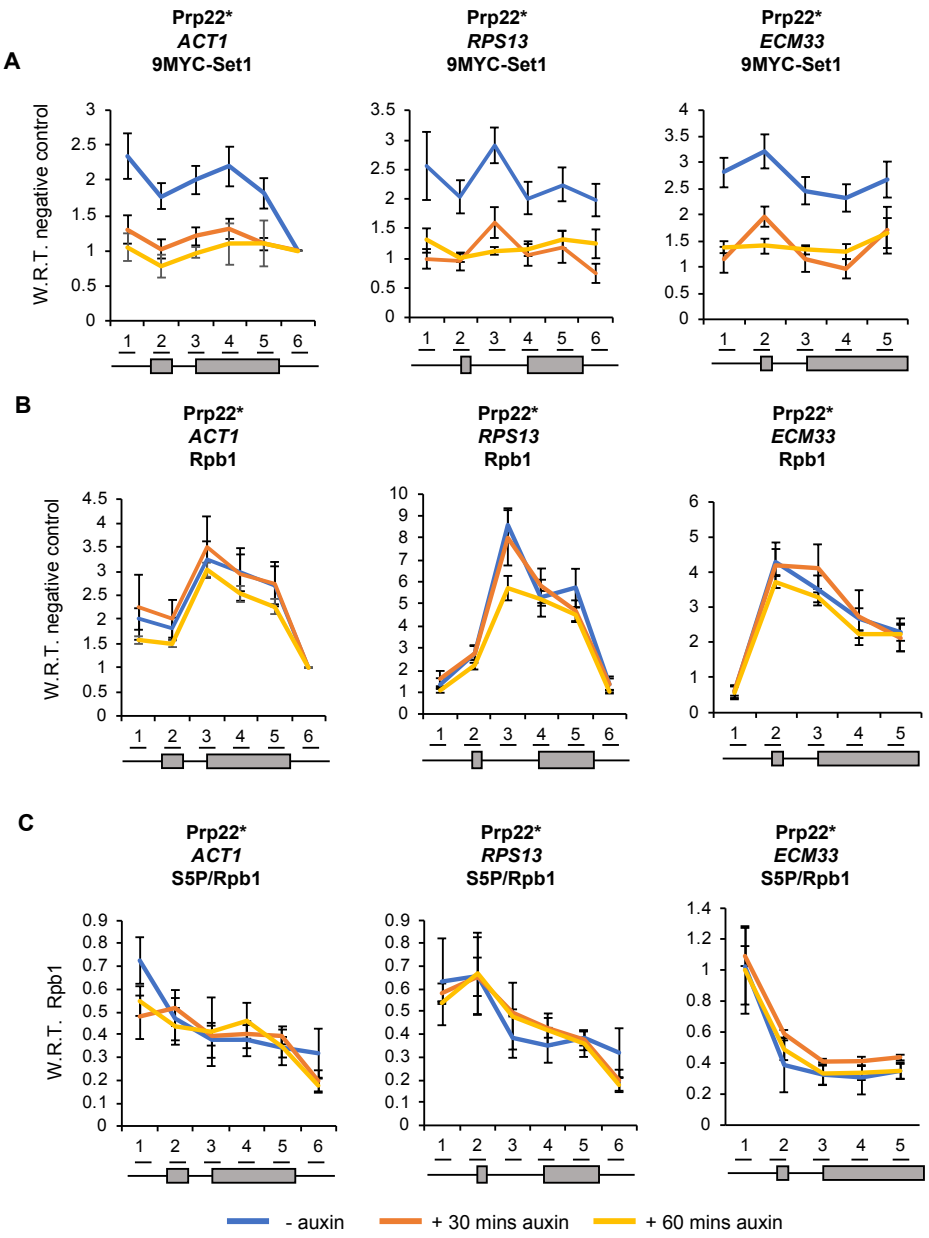


Figure 4.7. Depletion of Prp22 reduces Set1 recruitment to genes

(A) Anti-MYC (9MYC-Set1) ChIP followed by qPCR analysis of the intron-containing genes *ACT1, RPS13* and *ECM33* without (-) auxin (blue) or after 30 minutes (orange) and 60 minutes (yellow) of auxin addition to deplete Prp22-AID*. The data is presented as the mean percentage of input with respect to (W.R.T) a negative control amplicon.
(B) Anti-Rpb1 ChIP followed by qPCR analysis of the intron-containing genes *ACT1, RPS13* and *ECM33* without (-) auxin (blue) or after 30 minutes (orange) and 60 minutes (yellow) of auxin addition to deplete Prp22-AID*. The Rpb1 ChIP data is presented as the mean percentage of input W.R.T. a negative control region.
(C) Anti-S5P ChIP followed by qPCR analysis of the intron-containing genes *ACT1, RPS13* and *ECM33* without (-) auxin (blue) or after 30 minutes (orange) and 60 minutes (yellow) of auxin addition to deplete Prp22-AID*. The S5P ChIP data is presented as the mean percentage of input W.R.T. Rpb1.
ChIP data (A-C): The X-axis shows the positions of amplicons used for ChIP qPCR analysis – the exons are of each gene in grey. Mean of at least three biological replicates. Error bars = standard error of the mean.

D

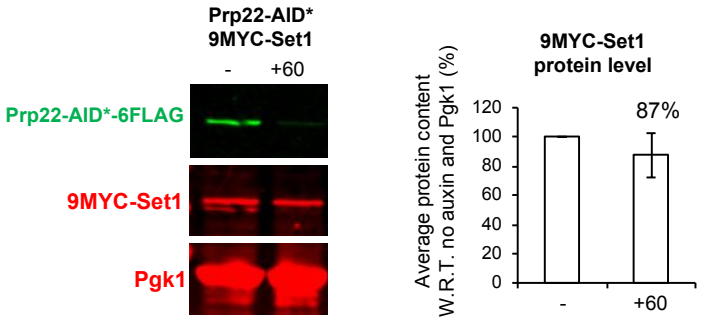


Figure 4.7 (continued). Depletion of Prp22 reduces Set1 recruitment to genes

(D) Western blot probed with anti-FLAG, anti-MYC and anti-Pgk1 as a loading control. Samples were taken without (-) and with 60 minutes (+60) of auxin treatment to deplete Prp22-AID*. 9MYC-Set1 was quantified and shown as the percentage mean of 3 biological replicates after 60 minutes of auxin addition with respect to (W.R.T) conditions with no auxin addition and normalised to the Pgk1 signal. Error bars = standard error of the mean.

4.7 Depletion of Prp22 reduces Spp1 recruitment

In *S. cerevisiae*, the Spp1 subunit of the Set1 complex is required specifically for H3K4me3 (not H3K4me1 or H3K4me2) by Set1. As Set1 and H3K4me3 were reduced upon Prp22 depletion, one would predict loss of Spp1 recruitment in these conditions, as Spp1 is in complex with Set1. To test whether Spp1 recruitment was reduced upon Prp22 depletion, Spp1 was C-terminally 13MYC epitope tagged in the Prp22-AID* strain, and recruitment of Spp1 was tested by ChIP-qPCR across the intron-containing genes *ACT1*, *RPS13* and *ECM33* after depletion of Prp22 for 60 minutes.

Spp1 recruitment was significantly reduced across *ACT1* on average to 23% (amplicons 1-5; 5' – exon 2) after depletion of Prp22 for 60 minutes (student's t-test, $P<0.05$), relative to conditions without depletion (Figure 4.8A). For *RPS13,* Spp1 recruitment was significantly reduced on average to 23% at amplicon 4 (exon 2) after depletion of Prp22 for 60 minutes (student's t-test, $P<0.05$), relative to conditions without depletion (Figure 4.8A). For *ECM33,* Spp1 recruitment was significantly reduced on average to 29% at amplicons 1 and 2 (5'-exon1) after depletion of Prp22

for 60 minutes (student's t-test, $P<0.05$), relative to conditions without depletion (Figure 4.8A).

The loss of Spp1 occupancy upon Prp22 depletion could be explained by a reduction in the total cellular level of the Spp1. To test this, western blotting was performed with an antibody against the MYC epitope tag (to detect Spp1-13MYC) using extracts with and without 60 minutes of Prp22 depletion. The average cellular level of Spp1 after depletion of Prp22 was 101% relative to conditions without depletion, and this difference was not statistically significant (student's t-test, $P>0.05$) (Figure 4.8B).

In summary, in addition to Set1 recruitment being reduced upon Prp22 depletion, Spp1 recruitment was reduced at three intron-containing genes tested.

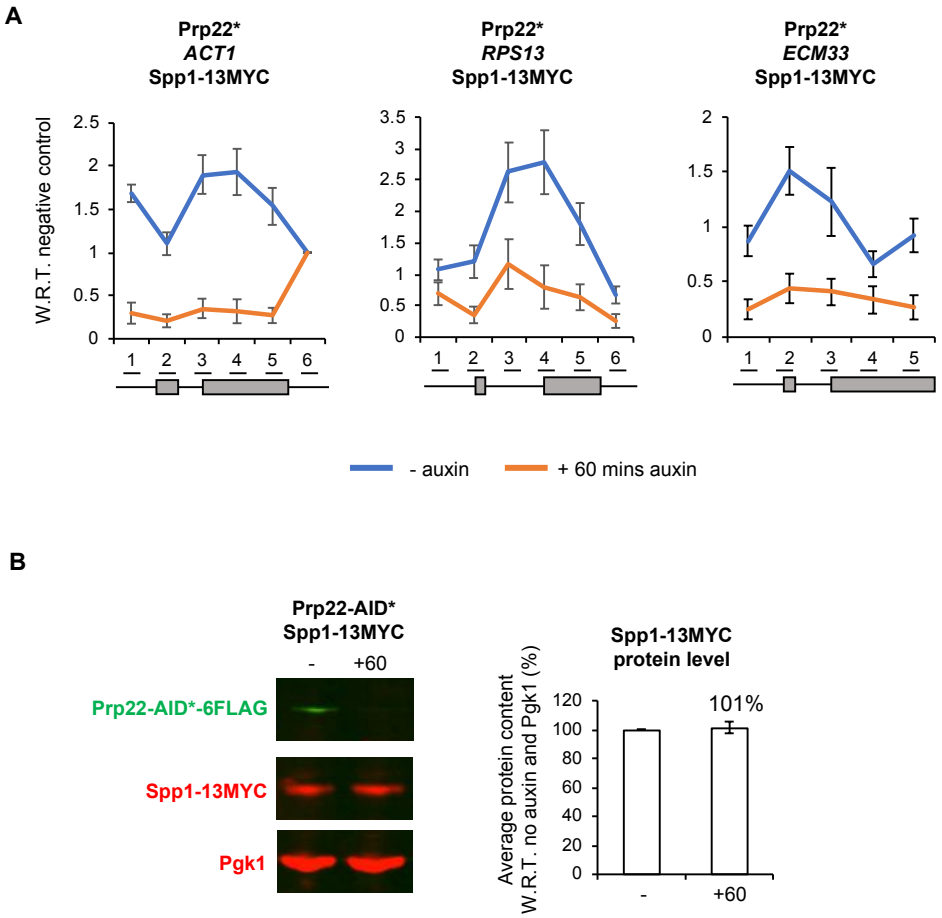


Figure 4.8. Depletion of Prp22 reduces Spp1 recruitment to genes

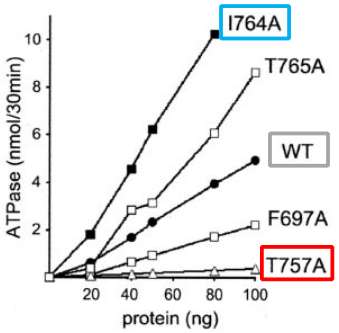
(A) Anti-MYC (Spp1-13MYC) ChIP followed by qPCR analysis of the intron-containing genes *ACT1, RPS13* and *ECM33* without (-) auxin (blue) or after 60 minutes (orange) of auxin addition to deplete Prp22-AID*. The X-axis shows the positions of amplicons used for ChIP qPCR analysis – the exons are of each gene in grey. The data is presented as the mean percentage of input with respect to the (W.R.T.) a negative control amplicon. Mean of at least three biological replicates. Error bars = standard error of the mean.
(B) Western blot probed with anti-FLAG, anti-MYC and anti-Pgk1 as a loading control. Samples were taken without (-) and with 60 minutes (+60) of auxin treatment to deplete Prp22-AID*. Spp1-13MYC was quantified and shown as the percentage mean of 3 biological replicates after 60 minutes of auxin addition with respect to (W.R.T) conditions with no auxin addition and normalised to the Pgk1 signal. Error bars = standard error of the mean.

4.8 Is the effect of Prp22 depletion on H3K4me3 dependent on its physical presence or ATPase activity?

The effect of Prp22 on H3K4me3 could be due to the loss of the physical presence of Prp22, or loss of its ATPase activity. To distinguish between these two possibilities, Prp22 ATPase mutants were used that were previously characterised *in vitro* (kind gift from Prof. Beate Schwer (Schneider *et al*., 2004)). As a DEX(H/D) box protein, Prp22 has seven conserved motifs (I-VII) that are important for ATP hydrolysis, coupling ATP hydrolysis to RNA unwinding and catalysing mRNA release. One mutant was selected that exhibited underactive ATPase activity (T757A) and the other selected exhibited overactive ATPase activity (I764A) *in vitro* (Figure 4.9A) (Schneider *et al*., 2004). Both mutations were in motif V of Prp22, and both mutants were able to bind RNA *in vitro*, however both had lost the ability to unwind RNA/DNA duplexes (Schneider *et al*., 2004). An ATPase assay showed that T757A had 8% of WT Prp22 ATPase activity, and I764A had 259% of WT Prp22 ATPase activity (Schneider *et al*., 2004) (Figure 4.9A). Whilst the ATPase activity of I764A was stimulated by the addition of an RNA co-factor, the ATPase activity of T757A was not. These mutants did not inhibit the second step of splicing but failed to release the mRNA from the spliceosome after splicing catalysis. Both mutants are lethal - T757A is lethal because the ATP hydrolysis of Prp22 is essential to release the mRNA, whereas I764A is lethal because ATP hydrolysis and RNA unwinding are uncoupled, and the mRNA cannot be released (Schneider *et al*., 2004).

As both mutants are lethal, full-length *PRP22* (WT), T757A or I764A were cloned into a centromeric plasmid that allowed conditional expression from a β-estradiol-inducible promoter (McIsaac *et al*., 2014; Mendoza-Ochoa *et al*., 2018). Additionally, each WT/mutant Prp22 was C-terminally V5 epitope tagged to allow immunodetection. Each plasmid with WT/mutant Prp22-V5 was transformed into the Prp22-AID* strain, so that each WT/mutant Prp22 could be expressed at the same time as depletion of endogenous Prp22-AID* (Figure 4.9B).

A



Prp22	ATPase (min^{-1}) -RNA	+RNA
WT	18	216
F697A	9	97
T757A	29	17
I764A	56	560
T765A	<1	358

(Schneider *et al*., 2004)

B

i) Endogenous: Prp22-AID*-6FLAG
Plasmid: Prp22-WT-V5

ii) Endogenous: Prp22-AID*-6FLAG
Plasmid: Prp22-T757A-V5

iii) Endogenous: Prp22-AID*-6FLAG
Plasmid: Prp22-I764A-V5

C

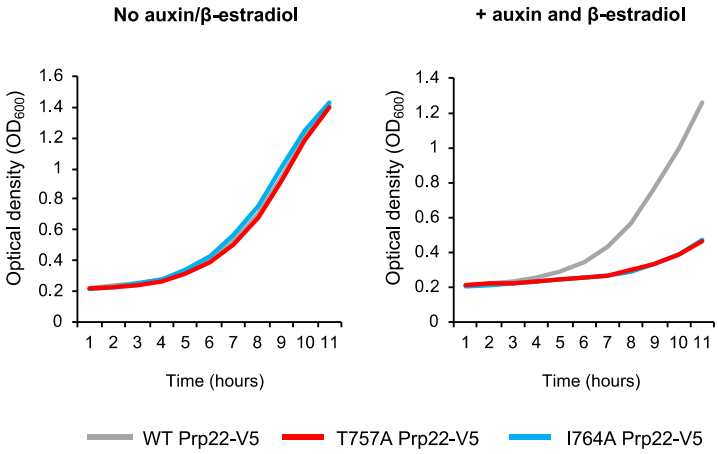


Figure 4.9. Is the effect of Prp22 depletion on H3K4me3 dependent on its physical presence or ATPase activity?

(A) ATPase activity of WT and mutant Prp22 in the absence (-RNA) and presence (+RNA) of poly(A) (+RNA). The underactive ATPase mutant of Prp22 used in the present work is highlighted in red, the overactive ATPase mutant of Prp22 highlighted in blue, and wild-type (WT) Prp22 highlighted in grey. Figure taken from Schneider *et al*., 2004.
(B) Schematic showing the strains used to determine effects of Prp22 ATPase activity on H3K4me3. The parental strain for each was the Prp22-AID* AID degron strain, which was transformed with either (i) WT Prp22-V5, (ii) T757A Prp22-V5 or (iii) I764A Prp22-V5 which were present on centromeric plasmids and inducible by β-estradiol addition.
(C) Growth of the Prp22-AID* strains containing either (i) WT Prp22-V5 (grey), (ii) T757A Prp22-V5 (red) or (iii) I764A Prp22-V5 (blue) was measured over a period of 11 hours without (-) auxin and β-estradiol or with (+) auxin and β-estradiol treatment. Mean of three biological replicates.

The effects of induction of WT or mutant Prp22-V5 at the same time as depleting endogenous Prp22-AID* was assessed by growth analysis (Figure 4.9C). In the absence of auxin and β-estradiol, the strains with the WT or mutant Prp22 plasmids grew comparably to each other (Figure 4.9C). With auxin and β-estradiol treatment, the growth of the strain with WT Prp22-V5 was unaffected, however the strains with T757A Prp22-V5 and I764A Prp22-V5 grew more slowly than the parental strain (Figure 4.9C). Therefore, induction of WT Prp22-V5 can prevent the growth defect normally found upon Prp22-AID* depletion, and the mutants cannot and, in fact, appear to make it worse.

Depletion and induction were checked by western blotting, splicing defects determined by splicing RT-qPCR, and effects on H3K4me3 tested by ChIP using antibodies against H3K4me3 and histone H3 followed by qPCR across intron-containing genes *ACT1*, *RPS13* and *ECM33*. Samples were taken: (i) without the addition of auxin (to deplete Prp22-AID*) and β-estradiol (to induce WT/mutant Prp22-V5), (ii) with the addition of auxin only for 60 minutes, and (iii) with the addition of β-estradiol for 30 minutes followed by auxin for 60 minutes. Each sample was taken at the same logarithmic growth stage. These depletion and induction timings were previously optimised (data not shown).

To confirm induction of the proteins by the β-estradiol system, western blotting was performed. Prp22-AID* was depleted on average to 14% and 7% after 90 minutes of auxin treatment and 90 minutes of auxin and β-estradiol treatment, respectively (student's t-test, $P<0.05$) (Figure 4.9D) (relative to conditions without auxin or β-estradiol addition). Wild-type (WT) Prp22-V5 was induced on average to 83% relative to the level of endogenous Prp22-AID* (Figure 4.9D). Underactive (T757A) Prp22-V5 was induced on average to 67% relative to the level of endogenous Prp22-AID* (Figure 4.9D). Overactive (I764A) Prp22-V5 was induced on average to 47% relative to the level of endogenous Prp22-AID* (Figure 4.9D). The difference in expression between induced WT Prp22-V5 and endogenous Prp22-AID* was not statistically significant (student's t-test, $P>0.05$), however the reduced expression of the mutants

T757A Prp22-V5 and I764A Prp22-V5 relative to Prp22-AID* was statistically significant (student's t-test, P<0.05).

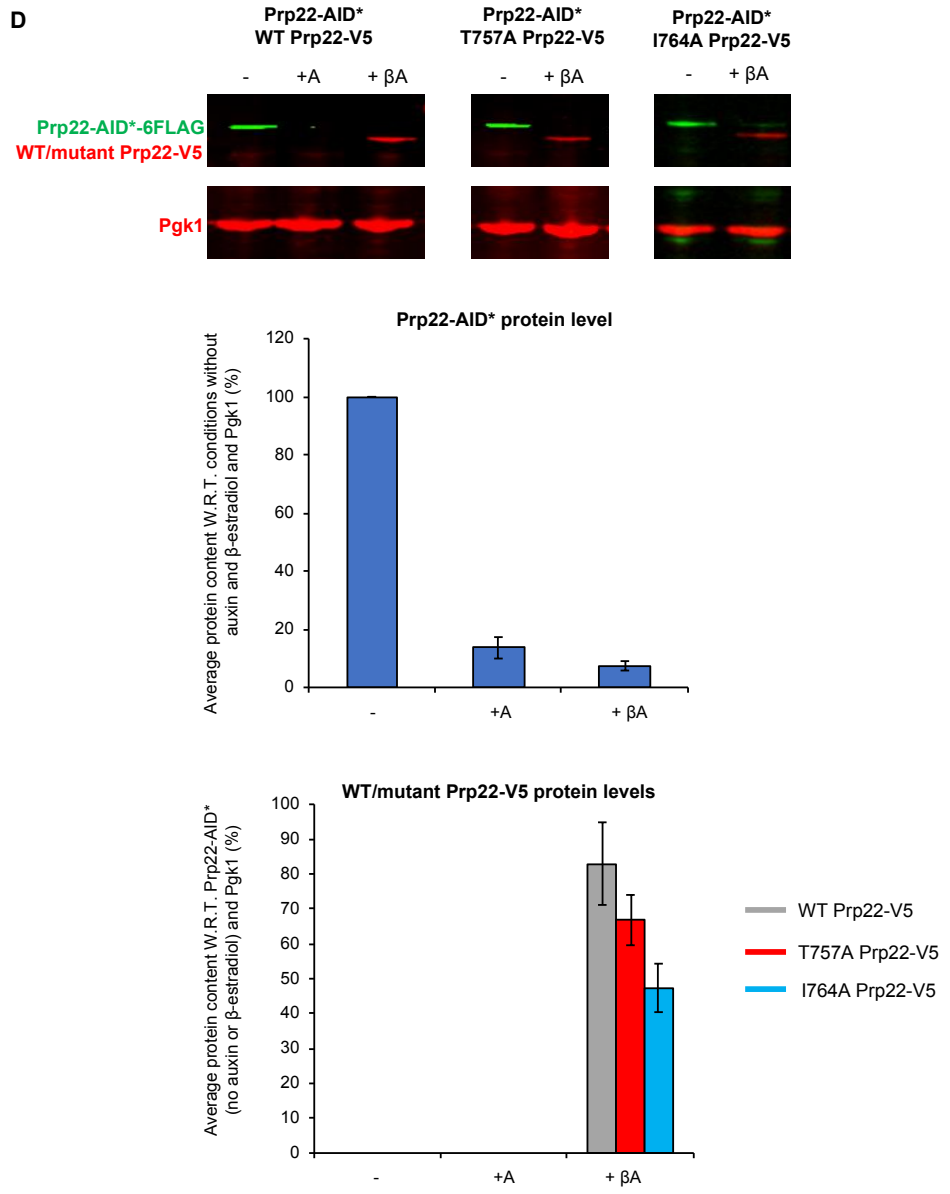


Figure 4.9 (continued). Is the effect of Prp22 depletion on H3K4me3 dependent on its physical presence or ATPase activity?

(D) Western blots probed with anti-FLAG, anti-V5 and anti-Pgk1 as a loading control. Samples were taken (i) without (-) auxin or β-estradiol, (ii) with 60 minutes of auxin treatment (+A), or (iii) with 30 minutes of β-estradiol treatment followed by 60 minutes of auxin treatment. Prp22-AID* depletion was quantified (upper graph, blue bars) and shown as the percentage mean of 3 biological replicates after auxin treatment with respect to (W.R.T) conditions without auxin and normalised to the Pgk1 signal. Induction of WT/mutant Prp22-V5 was quantified (lower graph, grey/red/blue bars) and shown as the percentage mean of 3 biological replicates after β-estradiol and auxin treatment with respect to (W.R.T) the level of endogenous Prp22-AID* prior to depletion. Error bars = standard error of the mean.

Prp22 is present at approximately 232 molecules/cell, therefore after 90 minutes of auxin treatment and 90 minutes of auxin and β-estradiol treatment, there were approximately 32 and 16 molecules/cell remaining, respectively (Kulak *et al*., 2014). WT Prp22-V5, T757A Prp22-V5 and I764A Prp22-V5 were present at approximately 193, 155 and 100 molecules/cell respectively. Longer inductions of WT/mutant Prp22-V5 were performed in an attempt to overexpress them, however this was apparently not successful within a three-hour time-frame (data not shown). At this stage, it is not clear why the mutant versions of Prp22 were less well induced in comparison to the WT version.

To test whether the chromatin-associated pool of Prp22-AID* was depleted, and whether each WT or mutant Prp22-V5 was recruited co-transcriptionally (to chromatin), ChIP was performed using antibodies against FLAG (Prp22-AID*) and V5 (WT/mutant Prp22-V5), followed by qPCR across the intron-containing genes *ACT1*, *RPS13* and *ECM33*. ChIP-qPCR showed that endogenous Prp22-AID* was significantly depleted from the genes tested after 60 minutes of auxin treatment (student's t-test, $P<0.05$) (Figure 4.9E). Additionally, ChIP-qPCR showed that each WT and mutant Prp22-V5 was recruited to chromatin upon induction with β-estradiol, with no significant difference observed between the WT and mutant proteins (student's t-test, $P>0.05$) (Figure 4.9F).

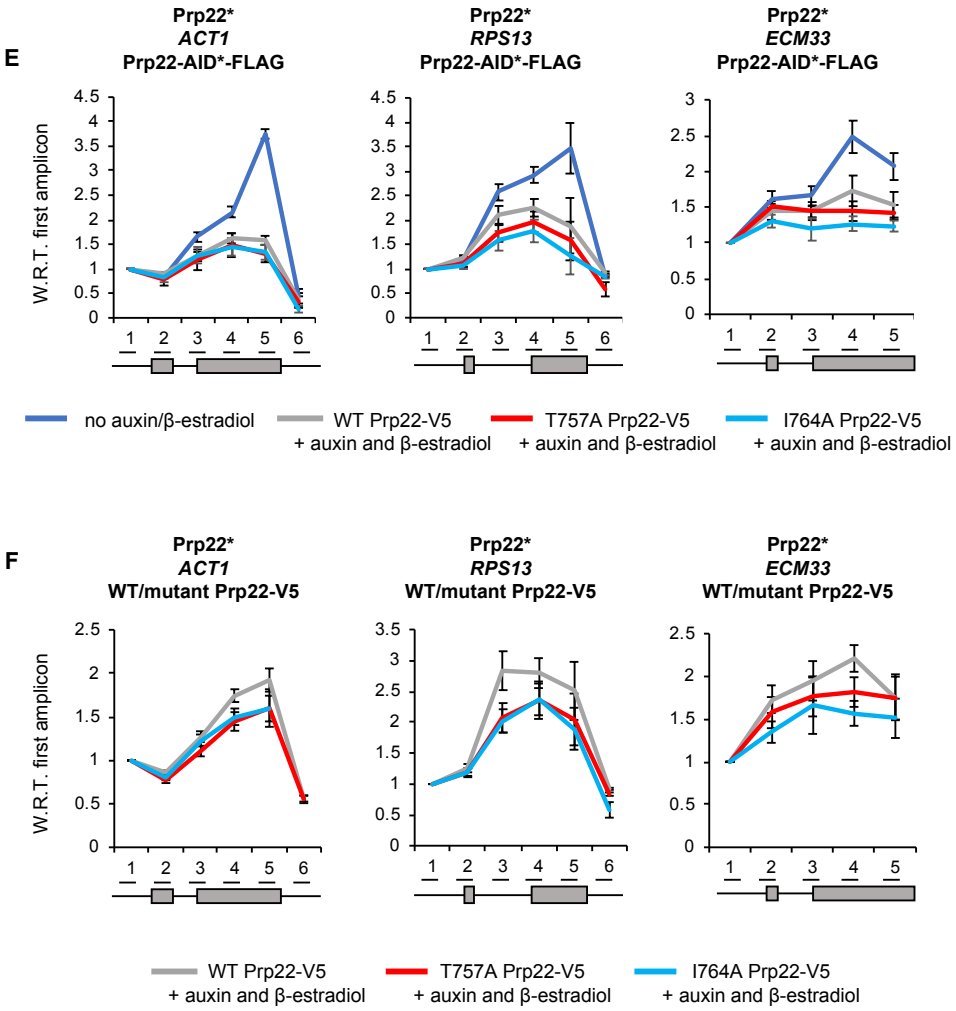


Figure 4.9 (continued). Is the effect of Prp22 depletion on H3K4me3 dependent on its physical presence or ATPase activity?

