Title:

The roles of Endonucleolytic Cleavage in RNA Metabolism and Transcriptional Termination

Volume 1 of 1

Submitted by Laura Francis to the University of Exeter as a thesis for the degree of Doctor of Philosophy by Research in Biological Sciences

April 2019

I certify that all material in this thesis which is not my own work has been identified and that no material has previously been submitted and approved for the award of a degree by this or any other University.

Signature: …….

Acknowledgements

Firstly, I would like to thank my supervisor, Steven West, for giving me this opportunity and supporting me throughout. All of his advice has been highly appreciated. In addition, thanks for his generation of the *CPSF73-AID* and *INTS11-SMASh* cell lines which were highly important in this work.

I would also like to thank all members of the West lab group for their valuable suggestions, encouragement, companionship and comedic relief: Lee Davidson, Josh Eaton, Chris Estell, Francesca Carlisle, Ryan Kelly. A special thank you to Lee Davidson who conducted the RNA-Seq experiments for the *DIS3-AID* cell line and patiently taught me bioinformatics, so that I could conduct my own analysis. Thanks also to the Exeter Sequencing Service who kindly shared their lab space and expertise and without whom the RNA sequencing experiments would have been exponentially more difficult.

Finally, I am extremely grateful to my friends and family. Thank you to my grandparents, Frank and Gloria, for everything. If it wasn't for them I would not be where I am today. To my Mum, thank you for making me believe that I can achieve anything and for always telling me that you're proud of me. Most importantly thank you again to Ryan, for moving across the country with me and always being there for me. I am so grateful for you being in my life.

Abstract

Eukaryotic gene expression begins with transcription of DNA into RNA by RNA polymerases and for protein coding genes is followed by translation into protein in the cytoplasm. Production of functioning mature RNA relies on proper processing events including 5' capping, splicing and 3' end processing. Endonucleases that cleave RNA are vital for these processing events and are involved in degradation pathways that may also be relevant for the turnover of aberrant transcripts. Studies investigating transcription, processing events and degradation pathways of RNA have generally focused on RNA polymerase II transcripts, which includes protein-coding genes. Many of these pathways were elucidated by studies in yeast due to the high conservation of the transcription process between yeast and metazoans.

The discovery and development of CRISPR/Cas9 mediated genome editing techniques have led to a more complete and direct approach to study specific protein functions, within human cells, than previous methods such as RNAi. In this study, a combination of CRISPR/Cas9 with protein tags including the auxin inducible degron and small molecule assisted shut-off, allowed rapid and conditional protein depletion in human cell lines for three endonucleases, DIS3, INTS11 and CPSF73. These endonucleases cooperate with accessory proteins and actively transcribing polymerase complexes to target a broad range of RNA transcripts, to ensure proper RNA processing and integrity of the transcriptome. Generation of these cell lines, coupled with high-throughput RNA sequencing analysis of nuclear transcriptomes, helped to elucidate specific substrates for each endonuclease. The following work shows the effects of aberrant processing in a variety of transcripts, their subsequent potential degradation and what happens when a major degradation pathway is disrupted. A major finding in this study was disruption of 3' end processing in protein coding mRNA resulted in extensive readthrough and termination defects, whereas 3' end misprocessing in smaller RNA species including snRNAs and replication dependent histones results in a much smaller extension and termination that occurs relatively close to the gene transcription end site. Additionally, this works shows the importance of the exosome subunit, DIS3, in maintaining appropriate gene expression and RNA environment, whilst suggesting aberrant RNA processing may commonly occur in human cells.

List of contents

List of tables

List of figures

Abbreviations

AID	Auxin-inducible degron
AMO	Antisense morpholino oligonucleotide
ARF	Auxin response factors
ATR	Auxin transcriptional repressors
Aux	Auxin / Indole-3-acetic acid / IAA
bp	Base pairs
BPS	Branch point sequence
CBC	Cap binding complex
CDK	Cyclin dependent kinases
cDNA	Complementary DNA
ChIP	Chromatin Immunoprecipitation
CIP	Calf Intestinal Alkaline Phosphatase
CMV	Cytomegalovirus
CPSF	Cleavage/Polyadenylation Specificity Factor
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
CTD	C-Terminal Domain of Rpb1
DCP	Downstream cleavage product
dH_2O	Distilled water
DMEM	Dulbecco's Modified Eagle Medium
DOC	Sodium Deoxycholate
dsDNA	Double Stranded DNA
DSE	Distal sequence element
ECL	Enhanced Chemi-Luminescence
eRNA	Enhancer RNA
FCS	Foetal Calf Serum
gb	Gene body
gRNA	Guide RNA
HCC	Histone cleavage complex

HCV	Hepatitis C virus
HDE	Histone downstream element
HDR	Homology directed repair
IEG	Immediate early genes
IR	Inverted repeats
Kb	Kilobase
kDa	Kilodalton
ml	Millilitres
mm	Millimetres
mM	Millimolar
mRNA	Messenger RNA
ncRNA	Non-coding RNA
NFR	Nucleosome-free region
ng	Nanogram
NHEJ	Non-homologous end joining
nM	Nanomolar
nt	Nucleotides
PAM	Protospacer Adjacent Motif
PAP	Poly(A) polymerase
PAS	Poly(A) site
PIC	Pre-initiation complex
Pol II	RNA Polymerase II
PROMPTs	Promoter Upstream Transcripts
PSE	Proximal sequence element
qPCR	Real Time Quantitative PCR
RDH	Replication-dependent histone
RISC	RNA inducing silencing complex
RNAi	RNA Interference
RNA-Seq	RNA Sequencing
RPKM	Reads per Kilobase of transcript Per Million mapped reads
rpm	Revolutions per Minute

rRNA	Ribosomal RNA
SB	Sleeping Beauty
SDS	Sodium Dodecyl Sulfate
SEC	Super elongation complex
SMASh	Small molecule assisted shutoff
snoRNA	Small nucleolar RNA
snRNA	Small Nuclear RNA
snRNP	Small nuclear ribonucleoprotein
TES	Transcription End Site
TFs	Transcription Factors
TIR1	Transport Inhibitor Response 1
TSS	Transcription Start Site
UC	Uncleaved
UTR	Untranslated region
WT	Wild type
x g	Relative centrifugal force
μg	Microgram
μl	Microlitre
μM	Micromolar

1. Introduction

1.1 RNA Polymerase II and transcription

RNA polymerase II (Pol II) is a 12 subunit complex that transcribes protein-coding RNA (messenger RNA / mRNA) and a variety of functional non-coding RNA (ncRNA) including small nuclear RNA (snRNA), small nucleolar RNA (snoRNA), histones and a plethora of as yet uncharacterised transcripts. The catalytic and largest component of Pol II is Rpb1, whose C terminal domain (hereafter referred to as CTD), contains numerous tandem heptad repeats. These repeats consist of the amino acid consensus sequence Tyr1 - Ser2 - Pro3 - Thr4 - Ser5 - Pro6 - Ser7, with humans having 52 repeats compared to yeast with 26 (Corden, 1990).

Post-transcriptional modifications occur frequently on the CTD, principally phosphorylation at Ser 2 and Ser 5 residues of the heptad repeats by cyclin-dependent kinases (CDKs), although phosphorylation at other residues has been observed (Heidemann et al, 2013). The phosphorylation state of the CTD influences transcription by acting as a platform for multiple transcription factors and other protein complexes. This in turn regulates the three different stages of transcription by Pol II: initiation, elongation and termination (Hsin and Manley, 2012).

1.1.1 Initiation

Transcription initiation requires recruitment of Pol II to the DNA promoter alongside transcription factors (TFs) and an open chromatin structure. Five TFs recognise the TATA-box domain located approximately 25 – 30 nucleotides (nt) upstream of many gene promoters and form the pre-initiation complex (PIC) by binding to the TATA-box. Pol II, with an unphosphorylated CTD, binds to the PIC and initiates recruitment of TFs including helicases that unwind the DNA and CDKs for phosphorylation (Krishnamurthy and Hampsey, 2009).

Pol II release from the initiation complex, allowing it to move into an early elongation phase, is believed to be initiated by CDK7 phosphorylation of the CTD

at Ser5 as Pol II moves along the promoter (Glover-Cutter et al, 2009). At 20 – 60 nts downstream of the transcription start site (TSS), Pol II is paused at a promoter-proximal pause site (Guenther et al, 2007; Kwak et al, 2013). Negative Elongation Factor (NELF) and DRB-sensitivity-inducing factor (DSIF) facilitate Pol II proximal-pausing in a large number of genes (Ping and Rana, 2001). During this pause, a 5' cap is added to nascent RNA and positive elongation factor b (P-TEFb) is recruited to reverse the elongation inhibition effects of NELF (Peterlin and Price, 2006). P-TEFb contains a kinase subunit CDK9 and is an important factor for Pol II release from the pause site and move to elongation. CDK9 phosphorylates NELF, DSIF and the CTD at Ser 2. These interactions in part regulate the release of paused Pol II by recruiting necessary processing factors, releasing NELF from Pol II, converting DSIF to a positive elongation factor and reorganising TFs. A CTD phosphorylated at both Ser 5 and Ser 2 is a hallmark of Pol II transition to elongation (Liu et al, 2015; Kwak and Lis, 2013).

1.1.2 Elongation

During elongation toward the 3' end, phosphorylation of Ser 5 is gradually removed whereas phosphorylation of Ser 2 accumulates, peaking towards the 3' end of genes (Davidson et al, 2014; Mayer et al, 2010; Tietjen et al, 2010; Kim et al, 2010a). Elongation factors are recruited to Pol II and enable it to elongate at a high rate (approximately 4 Kb / minute) (Singh and Padgett, 2009). However, throughout the gene there are variations in the transcription elongation rate and this may be due to a few factors. Firstly, the rate of transcription can be restricted by histone marks causing tightening of DNA binding around nucleosomes and vice versa. Secondly, transcription can be hindered by GC rich DNA areas, which may cause R-loops, or facilitated by elongation factors, histone chaperones and nucleosome remodellers maintaining elongation conducive chromatin (Jonkers and Lis, 2015).

Although Pol II rapidly elongates in a 5' to 3' direction, Pol II also performs retrograde motion during elongation, known as backtracking that is triggered by a weak DNA-RNA hybrid. During backtracking, the active site of Pol II becomes dissociated from the 3' end of RNA, leading to transcriptional arrest (Nudler et al, 1997). These backtracking-mediated pauses in transcription are important for

transcriptional regulation and processing of many genes (Nudler, 2012). In eukaryotic cells, backtracked Pol II elongation complexes can be corrected by transcript cleavage factors (TFIIS / SII) (Reinberg and Roeder, 1987). These factors promote cleavage of the extruded 3' transcript end to produce a new 3' end that realigns with the Pol II active site, allowing transcription to continue (Izban and Luse, 1992). Mutations in TFIIs inhibited intrinsic Pol II transcript cleavage and prevented both transcription through pause sites and elongation (Sigurdsson et al, 2010).

1.1.3 Termination

Termination pathways of Pol II transcription vary between mRNA and ncRNA genes. The termination pathway is thought to be defined by specific termination signals on the nascent RNA and distinctive phosphorylation patterns of the CTD. Currently there are three main pathways of Pol II termination described in metazoans that generate either mRNAs, snRNAs or replication-dependent histone encoding transcripts. However, the most studied termination pathway is that of mRNAs.

It is commonly believed that the poly(A) site (PAS) is required for termination of mRNAs, with Pol II pausing after transcription of the PAS increasing transcription termination efficiency and facilitating selection of alternative PAS sites (Fusby et al, 2016; Eaton et al, 2018). This is supported by studies that observed Pol II accumulation around the PAS (Gromak et al, 2006; Glover-Cutter et al, 2008). As previously mentioned, toward the 3' end of genes the CTD becomes highly phosphorylated on Ser 2. Inhibition of Ser 2 phosphorylation in metazoan cells leads to impaired recruitment of processing factors at 3' ends of genes and defects in mRNA polyadenylation (Ni et al, 2004). Therefore CTD Ser 2 phosphorylation enhances recruitment of processing factors.

Recruitment of processing factors include cleavage and polyadenylation specificity factor (CPSF) and cleavage stimulation factor (CstF) complexes, which recognise the AAUAAA hexamer and G / U rich sequences, respectively, of the PAS (Proudfoot et al, 2011). Co-transcriptional cleavage of transcripts occurs 18 – 30 nts downstream of the PAS by the endonuclease CPSF component,

CPSF73, which forms a heterodimer with CPSF100. This cleavage releases the nascent RNA, allowing polyadenylation factors to bind to the 3' end.

Observations that cleavage was required for termination lead to the development of the "torpedo" model of transcription termination. In this model, Pol II continues to transcribe a downstream transcript after cleavage. For termination to occur, this downstream transcript is degraded by a 5' – 3' exoribonuclease, namely Rat 1 in yeast (Kim et al, 2004) and the homolog XRN2 in humans (West et al, 2004). Upon the exoribonuclease reaching transcribing Pol II, it acts as a trigger to release Pol II from the DNA and therefore cause termination. XRN2 termination is enhanced by pausing of Pol II which may be caused by R-loops. R-loops are a nucleic acid structure consisting of two antiparallel DNA strands and a RNA strand, creating a DNA:RNA hybrid that particularly form over G-rich terminator elements (Skourti-Stathaki et al, 2011). Interestingly, the homolog of yeast Sen1, Senataxin (SETX), may also play a role in termination of some mRNAs (Suraweera et al, 2009; Wagschal et al, 2012). SETX is a RNA:DNA helicase and may facilitate XRN2 degradation of downstream transcripts by resolving R-loops and exposing DNA.

An alternative method for transcription termination is the "allosteric" model. It is proposed that transcription of the PAS causes a conformational change in the Pol II elongation complex. This change leads to termination by recruitment of termination factors and / or dissociation of elongation factors (Logan et al, 1987;). Support for this model has come from studies showing cleavage is not required for termination and thus disputing the "torpedo" model (Osheim et al, 1999; Osheim et al, 2002). Additionally, Zhang et al (2015a) observed PAS-dependent termination could occur without the requirement of cleavage. However, a more recent study argues against a cleavage independent method for transcription termination (Eaton et al, 2018). They showed CPSF73 loss caused extensive read-through transcription and that catalytically inactive CPSF73 could not restore termination to cells depleted of CPSF73.

1.1.4 Structure and biology of eukaryotic mRNA

In eukaryotes, after transcription of a gene, a pre-mRNA is produced that then undergoes multiple processing events to become a mature mRNA, some of which occur co-transcriptionally (Proudfoot et al, 2002). A mature mRNA contains a 5' cap, which consists of a guanine nucleotide connected to the mRNA via a 5' – 5' triphosphate linkage; the mRNA is polyadenylated at the 3' end, where approximately 200 adenosine residues are added to form the poly(A) tail; and mature mRNA are spliced, meaning the introns are removed from the pre-mRNA and exons ligated together to form mature mRNA (Proudfoot et al, 2002). Between the cap and coding sequence of the mRNA, there is a 5' untranslated region (UTR) that regulates translation of a transcript and is commonly not translated itself (Moore, 2005). Similarly, there is a 3' UTR found between the coding sequence and poly(A) tail of mRNA. These UTRs have roles in mRNA export, localisation, stability and translation efficiency (Matoulkova et al, 2012).

Mature mRNA is recognised by its processed modifications and exported from the nucleus into the cytoplasm, by cap binding proteins and the TREX complex, where it can be translated into a protein (Katahira, 2012). At the ribosome, the coding region of mature mRNA is translated into a protein. Upon the small ribosomal subunit, attached to the mRNA, reaching the start codon (commonly AUG) the large ribosomal subunit and the initiation tRNA join. The ribosome reads the coding sequence in a set of 3 nts, called a codon. A tRNA corresponding to the codon sequence transfers an amino acid to the growing polypeptide chain, continuing until the ribosomal subunits reach the stop codon and the polypeptide is released. The polypeptide then undergoes folding to become a functional protein (Cooper, 2000).

1.2 Diverse transcripts of Pol II

Pol II not only transcribes mRNAs, but it also produces RNA transcripts that lack polyadenylation including: snRNAs and replication-dependent histones (RDH). snRNAs play a critical role in mRNA processing, formation of the spliceosome, regulation of transcription factors, expression and processing of histone mRNA and ribosomal RNA (rRNA) biosynthesis. The majority of RDH proteins, which act to package newly replicated DNA into chromatin, are encoded by RDH genes which are physically linked in large genome clusters (Marzluff et al, 2002). RDH and snRNA transcripts have alternative processing and transcription pathways than described above. However, some mechanisms do overlap, for example the CTD appears to play a role in 3' end processing of both snRNAs and RDHs and CPSF73 is involved in cleavage of protein-coding mRNAs and RDH pre-mRNA (Jacobs et al, 2004; Hsin et al, 2011).

1.2.1 RDH processing

RDH genes are rapidly transcribed in the S-phase of the cell cycle, to coordinate with DNA replication and generally lack both introns and polyadenylation, instead having a conserved stem-loop at their 3' untranslated region (UTR). Downstream of the stem-loop, RDHs contain a purine-rich histone downstream element (HDE) and cleavage of RDH pre-mRNA occurs in-between the stem loop and HDE regions. For processing to occur, a stem-loop binding protein (SLBP) binds to the stem-loop region and a small nuclear ribonucleoprotein, U7 snRNP, binds to the HDE (Dominski and Marzluff, 2007). U7 snRNP contains a heptameric Sm ring and U7 snRNA with a complementary sequence to the HDE at its 5' end to allow base-pairing binding to the HDE. SLBP is thought to stabilise U7 snRNP binding to RDH pre-mRNA, possibly by interaction with the U7 snRNP subunit ZFP100 (Dominski et al, 2002).

Lsm10 and Lsm11 are U7 snRNP specific subunits, replacing spliceosomal SmD1 and SmD2 in the Sm ring (Pillai et al, 2001; Pillai et al; 2003). Lsm11 contains an extended N-terminal domain that interacts with Flice-associated huge protein (FLASH) and ZFP100 (Yang et al, 2009a; Wagner and Marzluff, 2006). Together, Lsm11 and FLASH form a docking platform that recruits the histone cleavage complex (HCC), consisting of multiple

polyadenylation subunits including CPSF100 (homolog to Integrator subunit 9), Symplekin and CPSF73 endonuclease. CPSF73 is part of the β-CASP (metallo-β-lactamase-associated CPSF Artemis SNM1/PSO2) family, whose protein members contain features amicable for endonuclease function. CPSF100 is also a β-CASP protein, however critical residues in the active site are altered suggesting it is catalytically inactive (Mandel et al, 2006; Callebaut et al, 2002). Similar to cleavage of protein-coding mRNAs, CPSF73 is responsible for the cleavage of RDH pre-mRNA (Dominski et al, 2005) (Figure 1.1).

As observed in both humans and *Drosophila melanogaster*, misprocessing of RDH pre-mRNA leads to their polyadenylation due to read-through and the usage of a secondary downstream polyadenylation signal (Kari et al, 2013; Romeo et al, 2014; Sullivan et al, 2009). In contrast to properly processed RDH mRNA, these polyadenylated histones are stable throughout the cell cycle (Levine et al, 1987).

The description above for RDH pre-mRNA processing may not be the full story however. Recently, Pettinati et al (2018) found another protein that appears to have a critical role in RDH 3' processing and showed that it has endoribonucleolytic activity in vitro. MBL domain containing protein 1 (MBLAC1) contains a MBL domain with similar di-zinc ion binding to CPSF73, although they have differing active site flanking loops and only CPSF73 contains a β-CASP domain. Depletion of MBLAC1 in HeLa cells caused a cell cycle defect, with accumulation of cells in G_1 / early S phase. Additionally, read-through of approximately 200 bp downstream of the RDH transcription end site (TES) was observed when CPSF73 and MBLAC1 were depleted, with both genes expressing a similar transcription termination defect pattern for RDHs. It was suggested that MBLAC1 and CPSF73 may selectively affect different RDH pre-mRNA 3' end processing or have varying impact on similar genes, with potential redundancy.

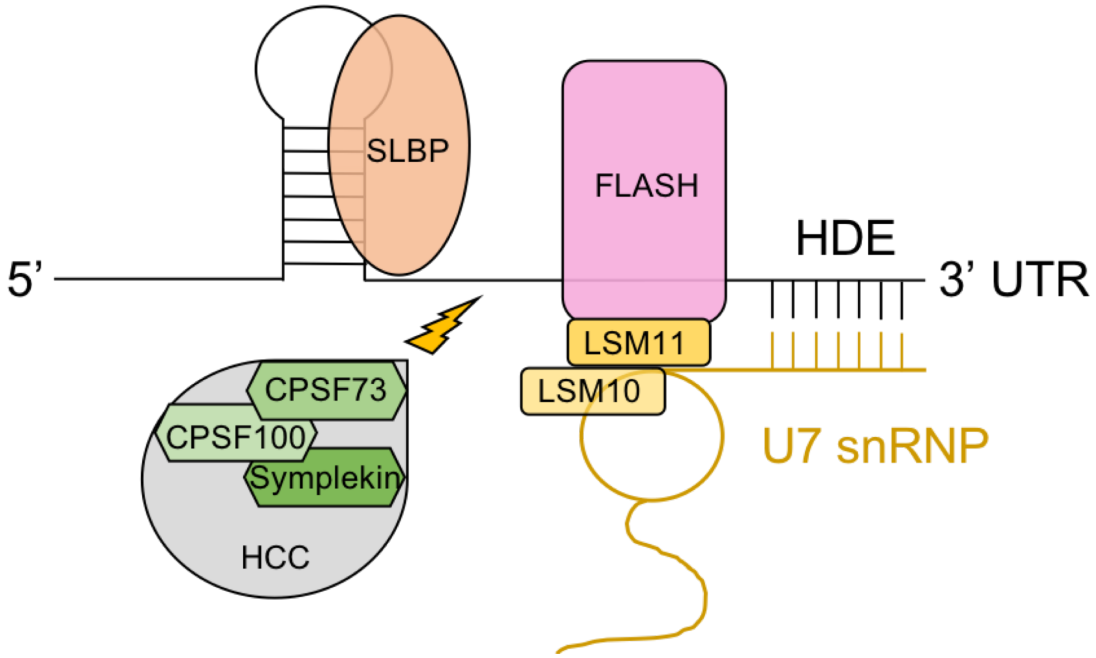

Figure 1.1: Replication-dependent histone (RDH) processing

The stem loop binding protein (SLBP) binds to the stem-loop of RDH pre-mRNA and aids in stabilization of U7 snRNP binding to the RDH pre-mRNA at the histone downstream element (HDE). Within the heptameric Sm ring structure of U7 snRNP are two spliceosomal subunits, Lsm10 and Lsm11. Lsm11 interacts with FLASH and together they recruit the histone cleavage complex (HCC). The HCC includes CPSF100, Symplekin and CPSF73, the latter of which is responsible for cleavage of the pre-mRNA between the stem loop and HDE. Overall this produces unpolyadenylated mature RDH mRNAs. For simplicity, other potential proteins involved in this process haven't been shown.

1.2.2 snRNAs

snRNAs are uridine-rich, approximately 60 – 200 nts long and play a critical role in spliceosome formation. They are transcribed by Pol II, with the exception of U6 snRNA which is transcribed by RNA polymerase III. snRNAs are not polyadenylated, they do not contain a TATA-box sequence and lack introns. Similar to histone genes, snRNAs are also found within clusters of the genome and have multiple copies (Chen and Wagner, 2010). The promoter of snRNAs contain two elements: an enhancer-like distal sequence element (DSE) that recruits transcription factors Oct1 and Sp1 and a proximal sequence element (PSE). The PSE, as well as specific phosphorylation of the CTD and a consensus sequence (3' box) located 9 – 19 nts downstream of the snRNA coding region, are required for 3' end snRNA processing (Chen and Wagner, 2010).

Transcription initiation and the phosphorylation pattern of the CTD differs at snRNA genes, compared to protein-coding genes as previously described. In brief, initiation is mediated by the snRNA activator protein complex binding to the PSE, which recruits Pol II to the promoter. After Pol II recruitment, Ser 5 is phosphorylated by the CDK7 subunit of TFIIH. In addition, CTD Ser 7 is also phosphorylated by CDK7, has been shown to be essential for processing and facilitates interactions with a snRNA processing complex (Egloff et al, 2007; Egloff et al, 2010). Ser 7 phosphorylation may allow interaction with RNA Polymerase II Associated Protein II (RPAPII), which dephosphorylates Ser 5 as Pol II transcribes the snRNA and recruits snRNA 3' end processing factors (Egloff et al 2011).

Conversely, Hsin et al (2014) mutated Ser 7 to an alanine in a chicken DT40 cell line and found no defects on snRNA levels or processing. This discrepancy may be a result of the use of chicken vs human cell lines, although this seems unlikely as there is a high degree of conservation in snRNA genes and their processing factors. Alternatively, findings from Egloff et al may be linked to their use of α-amanitin. Treatment with α-amanitin can increase degradation of some proteins, including DSIF, which plays a role in snRNA expression (Tsao et al, 2012; Yamamoto et al, 2014; Laitem et al, 2015). Thus the observed phenotype may be due to a reduction in DSIF accumulation and not due to the Ser 7 phosphorylation state. As snRNAs are relatively short transcripts, Ser 2 CTD phosphorylation for efficient elongation as seen in mRNAs is not required.

However, Ser 2 phosphorylation by P-TEFb instead plays an important role in snRNA 3' end formation (Medlin et al, 2005). It is thought PTEFb phosphorylates Ser 2 near the 3' end of snRNA genes and with Ser7 allows recruitment of necessary snRNA 3' end processing factors, specifically INTS9 and INTS11 subunits belonging to the Integrator complex (Zaborowska et al, 2016; Egloff et al 2010) (Figure 1.2).

1.2.3 snRNA 3' end processing and termination

Integral to snRNA processing is a complex called the Integrator (Baillat et al, 2005; Ezzeddine et al, 2011). The Integrator is formed of 12 - 14 subunits, including a homolog of CPSF73 (INTS11) and a homolog of CPSF100 (INTS9). In humans, these proteins are numbered in order of predicted molecular mass, with INTS1 having the largest mass (Chen and Wagner, 2010). The Integrator is recruited to the snRNA promoter, possibly through RPAPII, and associates with the CTD, travelling with Pol II as it transcribes the snRNA. Upon transcription and recognition of the 3' box, the nascent 3' snRNA is cleaved by the catalytic endonuclease subunit of the Integrator, INTS11 (Baillat et al, 2005). INTS9 and INTS11 form a heterodimeric complex that is thought to be functionally required for snRNA 3' end processing and are recruited later than other Integrator subunits (Dominski et al, 2005; Albrecht and Wagner, 2012; Egloff et al, 2011) (Figure 1.2). They are also members of the β-CASP family, however INTS9 contains the same amino acid changes that are suggested to inactivate catalytic activity in CPSF100 (Chen and Wagner, 2010). Depletion of INTS9 and INTS11 has been shown to cause accumulation of misprocessed snRNA (Ezzeddine et al, 2011; Baillat et al, 2005). Recently, depletion of the integrator subunit 4 (INTS4) was shown to have a similar defect in snRNA processing to that observed upon INTS9 or INTS11 depletion. It was also reported that INTS4 specifically interacts with the INTS9 / INTS11 heterodimer to potentially form a heterotrimeric integrator cleavage module (Albrecht et al, 2018).

Cleavage of precursor snRNA into mature snRNA is linked with efficient transcription termination, as demonstrated by disruption of snRNA termination causing inefficient snRNA processing and vice versa (Ramamurthy et al, 1996; O'Reilly et al, 2014).

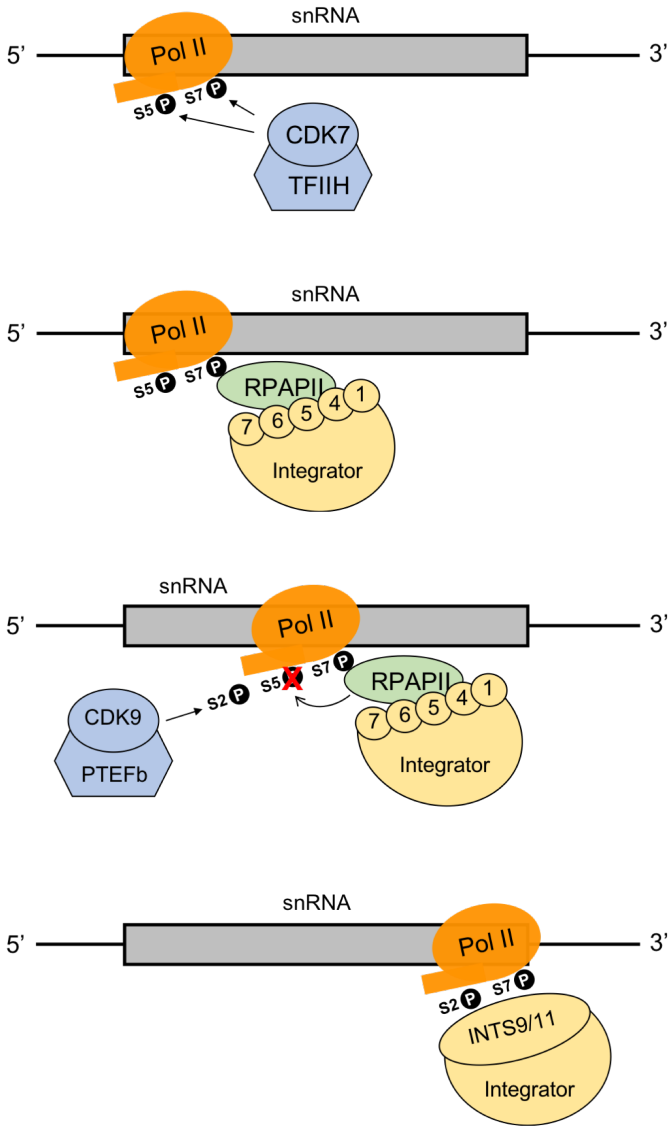

Figure 1.2 Pol II phosphorylation and recruitment of the Integrator at snRNA genes

After recruitment of Pol II to the TSS of a snRNA gene, TFIIH phosphorylates the Pol II CTD at Ser 5 and Ser 7 through its CDK7 subunit. Phosphorylated Ser7 interacts with RPAPII, which recruits the Integrator complex. It is thought that catalytic subunits INTS9 and INTS11 are not recruited at this time. As Pol II transcribes the snRNA, Ser5 is dephosphorylated by RPAPII and the CDK9 subunit of PTEFb phosphorylates Ser2 near the 3' end of the snRNA. This phosphorylation state of Pol II may then allow recruitment of the INTS9/INTS11 heterodimer and therefore snRNA 3' end processing. Figure adapted from Guiro and Murphy (2017).