(E) Anti-FLAG ChIP (Prp22-AID*) followed by qPCR analysis of the intron-containing genes *ACT1*, *RPS13* and *ECM33* (i) without (-) auxin and β-estradiol (dark blue) and with auxin to deplete endogenous Prp22-AID* and β-estradiol treatment to induce WT (grey), T757A (red) and I764A (light blue). The X-axis shows the positions of amplicons used for ChIP qPCR analysis – the exons are of each gene in grey. The data is presented as the mean percentage of input normalised to the first amplicon. Mean of at least three biological replicates. Error bars = standard error of the mean.
(F) Anti-V5 (WT/mutant Prp22-V5) ChIP followed by qPCR analysis of the intron-containing genes *ACT1*, *RPS13* and *ECM33* with auxin to deplete endogenous Prp22-AID* and β-estradiol treatment to induce WT (grey), T757A (red) and I764A (light blue). The X-axis shows the positions of amplicons used for ChIP qPCR analysis – the exons are of each gene in grey. The data is presented as the mean percentage of input normalised to the first amplicon. Mean of at least three biological replicates. Error bars = standard error of the mean.

Next, a splicing RT-qPCR assay was used to determine whether depletion of Prp22-AID* or depletion of Prp22-AID* and induction of WT/mutant Prp22-V5 resulted in the expected splicing defects. As in previous experiments, depletion of Prp22-AID* for 60 minutes resulted in significant accumulation of pre-mRNA (3'SS and BP) and lariat species of *ACT1* (student's t-test, $P<0.05$), pre-mRNA (5'SS and 3'SS) of *RPS13* (student's t-test, $P<0.05$), and pre-mRNA (5'SS) of *ECM33* (student's t-test, $P<0.05$), indicating a likely first and second step splicing defect in these conditions (Figure 4.9G). As expected, induction of WT Prp22-V5 prior to depleting Prp22-AID* prevented any splicing defects normally observed upon Prp22-AID* depletion, with no significant changes to any pre-mRNA species observed for ACT1, *RPS13* and *ECM33* relative to conditions without depletion (Figure 4.9H, grey bars). Induction of T757A/I764A mutant Prp22-V5 at the same time as depleting Prp22-AID* resulted in significant accumulation of pre-mRNA (3'SS and BP) and lariat species of *ACT1* (student's t-test, $P<0.05$), pre-mRNA (5'SS and 3'SS) of *RPS13* (student's t-test, $P<0.05$), and pre-mRNA (5'SS) of *ECM33* (student's t-test, $P<0.05$), indicating a likely first and second step splicing defect in these conditions (Figure 4.9H, red and blue bars). The induction of T757A/I764A mutant Prp22-V5 resulted in significant reduction in the levels of mRNA and exon 2 levels of all genes tested (which is analogous to depletion of Prp39, Prp9 and Slu7). The induction of the mutants results in a more severe defect in splicing catalysis than depletion of Prp22-AID* alone in this time-frame, meaning that these mutants are probably dominant, which fits with previous findings of Gonzalo Mendoza-Ochoa (unpublished) (Beggs lab).

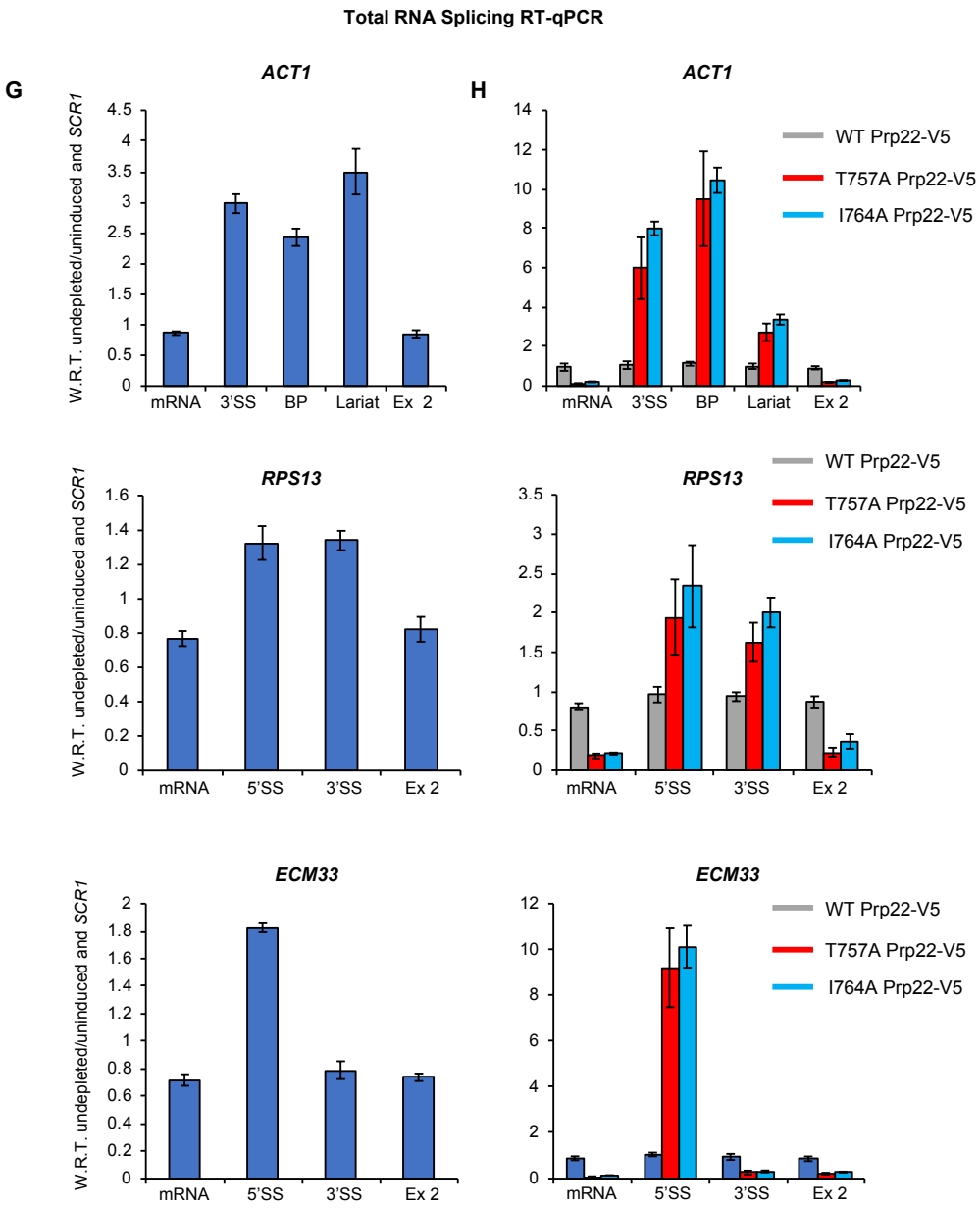


Figure 4.9 (continued). Is the effect of Prp22 depletion on H3K4me3 dependent on its physical presence or ATPase activity?

(G) RT-qPCR analysis of total RNA from the intron-containing genes *ACT1*, *RPS13* and *ECM33* after depletion of Prp22-AID* with 60 minutes of auxin treatment. Primers used detected pre-mRNA (3'SS or BP), lariat (excised or intron-exon 2), exon 2 (ex 2) and mRNA. Normalised to *SCR1* PolIII transcript and undepleted (without auxin addition). Mean of at least 3 biological replicates. Error bars = standard error of the mean.
(H) RT-qPCR analysis of total RNA from the intron-containing genes *ACT1*, *RPS13* and *ECM33* after induction of WT (grey), T757A (red) and I764A (blue) Prp22-V5 for 30 minutes of β-estradiol treatment followed by depletion of Prp22-AID* with 60 minutes of auxin treatment. Primers used detected pre-mRNA (3'SS or BP), lariat (excised or intron-exon 2), exon 2 (ex 2) and mRNA. Normalised to *SCR1* PolIII transcript and undepleted (without auxin addition). Mean of at least 3 biological replicates. Error bars = standard error of the mean.

In summary, depletion of Prp22-AID* by auxin addition and β-estradiol induction of WT or mutant Prp22-V5 worked well – western blotting showed the WT/mutant Prp22-V5 proteins were induced at the same time as endogenous Prp22-AID* was depleted. Additionally, all three WT/mutant Prp22-V5 proteins were recruited to chromatin and therefore acting co-transcriptionally. RT-qPCR showed expected splicing defects upon induction of the Prp22 mutants, and no splicing defects were observed upon induction of WT Prp22 – meaning induction of the WT protein could prevent splicing defects normally observed upon Prp22 depletion. Induction of T757A/I764A mutant Prp22 proteins resulted in a reduction in exon 2 and mRNA species, which was similarly observed upon depletion of Prp39, Prp9 and Slu7. As pre-mRNA species accumulate, it is not indicative of a transcription defect and is probably explained by degradation of spliced RNA.

4.9 An overactive ATPase mutant of Prp22 supports normal levels of H3K4me3 and Set1 recruitment

To test whether the effect of Prp22 depletion on H3K4me3 was dependent on its ATPase activity, the WT/mutant Prp22-V5 proteins were induced followed by depletion of endogenous Prp22-AID*. Each strain (i) WT (ii) T757A or (iii) I764A Prp22-V5 was induced by β-estradiol addition for 30 minutes, followed by the addition of auxin to deplete Prp22-AID* for 60 minutes. Effects on H3K4me3 were tested by ChIP using antibodies against H3K4me3 and histone H3 followed by qPCR across intron-containing genes *ACT1*, *RPS13* and *ECM33*.

Induction of WT Prp22-V5 prior to Prp22-AID* depletion prevented the loss in H3K4me3 previously observed upon Prp22-AID* depletion (Figure 4.5F), with no significant changes in H3K4me3 levels observed across the intron-containing genes tested, relative to conditions without auxin and β-estradiol addition (Figure 4.10A). Induction of the underactive ATPase mutant T757A Prp22-V5 prior to Prp22-AID* depletion did not prevent the significant reduction in H3K4me3 on average to 68% across amplicons 1-4 (5'-exon 2) of *ACT1*, 65% across amplicons 2-5 (exon 1-exon 2) of *RPS13* and 52% across amplicons 2 and 3 (exon 1 and 3'SS) of *ECM33* (student's t-test, $P<0.05$) (Figure 4.10A), relative to conditions without auxin and β-estradiol

addition. The induction of T757A Prp22-V5 at the same time as depletion of Prp22-AID* was comparable to depletion of Prp22-AID*. Significantly, induction of the overactive ATPase mutant I764A Prp22-V5 prior to Prp22-AID* depletion prevented the loss in H3K4me3 observed upon Prp22-AID* depletion – no significant changes in H3K4me3 were observed across the intron-containing genes tested, relative to conditions without auxin and β-estradiol addition (Figure 4.10A). This suggests that the ATPase activity of Prp22-AID* in the absence of RNA unwinding activity is sufficient to prevent the loss of H3K4me3 observed upon Prp22-AID* depletion.

To test whether induction of WT or I764A Prp22-V5 was also able to prevent the loss of 9MYC-Set1 observed upon Prp22 depletion (Figure 4.7A), ChIP of 9MYC-Set1 was performed followed by qPCR across the intron-containing genes *ACT1*, *RPS13* and *ECM33* (in the same experimental conditions described above). Induction of WT Prp22-V5 prevented the loss of Set1 recruitment previously observed upon Prp22 depletion – there was a significant increase in 9MYC-Set1 occupancy on average to 134% across all amplicons of *ACT1*, to 156%, 155% and 158% at amplicons 2, 4 and 5 (exon 1, 3'SS, exon 2) of RPS13, and to 131% and 151% at amplicons 4 and 5 (exon 2) of *ECM33* (student's t-test, $P<0.05$), relative to conditions without auxin and β-estradiol addition (Figure 4.10B). Induction of underactive T757A Prp22-V5 caused a significant decrease in 9MYC-Set1 occupancy on average to 48% across all amplicons of *ACT1*, to 34% across all amplicons of *RPS13* and to 49% across all amplicon of *ECM33* (student's t-test, $P<0.05$), relative to conditions without auxin and β-estradiol addition (Figure 4.10B). Induction of overactive I764A Prp22-V5 prior to depletion of Prp22-AID* prevented the loss of Set1 occupancy observed upon Prp22-AID* depletion – no significant changes in H3K4me3 were observed across the intron-containing genes tested, relative to conditions without auxin and β-estradiol addition (student's t-test, $P>0.05$) (Figure 4.10B). To test whether RNAPII occupancy was affected by induction of WT and mutants of Prp22-V5, ChIP was performed using an antibody against RNAPII (Rpb1) followed by qPCR across *ACT1*, *RPS13* and *ECM33* depletion (Figure 4.10C). No significant change in RNAPII occupancy was observed on induction of WT and mutants of Prp22-V5 (student's t-test, $P>0.05$) (Figure 4.10C).

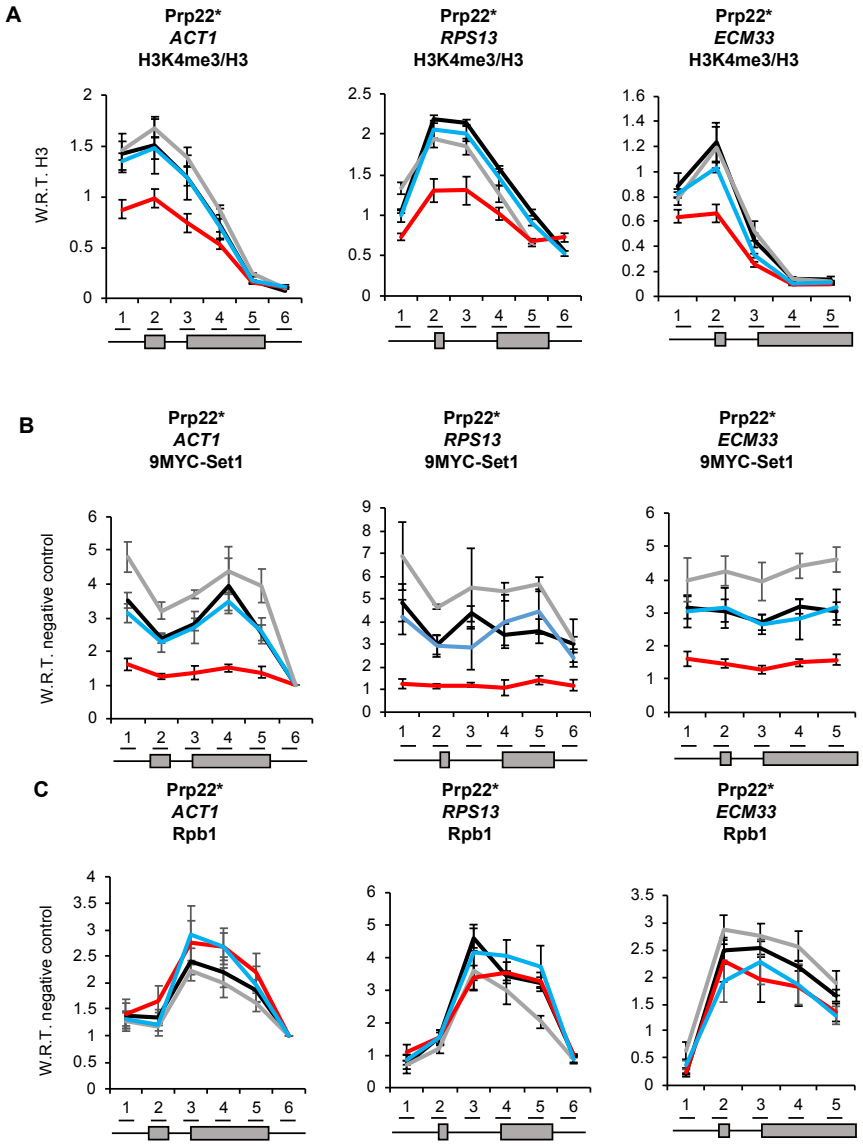


Figure 4.10. An overactive ATPase mutant of Prp22 supports normal levels of H3K4me3 and Set1 recruitment

(A) Anti-H3K4me3 ChIP followed by qPCR analysis of the intron-containing genes *ACT1, RPS13* and *ECM33* after induction of WT (grey), T757A (red) and I764A (blue) Prp22-V5 for 30 minutes of β-estradiol treatment followed by depletion of Prp22-AID* with 60 minutes of auxin treatment. Control conditions with no auxin and β-estradiol treatment is shown in black. The data is presented as the mean percentage of input with respect to (W.R.T.) total histone H3.
(B) Anti-MYC (9MYC-Set1) ChIP followed by qPCR analysis of the intron-containing genes *ACT1, RPS13* and *ECM33* after induction of WT (grey), T757A (red) and I764A (blue) Prp22-V5 for 30 minutes of β-estradiol treatment followed by depletion of Prp22-AID* with 60 minutes of auxin treatment. Control conditions with no auxin and β-estradiol treatment is shown in black. The data is presented as the mean percentage of input with respect to (W.R.T.) total histone H3.
(C) Anti-Rpb1 ChIP followed by qPCR analysis of the intron-containing genes *ACT1, RPS13* and *ECM33* after induction of WT (grey), T757A (red) and I764A (blue) Prp22-V5 for 30 minutes of β-estradiol treatment followed by depletion of Prp22-AID* with 60 minutes of auxin treatment. Control conditions with no auxin and β-estradiol treatment is shown in black. The data is presented as the mean percentage of input with respect to (W.R.T.) total histone H3.
ChIP data (A-C): The X-axis shows the positions of amplicons used for ChIP qPCR analysis – the exons are of each gene in grey. Mean of at least three biological replicates. Error bars = standard error of the mea

In summary, as predicted, induction of WT Prp22-V5 prior to depletion of Prp22-AID* prevents the reduction in Set1 recruitment and H3K4me3 previously observed upon Prp22-AID* depletion. Induction of an underactive Prp22 ATPase mutant, T757A Prp22-V5, did not prevent the reduction of Set1 and H3K4me3 observed by depletion of Prp22-AID*. Induction of the overactive Prp22 ATPase mutant, I764A Prp22-V5, prior to depletion of Prp22-AID* prevents the reduction in Set1 recruitment and H3K4me3 previously observed upon Prp22-AID* depletion, analogous to induction of WT Prp22-V5. The overactive ATPase mutant of Prp22 exhibits the same splicing defects as the underactive mutant, the difference between the two mutants being the ATPase activity. The overactive mutant of Prp22 is unable to couple ATP hydrolysis to RNA unwinding, and these data therefore suggest that the ATPase activity of Prp22 (normally coupled to RNA unwinding and splicing completion) may be required for proper Set1 recruitment and H3K4me3 on intron-containing genes.

4.10 Prp22 and Set1 interact

One way in which the loss of Prp22 might affect H3K4me3 and Set1 recruitment is by a physical interaction. Our collaborator, Prof. Vincent Géli (CRCM, Marseille, France) performed a yeast two-hybrid screen with full-length Set1, Set1 1-754 and Set1 754-1081 as baits and found a physical interaction between Prp22 and Set1 in all three screens (Figure 4.11A). The region of Prp22 common to all three baits is amino acids 508 to 762, which includes the ATP binding part of the helicase domain of Prp22 (Prof. Vincent Géli, personal communication).

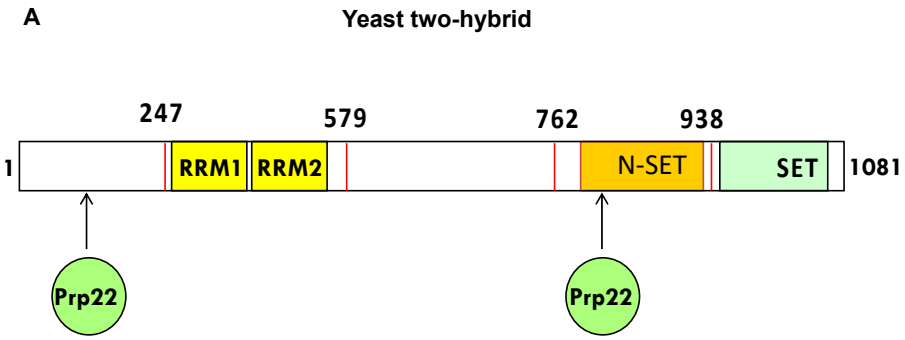


Figure 4.11. Prp22 and Set1 interact

(A) Diagram showing the functional domains of *S. cerevisiae* Set1. Two RNA recognition motifs (RRM) are shown toward the N-terminus, and the N-SET and SET domains at the C-terminus. The regions detected by yeast two-hybrid to interact with Prp22 are indicated by arrows: Set1 1-754 and Set1 754-1081. Set1 is 1081 amino acids long in *S. cerevisiae*. Data and diagram from Prof. Vincent Géli (CRCM, Marseille, France).

To confirm the physical interaction between Set1 and Prp22, co-immunoprecipitation experiments were performed in the Prp22-AID*-6FLAG in which Set1 had a N-terminal 9MYC epitope tag. Prp22-AID*-6FLAG was pulled down using a FLAG antibody, followed by Western blotting using an antibody against the MYC tag (Figures 4.11B and 4.11C). A reproducible interaction between Prp22 and Set1 was detected (Figures 4.11B and 4.11C). No pull down of Set1 was detected using a control strain with no FLAG-tagged Prp22, meaning the co-immunoprecipitation was specific. The interaction between Prp22 and Set1 was not RNA-dependent, as treatment with RNase did not abolish it (Figures 4.11B and 4.11C).

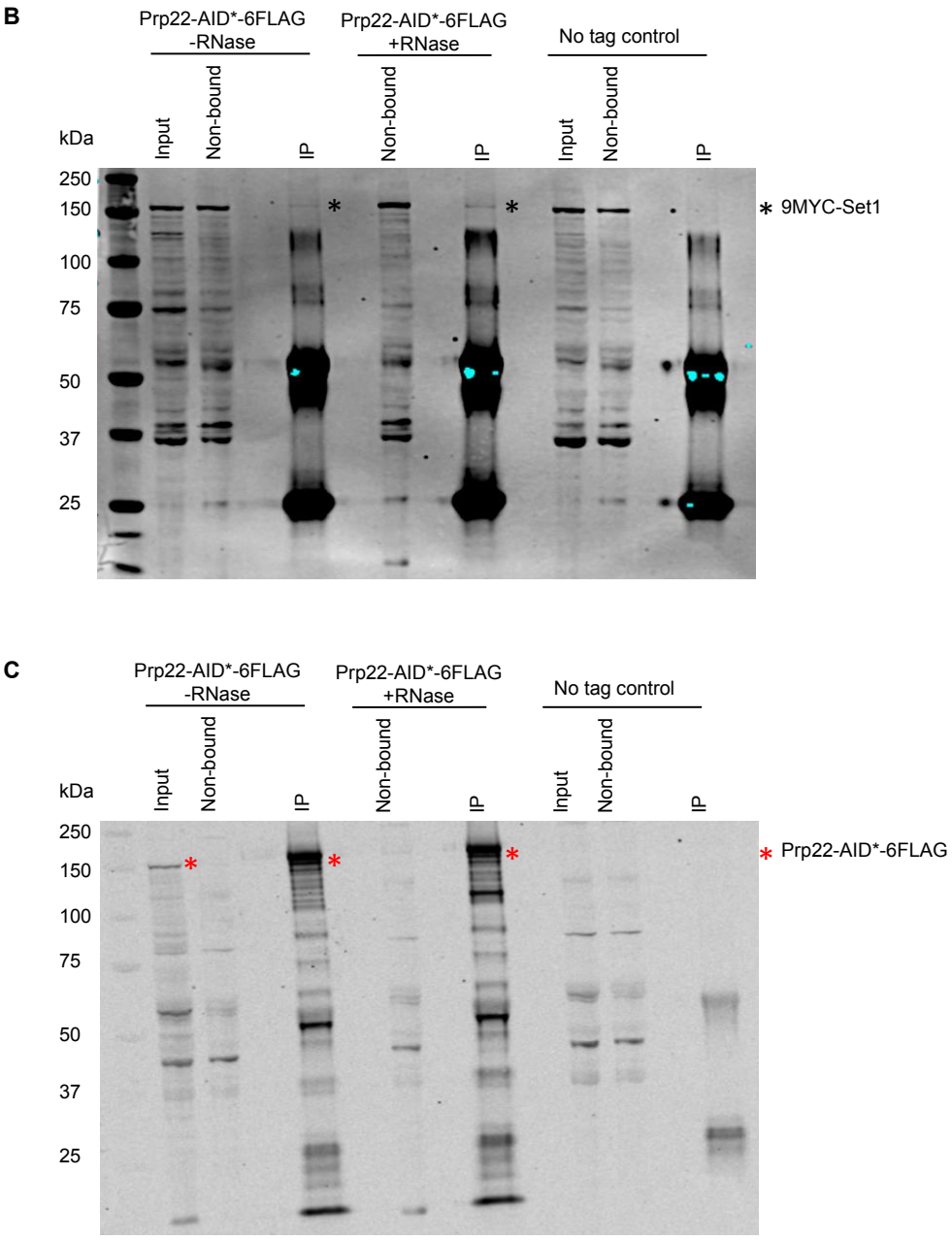


Figure 4.11 (continued). Prp22 and Set1 interact

(B) Western blots from a co-immunoprecipitation experiment in which Prp22-AID*-6FLAG was pulled down using a FLAG antibody with or without RNase treatment. Additionally, a strain with 9MYC-Set1 with no flag-tagged Prp22 was used as a negative control. Input, non-bound and immunoprecipitation (IP) samples were loaded. The blot was probed with an antibody against the MYC epitope tag (Set1).
9MYC-Set1 is indicated by the black asterisk.
(C) The blot in (B) was probed with a FLAG antibody to show that Prp22-AID*-6FLAG was successfully pulled-down in this assay. Prp22-AID*-6FLAG is indicated by the red asterisk.

As Set1 interacts with RNAPII, and splicing happens co-transcriptionally, the interaction observed between Prp22 and Set1 could be indirect, *via* RNAPII. Western blotting was performed using an antibody against RNAPII (Rpb1). In these conditions, Prp22-AID*-6FLAG did not pull down Rpb1 (Figure 4.11D).

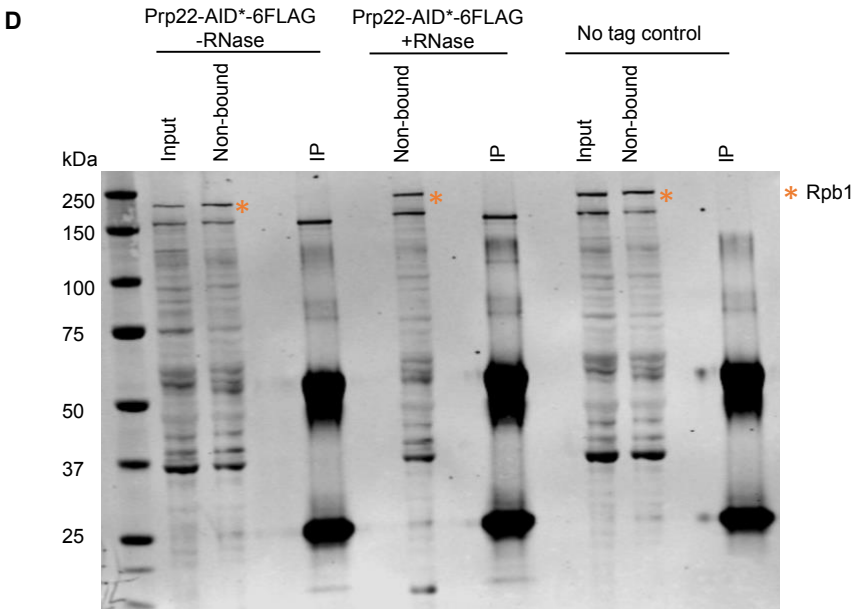


Figure 4.11 (continued). Prp22 and Set1 interact

(D) The blot in (B) was probed with an antibody against Rpb1 (RNAPII). Rpb1 is indicated by the orange asterisk.
Predicted molecular weights (kDa): Prp22-AID*-6FLAG = 143 kDa, 9MYC-Set1 = 140 kDa.

In summary, Set1 and Prp22 interact by yeast two-hybrid and co-immunoprecipitate in a manner stimulated by RNase addition. This interaction is not *via* RNAPII and gives a possible way by which Prp22 depletion could cause a reduction in Set1 recruitment to genes.

4.11 Use of the AID system to conditionally reduce the level of H3K4me3

An important question is what the biological significance of loss of H3K4me3 could be. One possibility was the level of H3K4me3 affects future rounds of co-transcriptional spliceosome assembly or pre-mRNA splicing, as has been reported in mammalian cells (Sims *et al*., 2007). To determine whether loss of H3K4me3 affects co-transcriptional spliceosome assembly or pre-mRNA splicing in *S. cerevisiae*, Spp1, which is a core member of the Set1 complex specifically required for H3K4me3, was C-terminally AID*-tagged in a strain that constitutively expresses Tir1 that enables conditional depletion of Spp1 in the presence of auxin (described in section 3.1).