Currently, the mechanisms of snRNA transcription termination are not fully understood although it has been suggested chromatin structure, polyadenylation factors and both DSIF and NELF play a role (Egloff et al, 2009; O'Reilly et al, 2014; Yamamoto et al 2014). Interestingly, NELF knockdown causes Pol II to transcribe past the 3' box, creating read-through transcripts and suggesting NELF is essential for proper transcription termination (Yamamoto et al, 2014). NELF was commonly believed to act only at promoter-proximal regions (Sun et al, 2011). Consistent with this, ChIP analysis confirmed NELF signal accumulated around the TSS of beta-actin mRNA compared to 300 bp downstream. Conversely, these findings weren't replicated in snRNA. NELF signal was higher at 180 and 370 bp downstream compared to the TSS of U1 snRNA, showing a difference in NELF localisation at these genes. In addition, NELF was found to interact with the Integrator and knockdown caused accumulation of uncleaved snRNAs (Yamamoto et al, 2014).

1.2.4 Additional functions of the Integrator

Findings from more recent studies have implicated a role for the Integrator in other aspects of transcriptional regulation. Firstly, Skaar et al (2015) found that the Integrator not only has a role in snRNA termination, but also termination of RDHs and genes with polyadenylated mRNAs. HIT-Seq and ChIP-Seq methods were utilised and the authors found extensive binding of the Integrator to the 3' end of RDHs. Depletion of Integrator subunit, INTS3, caused a significant increase in unprocessed RDH transcripts with poly(A) tails. Furthermore, INTS3 knockdown resulted in an increased localisation of Pol II downstream of RDH genes, suggesting a defect in Pol II termination. The Integrator was also found localised at the TSS of various gene types, reflecting binding of DSIF and NELF at these same locations. Binding of the Integrator at promoter-proximal sites was found to negatively regulate expression of genes with polyadenylated mRNAs (Skaar et al, 2015).

The Integrator also functions in initiation and Pol II pause-release at protein-coding genes. As discussed previously, P-TEFb is responsible for phosphorylation of multiple units at the proximal-pause site of protein-coding mRNAs which leads to Pol II pause-release. P-TEFb also exists as an active

factor of the larger multi-subunit super elongation complex (SEC) (Lin et al, 2010; Luo et al, 2012). Gardini et al (2014) used epidermal growth factor (EGF) in HeLa cells to promote transcription of immediate early genes (IEGs) by Pol II, which are known for their regulation through Pol II pause-release. They found the Integrator was necessary for recruitment of SEC-containing P-TEFb to paused Pol II, leading to Pol II pause-release and elongation. EGF stimulation caused a robust increase in Integrator occupancy at IEG TSS and 3' ends, as well as the TSS and body of EGF-responsive genes, suggesting the Integrator remains associated with elongating Pol II at these genes. Depletion of Integrator subunits INTS1 or INTS11 caused a loss of EGF-response. Diminished transcriptional activation, decrease in Pol II occupancy of nascent RNA and abolishment of two SEC components to IEGs was observed upon INTS11 depletion. Overall these findings suggest the Integrator has additional roles at protein coding genes, by association with the SEC complex and facilitating initiation and pause-release.

In support of an Integrator role in Pol II pause-release, Stadelmayer et al (2014) found the Integrator regulates NELF-mediated Pol II pause-release at coding genes. Genes bound by NELF and INTS3 showed a decreased pausing index and lower Pol II occupancy at the TSS. Depletion of INTS3 reduced Pol II occupancy over NELF-regulated genes, whereas INTS11 depletion increased Pol II occupancy at promoters bound by NELF and INTS3 but not at the 3' end. This resulted in defective RNA processing and is in accord with the Integrator recruiting SEC to promote elongation. However, Stadelmayer et al (2014) contrasts the findings of Gardini et al (2015) that showed INTS11 knockdown decreased Pol II occupancy. This may highlight the two functions of NELF in reducing transcription in non-induced conditions, whilst at the promoter helping to maintain open chromatin structure. Alternatively, this contrast may reflect the differences in cellular context and gene type, with Gardini investigating IEGs and Stadelmayer focusing on genes with increased transcription upon NELF and Integrator depletion. Regardless, both studies promote a role for the Integrator in transcriptional regulation of protein coding genes.

The additional functions of the Integrator discussed here are only some of the roles that have been postulated. Studies have reported a role of the Integrator in eRNA biogenesis, DNA damage response and viral miRNA biogenesis (Lai et al, 2015; Skaar et al, 2009; Cazalla et al, 2011; Xie et al, 2015). This list is not

extensive, but it does suggest the Integrator is important for a number of biological processes, with mutations or altered expression changes in Integrator genes being linked to several diseases (Rienzo and Casamassimi, 2016).

1.2.5 Cryptic transcripts

In addition to mRNA, Pol II also transcribes other types of polyadenylated transcripts. A cryptic transcript is a broad term for transcribed RNA that is highly unstable, meaning it is normally rapidly degraded and not detected in the cell. Upon depletion or defects in nuclear RNA surveillance pathways, these cryptic transcripts are revealed. In yeast these short, capped and polyadenylated transcripts are known as cryptic unstable transcripts (CUTs), which were found to be widely stabilised upon loss of a catalytic subunit of the exosome, Rrp6, that is responsible for degradation of RNA. CUTs are normally targets for degradation by the Nrd1-exosome-TRAMP complexes, immediately after synthesis (Wyers et al, 2005).

CUTs are derived from transcription of unannotated intergenic regions and transcription at bidirectional promoters. Studies in *S. cerevisiae* showed initiation sites for CUTs are often located in nucleosome-free regions (NFRs), which is common for sites around an active gene promoter. Additionally, it was shown many CUTs derived from transcription in the antisense direction to a protein-coding gene, with CUT initiation beginning near the TSS of active protein-coding genes (Neil et al, 2009). Another initiation site was found downstream of stop codons, which contains NFRs (Xu et al, 2009). Therefore, NFRs at 5' and 3' ends of protein-coding genes appear to be suitable locations for CUT transcription and bidirectional promoters may promote this pervasive transcription by maintaining NFRs.

Bidirectional promoter transcription can also be observed in humans and related to CUTs in yeast is the divergent transcription of Promoter Upstream Transcripts (PROMPTs). PROMPTs are generated between 500 and 2500 nts upstream of TSS of promoters for Pol II, RNA polymerase I and RNA polymerase III transcribed genes (Preker et al, 2011). Both CUTs and PROMPTs can be transcribed in a sense or antisense direction, depending on the downstream

gene; they are relatively small; they are polyadenylated and are only detectable upon depletion of components of the exosome (Preker et al, 2008).

PROMPTs are structurally similar to protein-coding mRNA transcripts, in that they contain a 5' cap and 3' adenosine tail, suggesting they are also processed by similar transcription machinery. In support of this, Pol II CTD phosphorylation was similar between PROMPTs and mRNAs at equal distances (Preker et al, 2011). However, PROMPT 3' adenylation has been shown to utilise PAPD5 (elsewhere referred to as Trf4-2), a homolog of yeast Trf4p that is part of the Trf4 / 5-Air1 / 2-Mtr4 Complex (TRAMP) (Preker et al, 2011). In yeast, the Nrd1-Nab3 pathway is used to terminate CUTs (Thiebaut et al, 2006). TRAMP polyadenylates terminated CUTs, rRNAs and snoRNAs and this facilitates their subsequent degradation or 3' end processing catalysed by the nuclear exosome (LaCava et al, 2005; Kadaba et al, 2006; Egecioglu et al, 2006). In contrast to CUTs, where both the exosome and TRAMP complex are required for degradation, 3' adenylation of PROMPTs is not required for their degradation by the exosome (Reis et al, 2007; Preker et al, 2011).

Whilst mRNA transcription from a bidirectional promoter is predominately elongation competent, the opposing-direction PROMPT transcription terminates early. This early termination and subsequent PROMPT degradation is affected by the location of proximal PASs that are more abundant upstream than downstream of the mRNA TSS (Ntini et al, 2013). PROMPTs were found to harbour PAS hexamers 10 – 30 nts upstream of their 3' end as well as CstF64 binding sites downstream of the 3' end (Ntini et al, 2013). In addition, 5' splice site sequences which are able to suppress PAS utilisation, are accumulated in proximal mRNAs compared to PROMPTs and therefore protect mRNAs from premature termination (Kaida et al, 2010).

Chen et al (2016) conducted genome-wide RNA profiling methods in HeLa cells and found a correlation of PROMPT stability and length with distance between mRNA promoters. Gene TSS can be closely positioned to each other, which can cause transcriptional overlap especially when bidirectional transcription occurs. The authors showed that neighbouring promoters with larger distances between them produce PROMPTs that are readily degraded by the exosome and whose 3' ends are believed to be defined by TSS proximal PASs. However, neighbouring promoters in close proximity cause PROMPT

transcription to overlap with mRNA sequences and instead these PROMPTs utilise the distal PAS site of the mRNA for 3' processing. These PROMPTs have been described as alternative mRNA isoforms.

1.3 Co-transcriptional RNA modifications

RNA modifications include 5' capping, splicing, cleavage and polyadenylation all of which are co-transcriptional events. These events act as methods of gene regulation to allow precise control of gene expression. Coordinating RNA processing events with Pol II transcription is facilitated by the CTD. Phosphorylation of the CTD modulates the interactions and actions of RNA processing factors, with the CTD also acting as a platform for these proteins (Figure 1.3).

1.3.1 5' Capping

Addition of a 5' cap to pre-mRNA is important for multiple reasons: to prevent 5' – 3' degradation of nascent mRNA, to aid recruitment of protein factors for splicing, polyadenylation and nuclear export and for recognition by initiation factors to facilitate and maintain efficient translation (Ramanathan et al, 2016). 5' capping occurs early in transcription, on Pol II transcripts specifically. Phosphorylation of the CTD aids this specificity by recruiting capping enzymes. Nascent pre-mRNAs are capped on their 5' end as the first 25 – 30 nts extrude from the active site of transcribing Pol II (Zhou et al, 2012). Two enzymes are responsible for 5' capping, a RNA guanylyltransferase, RNGTT, containing both triphosphatase and guanylyltransferase activity and a RNA guanine-7-methyltransferase, RNMT-RAM (Ramanathan et al, 2016). Firstly the 5' triphosphate end of mRNA is hydrolysed by triphosphatase activity to a diphosphate. The diphosphate is then capped with GMP by guanylyltransferase activity and finally this cap is converted to a 7-methylguanosine cap by RNMT-RAM (Shuman, 2001; Varshney et al, 2018). This process is reversible and decapping can generate an entry site for XRN2 degradation as well as causing premature transcription termination (Davidson et al, 2012; Brannan et al, 2012).

1.3.2 Splicing

Splicing involves the removal / excision of introns from pre-mRNA and ligation of exons, mediated by the spliceosome. For intron removal, cleavage occurs at conserved sequences known as splice sites, found at the 5' and 3' ends of introns (GU and AG respectively). Approximately 18 – 40 nts upstream of the 3' splice site is a branch point sequence (BPS) which is required for splicing, along with a polypyrimidine tract located between the BPS and 3' splice site in humans. Major components of the spliceosome are U1, U2, U4, U5 and U6 snRNPs and spliceosome assembly occurs anew for each splicing reaction (Ward and Cooper, 2010). The first step in splicing involves U1 snRNP binding to a complementary sequence found within the intron. This catalyses cleavage of the intron at the 5' end. The cut 5' end then forms a lariat through a transesterification process, pairing guanine and adenine nts of the 5' end and BPS. The other spliceosome snRNPs are recruited to form a functioning spliceosome, which contributes to positioning of the lariat, release of the lariat by cleavage at the 3' end and ligation of the adjoining exons (Herzel et al, 2017).

Splicing occurs co-transcriptionally, with spliceosome formation relying upon transcription of the 5' and 3' splice sites (Wang and Burge, 2008). Therefore, Pol II elongation is a rate-limiting step of splicing and can be regulated to allow or prevent alternative splicing (Bentley, 2014). In fact, a slow mutant Pol II has been shown to increase the inclusion of alternative exons (de la Mata et al, 2003) and it has been postulated that a specific Pol II elongation rate is required for alternative exon inclusion, potentially caused by nucleosome density slowing elongation (Saldi et al, 2016).

Alternative splicing has been attributed for the existence of multiple mRNA transcripts from single genes and can explain the numerous proteins produced from relatively few genes. It is suggested that at least 90 % of human genes undergo alternative splicing, with introns being retained or exons being extended or skipped (Wang et al, 2008). Consequently, alternatively spliced mRNAs will produce proteins with different amino acids sequences and often different function to their constitutively spliced counterparts. Splice-site selection is regulated by various proteins such as SR proteins, which contain long repeats of serine and arginine residues, and hnRNPs (Martinez-Contreras et al, 2007; Long and Caceres, 2009).

SR proteins recognise short RNA motifs in the pre-mRNA that when bound to exons commonly act as splicing enhancers and conversely repress splicing when bound to introns (Änkö et al, 2014). In constitutive splicing, SR proteins promote U1 snRNP and U2 snRNP binding to the 5' and 3' splice sites, respectively. SR proteins promote alternative splicing by promoting spliceosome formation at weaker 5' splice sites of alternative exons (Jeong, 2017). hnRNPs also recognise specific sequences of RNA but mainly act as splicing silencers, although some can also act as splicing activators i.e. hnRNPL (Martinez-Contreras et al, 2007). Typically, SR proteins bind to cis-acting elements i.e. exonic splicing enhancers or intronic splicing enhancers to promote splicing, whereas hnRNPs bind to exonic or intronic splicing silencers. It is the interplay of hnRNPs and SR proteins binding to enhancer or silencer sites, located within the vicinity of exon/intron junctions, that either promote or inhibit splicing at weak splice sites and therefore govern alternative splicing (Wang et al, 2015).

1.3.3 Cleavage and polyadenylation

Most protein-coding genes that have undergone cleavage by CPSF / CstF factors are then polyadenylated. Firstly, the pre-mRNA contains a 3' – OH which is polyadenylated by the poly(A) polymerase (PAP). To increase the affinity of PAP, poly(A) binding protein nuclear 1 (PABPN1) binds to the newly formed short poly(A) tail and suppresses proximal PASs (Jenal et al, 2012). PAP continues to increase the poly(A) tail length by addition of adenosine monophosphate units. PABPN1 interacts with CPSF and PAP to control poly(A) tail length and upon reaching a length of approximately 250 nts, PABPN1 stops or disrupts these interactions (Kuhn et al, 2009). Thus, the polyadenylation factors dissociate and polyadenylation terminates. The poly(A) tail length is important in initiation of translation and in protection of mRNAs from degradation (Eckmann et al, 2011). Cleavage and polyadenylation has been previously discussed in this work with regards to termination and processing of transcripts. However, there are some contradictory and interesting findings that will be mentioned in this section in addition to what has already been described.

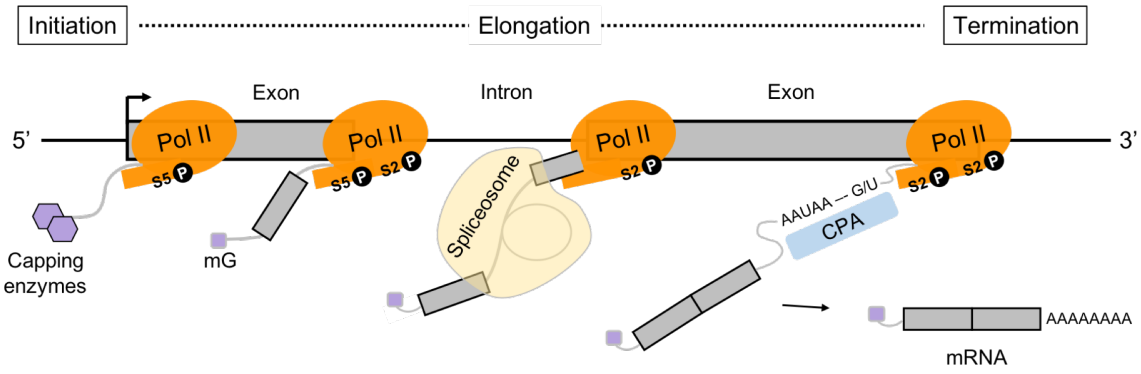

Figure 1.3 Co-transcriptional RNA modifications

Pol II transcription consists of initiation, elongation and termination phases with co-transcriptional RNA modifications. At the 5' end capping enzymes add a methylguanosine cap (mG); during elongation the spliceosome complex splices introns from the pre-mRNA and ligates exons; at the 3' end cleavage and polyadenylation factors (CPA) cleave the pre-mRNA and add a poly(A) tail. Throughout transcription the phosphorylation state of the CTD changes to aid recruitment of processing factors and overcome proximal-pausing of Pol II.

As previously mentioned and in support of the allosteric termination model, an in vitro study by Zhang et al (2015a) observed cleavage was not necessary for Pol II termination. However, another study refutes this by observing highly delayed termination of protein-coding mRNA upon CPSF73 depletion, suggesting PAS cleavage is indeed required for termination (Eaton et al, 2018). The authors used ChIP experiments in CPSF73 depleted HCT116 cells and found both decreased Pol II signal in the gene body, suggesting a strong reduction in transcription, and accumulation of Pol II after the TES showing a large termination defect. When comparing XRN2 and CPSF73, CPSF73 loss caused a greater termination defect than XRN2 depletion, suggesting cleavage of protein coding mRNA is important for promoting Pol II termination. These contrasting results may be due to the experimental systems used, with Zhang et al (2015a) using an in vitro system compared to a human cell line in Eaton et al (2018).

Eaton et al (2018) also found no role for XRN2 in snRNA or RDH gene termination, even though both undergo 3' processing. This is supported by another study who showed degradation of the downstream cleavage product (DCP), formed from cleavage of RDH pre-mRNA, does not require XRN2 (Yang et al, 2009b). Instead the study suggested CPSF73 was involved in degradation. Degradation patterns of the DCP demonstrate the DCP is degraded in a 5' – 3' direction, therefore utilising exonuclease activity. This exonuclease activity was blocked by inhibiting U7 snRNP binding to the HDE and inhibiting CPSF73 recruitment. UV cross-linking demonstrated that CPSF73 specifically interacts with the DCP, in a U7-dependent manner. Yang et al (2009b), concluded that CPSF73 exonuclease activity degraded DCP and that HDE distance requirements upstream and downstream of the cleavage site determine CPSF73 endonuclease or exonuclease activity for cleavage of RDH pre-mRNA. Interestingly, β-lactamase fold protein, Artemis, which contains the same β-CASP domain as CPSF73, shows both exonuclease and endonuclease activity (Ma et al, 2002). These are not the only enzymes suggested to have both exonuclease and endonuclease activities, indeed RNase J is another example and is involved in RNA processing and degradation (Mäder et al, 2008; Even et al, 2005; Mathy et al, 2007; Daou-Chabo and Condon, 2009). However, currently CPSF73 has not been directly shown to have exonuclease activity.

It has already been mentioned that NELF localises downstream of the TSS of snRNAs and knockdown causes snRNA processing defects. A similar role for NELF in RDH pre-mRNA processing, alongside another protein complex, has also been reported (Narita et al, 2007). Using immunoprecipitation in HeLa cells, it was shown NELF interacts with two cap binding complex (CBC) subunits, CBP80 and CBP20. The well-characterised function of the CBC is to bind the 5' cap of pre-mRNA to facilitate export for translation. Defects in cleavage of RDH pre-mRNA results in polyadenylated transcripts by utilising downstream PASs. Interestingly, knockdown of either NELF or CBC resulted in accumulation of polyadenylated RDHs, suggesting these proteins may play a role in 3' processing of RDH pre-mRNA (Narita et al, 2007). The CBC was shown to directly interact with SLBP and pull-down assay confirmed that the CBC is sandwiched between NELF and SLBP. SLBP knockdown also resulted in abnormal RDH mRNA processing and therefore it was hypothesised that NELF may recruit SLBP to the RDH stem-loop through an interaction with the CBC (Sullivan et al, 2001; Narita et al, 2007).

Similar to defects in 3' processing of RDH pre-mRNA, misprocessed snRNAs often become polyadenylated. However, snRNA genes don't commonly contain a PAS closely downstream of their 3' box, where transcription termination predominately occurs. Yamamoto et al (2014) questioned if aberrant polyadenylation could occur in a similar method to mRNA 3' end processing. Interestingly, CPSF73 knockdown caused a significant decrease in polyadenylated U1 snRNA. CPSF73 or CtsF-64 knockdown was also able to rescue accumulation of polyadenylated U1 snRNA caused by NELF knockdown. Overall this suggests that NELF, present at 3' end of snRNAs, may play a role in inhibition of CPSF and CstF mediated cleavage and polyadenylation.

1.3.4 Other RNA modifications

Modification of RNA can also occur internally, for example N^1-methyladenosine, 5-methylcytosine and isomerisation of Uridine (Roundtree et al, 2017; Desrosiers et al, 1974; Carlile et al, 2014). The most prevalent internal modification of both mRNA and long non-coding RNA is N^6-Methyladenosine (m6A), where the N^6 position of adenosine in mRNA is methylated (Perry and Kelley, 1974; Desrosiers et al, 1975; Wei et al, 1975). Lavi et al (1977) estimated poly(A) mRNA contained m6A modification every 1 per 700-800 nts. This essential modification has been found to accelerate mRNA processing and transport in mammalian cells (Camper et al, 1984; Finkel and Groner, 1983). The m6A modification is produced by a METTL3 and METTL14 heterodimer, with METTL3 providing catalytic activity, that is regulated by its association with a WTAP protein subunit (Liu et al, 2014; Ping et al, 2014; Wang et al, 2016). Previous research has suggested that methylation occurs preferentially in 3' UTRs, around the stop codon and also within intronic sequences. This could show that m6A modification occurs co-transcriptionally (Liu et al, 2014; Ping et al, 2014).

Biological functions of m6A are produced by interactions with m6A readers that specifically recognise the RNA modification. These include the YTH family of proteins that then allow m6A regulation of cellular processes (Dominissini et al, 2012) For example, YTHDC1 binding to m6A modifications of mRNA increases the inclusion of alternative exons through interactions with SR proteins (Xiao et al, 2016). Additionally, hnRNP proteins also interact with m6A modified RNAs to regulate alternative splicing. HNRNPC and HNRNPG recognise and bind m6A dependent structural switches to regulate splicing (Liu et al, 2017). These few examples demonstrate the importance of RNA modifications on post-transcriptional gene regulation.

1.4 Regulation of gene expression by degradation pathways

Degradation of RNA is an important stage in gene expression control and different classes of degradation can be characterised. Firstly, Pol II transcription generates a multitude of transcripts which undergo extensive processing. These processing events produce excised introns and spacer fragments that must undergo degradation. Secondly, regulated turnover of mRNA is important for gene expression control. Similarly, RDH degradation is important in cell cycle function; is tightly coupled to DNA replication to ensure proper chromatin formation and enhances recombination rates in response to DNA damage (Mullen and Marzluff, 2008; Hauer et al, 2017). Finally, degradation acts as a quality control mechanism. Due to the complexity of RNA processing mechanisms, errors can often occur that generate aberrant or defective transcripts. Additionally, mRNAs with premature translation termination codons are generated by alternative splicing. The levels of these defective RNAs must be controlled to prevent potential problems such as the saturation of RNA processing machinery and therefore they are rapidly degraded (Houseley and Tollervery, 2009).

There are different classes of RNA-degrading enzymes. Endonucleases cleave the phosphodiester bond between nucleotides, cleaving RNA internally. Exonucleases cleave RNA from the end, with one type hydrolysing RNA from the 5' end and another type hydrolysing from the 3' end (Houseley and Tollervey, 2009). In addition, some nucleases also exhibit kinase activity, such as NDK1 (Yoon et al, 2005). Pol II transcripts commonly obtain a 5' cap which protects RNA from degradation by 5' exonucleases such as XRN2 (Ramanathan et al, 2016). Therefore, RNA decapping is an important process in degradation. Dcp2 is predominately found in the cytoplasm but is able to shuttle into the nucleus and interact with XRN2 and transcription termination factors. This interaction allows Dcp2 to catalyse hydrolysis of the 5' cap, resulting in its removal (Piccirillo et al, 2003). XRN2 is not only involved near transcription termination of genes, but is associated with transcription machinery during initiation (Davidson et al, 2012; Jimeno-Gonzalez et al, 2010). Pol II aborted transcripts generated by promoter-proximal pausing, defectively spliced or capped transcripts are often retained at the TSS and can be degraded by XRN2. XRN2 accounts for the removal of some

transcripts, however other RNA transcripts are degraded by the exosome complex, which is one of the main focusses of this thesis.

1.4.1 Exosome complex

The human nuclear exosome has vital functions in processing, nuclear surveillance and degradation of nearly every class of RNA. The exosome is a multi-subunit complex composed of a 9 subunit barrel-like core lacking catalytic activity (EXO-9). These subunits are arranged as a hexamer (PH-like ring) capped with a trimeric S1/KH ring and they interact with 3' – 5' exonucleases DIS3 (homolog of yeast RRP44) and EXOSC10 (homolog of yeast RRP6) (Allmang et al, 1999; Mitchell, 2014). A central channel runs through the EXO-9 core, with EXOSC10 localised on top of the S1/KH cap and DIS3 at the opposite end (Figure 1.4). The channel is essential to exosome function as it mediates RNA binding to allow access of RNA substrates to DIS3 and EXOSC10 (Makino et al, 2015). During association with the exosome, the exonuclease domain of EXOSC10 is exposed whereas the exonuclease domain of DIS3 faces towards the channel at the exit pore. It is thought that RNA passes through the central channel to facilitate their interaction with DIS3 (Lorentzen et al, 2008).

EXOSC10, located at the entry pore of EXO-9, may regulate this RNA threading by widening the channel and thus allosterically mediating DIS3 activity (Wasmuth et al, 2014). Mutant catalytically-dead EXOSC10 is still able to enhance DIS3 activity in vitro. Additionally, mutations in the exosome complex that obstruct the channel inhibit DIS3 and EXOSC10 activities in yeast (Wasmuth and Lima, 2012), although RNA substrates can also be directed to the nuclease domains independently (Bonneau et al, 2009; Schneider et al, 2012). RNA threaded through the entire channel is degraded by DIS3, whereas RNA that enters the S1/KH ring before being deflected outwards is degraded by EXOSC10 (Zinder et al, 2016).

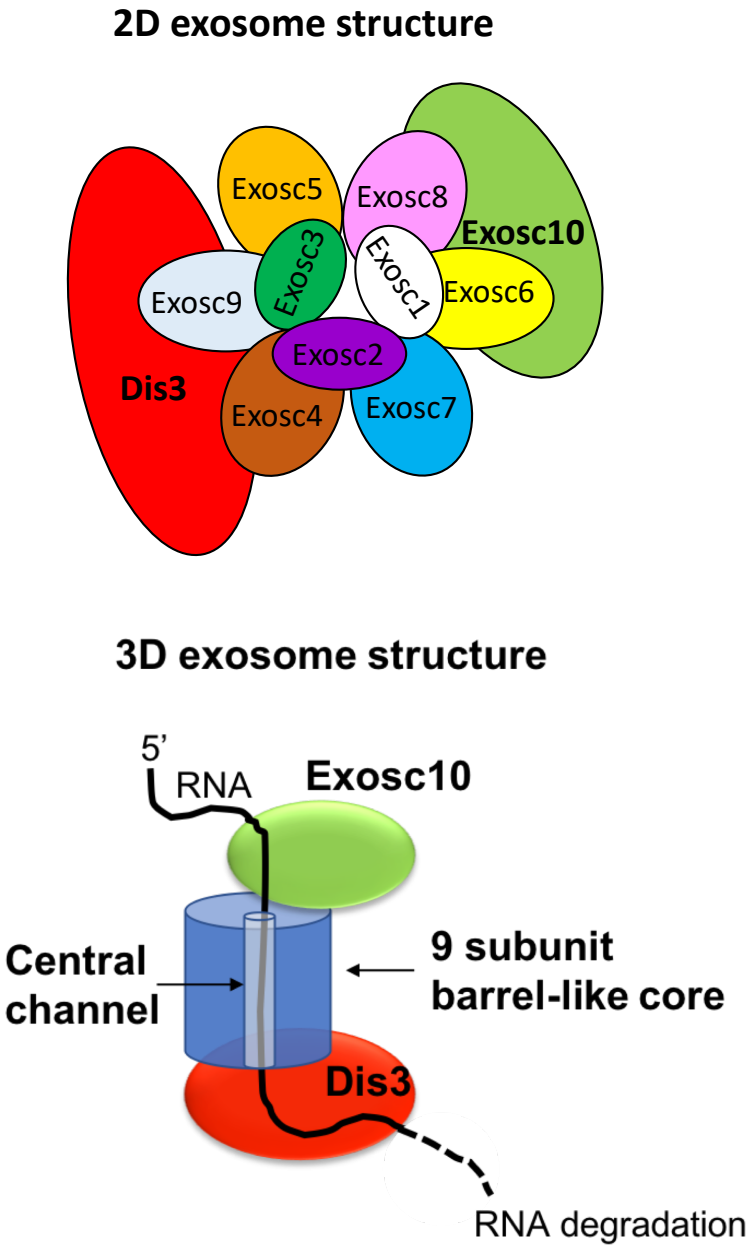

Figure 1.4 Exosome structure

The 2D structure of the exosome shows the core subunits, EXOSC1-9, that make up the barrel-like structure of the exosome, with EXOSC10 and DIS3 attached on either end. The exosome structure can exist in different formations with either EXOSC10, DIS3, both or neither subunits. The 3D structure shows EXOSC10 at the entry pore and DIS3 at the exit pore of the core, with the central channel running through the middle of the core. As shown, it is possible for RNA to thread through the central channel, facilitated by Exosc10, to allow its degradation by DIS3.

These structural exosome findings were conducted in yeast, however recent studies suggest the human nuclear exosome may show some slight differences. Through cyro-EM, Gerlach et al (2018) found the position of human DIS3 (hDIS3) on EXO-9 more closely resembles an open conformation of RNA binding directly to RRP44 in yeast than a closed conformation of RNA accessing the active site of RRP44 through the exosome channel. A long RNA channel path was still observed with RNA travelling through the EXO-9 channel to then bind DIS3, however the RNA path was longer in humans than yeast (Gerlach et al, 2018; Weick et al, 2018).

DIS3 encompasses two domains with different catalytic activities. Firstly, DIS3 includes a N-terminal PIN domain which is responsible for endoribonuclease activity and interacts with EXO-9 subunits RRP41 and RRP45. (Lebreton et al, 2008; Schneider et al, 2009; Schaeffer et al, 2009; Bonneau et al, 2009). Secondly, there is a ribonuclease domain (RNB) containing the active site for exoribonuclease activity (Lorentzen et al, 2008). Although knockdown of DIS3 is essential to cell growth (Mitchell et al, 1997), it was shown that inhibiting DIS3 exoribonuclease activity by mutating the RNB domain (D551N) is not lethal although a slower growth phenotype is observed (Dziembowski et al, 2007). Similarly mutating the PIN domain (D171N), thus preventing endoribonuclease activity, produced no obvious phenotype. However, expression of both these mutations together caused growth inhibition (Schaeffer et al, 2009). This suggests that catalytically inactive DIS3 results in a non-functional exosome however, at least one type of DIS3 nuclease activity is sufficient for cell viability. It is important to note that Schaeffer et al (2009) used RNAi methods to knockdown endogenous DIS3 levels whilst expressing mutant DIS3 constructs. Therefore, DIS3 may not have been fully depleted by RNAi and low levels of functioning DIS3 may be present. This could have a slight rescue effect on the mutant phenotypes and it is possible that mutations through genetic modification would instead be lethal.

DIS3 mutations not only affect cell viability in different ways but also exosome function. The D171N mutant produced no degradation intermediates whereas D551N mutation caused accumulation of degradation and processing intermediates. A combination of both mutations produced a similar phenotype to that of D551N alone. Therefore a viable RNB domain, but not PIN domain, is

essential for exosome function. These differences could be explained by exoribonuclease and endoribonuclease domains of DIS3 acting on separate specific substrates. Alternatively, both catalytic domains may increase the efficiency of each other and act synergistically on substrates (Schaeffer et al, 2009).

Although mutation of the PIN domain did not prevent degradation of exosome substrates, it may affect the speed and efficiency of degradation. The RNB domain aids in hydrolysis of single-stranded RNA in a 3' – 5' direction. Nucleotides are singularly released to produce an end product of only a few nucleotides. Normally DIS3 can unwind secondary structures of RNA provided there are unstructured regions of adequate length at the 3' end (Robinson et al, 2015). The PIN domain could act in releasing exosome substrates where degradation has stalled due to their secondary structure. PIN domain function may enhance exoribonuclease activity of DIS3 or EXOSC10 by providing them with alternative 3' end substrates and aiding exosome degradation functions when progression is blocked (Lebreton et al, 2008).

In yeast, RRP6 is located solely in the nucleus but both RRP44 and RRP6 can be found in the nucleoplasm and nucleolus. In comparison, EXOSC10 and DIS3 are located mainly in the nucleus of human cells, with exclusion of DIS3 from the nucleolus and enrichment of EXOSC10 in the nucleolar compartment (Tomecki et al, 2010). There are 2 other isoforms of DIS3 which are found exclusively in the cytoplasm, DIS3L and DIS3L2. DIS3L can associate with the exosome but it does not exhibit endoribonuclease activity as the two catalytic residues within the PIN domain are absent. Conversely, DIS3L2 lacks a PIN domain due to splicing of exon 2 and is not known to be part of any stable macromolecular assembly (Tomecki et al, 2010; Kumakura et al, 2013; Staals et al, 2010). Overall three potential exosome complexes may exist within the human nucleus; EXO-9 with EXOSC10, nucleoplasmic EXO-9 with DIS3 and nucleoplasmic EXO-9 with EXOSC10 and DIS3 (Lykke-Anderson et al, 2011). The differing subcellular distributions of these exosome complexes may allow each to perform specialised functions within the cellular compartments (Kilchert et al, 2016).

EXOSC10 and DIS3 may have specific substrates. It has been suggested that EXOSC10 is more involved in processing of RNAs than DIS3, specifically in

the processing of small RNAs. Additionally, EXOSC10 is more efficient in degrading substrates with more complex secondary substructures including small nucleolar RNAs and pre-rRNA (Januszyk et al, 2011). On the other hand, previous studies have suggested that DIS3 is the main catalytic subunit of the exosome for degrading nearly all classes of RNAs, including pervasive transcripts (Dziembowski et al, 2007). Szczepinska et al (2015) proposed that DIS3 also degrades enhancer RNAs (eRNAs) and snoRNAs.

In addition to the degradation of aberrant transcripts, the nuclear exosome has an important function in degradation of cryptic transcripts known as CUTs in yeast and PROMPTs in humans. As described earlier, cryptic transcripts are derived from transcription in the opposite direction to a protein-coding gene, at bidirectional promoters. Due to their quick turnover, they are only detectable in the cell upon exosome dysfunction (Preker et al, 2008). DIS3 is suggested as the predominant, if not only, degradation pathway for PROMPTs in humans. Szczepinska et al (2015) used PAR-CLIP techniques in HEK293 cells expressing a catalytically-dead DIS3 mutant and found that upon DIS3 dysfunction there was robust accumulation of PROMPTs. PROMPT accumulation was also observed when other exosome components were downregulated, including EXOSC10 and EXOSC3 (hRRP40) (Preker et al, 2008; Flynn et al, 2011).

1.4.2 Exosome co-factors

Nuclear exosome function is modulated by various cofactors and interacting partners. Of high significance is the yeast TRAMP complex which aids the exosome in substrate specificity (Schmidt and Butler, 2013). The TRAMP complex contributes to exosome RNA processing through Trf4p subunit addition of a short poly (A) tail (3 – 50 nts) to transcripts (Wyers et al, 2005). In addition to Trf4p, the TRAMP complex also contains the essential helicase MTR4 (LaCava et al, 2005). A TRAMP-like complex has been identified in humans which contains a MTR4 homolog and close orthologues such as Trf4p and PAPD5. However, unlike yeast, the activity of the TRAMP complex in humans is predominately restricted to the nucleolus due to TRAMP subunit nucleolar localisation (Lubas et al, 2011). Consistent with this localisation is the finding that PAPD5 polyadenylates snoRNA and pre-rRNA transcripts (Ogami et al, 2018).

MTR4 helicase activity is enhanced by Mpp6 binding to the EXO-9 subunit, RRP40. A secondary structure forms at the 3' end of RNA substrates, which is unwound by MTR4. This produces single-stranded RNA that is more capable of threading through the channel in a 3' – 5' direction (Falk et al, 2017). TRAMP polyadenylation activity may help prepare RNA as a substrate for degradation, by generating poly(A) tails long enough for binding by MTR4 (Zinder and Lima, 2017). In humans it was found that MTR4 binds to the exosome through contact with Mpp6 and exosome subunit EXOSC2 (Weick et al, 2018).

Human MTR4 is also part of the nuclear exosome targeting (NEXT) and poly(A) tail exosome targeting complexes (PAXT) (Lubas et al, 2011; Meola et al, 2016). The NEXT complex has been shown to promote degradation of PROMPTs and 3' extended RNAs (Lubas et al, 2011; Tseng et al, 2015; Hrossova et al, 2015), whereas PAXT promotes degradation of transcripts with larger poly(A) tails (Meola et al, 2016). Lubas et al (2011) found that depletion of NEXT components, Rbm7 and ZCCHC8, leads to accumulation of PROMPTs showing the importance of NEXT in exosome degradation of certain transcripts.

The exosome has also been observed to be tethered to nascent capped transcripts, through NEXT and PAXT interaction with the cap-binding complex containing ARS2 (CBCA) (Andersen et al, 2013; Meola et al, 2016). ARS2 binds to the CBC and acts as a scaffold protein, recruiting various protein complexes involved in 3' end processing, maturation, degradation and export to the 5' CBC (Gruber et al, 2009; Hallais et al, 2013; Andersen et al, 2013). Premature transcription termination produces RNA 3' ends within the first introns of protein-coding genes. These pervasive transcripts are exosome substrates and their turnover is supported by ARS2 function (Iasillo et al, 2017). In addition, Iasillo et al, (2017) found through ARS2 depletion in HeLa cells and RNA-Seq that ARS2 plays a role in transcription termination downstream of short snRNA, RDH, PROMPTs and eRNA.

1.4.3 Cytoplasmic mRNA degradation

Messenger RNAs in the cytoplasm are normally protected from endonucleases by their 5' cap and 3' poly(A) tail. The majority of cytoplasmic mRNAs are degraded in a deadenylation-depenent manner. Deadenylation is often the rate-limiting step of cytoplasmic mRNA decay and is conducted by 2 main deadenylases, CCR4-NOT and PAN2-PAN3 (Siwaszek et al, 2014). PAN2-PAN3 complex firstly shortens the poly(A) tail to approximately 110 nt and then the CCR4-NOT complex deadenylates the mRNA to a poly(A) tail length of approximately 10 nts (Yamashita et al, 2005; Chen et al, 2011). After deadenylation the mRNA may undergo decapping by DCP2, which can also decap mRNA in the nucleus as previously mentioned (Piccirillo et al, 2003). Decapped mRNA is then a substrate for 5' – 3' exonuclease degradation by XRN1 (Braun et al, 2012).

Alternatively, cytoplasmic mRNA can be degraded in a 3' – 5' direction by the cytoplasmic exosome. The cytoplasmic exosome is similar in structure to the nuclear exosome, except DIS3 is not present. Instead a paralogue DIS3L, which does not contain endonuclease activity due to mutations in the PIN domain, is responsible for the catalytic activity of the cytoplasmic exosome. (Tomecki et al, 2010). After degradation by the exosome a scavenging decapping enzyme, DcpS that has a specific for shorter RNA species, hydrolyses the residual 5' cap (Chen et al, 2005). Interestingly, DIS3L2 is another paralogue of DIS3 and is found specifically in the cytoplasm, doesn't interact with the exosome and lacks a PIN domain. It is found to preferentially degrade 3' uridylated RNAs in an exosome-independent manner (Malecki et al, 2013; Lubas et al, 2013). Depletion of DIS3L2 causes an accumulation of a multitude of mRNAs in the cytoplasm, suggesting DIS3L2 may be responsible for a third cytoplasmic degradation pathway (Malecki et al, 2013).

1.4.4 Nonsense-mediated decay

As discussed, formation of aberrant mRNAs can occur at multiple stages and can be hazardous to cells through the generation of potentially toxic proteins or by sequestering processing machinery. There are different posttranscriptional quality-control mechanisms to prevent this occurrence and the best characterised is the Nonsense-mediated decay (NMD) pathway. NMD removes aberrant mRNAs which contain a premature stop codon possibly due to mutations, transcriptional errors or splicing errors (Popp and Maquat, 2013). Translational termination, which involves eukaryotic release factor 1 and 3 (eRF1 and eRF3), is the first signal to trigger NMD.

As a consequence of pre-mRNA splicing, an exon-junction complex (EJC) is found on the mRNA approximately 20 – 24 nts upstream of the exon-exon junction. (Le Hir et al, 2000). During the first round of translation, EJC's are removed from the mRNA. After this, NMD occurs if any EJCs remain bound to the mRNA, which would occur if the ribosome was released before reaching the EJC i.e. if eRF1 and eRF3 assemble at a premature stop codon located ≥ 50 – 55 nts upstream of a EJC then NMD is triggered (Popp and Maquat, 2013). NMD is mediated by up-frameshift proteins UPF1, UPF2, UPF3A and UPF3B and aided by suppressors with morphological effects on genitalia, SMG1, SMG5-9. UPF1 is an ATP-dependent RNA helicase that with SMG1 kinase binds to eRF1 and eRF3 to form the SURF complex near a premature stop codon (Kashima et al, 2006; Chakrabarti et al, 2011).

UPF2 and UPF3 are found on the EJC and their subsequent contact with UPF1 results in phosphorylation of UPF1 by SMG1 and release of eRF1 and eRF3 (Kashima et al, 2006). Phosphorylated UPF1 becomes an active helicase that resolves mRNA secondary structure and removes bound proteins, as well as recruiting SMG5-7 and other general mRNA degradation factors including XRN1 (Fiorini et al, 2015; Okada-Katsuhata et al, 2012). SMG6 is capable of endonucleotytically cleaving the aberrant mRNA to generate RNA fragments that can be degraded by XRN1 or the cytoplasmic exosome (Huntzinger et al, 2008; Eberle et al, 2009).

1.5 Gene engineering using CRISPR/Cas9

Precise, targeted changes to the genome are important for many applications across science, including systemic interrogation of genetic elements and development of disease models. Previous methods have used zinc-finger nucleases or transcription-activator like effector nucleases, but a faster, cheaper, highly specific and more efficient gene editing method was developed, Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)/Cas9 (Gaj et al, 2013; Gupta and Musunuru, 2014; Cong et al, 2013).

CRISPR/Cas9 was developed from a naturally occurring gene editing system in bacteria, with the most commonly used CRISPR/Cas9 technology in human cells being adapted from *Streptococcus pyogenes*. CRISPR/Cas9 provides bacteria with adaptive immunity by acting as an immune memory of viral infections and preventing re-infection (Barrangou et al, 2007). The CRISPR loci consists of repetitive elements, 30 – 40 bp, which flank short sequences of DNA with viral and plasmid origins known as protospacers. In bacterial adaptation, new protospacers are introduced during infection and their DNA is homologous to bacteriophages or plasmids, to provide specific immunity (Mali et al, 2013a; Mali et al, 2013b).

Genome engineering by CRISPR/Cas9 requires a conserved 3' protospacer adjacent motif (PAM), that is associated downstream of every protospacer. Different CRISPR systems have various PAM sequences, for example the PAM sequence for Cas9 from *Streptococcus pyogenes* is 5'-NGG whereas the Cas9 ortholog in *Neisseria meningiditis* is 5'-NNNNGATT (Jinek et al, 2012; Zhang et al, 2013). For specific gene editing in mammalian cells, a human codon-optimised Cas9 must be expressed alongside a guideRNA (gRNA), consisting of DNA complementary to the genome target that associates with Cas9 and the genome. The Cas9 nuclease can therefore be targeted toward any part of the genome by altering the gRNA, as long as there is a PAM sequence located 3' of the target DNA.

The gRNA directs CRISPR/Cas9 to the DNA target, where Cas9 can cleave the DNA. Upon cleavage, a double stranded break (DSB) is formed and using the cell's own DNA repair machinery, is either repaired by non-homologous end joining (NHEJ) or high-fidelity homology-directed repair (HDR). With NHEJ,

the broken DNA strands are re-ligated creating insertion/deletion (indel) mutations, making NHEJ repair an effective way to study genetic variation by introducing random deleterious mutations (Bibikova et al, 2002). On the other hand, HDR occurs less frequently in vivo but creates more accurate repairs by using a repair template to ligate the DNA. Through the design of custom repair templates introduced to the cell, HDR can introduce large and precise DNA modifications (Chen et al, 2011).

1.5.1 Altering gene expression post-transcriptionally

Regulating gene expression in eukaryotes can be achieved by altering transcription levels and mRNA abundance, except these methods can be limited by their rate of downregulation, especially for proteins with a long half-life. Therefore, methods have been developed that instead modify protein levels more directly. A commonly used technique for post-transcriptional modifications is RNA interference (RNAi), that utilises complementary small RNA molecules to specifically target and degrade mRNA transcripts via the RNA inducing silencing complex (RISC) and thus prevent their translation (Elbashir et al, 2001). However, RNAi methods have been criticised for producing off-target effects, requiring long periods of time for gene downregulation and causing incomplete downregulation. Therefore, it has been important to find other methods that may combat these limitations and alter gene expression levels post-translationally. For this, various methods have been proposed that utilise the CRISPR/Cas9 system (Zhang et al, 2015b; Natsume et al, 2016; Lambrus et al, 2018; Chung et al, 2015).

1.5.2 The auxin system in plants

Plants contain a hormone, indole-3-acetic acid (IAA or auxin), which is detrimental to regulation of plant cell division, expansion and differentiation (Teale et al, 2006). Auxin enacts its role by regulating gene expression and to do this a ubiquitin-dependent proteolytic system is involved. Specifically, the F-box protein transport inhibitor response 1 (TIR1), which contains an auxin binding site, forms a functional E3 ubiquitin ligase complex with Skp1 and Cullin 1 (SCF^{TIR1}). The SCF^{TIR1} recruits an E2 ubiquitin conjugating enzyme that catalyses ubiquitination of proteins containing an auxin inducible degron (AID) (Gray et al, 1999). Auxin inducible genes are bound by auxin response factors (ARF), whose interaction with auxin transcriptional repressors (ATR) prevents gene expression (Tan et al, 2007). Auxin brings together the ATRs with SCF^{TIR1}, causing polyubiquitination of the ATRs and leading to their degradation. This in turn releases the inhibition of ARFs, causing activation of gene expression at auxin inducible genes (Gray et al, 2001) (Figure 1.5).

1.5.3 Implementation of the Auxin inducible degron system (AID) in eukaryotes

The plant auxin-regulated protein degradation system has since been exploited in various studies to allow ubiquitination of specific substrates and their subsequent degradation. Although non-plant eukaryotes express the ubiquitin ligase SCF, in which the F-box protein determines substrate specificity, they lack orthologs of TIR1 and auxin inducible degrons (Holland et al, 2012). However, due to the highly conserved Skp1, Nishimura et al (2009) was able to express the *Arabidopsis thaliana* TIR1 gene in budding yeast and find evidence for formation of SCF^{TIR1}. They also fused the AID, IAA17, to the N and C terminus of GFP and expressed these fusion proteins in cells expressing SCF^{TIR1}. Both AID-GFP-NLS and GFP-AID-NLS were depleted in a TIR1 and auxin-dependent manner. The AID system has also been implemented in mammalian cells and reversible protein degradation within minutes of auxin removal was observed (Holland et al, 2012; Nishimura et al, 2009) (Figure 1.6).

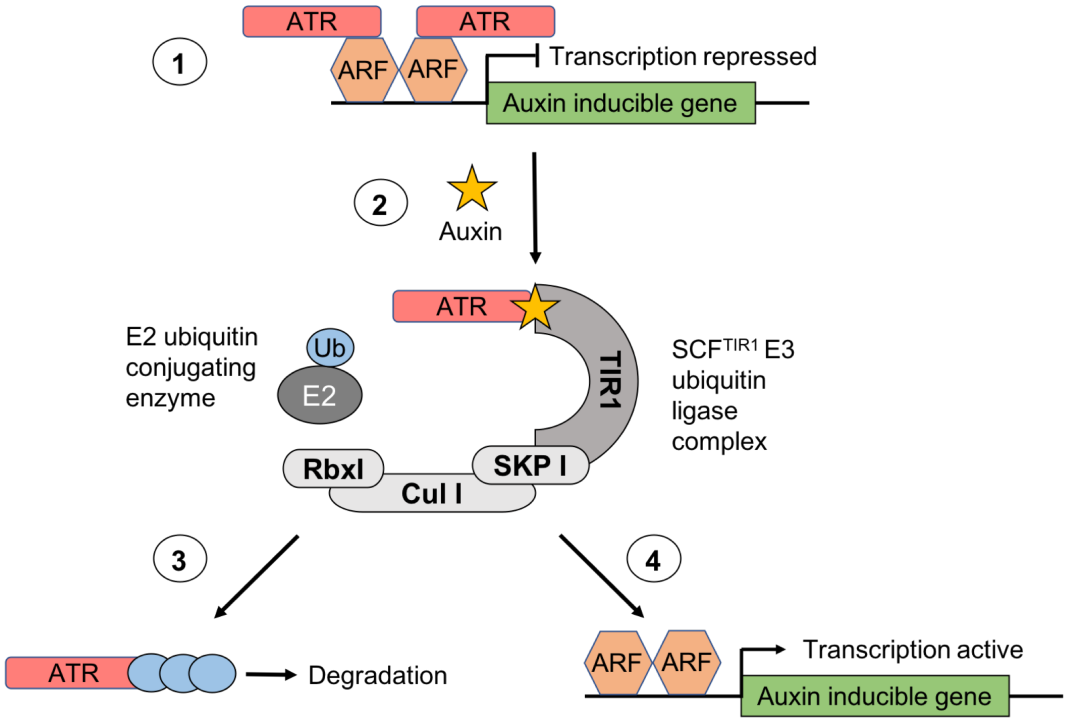

Figure 1.5: Auxin system in plants

1) Auxin inducible genes are bound by auxin response factors (ARF), which interact with auxin transcriptional repressors (ATR). 2) Auxin brings together the ATR and a E3 ubiquitin ligase complex, SCF^{TIR1}. SCF^{TIR1} recruits a E2 ubiquitin enzyme that causes ubiquitination of the ATR. 3) Polyubiquitination of ATR leads to its subsequent degradation. 4) ARFs are no longer inhibited and transcription of the auxin inducible gene can occur.

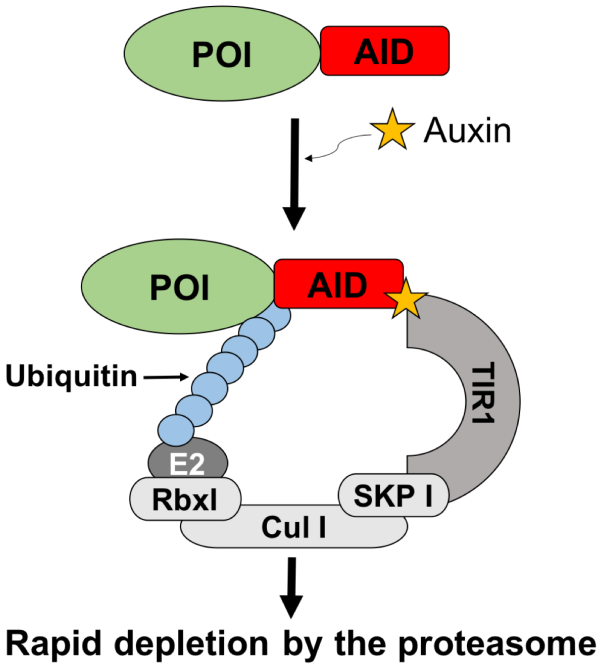

Figure 1.6: Auxin system in human cell lines

A protein of interest (POI) is tagged to an auxin inducible degron (AID) by CRISPR/Cas9 technology, in cells expressing plant TIR1. Upon addition of auxin, the SCF^{TIR1} complex binds to the POI and recruits a E2 ubiquitin enzyme. The POI is ubiquitinated and subsequently degraded by the proteasome.

1.5.4 AID system and CRISPR/Cas9

The AID system causes rapid depletion of a protein; however, its' implementation can be time consuming and difficult due to the necessity of tagging the endogenous target protein and co-expressing TIR1 within the desired eukaryote system. This caveat has been improved by the use of CRISPR/Cas9 technology. Zhang et al (2015b) used CRISPR/Cas9 genome editing to introduce the AID tag to a protein of interest (POI) in *C.elegans* and found a rapid degradation (20 minutes) of the POI in the presence of auxin. The authors also compared the AID system to RNAi depletion from a previous study (Kostrouchova et al, 2001) and found the AID system produced a highly pronounced phenotype (2 % progeny arrested in development compared to 100 %, respectively). This suggests that the AID system is able to produce a more robust phenotype than RNAi.

Natsume et al (2016) used a similar method in mammalian cells. They tagged endogenous genes with the AID-tag using donor vectors containing synthetic short homology arms as a repair cassette for HDR. This was done in human colorectal cancer (HCT116) cells due to their well-established diploid karyotype. Other studies have also used CRISPR/Cas9 to achieve biallelic insertion of the AID tag into human cells, a method which can be adapted to allow insertion of other tags (Lambrus et al, 2018).

1.5.5 Small Molecule Assisted Shutoff (SMASh)

Small Molecule Assisted Shutoff (SMASh) is another technique that can be used to modulate protein activity at a post-transcriptional level using chemical regulation (Chung et al, 2015). In contrast to the AID system, SMASh involves only a single component and is selective for new proteins. SMASh suppression of a protein works firstly by using CRISPR/CAS9 technology to fuse a SMASh tag to the target of interest via a hepatitis C virus (HCV) nonstructural protein 3 (NS3) protease recognition site. The SMASh tag consists of a NS3 protease and destabilising degron. After protein folding, the internal protease activity causes cleavage at the HCV NS3 recognition site, resulting in an unmodified protein product. The cleaved SMASh tag is then degraded due its' internal degron activity. Upon addition of a protease inhibitor, asunaprevir, the POI remains

tagged and is now targeted for proteasomal and / or autophagosomal degradation alongside the SMASh tag (Figure 1.7). Therefore asunaprevir causes the rapid degradation of newly synthesised POI.

Regulation by asunaprevir allows for stringent control and quick recovery of protein production. Chung et al (2015) showed the SMASh tag can be attached to the POI at either the N or C terminus and that due to the absence of protein structural modifications after cleavage, it is expected the POI will have normal functionality. However, as it is the protein's processing into a functional protein that is inhibited by asunaprevir, the SMASh system will work best when an accumulation of protein is required upon removal of asunaprevir or in cases where the POI is short-lived. This will prevent protein produced prior to asunaprevir addition from having an effect (Bondeson and Crews, 2017). This system has been previously used to regulate the replication of Influenza A Virus (IAV) in vitro and in vivo, without directly targeting viral proteins (Fay et al, 2019). In addition, Yan et al (2015) used the SMASh tag to alter expression of a reporter in a dual-reporter screen, thus increasing its statistical power and demonstrating endogenous yeast gene modification by the SMASh system.

Overall both the AID system and SMASh-tag are bioorthogonal and produce inducible and reversible protein degradation, making them efficacious methods for targeted protein degradation.

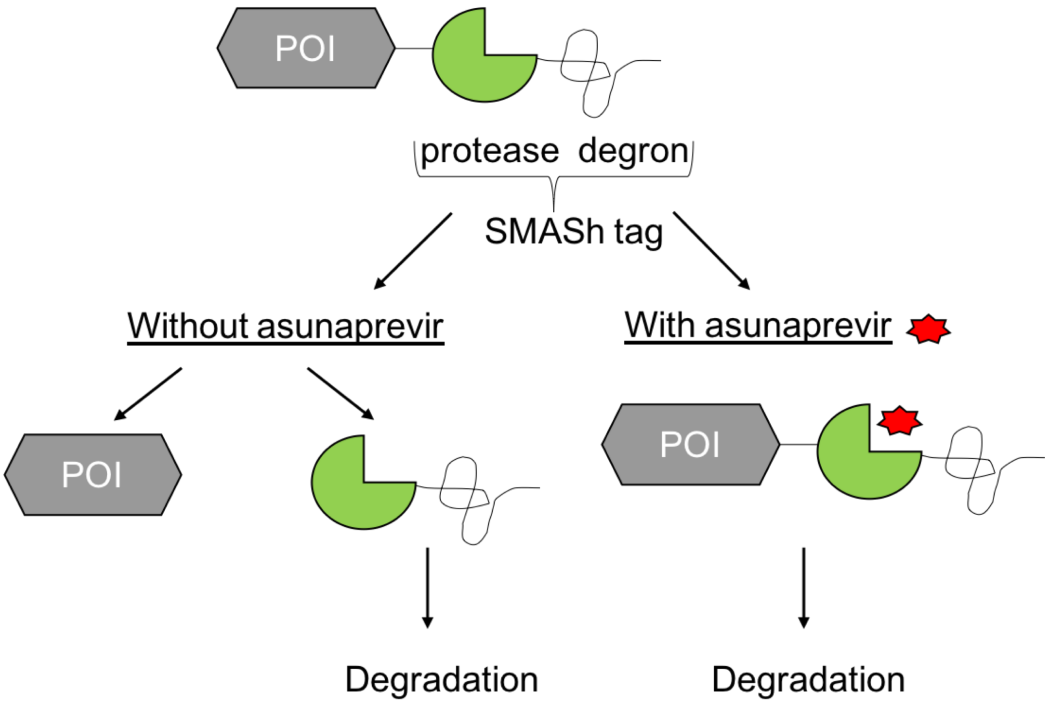

Figure 1.7: Small Molecule Assisted Shut-Off (SMASh)

The SMASh tag consists of a protease and degron linked to the protein of interest (POI) by a HCV NS3 protease recognition site. The SMASh tag internal protease activity cleaves the tag at the recognition site, when cells are untreated. The POI becomes untagged and is able to conduct its normal function whereas the SMASh tag is degraded due to internal degron activity. Upon addition of a protease inhibitor, asunaprevir, the protein remains tagged and is targeted for degradation.

1.6 Project Aims

This introduction has discussed the role of endonucleases in cleavage and polyadenylation events, 3' end processing and degradation of RNAs, with some endonucleases acting in more than one pathway. Although previous studies have attempted to elucidate the differing functions of the endonucleases CPSF73, INTS11 and DIS3, a lot is still unknown. Being able to determine their specific substrates would be beneficial in revealing functional roles.

Aforementioned work in these areas have mainly been conducted in yeast, due to their smaller genomes and therefore are more easily genetically modified. Studies conducted in human tissues have often used RNAi techniques to regulate expression of proteins of interest. RNAi methods utilise small interfering RNAs (siRNAs) to specifically target a mRNA. siRNAs associate with the RNA-inducing silencing complex (RISC) which unwinds the siRNA to produce a single-strand. The single-stranded RNA binds to the complementary mRNA target, allowing the RISC complex to cleave mRNA and in effect, silence gene expression. Although RNAi methods made manipulation of gene activity more accessible, with increased specificity and relative ease-of-use to previous methods, there are caveats.

RNAi mediated-knockdown of a gene is often time-consuming, with an adequate level of protein reduction taking multiple days depending on the half-life of the protein of interest. Over these long periods, RNAi has increased potential for off-target effects and complicates the interpretation of phenotypic effects. Indeed, siRNAs have been shown to have reduced specificity causing silencing of non-target genes (Jackson et al, 2003). In addition, RNAi does not always produce complete protein depletion. It was recently shown that because of these limitations, trace levels of protein remaining after RNAi may cause false negative results or a reduced phenotype (Eaton et al, 2018). Therefore gene editing techniques, such as CRISPR/Cas9 and the AID system, may be beneficial to produce an increased repertoire of protein functions. Furthermore it allows conditional depletion in a quicker manner than RNAi, which will be important when studying essential proteins.