Western blotting showed that depletion of Spp1 for 30 or 60 minutes resulted in a significant reduction, to 25% and 23% of Spp1 left relative to cells with no depletion ($P<0.05$) (Figure 4.12A). It is estimated that Spp1 is present at approximately 449 molecules/cell (Kulak *et al*., 2014), and therefore there are approximately 112 and 103 molecules/cell of Spp1 left after 30 and 60 minutes of depletion, respectively. Despite a significant reduction in Spp1 protein level, Western blotting showed that the level of H3K4me3 did not decrease accordingly, with on average 78% and 75% of H3K4me3 remaining after 30 and 60 minutes of depletion, respectively (Figure 4.12B).

To test whether there was an effect on H3K4me3 under these depletion conditions on a gene-by-gene basis, ChIP was performed using antibodies against H3K4me3 and histone H3 followed by qPCR across intron-containing genes *ACT1*, *RPS13* and *ECM33* after depletion of Spp1 for 30 minutes and 60 minutes (Figure 4.12C). Depletion of Spp1 for 30 minutes significantly reduced the level of H3K4me3 on *ACT1* on average at amplicons 2 to 5 (exon 1 to exon 2) to 78% after 30 minutes of depletion, relative to conditions without depletion (student's t-test, $P<0.05$) (Figure 4.12C). For *RPS13,* the level of H3K4me3 was significantly reduced on average at amplicons 4 to 6 (3'SS to 3') to 82%, 81% and 74%, respectively after 30 minutes of depletion, relative to conditions without depletion (student's t-test, $P<0.05$) (Figure 4.12C). For *ECM33*, the level of H3K4me3 was significantly reduced on average at amplicons 2, 3 and 4 (5', exon 1 and 3'SS) to 82%, 73% and 82% respectively after 30 minutes of depletion, relative to conditions without depletion (student's t-test,

P<0.05) (Figure 4.12C). Further depletion of Spp1 for 60 minutes significantly reduced the level of H3K4me3 on *ACT1* on average at amplicons 1 to 5 (5' to exon 2) to 67% after 60 minutes of depletion, relative to conditions without depletion (student's t-test, P<0.05) (Figure 4.12C). For *RPS13*, the level of H3K4me3 was significantly reduced at amplicons 2 to 6 (exon 1 to 3') on average to 65% after 60 minutes of depletion, relative to conditions without depletion (student's t-test, P<0.05) (Figure 4.12C). For *ECM33*, the level of H3K4me3 was significantly reduced on average across the gene to 62% after 60 minutes of depletion, relative to conditions without depletion (student's t-test, P<0.05) (Figure 4.12C).

Despite a significant reduction in Spp1 protein after 30 and 60 minutes of depletion, the reduction in H3K4me3 was modest. For the purposes of determining the effects of loss of H3K4me3 on co-transcriptional spliceosome assembly and splicing, a much greater loss of H3K4me3 was desired, so Spp1 was further depleted for 120 minutes (Figure 4.12D). Western blotting showed that depletion of Spp1 for 120 minutes resulted on average in a significant reduction to 12% of Spp1 left relative to cells with no depletion (student's t-test, P<0.05) (Figure 4.12D). Therefore, there were approximately 54 molecules/cell of Spp1 left after 120 minutes of depletion. Western blotting showed that the level of H3K4me3 significantly decreased to 22% after 120 minutes of Spp1 depletion (Figure 4.12E).

To test whether Spp1 was depleted from chromatin, ChIP was performed using a FLAG antibody followed by qPCR analyses across intron-containing genes *ACT1, RPS13* and *ECM33* (Figure 4.12F). ChIP-qPCR showed that, in addition to being depleted in whole cell extracts, each Spp1 was well depleted from chromatin after 120 minutes of auxin treatment (Figure 4.12F). Spp1 was significantly depleted to 62% across amplicons 2 to 5 (exon 1 to exon 2) of *ACT1,* 45% across amplicons 2 to 5 (exon 1 to exon 2) of *RPS13* and 56% across amplicons 2 to 4 (exon 1 to exon 2) of *ECM33* (student's t-test, P<0.05) (Figure 4.12F), relative to conditions without depletion.

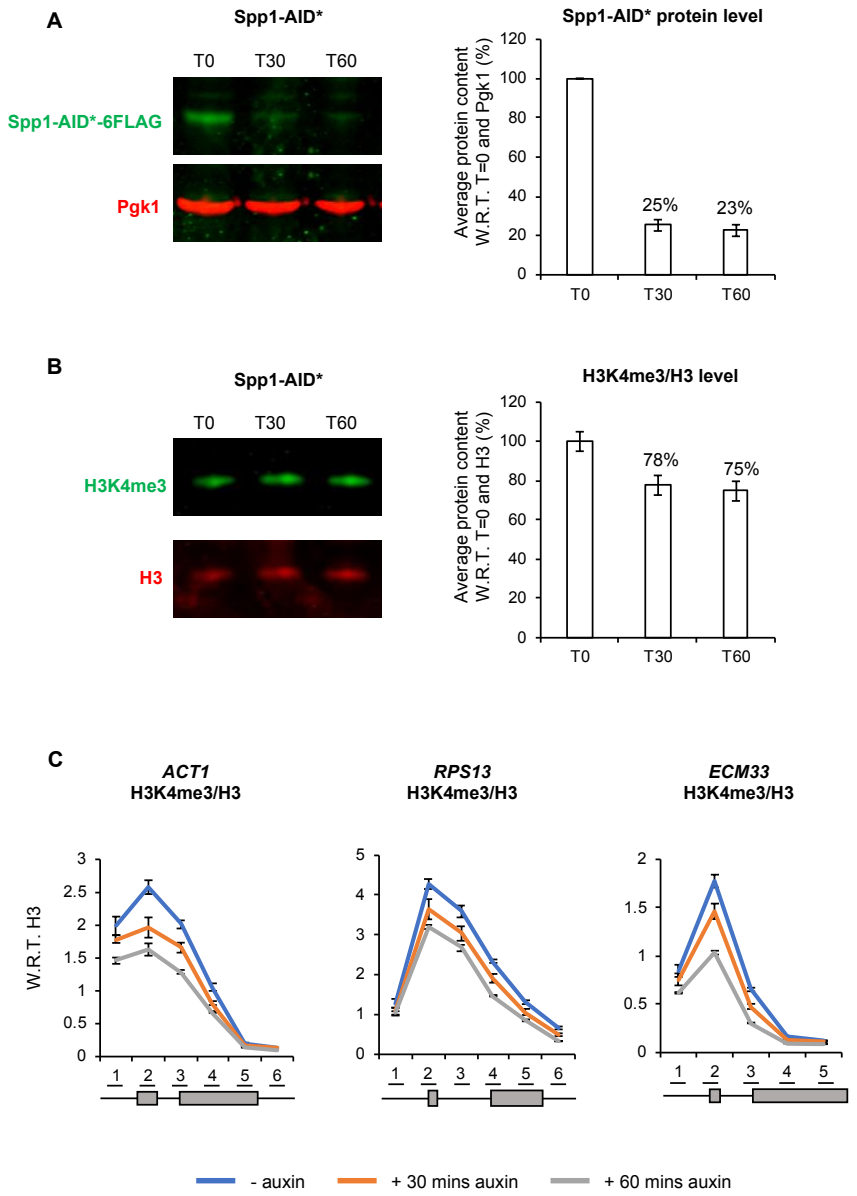


Figure 4.12. Use of the AID system to conditionally reduce the level of H3K4me3

A) Western blot probed with anti-FLAG and anti-Pgk1 as a loading control. Samples were taken before (T0), 30 minutes (T30) and 60 minutes (T60) after addition of auxin to deplete Spp1-AID*. Depletion was quantified and shown as the percentage mean of 3 biological replicates after 30 minutes of auxin addition with respect to (W.R.T) time zero and normalised to the Pgk1 signal. Error bars = standard error of the mean.
(B) Western blot probed with anti-H3K4me3 and anti-H3 as a loading control. Samples were taken before (T0), 30 minutes (T30) and 60 minutes (T60) after addition of auxin to deplete Spp1-AID*. H3K4me3 levels were normalised to H3 and shown as the percentage mean of 3 biological replicates after 30 or 60 minutes of auxin addition with respect to (W.R.T) time zero. Mean of three biological replicates. Error bars = standard error of the mean.
(C) Anti-H3K4me3 ChIP followed by qPCR analysis of the intron-containing genes *ACT1, RPS13* and *ECM33* before (- auxin, blue), 30 minutes (+ 30 mins auxin, orange) and 60 minutes (+ 60 mins auxin, grey) after addition of auxin to deplete Spp1-AID*. The X-axis shows the positions of amplicons used for ChIP qPCR analysis – the exons are of each gene in grey. The data is presented as the mean percentage of input normalised to total histone H3. Mean of at least three biological replicates. Error bars = standard error of the mean.

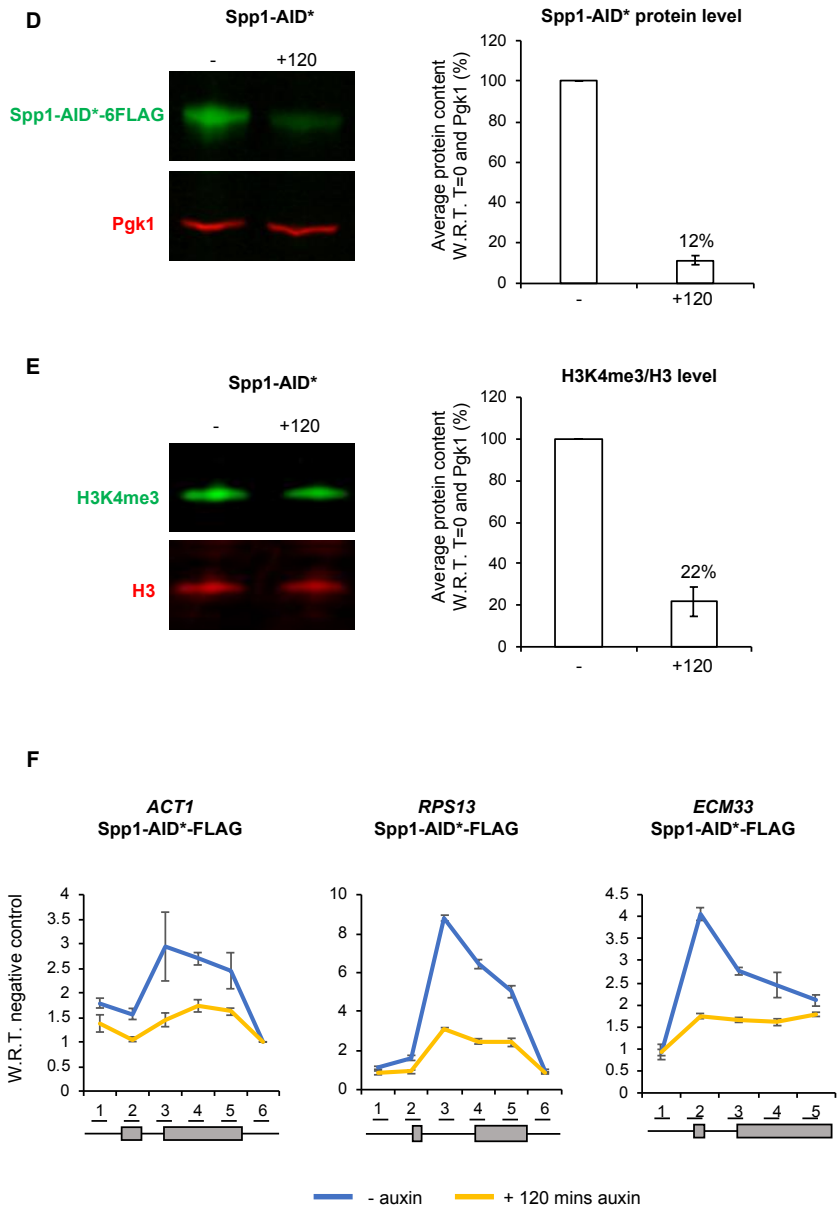


Figure 4.12 (continued). Use of the AID system to conditionally reduce the level of H3K4me3

D) Western blot as in (A) with samples taken without auxin (-) and with 120 minutes (120) of auxin treatment to deplete Spp1-AID*.
(E) Western blot as in (B) with samples taken without auxin (-) and with 120 minutes (120) of auxin treatment to deplete Spp1-AID*.
(F) Anti-FLAG ChIP followed by qPCR analysis of the intron-containing genes *ACT1*, *RPS13* and *ECM33* without (-) auxin (blue) or with (+) 120 minutes of auxin (yellow) addition to deplete Spp1-AID*. The X-axis shows the positions of amplicons used for ChIP qPCR analysis – the exons are of each gene in grey. The data is presented as the mean percentage of input normalised to a negative control amplicon. Mean of at least three biological replicates. Error bars = standard error of the mean.

To test whether depletion or Spp1 for 120 minutes was sufficient to reduce H3K4me3, ChIP was performed using antibodies against H3K4me3 and histone H3 followed by qPCR across intron-containing genes *ACT1*, *RPS13* and *ECM33* (Figure 4.12G). On *ACT1,* the level of H3K4me3 was significantly reduced on average at amplicons 1 to 5 (5' to exon 2) to 46%, 37%, 35%, 35% and 40% respectively after 120 minutes of depletion, relative to conditions without depletion (student's t-test, $P<0.05$) (Figure 4.12G). On *RPS13,* the level of H3K4me3 was significantly reduced on average at amplicons 1 to 5 (5' to exon 2) to 53%, 59%, 53%, 37% and 22% respectively after 120 minutes of depletion, relative to conditions without depletion (student's t-test, $P<0.05$) (Figure 4.12G). On *ECM33,* the level of H3K4me3 was significantly reduced on average at amplicons 1 to 5 (5' to exon 2) to 49%, 32%, 23%, 34% and 45% respectively after 120 minutes of depletion, relative to conditions without depletion (student's t-test, $P<0.05$) (Figure 4.12G). Therefore, depletion of Spp1 for 120 minutes was suitable for a greater depletion of H3K4me3, and these conditions were used to test for effects of loss of H3K4me3 on co-transcriptional spliceosome assembly and pre-mRNA splicing.

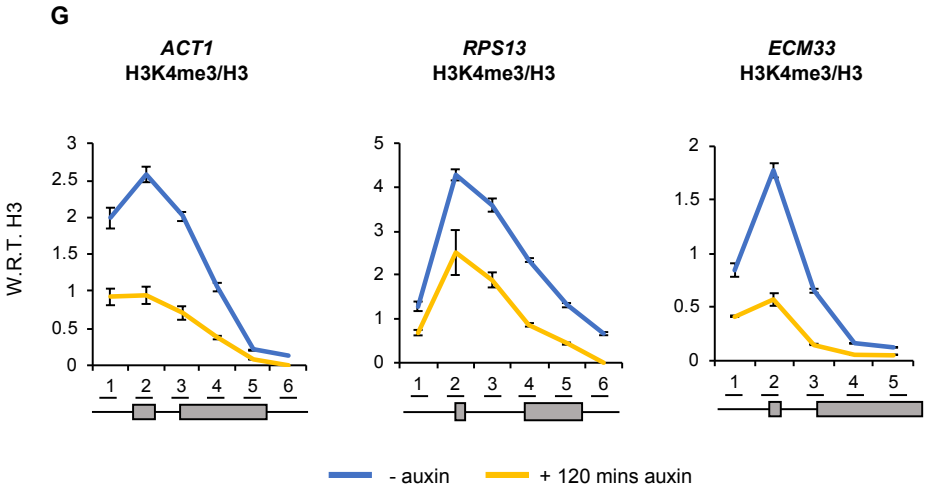


Figure 4.12 (continued). Use of the AID system to conditionally reduce the level of H3K4me3

(G) Anti-H3K4me3 ChIP followed by qPCR as in (C) with samples were taken without auxin (- auxin, blue) and with 120 minutes (+ 120 mins auxin, yellow) of auxin treatment to deplete Spp1-AID*. The X-axis shows the positions of amplicons used for ChIP qPCR analysis – the exons are of each gene in grey. The data is presented as the mean percentage of input normalised to a negative

In summary, it took 120 minutes of Spp1 depletion to greatly reduce H3K4me3 levels in whole cell extracts and across the genes tested. In *S. cerevisiae*, the average doubling time is 90 minutes, and the finding that it took 120 minutes of Spp1 depletion to greatly reduce H3K4me3 levels fits with previous findings that H3K4me3 is mostly lost by dilution of histones during DNA replication (Ng *et al*., 2003b; Radman-Livaja *et al*., 2010). The reduction in H3K4me3 observed after 60 minutes of Spp1 depletion was similar the reduction observed with Prp22 depletion.

4.12 Loss of H3K4me3 does not affect co-transcriptional spliceosome assembly

To determine whether the loss of H3K4me3 affects co-transcriptional spliceosome assembly, ChIP-qPCR was performed across intron-containing genes *ACT1*, *RPS13* and *ECM33*. Antibodies were used that detect core members of the spliceosome: the U1 snRNP (Prp40), U2 snRNP (Lea1-3HA) and U5 snRNP (Prp8), which allowed a detailed picture of which stage, if any, of co-transcriptional spliceosome assembly may be affected by loss of H3K4me3. In addition, ChIP using an antibody against RNAPII (Rpb1) was performed to test for changes in RNAPII occupancy.

ChIP-qPCR analyses showed no significant change in the occupancy of U1, U2 or U5 snRNPs upon Spp1 depletion for 120 minutes, relative to conditions without auxin (student's t-test, $P>0.05$) (Figures 4.13A-4.13C). ChIP-qPCR analyses showed no significant change in RNAPII occupancy upon Spp1 depletion across *ACT1* (student's t-test, $P>0.05$) (Figure 4.13D). There was a significant increase in RNAPII occupancy at amplicon 3 (3'SS) of RPS13 to 117% (student's t-test, $P<0.05$) (Figure 4.13D), relative to conditions without depletion. In contrast, for *ECM33* there was a significant decrease in RNAPII occupancy at amplicon 2 (exon 1) to 67% (student's t-test, $P<0.05$) (Figure 4.13D), relative to conditions without depletion.

Therefore, under these conditions, *in vivo* loss of H3K4me3 *via* Spp1 depletion does not affect co-transcriptional spliceosome assembly. This is in contrast to findings in mammalian cells where the recruitment of the U2 snRNP to chromatin is reduced by loss of H3K4me3 (Sims *et al*., 2007).

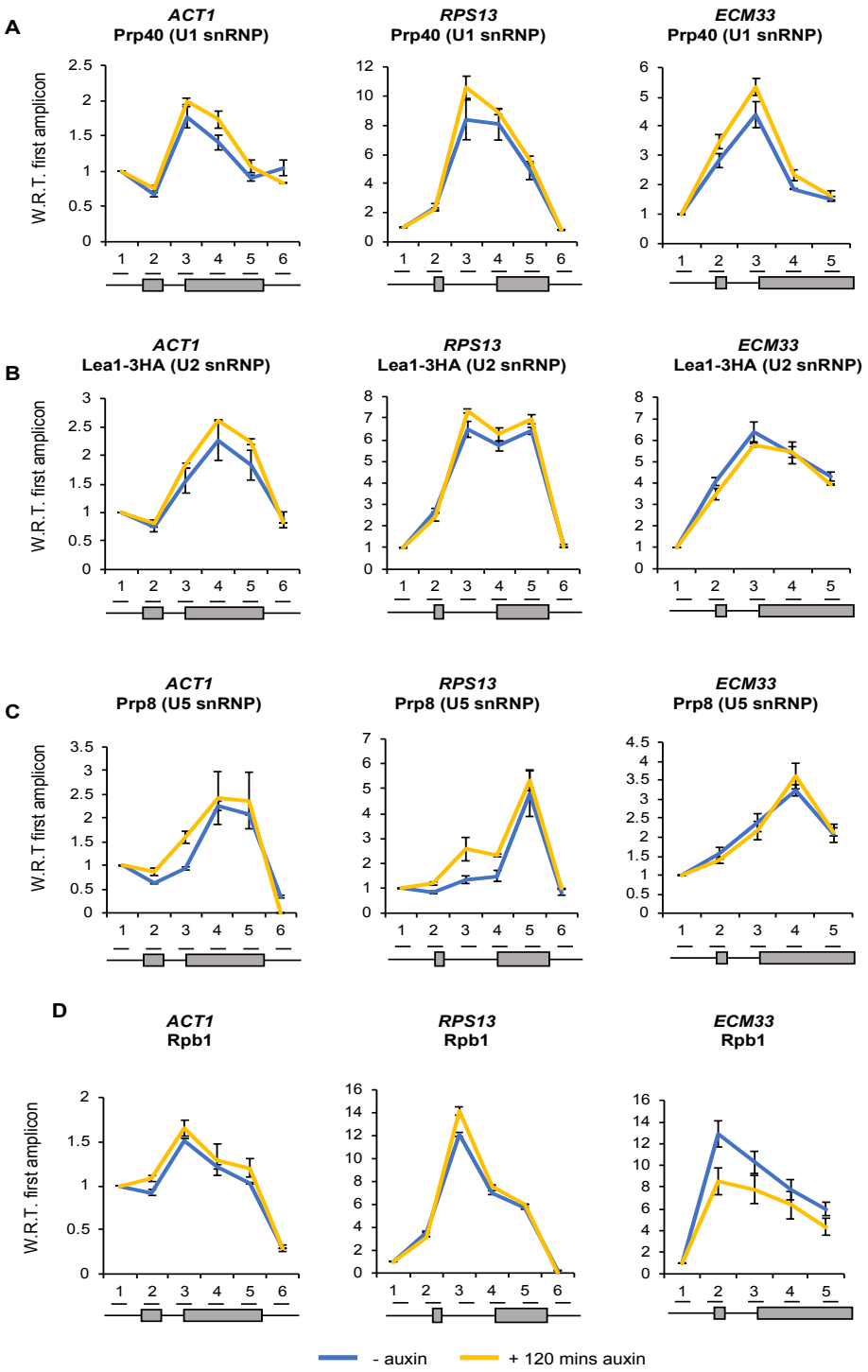


Figure 4.13. Loss of H3K4me3 does not affect co-transcriptional spliceosome assembly

(A) Anti-Prp40 (U1 snRNP) (B) Anti-Lea1-3HA (U2 snRNP) (C) Anti-Prp8 (U5 snRNP) and (D) Anti-Rpb1 ChIP followed by qPCR analysis of the intron-containing genes: *ACT1*, *RPS13*, *ECM33* without auxin (blue) and with 120 minutes (yellow) of auxin addition to deplete Spp1-AID*. The X-axis shows the positions of primers used for ChIP qPCR analysis – the exons are in grey. The data is presented as the mean percentage of input normalised to the first amplicon of each gene. Mean of at least three biological replicates. Error bars = standard error of the mean.

4.13 Loss of H3K4me3 does not affect pre-mRNA splicing

To test whether loss of H3K4me3 in these conditions affected pre-mRNA splicing, RT-qPCR was performed on total (steady-state) RNA, as described in section 3.5. No significant accumulation of pre-mRNA species upon depletion of Spp1 was observed for *ACT1*, *RPS13* and *ECM33* (student's t-test, P>0.05) (Figure 4.14A). For *ECM33*, there was a small but significant reduction in the levels of the 5'SS, mRNA and exon 2 (student's t-test, P<0.05) (Figure 4.14), indicating a possible transcription defect for this gene that was not observed for *ACT1* or *RPS13*.

Overall, depletion of Spp1 and loss of H3K4me3 does not result in defects in pre-mRNA splicing, which fits with the ChIP-qPCR data that showed no effect on co-transcriptional spliceosome assembly upon Spp1 depletion and loss of H3K4me3 (Figure 4.13). This is in contrast to the findings in mammalian cells where loss of H3K4me3 reduced pre-mRNA splicing efficiency and affects patterns of alternative splicing (Sims *et al*., 2007; Luco *et al*., 2010).

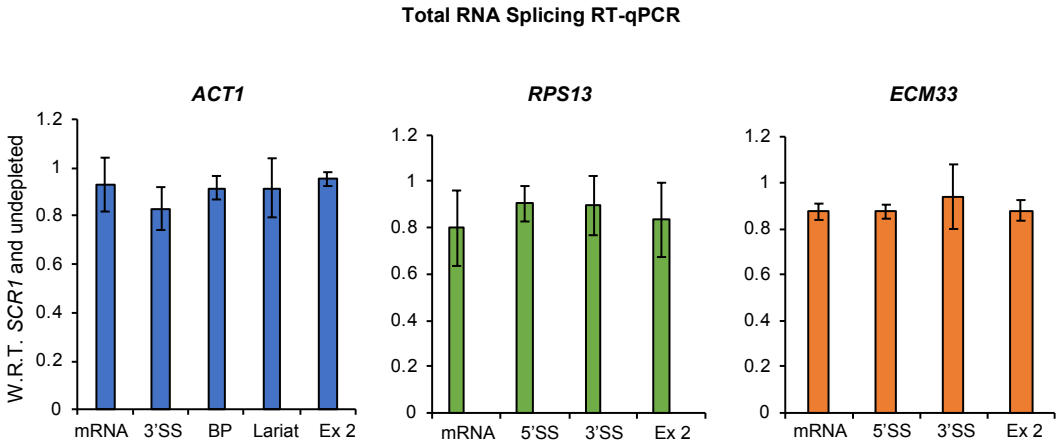


Figure 4.14. Loss of H3K4me3 does not affect pre-mRNA splicing

RT-qPCR analysis of total RNA from the intron-containing genes *ACT1*, *RPS13* and *ECM33* after 120 minutes of depletion of Paf1-AID* (A). Normalised to the *SCR1* PolIII transcript and conditions without auxin. Primers used detected pre-mRNA (3'SS or BP), lariat (excised or intron-exon 2), exon 2 (ex 2) and mRNA and are depicted in Figure 4.3D. Mean of 3 biological replicates. Error bars = standard error of the mean.

4.14 Discussion

Evidence that splicing affects H3K4me3 comes from experiments that make use of a splicing inhibitor drug (spliceostatin A) to block the first step of splicing of endogenous transcripts or transcripts of a reporter gene with a *cis*-acting 3'SS mutation that blocks the second step of splicing catalysis (Sims *et al*., 2007; Bieberstein *et al*., 2012). These studies effectively demonstrate that splicing is important for proper levels of H3K4me3. Spliceostatin inhibits SF3b (U2 snRNP) and inhibits spliceosome assembly between the A and B complex formation (Roybal and Jurica, 2010). In the case of a 3'SS mutant, the spliceosome can assemble, the first step of splicing occurs but the second step of splicing does not. These data point to spliceosome assembly and the second step as important for proper levels of H3K4me3 in mammalian cells, but do not pin-point precisely which factors may be important – particularly as the spliceosome assembles in a step-wise manner and inhibiting earlier stages by definition inhibits later stages and so it is difficult to distinguish between primary and secondary effects. Additionally, these experiments were all performed in mammalian systems, and it was not clear whether splicing affects H3K4me3 in *S. cerevisiae.* Experiments here describe the use of the AID system to conditionally deplete essential splicing factors that act throughout the splicing cycle to address whether splicing affects H3K4me3 in *S. cerevisiae*, and if so to pin-point a particular stage, or particular protein, that may be important for proper levels of H3K4me3.

Here, the AID system was effective in conditionally depleting the essential splicing factors from whole cell extracts and chromatin, and the expected splicing defects were observed upon their depletion. Depletion of first step splicing factors Prp39 (U1 snRNP) and Prp9 (U2 snRNP) for 30 minutes caused defects in the first step of splicing on the intron-containing genes tested, whilst depletion of second step splicing factor Slu7 caused defects in the first and second step of splicing (Figure 4.3). This is likely because a recycling defect, in which snRNPs are stuck in arrested spliceosomes, is caused by Slu7-AID* depletion, which feeds back to the first step of splicing, giving a secondary effect of a first step splicing defect. A recycling defect upon depletion of U5 snRNA has been previously demonstrated, and Gonzalo Mendoza-Ochoa (Beggs

lab) has also extensively demonstrated recycling defects upon splicing factor depletion by the AID system (Tardiff and Rosbash, 2006; Gonzalo Mendoza-Ochoa, personal communication). In 30 minutes, Prp22 was less well depleted than the other splicing factors, and this milder depletion did not result in defects in splicing catalysis, but the accumulation of the excised intron-lariat (Figure 4.3). This observation fits with the primary role of Prp22 in the release of the mRNA after splicing catalysis has occurred (Wagner *et al*., 1998; Aronova *et al*., 2007; Schwer, 2008; Fica *et al*., 2017).