The aims of this study were to further reveal the roles of three endonuclease proteins, DIS3, INTS11 and CPSF73, in human cells (Figure 1.8). Additionally, to provide insight into the substrates and mechanisms of the three endonucleases by investigating changes in transcription upon protein depletion. It was hypothesised that: 1) Using CRISPR/Cas9 technology to genetically modify gene targets with tags would allow conditional protein depletion; 2) Depletion of the exosome subunit DIS3 would cause an accumulation of RNA transcripts, due to loss of their degradation; 3) Depletion of CPSF73 would cause misprocessing and possible extension at protein-coding genes and RDHs, as CPSF73 is believed to play a major role in the 3' end processing of these genes; 4) Depletion of INTS11 would result in dysfunction of the Integrator complex, causing misprocessing at snRNA genes. The research questions that we asked throughout this study included: Are our protein-depletion cell lines capable of quick, specific and near complete depletion of our protein target? Can these cell lines, with the use of RNA-Seq, elucidate specific substrates of these endonucleases? If depletion of endonucleases responsible for 3' end processing of specific genes results in extension, where does this extension terminate? And do the findings support or refute the model of transcription termination? My objectives to address these aims and research questions were to use CRISPR/Cas9 technology to produce conditional-depletion cell lines of three endonucleases: CPSF73, INTS11 and DIS3. Upon generation of these cell lines, to utilise transcriptome-wide high-throughput RNA-Seq analysis of nascent RNA to determine specific substrates and effects of protein depletion. Finally, to validate RNA-Seq findings by other methods including RT-qPCR.

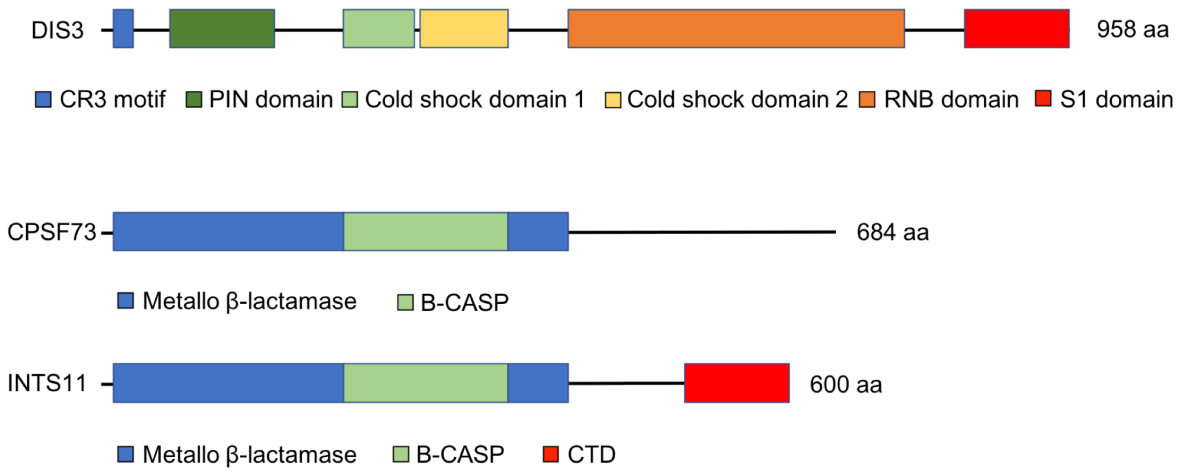

Figure 1.8 Domain organisation of DIS3, CPSF73 and INTS11

Domain organisations of human DIS3, CPSF73 and INTS11. The PIN domain of DIS3 contains endoribonuclease activity and the RNB domain contains the active site for exoribonuclease activity. As INTS11 and CPSF73 are homologs, their domain organisation is similar. CPSF73 also contains a CTD, but its sequence is highly divergent from INTS11 and the exact boundary is unknown (Robinson et al, 2015; Wu et al, 2017).

2. Materials and Methods

2.1 Buffer compositions

Before use, buffers were sterilised either by autoclave or filter-syringe, Millex-GP 0.22 μm filter (Sigma).

2.1.1 DNA/RNA Buffers

- **DNA Loading Buffer**: Gel loading dye purple (6 x) (B7024S, NEB)
- **1 x TBE buffer**: 10 mM Tris-HCl (pH 8) (ThermoFisher Scientific), 1 mM EDTA (pH 8) (ThermoFisher Scientific)
- **Total RNA Extraction**: TRI Reagent Solution (Sigma)

2.1.2 SDS-Polyacrylamide gel electrophoresis (PAGE) and Western blot buffers

- **RIPA Buffer**: 50 mM Tris-HCl (pH 7.4) (ThermoFisher Scientific), 150 mM NaCl (ThermoFisher Scientific), 0.5 % Sodium Deoxycholate (Sigma), 1 % NP40 (ThermoFisher Scientific), 0.1 % Sodium Dodecyl Sulphate (SDS) (ThermoFisher Scientific)
- **4 x SDS-PAGE sample buffer**: 40 % Glycerol (ThermoFisher Scientific), 8 % SDS (ThermoFisher Scientific), 0.006 % Bromophenol Blue (Sigma), 0.25 M Tris-HCl (pH 6.8) (ThermoFisher Scientific). Before use, 0.5 ml was separated and warmed to 50 °C, then 50 μl β-mercaptoethanol added.
- **4 x SDS-PAGE Stacking gel buffer**: 0.5 M Tris-HCl pH 6.8 (ThermoFisher Scientific), 0.4% SDS (ThermoFisher Scientific)
- **4 x SDS-PAGE Resolving gel buffer**: 1.5 M Tris-HCl pH 8.8 (ThermoFisher Scientific), 0.4% SDS (ThermoFisher Scientific)
- **SDS-PAGE running buffer**: 192 mM Glycine (ThermoFisher Scientific), 25 mM Tris, 0.1 % SDS (ThermoFisher Scientific)
- **Transfer buffer**: 25 mM Tris, 192 mM glycine, 20% methanol (All ThermoFisher Scientific)

- **5% Blocking Solution**: 2.5 g milk in 50 ml PBST
- **Enhanced Chemi-Luminescence (ECL) Solution 1**: 100 mM Tris-HCl (pH 8.5), 2.5 mM Luminol (Sigma), 400 μM p-Coumaric Acid (Sigma)
- **ECL Solution 2**: 100 mM Tris-HCl (pH 8.5) 5.3 mM Hydrogen Peroxide (Sigma)

2.1.3 Miscellaneous buffers

- **1 x PBS**: 137 mM NaCl (ThermoFisher Scientific), 10 mM Na_2HPO_4 (Sigma), 2.7 mM KCl (ThermoFisher Scientific), 1.8 mM KH_2PO_4 (pH 7.4 with HCl) (Sigma)
- **1 x PBST**: Same as 1 x PBS, except with addition of 0.05% Tween 20 (Sigma)
- **Trypsin PBS-EDTA**: 500 ml 1 x PBS, 1 mM EDTA (ThermoFisher Scientific), 0.25% Trypsin (Sigma)
- **2 x Oligo annealing buffer**: 100 mM NaCl (ThermoFisher Scientific), 20 mM Tris-HCl (pH 7.5) (ThermoFisher Scientific), 1mM EDTA (pH 8) (ThermoFisher Scientific)
- **qPCR Master Mix**: Agilent Brilliant III Ultra-Fast SYBR Green qPCR Master Mix (Agilent)

2.1.4 RNA-Seq buffers and kits

- **HLBN**: 10 mM Tris-HCl (pH 7.5) (ThermoFisher Scientific), 10 mM NaCl (ThermoFisher Scientific), 2.5 mM $MgCl_2$ (Sigma), 0.5 % NP40 (ThermoFisher Scientific)
- **HLBNS**: Same as HLBN with addition of 10 % sucrose.
- **Ribosomal RNA Depletion**: Illumina Ribo-Zero Gold rRNA Removal Kit
- **RNA-Seq Library Generation**: Illumina TruSeq Stranded Total RNA Library Prep Kit
- **RNA Purification**: Beckman Coulter Agencourt RNAClean XP Beads
- **DNA Purification**: Beckman Coulter Agencourt AMPure XP Beads

- **QC Analysis of RNA and DNA**: Agilent ScreenTape RNA; High Sensitivity RNA; D1000 Assay for TapeStation

2.1.5 ChIP buffers

- **RIPA ChIP**: 1 % NP40 (ThermoFisher Scientific), 150 mM NaCl (ThermoFisher Scientific), 0.5 % Sodium Deoxycholate (DOC) (Sigma), 0.1 % Sodium Dodecyl Sulfate (SDS) (Sigma), 50 mM Tris (pH 8) (ThermoFisher Scientific), 5 mM EDTA (ph 8) (ThermoFisher Scientific)
- **ChIP wash**: 500 mM NaCl (ThermoFisher Scientific), 1 % NP40 (ThermoFisher Scientific), 1 % DOC (Sigma), 100 mM Tris (pH 8.5) (ThermoFisher Scientific)
- **Elution buffer**: 1 % SDS (Sigma), 0.1 M $NaHCO_3$ (Sigma)

2.1.6 Molecular biology kits

- **Plasmid extraction from E.coli**: Qiagen QIAprep Spin Miniprep Kit

2.2 Antibodies

Western Blot Analysis detected levels of proteins using the following antibodies:

Table 2.1 Antibodies used for Western Blot

Protein Detected	Antibody name	Code	Manufacturer
DIS3	Rabbit Anti-DIS3	A303-764A	Bethyl
DIS3	Rabbit Anti-DIS3	A303-765A	Bethyl
AID	Mouse Anti-AID-tag	M214-3	MBL
INTS11	Rabbit anti-CPSF3L	Abx005038	Abbexa
CPSF73	Mouse Anti-CPSF73	A301-090A	Bethyl
RNA Pol II	Mouse anti-RNA Polymerase II CTD	MABI0601	MBL
Alpha tubulin	Mouse anti-alpha tubulin	Ab7291	Abcam
Anti-rabbit secondary	Anti-rabbit IgG, HRP-linked	7074	Cell Signaling Technology
Anti-mouse secondary	Rabbit Anti-Mouse IgG (HRP)	Ab97046	Abcam

2.3 Vectors

All vectors were supplied by Addgene and can be seen in Table 2.2.

Table 2.2 Vectors

Plasmid name	Description
pUC19	Cloning vector with empty backbone
pX330-U6-Chimeric_BB-CBh-hSpCas9	Cloning vector for gRNA with U6 driven expression, containing human codon optimised Cas9
pBABE osTIR1	Human codon optimised TIR1
pMK243 (Tet-osTIR1-Puro)	Plasmid expressing OsTIR1 under control of the Tet promoter
pCMV(CAT) T7SB100	SB-transposase
pSBbi-Blast	Empty SB-transposon with Blasticidin resistance gene

2.4 Bacterial strains

Molecular cloning / genetic recombination was conducted using high efficiency NEB 5 - alpha competent *Escherichia coli*, with the following conditions.

2.4.1 Bacterial growth media

Bacterial growth media was autoclave sterilised and stored at room temperature.

- **Luria Bertani (LB) Broth**: 5 % Yeast Extract (Sigma), 10 % tryptone (Sigma) and 10 % NaCl (ThermoFisher Scientific)
- **LB agar**: Same as LB Broth with addition of 2 % Agar (Sigma)

2.4.2 Antibiotic selection in bacteria

Bacteria were grown in the presence of antibiotics, to select for positively transformed bacterial clones. The final concentrations of selective antibiotics used are as follows:

- **Ampicillin**: 100 μg / ml
- **Kanamycin**: 50 μg / ml

2.5 Molecular Cloning

2.5.1 Polymerase chain reaction (PCR)

PCR was used to amplify specific sequences of DNA, using Q5 High-Fidelity DNA Polymerase (NEB). Typically, a 50 μl reaction consisted of 20 ng template DNA, 5 μl of 5 x Q5 reaction buffer, 0.5 μl of 10 mM dNTPs (NEB), 1.25 μl of 10 μM each primer, 0.25 μl Q5 High-Fidelity DNA Polymerase, 5 μl of Q5 High GC Enhancer (optional), made up to 25 μl total volume with nuclease-free water. The PCR reaction was then set up as follows:

Table 2.3 PCR Protocol

Step	Temperature	Time	Number of cycles
Initial denaturation	98 °C	30 seconds	1
Denaturation Annealing Extension	98 °C 50 – 72 °C * 72 °C	10 seconds 30 seconds 30 seconds / Kb	25 - 30
Final extension	72 °C	2 minutes	1
Hold	4 °C	-	-

*Annealing temperature depended on optimal temperature for primers used

If original template DNA was from bacteria, the PCR product underwent a 1 hour incubation at 37 °C with 0.5 μl DPN1 (NEB), to remove bacterial template DNA. DNA was then purified by a DNA phenol-chloroform extraction and ethanol precipitation (see section 2.5.4).

For colony screening transformed competent cells, a PCR reaction was set up using Taq polymerase (NEB). In a 25 μl volume this contained: variable template DNA (< 500ng), 200 μM dNTPS, 0.2 μM each of forward and reverse primer, 1 x Standard Taq Reaction Buffer and 1.25 Units Taq DNA Polymerase. The same PCR protocol was used as above, except with 30 – 32 cycles, an annealing temperature between 50 – 65 °C and an extension temperature of 68 °C for 1 minute / Kb.

2.5.2 Agarose gel Electrophoresis

To perform agarose gel electrophoresis, agarose gels were prepared using 1 x TBE buffer containing 1 - 2 % (w/v) agarose that was heated until dissolved. 5 % of Midori Green Advanced DNA Stain (Geneflow) was added and the solution was left to cool and set in a Owl™ EasyCast™ Gasketed UVT gel tray (ThermoFisher Scientific). Gels were placed into an electrophoresis tank filled with 0.5 x TBE buffer and DNA samples containing 10 % DNA loading buffer were loaded into the wells alongside an appropriate DNA ladder. To separate DNA or RNA bands, gels were run under 120 V or 180 V respectively. Afterwards, visualisation by UV light was conducted using a Gel Doc XR + System (Bio-Rad).

If gel separated DNA was to be used in downstream applications, a scalpel was carefully used to extract the required DNA from the gel. UV light was used to visualise the bands of DNA and extraction conducted swiftly to minimise UV exposure. DNA was then filtered for 20 seconds at 10,000 rpm into a 0.5 ml Eppendorf. Agarose gel electrophoresis was conducted on 3 μl of eluted DNA for validation.

2.5.3 Restriction Digest

Commercial enzymes were used to digest DNA according to manufacturer's protocol, unless otherwise stated.

2.5.4 Phenol-Chloroform extraction and Ethanol precipitation

A 1:1 ratio of either DNA (pH 8) or RNA (pH 4.3) phenol-chloroform solution (Sigma) was added to DNA or RNA respectively, with a minimum total volume of 200 μl. Samples were vortexed and centrifuged for 5 minutes at 13,000 rpm. The upper-phase solution was transferred to an Eppendorf containing 2.5 times the volume of 100% Ethanol and a 10% volume of 3 M Sodium Acetate (pH 5.4). Samples were vortexed again and centrifuged for 15 minutes at 13,000 rpm. The supernatant was completely removed and samples allowed to air-dry for 5 minutes at room temperature. Cell pellets were resuspended in dH_2O.

2.5.5 Ligation with T4 DNA Ligase

Ligation of linearised or restriction digested DNA occurred using T4 DNA Ligase (NEB). DNA concentrations were determined by a NanoDrop 2000 Spectrophotometer (ThermoFisher) and 100 ng of DNA used in a 20 μl reaction containing T4 DNA Ligase Buffer and T4 DNA Ligase as stated in the manufacturer's protocol. After incubation at 16 °C for 2 hours to overnight, 4 μl of reaction was transformed into competent bacterial cells.

2.5.6 Gibson Assembly

Gibson Assembly (NEB) was used to anneal DNA where vectors had been amplified as multiple fragments. To create blunt ended vectors, plasmid cassettes were amplified using divergent PCR. Cut vectors were treated with 1U Calf Intestinal Alkaline Phosphatase (CIP) (NEB) and 2 μl CutSmart Buffer for 30 minutes at 37 °C to prevent re-ligation. DNA was purified by phenol-chloroform extraction and ethanol precipitation.

Insert fragments for Gibson Assembly were generated either by PCR or synthesised DNA oligos. In the case of small inserts, including gRNAs, DNA oligos were produced with homologous 5' and 3' arms to the blunt ends of the vector backbone. The complementary oligos were annealed together using 1 x Oligo annealing buffer and incubated for 5 minutes at 90 °C. Hybridisation occurred by gradual cooling to room temperature, forming a dsDNA insert. Other insert fragments, generated by PCR, were designed with 5' and 3' sequence complementarity with the cut vector. All insert fragments were purified by DNA phenol-chloroform extraction and ethanol precipitation.

Typically for ligation, a 1:3 volume ratio (depending on relative size) of vector to insert was used with 1 x Gibson Reaction Master Mix (NEB). Ligations were incubated for 1 hour at 50 °C before subsequent transformation into competent bacterial cells.

2.5.7 Transformation of plasmids into bacteria

For transformation into bacterial cells, 10 - 20 ng of purified plasmid DNA or 4 μl of a ligation reaction was used. DNA was equilibrated to 4 °C by placing on ice for 5 minutes; concurrently 60 μl of bacterial cells were thawed on ice. After this incubation, 60 μl of thawed cells were added to the DNA and mixed once by pipetting. This was kept on ice for 5 minutes, before undergoing a heat-shock at 42 °C for 90 seconds and immediately being transferred to ice for 2 minutes. 500 μl of Super Optimal broth with Catabolite repression (SOC) medium (NEB) was added and mixed gently by inversion. Cells underwent a recovery step for 1 hour at 37 °C to allow expression of the antibiotic resistance gene, present in the transformed plasmid. After this time, approximately 250 μl of cells were pipetted onto a LB agar plate supplemented with the appropriate antibiotic and stored overnight at 37 °C.

2.5.8 Plasmid purification from bacteria

Plasmids were isolated from single bacterial colonies that had grown on LB agar plates supplemented with the appropriate selection drug. Single colonies

were used to inoculate 6 ml of LB media supplemented with the relevant antibiotic, then stored at 37 °C overnight in a shaking incubator (180 rpm) to allow for bacterial growth. Following this, a cell pellet was obtained by centrifugation for 5 minutes at 13,000 rpm. Plasmids were purified from cell pellets using the QIAprep Spin Miniprep Kit following manufacturer's instructions.

2.5.9 Plasmid construction for CRISPR/Cas9

Repair templates generated for CRISPR/Cas9, were assembled using the empty pUC19 backbone. Homology arms for the gene of interest were synthesised flanking the poly(A) site using Integrated DNA Technologies (IDT) and ligated into pUC19. Using Gibson Assembly (NEB), a pre-synthesised (IDT) AID-P2A was inserted into the vector with either a hygromycin or neomycin resistance gene. For this incorporation, the vector was linearised between the penultimate and stop codon.

IDT synthesised gRNA oligos to our gene of interest were inserted into a Cas9 expression plasmid (pX330-U6-Chimeric_BB-CBh-hSpCas9) using Gibson Assembly (NEB). Sequences to create gRNA oligos were obtained using the online Benchling software (https://benchling.com).

HCT116:TIR1 cells had been previously made in the West lab by isolating human codon optimised TIR1 from pBABE osTIR1 and using SFI1 restriction sites in the pSBbi-Blast empty vector. CPSF73-AID doxycycline inducible and Ints11-SMASh cell lines were generated by Professor Steven West using the same approach as above except the Ints11-SMASh repair template contained a SMASh-tag instead of AID-P2A. Additionally, for the CPSF73-AID doxycycline inducible cell line osTIR1 was expressed under a tetracycline promoter using the pMK243 plasmid.

2.6 Tissue culture

2.6.1 Cell lines

Multiple cells lines were created using HCT116 cells, by simple transfection protocols or CRISPR/Cas9 genome engineering. HCT116 cells were used due to their obligate diploid karyotype, to increase efficiency of CRISPR/Cas9 methods. A description of the cell lines generated are shown in Table 2.4. HCT116:TIR1, INTS11-SMASh, XRN2-AID and CPSF73-AID cell lines were made by Steven West.

Table 2.4 Cell lines

Cell line name	Description
HCT116	Unmodified human colon carcinoma cells; parental cells
HCT116:TIR1	SB - integrated TIR1 in HCT116 cells
DIS3-AID	SB - integrated TIR1; homozygous 3' AID tagged DIS3
INTS11-SMASh	Homozygous 3' SMASh tagged INTS11
CPSF73-AID (doxycycline inducible)	Homozygous 3' AID tagged CPSF73; osTIR1 expressed under a Tet promoter
XRN2-AID	SB-integrated TIR1; homozygous 3' AID tagged XRN2

For depletion of the protein of interest, 500 μM of auxin was added to DIS3-AID for 60 minutes or XRN2-AID for 120 minutes. In CPSF73-AID doxycycline inducible cells, TIR1 expression was induced by addition of 2 μg / ml doxycycline for 16 hours, followed by 500 μM auxin addition for 2 hours to deplete CPSF73. In INTS11-SMASh cells, 2 μM of asunaprevir was added for 48 hours. Untreated controls were incubated with ethanol (solvent), for an equivalent time to treated cells.

2.6.2 Cell growth and maintenance

All cells were incubated at 37 °C with 5 % CO_2 and maintained in T75 flasks containing Dulbecco's modified Eagle's medium (DMEM) supplemented with 10 % foetal calf serum (FCS) and 1 % Penicillin Streptomycin. Additionally, cells expressing TIR1 were maintained with blasticidin (5 μg / ml), to prevent loss of the TIR1 Sleeping Beauty (SB) plasmid.

Once cells had grown to approximately 80 % confluency, they were passaged. For passaging, cells were washed with 1 x PBS, then washed with Trypsin PBS-EDTA and incubated for 3 minutes. 10 ml of DMEM media was added to cells to neutralise trypsinisation and pipetted up and down to remove cells from the flask wall. Approximately 1 ml of cells were seeded into a T75 flask containing 12 ml of DMEM media, then allowed to grow at 37 °C with 5 % CO_2.

2.6.3 Long-term storage of cultured cell lines

Confluent T75 flasks (Greiner) of cultured cell lines were passaged as above, with cells resuspended in 10 ml of DMEM and centrifuged at 500 x g for 5 minutes. For long-term storage at - 80 °C, cell pellets were resuspended in 1 ml of FCS supplemented with 10 % DMSO and transferred to a cryovial. To recover cells from storage, cells were slowly thawed to room temperature and homogenised in 5 ml of DMEM. After centrifugation for 5 minutes at 500 x g, all media was removed and cells resuspended in 12 ml of DMEM (Sigma) and placed into a T75 for growth.

2.6.4 Generation of the HCT116:TIR1 cell-line.

Using the SB transposon system, HCT116 cells expressing TIR1 (*HCT116:TIR1*) were generated (Hackett et al 2010, Skipper et al 2013, Hou et al 2015). 400 ng of SB transposon and 50 ng of transposase plasmids were transfected using JetPrime Reagent according to manufacturer's protocol. After 48 hours, cells were passaged into 10 cm dishes containing DMEM supplemented with 20 μg / ml blasticidin, until single colonies could be picked and allowed to grow.

2.6.5 Generation of stable cell-lines

For generating CRISPR/Cas9 stable cell-lines with integration of either 3' AID tag or SMASh tag, cells were split into 3 cm plates at approximately 30 % confluency. 2 μg of repair plasmid containing hygromycin or neomycin resistance and 2 μg of gRNA were transfected into cells using JetPrime Reagent (Polyplus). After 24 hours, media was changed and at 48 hours cells were passaged into 10 cm plates.

Stable cell-lines that had been successfully transfected with either a 3' AID tag or SMASh tag on the target gene and, where stated, TIR1 integration at SB loci or doxycycline inducible TIR1 transfection, were selected using antibiotic resistance. Single colonies that grew in the presence of antibiotics were picked, transferred to 24 well plates and screened for homozygous integration. Final concentrations of antibiotics used are as follows:

- **Hygromycin**: 150 μg / ml
- **Neomycin**: 800 μg / ml
- **Blasticidin**: 20 μg / ml
- **Puromycin**: 1 μg / ml

2.6.6 Genomic DNA isolation from stable cell-lines

Positive cultured cells were screened for homozygous CRISPR/Cas9 repair cassette incorporation by extracting genomic DNA using QuickExtract DNA Extraction Solution (Cambio). Cells were grown in 3 cm dishes to 80 % confluency, before being washed with 4 °C 1 x PBS. Cells were scraped and spun down in 1 ml of 1 x PBS, for 5 minutes at 500 x g. Supernatant was discarded and cell pellets resuspended in QuickExtract depending on pellet size. This was incubated for 6 minutes at 65 °C, then samples vortexed before incubation for 2 minutes at 98 °C to denature the QuickExtract. Subsequently, 1 μl was used for future PCR and DNA stored at - 20 °C.

2.6.7 RNAi transfections

To deplete proteins using RNAi methods, cells were firstly split into 6-well dishes at approximately 30 % confluency and grown in DMEM media without antibiotics. Appropriate siRNA was transfected into cells using Lipofectamine RNAiMax (Life Technologies) following the manufacturer's guidelines. Transfection was repeated after 48 hours and RNA was isolated 24 hours later.

2.6.8 Transfection by electroporation

Antisense morpholino oligonucleotides (AMO) were transfected into cells by electroporation. To do this, 1 x T75 flask of cells were grown in DMEM media without FCS or antibiotics. Cells were trypsinised, resuspended in 10 ml of media and centrifuged for 5 minutes at 300 x g. The cell pellet was resuspended in 800 μl media. 10 μM of control morpholino was added to half of the cell volume and 10 μM of U7 snRNA AMO was added to the remaining cells, ensuring full resuspension. Afterwards cells were placed into a 4mm cuvette and electroporated at 280 V, with a capacitance of 950 μF and infinity resistance, using a BioRad Gene Pulser Xcell. Electroporated cells were resuspended in 6 ml of media before allowing to grow in a 3 cm plate for 5 hours. After this time RNA was extracted from cells following the protocol in Section 2.7.4.

2.7 Molecular Biology

2.7.1 Protein extraction for Western blot

A confluent 3 cm plate of cells was used for protein extraction. Cells were washed with PBS, then scraped off the plate in 1 ml of 1 x PBS and added to a 1.5 ml Eppendorf. Cells were spun for 5 minutes at 500 x g to create a cell pellet. The cell pellet was resuspended in RIPA buffer using 10 x volume of the cell pellet; the samples were vortexed then placed on ice for 20 minutes. Finally, samples were spun for 10 minutes at 13,000 rpm. The supernatant containing protein was removed and stored at - 20 °C.

2.7.2 SDS-PAGE

Protein samples were separated by molecular weight using Sodium Dodecyl Sulphate PolyAcrylamide Gel Electrophoresis (SDS-PAGE). Gels were made using a 5 % stacking gel and a resolving gel, with the resolving gel varying in percentage of acrylamide depending on the size of the POI. Most commonly a 10 % resolving gel was made as shown in Table 2.5. After addition of TEMED and 10 % Ammonium Persulphate (APS), the resolving gel was poured into the assembled Mini-PROTEAN Tetra Cell Casting Module (Bio-Rad). 500 μl of dH_2O was pipetted on top of the resolving gel to minimise bubbles and then left until set, approximately 15 minutes. The stacking gel was poured on top and a 1.5 mm comb inserted, then the gel left to set. After setting, the comb was removed and wells were washed with 1 x SDS running buffer.

Cast gels were placed into a Mini-PROTEAN system (Bio-Rad) inside a Buffer tank (Bio-Rad). The Buffer tank was then half filled with 1 x SDS-PAGE Running Buffer. Protein samples were prepared by addition of 4 x SDS-PAGE sample buffer containing β-mercaptoethanol and heated at 95 °C for 3 minutes to denature proteins. An appropriate protein marker was loaded and up to 20 μl of each sample. The gel was then run at 25 mA until the dye front passed through the stacking gel, upon which the gel was run at 50 mA until passed through the resolving gel.

2.7.3 Western Blot

Following from SDS-PAGE, proteins were transferred from the gel to a nitrocellulose membrane (GE Healthcare) using the Trans-Blot Turbo Transfer System (Bio-Rad) and Transfer Buffer. For 1 hour the membrane was blocked in 5% blocking solution, whilst shaking. After blocking, the membrane was incubated on a shaker with 2% blocking solution and primary antibody for 1 hour. The membrane was then washed 3 times for 5 minutes each in PBST. Membranes were incubated for 1 hour with 2 % blocking solution containing a 1:10,000 concentration of secondary antibody. Afterwards the membrane was washed 3 times in PBST for 5 minutes each. To visualise proteins, an equal volume of ECL 1 and ECL 2 solution were added to the membrane and images captured on a Gel Doc XR + system (Bio-Rad).

Table 2.5 Solutions and amounts to make 10 ml of Resolving Gel or 6 ml Stacking Gel.

	8%	10%	12%	Stacking Gel 5%
dH_2O	4.73 ml	4.07 ml	3.35 ml	3.44 ml
4 x Resolving Gel	2.5 ml	2.5 ml	2.5 ml	*1.5 ml
30 % acrylamide (Protogel)	2.67 ml	3.33 ml	4 ml	1ml
TEMED	100 μl	100 μl	100 μl	60 μl
10 % APS	6 μl	10 μl	15 μl	6 μl

* 4 x Stacking Gel was used instead of 4 x Resolving Gel

2.7.4 Total RNA Extraction

Total RNA was extracted from cells grown to 80 % confluency in a 6-well plate. All media was removed and cells were incubated with 1 ml of TRI Reagent (Sigma) for 5 minutes. Cells were transferred to a 1.5 ml Eppendorf, containing 200 μl chloroform (Sigma). This was vortexed for 10 seconds then left at room temperature for 5 minutes. Afterwards, cells were centrifuged for 15 minutes at 13,000 rpm. The upper aqueous layer was transferred to a new 1.5 ml Eppendorf containing a 1:1 (v/v) ratio of isopropanol (Sigma), then briefly vortexed before a 10 minute spin at 13,000 rpm. Supernatant was completely removed and 650 μl of 70% ethanol added. This was centrifuged again for 10 minutes at 13,000 rpm. The supernatant was removed and pellet air-dried for 5 minutes to remove residual ethanol, before resuspension in 87 μl of dH_2O. RNA samples were treated for 1 hour at 37 °C with 2 μl Turbo DNase (ThermoFisher), 10 μl Turbo DNase buffer and 1 μl RNase Inhibitor Murine (NEB) to remove contaminating DNA whilst preventing RNA degradation. Subsequently, DNase treatment was inactivated by RNA phenol-chloroform extraction and ethanol precipitation as described in Section 2.5.4, after which samples were stored at - 20 °C. A 3 μl aliquot of each RNA sample was run on a 1% agarose gel to ensure RNA obtained was not degraded.

2.7.5 Reverse Transcription (RT-PCR)

RT-PCR was conducted on purified and genomic DNA depleted RNA, to produce cDNA for further use in Real-time Quantitative PCR (qPCR). 1 μg of RNA was used for each RT-PCR reaction, with a RT-PCR control for each sample and RNA concentrations being determined by a NanoDrop 2000 spectrophotometer (ThermoFisher). To 1 μg of RNA, 1 μl random hexamers (Bioline) were added and total volume made up to 10 μl. This was primed for 5 minutes at 70 °C, then immediately placed on ice, before addition of 10 μl reverse transcription master mix or reverse transcription control mix for RT-PCR controls (see Table 2.6). The reactions were incubated in a PCR machine for 5 minutes at 25 °C, 1 hour at 42 °C, 20 minutes at 70 °C and held at 10 °C. Afterwards, the cDNA samples were diluted in 30 μl of dH_2O and stored at - 20 °C.