Depletion of factors that affected the first (Prp39 and Prp9), or first and second (Slu7), step of splicing caused a significant reduction in H3K4me3 on the intron-containing genes tested (Figure 4.4). This observation fits with the evidence from mammalian cells that inhibiting the first or second step of splicing reduces the level of H3K4me3 (Sims *et al*., 2007; Bieberstein *et al*., 2012). Unexpectedly, depletion of Prp22 to a level where no defects in splicing catalysis were observed by splicing RT-qPCR also caused a significant reduction in H3K4me3 (Figure 4.5A). No splicing defects observed upon mild depletion of Prp22 was verified by a time course experiment with more thorough depletion of Prp22 and 4-thiouracil labelling of newly synthesised RNA (Figures 4.5B-4.5E). These experiments showed again that milder depletion of Prp22 did not cause defects in splicing catalysis, but longer depletion of Prp22 showed first and second step splicing defects, likely due to a recycling defect (as in the case of Slu7 depletion). From this point I decided to focus on Prp22 as the factor that may regulate the level of H3K4me3 on intron-containing genes because the spliceosome is well-established to assemble on pre-mRNA substrates in a step-wise manner, and depletion of the earlier acting splicing factors (Prp39, Prp9 and Slu7) caused a reduction in the co-transcriptional recruitment of Prp22 to the intron-containing genes tested (Figure 4.6).

Analysis of the effects of depletion of Prp22 on other histone modifications (H3K36me3, H3K4me1, H3K4me2) revealed most consistently that these modifications were reduced after an hour of Prp22 depletion (where defects in the first and second steps of splicing were detected), not in the milder 30 minute depletion (no defects in splicing catalysis detected) (Figures 4.5G-4.5I). The finding that a mild

depletion of Prp22 did not affect H3K36me3 fits with the findings of Dr. Emanuela Sani (Beggs lab), who has investigated the effects of splicing on H3K36me3 and found that blocking pre-spliceosome formation, but not later steps, leads to loss of H3K36me3 and recruitment of H3K36 methyltransferase Set2 (Dr. Emanuela Sani, personal communication). The delay in loss of H3K4me2 and H3K4me1 is consistent with previous studies that show loss of H3K4me3 and Set1 without loss of H3K4me2 and H3Kme1 (Schlichter and Cairns, 2005; Luciano *et al*., 2017; Sayou *et al*., 2017). These data suggest specificity of the effect of Prp22 depletion to H3K4 methylation states.

How might depletion of Prp22 result in a reduction in H3K4me3? One possibility was that the recruitment of Set1, the only histone methyltransferase responsible for H3K4me3 in *S. cerevisiae,* was reduced. ChIP experiments showed that Set1 recruitment was reduced on the intron-containing genes tested upon Prp22 depletion, giving an explanation as to why H3K4me3 was reduced after Prp22 depletion (Figure 4.7A). Set1 is recruited to the CTD of transcribing RNAPII *via* the Paf complex which interacts with the serine 5 phosphorylated CTD of RNAPII (Krogan *et al*., 2003; Ng *et al*., 2003b). It was possible that Set1 recruitment was reduced because RNAPII was reduced, however no significant changes to RNAPII occupancy or serine 5 phosphorylation of the CTD were observed in the present work, and therefore I discount this possibility (Figures 4.7B and 4.7C). To fit with the reduction of Set1 recruitment upon Prp22 depletion, recruitment of Spp1, a member of the *S. cerevisiae* Set1 complex required for H3K4me3 was also reduced after Set1 depletion (Figure 4.8).

Currently, it is unclear how widely in the genome loss of Set1 recruitment and H3K4me3 occurs upon Prp22 depletion, though we have some evidence that it is specific to intron-containing genes and not intronless genes (Figure 4.5J). To address this, ChIP-sequencing has been performed of 9MYC-Set1 after Prp22 depletion by our collaborator, Vincent Géli, and we are awaiting the results (Prof. Vincent Géli, personal communication). The observation that blocking spliceosome assembly and

splicing-dependent Prp22 recruitment reduces H3K4me3 suggests that the phenomenon is splicing-dependent.

The loss of Set1 recruitment and H3K4me3 upon Prp22 depletion could be solely due to loss of the physical presence of Prp22 (presumably co-transcriptionally) or due to the loss of its ATPase activity. To determine this, WT Prp22-V5, underactive ATPase T757A Prp22-V5 and overactive ATPase I764A Prp22-V5 were each conditionally expressed (using the β-estradiol system) prior to depletion of endogenous Prp22. ChIP experiments showed each WT/mutant Prp22 protein were recruited co-transcriptionally to the intron-containing genes tested (Figure 4.9F). Splicing RT-qPCR demonstrated *in vivo* defects in the first and second step of splicing upon induction of the mutants, and not in the expression of WT Prp22-V5 (Figure 4.9H). Therefore, WT Prp22-V5 could prevent splicing defects observed upon Prp22 depletion, and the mutants appeared to have a dominant affect and compound the splicing defects, probably because the spliceosome cannot disassemble, and recycling defects occur (as was observed with depletion of Slu7). These data agree with the *in vitro* data that showed that the ATPase mutants allowed the second step of splicing but caused defects in mRNA release (Schneider *et al*., 2004), and the *in vivo* data of Gonzalo Mendoza-Ochoa (Beggs lab), who similarly used the mutants. As the mutant Prp22 proteins were recruited co-transcriptionally to the intron-containing genes tested, this means the proteins are recruited, this blocks spliceosome recycling and defects in pre-mRNA splicing occur.

These experiments demonstrated that expression of WT Prp22-V5 can prevent the reduction in Set1 recruitment and H3K4me3 observed upon Prp22 depletion, and so can expression of an overactive ATPase mutant of Prp22 (I764A) that can hydrolyse ATP but not couple ATP hydrolysis to RNA unwinding (Figures 4.10A and 4.10B). Expression of an underactive ATPase mutant (T757A Prp22-V5) could not prevent the reduction in Set1 recruitment and H3K4me3 observed upon Prp22 depletion (Figures 4.10A and 4.10B). No significant differences in RNAPII occupancy were observed, and therefore this cannot explain the differences (Figure 4.10C). The main difference between the Prp22 mutants was their ability to hydrolyse ATP – both can bind RNA

but lack RNA unwinding activity. Therefore, the ATPase activity of Prp22 without its RNA unwinding activity is sufficient to prevent loss of Set1 recruitment and H3K4me3 normally observed upon Prp22 depletion. Of course, this is an abnormal situation – the ATP hydrolysis of Prp22 is normally coupled to 3'-5' RNA unwinding and mechanical work to disrupt the contacts between the U5 snRNP and mRNA. The ATPase activity of Prp22 is not required for the second step, and therefore any effects of Prp22 depletion are not due to second step splicing defects. This latter statement fits with the fact that depletion of Prp22 without defects in splicing catalysis reduced Set1 recruitment and H3K4me3, and also fits with the finding that induction of the overactive ATPase mutant could prevent loss of H3K4me3 in the presence of first and second step defects in pre-mRNA splicing.

An interesting observation was that expression of mutant Prp22 proteins resulted in reduction in exon 2 and mRNA levels of the intron-containing genes tested, whilst pre-mRNA species accumulate (Figure 4.9H). As pre-mRNA accumulates, this indicates that specifically spliced RNA is degraded. This is likely important, as aberrant mRNAs produced may translate into aberrant proteins that have negative consequences to the cell. It is curious that pre-mRNAs are not degraded, but it suggests that pre-mRNA that is spliced in these conditions is degraded in a manner perhaps triggered by failure to satisfy a Prp22-dependent checkpoint. This effect was much more dramatic upon expression of the Prp22 mutants in comparison to depletion of Prp22 alone (Figure 4.9G), which supports the idea that the mutants can trigger degradation. As there are secondary defects in splicing catalysis upon expression of the mutants (due to spliceosome recycling defects), eventually no mRNA would be produced which means mRNA and associated exon 2 levels decrease and pre-mRNA species accumulate as they are protected from degradation by the spliceosome or synthesised at a rate such that the degradation machinery cannot keep up (Schwer and Meszaros, 2000; Hicks *et al*., 2006; Mayas *et al*., 2006). Previously, Gonzalo Mendoza-Ochoa also observed this phenomenon with the same Prp22 mutants. In this case, *UPF1* was deleted, and therefore this phenomenon is not Upf1-dependent (Gonzalo Mendoza-Ochoa, Beggs lab, personal communication). It is possible that failure to satisfy a Prp22-dependent checkpoint recruits the nuclear exosome. This could be tested by deletion of nuclear

3’ to 5’ exonuclease Rrp6 in the strains with Prp22 WT/mutant. It is suggested that splicing factors could facilitate recruitment of the exosome (reviewed in Bresson and Tollervey, 2018) and it has been shown in *S. pombe* that sequences within a subset of introns can promote recruitment of the exosome and degradation of pre-mRNA, which is important for regulation of gene expression in different environmental conditions, including stress conditions (Kilchert *et al*., 2015). It is not known whether intron-dependent recruitment of the exosome could degrade mRNA.

Yeast-2-hybrid experiments found that Prp22 and Set1 interact, and this seemed quite robust as Prp22 was found in all three screens, using full-length Set1, Set1 1-754 or Set1 754-1081 as baits (Figure 4.11A) (Prof. Vincent Géli, personal communication). The region of Prp22 common to all three screens is amino acids 508 to 762, which is the ATP binding part of the helicase domain of Prp22. Validating the yeast-2-hybrid results, co-immunoprecipitation experiments showed that Prp22 pulled-down Set1 (Figures 4.11B and 4.11C). Further, this interaction was not abolished by RNase treatment. Though the co-immunoprecipitation experiments show that Set1 co-immunoprecipitates with Prp22, and yeast-2-hybrid data (Prof. Vincent Géli, personal communication) show interaction between Set1 and Prp22, at this stage I cannot determine whether this interaction is direct or indirect. As Set1 is recruited to transcribing RNAPII, one possibility was that the interaction was indirect, *via* RNAPII. There are two reasons I discard this possibility. Firstly, as Prp22 does not pull-down Rpb1 (the largest subunit of RNAPII, containing the CTD), this suggests that the co-immunoprecipitation is not *via* RNAPII (Figure 4.11D). Secondly, the reduction in Set1 recruitment observed upon Prp22 depletion occurred without changes to RNAPII occupancy or serine 5 phosphorylation of the CTD. Another possibility is that the interaction between Prp22 and Set1 is indirect, *via* other splicing factors, as has been observed in mammalian cells where Set1 was detected by mass spectrometry to interact with the U2 snRNP (Allemand *et al*., 2016). Further, Chd1 which binds to H3K4me3, was shown to interact with SF3a (U2 snRNP) (Sims *et al*., 2007). To test this, co-immunoprecipitation experiments with other splicing factors to look for interaction with Set1 should be performed. Another possibility was that the interaction is independent of splicing and the spliceosome. To test this splicing could

be blocked by depletion of an earlier splicing factor that is required for the first step of splicing (such as Prp39 or Prp9) using the AID system. Experiments here showed that co-transcriptional recruitment of Prp22 was reduced upon their depletion, and therefore pull-down Prp22 in these conditions would test whether the pull-down of Set1 by Prp22 is dependent on splicing. As the expression of an ATPase mutant of Prp22 was able to reverse the effects of Prp22 depletion on Set1 recruitment and H3K4me3, an intriguing hypothesis is that ATP hydrolysis by Prp22 stimulates its interaction with Set1, possibly by conformational changes to Prp22. The fact that the ATP binding part of the helicase domain of Prp22 interacts with Set1 fits with this hypothesis. This could be tested by performing co-immunoprecipitation experiments with the WT and ATPase mutant Prp22 proteins. Presumably, Prp22 lets go of the mRNA after ATP hydrolysis and as Prp22 is an RNA-stimulated ATPase, this may expose the ATP binding site in Prp22, allowing it to bind set1.

Together, these data suggest a way by which at least the physical presence of Prp22 could affect Set1 recruitment and subsequently H34Kme3 – by a physical interaction that is possibly stimulated by the ATP hydrolysis of Prp22 and therefore its ATP-dependent RNA helicase activity (Figure 4.15). In agreement with the role of Prp22 in modulating chromatin, a previous study in *Drosophila* oocytes found that RNAi knockdown of *peanuts* (Prp22) resulted in splicing factor recycling defects and failures of chromatin decondensation in nurse cells (Klusza *et al*., 2013). The mechanism and significance of chromatin decondensation in *Drosophila* nurse cells is unclear, however in other systems chromatin decondensation is associated with H3K4me3 and active transcription (reviewed in Cremer *et al*., 2015).

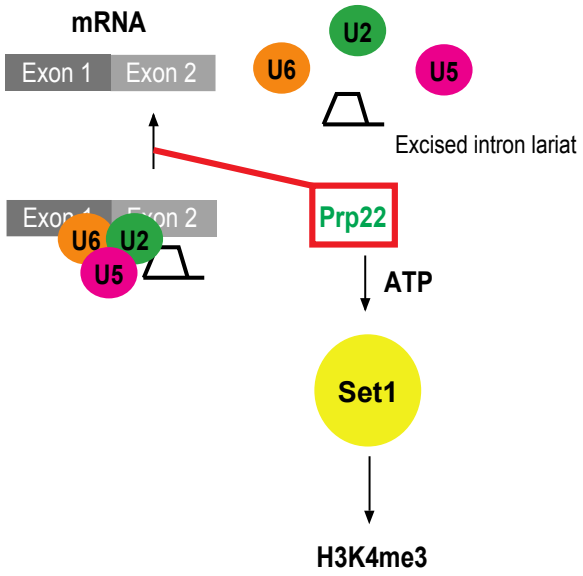


Figure 4.15. Prp22 affects Set1 recruitment and H3K4me3 in *S. cerevisiae*

Depletion of Prp22 in the absence of defects in splicing catalysis reduces Set1 recruitment and H3K4me3. This phenomenon seems to be dependent on the ATPase activity of Prp22. Prp22 also interacts (genetically and in co-immunoprecipitation experiments) with Set1, suggesting that Prp22 (directly or indirectly) helps to recruit Set1 to genes. We propose that Prp22 hydrolyses ATP and signals splicing has occurred. Without Prp22 and its ATPase activity, Set1 recruitment and H3K4me3 are reduced.

In the present work, ChIP data of Prp22 occupancy show that Prp22 peaks towards the 3'end of genes (Figure 4.3C), whereas Set1 is in gene bodies and H3K4me3 peaks at the 5' end of genes. The ChIP data here raise the question as to how Prp22 toward the 3'end of the gene might affect H3K4me3 at the 5'end? Firstly, ChIP is an indirect assay that provides a delayed snapshot of co-transcriptional splicing factor occupancy, meaning splicing factors are detected by ChIP downstream to where they bind pre-mRNA (this could be because the pre-mRNA is not aligned in parallel with the DNA). This means that Prp22 is likely present upstream to where it is detected by ChIP. It is known from *in vitro* cross-linking studies that Prp22 binds 13-33 nucleotides downstream of the 3'SS (Schwer, 2008), and the recent structure of the postcatalytic splicing complex shows Prp22 binding 15-21 nucleotides downstream of the exon-exon junction (Liu *et al*., 2017). We have performed ChIP-sequencing of Prp22 which shows that although Prp22 occupancy peaks toward the 3'ends of some genes, on many it is present within introns (Dr. Alastair Kerr, Dr Gabriele Schweikert and Dr Guido Sanguinetti personal communication). The sequencing data also show that the

level of Prp22 occupancy correlates with RNAPII occupancy which is consistent with ChIP-sequencing of earlier-acting splicing factors in *S. cerevisiae* (Moore *et al*., 2006). Set1 also associates with RNAPII and was recently shown to associate with nascent RNAPII transcripts (Battaglia *et al*., 2017; Luciano *et al*., 2017; Sayou *et al*., 2017). The CTD of RNAPII is proposed to be a 'landing pad' for co-transcriptional recruitment of splicing factors, and it is possible that Prp22 (with other splicing factors) generally associates with RNAPII, and in this way influences Set1 recruitment and therefore H3K4me3 over multiple rounds of transcription. Another possibility is that the genes are looped – where the 5' and 3' ends of a gene come into close contact during transcription (O'Sullivan *et al*., 2004; Ansari and Hampsey, 2005). It is hypothesized that the presence of an intron may facilitate gene looping (Moabbi *et al*., 2012).

What is the biological significance of this phenomenon whereby Prp22 affects Set1 recruitment and H3K4me3? Prp22 is an ATPase that acts at the very late stages of the splicing cycle to release the mRNA after splicing catalysis is complete. Prp22 was shown to bind in the second exon of pre-mRNA substrates 17 nucleotides downstream of the exon-exon junction, and then acts as a 3' to 5' ATP-dependent RNA helicase to unwind contacts between the mRNA and U5 snRNA (Wagner *et al*., 1998; Aronova *et al*., 2007; Schwer, 2008; Fica *et al*., 2017; Liu *et al*., 2017). Additionally, Prp22 was reported to have roles in fidelity; rejecting sub-optimal splice sites *via* its ATP-dependent RNA unwinding activity before exon ligation has occurred (Mayas *et al*., 2006). Both these roles of Prp22 are dependent on its ATPase activity, and the finding that the effect of Prp22 depletion on H3K4me3 and Set1 recruitment can be prevented by expression of an overactive ATPase mutant of Prp22 that is unable to couple ATP hydrolysis to RNA unwinding suggests that the phenomenon is dependent on the ATPase activity of Prp22. This indicates either that ATP hydrolysis by Prp22 could be a signal that splicing is complete and mRNA released, or it could be a signal that the substrate is sub-optimal and rejected.

As H3K4me3 is generally thought of as an 'activating' chromatin mark, it would make more sense to stimulate transcription of a gene the transcript of which has been

correctly spliced than one that has been incorrectly spliced, although you could argue either way. H34Kme3 stimulated by splicing might be important in serving as a memory that stimulates future rounds of transcription and splicing, perhaps by providing a platform to enhance co-transcriptional recruitment of splicing factors. There is precedent for this - in mammalian cells, loss of H3K4me3 or H3K4me3 reader Chd1 was shown to cause a reduction in pre-mRNA splicing efficiency and U2 snRNP recruitment to chromatin (Sims *et al*., 2007). Further, in mammalian cells it was shown by metabolic labelling that the loss of an intron in a reporter gene reduced transcriptional output, and treatment with splicing inhibitor drug spliceostatin globally reduced transcription (Bieberstein *et al*., 2012). A caveat here is that splicing is sensitive to transcription speed, and any changes to transcription could lead to changes in co-transcriptional spliceosome assembly and splicing, and a steady-state is reached for each gene. In the present work, no consistent changes in RNAPII occupancy were observed upon loss of H3K4me3 (Figure 4.13D), though it could be during steady-state differences are difficult to detect. There is precedent this – a study found that effects of Set2 (methyltransferase for H3K36) on gene expression were only detectable on inducible genes and in certain environmental conditions (Kim *et al*., 2016). It may be informative to perform experiments on an inducible reporter gene that could be more sensitive to changes in H3K4me3.

In order to gauge whether H3K4me3 was important for co-transcriptional spliceosome assembly and splicing in *S. cerevisiae*, Spp1 was conditionally depleted by the AID system (Figure 4.12). The average doubling time of *S. cerevisiae* is 90 minutes, and the finding that it took 120 minutes of Spp1 depletion to greatly reduce H3K4me3 levels fits with previous findings that H3K4me3 is mostly lost by dilution of histones during DNA replication (Ng *et al*., 2003b; Radman-Livaja *et al*., 2010). The reduction in H3K4me3 observed after 60 minutes of Spp1 depletion was similar the reduction observed by Prp22 depletion, meaning the relative kinetics of H3K4me3 loss are comparable. No effects on co-transcriptional spliceosome assembly or splicing were observed, nor any effects on RNAPII occupancy (Figures 4.13 and 4.14). This is consistent with findings that deletion of Set1 did not affect co-transcriptional spliceosome assembly in yeast and only affected splicing of a subset of yeast genes

(approximately 40) (Hérissant *et al*., 2014), but is in contrast to findings in mammalian cells (as discussed above) (Bieberstein *et al*., 2012). These differences could be due to differences in splicing complexity between mammalian cells and yeast cells in terms of splicing – the core splicing factors (approximately 90) are conserved between *S. cerevisiae* and mammals, but mammals have approximately double this number due to alternative splicing, which is rare in *S. cerevisiae* (Fabrizio *et al*., 2009). The finding here that loss of H3K4me3 does not affect RNAPII occupancy fits with previous studies in *S. cerevisiae* where loss of H3K4me3 did not result in significant changes to gene expression (Briggs *et al*., 2001; Miller *et al*., 2001) and the role for H3K4me3 in transcription is currently unclear (Howe *et al*., 2017).

It is possible that any effects of H3K4me3 on co-transcriptional spliceosome assembly and splicing are undetected by the experiments described here because the level of H3K4me3 was not depleted enough. However, Set1 is non-essential and its deletion did not show wide-spread splicing defects (Hérissant *et al*., 2014). Further, David Barrass (Beggs lab) performed 4-thiouracil labelling of newly synthesised RNA in Set1 and Spp1 deletion strains, which did not show defects in transcription or splicing catalysis relative to wild-type conditions (data not shown). It may be that H3K4me3 could be particularly important in certain conditions – such as stress or meiosis. In comparison to other chromatin marks, H3K4me3 was shown to be a very stable mark in *S. cerevisiae* - remaining on genes after transcription shutdown during meiosis and osmotic stress, and it is suggested that H3K4me3 could serve as a transcriptional memory that allows more rapid changes in gene expression (Borde *et al*., 2009; Magraner-Pardo *et al*., 2014; Weiner *et al*., 2015; Howe *et al*., 2017). Therefore, it may be that testing for splicing defects upon loss of H34Kme3 in stress conditions or meiosis would be informative. Alternatively, any defects in splicing could be hard to detect due to pre-mRNA species being degraded by the degradation machinery, which has been demonstrated previously (Hossain *et al*., 2011; Kawashima *et al*., 2014). Most recently, a role for H2A.Z in splicing was only apparent in cells in which components of the degradation machinery (Upf1 and Xrn1) were deleted (Neves *et al*., 2017).

Whilst the finding that splicing affects H3K4me3 is conserved between *S. cerevisiae* and mammalian cells, whether the observed links between Prp22, Set1 and H3K4me3 are conserved is currently unclear. Prp22 is conserved from yeast to mammals, and experiments could be performed in mammalian systems to test whether the links are conserved. Certainly, effects of loss of H3K4me3 on splicing and co-transcriptional spliceosome assembly are clearer in mammalian systems (Sims *et al*., 2007; Bieberstein *et al*., 2012).

It is currently unclear whether in *S. cerevisiae* there is an increased level of H3K4me3 on intron-containing genes in comparison to intronless genes, and the loss of H3K4me3 simply reduces the level of H3K4me3 to the level of intronless genes. This may be a difficult question to address, as intron-containing genes are more highly expressed than intronless genes, however we are currently investigating this question bioinformatically (in collaboration with Dr Alastair Kerr, Dr Gabriele Schweikert and Dr Guido Sanguinetti). It is also hard to distinguish whether splicing promotes H3K4me3 or loss of splicing removes H3K4me3. It is difficult to address this question, as during gene expression genes are expressed and spliced. Importantly, in the mild depletion of Prp22, transcription and splicing are both wild-type but H3K4me3 is still reduced. This suggests that a stage of splicing Prp22 and beyond is important for this phenomenon.

In summary, depletion of factors that result in first (Prp39, Prp9), or first and second (Slu7), step splicing defects resulted in reduced H3K4me3 on the intron-containing genes tested. Most interestingly, a mild depletion of Prp22 reduced Set1 recruitment and H3K4me3 without causing defects in splicing catalysis or transcription. As the spliceosome assembles on pre-mRNA substrates in a step-wise manner, depletion of earlier-acting factors also reduces co-transcriptional recruitment of Prp22, and for this reason we focused on Prp22 as the latest-acting factor required for proper levels of H3K4me3 on intron-containing genes. Notably, expression of an overactive ATPase mutant of Prp22 prevented the loss of Set1 recruitment and H3K4me3 normally observed upon Prp22 depletion. This suggests that the effect of Prp22 depletion on

Set1 and H3K4me3 is due to the loss of Prp22 ATPase activity. Together, these data highlight a previously unknown link between a late-acting splicing factor Prp22, Set1 and H3K4me3. We favour the idea that ATP hydrolysis by Prp22 after correct pre-mRNA splicing has occurred stimulates an interaction between Prp22 and Set1 (brought about perhaps by conformational changes in Prp22 during ATP hydrolysis and release of mRNA) which affects the recruitment of Set1 and subsequently H3K4me3. Our hypothesis therefore is that H3K4me3 may reflect the status of co-transcriptional splicing, as signalled by the ATPase activity of Prp22 (Figure 4.15). Currently, the biological significance of such a phenomenon is unclear, but it may be important in stress conditions or meiosis and difficult to detect by assays used in the present work. We hypothesise that the level of H3K4me3 may act as a memory that the gene has been spliced and promote transcription through the intron or nucleosomes in the next round of transcription, and thereby enhance co-transcriptional splicing.

Chapter 5. Closing remarks

In the present study, the use of the AID system has been an effective and novel way to characterise the links between splicing and transcription and splicing and chromatin in *S. cerevisiae*. Until now, most studies examining the functions of transcription elongation factors have made use of mutants or deletions if non-essential. Splicing inhibitor drugs are not available for use in *S. cerevisiae* so that most studies have made use of mutants. Such studies have undoubtedly been fruitful, however in some cases adaptation may have occurred and important functions may have been missed or effects observed indirectly. The AID system bypasses these problems and has the potential to provide novel insights into the functions of both non-essential and essential proteins in a wide-variety of biological processes in a variety of systems.

It was already known that mutations or depletion of particular elongation factors resulted in defects in pre-mRNA splicing, but it was not known mechanistically how these effects came about. For example, previous studies had shown mutants and depletion of core transcription elongation factor Spt5 caused splicing defects in yeast, but why this should be was unknown. The present work has shown that Spt5 is required for proper co-transcriptional recruitment/stability of the U5 snRNP and that, without Spt5, the B complex fails to form co-transcriptionally. Further, Spt5 co-immunoprecipitates with U5 snRNP, showing a possible mechanism as to how Spt5 could affect U5 snRNP recruitment and B complex formation co-transcriptionally; this also provides mechanistic insight into how Spt5 depletion/mutation could cause splicing defects. The AID system proved an effective way to examine the functions of the core elongation machinery not only in co-transcriptional spliceosome assembly, but also their documented functions. In the present work, some findings did not agree with the literature; for example, finding that depletion of Bur1 or Bur2 did not affect serine 2 phosphorylation of the CTD of RNAPII. These differences are probably due to the fact that most studies have made use of mutants or deletions, which may have adapted over time.

The present work also highlighted apparent differences between lower and higher eukaryotes. For example, serine 2 phosphorylation was shown to be important for pre-mRNA splicing in mammals, however in the present work depletion of Ctk1 in *S. cerevisiae* effectively abolished serine 2 phosphorylation of the CTD and did not affect co-transcriptional spliceosome assembly and splicing. As splicing is more complicated in mammals, this finding may reflect the need for more advanced regulation of co-transcriptional splicing. Collectively, these data provide systematic and mechanistic insight into the contribution of the core transcription elongation machinery to co-transcriptional spliceosome assembly and splicing in *S. cerevisiae*.

The AID system was a particularly powerful approach for depleting essential splicing factors that act at different stages of the splicing cycle and asking whether splicing affects H3K4me3 in *S. cerevisiae* and, if so, which stage of splicing. Whilst there were known links between H3K4me3 and splicing in mammals, no such links had been observed in *S. cerevisiae* and mechanistically, how splicing might affect H3K4me3. The present work has shown that defects in the first or first and second steps of pre-mRNA splicing reduces H3K4me3 on intron-containing genes. Furthermore, depletion of late-acting factor Prp22, without defects in splicing catalysis, reduces H3K4me3. The AID system pin-pointed Prp22 as the latest-acting factor required for proper levels of H3K4me3 on the intron-containing genes tested. This work has also shown that Prp22 depletion reduces Set1 methyltransferase recruitment and that Prp22 interacts with Set1, showing how the loss of Prp22 could affect H3K4me3 – via Set1 recruitment. Use of Prp22 ATPase mutants showed it was not just the physical presence of Prp22, but its ATP hydrolysis that is important for Set1 recruitment and H3K4me3 on intron-containing genes in S. cerevisiae. The biological significance of these links is currently unclear. It is possible that these links become more important under certain environmental conditions.

Together, work presented here shows new insight into the connections between splicing and transcription and splicing and chromatin in a relatively simple biological system. It is clear from the present work and the literature that there are key differences in the links between splicing and transcription and splicing and chromatin from lower

to higher eukaryotes, likely due to differences in complexity of these biological processes. The work described here was carried out on a gene-by-gene basis which has obvious limitations and important insights may have been missed. Genome-wide studies would ensure that important phenomena do not go unmissed.

Several key questions remain, including: how conserved are these connections between splicing and transcription and splicing and chromatin? What is their biological significance? Are the effects genome-wide? The AID system is applicable in yeast and in mammals, and it would be interesting to perform genome-wide studies in both systems with and without depletion of some of the transcription elongation factors/chromatin modifiers examined in the present work to gauge effects on co-transcriptional spliceosome assembly and splicing. The degree of conservation of this phenomenon would also be examined. It would be interesting to conditionally deplete conserved splicing factors and examine effects on a variety of histone modifications genome-wide. Further, whether the links between splicing and transcription and splicing and chromatin are more important in certain environmental conditions is unclear, and it would be interesting to perform genome-wide studies to address this. Such genome-wide studies could be complimented by mechanistic studies, and together these data would combine to provide clearer insight into the links between splicing and transcription and splicing and chromatin in a variety of different biological systems.

Appendix 1. Table of *S. cerevisiae* strains used in this study

Name	Genotype	Source
W303	*MATα ade2-1 ura3-1 his3-11,15 trp1-1 leu2-3,112 can1-100*	Beggs lab
YBRT	*MATα ade2-1 ura3-1 trp1-1 leu2-3,112 can1-100 his3-11,15::PADH1-397-OsTIR1*	Barbara Terlow (Beggs lab)
YBRT Prp22-AID*-6FLAG	*MATα ade2-1 ura3-1 trp1-1 leu2-3,112 can1-100 his3-11,15::PADH1-397-OsTIR1 PRP22::PRP22-AID*-6FLAG-HygMX*	Ema Sani (Beggs lab)
YBRT Prp22-AID*-6FLAG 9MYC-Set1	*MATα ade2-1 ura3-1 trp1-1 leu2-3,112 can1-100 his3-11,15::PADH1-397-OsTIR1 PRP22::PRP22-AID*-6FLAG-HygMX SET1::9MYCSET1-TRP1*	This study
YBRT Prp22-AID*-6FLAG Spp1-9MYC	*MATα ade2-1 ura3-1 trp1-1 leu2-3,112 can1-100 his3-11,15::PADH1-397-OsTIR1 PRP22::PRP22-AID*-6FLAG-HygMX SPP1::SPP1-9MYC-KAN*	This study
YBRT 9MYC-Set1	*MATα ade2-1 ura3-1 trp1-1 leu2-3,112 can1-100 his3-11,15::PADH1-397-OsTIR1 SET1::9MYCSET1-TRP1*	This study
YBRT Bur1-AID*-6FLAG Lea1-3HA	*MATα ade2-1 ura3-1 trp1-1 leu2-3,112 can1-100 his3-11,15::PADH1-397-OsTIR1 BUR1::BUR1-AID*-6FLAG-HygMX LEA1::LEA1-3HA-KAN*	This study
YBRT Bur1-AID*-6FLAG Lea1-3HA Rpb3-TAP	*MATα ade2-1 ura3-1 trp1-1 leu2-3,112 can1-100 his3-11,15::PADH1-397-OsTIR1 BUR1::BUR1-AID*-6FLAG-HygMX LEA1::LEA1-3HA-KAN RPB3::TAP-HIS*	This study
YBRT Bur2-AID*-6FLAG Lea1-3HA	*MATα ade2-1 ura3-1 trp1-1 leu2-3,112 can1-100 his3-11,15::PADH1-397-OsTIR1 BUR2::BUR2-AID*-6FLAG-HygMX LEA1::LEA1-3HA-KAN*	This study
YBRT Ctk1-AID*-6FLAG Lea1-3HA	*MATα ade2-1 ura3-1 trp1-1 leu2-3,112 can1-100 his3-11,15::PADH1-397-OsTIR1 CTK1::CTK1-AID*-6FLAG-HygMX LEA1::LEA1-9MYC-KAN*	This study
YBRT Paf1-AID*-6FLAG Lea1-3HA	*MATα ade2-1 ura3-1 trp1-1 leu2-3,112 can1-100 his3-11,15::PADH1-397-OsTIR1 PAF1::PAF1-AID*-6FLAG-HygMX LEA1::LEA1-3HA-KAN*	This study
W303 Spt5-AID*-6FLAG Lea1-3HA pBest-TIR1-LEU2	*MATα ade2-1 ura3-1 trp1-1 leu2-3,112 can1-100 his3-11,15 SPT5::SPT5-AID*-6FLAG-HygMX LEA1::LEA1-3HA-KAN [pBest-TIR1-LEU2]*	This study
YBRT Spp1-AID*-6FLAG Lea1-3HA	*MATα ade2-1 ura3-1 trp1-1 leu2-3,112 can1-100 his3-11,15::PADH1-397-OsTIR1 SPP1::SPP1-AID*-6FLAG-HygMX LEA1::LEA1-3HA-KAN*	This study
YBRT Prp39-AID*-6FLAG Prp22-9MYC	*MATα ade2-1 ura3-1 trp1-1 leu2-3,112 can1-100 his3-11,15::PADH1-397-OsTIR1 PRP39::PRP39-AID*-6FLAG-HygMX PRP22::PRP22-AID-9MYC-KAN*	Ema Sani (Beggs lab)/this study

YBRT Prp9-AID*-6FLAG Prp22-9MYC	*MATα ade2-1 ura3-1 trp1-1 leu2-3,112 can1-100 his3-11,15::PADH1-397-OsTIR1 PRP9::PRP9-AID*-6FLAG-HygMX PRP22::PRP22-AID-9MYC-KAN*	Ema Sani (Beggs lab)/this study
YBRT Slu7-AID*-6FLAG Prp22-9MYC	*MATα ade2-1 ura3-1 trp1-1 leu2-3,112 can1-100 his3-11,15::PADH1-397-OsTIR1 Slu7::Slu7-AID*-6FLAG-HygMX PRP22::PRP22-AID-9MYC-KAN*	Ema Sani (Beggs lab)/this study
H4759 Spt5 WT Lea1-V5	*MATα his3-del'1 leu2-del'0 met15-del'0 ura3-del'0 bur1-as PAF1-13XMYC::HIS3 spt5-del'::NatMX pHQ1494 [SPT5-3XHA, LEU2, CEN4] LEA1::LEA1-V5-HygMX*	Qiu *et al*., 2012 Kind gift from the Hinnebusch lab
H4756 Spt5 S->A Lea1-V5	*MATα his3-del'1 leu2-del'0 met15-del'0 ura3-del'0 bur1-as PAF1-13XMYC::HIS3 spt5-del'::NatMX pHQ1494 [spt5-S1-15A-3XHA, LEU2, CEN4] LEA1-V5-HygMX*	Qiu *et al*., 2012 Kind gift from the Hinnebusch lab
H4758 Spt5 S->E Lea1-V5	*MATα his3-del'1 leu2-del'0 met15-del'0 ura3-del'0 bur1-as PAF1-13XMYC::HIS3 spt5-del'::NatMX pHQ1494 [spt5-S1-15E-3XHA, LEU2, CEN4], LEU2, CEN4] LEA1-V5-HygMX*	Qiu *et al*., 2012 Kind gift from the Hinnebusch lab
YBRT Prp22-AID*-6FLAG 9MYC-Set1 pBest-PRP22WT-LEU2	*MATα ade2-1 ura3-1 trp1-1 leu2-3,112 can1-100 his3-11,15::PADH1-397-OsTIR1 PRP22::PRP22-AID*-6FLAG-HygMX SET1::9MYCSET1-TRP1 [pBest-PRP22WT-LEU2]*	This study
YBRT Prp22-AID*-6FLAG 9MYC-Set1 pBest-PRP22T757A-LEU2	*MATα ade2-1 ura3-1 trp1-1 leu2-3,112 can1-100 his3-11,15::PADH1-397-OsTIR1 PRP22::PRP22-AID*-6FLAG-HygMX SET1::9MYCSET1-TRP1 [pBest-PRP22T757A-LEU2]*	This study
YBRT Prp22-AID*-6FLAG 9MYC-Set1 pBest-PRP22I764A-LEU2	*MATα ade2-1 ura3-1 trp1-1 leu2-3,112 can1-100 his3-11,15::PADH1-397-OsTIR1 PRP22::PRP22-AID*-6FLAG-HygMX SET1::9MYCSET1-TRP1 [pBest-PRP22I764A-LEU2]*	This study
YBRT Prp22-AID*-6FLAG pRS426–FUI1	*MATα ade2-1 ura3-1 trp1-1 leu2-3,112 can1-100 his3-11,15::PADH1-397-OsTIR1 PRP22::PRP22-AID*-6FLAG-HygMX [pRS426–FUI1-URA3]*	This study

Appendix 2. Oligonucleotides used for strain construction and cloning

Name	Sequence (5'-3')	Use
Spt5-AID-F	GCTTGGAACAACCAAGGAAATAAGTCAAACTATG GTGGTAACAGTACATGGGGAGGTCATCGTACGC TGCAGGTCGAC	Forward primer to AID-degron tag Spt5
Spt5-AID-R	TAATATTAAAGTCTTTTTTATTGATTTCTTCTTGG GTGATATTGGTTCTCCTTTTGGTGAATCGATGAA TTCGAGCTCG	Reverse primer to AID-degron tag Spt5
Spt5-CF	GGGGTGGCCAAGGTAATGGAGG	To check tagging of Spt5
Spt5-CR	ACATTACAATAATCAACTCCCTC	To check tagging of Spt5
Bur1-AID-F	GAAAGAGCGAAACCAGATGAGTCTAAGGAGTTC CAAAATAGTGATATTGCAGATCTATATCGTACGC TGCAGGTCGAC	Forward primer to AID-degron tag Bur1
Bur1-AID-R	TCAGTATACAACTTTCCGTAATTAGCCACGAGGC CAGAAAGGAAGAGAGAATAGTATAACATCGATG AATTCGAGCTCG	Reverse primer to AID-degron tag Bur1
BUR1-CF	CGGTTCTCGCAATATGGCTGGT	To check tagging of Bur1
Bur1-CR	ATCCGCGCAGGATAACGC	To check tagging of Bur1
Bur2-AID-F	AGTATGAAGAGAAAGGCTAAAGATCCTATAAGAA CCCCAGATGCCAAAAAACCTAAAATACGTACGCT GCAGGTCGAC	Forward primer to AID-degron tag Bur2
Bur2-AID-R	GTCACGTATATATCAAACCTTTACTGATCCCTCC AATTAAACATAACTTGTACTCTATTTATCGATGAA TTCGAGCTCG	Reverse primer to AID-degron tag Bur2
Bur2-CF	CCCTATCGAAGTAACGCCGGA	To check tagging of Bur2
Bur2-CR	GCAGAAACTGCGCTTTGACTCACC	To check tagging of Bur2
CTK1-AID-F	AATAAGGGTAATGGTAATAGTAATAATAATAATAA TAATAATAATGACGATGATGATAAACGTACGCTG CAGGTCGAC	Forward primer to AID-degron tag Ctk1
CTK1-AID-R	GTAATAAATAAGTTATTAATCTATTTTTTGTGTCT ACTTATTTCAATTGGCTATATATCCATCGATGAAT TCGAGCTCG	Reverse primer to AID-degron tag Ctk1
CTK1-CF	GCTGGCCAACGCTTTACGATAT	To check tagging of Ctk1
CTK1-CR	GGTAAGACAGTTAGTGGCCAGGC	To check tagging of Ctk1
Paf1-AID-F	TCAGATGCTGTTCATACTGAACAAAAACCAGAGG AAGAAAAGGAAACTTTACAAGAAGAACGTACGCT GCAGGTCGAC	Forward primer to AID-degron tag Paf1
Paf1-AID-R	CCAAACAAATGTAAAAAGAACTACAGGTTTAAAA TCAATCTCCCTTCACTTCTCAATATTATGAATTCG AGCTCG	Reverse primer to AID-degron tag Paf1
Paf1-CF	GAAGGCGATTCAAAAACAGAAGGT	To check tagging of Paf1
Paf1-CR	GTGTAACTGCTGGTGGCTGG	To check tagging of Paf1
Lea1-TAG-F	CTTCTTTAGAAGAGATTGCCAGGCTGGAAAAACT ACTCTCTGGTGGTGTTCGGATCCCCGGGTTAAT TAA	Forward primer to epitope tag Lea1

Lea1-TAG-R	TTTATAATTCTTTTTTTTTAAGTCATTGAACAGTCGCACTAACCAAAAGAATCGATGAATTCGAGCTCG	Reverse primer to epitope tag Lea1
Lea1-CF	CGATGGAGATCATGAATTTGG	To check tagging of Lea1
Lea1-CR	GAGCACCATTATTTGTTTTCGTT	To check tagging of Lea1
PRP22-TAG-F	GACTAAGCTCAATAAGGCAGTCAAGGGAAAGGGCATTAGGTATCAAGAGGCGGATCCCCGGGTTAATTAA	Forward primer to epitope tag Prp22
PRP22-TAG-R	ATATAGGTCTATAAAACTCGATAATTATAATGCATAAAAAGCTAACAATGGAATTCGAGCTCGTTTAAAC	Reverse primer to epitope tag Prp22
PRP22-CF	ATTAACTAGCCGGGAATACA	To check tagging of Prp22
PRP22-CR	AATTGGTGTTCTTCGCGAAT	To check tagging of Prp22
Spp1-Tag-F	AAAAGCAACTAAATATACAATACTATGAGGAAATTTTAAGAAGAGGTTTGCGGATCCCCGGGTTAATTAA	Forward primer to epitope tag Spp1
Spp1-Tag-R	CGAAGTATATATATATGTAGAAACTGATATTTGATTAGGCTCCAACGCCGATCGATGAATTCGAGCTCG	Reverse primer to epitope tag Spp1
SPP1-AID*-F	AAAAGCAACTAAATATACAATACTATGAGGAAATTTTAAGAAGAGGTTTGCGTACGCTGCAGGTCGAC	Forward primer to AID-degron tag Spp1
SPP1-AID*-R	CGAAGTATATATATATGTAGAAACTGATATTTGATTAGGCTCCAACGCCGATCGATGAATTCGAGCTCG	Reverse primer to AID-degron tag Spp1
Spp1-CF	CTACATACGAGCGAATACCC	To check tagging of Spp1
Spp1-CR	AAAACTATACTCCTAAGTGT	To check tagging of Spp1
Set1-CF	ACAATGTTGGACTTGTTGCA	To check tagging of Set1
Set1-CR	ATGGCCGAGATTACTTATAG	To check tagging of Set1
Bur1-TAG-F	GAAAGAGCGAAACCAGATGAGTCTAAGGAGTTCCAAAATAGTGATATTGCAGATCTATATCGGATCCCCGGGTTAATTAA	Forward primer to epitope tag Bur1
Bur1-TAG-R	TCAGTATACAACTTTCCGTAATTAGCCACGAGGCCAGAAAGGAAGAGAGAATAGTATAACATCGATGAATTCGAGCTCG	Reverse primer to epitope tag Bur1
Rpb3-TAG-	CCAAACGAAGGTGACCCCTTCG	Forward primer to TAP tag RPB3 (designed by Dr Susana de Lucas, Beggs lab)
Rpb3-TAG-R	AAGCAAGGTTAGCCCTTCCTTTC	Reverse primer to TAP tag RPB3 (Designed by Dr Susana de Lucas, Beggs lab)
Prp22-SFI-F	GGGAATTCGGCCATTCGGGCcATGTCTGATATATCGAAACT	Forward primer for Gibson assembly to amplify Prp22 from p360-PRP22-V5,

		p360-T757A-V5 or p360-I764A-V5 and adding SfiI restriction sites
Prp22-SFI-R	GAATTCCCGGCCGTACTGGCCGAATATCTCTATTGTTCTCTCT	Reverse primer for Gibson assembly to amplify Prp22 from p360-PRP22-V5, p360-T757A-V5 or p360-I764A-V5 and adding SfiI restriction sites

Appendix 3. Table of plasmids used in this study

Name	Description	Source
pHYG-AID*-6FLAG	To C-terminally AID*-6FLAG tag a gene-of-interest (HYG selection)	Ulrich lab (Morawska and Ulrich, 2013)
pHYG-AID*-9MYC	To C-terminally AID*-9MYC tag a gene-of-interest (HYG selection)	Ulrich lab (Morawska and Ulrich, 2013)
pHYG-AID*-6HA	To C-terminally AID*-6HA tag a gene-of-interest (HYG selection)	Ulrich lab (Morawska and Ulrich, 2013)
pZTRL	To express OsTIR1 using the B-estradiol system on a centromeric plasmid (LEU2 selection)	Gonzalo Mendoza-Ochoa (Beggs lab) (Mendoza-Ochoa et al., 2018)
pRS426–FUI1	Yeast uracil permease (FUI1) on the 2 micron plasmid pRS425 (URA3)	David Barrass (Beggs lab) (Barrass et al., 2015)
p360-PRP22-V5	C-terminally V5 tagged version of WT PRP22 in a centromeric plasmid (TRP selection)	Prof. Beate Schwer (Schneider *et al*., 2004) and Gonzalo Mendoza-Ochoa (Beggs lab)
p360-T757A-V5	C-terminally V5 tagged version of T757A PRP22 in a centromeric plasmid (TRP selection)	Prof. Beate Schwer (Schneider *et al*., 2004) and Gonzalo Mendoza-Ochoa (Beggs lab)
p360-I764A-V5	C-terminally V5 tagged version of I764A PRP22 in a centromeric plasmid (TRP selection)	Prof. Beate Schwer (Schneider *et al*., 2004) and Gonzalo Mendoza-Ochoa (Beggs lab)
pZL-PRP22WT-V5	C-terminally V5 tagged version of WT PRP22 in a centromeric plasmid that allows β-estradiol inducible expression (LEU2 selection) Derived from pZTRL and p360-PRP22-V5	This study
pZL-PRP22T757A-V5	C-terminally V5 tagged version of T757A PRP22 in a centromeric plasmid that allows β-estradiol inducible expression (LEU2 selection). Derived from pZTRL and p360-T757A-V5	This study
pZL-PRP22I764A-V5	C-terminally V5 tagged version of I764A PRP22 in a centromeric plasmid that allows β-estradiol inducible expression (LEU2 selection). Derived from pZTRL and p360-I764A-V5	This study
pFA6a-6xGLY-13MYC-kanMX6	To C-terminally 13XMYC tag a gene-of-interest (KAN selection)	Mark Hochstrasser (Addgene plasmid #20769) PubMed 19243080
pFA6a-3HA-kanMX6	To C-terminally 3XHA tag a gene-of-interest (KAN selection)	Jurg Bahler & John Pringle (Addgene plasmid #39295) PubMed 9717240
pFA6a-6xGLY-V5-HIS3MX6	To C-terminally 3XV5 tag a gene-of-interest (KAN selection)	Mark Hochstrasser (Addgene plasmid #20777)

		PubMed 19243080
p9MycSET1-TRP1	To N-terminally 9XMYC tag a Set1 (TRP selection)	Vincent Géli (CNRS, Marseille) (Dehé *et al*., 2006)

Appendix 4. Table of oligonucleotides used for Splicing RT-qPCR

Name	Sequence (5'-3')	Use
ALG9_F	TAAGCTGGCATGTGCTGCATTC	To detect *ALG9* mRNA using RT-qPCR
ALG9_R	TTTGCATGATTCGGTTGATTGG	To detect *ALG9* mRNA using RT-qPCR
SCR1-F	TCTCTGTCTGGTGCGGCAAG	To detect SCR1 mRNA using RT-qPCR
SCR1-R	TCACGGGTCACCTTTGCTGA	To detect *SCR1* mRNA using RT-qPCR
YFL039C_ACT_BL_F	AGGGGCTTGAAATTTGGAAAAA	To detect the *ACT1* BP/ lariat using RT-qPCR
YFL039C_ACT_B_R	GCAACAAAAAGAATGAAGCAATCG	To detect the *ACT1* BP using RT-qPCR
YFL039C_ACT_L_R	GCAAGCGCTAGAACATACATAGTACA	To detect the *ACT1* lariat using RT-qPCR
YFL039C_ACT_3_F	TTGCTTCATTCTTTTTGTTGCT	To detect *ACT1* 3'SS using RT-qPCR
YFL039C_ACT_3m_R	GCAAAACCGGCTTTACACAT	To detect *ACT1* 3'SS/mRNA using RT-qPCR
YFL039C_ACT_m_F	CTGCTTTTTTCTTCCCAAGATCGA	To detect *ACT1* mRNA using RT-qPCR
YFL039C_ACT_E_F	GCTGCTTTGGTTATTGATAACGGTTC	To detect *ACT1* exon 2 using RT-qPCR
YFL039C_ACT_E_R	GATGGGAAGACAGCACGAGGAG	To detect *ACT1* exon 2 using RT-qPCR
YGL103W_RPL28_V_R	TTGGTTCTTTCATTCCCTCTTCCA	To detect *RPL28* 5'SS using RT-qPCR
YGL103W_RPL28_Vm_F	TCCAGATTCACTAAGACTAGAAAGCACAGA	To detect *RPL28* 5'SS/mRNAusing RT-qPCR
YGL103W_RPL28_m_R	TGACCACCGGCCATACCTCT	To detect *RPL28* mRNA using RT-qPCR
YGL103W_RPL28_3_F	TTTTTGTACAGCCGGTAAAGGTCGT	To detect *RPL28* 3'SS using RT-qPCR
YGL103W_RPL28_3_R	GATGTTGACCACCGGCCATAC	To detect *RPL28* 3'SS using RT-qPCR
YGL103W_RPL28_E_F	AGAGGTATGGCCGGTGGTCA	To detect *RPL28* exon 2 using RT-qPCR
YGL103W_RPL28_E_R	CAGAAATGAGCTTGTTGCTTGTGG	To detect *RPL28* exon 2 using RT-qPCR
YGL103W_RPL28_L_F	GAGCGCAATTATGAAAAAGAGTTACCA	To detect *RPL28* lariat using RT-qPCR
YGL103W_RPL28_L_R	TTCCAAATGGAACTACATACATAGTAAAACAG	To detect *RPL28* lariat using RT-qPCR
YDR064W_RPS13_V_F	TCGTATGCACAGTGCCGTATGTT	To detect *RPS13* 5'SS using RT-qPCR
YDR064W_RPS13_V_R	TGATTTAGCGAACTATTCAATGCAACTTT	To detect *RPS13* 5'SS using RT-qPCR
YDR064W_RPS13_3_F	TCCAATTCCACTAAATATTACTTTAAACAGGGTA	To detect *RPS13* 3'SS using RT-qPCR
YDR064W_RPS13_3_R	CTTGAACCAAGCTGGAGCATTTCT	To detect *RPS13* 3'SS using RT-qPCR

YDR064W_RPS13_m_F	TCGTATGCACAGTGCCGGTAA	To detect *RPS13* mRNA using RT-qPCR
YDR064W_RPS13_m_R	AGGACAACTTGAACCAAGCTGGAG	To detect *RPS13* mRNA using RT-qPCR
YDR064W_RPS13_E_F	CTAGAAATGCTCCAGCTTGGTTCAA	To detect *RPS13* exon 2 using RT-qPCR
YDR064W_RPS13_E_R	TCAAACCCTTTCTCGCGTACTTG	To detect *RPS13* exon 2 using RT-qPCR
YBR078W_ECM33_V_F	AGTGCCTCCGCTCTAGCTGGT	To detect *ECM33* 5'SS using RT-qPCR
YBR078W_ECM33_V_R	CGAGATTTGTGAGGAAAGAGGCAAA	To detect *ECM33* 5'SS using RT-qPCR
YBR078W_ECM33_3_F	TCTAAGCAGCTAACTCAACTACTTCTATTCCATC	To detect *ECM33* 3'SS using RT-qPCR
YBR078W_ECM33_3_R	TTTTGTCCAAATCAGCTTGAGCAGT	To detect *ECM33* 3'SS using RT-qPCR
YBR078W_ECM33_m_F	GCCTCCGCTCTAGCTGCTAACTC	To detect *ECM33* mRNA using RT-qPCR
YBR078W_ECM33_m_R	TTGAGCAGTAGCAGTGGCAGAAGT	To detect *ECM33* mRNA using RT-qPCR
YBR078W_ECM33_E_F	CTTCTGCCACTGCTACTGCTCAAG	To detect *ECM33* exon 2 using RT-qPCR
YBR078W_ECM33_E_R	AGCAGCGGAACCCAAGTCAC	To detect *ECM33* exon 2 using RT-qPCR

Appendix 5. Table of oligonucleotides used for ChIP-qPCR

Name	Sequence (5'-3')	Use
YDR064W_RPS13_5lG_F	CCCATAAACCATAAAGTAGACCCAAACA	To amplify the promoter of *RPS13* (amplicon 1) using qPCR
YDR064W_RPS13_5lG_R	TCATTCGTATCAAAATTTCACCGCTAC	To amplify the promoter of *RPS13* (amplicon 1) using qPCR
YDR064W_RPS13_Pm_F	AGTCGTGATTGAATTAACAATTTCTTTCTCA	To amplify *RPS13* (amplicon 2) using qPCR
YDR064W_RPS13_Pm_R	GCACTGTGCATACGACCCATTTT	To amplify *RPS13* (amplicon 2) using qPCR
YDR064W_RPS13_I_F	GCTGGGTGATTCCAATTTCTTTTACA	To amplify *RPS13* (amplicon 3) using qPCR
YDR064W_RPS13_I_R	CATAAAGGCGGCTAGCCATCAG	To amplify *RPS13* (amplicon 3) using qPCR
YDR064W_RPS13_3_F	TCCAATTCCACTAAATATTACTTTAAACAGGGTA	To amplify *RPS13* (amplicon 4) using qPCR
YDR064W_RPS13_3_R	CTTGAACCAAGCTGGAGCATTTCT	To amplify *RPS13* (amplicon 4) using qPCR
YDR064W_RPS13_Eb_F	TTCACAGATTGGCCAGATACTACAGAAC	To amplify *RPS13* (amplicon 5) using qPCR
YDR064W_RPS13_Eb_R	TTGACCAAAGCGGAGGCAGT	To amplify *RPS13* (amplicon 5) using qPCR
RPS13 +1288_F	GGACCGTTCTCAGAAACATTCCA	To amplify *RPS13* (amplicon 6) using qPCR
RPS13 +1400_R	CCAACTTGACTTGTCATGCTTGTGT	To amplify *RPS13* (amplicon 6) using qPCR
Act1 -139 F	TTCCCCTTTCTACTCAAACCAAGAAG	To amplify the promoter of *ACT1* (amplicon 1) using qPCR

Act1 -43 R	AAGCGTGAAAAATCTAAAAGCTGATG	To amplify the promoter of *ACT1* (amplicon 1) using qPCR
Act1 -71 (5SS) F	TACATCAGCTTTTAGATTTTTCACGCTTACTGCTT	To amplify the first exon of *ACT1* (amplicon 2) using qPCR
Act1 +34 (5SS) R	GATGGTGCAAGCGCTAGAACATACCAGAAT	To amplify the first exon of *ACT1* (amplicon 2) using qPCR
Act1 +368 (3SS) F	TGTACTAACATCGATTGCTTCATTCTTTTTGTTGC	To amplify the 3'ss of *ACT1* (amplicon 3) using qPCR
Act1 +413 (3SS) R	GACGATAGATGGGAAGACAGCACGAGGA	To amplify the 3'ss of *ACT1* (amplicon 3) using qPCR
Act1 +561 F	ATCTGGCATCATACCTTCTACAACGA	To amplify a amplicon in the second exon of *ACT1* (amplicon 4) using qPCR
Act1 +653 R	GTTTGATTTAGGGTTCATTGGAGCTT	To amplify a amplicon in the second exon of *ACT1* (amplicon 4) using qPCR
Act1 +1119 F	TCTGCCGGTATTGACCAAACTACTTA	To amplify a amplicon in the second exon of *ACT1* (amplicon 5) using qPCR
Act1 +1210 R	CCGGACATAACGATGTTACCGTATAA	To amplify a amplicon in the second exon of *ACT1* (amplicon 5) using qPCR
Act1 +1595 F	ATGTGTTTTGTCTCTCCCTTTTCTACGAAAATTTC	To amplify a amplicon in the 3' UTR of *ACT1* (amplicon 6) using qPCR
Act1 +1734 R	TGATCATATGATACACGGTCCAATGGATAAACAT	To amplify a amplicon in the 3' UTR of *ACT1* (amplicon 6) using qPCR
ECM33 -592_F	GCAGTATCATCCTTCACGACCC	To amplify the promoter of *ECM33* (amplicon 1) using qPCR
ECM33 -510_R	GCGTCTTTCCCGTTTTTGC	To amplify the promoter of