Table 2.6 Reagents and amounts required for RT-PCR

Master mix components for reverse transcription	**Amount required for 1 x 10 μl reaction**
10 mM dNTP mix (ThermoFisher)	1 μl
10 x DTT (NEB)	2 μl
RNase free water	2.5 μl
Protoscript II RT reaction buffer (NEB)	4 μl
*Protoscript II RT enzyme (NEB)	0.5 μl
*Same component volumes used to make Reverse Transcription control mix, except Protoscipt II RT enzyme is omitted and instead 0.5 μl dH_2O added.	

2.7.6 Real-Time Quantitative PCR (RT-qPCR)

For RT-qPCR, all reactions were set-up in triplicate for both RT-PCR and RT-PCR control samples. In each reaction, 20 – 50 ng of cDNA was added to a master-mix containing the following to give a total reaction volume of 8 μl: 100

nM of reverse primer, 100 nM of forward primer, 4 μl of 2 x Brilliant III SYBR Green Master Mix (Agilent) and 2.8 μl dH_2O. RT-qPCR was conducted in a Rotor-Gene Q (Qiagen) using the incubation steps shown in Table 2.7, to detect amplicons of < 150 - 200 nt in length. Data acquisition occurred on the green channel, during the incubation for 10 seconds at 60 °C. A minus RT control was included in all RT-qPCR experiments to check for contaminating DNA.

Table 2.7 RT-qPCR incubation steps

Temperature	Time	Number of cycles
95 °C	3 minutes	1
95 °C 60 °C	10 seconds 10 seconds	40 - 50 cycles

For normalisation, spliced β-actin primers were used as a housekeeping control gene that had previously been shown in the laboratory to have stable expression in HCT116 cells. Rotor-Gene Q Series Software v2.3.1 was used for analysis to calculate fold enrichment relative to a control sample and melt curves were investigated to check primer specificity through the amplification of a single DNA product. Data was analysed to determine the delta delta C_T relative quantitation values, with each sample normalised by comparison to the housekeeping gene, for the amount of template cDNA, then further normalised to a control sample i.e. non-treated cells. All RT-qPCR experiments were replicated in triplicate and figures containing RT-qPCR data show the mean of three independent RT-qPCR experiments.

2.7.7 Cell colony formation assay

To conduct a cell colony formation assay, 300 cells were seeded into 10 cm plates with 8 ml of DMEM media supplemented with either 500 μM auxin or ethanol (solvent) for 10 days. Media was replenished every 3 - 4 days. After 10 days, media was removed and cells placed on ice. Cell plates were washed twice in 4 °C PBS and cells were fixed by incubation of 10 ml methanol for 10 minutes. Plates were stained with 0.5 % crystal violet and 25 % methanol solution for 10

minutes. Excess crystal violet was removed by washing with dH_2O, plates were air-dried and images taken. For analysis, Image J particle analyser function (Schindelin et al 2012) was used to count cell colonies. Colonies with a density range between 100-600 pixels and a circularity rating of 0.75 – 1 were counted.

2.7.8 Chromatin immunoprecipitation protocol (ChIP)

To bind antibody to Sheep Antimouse Dynabeads M280 (ThermoFisher), firstly 40 μl of beads per sample were rinsed in 500 μl RIPA ChIP buffer, resuspended in 1 ml of RIPA ChIP buffer then split equally into two tubes. The volume was made up to 1 ml using RIPA ChIP buffer with protease inhibitors and to one tube, 2 μg of RNA Polymerase II CTD antibody per sample was added. The Dynabeads were then rotated at 18 rpm in the cold room for 2 – 3 hours.

Cells were seeded in 10 cm plates with DMEM media supplemented with 10 % FCS. Cell plates were rinsed in 4°C 1 x PBS solution before addition of 10 ml PBS directly to cells. Formaldehyde was added to a final concentration of 0.5 % and plates were placed on a shaking platform (60 rpm) for a maximum of 10 minutes. To quench the crosslinks, 1 M glycine was added to a final concentration of 125 mM, then cells left on the shaking platform for 5 minutes. Afterwards, cells were scraped into a 15 ml falcon tube and centrifuged at 500 x g for 5 minutes at 4 °C. Pelleted cells were resuspended in 5 ml of PBS and centrifuged again at 500 x g for 5 minutes at 4 °C. All supernatant was removed and cells resuspended in 300 μl of RIPA ChIP buffer per 30 μl cell pellet volume. The cell suspension was transferred to a sonication tube, before being placed in the Bioruptor Plus sonication device (Diagenode) for 10 minutes on high setting, 30 seconds on and 30 seconds off. Sonicated cells were centrifuged for 10 minutes, at 13000 rpm and 4 °C. Supernatant was aliquoted so that two Eppendorfs contained 45 % of the sonicated cell volume each. The remaining 10 % cell volume was stored at -20 °C. For each Eppendorf, the volume was made up to 1 ml with RIPA ChIP buffer.

Dynabeads incubated with or without antibody were placed in a magnetic rack and all supernatant removed. The beads were then resuspended in 10 μl of RIPA ChIP buffer per sample before 10 μl of bead suspension was added to each

tube, being careful to distinguish between plus and minus antibody samples. For immunoprecipitation, suspensions were rotated at 18 rpm in the cold room for 2 hours. Using a magnetic rack all supernatant was removed from samples. The beads then underwent a series of washes performed in the cold room, as described in Table 2.8.

Table 2.8 ChIP bead-wash protocol

Solution	Amount	Repetition	Rotation
RIPA ChIP buffer	500 μl	x 2	No
ChIP wash buffer	500 μl	x 1	No
ChIP wash buffer	500 μl	x 3	5 minutes of 18 rpm rotation in between washes
RIPA ChIP buffer	500 μl	x 2	No

1.5 ml of Elution buffer per sample was freshly prepared. Beads were resuspended in 250 μl of elution buffer and rotated on a wheel for 15 minutes at room temperature. Using a magnetic rack, the supernatant was transferred to a new Eppendorf. The elution process was then repeated, adding the second 250 μl of eluate to the first. For an input sample control, 10 μl of the 10 % sample stored in the freezer previously was placed into a new Eppendorf. To all samples, 25 μl of 4 M NaCl was added and incubated for 4.5 hours at 68 °C. DNA was purified by DNA phenol-chloroform extraction and ethanol precipitation before samples were stored at - 20 °C.

2.8 RNA-Seq

2.8.1 Seeding cells

Cells were seeded in 10 cm plates with DMEM media supplemented with 10 % FCS. For CPSF73-AID doxycycline inducible HCT116 cells, 2 μg / ml of doxycycline was added the day of seeding and cells incubated for 16 hours. Following this, 500 μM auxin was added to the appropriate cells for 2 hours before RNA extraction. For Ints11-SMASh HCT116 cells, cells were seeded in 10 cm plates as above. Asunaprevir was added at a final concentration of 2 μM for 30 hours before RNA extraction occurred.

2.8.2 Nuclear RNA extraction for RNA-Seq

Cells grown on 10 cm plates were at approximately 80 % confluency upon nuclear RNA extraction. Firstly all media was removed and cells washed in 4 °C 1 x PBS solution. Cells were scraped off the plate in 10 ml of 1 x PBS and placed into a 15 ml falcon tube. This underwent centrifugation for 5 minutes at 500 x g. Cell pellets were resuspended in 4 ml of HLBN and incubated on ice for 5 minutes. The solution was then carefully underlayed with 1 ml of HLBNS. This was centrifuged for 5 minutes at 500 x g to obtain a nuclear pellet. The nuclear pellet was resuspended in 5 ml HLBN to remove any traces of cytoplasmic material and centrifuged for 5 minutes at 500 x g, after which the supernatant was discarded. To extract nuclear RNA from the isolated nuclei, Tri Reagent (Sigma) was used as previously described in section 2.7.4 and samples stored at - 80 °C.

2.8.3 Ribosomal RNA (rRNA) removal

RNA quality was determined by using the TapeStation apparatus (Agilent). rRNA was extracted from 1 μg of genomic DNA depleted, nuclear RNA using the Ribo-Zero Gold rRNA removal kit following manufacturers protocol.

2.8.4 Purify RNA beads using Agencourt RNAClean XP Kit

Before sample library preparation, RNA samples depleted of rRNA were purified to remove remaining salts and buffers and to concentrate samples. To each sample, 160 μl of RNAClean XP beads (Beckman Coulter) were added and mixed by pipetting. After a room temperature incubation of 15 minutes, samples were placed on a magnetic stand until the beads were captured and liquid was clear. The supernatant was removed and beads washed with 200 μl of 80 % ethanol. This wash step was repeated, all ethanol removed and the pellet air-dried for 3 minutes. Dried beads were resuspended in 11.5 μl of Resuspension Buffer, incubated at room temperature for 2 minutes then placed on the magnetic rack for 5 minutes. All supernatant was removed, placed into a 0.5 ml Eppendorf and stored at - 80 °C. Before proceeding, the quality and quantity of depleted RNA samples were checked using 2 μl from each sample on a Tapestation High Sensitivity RNA ScreenTape (Agilent).

2.8.5 TruSeq Stranded mRNA

Purified RNA was used to create a library of template molecules using the TruSeq Stranded Total RNA Sample Preparation Kits (Illumina). 8.5 μl of Elute, Prime, Fragment High Mix was added to each sample and mixed by pipetting. This was incubated for 8 minutes at 94 °C, then held at 4 °C. Following manufacturer's instructions, Actinomycin was added to the First Strand Synthesis Act D mix to prevent spurious DNA synthesis. A master mix of Superscript II and First Strand Synthetic Act D was made with a volume ratio of 1:9 respectively, for each sample. Of this master mix, 8 μl was added to each sample then incubated for the following to synthesise the first cDNA strand: 10 minutes at 25 °C, 15 minutes at 42 °C, 15 minutes at 70 °C and held at 4 °C.

To synthesise the second strand cDNA, firstly 20 μl of Second Strand Marking Master Mix was added to each sample and mixed by pipetting. This was incubated for 1 hour at 16 °C. For separation of the double-stranded cDNA from the second strand reaction mix, 90 μl of AMPure XP beads (Beckman Coulter) were added and incubated at room temperature for 15 minutes. Beads were captured on a magnetic rack and supernatant discarded. Samples were washed

with 200 μl of 80 % ethanol twice then beads air-dried for 3 minutes. Dried beads were resuspended in 17.5 μl Resuspension Buffer, then placed on the magnetic rack and 15 μl of supernatant, now containing blunt-ended cDNA, transferred to a new 1.5 ml Eppendorf.

To prevent ligation of blunt fragments, a single 'A' nucleotide was added to 3' ends. 12.5 μl of A-Tailing Mix was added to samples and incubated at 37 °C for 30 minutes, 70 °C for 5 minutes and held at 4 °C. Index adaptors were then ligated to the ends of double stranded cDNA. Firstly 2.5 μl of Ligation Mix and 2.5 μl RNA Adaptor Index was added to each sample and mixed by pipetting. An incubation of 10 minutes at 30 °C occurred before 5 μl of Stop Ligation Buffer was used to stop the ligation reaction. 42 μl of AMPure XP beads (Beckman Coulter) were added, incubated at room temperature for 15 minutes then beads captured on a magnetic rack. The supernatant was removed and beads washed with 200 μl of 80 % ethanol twice. Beads were air-dried for 3 minutes and resuspended in 52.5 μl of Resuspension Buffer, then placed on a magnetic rack and 50 μl of supernatant was transferred to a new 1.5 ml Eppendorf. Another 50 μl of AMPure XP beads (Beckman Coulter) were added to the supernatant and incubated at room temperature for 15 minutes. The magnetic rack was used to remove the supernatant and beads washed twice with 80 % ethanol. After air-drying the beads, they were resuspended in 22.5 μl Resuspension Buffer, then captured on a magnetic rack and 20 μl of supernatant transferred to a new Eppendorf.

To enrich for DNA fragments with adaptor molecules on either end and amplify DNA amounts, 5 μl of PCR Primer Cocktail and 25 μl of PCR Master Mix were added to each sample. This was incubated as shown in Table 2.9. Afterwards, 40 μl of AMPure XP beads (Beckman Coulter) were added and incubated at room temperature for 15 minutes. Beads were captured on a magnetic rack, supernatant discarded and beads washed twice with 200 μl of 80 % ethanol. Air-dried beads were resuspended in 32.5 μl Resuspension Buffer then placed on a magnetic rack. 30 μl of supernatant was transferred to a new tube and stored at - 80 °C to be further used in a sequencing reaction.

Table 2.9 DNA fragment enrichment PCR for RNA-Seq

Step	Temperature	Time	Number of cycles
Initial denaturation	98 °C	30 seconds	1
Denaturation Annealing Extension	98 °C 60 °C 72 °C	10 seconds 30 seconds 30 seconds	15
Final extension	72 °C	5 minutes	1
Hold	4 °C	-	-

2.8.6 Sequencing

Before sequencing, DNA fragment size and concentration were determined by running each sample on a Tapestation D1000. Samples were given to the Exeter Sequencing Service facility to conduct a HiSeq 2500 (Illumina) rapid run single-end with a read length of 50 bp.

Table 2.10 Total mapped reads for all RNA-Seq data using merged replicate libraries*

RNA-Seq cell lines	Total mapped reads	Total exon mapped reads
DIS3-AID	76561054	32900452
DIS3-AID with auxin	56625921	22512088
INTS11:SMASh	28406998	6241067
INTS11:SMASh with asunaprevir	27875162	6563014
CPSF73-AID	81473198	22067230
CPSF73-AID + auxin	81136711	17050608

*There is only one replicate for both INTS11:SMASh and INTS11:SMASh with asunaprevir

2.9 Bioinformatic analysis

Bioinformatic software used in the analysis of sequencing data obtained through RNA-Seq is shown in Table 2.11. Software derived from a Bioconductor package in the R environment is denoted by a *. Bioinformatical analysis on RNA-Seq data derived from the DIS3-AID cell line only, was conducted by Lee Davidson.

2.9.1 Read alignment of RNA-Seq data

The sequencing quality of unprocessed single-end 50 bp reads obtained from RNA-Seq data was determined by FastQC (Andrews, 2010). TrimGalore! was used to remove adaptors from reads and to discard reads shorter than 20 bp (Krueger, 2012). Reads were aligned to the human genome using Hisat2 with a known splice sites file and all analysis conducted used the Ensembl GRCh38.p10 and GRCh38.90 human gene annotations. All non-mapped reads and reads with a mapping quality less than 20 were removed using SAMtools (Li, 2011).

2.9.2 Differential Expression Analysis

featureCounts was used to count all reads over a gene or transcript and determine expression levels, only counting reads with a minimum mapping quality score of 30 (Liao et al, 2014). The Integrated Genome Viewer (IGV) suite was used to visualise normalised coverage plots (Reads per Kilobase of transcript per Million mapped reads, RPKM) throughout the genome (Robinson et al, 2011).

Table 2.11 Bioinformatic software

Software name	Version	Description	Reference
BamTools	v2.4	Processes BAM files	Barnett et al, 2011
BEDTools	v2.2.5	Allows comparison of large sets of genomic features; tools for genomic analysis	Quinlan and Hall, 2010
BEDOPS	v2.4.33	Toolkit for processing genomic data	Neph et al, 2012
CutAdapt	v1.14	Removes adaptor sequences from sequencing reads	Martin, 2011
Deeptools	v3.0.2	Python tools for analysis of high-throughput sequencing	Ramírez et al, 2016
FastQC	v0.11.5	Quality control of raw sequence data	Andrews, 2010
FeatureCounts	v1.6.0	Counts reads to genomic features	Liao et al, 2014
GenomicRanges*	v1.30.2	Storage and manipulation of genomic intervals and variables	Lawrence et al, 2013
HISAT2	v2.1.0	Alignment programme for mapping sequencing reads	Kim et al, 2015
IGV	v2.4.3	Visualisation tool for genomic datasets	Robinson et al, 2011
R	v3.4.4	Software environment for statistics and graphics	http://www.R-project.org
Rtracklayer*	v1.38.3	Interface for manipulating annotation tracks	Lawrence et al, 2009
SAMtools	v1.6	Manipulates alignments in the SAM format	Li, 2011
TrimGalore!	v0.4.4	Applies quality and adapter trimming to FastQ files (via CutAdapt)	Krueger, 2012

2.9.3 Metagene

Using the gene count file created via featureCounts, which counted the number of reads aligned to a gene or transcript, any gene or transcript with less than 200 reads were removed. To each gene, an increased transcriptional window of 1 Kb upstream of the TSS was added. Additionally either 7 Kb, 50 Kb or 100 Kb was added downstream of the TES depending on the metaplot graph. Any genes that overlapped after this extension were removed to prevent repetitive counting of mapped reads using BEDTools merge (Quinlan and Hall, 2010). RPKM normalised reads from the remaining genes were used to generate metaplot profiles. This was conducted using the deeptools suite (Ramírez et al, 2016) and the R environment for graphical design (http://www.R-project.org).

2.10 Primers and oligonucleotides

Table 2.12 Primers and oligonucleotides for creation of the DIS3:AID cell line

Primer name	**Sequence**	**Description**
DIS3 gRNA F	CACCGTCCATGTTTGAAGTATCAGT	Used to make the gRNA for creation of the DIS3:AID cell line.
DIS3 gRNA R	AAACACTGATACTTCAAACATGGAC	
DIS3 screening primer 1 F	TCTTTAGGCCACGGGATTCT	First set of screening primers to determine insertion of the AID tag to the DIS3 gene. Designed outside of the homology arms and used in a nested PCR screen.
DIS3 screening primer 1 R	TGCCTTTCTACCAATTCCCAA	
DIS3 screening primer 2 F	TCCATTCTCCTGCCTAGTCT	Second set of screening primers to determine insertion of the AID tag to the DIS3 gene. Designed outside of the homology arms and used in a nested PCR screen.
DIS3 screening primer 2 R	CCTCAACACTGACAGCTTCC	

Table 2.13 Homology arm sequences for DIS3:AID

DIS3 HDR 5’	CTTGAAATCAACTCTGATTCTGTCAATCACAGTGGCTCCCC ATTGGGAAGGCTGTTTTGTAGTTAAAAAGAACACTTCCTAA ATGACATGCTTCTCACCTGTTGAGACCATGTCTAGCTTTTA CATTTTTGAACCACTGCTACTTTGTAAAATACCTTCTGTGTA TAAAACCTTTAATTAGCCCCCTTTCCCCTCCCTACCACTACA TCCTTTTAAATTTGAAGCTGGCAGTGGGGAAGGGGAGGATG AGGTTGAGATGTATTCTATCCTTTAAATCACCTTATTTCCCCC CATTTGCATTACTTTAGATACCAGGAATAAGCATTCCTACAGA CACATCTAACATGGACCTTAATGGACCAAAGAAAAAGAAGAT GAAGCTTGGAAAA
DIS3 HDR 3’	TAGCTATATTCAACAAAAATCTTCAAAGACTGGTTTCTTTTTT AAAAGAAAAAACTTGAAAGAACACTTCTAAGCCTAAGTGTGT GATACAGTTTGTTACTTTTAAGTACATTTTAATAATTTCAGAC ATCTGCATTTTTATTGAACAGTTGACTGTATCTGACCCATCAT ACTACTATACTTCTGGGTTGAACAGAATTATTTATGCAGAATA ATTCAATTGAATATCCATCACTTAAATACAGTGACAGGACAGC AACTTCAGGG ATCTGTAAAGATCATTTAAATGGAGT

Table 2.14 Primers for PROMPT detection by qRT-PCR

Primer name	Sequence	Description
STK II F	GGGAGTCTAAGGAAAAGGAG	Primers designed to detect a PROMPT upstream of the STKII gene TSS.
STK II R	CAGTGAAAGGAGAGCGTATC	
SERPINB8 F	CTACTGATCACACCCTCCTC	Primers designed to detect a PROMPT upstream of the SERPINB8 gene TSS.
SERPINB8 R	CATTCTGGATGCATGTGTAG	
FOX P4 F	TGCACAATTTCACACCTAGA	Primers designed to detect a PROMPT upstream of the FOX P4 gene TSS.
FOX P4 R	ATGTTAGTGACACCTGCACA	
RBM39 F	GGAAATAGTGGAGAAAAGCA	Primers designed to detect a PROMPT upstream of the RBM39 gene TSS.
RBM39 R	CATTTTTGAAGGAACGGTAG	

Table 2.15 Primers for detection of abortive transcripts by qRT-PCR

Primer name	Sequence	Description
NFU1 in1-in1 F	GGCTCAGAGACCCAGTTCTT	Primers to detect prematurely terminated NFU1 transcripts, by measuring RNA levels over the first intron.
NFU1 in1-in1 R	CCTTGGACATGTCACCTCCT	
NFU1 ex2-in2 F	ACACCATTAAGAAACAGCCTCT	Primers measuring NFU1 RNA levels over the exon-intron junction as a control.
NFU1 ex2-in2 R	TGATCCACAAAATCCTAGCACAG	
CLIP4 in1-in1 F	GTCAGGCTGTTCACGTCATC	Primers to detect prematurely terminated

CLIP4 in1-in1 R	TTTCAAAGGCGCCCGTTTTA	CLIP4 transcripts, by measuring RNA levels over the first intron.
CLIP4 ex2-in2 F	TCCTTTGTTTGGAAGATACCCA	Primers measuring CLIP4 RNA levels over the exon-intron junction as a control.
CLIP ex2-in2 R	GGCGTAACAGAGAAGTCAAGT	
PCBP1-AS1 ex1-in1 F	CCACCTCCGCGAGTTTTATG	Primers to detect prematurely terminated PCBP1-AS1 transcripts, by measuring RNA levels over the first intron.
PCBP1-AS1 ex1-in1 R	ATGCTTTGGTACTGTGGGGA	
PCBP1-AS1 ex3-in3 F	AATGTGACTTTGGAGCCAGC	Primers measuring PCBP1-AS1 RNA levels over the exon-intron junction as a control.
PCBP1-AS1 ex3-in3 R	ACCGAGATGAAACTGAGGGA	
C2orf42 in1-in1 F	TTCCAACACCAGTCCCTTGA	Primers to detect prematurely terminated C2orf42 transcripts, by measuring RNA levels over the first intron.
C2orf42 in1-in1 R	CGACATGGGATTTGGGAAACA	
C2orf42 ex2-in2 F	ATTGGCTGGTGGAGAAAGGAG	Primers measuring C2orf42 RNA levels over the exon-intron junction as a control.
C2orf42 ex2-in2 R	TCCCTTCCATCATTCCCCAC	

Table 2.16 Primers for detection of snRNAs and RNA levels downstream of their TES

Primer name	Sequence	Description
RNU5B-1 300 bp F	CCGGTAATCCCACTGCATTG	Detects RNA levels 300 bp downstream of the RNU5B-1 TES
RNU5B-1 300 bp R	CATTGTCCATGTGTGCCGAT	
RNU5B-1 1.5 Kb F	AGAATCGCTTGAACCTGGGA	Detects RNA levels 1.5 Kb downstream of the RNU5B-1 TES
RNU5B-1 1.5 Kb R	CCAGCCTGTGTGATAAAGCC	
RNU5D-1 200 bp F	TGTTTGTTGCGAGGTGTGAG	Detects RNA levels 200 bp downstream of the RNU5D-1 TES
RNU5D-1 200 bp R	GGAAAATCCCTTGAAGCCGG	
RNU5D-1 3.5 Kb F	TAGCTGAATGTGGTCGTGGT	Detects RNA levels 3.5 Kb downstream of the RNU5D-1 TES
RNU5D-1 3.5 Kb R	TCCTGACCTCATGATCTGCC	
RNU1-28P 300 bp F	GTGCTTTCTCCAGGCCAAAG	Detects RNA levels 300 bp downstream of the RNU1-28P TES
RNU1-28P 300 bp R	GGACCAGGATTAATTGCCCG	
RNU1-28P 500 bp F	TCCGGCTTAGAGGTTTAGGA	Detects RNA levels 500 bp downstream of the RNU1-28P TES
RNU1-28P 500 bp R	AGTCTCCTGTTCTTGAGGGC	
RNU1-28P 1 Kb F	GAATTGCTTGAACCCGGGAG	Detects RNA levels 1 Kb downstream of the RNU1-28P TES
RNU1-28P 1 Kb R	AATGCACATTCGGACTCAGC	
RNU1-28P 2 Kb F	TCCCTTCACCTGCTTCAAGT	Detects RNA levels 2 Kb downstream of the RNU1-28P TES
RNU1-28P 2 Kb R	AATCTACACCGGGCTGCATA	
RNU1-28P 2.5 Kb F	TTTCACCGTGTCATCCAGGA	Detects RNA levels 2.5 Kb downstream of the RNU1-28P TES
RNU1-28P 2.5 Kb R	GGGTGACAGCGAGACTTAGT	
RNU1-28P 3 Kb F	GCGGTGCAGGGTTATCTTTT	Detects RNA levels 3 Kb downstream of the RNU1-28P TES
RNU1-28P 3 Kb R	CCCCTGTTGTTCCAGCTACT	

RNU1-1 500 bp F	TCTCTGGGAAGAAAGCAGGG	Detects RNA levels 500 bp downstream of the RNU1-1 TES
RNU1-1 500 bp R	ACGGCAGGAGATAGTAGGGA	
RNU1-1 1 Kb F	GGTTTTGTCCCTGCACTACA	Detects RNA levels 1 Kb downstream of the RNU1-1 TES
RNU1-1 1 Kb R	AGGCTGGTCTTGAACTCCTG	
RNU1-1 2 Kb F	TCTCTGTTGGGTCGTGTTGA	Detects RNA levels 2 Kb downstream of the RNU1-1 TES
RNU1-1 2 Kb R	GCCACTCTTGCAGATATTGACA	
RNU1-1 3 Kb F	CACCACGCCCAGCTAATTTT	Detects RNA levels 3 Kb downstream of the RNU1-1 TES
RNU1-1 3 Kb R	TCAAGCATAAGGAGCCTGGG	
RNU4-2 500 bp F	ACACTATGTTGGGAACTGGGT	Detects RNA levels 500 bp downstream of the RNU4-2 TES
RNU4-2 500 bp R	GGAAACAGCGAAAACTCCGT	
RNU4-2 1 Kb F	CACTACACCAGCCTCTTCCA	Detects RNA levels 1 Kb downstream of the RNU4-2 TES
RNU4-2 1 Kb R	TTTTCCCAGCACCGTCTTTG	
RNU4-2 2 Kb F	ACTGCAATCTCCACTTCCCA	Detects RNA levels 2 Kb downstream of the RNU4-2 TES
RNU4-2 2 Kb R	TGAGCCCAGGAGTTTGAGAC	
RNU4-2 3 Kb F	TATTGGTCAGGCTGGTCTCG	Detects RNA levels 3 Kb downstream of the RNU4-2 TES
RNU4-2 3Kb R	AACCTTCTCCAGCTGTCCTC	
RNU5A-1 precursor F	CTGGTTTCTCTTCAGATCGCA	Detects levels of RNU5A-1 precusor RNA
RNU5A-1 precursor R	CAGAATCTGCTAGTCACTGCT	
RNU4-1 precursor F	CCAATACCCCGCCGTGAC	Detects levels of RNU4-1 precursor RNA
RNU4-1 precursor R	TGCGAACAAGTACTCTTCAACC	
RNU1-1 precursor F	TCCATTGCACTCCGGATGT	Detects levels of RNU1-1 precursor RNA
RNU1-1 precursor R	ACCAACCAAGACACAAACCA	
INTS1 spliced F	CCTCATGTACCTGGCCAAGA	

INTS1 spliced R	CATGAGGAGGTTACAGGCCA	Detects levels of spliced INTS1 mRNA.

Table 2.17 Primers to detect RDHs and RNA levels downstream of their TES

Primer name	Sequence	Description
HIST1H4H gene body F	GTTTGGGTAAGGGAGGAGCT	Primers to detect levels of HIST1H4H
HIST1H4H gene body R	TCAGAACACCACGAGTCTCC	
HIST1H4H uncleaved F	GACGCACTCTTTACGGCTTC	Primers to detect uncleaved transcripts of HIST1H4H
HIST1H4H uncleaved R	GCCCAAATCCTAAACATGCG	
HIST1H4H 150 bp F	TTACTCGTGCTTAATCTCGCA	Primers to detect RNA levels 150 bp downstream of the HIST1H4H TES
HIST1H4H 150 bp R	TGTCACAATCCAGCTTACTCAC	
HIST1H4H 600 bp F	CTACAAAAGGCAGTGTGGGG	Primers to detect RNA levels 600 bp downstream of the HIST1H4H TES
HIST1H4H 600 bp R	CAGCCTGGATGAAAGAGCAA	
HIST1H4H 1 Kb F	TCCCAAGTGACTACAGGCTC	Primers to detect RNA levels 1 Kb downstream of the HIST1H4H TES
HIST1H4H 1 Kb R	CACGCCTGTAATCCCAACAC	
HIST1H4H 2 Kb F	TAGGGTCTTGCTCTGTTGCC	Primers to detect RNA levels 2 Kb downstream of the HIST1H4H TES
HIST1H4H 2 Kb R	GGACCAGCCTAACCCCATAA	
HIST1H3B gene body F	GGCTCGTACTAAACAGACAGC	Primers to detect levels of HIST1H3B
HIST1H3B gene body R	AGCAACTCGGTCGACTTTTG	
HIST1H3B uncleaved F	AGGGCTCTTTGAGGACACAA	Primers to detect uncleaved transcripts of HIST1H3B
HIST1H3B uncleaved R	AGTGGGTGGCTCTGAAAAGA	
HIST1H3B 150 bp F	TCTTTTCAGAGCCACCCACT	Primers to detect RNA levels 150 bp downstream of the HIST1H3B TES
HIST1H3B 150 bp R	GCAAGACTGACCAAACCGTT	
HIST1H3B 300 bp F	AACGGTTTGGTCAGTCTTGC	

HIST1H3B 300 bp R	TGCCTAGTAAGCGCCAGTTA	Primers to detect RNA levels 300 bp downstream of the HIST1H3B TES
HIST1H3B 2 KB F	ATGCTCTGCTTGTACCAGGT	Primers to detect RNA levels 2 Kb downstream of the HIST1H4H TES
HIST1H3B 2 Kb R	GAGAGGCAATTGTGGGAAAGT	
HIST1H3B 3Kb F	AGTCTCTTCTCATGCCTCGT	Primers to detect RNA levels 3 Kb downstream of the HIST1H4H TES
HIST1H3B 3Kb R	GGATGGGAGTGGAGTTTTGC	

3. Results Chapter 1 : The role of DIS3 in the nucleus of human cells

DIS3, also known as RRP44, is a major component of the nuclear exosome complex which has a vital role in the processing and degradation of a broad range of RNA transcripts (Allmang et al, 1999; Mitchell et al, 2014). As previously mentioned, DIS3 has two catalytically active domains, a RNB and a PIN domain, which constitute the 3' – 5' exonuclease and endonuclease activities of DIS3 respectively. DIS3 is able to act independently or as part of the exosome, where it associates with the exosome EXO-9 core structure at the exit pore (Lebreton et al, 2008; Schneider et al, 2009; Schaeffer et al, 2009; Bonneau et al, 2009; Lorentzen et al, 2008; Gerlach et al, 2018). At the opposing entry pore end of EXO-9 is where the other 3' – 5' exonuclease of the exosome resides, EXOSC10. Degradation of substrates by DIS3 is facilitated by: EXOSC10 mediated threading of transcripts into the central channel allowing them to reach the active site of DIS3; MTR4 helicase unwinding of RNA; and potentially DIS3 endonuclease function resolving complex secondary structures by cleavage, providing alternative 3' ends for DIS3 and EXOSC10 (Wasmuth et al, 2014; Zinder et al, 2016; Falk et al, 2017; Lebreton et al, 2008).