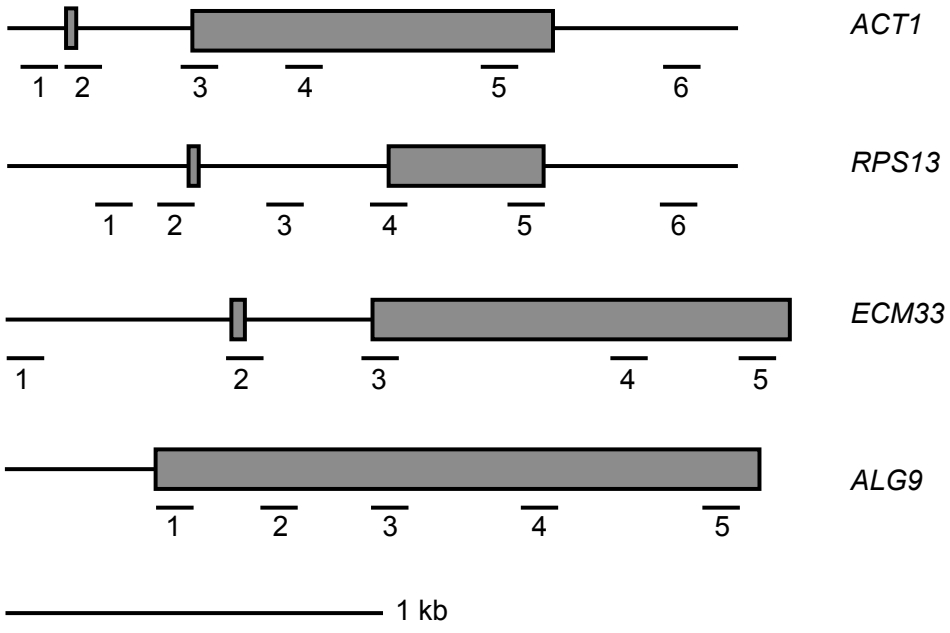
		ECM33 (amplicon 1) using qPCR
ECM33 +9_F	CAAGAACGCTTTGACTGCTACTG	To amplify the 5'ss of *ECM33* (amplicon 2) using qPCR
ECM33 +145 _R	GAAGAGGACCACGAATCTACTCG	To amplify the 5'ss of *ECM33* (amplicon 2) using qPCR
ECM33 +430_F	ACTTCTGCCACTGCTACTGCTC	To amplify the 3'ss of *ECM33* (amplicon 3) using qPCR
ECM33 +562_R	AGGAACCATCAATCTCTTGGATAC	To amplify the 3'ss of *ECM33* (amplicon 3) using qPCR
ECM33 +1073_F	TTGGTCAATCTTTGTCTATCGTCTC	To amplify a amplicon in the second exon of *ECM33* (amplicon 4) using qPCR
ECM33 +1173_R	TGTGTTGTTAGCAATGATGAAACC	To amplify a amplicon in the second exon of *ECM33* (amplicon 4) using qPCR
ECM33 +1531_F	TCTAAGAAGTCTAAGGGTGCTGCTC	To amplify a amplicon in the second exon of *ECM33* (amplicon 5) using qPCR
ECM33 +1582_R	TGAATGAAGTGGCTGGAACAAG	To amplify a amplicon in the second exon of *ECM33* (amplicon 5) using qPCR
ALG9+3F	GAATTGCAAGGCGGTAACCA	To amplify *ALG9* (amplicon 1) using qPCR
ALG9+86R	GAAATTAACGAGAATGTCGGCTGAA	To amplify *ALG9* (amplicon 1) using qPCR
ALG9+326F	TGAAATTGCAGGCAGCTTGG	To amplify *ALG9* (amplicon 2) using qPCR
ALG9+437R	GCAACGGCAGAAGGCAATAA	To amplify *ALG9* (amplicon 2) using qPCR

ALG9+630F	CGCATTCGACTGCTGTTTGA	To amplify *ALG9* (amplicon 3) using qPCR
ALG9+715R	CAGGAGCAAGCTTCCCGTAA	To amplify *ALG9* (amplicon 3) using qPCR
ALG9+1079F	GCCAGGCAATGTCACGGATA	To amplify *ALG9* (amplicon 4) using qPCR
ALG9+1118F	GGTGCCTTCACACCACCTTG	To amplify *ALG9* (amplicon 4) using qPCR
ALG9+1486F	AAGCTGGCATGTGCTGCATT	To amplify *ALG9* (amplicon 5) using qPCR
ALG9+1567R	GGTTGATTGGCTCCGGTACG	To amplify *ALG9* (amplicon 5) using qPCR
RPL28_P_F	GTACTGTAACCAATGTAACATCTGT	To amplify *RPL28* (amplicon 1) using qPCR
RPL28_P_R	AAAAACTGCGAGGAAAATTGTTTG	To amplify *RPL28* (amplicon 1) using qPCR
RPL28_7_F	TCCAGATTCACTAAGACTAGAAAGCACAGA	To amplify *RPL28* (amplicon 2) using qPCR
RPL28_88_R	TTGGTTCTTTCATTCCCTCTTCCA	To amplify *RPL28* (amplicon 2) using qPCR
RPL28_605_F	AGAGGTATGGCCGGTGGTCA	To amplify *RPL28* (amplicon 3) using qPCR
RPL28_717_R	CAGAAATGAGCTTGTTGCTTGTGG	To amplify *RPL28* (amplicon 3) using qPCR
RPL28_817_F	TGACACTTTGGCAGCCGGTTACG	To amplify *RPL28* (amplicon 4) using qPCR
RPL28_914_R	CAGCCAACTTGGAGACGAATCTAGCT	To amplify *RPL28* (amplicon 4) using qPCR
RPL28_1032_F	ACTTCTGTATTATTTCTATCACCCATTTTCACCGT	To amplify *RPL28* (amplicon 5) using qPCR
RPL28_1114_R	AGCGTTGGGATCTGCCACTGG	To amplify *RPL28*

		(amplicon 5) using qPCR
RPL28_1208_F	AAAGACCAGGAAAAACAAAATGACTGCAGTT	To amplify *RPL28* (amplicon 6) using qPCR
RPL28_1293_R	TCGAAAGAATTGCCAAAATCTCCCTAGGTATATT	To amplify *RPL28* (amplicon 6) using qPCR
FMP27 -427 F	TTCAATTGATCAAATTTATGGAAGATCCTCAAGAA	To amplify the promoter of *FMP27* (amplicon 1) using qPCR
FMP27 -336 R	GGAACAAAAACTTGACAAATTTGAAACTCTGGAT	To amplify the promoter of *FMP27* (amplicon 1) using qPCR
FMP27 -44 F	GAACATAAGAATCCTTAGAAAAGCCCTTTACCTCG	To amplify a amplicon in the exon of *FMP27* (amplicon 2) using qPCR
FMP27 +63 R	CCATAAGAAAGTCACTGCAAATATAAGCCACTTGT	To amplify a amplicon in the exon of *FMP27* (amplicon 2) using qPCR
FMP27 +474 F	AAGATTTGATTCCTTTTTGAGAAAACTTCTTTGGA	To amplify a amplicon in the exon of *FMP27* (amplicon 3) using qPCR
FMP27 +593 R	CCATCCTTCAGAGGATTCATAATTTCACCAATT	To amplify a amplicon in the exon of *FMP27* (amplicon 3) using qPCR
FMP27 +1051 F	TTGAATCTAAATCGAAAACATCAAAGCCACG	To amplify a amplicon in the exon of *FMP27* (amplicon 4) using qPCR
FMP27 +1154 R	AATTTTTGAGAGAACAAATTGGTTTCGCCA	To amplify a amplicon in the exon of *FMP27* (amplicon 4) using qPCR
FMP27 +4525 F	AGACCTAGTACCAATACAATGTTCATTCCAAACCA	To amplify a amplicon in the exon of *FMP27* (amplicon 5) using qPCR

FMP27 +4642 R	CCTTGTCTGCTTTTTCGTTTTTACTTGATG TAGTG	To amplify a amplicon in the exon of *FMP27* (amplicon 5) using qPCR
FMP27 +7589 F	TGAACAGCTTCAAACTTTGTATCAGTTATA AGGGC	To amplify a amplicon in the exon of *FMP27* (amplicon 6) using qPCR
FMP27 +7673 R	GGGAAATTGAAAACAAAGTTAGTAACGTT AGCCAA	To amplify a amplicon in the exon of *FMP27* (amplicon 6) using qPCR

Appendix 6. Gene maps showing positions of amplicons used for ChIP-qPCR



Reference list

Adelman K & Lis JT (2012) Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. *Nat. Rev. Genet.* **13:** 720–731

Ahn SH, Kim M & Buratowski S (2004) Phosphorylation of Serine 2 within the RNA Polymerase II C-Terminal Domain Couples Transcription and 3′ End Processing. *Mol. Cell* **13:** 67–76

Alexander RD, Barrass JD, Dichtl B, Kos M, Obtulowicz T, Robert MC, Koper M, Karkusiewicz I, Mariconti L, Tollervey D, Dichtl B, Kufel J, Bertrand E & Beggs JD (2010a) 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

Alexander RD, Innocente SA, Barrass JD & Beggs JD (2010b) Splicing-Dependent RNA polymerase pausing in yeast. *Mol. Cell* **40:** 582–593

Allemand E, Myers MP, Garcia-Bernardo J, Harel-Bellan A, Krainer AR & Muchardt C (2016) A Broad Set of Chromatin Factors Influences Splicing. *PLoS Genet.* **12:**

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

Amit M, Donyo M, Hollander D, Goren A, Kim E, Gelfman S, Lev-Maor G, Burstein D, Schwartz S, Postolsky B, Pupko T & Ast G (2012) Differential GC Content between Exons and Introns Establishes Distinct Strategies of Splice-Site Recognition. *Cell Rep.* **1:** 543–556

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

Ansari A & Hampsey M (2005) A role for the CPF 3′ -end processing machinery in RNAP II-dependent gene looping. *Genes Dev.* **19:** 2969–2978

Ardehali MB & Lis JT (2009) Tracking rates of transcription and splicing in vivo. *Nat. Struct. Mol. Biol.* **16:** 1123–1124

Ares M. J, Grate L & Pauling MH (1999) A handful of intron-containing genes produces the lion's share of yeast mRNA [2]. *Rna* **5:** 1138–1139

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

Aslanzadeh V, Huang Y, Sanguinetti G & Beggs JD (2018) Transcription rate strongly affects splicing fidelity and cotranscriptionality in budding yeast. *Genome Res.* **28:** 203–213

Auboeuf D, Hönig A, Berget SM & O'Malley BW (2002) Coordinate regulation of transcription and splicing by steroid receptor coregulators. *Science.* **298:** 416–419

Bannister AJ & Kouzarides T (2011) Regulation of chromatin by histone modifications. *Cell Res.* **21:** 381–395

Barboric M, Lenasi T, Chen H, Johansen EB, Guo S & Peterlin BM (2009) 7SK snRNP/P-TEFb couples transcription elongation with alternative splicing and is essential for vertebrate development. *Proc. Natl. Acad. Sci. U. S. A.* **106:** 7798–7803

Barrass JD, Reid JEA, Huang Y, Hector RD, Sanguinetti G, Beggs JD & Granneman S (2015) Transcriptome-wide RNA processing kinetics revealed using extremely short 4tU labeling. *Genome Biol.* **16:**282

Bartkowiak B, MacKellar AL & Greenleaf AL (2011) Updating the CTD Story: From Tail to Epic. *Genet. Res. Int.* **2011:** 1–16 Available at: http://www.hindawi.com/journals/gri/2011/623718/

Bartolomei MS, Halden NF, Cullen CR & Corden JL (1988) Genetic analysis of the repetitive carboxyl-terminal domain of the largest subunit of mouse RNA polymerase II. *Mol. Cell. Biol.* **8:** 330–9

Baskaran R, Chiang GG, Mysliwiec T, Kruh GD & Wang JYJ (1997) Tyrosine phosphorylation of RNA polymerase II carboxyl-terminal domain by the Abl-related gene product. *J. Biol. Chem.* **272:** 18905–18909

Batsché E, Yaniv M & Muchardt C (2006) The human SWI/SNF subunit Brm is a regulator of alternative splicing. *Nat. Struct. Mol. Biol.* **13:** 22–29

Battaglia S, Lidschreiber M, Baejen C, Torkler P, Vos SM & Cramer P (2017) RNA-dependent chromatin association of transcription elongation factors and pol II CTD kinases. *Elife* **6:**

Benayoun BA, Pollina EA, Ucar D, Mahmoudi S, Karra K, Wong ED, Devarajan K, Daugherty AC, Kundaje AB, Mancini E, Hitz BC, Gupta R, Rando TA, Baker JC, Snyder MP, Cherry JM & Brunet A (2014) H3K4me3 breadth is linked to cell identity and transcriptional consistency. *Cell* **158:** 673–688

Bentley DL (2005) Rules of engagement: Co-transcriptional recruitment of pre-mRNA processing factors. *Curr. Opin. Cell Biol.* **17:** 251–256

Bentley DL (2014) Coupling mRNA processing with transcription in time and space. *Nat. Rev. Genet.* **15:** 163–175

Berget SM, Moore C & Sharp PA (1977) Spliced segments at the 5′ terminus of adenovirus 2 late mRNA. *Proc. Natl. Acad. Sci.* **74:** 3171–3175

Bertram K, Agafonov DE, Liu WT, Dybkov O, Will CL, Hartmuth K, Urlaub H, Kastner B, Stark H & Lührmann R (2017) Cryo-EM structure of a human spliceosome activated for step 2 of splicing. *Nature* **542:** 318–323

Betz JL, Chang M, Washburn TM, Porter SE, Mueller CL & Jaehning JA (2002) Phenotypic analysis of Paf1/RNA polymerase II complex mutations reveals connections to cell cycle regulation, protein synthesis, and lipid and nucleic acid metabolism. *Mol. Genet. Genomics* **268:** 272–285

Beyer AL, Bouton AH & Miller OL (1981) Correlation of hnRNP structure and nascent transcript cleavage. *Cell* **26:** 155–165

Bieberstein NI, Oesterreich FC, Straube K & Neugebauer KM (2012) First exon length controls active chromatin signatures and transcription. *Cell Rep.* **2:** 62–68

Bitton DA, Atkinson SR, Rallis C, Smith GC, Ellis DA, Chen YYC, Malecki M, Codlin S, Lemay JF, Cotobal C, Bachand F, Marguerat S, Mata J & Bähler J (2015) Widespread exon skipping triggers degradation by nuclear RNA surveillance in fission yeast. *Genome Res.* **25:** 884–896

Bon E, Casaregola S, Blandin G, Llorente B, Neuvéglise C, Munsterkotter M, Guldener U, Mewes HW, Van Helden J, Dujon B & Gaillardin C (2003) Molecular evolution of eukaryotic genomes: Hemiascomycetous yeast spliceosomal introns. *Nucleic Acids Res.* **31:** 1121–1135

Borde V, Robine N, Lin W, Bonfils S, Géli V & Nicolas A (2009) Histone H3 lysine 4 trimethylation marks meiotic recombination initiation sites. *EMBO J.* **28:** 99–111

Brès V, Gomes N, Pickle L & Jones KA (2005) A human splicing factor, SKIP, associates with P-TEFb and enhances transcription elongation by HIV-1 Tat. *Genes Dev.* **19:** 1211–1226

Braberg H, Jin H, Moehle EA, Chan YA, Wang S, Shales M, Benschop JJ, Morris JH, Qiu C, Hu F, Tang LK, Fraser JS, Holstege FCP, Hieter P, Guthrie C, Kaplan CD & Krogan NJ (2013) From structure to systems: High-resolution, quantitative genetic analysis of RNA polymerase II. *Cell* **154:** 775–788

Bresson S & Tollervey D (2018) Surveillance-ready transcription: Nuclear RNA decay as a default fate. *Open Biol.* **8:**

Briggs SD, Bryk M, Strahl BD, Cheung WL, Davie JK, Dent SYR, Winston F & David Allis C (2001) Histone H3 lysine 4 methylation is mediated by Set1 and required for cell growth and rDNA silencing in Saccharomyces cerevisiae. *Genes Dev.* **15:** 3286–3295

Brinster RL, Allen JM, Behringer RR, Gelinas RE & Palmiter RD (1988) Introns increase transcriptional efficiency in transgenic mice. *Proc. Natl. Acad. Sci.* **85:** 836–840

Brodsky AS, Meyer C a, Swinburne I a, Hall G, Keenan BJ, Liu XS, Fox E a & Silver P a (2005) Genomic mapping of RNA polymerase II reveals sites of co-transcriptional regulation in human cells. *Genome Biol.* **6:** R64

Brogaard K, Xi L, Wang JP & Widom J (2012) A map of nucleosome positions in yeast at base-pair resolution. *Nature* **486:** 496–501

Brugiolo M, Herzel L & Neugebauer KM (2013) Counting on co-transcriptional splicing. *F1000Prime Rep.* **5:** Available at: http://www.f1000.com/prime/reports/b/5/9

Buratowski S (2003) The CTD code. *Nat. Struct. Biol.* **10:** 679–680

Buratowski S (2009) Progression through the RNA Polymerase II CTD Cycle. *Mol. Cell* **36:** 541–546

Burckin T, Nagel R, Mandel-Gutfreund Y, Shiue L, Clark TA, Chong JL, Chang TH, Squazzo S, Hartzog G & Ares M (2005) Exploring functional relationships between components of the gene expression machinery. *Nat. Struct. Mol. Biol.* **12:** 175–182

Burge CB, Padgett RA & Sharp PA (1998) Evolutionary fates and origins of U12-type introns. *Mol. Cell* **2:** 773–785

Burgess SM & 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

Cao Q-F, Yamamoto J, Isobe T, Tateno S, Murase Y, Chen Y, Handa H & Yamaguchi Y (2015) Characterization of the Human Transcription Elongation Factor Rtf1: Evidence for Nonoverlapping Functions of Rtf1 and the Paf1 Complex. *Mol. Cell. Biol.* **35:** 3459–3470

Cao Y, Wen L, Wang Z & Ma L (2015) SKIP Interacts with the Paf1 Complex to Regulate Flowering via the Activation of FLC Transcription in Arabidopsis. *Mol. Plant* **8:** 1816–1819

Carlone DL & Skalnik DG (2001) CpG Binding Protein Is Crucial for Early Embryonic Development. *Mol. Cell. Biol.* **21:** 7601–7606

Carlone DL, Lee J-H, Young SRL, Dobrota E, Butler JS, Ruiz J & Skalnik DG (2005) Reduced Genomic Cytosine Methylation and Defective Cellular Differentiation in Embryonic Stem Cells Lacking CpG Binding Protein. *Mol. Cell. Biol.* **25:** 4881–4891

Carrillo Oesterreich F, Preibisch S & Neugebauer KM (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 & Neugebauer KM (2016) Splicing of Nascent RNA Coincides with Intron Exit from RNA Polymerase II. *Cell* **165:** 372–381

Carrozza MJ, Li B, Florens L, Suganuma T, Swanson SK, Lee KK, Shia WJ, Anderson S, Yates J, Washburn MP & Workman JL (2005) Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. *Cell* **123:** 581–592

Cassart C, Drogat J, Migeot V, & Hermand D (2012). Distinct requirement of RNA polymerase II CTD phosphorylations in budding and fission yeast. *Transcription*. **3:** 231-4.

Castelnuovo M, Zaugg JB, Guffanti E, Maffioletti A, Camblong J, Xu Z, Clauder-Münster S, Steinmetz LM, Luscombe NM & Stutz F (2014) Role of histone modifications and early termination in pervasive transcription and antisense-mediated gene silencing in yeast. *Nucleic Acids Res.* **42:** 4348–4362

Chan SP, Kao DI, Tsai WY & Cheng SC (2003) The Prp19p-associated complex in spliceosome activation. *Science.* **302:** 279–282

Chan SP & Cheng SC (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

Chanfreau G & Jacquier A (1993) Interaction of intronic boundaries is required for the second splicing step efficiency of a group II intron. *EMBO J.* **12:** 5173–80

Chanfreau G, Legrain P, Dujon B & Jacquier A (1994) Interaction between the first and last nucleotides of pre-mRNA introns is a determinant of 3′ splice site selection in S.cerevisiae. *Nucleic Acids Res.* **22:** 1981–1987

Chathoth KT, Barrass JD, Webb S & Beggs JD (2014) A Splicing-Dependent Transcriptional Checkpoint Associated with Prespliceosome Formation. *Mol. Cell* **53:** 779–790

Chen FX, Woodfin AR, Gardini A, Rickels RA, Marshall SA, Smith ER, Shiekhattar R & Shilatifard A (2015) PAF1, a Molecular Regulator of Promoter-Proximal Pausing by RNA Polymerase II. *Cell* **162:** 1003–1015

Chen K, Chen Z, Wu D, Zhang L, Lin X, Su J, Rodriguez B, Xi Y, Xia Z, Chen X, Shi X, Wang Q & Li W (2015) Broad H3K4me3 is associated with increased transcription elongation and enhancer activity at tumor-suppressor genes. *Nat. Genet.* **47:** 1149–1157

Cheng H, He X & Moore C (2004) The essential WD repeat protein Swd2 has dual functions in RNA polymerase II transcription termination and lysine 4 methylation of histone H3. *Mol. Cell. Biol.* **24:** 2932–2943

Cho EJ & Buratowski S (1999) Evidence that transcription factor IIB is required for a post-assembly step in transcription initiation. *J. Biol. Chem.* **274:** 25807–25813

Cho EJ, Kobor MS, Kim M, Greenblatt J & Buratowski S (2001) Opposing effects of Ctk1 kinase and Fcp1 phosphatase at Ser 2 of the RNA polymerase II C-terminal domain. *Genes Dev.* **15:** 3319–3329

Chodavarapu RK, Feng S, Bernatavichute Y V., Chen PY, Stroud H, Yu Y, Hetzel JA, Kuo F, Kim J, Cokus SJ, Casero D, Bernal M, Huijser P, Clark AT, Krämer U, Merchant SS, Zhang X, Jacobsen SE & Pellegrini M (2010) Relationship between nucleosome positioning and DNA methylation. *Nature* **466:** 388–392

Chow LT, Roberts JM, Lewis JB & Broker TR (1977) A map of cytoplasmic RNA transcripts from lytic adenovirus type 2, determined by electron microscopy of RNA:DNA hybrids. *Cell* **11:** 819–836

Chu Y, Simic R, Warner MH, Arndt KM & Prelich G (2007) Regulation of histone modification and cryptic transcription by the Bur1 and Paf1 complexes. *EMBO J.* **26:** 4646–4656

Churchman LS & Weissman JS (2011) Nascent transcript sequencing visualizes transcription at nucleotide resolution. *Nature* **469:** 368–373

Clapier CR, Iwasa J, Cairns BR & Peterson CL (2017) Mechanisms of action and regulation of ATP-dependent chromatin-remodelling complexes. *Nat. Rev. Mol. Cell Biol.* **18:** 407–422

Clouaire T, Webb S, Skene P, Illingworth R, Kerr A, Andrews R, Lee JH, Skalnik D & Bird A (2012) Cfp1 integrates both CpG content and gene activity for accurate H3K4me3 deposition in embryonic stem cells. *Genes Dev.* **26:** 1714–1728

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

Cremer T, Cremer M, Hübner B, Strickfaden H, Smeets D, Popken J, Sterr M, Markaki Y, Rippe K & Cremer C (2015) The 4D nucleome: Evidence for a dynamic nuclear landscape based on co-aligned active and inactive nuclear compartments. *FEBS Lett.* **589:** 2931–2943

Crotti LB, Bačíková D & Horowitz DS (2007) The Prp18 protein stabilizes the interaction of both exons with the U5 snRNA during the second step of pre-mRNA splicing. *Genes Dev.* **21:** 1204–1216

Dangkulwanich M, Ishibashi T, Liu S, Kireeva ML, Lubkowska L, Kashlev M & Bustamante CJ (2013) Complete dissection of transcription elongation reveals slow translocation of RNA polymerase II in a linear ratchet mechanism. *Elife* **2013:**

Danino YM, Even D, Ideses D & Juven-Gershon T (2015) The core promoter: At the heart of gene expression. *Biochim. Biophys. Acta - Gene Regul. Mech.* **1849:** 1116–1131

Das R, Dufu K, Romney B, Feldt M, Elenko M & Reed R (2006) Functional coupling of RNAP II transcription to spliceosome assembly. *Genes Dev.* **20:** 1100–1109

David CJ, Boyne AR, Millhouse SR & Manley JL (2011) The RNA polymerase II C-terminal domain promotes splicing activation through recruitment of a U2AF65-Prp19 complex. *Genes Dev.* **25:** 972–982

Davis CA, Grate L, Spingola M & Ares M (2000) Test of intron predictions reveals novel splice sites, alternatively spliced mRNAs and new introns in meiotically regulated genes of yeast. *Nucleic Acids Res.* **28:** 1700–1706

De Almeida SF, Grosso AR, Koch F, Fenouil R, Carvalho S, Andrade J, Levezinho H, Gut M, Eick D, Gut I, Andrau JC, Ferrier P & Carmo-Fonseca M (2011) Splicing enhances recruitment of methyltransferase HYPB/Setd2 and methylation of histone H3 Lys36. *Nat. Struct. Mol. Biol.* **18:** 977–983

De Graaf K, Czajkowska H, Rottmann S, Packman LC, Lilischkis R, Lüscher B & Becker W (2006) The protein kinase DYRK1A phosphorylates the splicing factor SF3b1/SAP155 at Thr434, a novel in vivo phosphorylation site. *BMC Biochem.* **7:**

De La Mata M, Alonso CR, Kadener S, Fededa JP, Blaustein M, Pelisch F, Cramer P, Bentley D & Kornblihtt AR (2003) A slow RNA polymerase II affects alternative splicing in vivo. *Mol. Cell* **12:** 525–532

Dehé PM, Dichtl B, Schaft D, Roguev A, Pamblanco M, Lebrun R, Rodríguez-Gil A, Mkandawire M, Landsberg K, Shevchenko A, Shevchenko A, Rosaleny LE, Tordera V, Chávez S, Stewart AF & Géli V (2006) Protein interactions within the Set1 complex and their roles in the regulation of histone 3 lysine 4 methylation. *J. Biol. Chem.* **281:** 35404–35412

Descostes N, Heidemann M, Spinelli L, Schüller R, Maqbool MA, Fenouil R, Koch F, Innocenti C, Gut M, Gut I, Eick D & Andrau JC (2014) Tyrosine phosphorylation of RNA Polymerase II CTD is associated with antisense promoter transcription and active enhancers in mammalian cells. *Elife* **2014:**

Dhami P, Saffrey P, Bruce AW, Dillon SC, Chiang K, Bonhoure N, Koch CM, Bye J, James K, Foad NS, Ellis P, Watkins NA, Ouwehand WH, Langford C, Andrews RM, Dunham I & Vetrie D (2010) Complex exon-intron marking by histone modifications is not determined solely by nucleosome distribution. *PLoS One* **5:**

Diamant G, Amir-Zilberstein L, Yamaguchi Y, Handa H & Dikstein R (2012) DSIF Restricts NF-κB Signaling by Coordinating Elongation with mRNA Processing of Negative Feedback Genes. *Cell Rep.* **2:** 722–731

Diamant G, Bahat A & Dikstein R (2016) The elongation factor Spt5 facilitates transcription initiation for rapid induction of inflammatory-response genes. *Nat. Commun.* **7:**

Ding B, LeJeune D & Li S (2010) The C-terminal repeat domain of Spt5 plays an important role in suppression of Rad26-independent transcription coupled repair. *J. Biol. Chem.* **285:** 5317–5326

Doamekpor SK, Sanchez AM, Schwer B, Shuman S & Lima CD (2014) How an mRNA capping enzyme reads distinct RNA polymerase II and Spt5 CTD phosphorylation codes. *Genes Dev.* **28:** 1323–1336

Doamekpor SK, Schwer B, Sanchez AM, Shuman S & Lima CD (2015) Fission yeast RNA triphosphatase reads an Spt5 CTD code. *Rna* **21:** 113–123

Dronamraju R & Strahl BD (2014) A feed forward circuit comprising Spt6, Ctk1 and PAF regulates Pol II CTD phosphorylation and transcription elongation. *Nucleic Acids Res.* **42:** 870–881

Dujardin G, Lafaille C, de la Mata M, Marasco LE, Muñoz MJ, Le Jossic-Corcos C, Corcos L & Kornblihtt AR (2014) How Slow RNA Polymerase II Elongation Favors Alternative Exon Skipping. *Mol. Cell* **54:** 683–690

Egloff S, O'Reilly D, Chapman RD, Taylor A, Tanzhaus K, Pitts L, Eick D & Murphy S (2007) Serine-7 of the RNA polymerase II CTD is specifically required for snRNA gene expression. *Science.* **318:** 1777–1779

Eick D & Geyer M (2013) The RNA polymerase II carboxy-terminal domain (CTD) code. *Chem. Rev.* **113:** 8456–8490

Ezkurdia I, Juan D, Rodriguez JM, Frankish A, Diekhans M, Harrow J, Vazquez J, Valencia A & Tress ML (2014) Multiple evidence strands suggest that theremay be as few as 19 000 human protein-coding genes. *Hum. Mol. Genet.* **23:** 5866–5878

Fabrizio P, Dannenberg J, Dube P, Kastner B, Stark H, Urlaub H & Lührmann R (2009) The Evolutionarily Conserved Core Design of the Catalytic Activation Step of the Yeast Spliceosome. *Mol. Cell* **36:** 593–608