Whether DIS3 and EXOSC10 have their own specific substrates or work together is currently unclear, although it has been proposed that DIS3 provides the main catalytic activity of the exosome (Januszyk et al, 2011; Dziembowski et al, 2007). Through studies using either catalytically dead DIS3 or depleted DIS3 levels, some clear DIS3 substrates have been elucidated. Upon DIS3 depletion, there is an accumulation of short transcripts derived from promoter upstream regions, due to bidirectional transcription. These are known as PROMPTs and are only detectable upon exosome dysfunction (Preker et al, 2008; Preker et al, 2011). In addition, eRNAs, snoRNAs and prematurely terminated protein-coding transcripts have been suggested as DIS3 substrates (Szczepinska et al, 2015).

Functions of the exosome have largely either been revealed in human cells with protein depletion by RNAi or through studies in yeast, due to the ease of generating gene knockout mutants. However, findings in yeast are not always translational to humans and RNAi methods are slow with indirect effects (Tomecki et al, 2010; Jackson et al, 2003; Boutros and Ahringer, 2008). Therefore, a

method producing rapid protein depletion would be beneficial to investigate the immediate effects of DIS3 loss. To this end and to further understand DIS3 function in human cells, I aimed to generate conditional DIS3 depletion cells using CRISPR/Cas9 and AID technologies. This allowed rapid depletion of DIS3 and through the use of RNA-Seq I was able to investigate direct substrates of DIS3.

3.1 Production of the *DIS3-AID* cell line

For rapid and reversible knockdown of DIS3 protein we used CRISPR/Cas9 to produce a *DIS3-AID* cell line in HCT116 cells. These cells are derived from human colon carcinoma and have a diploid karyotype, unlike other standard mammalian cell culture models (Haigis et al, 2002; Horii et al, 2015). This allows for easier selection of homozygous tagged cell populations using only two drug resistance markers, making them highly suitable for genome manipulation. In addition, HCT116 cells have a high efficiency and ease of plasmid transfection. Both alleles of *DIS3* were genetically modified by the addition of an AID tag at the 3' end. This approach allowed us to overcome some of the limitations of RNAi based protein depletion.

3.1.1 Plant specific TIR1 expression in HCT116 cells

For the auxin-degron system to function in human cells, our cell line required the expression of the plant specific TIR1 F-box protein. This would allow TIR1 to recognise an AID tagged protein and promote its ubiquitination through the SCF^{TIR1} complex and recruited E2 ubiquitin ligase, leading to degradation (Gray et al, 1999; Nishimura et al, 2009; Holland et al, 2012). For stable TIR1 integration into transcriptionally active loci of HCT116 cells, the sleeping beauty transposon system was exploited.

The sleeping beauty (SB) system utilises a "cut-and-paste" DNA transposon and a transposase. Transposition of a DNA transposon is the direct movement of DNA through transposase-mediated excision from a donor locus and reinsertion into the cell genome. DNA sequences are flanked by terminal inverted repeats (IR) which contain transposase binding sites (Ivics and Izsvak, 2015). Although there are several transposon delivery systems, the sleeping

beauty transposon was selected due to its' ability to integrate a transposon of up to 10 Kb in length and is less likely to integrate into genes than HIV- or AVV-based vectors (Izsvak et al, 2000; Izsvak and Ivics, 2004). In addition, the SB transposon system has been shown to produce transposon expression for a prolonged period of time and at an adequate level (Yant et al, 2000; Belur et al, 2003; Kowarz et al, 2015).

TIR1 under the control of an ON CMV promoter was expressed inside a SB transposon vector with flanking IR sites. For selection of transfected cells, the vector contained a blasticidin resistance gene. A single colony grown under drug selection was cultivated to produce HCT116 cells expressing TIR1, which from now on are referred to as *HCT116:TIR1* cells.

3.1.2 AID tagging of *DIS3* using CRISPR/Cas9 and HDR

To produce DIS3 protein tagged with AID, both *DIS3* alleles were targeted for CRISPR/Cas9. Firstly, two repair templates were generated containing the AID tag, a self-cleaving peptide (P2A), a drug resistance selection marker of either hygromycin or neomycin to ensure tagging of both alleles simultaneously, a SV40 PAS and flanking sequences homologous to the 3' ends of the endogenous *DIS3* gene (Figure 3.1). Secondly a gRNA plasmid was constructed with sequence homology to *DIS3* and containing Cas9. Both selectable marker constructs were integrated into the cell with the gRNA directing Cas9 specifically to the *DIS3* gene. Cas9 cleaved the *DIS3* gene resulting in a double-stranded break, which was then repaired using HDR and the repair templates. This resulted in integration of the AID tag with a P2A site, drug selection marker and SV40 PAS at the 3' end of *DIS3*. Homozygous tagged *DIS3* cell selection was aided by the diploid karyotype of HCT116 and obtained through drug selection. Transcription of the newly tagged *DIS3* generates a single mRNA transcript using the endogenous promoter and SV40 PAS. The resulting AID-tagged protein is released after self-cleavage at the P2A site which removes the drug selection marker. The resulting cell line is referred to as *DIS3-AID*.

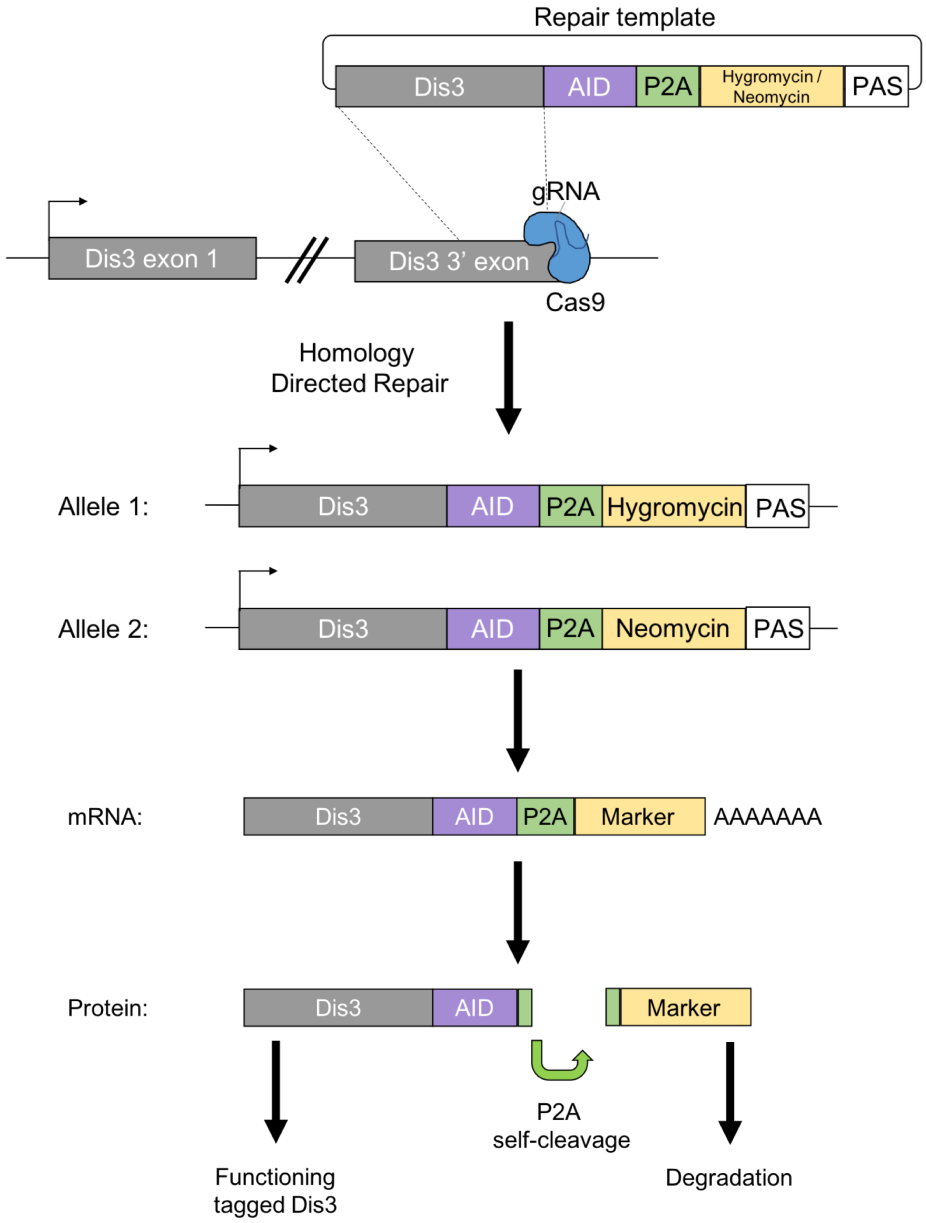

Figure 3.1 Generation of *DIS3-AID* using HDR and CRISPR/Cas9

A gRNA with homology to *DIS3* directs Cas9 to create a double-stranded break in the 3' end of the *DIS3* gene. The break is repaired by HDR using repair templates containing the AID tag, a P2A site, drug selection marker and SV40 PAS. Both alleles are altered and mRNA is produced using the endogenous promoter and a SV40 PAS downstream of the selection marker. Following translation into a protein, the P2A peptide self cleaves to produce two distinct proteins: the AID-tagged DIS3 and drug resistant protein.

3.1.3 Genomic PCR validation of *DIS3-AID*

After antibiotic selection of *DIS3-AID* cells, single colonies were isolated and allowed to grow. Homozygous integration was validated by a genomic DNA PCR screen. A nested-PCR approach was used with primers designed to flank the homology arms. All three clones investigated showed inclusion of *DIS3* homozygous modification compared to a control *HCT116:TIR1* cell line. This can be seen by the single large band at the expected size for tag incorporation in *DIS3* clones, compared to the single smaller PCR product only present in the control cell line (Figure 3.2). Overall this confirmed modification of both *DIS3* alleles and from this we decided to continue all further experiments with clone #1.

3.2 Conditional depletion of DIS3 by auxin addition

Following validation of the AID-tag being incorporated into both alleles of *DIS3*, *DIS3-AID* cells underwent western blot screening to determine if auxin treatment had an effect on DIS3 protein levels. An antibody binding to the C terminus of DIS3 was used in a western blot with *HCT116:TIR1* or *DIS3-AID* cells treated or not with auxin (Figure 3.3A). Endogenous DIS3 was easily detected in *HCT116:TIR1* cells and levels were unchanged by auxin addition. However, no DIS3 was detected in the *DIS3-AID* cell line. As the AID tag is present on the C terminus of DIS3, this absence of detection may have been due to the AID tag effecting the efficacy of antibody binding and leading to a false negative detection.

To overcome this issue a different antibody to DIS3 was used that recognises an internal amino acid sequence, thus allowing detection of both endogenous DIS3 and AID-tagged DIS3 (Figure 3.3B). An unmodified parent cell line, *HCT116:TIR1*, was used as a control and the endogenous DIS3 protein was detected at approximately 117 kDa. In the *DIS3-AID* cell line a larger DIS3 specific band was observed at approximately 150 kDa, suggesting incorporation of the AID-tag.

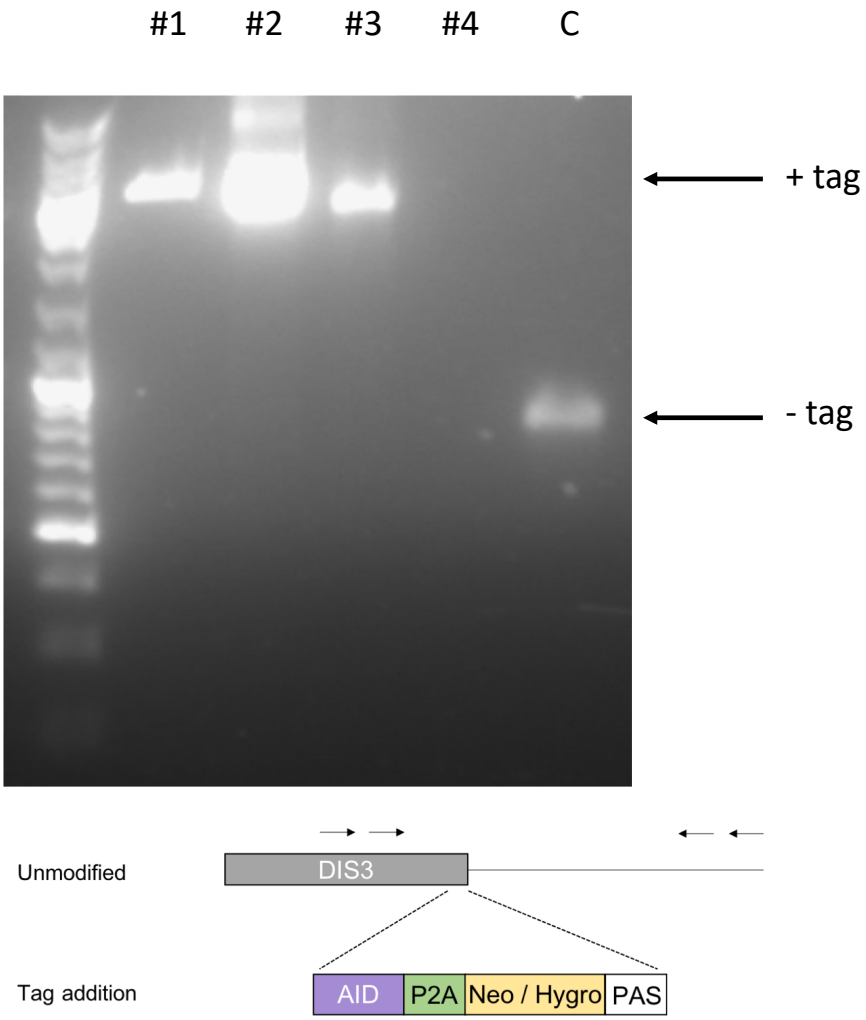

Figure 3.2 Genomic PCR validation of *DIS3-AID*

Nested PCR of genomic DNA obtained from 3 *DIS3-AID* clones after undergoing antibiotic selection (#1 - #3) and a control HCT116 cell line (C). Products were produced using primers designed outside the homology arm sequences as shown by the arrows. A single small endogenous *DIS3* band can be seen exclusively in the control cell line, whereas a larger band with predicted size of *DIS3* with tag incorporation is seen in all 3 *DIS3-AID* colonies. The endogenous band is not present in the *DIS3-AID* colonies suggesting both alleles were genetically modified. Clone #1 was taken forward for all future experiments.

This band was slightly bigger than the predicted size for AID-tag incorporation, however I believe it represents DIS3-AID protein as often modified proteins migrate during SDS-PAGE at rates inconsistent with their molecular mass, known as gel-shifting (Rath et al, 2009; Shi et al, 2012; Guan et al, 2015).

The lack of an endogenous DIS3 band in the *DIS3-AID* cells further supports the previous genomic PCR screen, suggesting a homozygous tagging of *DIS3*. Upon addition of auxin to cell media for 1 hour, the higher DIS3-AID band observed in the *DIS3-AID* cells is no longer detectable. However, auxin treatment of *HCT116:TIR1* cells had no effect. Therefore, auxin is able to specifically deplete AID-tagged DIS3 whilst having little / no effect on endogenous DIS3 protein levels. For protein depletion by auxin to occur, both TIR1 expression and inclusion of the AID-tag at the 3' end of the protein of interest is required. In addition, auxin treatment was conducted for 1 hour and was able to deplete tagged DIS3 levels to near complete absence. This shows that the rate of protein depletion following auxin treatment is rapid. Interestingly, comparing the two cell lines without auxin treatment there appeared be to less DIS3-AID protein expressed than endogenous DIS3. A qRT-PCR was conducted to investigate whether DIS3 mRNA levels were also altered (Figure 3.3C). In *DIS3-AID* cells there was a significant depletion of spliced DIS3 mRNA levels, probably caused by inclusion of the AID tag, that explains the reduced protein expression observed. An auxin treatment time course was not carried out, due to the near complete depletion of DIS3 protein levels at 1 hour. Longer auxin treatment would have increased the likelihood of confounding secondary / downstream effects and possible redundant pathway activation. Shorter auxin treatment times may not have been long enough for a strong DIS3 depletion, although this was not tested.

For further validation an antibody to detect the AID-tag was used in both control and *DIS3-AID* cells (Figure 3.3D). As expected, no detectable band was observed in the control cell line. However, the AID-tag could be readily observed in untreated *DIS3-AID* cells with a band of corresponding size absent upon 1 hour of auxin treatment. This further supports our findings that auxin conditionally depletes AID-tagged DIS3 in a time-effective manner.

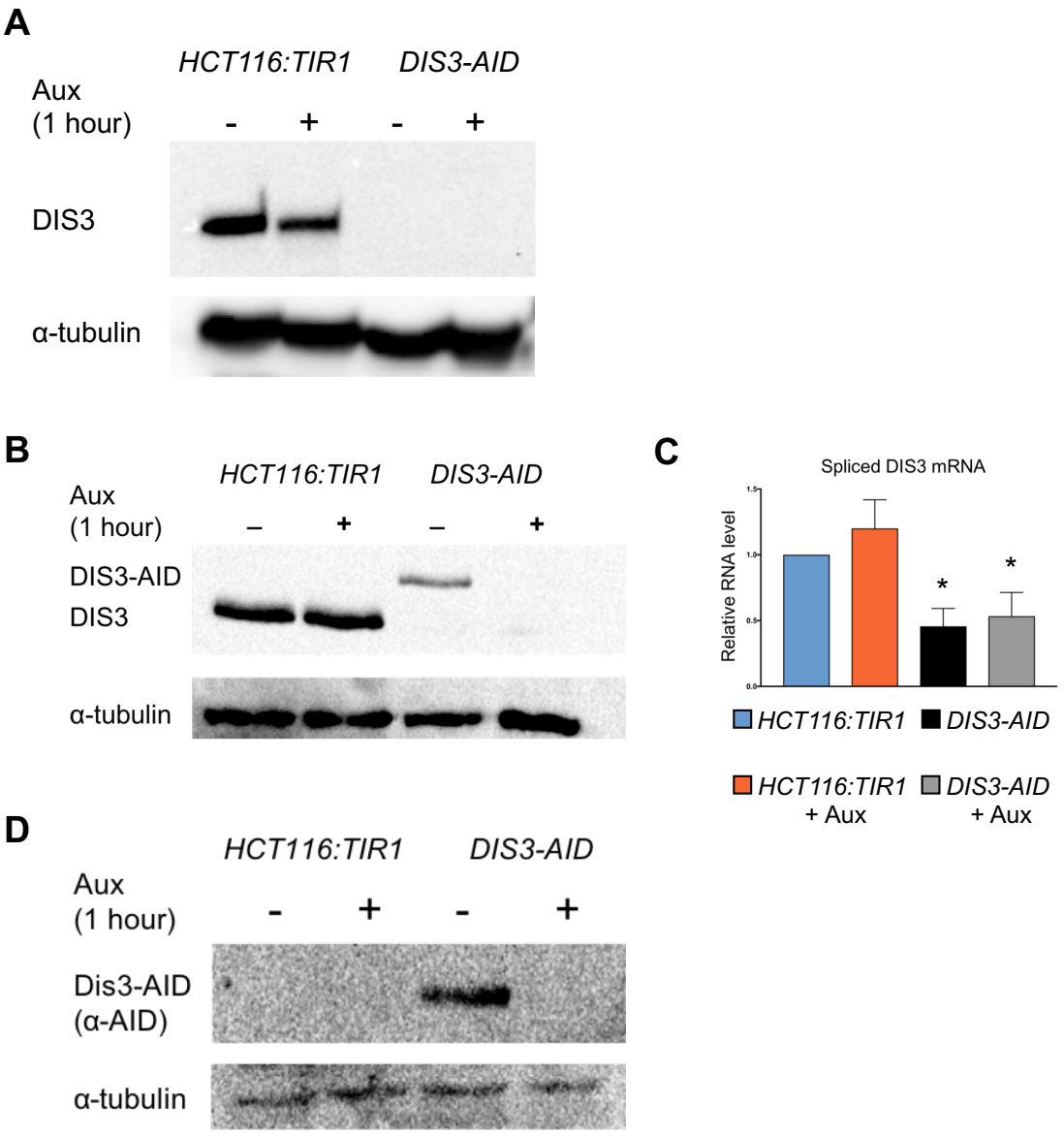

Figure 3.3 Western blots of endogenous DIS3, AID-tagged DIS3 and α-AID

A,B,D) Western blots showing the levels of specific proteins in control *HCT116:TIR1* cells and *DIS3-AID* cells that had either been untreated or treated with auxin for 1 hour. Anti-α-tubulin was used as a loading control. **A**) Antibody to the C terminus of DIS3 detected levels of endogenous DIS3 protein. **B**) Antibody to internal sequence of DIS3 protein used to detect endogenous and AID-tagged DIS3 protein. **C**) qRT-PCR detected levels of spliced DIS3 mRNA in *HCT116:TIR1* and *DIS3-AID* cells, treated or not with auxin. **D**) Anti-α-AID antibody used to detect the levels of the AID-tag. The AID-tag was detected in *DIS3-AID* cells only and upon auxin addition levels became undetectable.

3.2.1 Rapid depletion of DIS3 leads to accumulation of PROMPTs

Modification of *DIS3* by addition of the AID tag could potentially affect DIS3 function, by causing substrate recognition issues or reducing catalytic activity. To investigate these potential issues, the function of DIS3-AID protein and its depletion were tested using known substrates. As previously discussed, PROMPTs are well-characterised substrates of the exosome and more specifically DIS3 (Preker et al, 2008; Szczepinska et al, 2015). The levels of four different PROMPTs (STK11-IP, SERPINB8, FOXP4 and RBM39) were analysed by qRT-PCR (Figure 3.4A). These four PROMPTs were chosen specifically, as their accumulation upon DIS3 depletion had been previously shown (Preker et al, 2008). β-actin was used as a normalising gene for RT-qPCR and β-actin mRNA was shown to be stable in DIS3-AID cells upon auxin dependent DIS3 depletion by Steven West.

DIS3-AID and *HCT116:TIR1* cells were treated or not with auxin for 1 hour. Auxin had no effect on PROMPT levels in *HCT116:TIR1* cells. Additionally, untreated *DIS3-AID* cells showed similar PROMPT levels to controls. Therefore, AID modification of DIS3 does not impact on its ability to degrade PROMPTs. However, upon depletion of tagged DIS3 by auxin addition there is a significant accumulation of all four PROMPTs tested. These results show that PROMPTs are acutely sensitive to depletion of DIS3-AID.

3.2.2 Wild-type DIS3 is able to rescue auxin-dependent effects

As auxin dependent DIS3-AID depletion lead to a strong increase in PROMPT levels, I next investigated whether expression of wildtype (WT) DIS3 could rescue these effects. WT DIS3 was transfected into *DIS3-AID* cells using the SB system and PROMPT levels were detected by qRT-PCR (Figure 3.4B). Expression of PROMPTs is the same in untreated *DIS3-AID* and *DIS3-AID* cells transfected with WT DIS3. As previously shown, PROMPTs accumulate upon DIS3-AID depletion by auxin. This accumulation is rescued when WT DIS3 is expressed, with levels returning to the same as in untreated *DIS3-AID* cells. It is important to note here that there is no evidence that the WT DIS3 protein has been expressed, other than the observed rescue effect.

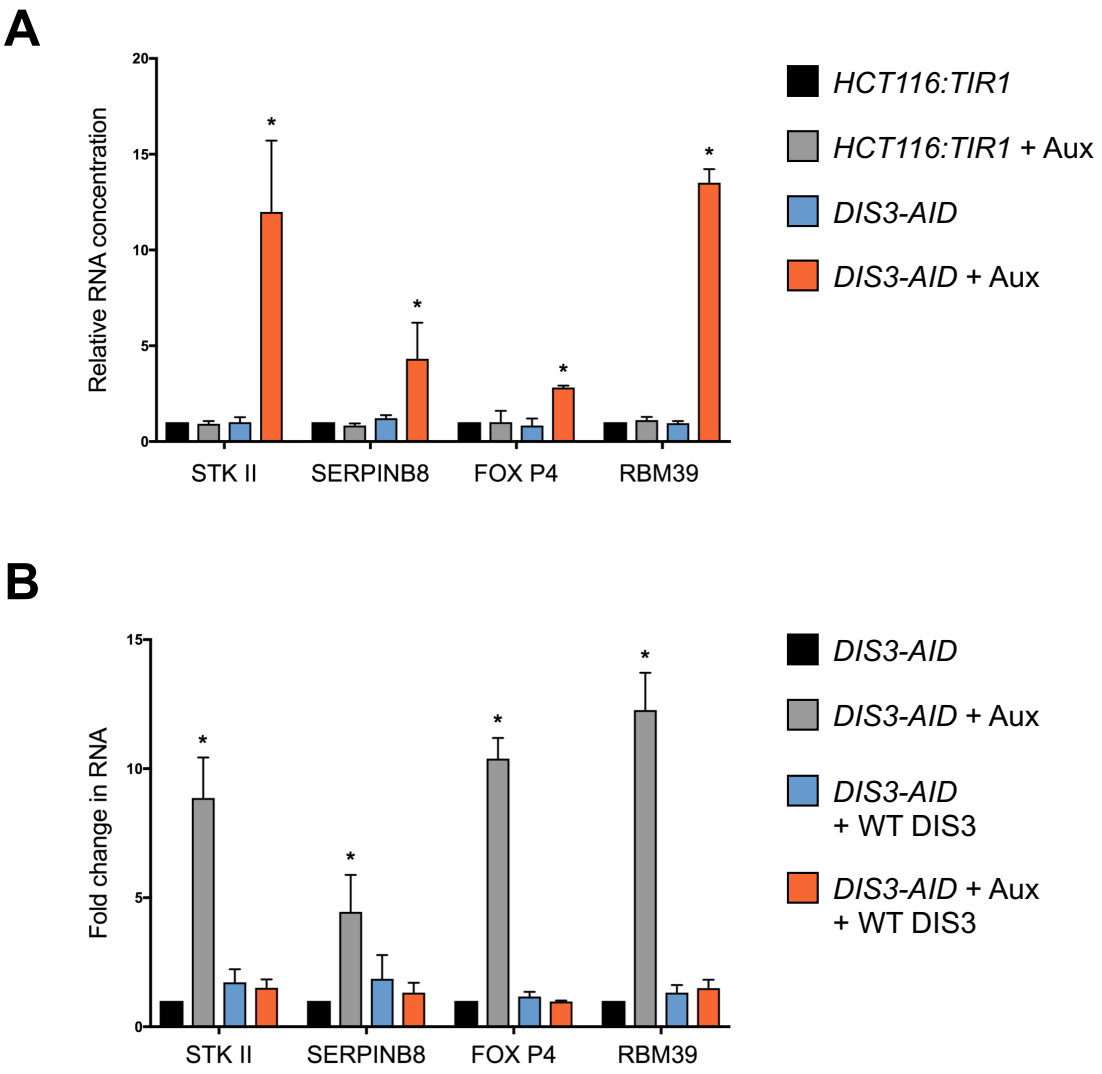

Figure 3.4 qRT-PCR of PROMPT levels in *DIS3-AID* cells

qRT-PCR detection of four PROMPTs (STK11-IP, SERPINB8, FOXP4 and RBM3). All levels were normalised to β actin, * denotes $p < 0.05$ and standard deviation is plotted by error bars. Data is the mean of three independent experiments with samples run in triplicate each time. **A**) PROMPT levels in *HCT116:TIR1* and *DIS3-AID* cells treated or not with auxin (Aux) for 1 hour. PROMPT levels remain the same except in *DIS3-AID* cells treated with auxin, where PROMPTs significantly accumulate. Quantitation is expressed as relative RNA level to untreated parental *HCT116:TIR1* cells. **B**) PROMPT levels in *DIS3-AID* and *DIS3-AID* cells transfected with WT DIS3 and treated or not with Aux for 1 hour. WT DIS3 is able to rescue PROMPT accumulation. Quantitation is expressed as fold change in RNA relative to untreated *DIS3-AID* cells.

However, overall this data suggests that it is the specific depletion of DIS3 causing accumulation of PROMPTs and expression of WT DIS3 is able to rescue the effects of DIS3 loss.

3.3 DIS3 is essential for cell viability

Previous studies have shown that DIS3 is essential for cell growth in yeast and for cell survival in a chicken DT40 cell line (Mitchell et al, 1997; Tomecki et al, 2014). To establish if the same was true for DIS3 in human cells, the AID system allowed investigation of DIS3 protein depletion on cell viability. A cell colony formation assay was conducted on *HCT116:TIR1* cells as a control and *DIS3-AID* cells, both in the presence and absence of auxin. After 10 days of growth in the presence of auxin there were no adverse effects on cell viability of control cells (Figure 3.5). *DIS3-AID* cells grown in the absence of auxin formed a similar number of cell colonies to controls, showing DIS3-AID does not impact cell viability. However, a slower growth phenotype was observed with smaller sized colonies. Treatment of *DIS3-AID* cells with auxin prevented colony formation, suggesting that DIS3 is essential for cell survival. Importantly, this lethality is specifically due to the loss of DIS3 as prolonged auxin treatment of *HCT116:TIR1* had no effect.

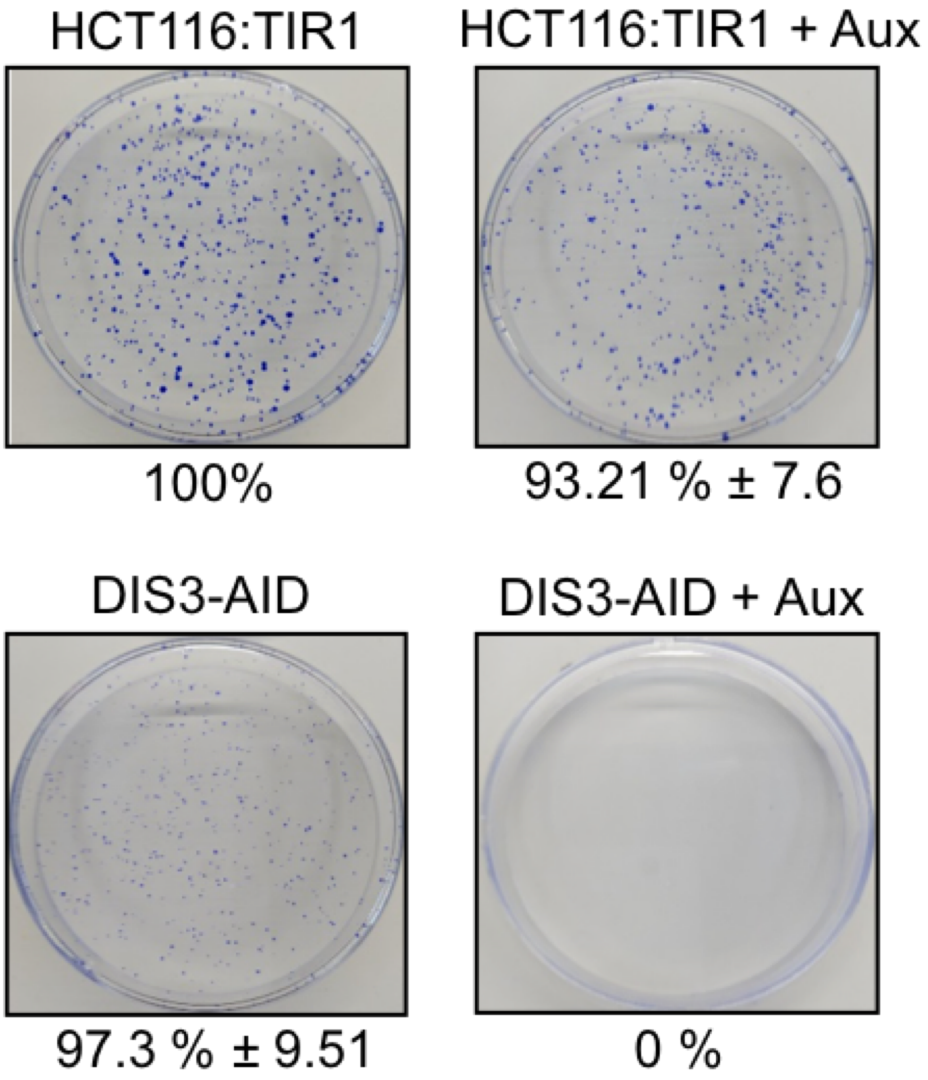

Figure 3.5 Cell colony formation assay of *DIS3-AID* and *HCT116:TIR1*

Approximately 300 cells of either *HCT116:TIR1* or *DIS3-AID* were seeded and grown in the presence or absence of auxin. After 10 days, cells were fixed and stained before counting using ImageJ software. No colonies grew upon DIS3 depletion by auxin. Number of colonies are expressed as a percentage of those grown from *HCT116:TIR1* cells without auxin. n = 3 and standard deviation is shown.

3.4 RNA-Seq investigation of DIS3 substrates

To elucidate direct substrates of DIS3 and the effects of DIS3 loss in human cells, a transcriptome-wide RNA-Seq analysis using single-end 50 bp reads was conducted. To generate RNA-Seq libraries, nascent nuclear RNA was extracted from *DIS3-AID* cells treated or not with 1 hour of auxin for DIS3 protein depletion. Reads were filtered and aligned to the genome, with the expression levels at each gene counted. This work was done in collaboration with Dr. Lee Davidson, who conducted the RNA-Seq and bioinformatic analyses of results. In addition, this and further work has been published in Davidson et al (2019).

3.4.1 Metagene profile of DIS3 loss shows stabilisation of PROMPTs

DIS3-dependent changes were first investigated by the production of a metagene read coverage profile. To do this, only annotated genes with an expression level higher than 50 reads per gene were used. The inclusion window was extended 3 Kb upstream of the TSS and 7 Kb downstream of the TES to ensure PROMPTs and other gene effects could be clearly visualised. Due to this extension, any overlapping genes were removed to decrease false-positive discovery of DIS3 loss effects. Therefore 4701 genes were included in the metagene plot, which represents the average transcription profile over these genes. The metagene figure only shows 3 Kb downstream of the TES for clarity (Figure 3.6 and 3.7).

From the metagene plot it can be seen that upon auxin induced DIS3 loss, there is an accumulation of reads before the TSS which is indicative of PROMPT accumulation (as shown by the red arrow in Figure 3.6). PROMPTs are transcribed in the opposing direction to their associated coding gene and are a result of bidirectional promoter transcription. Previous studies have found PROMPT transcription occurs up to 3 Kb upstream of the TSS (Flynn et al, 2011; Preker et al, 2008; Szczepinska et al, 2015). From the metagene plot, PROMPT expression peaked at 0.5 – 1 Kb upstream of the TSS and gradually decreased to near background levels at 3 Kb upstream. These findings correspond with the short length of PROMPTs and their termination proximal to the TSS, as well as further verifying loss of DIS3 function upon auxin addition. In addition, this widespread increase of PROMPTs validates their acute instability.

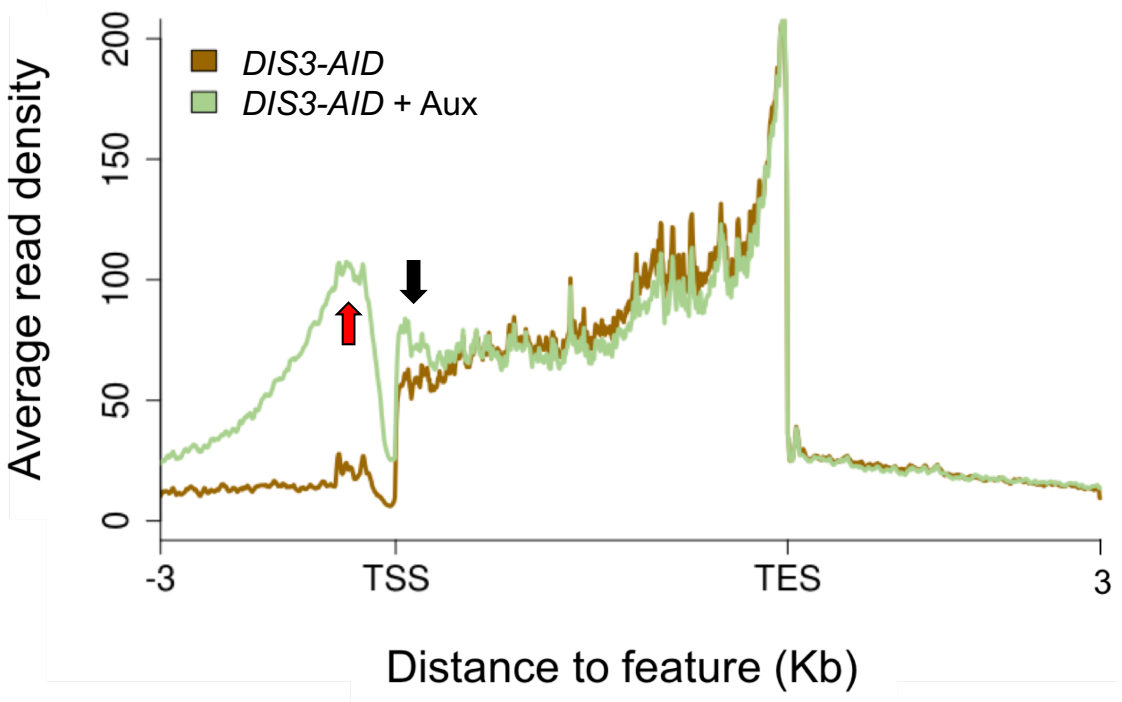

Figure 3.6 *DIS3-AID* metagene plot

Metagene plot profile of non-overlapping expressed protein coding genes in *DIS3-AID* cells with or without auxin treatment. The inclusion window is 3 Kb upstream of the TSS and 7 Kb downstream of the TES, with the gene body scaled to 5 Kb (n = 4701). The red arrow highlights the peak corresponding to accumulation of PROMPTs. The black arrow highlights a peak potentially corresponding to prematurely terminated transcripts. Profile is representative of 1 biological replicate; an additional replicate is shown in Figure 3.7. Produced from RNA-Seq analysis conducted by Lee Davidson.

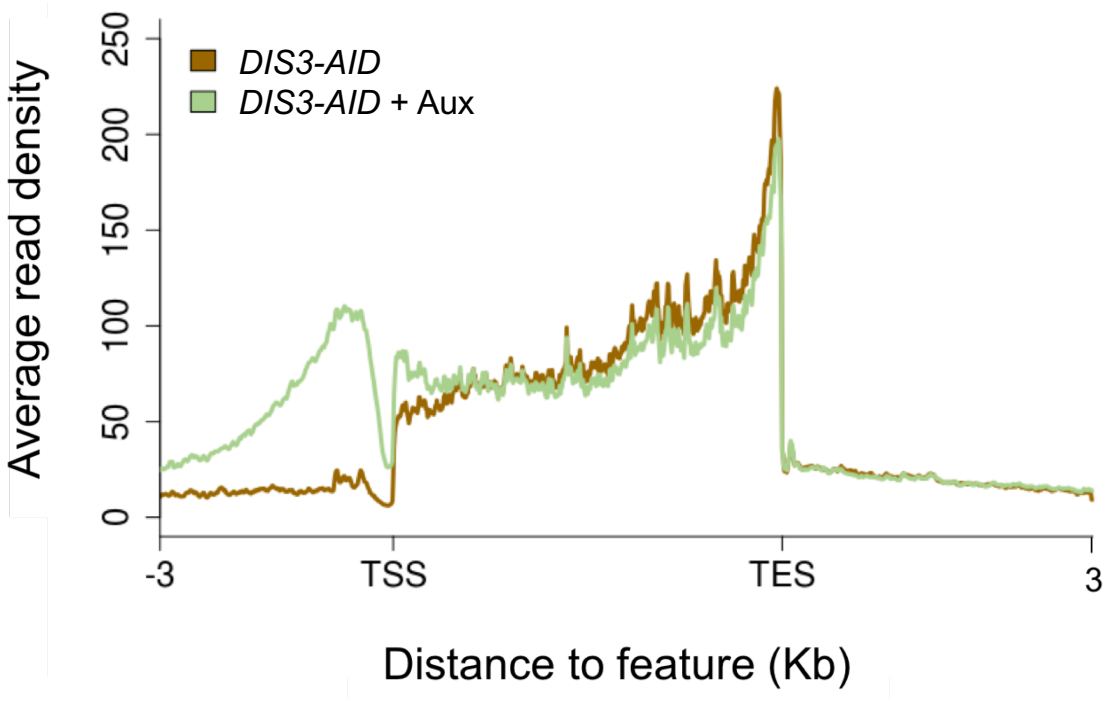

Figure 3.7 Second replicate of DIS3-AID metagene plot

Second biological replicate for metagene plot profile of non-overlapping expressed genes in *DIS3-AID* cells with or without auxin treatment. The inclusion window is 3 Kb upstream of the TSS and 7 Kb downstream of the TES, with the gene body scaled to 5 Kb (n = 4701). Produced from RNA-Seq analysis conducted by Lee Davidson.

Interestingly, DIS3 loss did not cause an observable effect on read density at the gene body or downstream of the TES. Previously it has been shown that PROMPT degradation enhances transcription of associated coding genes (Ntini et al, 2013), however an indicative increase in gene body reads was not observed in our data. It is possible that longer auxin treatment times, to further deplete DIS3 and / or prolong accumulation of PROMPTs through their enhanced stability, may result in an observable phenotype for downregulation of gene transcription.

3.4.2 DIS3 degrades prematurely terminated transcripts

In addition to the PROMPT accumulation peak upstream of the TSS, the metagene plot also showed a slight increase in read density immediately downstream of the TSS (as shown by the black arrow in Figure 3.6). It was hypothesised that this peak could be caused by an accumulation of RNA derived from premature transcription termination. To investigate this, three genes analysed in the metagene plot were chosen at random: C2orf42, NFU1 and PCBP1-AS1. Firstly, RPKM normalised coverage tracks were used to visualise these genes individually (Figure 3.8). For both C2orf42 and NFU1 a PROMPT transcript was observable before the TSS, on the sense strand (shown in red), that is detectable only upon DIS3 depletion. For PCBP1-AS1 the reads observable on the sense strand could be either from PROMPT transcription or transcription of another gene, PCBP1, as shown. The origin of these reads is difficult to determine, although the presence of these reads at a similar level when DIS3 is present suggests they correspond to PCBP1 transcription. Importantly, all three genes showed an increased number of reads near the TSS and over early intronic regions, as highlighted by the dashed box. This increase did not continue over the full length of the gene suggesting an accumulation of prematurely terminated transcripts occurs upon DIS3 depletion.

Secondly, qRT-PCR was used to validate these findings in the same three genes as above and in an additional gene, CLIP4. A relative RNA concentration was determined over the first intron of these genes, where promoter proximal transcripts might terminate, and over the exon – intron junction as a control (Figure 3.9).

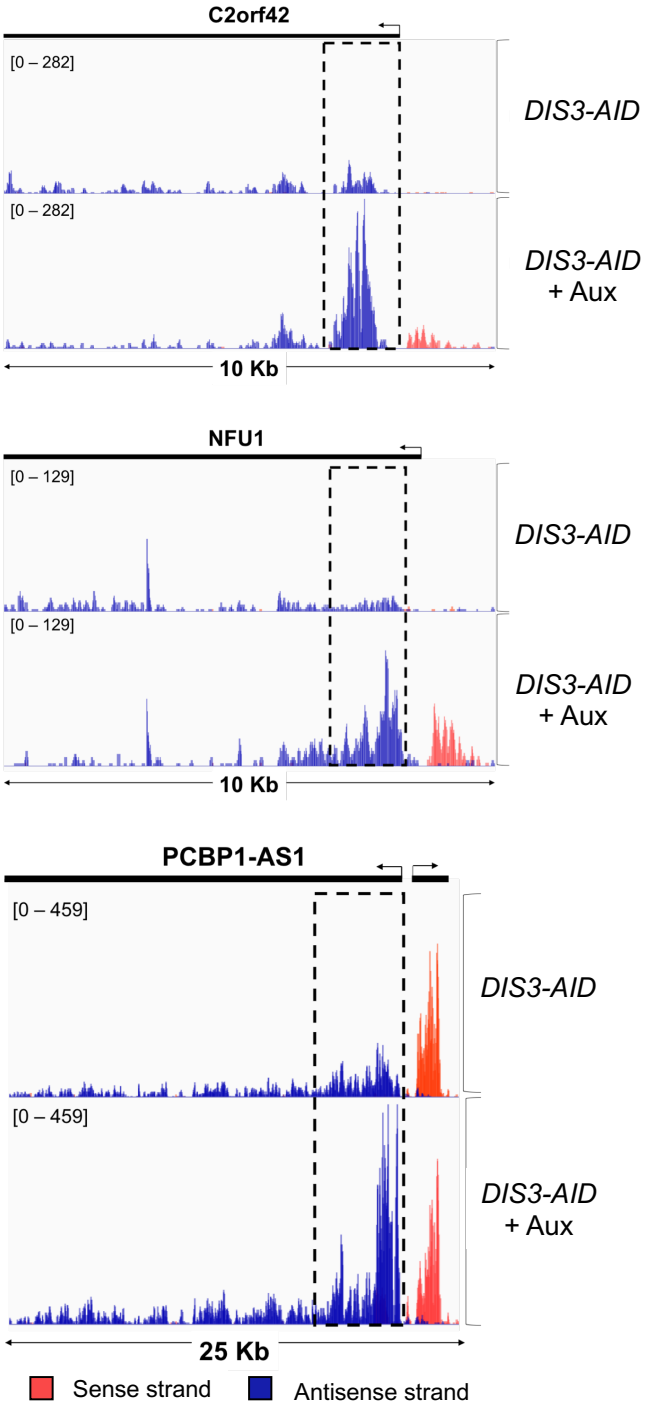

Figure 3.8 RPKM coverage tracks showing prematurely terminated transcripts in *DIS3-AID* cells

RPKM normalised coverage tracks of three protein-coding genes, C2orf42, PCBP1-AS1 and NFU1, in *DIS3-AID* cells treated or not with auxin. The dashed box shows an increase of reads near the TSS, corresponding to accumulation of prematurely terminated transcripts. For C2orf42 and NFU1 the sense reads shown in red correspond to PROMPTs. Sense and antisense strands are overlapped and the figure represents two replicates. The numbers in brackets show the RPKM normalised read count range.

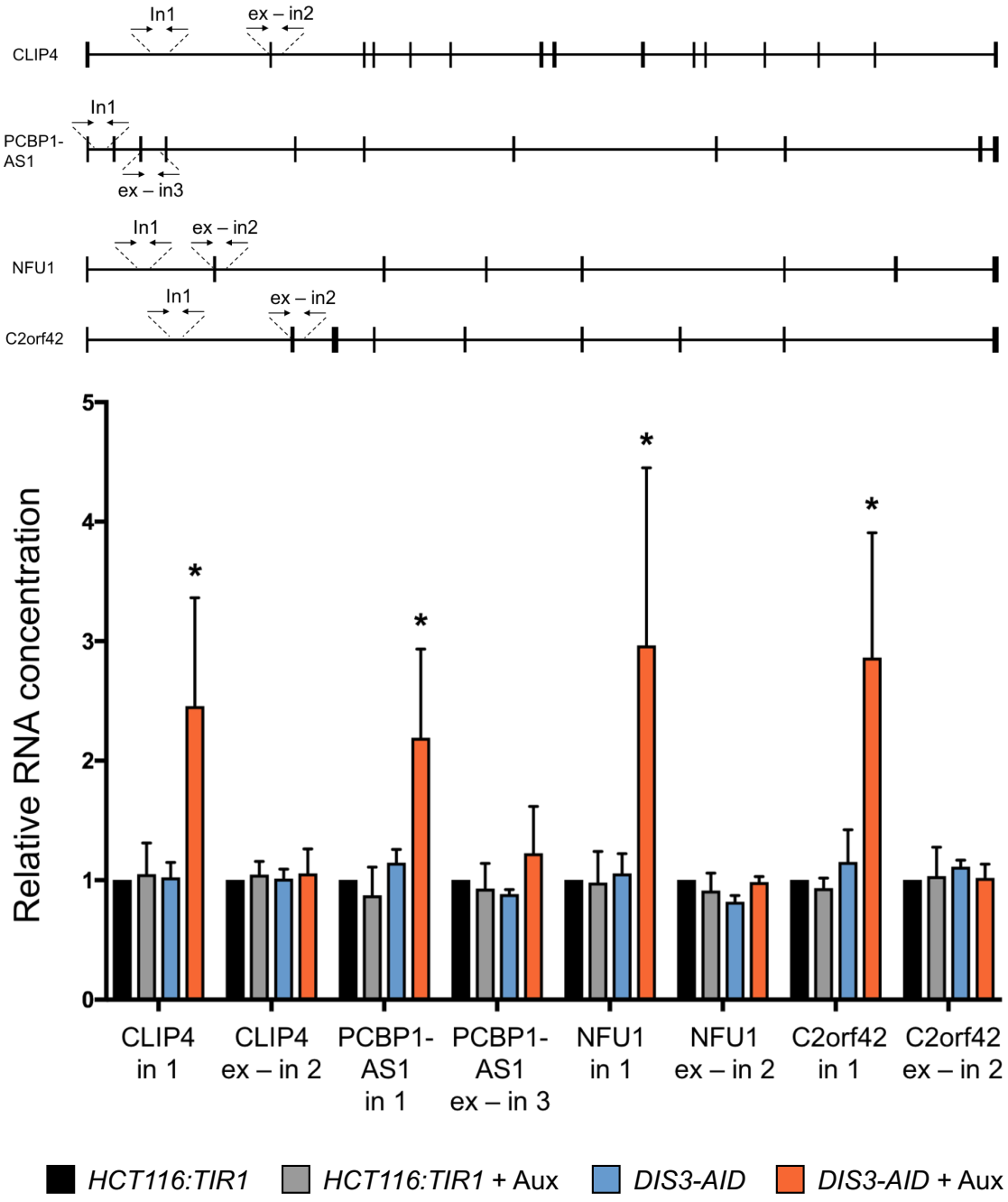

Figure 3.9 qRT-PCR investigating levels of prematurely terminated transcripts

In *HCT116:TIR1* and *DIS3-AID* cells treated or not with auxin for 1 hour, the levels of four transcripts were compared by qRT-PCR. In CLIP4, PCBP1-AS1, NFU1 and C2orf42 genes, qRT-PCR using primers spanning the first intron (in 1) were compared to primers spanning an exon – intron (ex – in) junction. Quantitation is expressed as relative RNA concentration to *HCT116:TIR1* cells. n = 3, * denotes $p < 0.05$, standard deviation is plotted as error bars. Data is the mean of three independent experiments with samples run in triplicate each time.

For each gene, RNA levels were not significantly different between control *HCT116:TIR1* cells, with or without auxin treatment, and untreated *DIS3-AID* cells. However, upon auxin-induced DIS3 loss there was a significant accumulation of RNA over intron 1. This increase was not observed over the exon – intron junction. Therefore, DIS3 loss caused accumulation of promoter proximal transcripts, explaining the peak near the 5' site on the earlier metagene plot and emphasising the role of DIS3 during multiple stages of transcription. Furthermore, this may suggest the exosome is recruited to promoter – proximal sites, possibly before transcription initiation and via previously discussed mechanics such as interactions between MTR4, NEXT complex and ARS2, to rapidly degrade pervasive or abortive transcripts. The rapid accumulation of these transcripts within 1 hour of auxin treatment suggests a large number of genes frequently undergo premature termination and DIS3 normally aids in degradation of these transcripts.

As no differences in RNA levels were seen over the gene body after DIS3 loss, this corroborates that these transcripts have arisen through premature termination. An alternative suggestion is that upregulated early intronic regions may be stabilised in some genes by readthrough from nearby PROMPT transcription. If intron 1 levels accumulate due to overlapping upstream PROMPTs, it could be assumed that PROMPTs from downstream neighbouring genes could overlap the TES of genes. However, no changes are observed over the TES as shown in Figure 3.6, suggesting our hypothesis of the described transcripts arising from premature transcription is more likely. This data also corresponds with Szczepinska et al (2015), who showed Dis3 was highly involved in degrading prematurely terminated protein-coding transcripts.

3.4.3 DIS3 depletion causes increased levels of unannotated genes

Szczepinska et al (2015) also found that catalytically dead DIS3 caused accumulation of RNAs originating from unannotated genomic regions and known enhancer RNAs (eRNAs). Enhancer elements are highly conserved sequences that bind to transcription factors and enhance gene transcription (Lee et al, 2015; Banerji et al, 1981). In 2010 it was discovered that Pol II transcription can occur at these enhancer elements, to produce eRNAs that play a role in gene

transcription regulation (Kim et al, 2010b; De Santa et al, 2010). eRNAs are short transcripts (< 2 Kb) that undergo rapid turnover by the exosome and arise from bidirectional promoters (Andersson et al, 2014).

Through RNA-Seq we were able to detect *de novo* transcripts that accumulated upon DIS3 loss and did not overlap with any known transcripts. Potential PROMPT transcripts were removed from this list by their proximity to annotated genes and short length, leaving 960 transcripts originating > 3 Kb from an annotated gene and aligning to distal intergenic regions. From visualisation of a multitude of these transcripts by Genome Viewer we established that each transcript consisted of two separate RNAs from the same bidirectional promoter-like region but on opposite strands (Figure 3.10A). Between the opposing transcripts there is a clear separation consistent with the presence of a nucleosome separating a promoter boundary (Andersson et al, 2014). Therefore, these *de novo* transcripts arise from regions where bidirectional transcription occurs, similar to enhancer sequences.

To ensure these transcripts were not artefacts of RNA-Seq, qRT-PCR validation was performed for three of the *de novo* transcripts chosen at random (Figure 3.10B). Primers were designed upstream and downstream of the region between the sense and antisense transcripts. Relative RNA concentration were similar in *HCT116:TIR1* cells with or without auxin and untreated *DIS3-AID* cells. Upon DIS3 loss there was a significant accumulation of RNA in both the upstream and downstream regions in all cases, corresponding to the *de novo* transcripts in both directions. Therefore, with the DIS3-AID system we were able to detect novel transcripts of DIS3 originating from genomic regions that are normally rapidly degraded.

The novel transcripts we identified could potentially be uncharacterised eRNAs or derive from spurious transcription from open chromatin loci. Further work is needed to elucidate their true characterisation. Interestingly, the use of the DIS3-AID system was able to detect more eRNAs and novel transcripts than previous RNAi experiments (Szczepinska et al, 2015). These differences could be due to the use of different cell lines, however the AID system may be able to enhance detection of unstable RNAs.

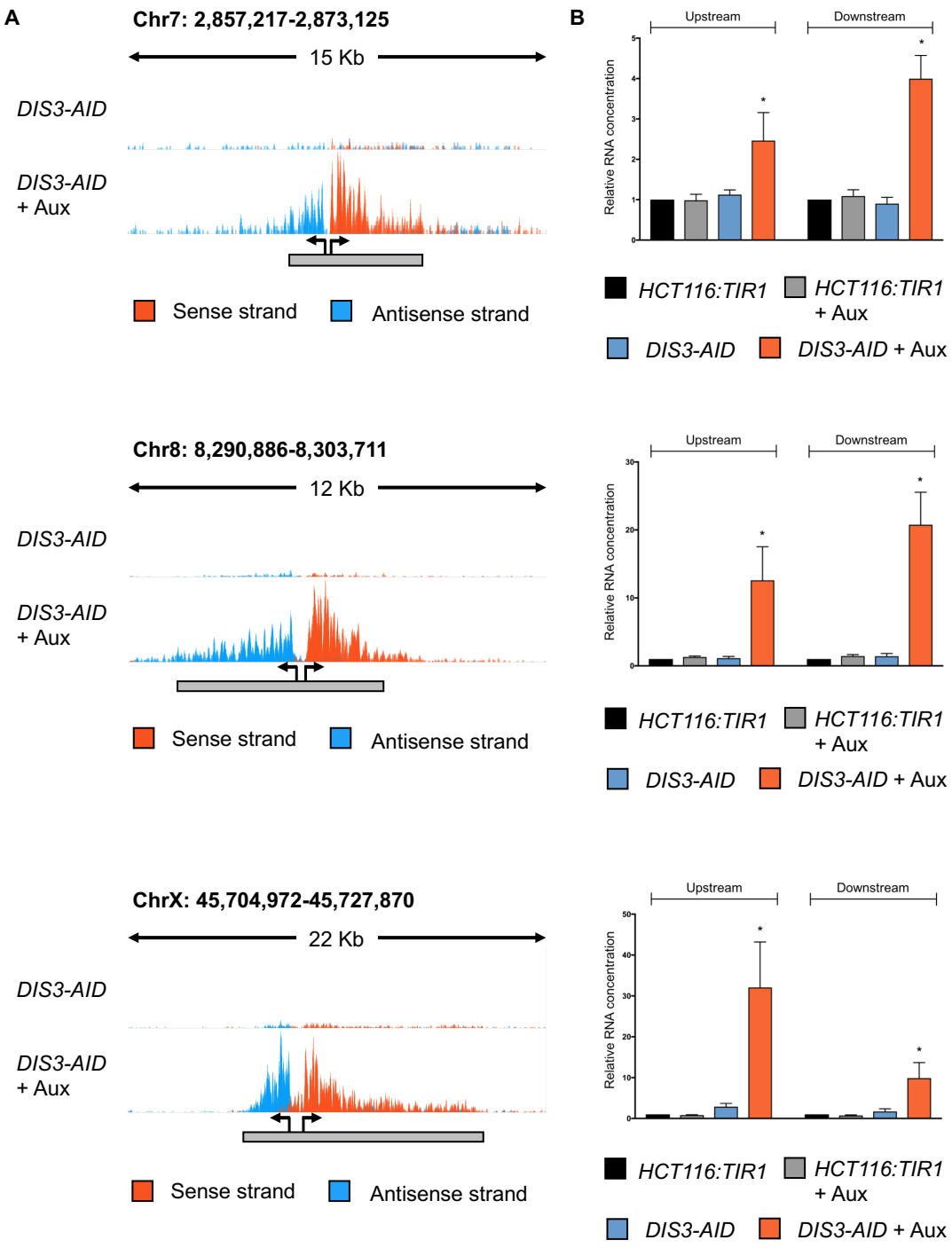

Figure 3.10 DIS3 depletion effects levels of *de novo* transcripts

A) RPKM normalised coverage tracks of three *de novo* transcripts from RNA-seq analysis of *DIS3-AID* cells with or without auxin. Transcripts were detected over intergenic intervals. **B**) qRT-PCRs of the same *de novo* transcripts with primers designed upstream and downstream of the region separating the opposing transcripts. Conducted in *HCT116:TIR1* and *DIS3-AID* cells treated or not with auxin for 1 hour. Quantitation is relative RNA concentration to untreated *HCT116:TIR1* cells. n = 3, * denotes $p < 0.05$, error bars plot standard deviation.

3.5 Summary

In this chapter I have shown that a protein of interest can be rapidly and significantly degraded using the AID system (Figure 3.3). Fusion of an AID tag to the 3' end of DIS3 using CRISPR/Cas9 is relatively simple and easily reproducible. The AID system causes conditional protein degradation with the necessary expression of a plant specific TIR1 F-box protein, by utilising human proteasome mediated degradation pathways. This inducible degradation is controlled by the addition of auxin and can be reversed by removal of auxin from growth media.