Fica SM, Oubridge C, Galej WP, Wilkinson ME, Bai XC, Newman AJ & Nagai K (2017) Structure of a spliceosome remodelled for exon ligation. *Nature* **542:** 377–380

Fica SM, Tuttle N, Novak T, Li NS, Lu J, Koodathingal P, Dai Q, Staley JP & Piccirilli JA (2013) RNA catalyses nuclear pre-mRNA splicing. *Nature* **503:** 229–234

Field Y, Kaplan N, Fondufe-Mittendorf Y, Moore IK, Sharon E, Lubling Y, Widom J & Segal E (2008) Distinct modes of regulation by chromatin encoded through nucleosome positioning signals. *PLoS Comput. Biol.* **4:**

Fischl H, Howe FS, Furger A & Mellor J (2017) Paf1 Has Distinct Roles in Transcription Elongation and Differential Transcript Fate. *Mol. Cell* **65:** 685–698.e8

Fiszbein A, Godoy Herz MA, Gomez Acuña LI & Kornblihtt AR (2016) Interplay Between Chromatin and Splicing. In *Chromatin Regulation and Dynamics*, Göndör ABT-CR and D (ed) pp 191–209. Boston: Academic Press Available at: http://www.sciencedirect.com/science/article/pii/B9780128033951000083

Fitz J, Neumann T & Pavri R (2018) Regulation of RNA polymerase II processivity by Spt5 is restricted to a narrow window during elongation. *EMBO J.***:** e97965

Fong N, Kim H, Zhou Y, Ji X, Qiu J, Saldi T, Diener K, Jones K, Fu XD & Bentley DL (2014) Pre-mRNA splicing is facilitated by an optimal RNA polymerase II elongation rate. *Genes Dev.* **28:** 2663–2676

Fong YW & Zhou Q (2001) Stimulatory effect of splicing factors on transcriptional elongation. *Nature* **414:** 929–933

Fox-Walsh KL & Hertel KJ (2009) Splice-site pairing is an intrinsically high fidelity process. *Proc. Natl. Acad. Sci.* **106:** 1766–1771

Fuchs G, Voichek Y, Benjamin S, Gilad S, Amit I & Oren M (2014) 4sUDRB-seq: measuring genomewide transcriptional elongation rates and initiation frequencies within cells. *Genome Biol.* **15:** R69

Galej WP, Wilkinson ME, Fica SM, Oubridge C, Newman AJ & Nagai K (2016) Cryo-EM structure of the spliceosome immediately after branching. *Nature* **537:** 197–201

Gasch A, Wiesner S, Martin-Malpartida P, Ramirez-Espain X, Ruiz L & Macias MJ (2006) The structure of Prp40 FF1 domain and its interaction with the crn-TPR1 motif of Clf1 gives a new insight into the binding mode of FF domains. *J. Biol. Chem.* **281:** 356–364

Ghosh S & Garcia-Blanco MA (2000) Coupled in vitro synthesis and splicing of RNA polymerase II transcripts. *Rna* **6:** 1325–1334

Gietz RD & Schiestl RH (2007) Quick and easy yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nat. Protoc.* **2:** 35–37

Glaser S (2006) Multiple epigenetic maintenance factors implicated by the loss of Mll2 in mouse development. *Development* **133:** 1423–1432 Available at: http://dev.biologists.org/cgi/doi/10.1242/dev.02302

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

Gottschalk A, Neubauer G, Banroques J, Mann M, Lührmann R & Fabrizio P (1999) Identification by mass spectrometry and functional analysis of novel proteins of the yeast [U4/U6.U5] tri-snRNP. *EMBO J.* **18:** 4535–48

Gottschalk A, Tang J, Puig O, Salgado J, Neubauer G, Colot H V, Mann M, Seraphin B, Rosbash M, Lührmann R & Fabrizio P (1998) A comprehensive biochemical and genetic analysis of the yeast U1 snRNP reveals five novel proteins. *Rna* **4:** 374–393

Grainger RJ & Beggs JD (2005) Prp8 protein: At the heart of the spliceosome. *Rna* **11:** 533–557

Grosso AR, De Almeida SF, Braga J & Carmo-Fonseca M (2012) Dynamic transitions in RNA polymerase II density profiles during transcription termination. *Genome Res.* **22:** 1447–1456

Grünberg S & Hahn S (2013) Structural insights into transcription initiation by RNA polymerase II. *Trends Biochem. Sci.* **38:** 603–611

Gu B, Eick D & Bensaude O (2013) CTD serine-2 plays a critical role in splicing and termination factor recruitment to RNA polymerase II in vivo. *Nucleic Acids Res.* **41:** 1591–1603

Gunderson FQ, Merkhofer EC & Johnson TL (2011) Dynamic histone acetylation is critical for cotranscriptional spliceosome assembly and spliceosomal rearrangements. *Proc. Natl. Acad. Sci. U. S. A.* **108:** 2004–2009

Gunderson FQ & Johnson TL (2009) Acetylation by the transcriptional coactivator Gcn5 plays a novel role in co-transcriptional spliceosome assembly. *PLoS Genet.* **5:**

Hall SL & Padgett RA (1994) Conserved sequences in a class of rare eukaryotic nuclear introns with non-consensus splice sites. *J. Mol. Biol.* **239:** 357–365

Hammond CM, Strømme CB, Huang H, Patel DJ & Groth A (2017) Histone chaperone networks shaping chromatin function. *Nat. Rev. Mol. Cell Biol.* **18:** 141–158

Haque N & Oberdoerffer S (2014) Chromatin and splicing. *Methods Mol. Biol.* **1126:** 97–113

Hardy SF, Grabowski PJ, Padgett RA & Sharp PA (1984) Cofactor requirements of splicing of purified messenger RNA precursors. *Nature* **308:** 375–377

Harlen KM, Trotta KL, Smith EE, Mosaheb MM, Fuchs SM & Churchman LS (2016) Comprehensive RNA Polymerase II Interactomes Reveal Distinct and Varied Roles for Each Phospho-CTD Residue. *Cell Rep.* **15:** 2147–2158

Hartzog GA & Fu J (2013) The Spt4-Spt5 complex: A multi-faceted regulator of transcription elongation. *Biochim. Biophys. Acta - Gene Regul. Mech.* **1829:** 105–115

Hartzog GA, Wada T, Handa H & Winston F (1998) Evidence that Spt4, Spt5, and Spt6 control transcription elongation by RNA polymerase II in Saccharomyces cerevisiae. *Genes Dev.* **12:** 357–369

He Q, Johnston J & Zeitlinger J (2015) ChIP-nexus enables improved detection of in vivo transcription factor binding footprints. *Nat. Biotechnol.* **33:** 395–401

Hebbes TR, Clayton AL, Thorne AW & Crane-Robinson C (1994) Core histone hyperacetylation co-maps with generalized DNase I sensitivity in the chicken beta-globin chromosomal domain. *EMBO J.* **13:** 1823–30

Hérissant L, Moehle EA, Bertaccini D, Van Dorsselaer A, Schaeffer-Reiss C, Guthrie C & Dargemont C (2014) H2B ubiquitylation modulates spliceosome assembly and function in budding yeast. *Biol. Cell* **106:** 126–138

Hernandez N & Keller W (1983) Splicing of in vitro synthesized messenger RNA precursors in HeLa cell extracts. *Cell* **35:** 89–99

Hesselberth JR (2013) Lives that introns lead after splicing. *Wiley Interdiscip. Rev. RNA* **4:** 677–691

Hicks MJ, Yang CR, Kotlajich M V. & Hertel KJ (2006) Linking splicing to Pol II transcription stabilizes pre-mRNAs and influences splicing patterns. *PLoS Biol.* **4:** 0943–0951

Hintermair C, Heidemann M, Koch F, Descostes N, Gut M, Gut I, Fenouil R, Ferrier P, Flatley A, Kremmer E, Chapman RD, Andrau JC & Eick D (2012) Threonine-4 of mammalian RNA polymerase II CTD is targeted by Polo-like kinase 3 and required for transcriptional elongation. *EMBO J.* **31:** 2784–2797

Hirose Y, Tacke R & Manley JL (1999) Phosphorylated RNA polymerase II stimulates pre-mRNA splicing. *Genes Dev.* **13:** 1234–1239

Hirtreiter A, Damsma GE, Cheung ACM, Klose D, Grohmann D, Vojnic E, Martin ACR, Cramer P & Werner F (2010) Spt4/5 stimulates transcription elongation through the RNA polymerase clamp coiled-coil motif. *Nucleic Acids Res.* **38:** 4040–4051

Hodges C, Bintu L, Lubkowska L, Kashlev M & Bustamante C (2009) Nucleosomal fluctuations govern the transcription dynamics of RNA polymerase II. *Science.* **325:** 626–628

Hodges C, Bintu L, Lubkowska L, Kashlev M & Bustamante C (2009) Nucleosomal fluctuations govern the transcription dynamics of RNA polymerase II. *Science.* **325:** 626–628

Hodges PE & Beggs JD (1994) RNA Splicing: U2 fulfils a commitment. *Curr. Biol.* **4:** 264–267

Hodgkin J (2001) What does a worm want with 20,000 genes? *Genome Biol.* **2:** COMMENT2008

Hoskins AA, Friedman LJ, Gallagher SS, Crawford DJ, Anderson EG, Wombacher R, Ramirez N, Cornish VW, Gelles J & Moore MJ (2011) Ordered and dynamic assembly of single spliceosomes. *Science.* **331:** 1289–1295

Hossain MA, Claggett JM, Edwards SR, Shi A, Pennebaker SL, Cheng MY, Hasty J & Johnson TL (2016) Posttranscriptional Regulation of Gcr1 Expression and Activity Is Crucial for Metabolic Adjustment in Response to Glucose Availability. *Mol. Cell* **62:** 346–358

Hossain MA & Johnson TL (2014) Using Yeast Genetics to Study Splicing Mechanisms. *Methods Mol. Biol.***:** 285–298

Hossain MA, Rodriguez CM & Johnson TL (2011) Key features of the two-intron Saccharomyces cerevisiae gene SUS1 contribute to its alternative splicing. *Nucleic Acids Res.* **39:** 8612–8627

Howe FS, Fischl H, Murray SC & Mellor J (2017) Is H3K4me3 instructive for transcription activation? *BioEssays* **39:** 1–12

Howe KJ, Kane CM & Ares M (2003) Perturbation of transcription elongation influences the fidelity of internal exon inclusion in Saccharomyces cerevisiae. *Rna* **9:** 993–1006

Hsin JP, Li W, Hoque M, Tian B & Manley JL (2014) RNAP II CTD tyrosine 1 performs diverse functions in vertebrate cells. *Elife* **2014:**

Hsin JP, Sheth A & Manley JL (2011) RNAP II CTD phosphorylated on threonine-4 is required for histone mRNA 3' end processing. *Science.* **334:** 683–686

Huff JT, Plocik AM, Guthrie C & Yamamoto KR (2010) Reciprocal intronic and exonic histone modification regions in humans. *Nat. Struct. Mol. Biol.* **17:** 1495–1499

Hyun K, Jeon J, Park K & Kim J (2017) Writing, erasing and reading histone lysine methylations. *Exp. Mol. Med.* **49:**

Iyer V & Struhl K (1995) Poly(dA:dT), a ubiquitous promoter element that stimulates transcription via its intrinsic DNA structure. *EMBO J.* **14:** 2570–9

Jonkers I, Kwak H & Lis JT (2014) Genome-wide dynamics of Pol II elongation and its interplay with promoter proximal pausing, chromatin, and exons. *Elife* **2014:** e02407

Juneau K, Miranda M, Hillenmeyer ME, Nislow C & Davis RW (2006) Introns regulate RNA and protein abundance in yeast. *Genetics* **174:** 511–518

Juneau K, Nislow C & Davis RW (2009) Alternative splicing of PTC7 in Saccharomyces cerevisiae determines protein localization. *Genetics* **183:** 185–194

Kadowaki T (1994) Isolation and characterization of Saccharomyces cerevisiae mRNA transport-defective (mtr) mutants [published erratum appears in J Cell Biol 1994 Sep;126(6):1627]. *J. Cell Biol.* **126:** 649–659

Kaplan N, Moore IK, Fondufe-Mittendorf Y, Gossett AJ, Tillo D, Field Y, LeProust EM, Hughes TR, Lieb JD, Widom J & Segal E (2009) The DNA-encoded nucleosome organization of a eukaryotic genome. *Nature* **458:** 362–366

Käufer NF & Potashkin J (2000) Analysis of the splicing machinery in fission yeast: a comparison with budding yeast and mammals. *Nucleic Acids Res* **28:** 3003–3010

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

Kawecki TJ & Ebert D (2004) Conceptual issues in local adaptation. *Ecol. Lett.* **7:** 1225–1241

Keren-Shaul H, Lev-Maor G & Ast G (2013) Pre-mRNA Splicing Is a Determinant of Nucleosome Organization. *PLoS One* **8:**

Khodor YL, Menet JS, Tolan M & Rosbash M (2012) Cotranscriptional splicing efficiency differs dramatically between Drosophila and mouse. *Rna* **18:** 2174–2186

Kilchert C, Wittmann S, Passoni M, Shah S, Granneman S & Vasiljeva L (2015) Regulation of mRNA Levels by Decay-Promoting Introns that Recruit the Exosome Specificity Factor Mmi1. *Cell Rep.* **13:** 2504–2515

Kim J, Guermah M, McGinty RK, Lee JS, Tang Z, Milne TA, Shilatifard A, Muir TW & Roeder RG (2009) RAD6-Mediated Transcription-Coupled H2B Ubiquitylation Directly Stimulates H3K4 Methylation in Human Cells. *Cell* **137:** 459–471

Kim JH, Lee BB, Oh YM, Zhu C, Steinmetz LM, Lee Y, Kim WK, Lee SB, Buratowski S & Kim TS (2016) Modulation of mRNA and lncRNA expression dynamics by the Set2-Rpd3S pathway. *Nat. Commun.* **7:** 13534

Kim S, Kim H, Fong N, Erickson B & Bentley DL (2011) Pre-mRNA splicing is a determinant of histone H3K36 methylation. *Proc. Natl. Acad. Sci.* **108:** 13564–13569

Kireeva ML, Hancock B, Cremona GH, Walter W, Studitsky VM & Kashlev M (2005) Nature of the nucleosomal barrier to RNA polymerase II. *Mol. Cell* **18:** 97–108

Klusza S, Novak A, Figueroa S, Palmer W & Deng WM (2013) Prp22 and spliceosome components regulate chromatin dynamics in germ-line polyploid cells. *PLoS One* **8:** e79048

Knezetic JA & Luse DS (1986) The presence of nucleosomes on a DNA template prevents initiation by RNA polymerase II in vitro. *Cell* **45:** 95–104

Koga M, Hayashi M & Kaida D (2015) Splicing inhibition decreases phosphorylation level of Ser2 in Pol II CTD. *Nucleic Acids Res.* **43:** 8258–8267

Kolasinska-Zwierz P, Down T, Latorre I, Liu T, Liu XS & Ahringer J (2009) Differential chromatin marking of introns and expressed exons by H3K36me3. *Nat. Genet.* **41:** 376–381

Kornblihtt AR, Schor IE, Alló M, Dujardin G, Petrillo E & Muñoz MJ (2013) Alternative splicing: A pivotal step between eukaryotic transcription and translation. *Nat. Rev. Mol. Cell Biol.* **14:** 153–165

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

Krogan NJ, Dover J, Wood A, Schneider J, Heidt J, Boateng MA, Dean K, Ryan OW, Golshani A, Johnston M, Greenblatt JF & Shilatifard A (2003) The Paf1 complex is required for histone H3 methylation by COMPASS and Dot1p: Linking transcriptional elongation to histone methylation. *Mol. Cell* **11:** 721–729

Kulak NA, Pichler G, Paron I, Nagaraj N & Mann M (2014) Minimal, encapsulated proteomic-sample processing applied to copy-number estimation in eukaryotic cells. *Nat. Methods* **11:** 319–324

Kunkel GR & Martinson HG (1981) Nucleosomes will not form on double-stranded RNA or over poly(dA)-poly(dT) tracts in recombinant DNA. *Nucleic Acids Res.* **9:** 6869–6888

Kurdistani SK & Grunstein M (2003) Histone acetylation and deacetylation in yeast. *Nat. Rev. Mol. Cell Biol.* **4:** 276–284

Kwak H, Fuda NJ, Core LJ & Lis JT (2013) Precise maps of RNA polymerase reveal how promoters direct initiation and pausing. *Science.* **339:** 950–953

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

Lacadie SA, Tardiff DF, Kadener S & Rosbash M (2006) In vivo commitment to yeast cotranscriptional splicing is sensitive to transcription elongation mutants. *Genes Dev.* **20:** 2055–2066

Langford CJ, Klinz FJ, Donath C & Gallwitz D (1984) Point mutations identify the conserved, intron-contained TACTAAC box as an essential splicing signal sequence in yeast. *Cell* **36:** 645–653

Lardelli RM, Thompson JX, Yates JR & Stevens SW (2010) Release of SF3 from the intron branchpoint activates the first step of pre-mRNA splicing. *Rna* **16:** 516–528

Laribee RN, Krogan NJ, Xiao T, Shibata Y, Hughes TR, Greenblatt JF & Strahl BD (2005) BUR kinase selectively regulates H3 K4 trimethylation and H2B ubiquitylation through recruitment of the PAF elongation complex. *Curr. Biol.* **15:** 1487–1493

Lee JS, Shukla A, Schneider J, Swanson SK, Washburn MP, Florens L, Bhaumik SR & Shilatifard A (2007) Histone Crosstalk between H2B Monoubiquitination and H3 Methylation Mediated by COMPASS. *Cell* **131:** 1084–1096

Lee J, Saha PK, Yang Q-H, Lee S, Park JY, Suh Y, Lee S-K, Chan L, Roeder RG & Lee JW (2008) Targeted inactivation of MLL3 histone H3-Lys-4 methyltransferase activity in the mouse reveals vital roles for MLL3 in adipogenesis. *Proc. Natl. Acad. Sci.* **105:** 19229–19234

Lee Y, Park D & Iyer VR (2017) The ATP-dependent chromatin remodeler Chd1 is recruited by transcription elongation factors and maintains H3K4me3/H3K36me3 domains at actively transcribed and spliced genes. *Nucleic Acids Res.* **45:** 7180–7190

Leeds NB, Small EC, Hiley SL, Hughes TR & Staley JP (2006) The Splicing Factor Prp43p, a DEAH Box ATPase, Functions in Ribosome Biogenesis. *Mol. Cell. Biol.* **26:** 513–522

Lenasi T & Barboric M (2010) P-TEFb stimulates transcription elongation and pre-mRNA splicing through multilateral mechanisms. *RNA.* **7:** 145-150

Li B, Carey M & Workman JL (2007) The Role of Chromatin during Transcription. *Cell* **128:** 707–719

Li G & Reinberg D (2011) Chromatin higher-order structures and gene regulation. *Curr. Opin. Genet. Dev.* **21:** 175–186

Li Y, Xia C, Feng J, Yang D, Wu F, Cao Y, Li L & Ma L (2016) The SNW Domain of SKIP Is Required for Its Integration into the Spliceosome and Its Interaction with the Paf1 Complex in Arabidopsis. *Mol. Plant* **9:** 1040–1050

Lidschreiber M, Leike K & Cramer P (2013) Cap Completion and C-Terminal Repeat Domain Kinase Recruitment Underlie the Initiation-Elongation Transition of RNA Polymerase II. *Mol. Cell. Biol.* **33:** 3805–3816

Lin S, Coutinho-Mansfield G, Wang D, Pandit S & Fu X-D (2008) The splicing factor SC35 has an active role in transcriptional elongation. *Nat. Struct. Mol. Biol.* **15:** 819–826

Lindstrom DL, Squazzo SL, Muster N, Burckin TA, Wachter KC, Emigh CA, McCleery JA, Yates JR & Hartzog GA (2003) Dual Roles for Spt5 in Pre-mRNA Processing and Transcription Elongation Revealed by Identification of Spt5-Associated Proteins. *Mol. Cell. Biol.* **23:** 1368–1378

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

Liu Y, Warfield L, Zhang C, Luo J, Allen J, Lang WH, Ranish J, Shokat KM & Hahn S (2009) Phosphorylation of the Transcription Elongation Factor Spt5 by Yeast Bur1 Kinase Stimulates Recruitment of the PAF Complex. *Mol. Cell. Biol.* **29:** 4852–4863

Liu CR, Chang CR, Chern Y, Wang TH, Hsieh WC, Shen WC, Chang CY, Chu IC, Deng N, Cohen SN & Cheng TH (2012) Spt4 is selectively required for transcription of extended trinucleotide repeats. *Cell* **148:** 690–701

Liu X, Bushnell DA & Kornberg RD (2013) RNA polymerase II transcription: Structure and mechanism. *Biochim. Biophys. Acta - Gene Regul. Mech.* **1829:** 2–8

Liu X, Kraus WL & Bai X (2015) Ready, pause, go: Regulation of RNA polymerase II pausing and release by cellular signaling pathways. *Trends Biochem. Sci.* **40:** 516–525

Liu S, Li X, Zhang L, Jiang J, Hill RC, Cui Y, Hansen KC, Zhou ZH & Zhao R (2017) Structure of the yeast spliceosomal postcatalytic P complex. *Science.* **358:** 1278–1283

Lockhart SR & Rymond BC (1994) Commitment of yeast pre-mRNA to the splicing pathway requires a novel U1 small nuclear ribonucleoprotein polypeptide, Prp39p. *Mol. Cell. Biol.* **14:** 3623–3633

Lorch Y, LaPointe JW & Kornberg RD (1987) Nucleosomes inhibit the initiation of transcription but allow chain elongation with the displacement of histones. *Cell* **49:** 203–210

Lu H, Flores O, Weinmann R & Reinberg D (1991) The nonphosphorylated form of RNA polymerase II preferentially associates with the preinitiation complex. *Proc. Natl. Acad. Sci. U. S. A.* **88:** 10004–10008

Luciano P, Jeon J, El-kaoutari A, Challal D, Bonnet A, Barucco M, Candelli T, Jourquin F, Lesage P, Kim J, Libri D & Géli V (2017) Binding to RNA regulates Set1 function. *Cell Discov.* **3:** 17040

Luco RF, Pan Q, Tominaga K, Blencowe BJ, Pereira-Smith OM & Misteli T (2010) Regulation of alternative splicing by histone modifications. *Science.* **327:** 996–1000

Luger K, Mäder AW, Richmond RK, Sargent DF & Richmond TJ (1997) Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* **389:** 251–260

Luse DS (2013) The RNA polymerase II preinitiation complex: Through what pathway is the complex assembled? *Transcription* **5:** e27050

Lustig AJ, Lin RJ & Abelson J (1986) The yeast RNA gene products are essential for mRNA splicing in vitro. *Cell* **47:** 953–963

Ma P & Xia X (2011) Factors affecting splicing strength of yeast genes. *Comp. Funct. Genomics* **2011:** 212146

Madhani HD & Guthrie C (1992) A novel base-pairing interaction between U2 and U6 snRNAs suggests a mechanism for the catalytic activation of the spliceosome. *Cell* **71:** 803–817

Maeder C, Kutach AK & Guthrie C (2009) ATP-dependent unwinding of U4/U6 snRNAs by the Brr2 helicase requires the C terminus of Prp8. *Nat. Struct. Mol. Biol.* **16:** 42–48

Magraner-Pardo L, Pelechano V, Coloma MD & Tordera V (2014) Dynamic remodeling of histone modifications in response to osmotic stress in Saccharomyces cerevisiae. *BMC Genomics* **15:** e247

Malik I, Qiu C, Snavely T & Kaplan CD (2017) Wide-ranging and unexpected consequences of altered Pol II catalytic activity in vivo. *Nucleic Acids Res.* **45:** 4431–4451

Margaritis T, Oreal V, Brabers N, Maestroni L, Vitaliano-Prunier A, Benschop JJ, van Hooff S, van Leenen D, Dargemont C, Géli V & Holstege FCP (2012) Two Distinct Repressive Mechanisms for Histone 3 Lysine 4 Methylation through Promoting 3′-End Antisense Transcription. *PLoS Genet.* **8:** e1002952

Martin A, Schneider S & 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

Martinez-Rucobo FW, Sainsbury S, Cheung ACM & Cramer P (2011) Architecture of the RNA polymerase-Spt4/5 complex and basis of universal transcription processivity. *EMBO J.* **30:** 1302–1310

Mason PB & Struhl K (2005) Distinction and relationship between elongation rate and processivity of RNA polymerase II in vivo. *Mol. Cell* **17:** 831–840

Mathew R, Hartmuth K, Möhlmann S, Urlaub H, Ficner R & Lührmann R (2008) Phosphorylation of human PRP28 by SRPK2 is required for integration of the U4/U6-U5 tri-snRNP into the spliceosome. *Nat. Struct. Mol. Biol.* **15:** 435–443

Mavrich TN, Ioshikhes IP, Venters BJ, Jiang C, Tomsho LP, Qi J, Schuster SC, Albert I & Pugh BF (2008) A barrier nucleosome model for statistical positioning of nucleosomes throughout the yeast genome. *Genome Res.* **18:** 1073–1083

Mayas RM, Maita H, Semlow DR & Staley JP (2010) Spliceosome discards intermediates via the DEAH box ATPase Prp43p. *Proc. Natl. Acad. Sci.* **107:** 10020–10025

Mayas RM, Maita H & Staley JP (2006) Exon ligation is proofread by the DExD/H-box ATPase Prp22p. *Nat. Struct. Mol. Biol.* **13:** 482–490

Mayer A, Di Iulio J, Maleri S, Eser U, Vierstra J, Reynolds A, Sandstrom R, Stamatoyannopoulos JA & Churchman LS (2015) Native elongating transcript sequencing reveals human transcriptional activity at nucleotide resolution. *Cell* **161:** 541–544

Mayer A, Heidemann M, Lidschreiber M, Schreieck A, Sun M, Hintermair C, Kremmer E, Eick D & Cramer P (2012) CTD Tyrosine phosphorylation impairs termination factor recruitment to RNA polymerase II. *Science.* **336:** 1723–1725

Mayer A, Lidschreiber M, Siebert M, Leike K, Söding J & Cramer P (2010) Uniform transitions of the general RNA polymerase II transcription complex. *Nat. Struct. Mol. Biol.* **17:** 1272–1278

McCracken S, Fong N, Yankulov K, Ballantyne S, Pan G, Greenblatt J, Patterson SD, Wickens M & Bentley DL (1997) The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. *Nature* **385:** 357–360

McIsaac RS, Gibney PA, Chandran SS, Benjamin KR & Botstein D (2014) Synthetic biology tools for programming gene expression without nutritional perturbations in Saccharomyces cerevisiae. *Nucleic Acids Res.* **42:** e48-e48

McKay SL & Johnson TL (2010) A bird's-eye view of post-translational modifications in the spliceosome and their roles in spliceosome dynamics. *Mol. Biosyst.* **6:** 2093–2102

McKenzie RW & Brennan MD (1996) The two small introns of the Drosophila affinidisjuncta Adh gene are required for normal transcription. *Nucleic Acids Res.* **24:** 3635–42

Meinhart A, Kamenski T, Hoeppner S, Baumli S & Cramer P (2005) A structural perspective of CTD function. *Genes Dev.* **19:** 1401–1415

Meininghaus M, Chapman RD, Horndasch M & Eick D (2000) Conditional expression of RNA polymerase II in mammalian cells. Deletion of the carboxyl-terminal domain of the large subunit affects early steps in transcription. *J. Biol. Chem.* **275:** 24375–24382

Mendoza-Ochoa G, Barrass JD, Terlouw BR, Maudlin IE, de Lucas S, Sani E, Aslanzadeh V, Reid JAE, Beggs JD (2018) A fast and tuneable auxin-inducible degron for depletion of target proteins in budding yeast. *Yeast. Manuscript in review*

Merkhofer EC, Hu P & Johnson TL (2014) Introduction to cotranscriptional RNA splicing. *Methods Mol. Biol.* **1126:** 83–96

Miller T, Krogan NJ, Dover J, Erdjument-Bromage H, Tempst P, Johnston M, Greenblatt JF & Shilatifard A (2001) COMPASS: a complex of proteins associated

with a trithorax-related SET domain protein. *Proc. Natl. Acad. Sci. U. S. A.* **98:** 12902–7

Milligan L, Huynh-Thu VA, Delan-Forino C, Tuck A, Petfalski E, Lombraña R, Sanguinetti G, Kudla G & Tollervey D (2016) Strand-specific, high-resolution mapping of modified RNA polymerase II. *Mol. Syst. Biol.* **12:** 874

Milligan L, Sayou C, Tuck A, Auchynnikava T, Reid JEA, Alexander R, de Lima Alves F, Allshire R, Spanos C, Rappsilber J, Beggs JD, Kudla G & Tollervey D (2017) RNA polymerase II stalling at pre-mRNA splice sites is enforced by ubiquitination of the catalytic subunit. *Elife* **6:** e27082

Moabbi AM, Agarwal N, El Kaderi B & Ansari A (2012) Role for gene looping in intron-mediated enhancement of transcription. *Proc. Natl. Acad. Sci.* **109:** 8505–8510

Moehle EA, Ryan CJ, Krogan NJ, Kress TL & Guthrie C (2012) The Yeast SR-Like Protein Npl3 Links Chromatin Modification to mRNA Processing. *PLoS Genet.* **8:**

Moore MJ, Schwartzfarb EM, Silver PAA & Yu MC (2006) Differential Recruitment of the Splicing Machinery during Transcription Predicts Genome-Wide Patterns of mRNA Splicing. *Mol. Cell* **24:** 903–915

Morawska M & Ulrich HD (2013) An expanded tool kit for the auxin-inducible degron system in budding yeast. *Yeast* **30:** 341–351

Morillon A, Karabetsou N, O'Sullivan J, Kent N, Proudfoot N & Mellor J (2003) Isw1 Chromatin Remodeling ATPase Coordinates Transcription Elongation and Termination by RNA Polymerase II. *Cell* **115:** 425–435

Morris DP & Greenleaf AL (2000) The splicing factor, Prp40, binds the phosphorylated carboxyl-terminal domain of RNA Polymerase II. *J. Biol. Chem.* **275:** 39935–39943

Mortillaro MJ, Blencowe BJ, Wei X, Nakayasu H, Du L, Warren SL, Sharp PA & Berezney R (1996) A hyperphosphorylated form of the large subunit of RNA polymerase II is associated with splicing complexes and the nuclear matrix. *Proc. Natl. Acad. Sci. U. S. A.* **93:** 8253–7

Mueller CL, Porter SE, Hoffman MG & Jaehning JA (2004) The Paf1 complex has functions independent of actively transcribing RNA polymerase II. *Mol. Cell* **14:** 447–456

Nakanishi S, Jung SL, Gardner KE, Gardner JM, Takahashi YH, Chandrasekharan MB, Sun ZW, Osley MA, Strahl BD, Jaspersen SL & Shilatifard A (2009) Histone H2BK123 monoubiquitination is the critical determinant for H3K4 and H3K79 trimethylation by COMPASS and Dot1. *J. Cell Biol.* **186:** 371–377

Natsume T, Kiyomitsu T, Saga Y & Kanemaki MT (2016) Rapid Protein Depletion in Human Cells by Auxin-Inducible Degron Tagging with Short Homology Donors. *Cell Rep.* **15:** 210–218

Nelson HCM, Finch JT, Luisi BF & Klug A (1987) The structure of an oligo(dA)·oligo(dT) tract and its biological implications. *Nature* **330:** 221–226

Neugebauer KM (2002) On the importance of being co-transcriptional. *J. Cell Sci.* **115:** 3865–3871

Neves LT, Douglass S, Spreafico R, Venkataramanan S, Kress TL & Johnson TL (2017) The histone variant H2A.Z promotes efficient cotranscriptional splicing in s. cerevisiae. *Genes Dev.* **31:** 702–717

Newman AJ & Norman C (1992) U5 snRNA interacts with exon sequences at 5′ and 3′ splice sites. *Cell* **68:** 743–754

Ng HH, Dole S & Struhl K (2003a) The Rtf1 Component of the Paf1 Transcriptional Elongation Complex Is Required for Ubiquitination of Histone H2B. *J. Biol. Chem.* **278:** 33625–33628

Ng HH, Robert F, Young RA & Struhl K (2003b) Targeted recruitment of Set1 histone methylase by elongating Pol II provides a localized mark and memory of recent transcriptional activity. *Mol. Cell* **11:** 709–719

Nicholson TB, Veland N & Chen T (2015) Writers, Readers, and Erasers of Epigenetic Marks. In *Epigenetic Cancer Therapy* pp 31–66.

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

Nislow C, Ray E & Pillus L (1997) *SET1* , A Yeast Member of the *Trithorax* Family, Functions in Transcriptional Silencing and Diverse Cellular Processes. *Mol. Biol. Cell* **8:** 2421–2436

Nogués G, Kadener S, Cramer P, Bentley D & Kornblihtt AR (2002) Transcriptional activators differ in their abilities to control alternative splicing. *J. Biol. Chem.* **277:** 43110–43114

Nojima T, Gomes T, Grosso ARF, Kimura H, Dye MJ, Dhir S, Carmo-Fonseca M & Proudfoot NJ (2015) Mammalian NET-seq reveals genome-wide nascent transcription coupled to RNA processing. *Cell* **161:** 526–540

Nordick K, Hoffman MG, Betz JL & Jaehning JA (2008) Direct interactions between the Paf1 complex and a cleavage and polyadenylation factor are revealed by dissociation of Paf1 from RNA polymerase II. *Eukaryot. Cell* **7:** 1158–1167

O'Sullivan JM, Tan-Wong SM, Morillon A, Lee B, Coles J, Mellor J & Proudfoot NJ (2004) Gene loops juxtapose promoters and terminators in yeast. *Nat. Genet.* **36:** 1014–1018

Osheim YN, O.L. Miller J & Beyer AL (1985) RNP particles at splice junction sequences on Drosophila chorion transcripts. *Cell* **43:** 143–151

Padgett RA, Mount SM, Steitz JA & Sharp PA (1983) Splicing of messenger RNA precursors is inhibited by antisera to small nuclear ribonucleoprotein. *Cell* **35:** 101–107

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

Pandit S, Lynn B & Rymond BC (2006) Inhibition of a spliceosome turnover pathway suppresses splicing defects. *Proc. Natl. Acad. Sci.* **103:** 13700–13705

Parenteau J, Durand M, Veronneau S, Lacombe A-A, Morin G, Guerin V, Cecez B, Gervais-Bird J, Koh C-S, Brunelle D, Wellinger RJ, Chabot B & Abou Elela S (2008) Deletion of Many Yeast Introns Reveals a Minority of Genes that Require Splicing for Function. *Mol. Biol. Cell* **19:** 1932–1941

Parua PK, Booth GT, Sansó M, Benjamin B, Tanny JC, Lis JT & Fisher RP (2018) A Cdk9-PP1 switch regulates the elongation-termination transition of RNA polymerase II. *Nature* **558:** 460–464

Patterson B & Guthrie C (1991) A U-rich tract enhances usage of an alternative 3′ splice site in yeast. *Cell* **64:** 181–187

Pavri R, Zhu B, Li G, Trojer P, Mandal S, Shilatifard A & Reinberg D (2006) Histone H2B Monoubiquitination Functions Cooperatively with FACT to Regulate Elongation by RNA Polymerase II. *Cell* **125:** 703–717

Pavri R, Gazumyan A, Jankovic M, Di Virgilio M, Klein I, Ansarah-Sobrinho C, Resch W, Yamane A, San-Martin BR, Barreto V, Nieland TJ, Root DE, Casellas R & Nussenzweig MC (2010) Activation-induced cytidine deaminase targets DNA at sites of RNA polymerase II stalling by interaction with Spt5. *Cell* **143:** 122–133

Perales R & Bentley D (2009) 'Cotranscriptionality': The Transcription Elongation Complex as a Nexus for Nuclear Transactions. *Mol. Cell* **36:** 178–191

Petesch SJ & Lis JT (2012) Overcoming the nucleosome barrier during transcript elongation. *Trends Genet.* **28:** 285–294

Ping XL, Sun BF, Wang L, Xiao W, Yang X, Wang WJ, Adhikari S, Shi Y, Lv Y, Chen YS, Zhao X, Li A, Yang Y, Dahal U, Lou XM, Liu X, Huang J, Yuan WP, Zhu XF, Cheng T, et al (2014) Mammalian WTAP is a regulatory subunit of the RNA N6-methyladenosine methyltransferase. *Cell Res.* **24:** 177–189

Porrua O & Libri D (2015) Transcription termination and the control of the transcriptome: Why, where and how to stop. *Nat. Rev. Mol. Cell Biol.* **16:** 190–202

Porrua O, Boudvillain M & Libri D (2016) Transcription Termination: Variations on Common Themes. *Trends Genet.* **32:** 508–522

Qiu H, Hu C & Hinnebusch AG (2009) Phosphorylation of the Pol II CTD by KIN28 Enhances BUR1/BUR2 Recruitment and Ser2 CTD Phosphorylation Near Promoters. *Mol. Cell* **33:** 752–762

Qiu H, Hu C, Gaur NA & Hinnebusch AG (2012) Pol II CTD kinases Bur1 and Kin28 promote Spt5 CTR-independent recruitment of Paf1 complex. *EMBO J.* **31:** 3494–3505

Quan TK & Hartzog GA (2010) Histone H3K4 and K36 methylation, Chd1 and Rpd3S oppose the functions of Saccharomyces cerevisiae Spt4-Spt5 in transcription. *Genetics* **184:** 321–334

Radman-Livaja M, Liu CL, Friedman N, Schreiber SL & Rando OJ (2010) Replication and active demethylation represent partially overlapping mechanisms for erasure of H3K4me3 in budding yeast. *PLoS Genet.* **6:** e1000837

Reed R (1989) The organization of 3' splice-site sequences in mammalian introns. *Genes Dev.* **3:** 2113–2123

Rosonina E, Yurko N, Li W, Hoque M, Tian B & Manley JL (2014) Threonine-4 of the budding yeast RNAP II CTD couples transcription with Htz1-mediated chromatin remodeling. *Proc. Natl. Acad. Sci.* **111:** 11924–11931

Rossetto D, Avvakumov N & Côté J (2012) Histone phosphorylation: A chromatin modification involved in diverse nuclear events. *Epigenetics* **7:** 1098–1108

Roybal GA & Jurica MS (2010) Spliceostatin A inhibits spliceosome assembly subsequent to prespliceosome formation. *Nucleic Acids Res.* **38:** 6664–6672

Samanfar B, Omidi K, Hooshyar M, Laliberte B, Alamgir M, Seal AJ, Ahmed-Muhsin E, Viteri DF, Said K, Chalabian F, Wainer G, Burnside D, Shostak K, Bugno M, Willmore WG, Smith ML, Golshani A & Golshani A (2013) Large-scale investigation of oxygen response mutants in Saccharomyces cerevisiae. *Mol. Biosyst.* **9:** 1351–1359

Santangelo GM & Tornow J (1990) Efficient transcription of the glycolytic gene ADH1 and three translational component genes requires the GCR1 product, which can act through TUF/GRF/RAP binding sites. *Mol. Cell. Biol.* **10:** 859–62

Satchwell SC, Drew HR & Travers AA (1986) Sequence periodicities in chicken nucleosome core DNA. *J. Mol. Biol.* **191:** 659–675

Sayou C, Millán-Zambrano G, Santos-Rosa H, Petfalski E, Robson S, Houseley J, Kouzarides T & Tollervey D (2017) RNA Binding by Histone Methyltransferases Set1 and Set2. *Mol. Cell. Biol.* **37:** e00165-17

Schneider S, Campodonico E & 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

Schneider J, Wood A, Lee JS, Schuster R, Dueker J, Maguire C, Swanson SK, Florens L, Washburn MP & Shilatifard A (2005) Molecular regulation of histone H3 trimethylation by COMPASS and the regulation of gene expression. *Mol. Cell* **19:** 849–856

Schor IE, Rascovan N, Pelisch F, Allo M & Kornblihtt AR (2009) Neuronal cell depolarization induces intragenic chromatin modifications affecting NCAM alternative splicing. *Proc. Natl. Acad. Sci.* **106:** 4325–4330

Schreieck A, Easter AD, Etzold S, Wiederhold K, Lidschreiber M, Cramer P & Passmore LA (2014) RNA polymerase II termination involves C-terminal-domain tyrosine dephosphorylation by CPF subunit Glc7. *Nat. Struct. Mol. Biol.* **21:** 175–179

Schüller R, Forné I, Straub T, Schreieck A, Texier Y, Shah N, Decker TM, Cramer P, Imhof A & Eick D (2016) Heptad-Specific Phosphorylation of RNA Polymerase II CTD. *Mol. Cell* **61:** 305–314

Schwartz S, Meshorer E & Ast G (2009) Chromatin organization marks exon-intron structure. *Nat. Struct. Mol. Biol.* **16:** 990–995

Schwer B & Guthrie C (1992) A conformational rearrangement in the spliceosome is dependent on PRP16 and ATP hydrolysis. *EMBO J.* **11:** 5033–9

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

Schwer B & Meszaros T (2000) RNA helicase dynamics in pre-mRNA splicing. *EMBO J.* **19:** 6582–6591

Schwer B (2008) A Conformational Rearrangement in the Spliceosome Sets the Stage for Prp22-Dependent mRNA Release. *Mol. Cell* **30:** 743–754

Schwer B & Shuman S (2011) Deciphering the RNA Polymerase II CTD Code in Fission Yeast. *Mol. Cell* **43:** 311–318

Segal E, Fondufe-Mittendorf Y, Chen L, Thåström A, Field Y, Moore IK, Wang JPZ & Widom J (2006) A genomic code for nucleosome positioning. *Nature* **442:** 772–778

Semlow DR & Staley JP (2012) Staying on message: Ensuring fidelity in pre-mRNA splicing. *Trends Biochem. Sci.* **37:** 263–273

Semlow DR, Blanco MR, Walter NG & Staley JP (2016) Spliceosomal DEAH-Box ATPases Remodel Pre-mRNA to Activate Alternative Splice Sites. *Cell* **164:** 985–998

Shandilya J & Roberts SGE (2012) The transcription cycle in eukaryotes: From productive initiation to RNA polymerase II recycling. *Biochim. Biophys. Acta - Gene Regul. Mech.* **1819:** 391–400

Shetty A, Kallgren SP, Demel C, Maier KC, Spatt D, Alver BH, Cramer P, Park PJ & Winston F (2017) Spt5 Plays Vital Roles in the Control of Sense and Antisense Transcription Elongation. *Mol. Cell* **66:** 77–88.e5

Shilatifard A (2004) Transcriptional elongation control by RNA polymerase II: a new frontier. *Biochim. Biophys. Acta.* **1677:** 79–86.

Shilatifard A (2012) The COMPASS Family of Histone H3K4 Methylases: Mechanisms of Regulation in Development and Disease Pathogenesis. *Annu. Rev. Biochem.* **81:** 65–95

Shivaswamy S, Bhinge A, Zhao Y, Jones S, Hirst M & Iyer VR (2008) Dynamic remodeling of individual nucleosomes across a eukaryotic genome in response to transcriptional perturbation. *PLoS Biol.* **6:** 0618–0630

Simic R, Lindstrom DL, Tran HG, Roinick KL, Costa PJ, Johnson AD, Hartzog GA & Arndt KM (2003) Chromatin remodeling protein Chd1 interacts with transcription elongation factors and localizes to transcribed genes. *EMBO J.* **22:** 1846–1856

Sims RJ, Chen CF, Santos-Rosa H, Kouzarides T, Patel SS & Reinberg D (2005) Human but not yeast CHD1 binds directly and selectively to histone H3 methylated at lysine 4 via its tandem chromodomains. *J. Biol. Chem.* **280:** 41789–41792

Sims RJ, Millhouse S, Chen CF, Lewis BA, Erdjument-Bromage H, Tempst P, Manley JL & Reinberg D (2007) Recognition of Trimethylated Histone H3 Lysine 4 Facilitates the Recruitment of Transcription Postinitiation Factors and Pre-mRNA Splicing. *Mol. Cell* **28:** 665–676

Skaar DA & Greenleaf AL (2002) The RNA polymerase II CTD kinase CTDK-I affects pre-mRNA 3′ cleavage/polyadenylation through the processing component Pti1p. *Mol. Cell* **10:** 1429–1439

Soares LM, He PC, Chun Y, Suh H, Kim TS & Buratowski S (2017) Determinants of Histone H3K4 Methylation Patterns. *Mol. Cell* **68:** 773–785.e6

Sorenson MR, Jha DK, Ucles SA, Flood DM, Strahl BD, Stevens SW & Kress TL (2016) Histone H3K36 methylation regulates pre-mRNA splicing in Saccharomyces cerevisiae. *RNA Biol.* **13:** 412–426

Spies N, Nielsen CB, Padgett RA & Burge CB (2009) Biased Chromatin Signatures around Polyadenylation Sites and Exons. *Mol. Cell* **36:** 245–254

Spingola M, Grate L, Haussler D & Manuel A (1999) Genome-wide bioinformatic and molecular analysis of introns in Saccharomyces cerevisiae. *Rna* **5:** 221–234

St. Amour C V., Sanso M, Bosken CA, Lee KM, Larochelle S, Zhang C, Shokat KM, Geyer M & Fisher RP (2012) Separate Domains of Fission Yeast Cdk9 (P-TEFb) Are Required for Capping Enzyme Recruitment and Primed (Ser7-Phosphorylated) Rpb1 Carboxyl-Terminal Domain Substrate Recognition. *Mol. Cell. Biol.* **32:** 2372–2383

Stepankiw N, Raghavan M, Fogarty EA, Grimson A & Pleiss JA (2015) Widespread alternative and aberrant splicing revealed by lariat sequencing. *Nucleic Acids Res.* **43:** 8488–8501

Stiller JW, Mcconaughy BL & Hall BD (2000) Evolutionary complementation for polymerase II CTD function. *Yeast* **16:** 57–64

Stiller JW & Cook MS (2004) Functional unit of the RNA polymerase II C-terminal domain lies within heptapeptide pairs. *Eukaryot. Cell* **3:** 735–740

Struhl K & Segal E (2013) Determinants of nucleosome positioning. *Nat. Struct. Mol. Biol.* **20:** 267–273

Suh H, Ficarro SB, Kang UB, Chun Y, Marto JA & Buratowski S (2016) Direct Analysis of Phosphorylation Sites on the Rpb1 C-Terminal Domain of RNA Polymerase II. *Mol. Cell* **61:** 297–304

Suter B (2000) Poly(dAmiddle dotdT) sequences exist as rigid DNA structures in nucleosome-free yeast promoters in vivo. *Nucleic Acids Res.* **28:** 4083–4089

Szerlong HJ & Hansen JC (2011) Nucleosome distribution and linker DNA: connecting nuclear function to dynamic chromatin structure. *Biochem. Cell Biol.* **89:** 24–34 Available at: http://www.nrcresearchpress.com/doi/abs/10.1139/O10-139

Tardiff DF & Rosbash M (2006) Arrested yeast splicing complexes indicate stepwise snRNP recruitment during in vivo spliceosome assembly. *Rna* **12:** 968–979

Tarn WY & Steitz JA (1996) A novel spliceosome containing U11, U12, and U5 snRNPs excises a minor class (AT-AC) intron in vitro. *Cell* **84:** 801–811

Teigelkamp S, Newman a J & Beggs JD (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:** 2602–12 Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=398374&tool=pmcentrez&rendertype=abstract

Tilgner H, Nikolaou C, Althammer S, Sammeth M, Beato M, Valcárcel J & Guigó R (2009) Nucleosome positioning as a determinant of exon recognition. *Nat. Struct. Mol. Biol.* **16:** 996–1001

Tilgner H, Knowles DG, Johnson R, Davis CA, Chakrabortty S, Djebali S, Curado J, Snyder M, Gingeras TR & Guigó R (2012) Deep sequencing of subcellular RNA fractions shows splicing to be predominantly co-transcriptional in the human genome but inefficient for lncRNAs. *Genome Res.* **22:** 1616–1625

Tollervey D & Mattaj IW (1987) Fungal small nuclear ribonucleoproteins share properties with plant and vertebrate U-snRNPs. *EMBO J.* **6:** 469–76

Treisman R, Orkin SH & Maniatis T (1983) Specific transcription and RNA splicing defects in five cloned beta-thalassaemia genes. *Nature* **302:** 591–6 Available at: http://www.ncbi.nlm.nih.gov/pubmed/6188062

Tsai RT, Fu RH, Yeh FL, Tseng CK, Lin YC, Huang YH & Cheng SC (2005) Spliceosome disassembly catalyzed by Prp43 and its associated components Ntr1 and Ntr2. *Genes Dev.* **19:** 2991–3003

Tseng CK, Liu HL & Cheng SC (2011) DEAH-box ATPase Prp16 has dual roles in remodeling of the spliceosome in catalytic steps. *Rna* **17:** 145–154

Tseng CK & Cheng SC (2008) Both catalytic steps of nuclear pre-mRNA splicing are reversible. *Science.* **320:** 1782–1784

Tyagi A, Ryme J, Brodin D, Farrants AKÖ & Visa N (2009) SWI/SNF associates with nascent pre-mRNPs and regulates alternative pre-mRNA processing. *PLoS Genet.* **5:** e1000470

Van Nues RW & Beggs JD (2001) Functional contacts with a range of splicing proteins suggest a central role for Brr2p in the dynamic control of the order of events in spliceosomes of Saccharomyces cerevisiae. *Genetics* **157:** 1451–1467

Veloso A, Kirkconnell KS, Magnuson B, Biewen B, Paulsen MT, Wilson TE & Ljungman M (2014) Rate of elongation by RNA polymerase II is associated with specific gene features and epigenetic modifications. *Genome Res.* **24:** 896–905

Vermeulen M, Eberl HC, Matarese F, Marks H, Denissov S, Butter F, Lee KK, Olsen J V., Hyman AA, Stunnenberg HG & Mann M (2010) Quantitative Interaction Proteomics and Genome-wide Profiling of Epigenetic Histone Marks and Their Readers. *Cell* **142:** 967–980

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

Volanakis A, Passoni M, Hector RD, Shah S, Kilchert C, Granneman S & Vasiljeva L (2013) Spliceosome-mediated decay (SMD) regulates expression of nonintronic genes in budding yeast. *Genes Dev.* **27:** 2025–2038

Wagner JDO, Jankowsky E, Company M, Pyle AM & Abelson JN (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 EWJ & Beggs JD (2017) Extremely fast and incredibly close: Cotranscriptional splicing in budding yeast. *Rna* **23:** 601–610

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

Wang C, Chua K, Seghezzi W, Lees E, Gozani O & Reed R (1998) Phosphorylation of spliceosomal protein SAP 155 coupled with splicing catalysis. *Genes Dev.* **12:** 1409–1414

Wang GS & Cooper TA (2007) Splicing in disease: Disruption of the splicing code and the decoding machinery. *Nat. Rev. Genet.* **8:** 749–761

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

Weake VM & Workman JL (2008) Histone Ubiquitination: Triggering Gene Activity. *Mol. Cell* **29:** 653–663

Weiner A, Hsieh THS, Appleboim A, Chen H V., Rahat A, Amit I, Rando OJ & Friedman N (2015) High-resolution chromatin dynamics during a yeast stress response. *Mol. Cell* **58:** 371–386

Wen Y & Shatkin AJ (1999) Transcription elongation factor hSPT5 stimulates mRNA capping. *Genes Dev.* **13:** 1774–1779

West ML & Corden JL (1995) Construction and analysis of yeast RNA polymerase II CTD deletion and substitution mutations. *Genetics* **140:** 1223–1233

Widom J (2001) Role of DNA sequence in nucleosome stability and dynamics. **34**: 269-324

Wier AD, Mayekar MK, Héroux A, Arndt KM & VanDemark AP (2013). Structural basis for Spt5-mediated recruitment of the Paf1 complex to chromatin. *PNAS*. **110**:17290–17295.

Wiest DK, O'Day CL & Abelson J (1996) In vitro studies of the Prp9·Prp11·Prp21 complex indicate a pathway for U2 small nuclear ribonucleoprotein activation. *J. Biol. Chem.* **271:** 33268–33276

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

Will CL & Lührmann R (2011) Spliceosome structure and function. *Cold Spring Harb. Perspect. Biol.* **3:** 1–2

Wood A, Schneider J, Dover J, Johnston M & Shilatifard A (2003) The Paf1 complex is essential for histone monoubiquitination by the Rad6-Bre1 complex, which signals for histone methylation by COMPASS and Dot1p. *J. Biol. Chem.* **278:** 34739–34742

Wood A & Shilatifard A (2006) Bur1/Bur2 and the Ctk complex in yeast: The split personality of mammalian P-TEFb. *Cell Cycle* **5:** 1066–1068

Xiao T, Kao C-F, Krogan NJ, Sun Z-W, Greenblatt JF, Osley MA & Strahl BD (2005) Histone H2B ubiquitylation is associated with elongating RNA polymerase II. *Mol. Cell. Biol.* **25:** 637–51

Xu YZ & Query CC (2007) Competition between the ATPase Prp5 and Branch Region-U2 snRNA Pairing Modulates the Fidelity of Spliceosome Assembly. *Mol. Cell* **28:** 838–849

Xu Y, Bernecky C, Lee CT, Maier KC, Schwalb B, Tegunov D, Plitzko JM, Urlaub H, Cramer P (2017) Architecture of the RNA polymerase II-Paf1C-TFIIS transcription elongation complex. *Nat.* Commun. **8:** 15741

Yan D, Perriman R, Igel H, Howe KJ, Neville M & Ares MJ (1998). CUS2, a yeast homolog of human Tat- SF1, rescues function of misfolded U2 through an unusual RNA recognition motif. *Mol. Cell. Biol.* **18:** 5000-9.

Yang F, Wang XY, Zhang ZM, Pu J, Fan YJ, Zhou J, Query CC & Xu YZ (2013) Splicing proofreading at 5′ splice sites by ATPase Prp28p. *Nucleic Acids Res.* **41:** 4660–4670

Yang Y, Li W, Hoque M, Hou L, Shen S, Tian B & Dynlacht BD (2016) PAF Complex Plays Novel Subunit-Specific Roles in Alternative Cleavage and Polyadenylation. *PLoS Genet.* **12:** e1005794

Yao S, Neiman A & Prelich G (2000) BUR1 and BUR2 encode a divergent cyclin-dependent kinase-cyclin complex important for transcription in vivo. *Mol. Cell. Biol.* **20:** 7080–7087

Yu M, Yang W, Ni T, Tang Z, Nakadai T, Zhu J & Roeder RG (2015) RNA polymerase II-associated factor 1 regulates the release and phosphorylation of paused RNA polymerase II. *Science.* **350:** 1383–1386

Yuan GC, Liu YJ, Dion MF, Slack MD, Wu LF, Altschuler SJ & Rando OJ (2005) Molecular biology: Genome-scale identification of nucleosome positions in S. cerevisiae. *Science.* **309:** 626–630

Yurko N, Liu X, Yamazaki T, Hoque M, Tian B & Manley JL (2017) MPK1/SLT2 Links Multiple Stress Responses with Gene Expression in Budding Yeast by Phosphorylating Tyr1 of the RNAP II CTD. *Mol. Cell* **68:** 913–925.e3

Yurko NM & Manley JL (2018) The RNA polymerase II CTD "orphan" residues: Emerging insights into the functions of Tyr-1, Thr-4, and Ser-7. *Transcription* **9:** 30–40

Yuryev A, Patturajan M, Litingtung Y, Joshi R V, Gentile C, Gebara M & Corden JL (1996) The C-terminal domain of the largest subunit of RNA polymerase II interacts with a novel set of serine/arginine-rich proteins. *Proc. Natl. Acad. Sci. U. S. A.* **93:** 6975–80

Zaret KS & Carroll JS (2011) Pioneer transcription factors: Establishing competence for gene expression. *Genes Dev.* **25:** 2227–2241

Zhan X, Yan C, Zhang X, Lei J & Shi Y (2018) Structure of a human catalytic step I spliceosome. *Science.* **359:** 537–545

Zhang Y, Moqtaderi Z, Rattner BP, Euskirchen G, Snyder M, Kadonaga JT, Liu XS & Struhl K (2009) Intrinsic histone-DNA interactions are not the major determinant of nucleosome positions in vivo. *Nat. Struct. Mol. Biol.* **16:** 847–852

Zhang DW, Mosley AL, Ramisetty SR, Rodríguez-Molina JB, Washburn MP & Ansari AZ (2012) Ssu72 phosphatase-dependent erasure of phospho-Ser7 marks on the RNA polymerase II C-terminal domain is essential for viability and transcription termination. *J. Biol. Chem.* **287:** 8541–8551

Zhang T, Cooper S & Brockdorff N (2015) The interplay of histone modifications - writers that read. *EMBO Rep.* **16:** 1467–1481

Zhou K, Kuo WHW, Fillingham J & Greenblatt JF (2009) Control of transcriptional elongation and cotranscriptional histone modification by the yeast BUR kinase substrate Spt5. *Proc. Natl. Acad. Sci.* **106:** 6956–6961

Zhou H-L, Hinman MN, Barron VA, Geng C, Zhou G, Luo G, Siegel RE & Lou H (2011) Hu proteins regulate alternative splicing by inducing localized histone hyperacetylation in an RNA-dependent manner. *Proc. Natl. Acad. Sci.* **108:** E627–E635