The findings described in this chapter are largely consistent with data from previous studies using RNAi techniques to indirectly deplete exosome subunits by targeting mRNA (Szczepinska et al, 2015). However, the AID system causes protein depletion in a rapid manner, allowing easier investigation of the effects of immediate protein loss in a shorter time frame than RNAi. In addition, the AID system uncovered more PROMPT and other DIS3 substrate changes than typically reported, suggesting a more complete protein depletion by the AID system may be beneficial for studying RNA turnover. From our results gene fusion with the AID tag, expression of TIR1 or untagged cell growth in the presence of auxin has few deleterious effects on HCT116 cell function although *DIS3-AID* cells showed a slight defect in growth rate and reduced levels of spliced DIS3 mRNA (Figure 3.5 and Figure 3.3C). A potential reason that untreated *DIS3-AID* cells showed a slow growth phenotype might be due to the levels of tagged DIS3 present. Results from the western blot suggested a reduction in DIS3-AID levels compared to endogenous DIS3 in *HCT116:TIR1* cells (Figure 3.3B). A qRT-PCR was conducted to investigate this further and found DIS3-AID mRNA was present at approximately 50 % of the level of endogenous DIS3 mRNA (Figure 3.3C). Therefore, the AID tag may cause a decrease in DIS3-AID transcription or increase in its degradation and this reduction might explain why growth of *DIS3-AID* colonies is slower than that of *HCT116:TIR1* cells.

To further prevent AID tag effects and improve protein stability a smaller mini-AID tag can be used in place of the larger AID tag (Natsume et al, 2016). A more recently described method involves expression of the auxin response transcription factor (ARF), that in the absence of auxin binds to the AID in plants (Sathyan et al, 2019). Expressing ARF in human cells utilising the AID system

resulted in a decrease of AID tag effects. Additionally, I have shown that DIS3 is vital for cell survival with auxin-conditional depletion of DIS3 causing 100% cell death (Figure 3.5). This finding is analogous with previous reports by Mitchell et al (1997).

PROMPTs originating from bi-directional promoters of protein-coding genes were detectable after only 60 minutes of DIS3 depletion by auxin addition. As previously described, these PROMPTs were transcribed in the opposing direction to the protein-coding gene (Preker et al, 2008; Flynn et al, 2011). From Figure 3.6, PROMPT transcription was observed upstream of the protein-coding gene TSS with a decrease in reads occurring within 3 Kb upstream. This suggests PROMPT transcription is not finite and transcription termination happens at approximately 3 Kb of length. PROMPT termination may still occur by conventional cleavage at the PAS, as poly(A) sites are more abundant upstream than downstream of the mRNA TSS, and PAS hexamers are located 10 – 30 nts upstream of PROMPT 3' ends (Ntini et a, 2013). Termination could also occur by PROMPT readthrough into a neighbouring gene and the use of that genes' PAS (Chen et al, 2016). Either way, this termination mechanism would provide a free 3' end allowing rapid degradation by DIS3 mediated pathways. DIS3 is vital for maintaining proper promoter directionality and preventing accumulation of redundant transcripts produced by bi-directional transcription.

In addition to PROMPTs, RNA-Seq data revealed a small increase in reads immediately downstream of the protein-coding gene TSS (Figure 3.6). Through further investigation I was able to determine this peak corresponded to an increase in reads over the early intronic regions of genes, suggesting an overall accumulation of prematurely terminated / abortive transcripts (Figure 3.8 and 3.9). Importantly, these findings support a role for DIS3 in multiple stages of transcription. Under normal conditions, it appears many genes frequently undergo premature termination and these abortive transcripts are rapidly degradation by DIS3, as shown by their accumulation within 60 minutes of DIS3 depletion. This degradation potentially occurs co-transcriptionally and through close association with the transcribing Pol II complex. Exosome recruitment to promoter-proximal sites may occur through MTR4, NEXT and ARS2 interactions as previously mentioned.

Finally, accumulation of *de novo* transcripts originating from unannotated genome regions were observed upon DIS3 depletion (Figure 3.10). Analysis revealed that these transcripts potentially originate from bidirectional promoter-like regions, similar to eRNAs. These *de novo* transcripts may be as yet unidentified enhancers, however further work is required for their characterisation. Overall, in this chapter I have been able to identify the importance of DIS3 in the degradation of a multitude of transcripts and its vital role in maintaining appropriate gene expression. In the following chapter I will assess the role of DIS3 and other endonucleases, including the Integrator, in the transcription of small nuclear RNAs (snRNAs).

4. Results Chapter 2: Endonuclease function in snRNA transcription and processing

Pol II transcribes a number of non-coding transcripts including snRNAs. snRNAs are commonly 60 – 200 nts in length, are not polyadenylated and lack introns. They have a primary function in splicing of mRNA through formation of the spliceosome and U7 snRNA has a major role in the 3' end formation of RDH mRNA (Chen and Wagner, 2010). Processing of pre-snRNAs into mature snRNAs involves endonuclease activity, in particular from the Integrator complex. The Integrator consists of 12 – 14 subunits, with catalytic activity provided by the endonuclease subunit INTS11. INTS11 is a homolog of CPSF73 and forms a heterodimer with another Integrator subunit INTS9, which is thought to be necessary for snRNA 3' end processing (Dominski et al, 2005). INTS9 is similar to CPSF100, possessing an incomplete catalytic centre. Recent findings suggest INTS4 is also necessary for snRNA processing and may in fact form a heterotrimeric structure with INTS11 and INTS9 (Albrecht et al, 2018). Either way, the Integrator is able to recognise the 3' box consensus sequence located 9 – 19 nts downstream of the snRNA coding region and it is thought that INTS11 is then responsible for cleavage of snRNA near this site (Baillat et al, 2005).

As previously described, 3' end processing of Pol II protein-coding genes is tightly coupled to their transcription termination. Similarly, snRNA 3' end processing has been linked to efficient termination, however the actual mechanisms of snRNA termination are not fully understood (Ramamurthy et al, 1996; O'Reilly et al, 2014). Polyadenylation factors have also been suggested to play a role in snRNA termination (O'Reilly et al, 2014). In addition to snRNAs, the Integrator has been implicated in other aspects of transcriptional regulation including transcription initiation at protein-coding genes, termination of RDHs and Pol II pause-release (Gardini et al, 2014; Skaar et al, 2015; Stadelmayer et al, 2014).

To investigate snRNA transcription and termination, as well as the further characterisation of Integrator function we utilised RNA-Seq methods in a number of cell models allowing conditional depletion of specific endonuclease proteins including INTS11, CPSF73 and as discussed in the previous chapter, DIS3.

4.1 Production of the Ints11-SMASh cell line

To investigate the role of INTS11 we aimed to generate an inducible INTS11 knockdown cell line, similar to the production of the *DIS3-AID* cells. However, the last 10 amino acids of INTS11 are necessary for its interaction with INTS9. Abrogation of this heterodimer formation has effects equivalent to mutations disrupting the active site of INTS11, including interference of Integrator function (Wu et al, 2017). Therefore, the addition of an AID-tag to the C terminus of INTS11 might disrupt INTS11 function, even without addition of auxin. To overcome this issue, we decided to utilise the SMASh-tag system, which contains a NS3 protease, HCV NS3 recognition site and a destabilising degron. Using CRISPR/Cas9 techniques as previously described (Figure 3.1), the SMASh-tag was genetically integrated at the C terminus of both *INTS11* alleles to produce *INTS11-SMASh* cells (Figure 4.1). The protease function of the SMASh-tag cleaves the NS3 recognition site under normal conditions, causing INTS11-SMASh to become untagged and allowing normal INTS11 function. Thus, solving the potential issues around INTS11 protein interactions if having used the AID-tag. Upon addition of a protease inhibiting drug, asunaprevir (asn), the SMASh-tag is no longer cleaved causing tagged INTS11 to be degraded due to SMASh-tag internal degron activity (Figure 1.5). Generation of the *INTS11-SMASh* cells was conducted by Steven West.

4.1.1 Genomic PCR validation of Ints11-SMASh

INTS11-SMASh colonies were grown under hygromycin and neomycin drugs to select for homozygous integration of the SMASh-tag. After selection, the cells were validated by genomic PCR with primers flanking the homology arms (Figure 4.2). *HCT116:TIR1* cells were used as a control and show a band at the expected endogenous *INTS11* size. Whereas the *INTS11-SMASh* cells show a much larger band of a size expected for SMASh-tag inclusion. As no other bands were observed for the *INTS11-SMASh* cells, it was concluded that both *INTS11* alleles had been successfully modified.

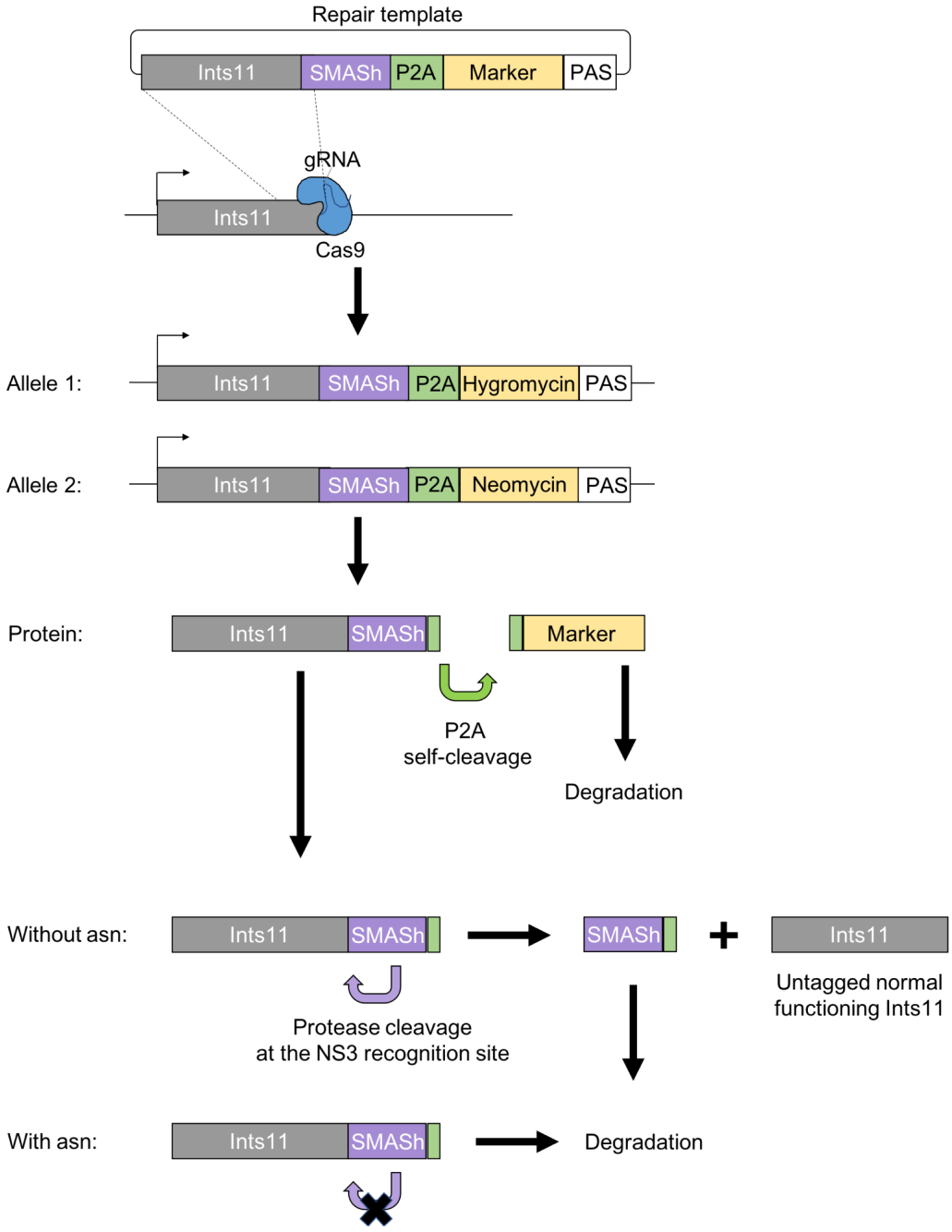

Figure 4.1 Generation of *INTS11-SMASh* using CRISPR/Cas9

INTS11 homologous gRNA directs Cas9 to create a double-stranded break in the 3' *INTS11* gene. The break is repaired using repair templates consisting of a SMASh-tag, P2A site, selection marker and SV40 PAS. After translation, the P2A peptide self cleaves to produce the SMASh-tagged INTS11 protein. Under normal conditions, the protease activity of the SMASh-tag cleaves at the NS3 recognition site to produce an untagged INTS11 protein capable of normal function. Upon addition of asunaprevir (asn), the protease activity is inhibited and the tagged INTS11 protein is targeted for degradation due to the internal degron activity of the SMASh-tag.

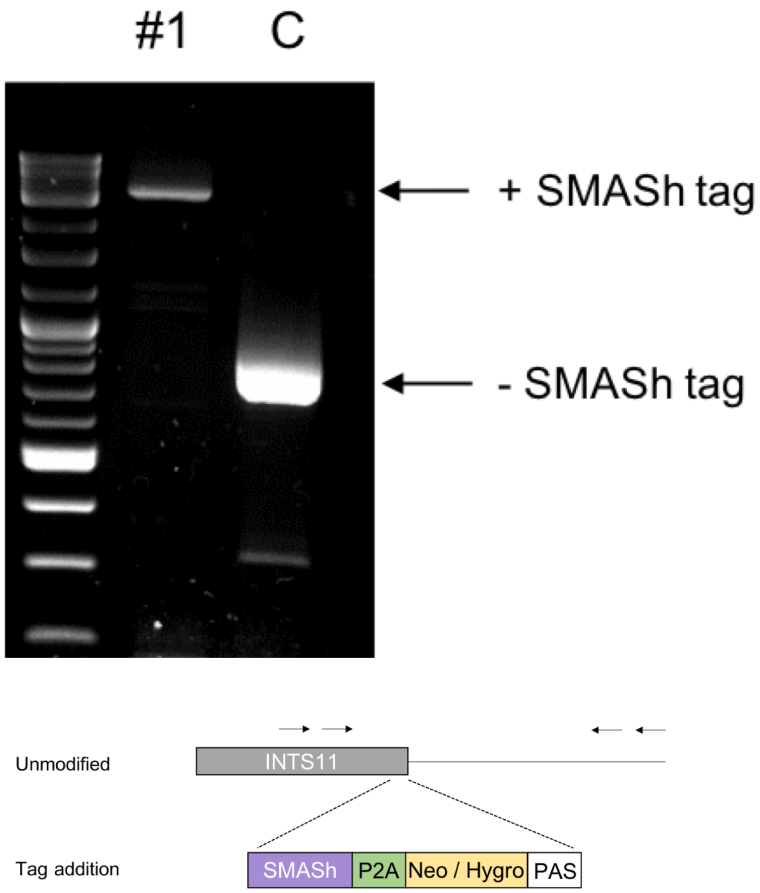

Figure 4.2 Genomic PCR validation of *INTS11-SMASh*

Nested genomic PCR in control HCT116 cells (C) and *INTS11-SMASh* cells grown under drug selection (#1) using primers flanking the *INTS11* homology arms, as shown in the diagram by the arrows. A strong band is present in the control cells corresponding to endogenous *INTS11*. In the *INTS11-SMASh* cells a much higher band is present at the predicted size for *INTS11* with SMASh-tag incorporation. No endogenous *INTS11* band is observed in the *INTS11-SMASh* cell line, suggesting both alleles were genetically modified. Figure from Steven West.

4.1.2 Conditional depletion of INTS11 by asunaprevir addition

Following validation of homozygous integration in *INTS11-SMASh* cells, I wanted to determine if addition of asunaprevir to cell media caused inducible INTS11 depletion. A western blot using an antibody to INTS11 was conducted for *HCT116:TIR1* cells as a control and *INTS11-SMASh* cells, both with and without 48 hours of asunaprevir treatment (Figure 4.3). An INTS11 specific band was detected at approximately 65 kDa and an antibody to alpha tubulin was used as a loading control. In *HCT116:TIR1* cells the levels of INTS11 did not alter upon asunaprevir addition, showing asunaprevir alone does not affect INTS11 protein levels. *INTS11-SMASh* cells without asunaprevir showed similar INTS11 protein levels to control cells. After 48 hours of drug treatment there was a near complete depletion of INTS11 protein. Therefore, inducible INTS11 protein depletion is capable upon addition of asunaprevir to *INTS11-SMASh* cells. Asunaprevir induced INTS11 depletion is not as rapid as AID protein depletion, i.e. 1 hour of auxin treatment significantly depletes DIS3 in the *DIS3-AID* cell line. The reason for this is asunaprevir treatment prevents protease cleavage at the NS3 recognition site of newly synthesised tagged INTS11 protein. This results in rapid degradation of tagged INTS11 protein, however untagged INTS11 that had been previously cleaved from the SMASh tag is still present. Therefore, the half-life of untagged INTS11 protein is important for complete degradation and explains the longer treatment times necessary for this methodology. As 48 hours of asunaprevir treatment produced a significant depletion of INTS11 protein levels, all further experiments were conducted for this length of time. Longer asunaprevir treatment time courses to produce further INTS11 depletion were not investigated due to the increased possibility of any observed depletion effects being due to secondary effects and not the immediate loss of INTS11. Although not shown here, shorter treatment times were analysed by western blot and did not produce as pronounced a decrease in INTS11 protein levels.

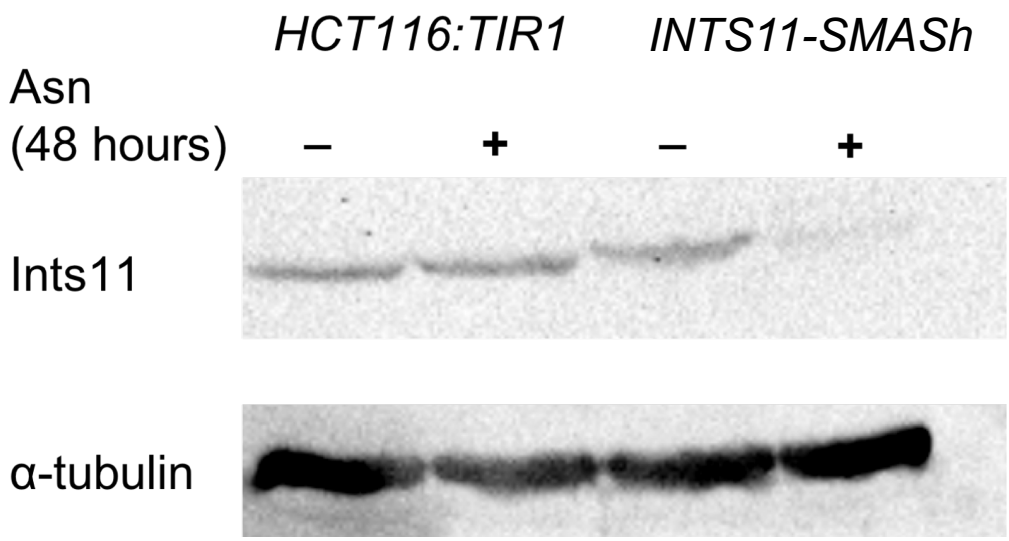

Figure 4.3 Western blot of INTS11

Western blot showing the levels of INTS11 protein in *HCT116:TIR1* and *INTS11-SMASh* cells, treated or not with asunaprevir for 48 hours. INTS11 levels are comparable between *HCT116:TIR1* cells in both conditions and *INTS11-SMASh* cells without asunaprevir. Asunaprevir significantly and specifically reduced INTS11 protein levels in *INTS11-SMASh* cells after 48 hours. Alpha tubulin was used as a loading control.

4.1.3 Depletion of INTS11 causes accumulation of extended snRNAs

Before further investigation into the endonuclease function of INTS11 by RNA-Seq, it was determined whether INTS11 depletion by asunaprevir caused an effect on snRNA processing. Previous studies found a coupling between transcription termination and Integrator cleavage of snRNAs, with disruption of one negatively affecting the other (Ramamurthy et al, 1996; O'Reilly et al, 2014). Therefore, RNA levels downstream of three snRNAs chosen at random, RNU5B-1, RNU5D-1 and RNU1-28P, were investigated using qRT-PCR (Figure 4.4). *HCT116:TIR1* cells were used as a control to *INTS11-SMASh* cells, treated or not with asunaprevir.

There were no significant differences in RNA levels between *HCT116:TIR1* cells with and without asunaprevir or untreated *INTS11-SMASh* cells (Figure 4.4). Upon depletion of INTS11, there is a significant accumulation of RNA immediately downstream of the TES of all three snRNAs (200 – 300 bp). As INTS11 would normally cleave snRNAs at their 3' end, this accumulation of misprocessed snRNA is likely due to INTS11 depletion. RNA levels are comparable to the control by 1.5 Kb - 3.5 Kb downstream of the snRNA TES, suggesting that readthrough of these snRNAs is not finite and that they still undergo transcriptional termination when processing is impaired. Overall, INTS11 depletion in *INTS11-SMASh* cells is sufficient for Integrator dysfunction as shown by the aberrant processing of snRNAs. Unfortunately, RNA levels downstream of these snRNAs could not be determined at other intervals, such as 500 bp, 1Kb etc, due to primer design issues with primer specificity. Therefore the locations downstream of the snRNAs analysed were determined by using only primers that had high specificity as shown through RT-qPCR melt curve analysis.

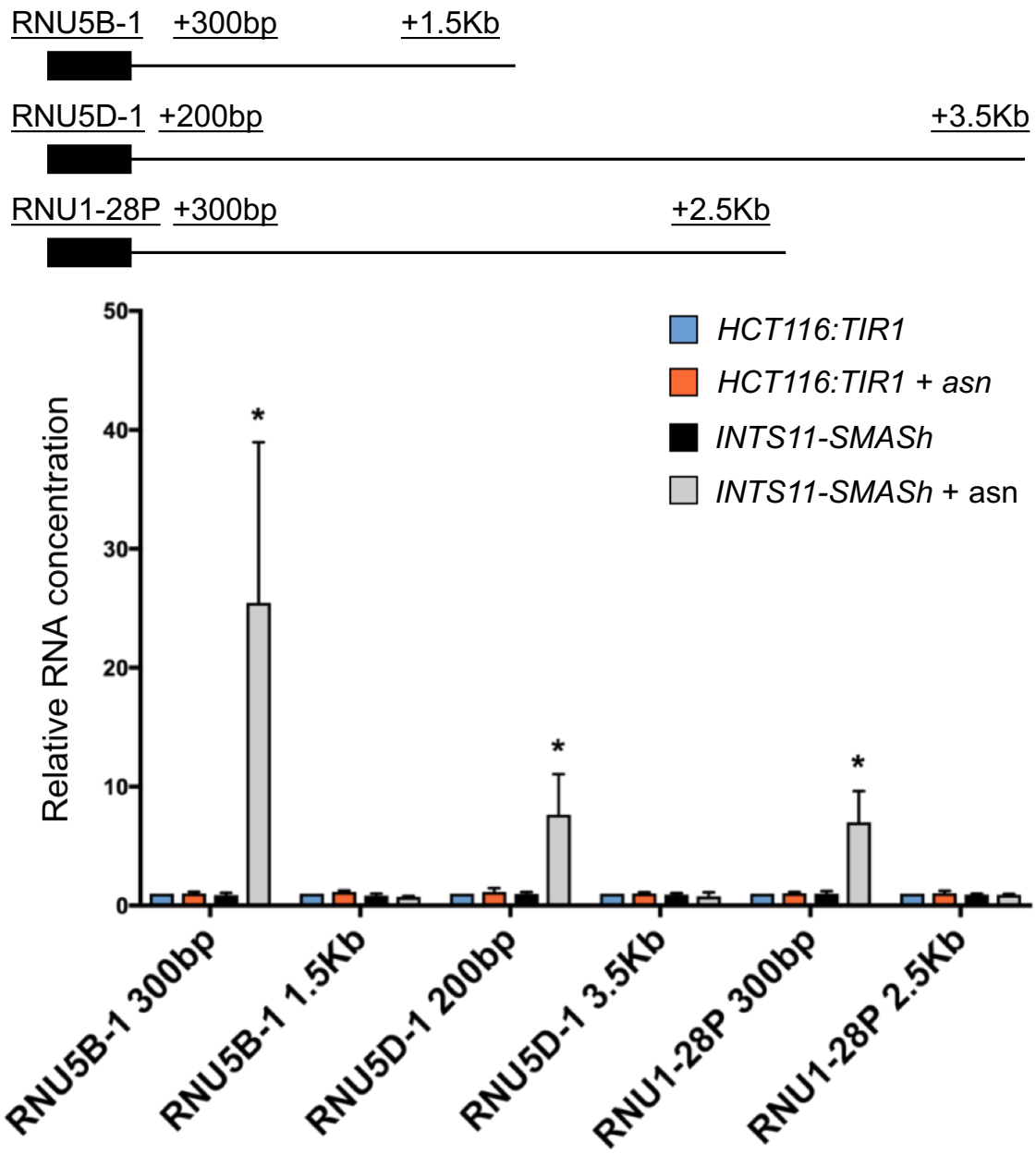

Figure 4.4 RNA concentration downstream of snRNAs

qRT-PCR detection of RNA levels downstream of the TES of three snRNAs: RNU5B-1, RNU5D-1 and RNU1-28P. Conducted in *HCT116:TIR1* and *INTS11-SMASh* cells with and without 48 hour asunaprevir treatment. Quantitation of RNA is expressed as fold change relative to untreated *HCT116:TIR1* cells. All levels were normalised to β actin. n = 3, * denotes $p < 0.05$, error bars show standard deviation. Data is the mean of three independent experiments with samples run in triplicate each time.

4.2 INTS11 depletion does not prevent snRNA termination

To further investigate the effects of INTS11 on snRNA transcription, RNA-Seq with single-end 50 bp reads was conducted on nuclear RNA obtained from *INTS11-SMASh* cells with or without asunaprevir treatment. Reads were aligned to the genome and filtered, with Table 4.1 showing details of the sequencing depth and coverage. To conduct a snRNA metagene plot, I firstly removed any genes with low expression (< 50 reads per gene). Not all snRNAs were enriched for in the RNA-Seq dataset which utilised 50 nt reads, reducing the resolution for small transcripts such as snRNAs (median length = 150 nt). Another reason why not all snRNAs are represented in the dataset is because there are variants of almost all snRNAs that have very similar sequences to one another (Kyriakopoulou et al, 2006; Sontheimer and Steitz, 1992; O'Reilly et al, 2013). This can prevent unambiguous mapping of reads, which are instead removed. Therefore, after filtering genes for expression levels, 95 non-overlapping snRNAs were used to generate the metagene plot with an inclusion window of 100 bp upstream of the snRNA TSS and 5 Kb downstream of the TES (Figure 4.5). For clarity, only 2 Kb downstream of the snRNA TES is shown.

Table 4.1 RNA-Seq statistical information for INTS11:SMASh cells

	INTS11:SMASh without asunaprevir	INTS11:SMASh with asunaprevir
Sequencing depth over all exons	10.0	9.8
Sequencing coverage over all exons	0.9	1.0

Sequencing depth over all exons = (Total number of mapped reads * average read length (bp)) / total length of exons

Sequencing coverage over all exons = (Total number of mapped reads to exons * average read length (bp)) / total length of all exons

Figure 4.5 shows an obvious increase in RNA reads immediately downstream of the snRNA TES upon INTS11 depletion, corresponding to readthrough of snRNA due to disruption of their cleavage by the Integrator. The number of reads then decrease, returning back to baseline values by 2 Kb and often much sooner. This suggests that extended snRNAs are terminated relatively close to the TES, even when not endonucleolytically cleaved by INTS11. To further validate these findings, individual RPKM normalised coverage tracks of the three snRNAs analysed in Figure 4.4, RNU5D-1, RNU5B-1 and RNU1-28P, were used to better visualise snRNA 3' extension and further validate RT-qPCR findings (Figure 4.6). For all three snRNAs there is a slight extension of reads past the TES. However, this readthrough stops by 2 Kb downstream suggesting snRNA termination occurs within this downstream window and corroborates both the metagene and qRT-PCR findings (Figure 4.5 and 4.4).

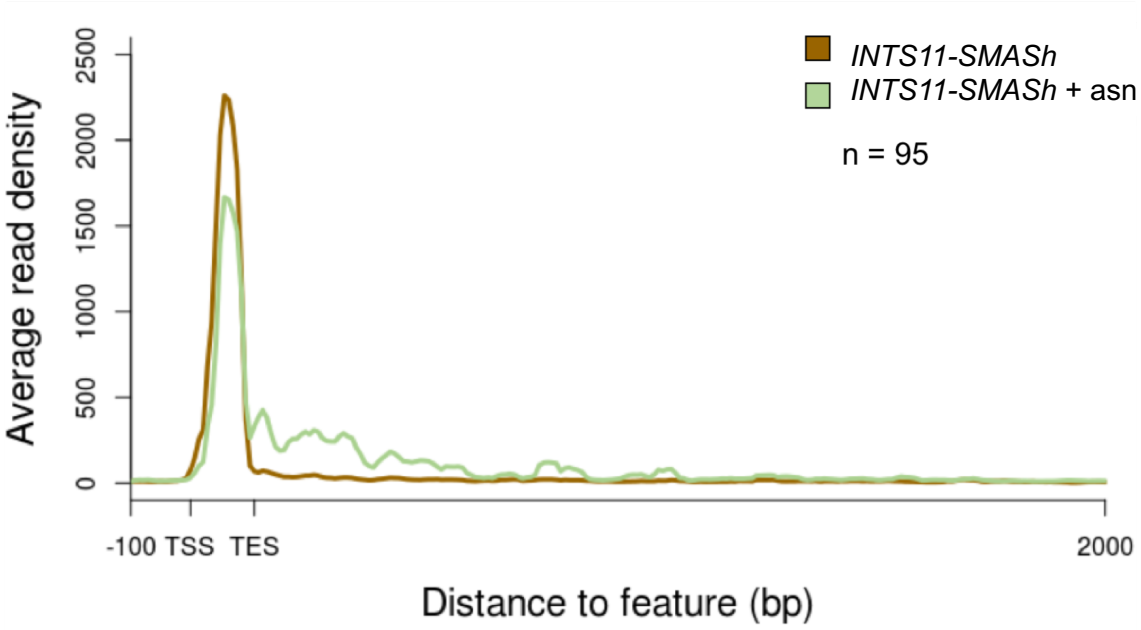

Figure 4.5 *INTS11-SMASh* snRNA metagene plot

Metagene coverage plot for 95 non-overlapping snRNAs from RNA-Seq data of *INTS11-SMASh* cells treated or not with asunaprevir. Inclusion window contains 100 nt upstream of the snRNA TSS and 2000 bp downstream of the TES, with a gene body scaled to 100 bp (n = 95). Figure represents one biological replicate.

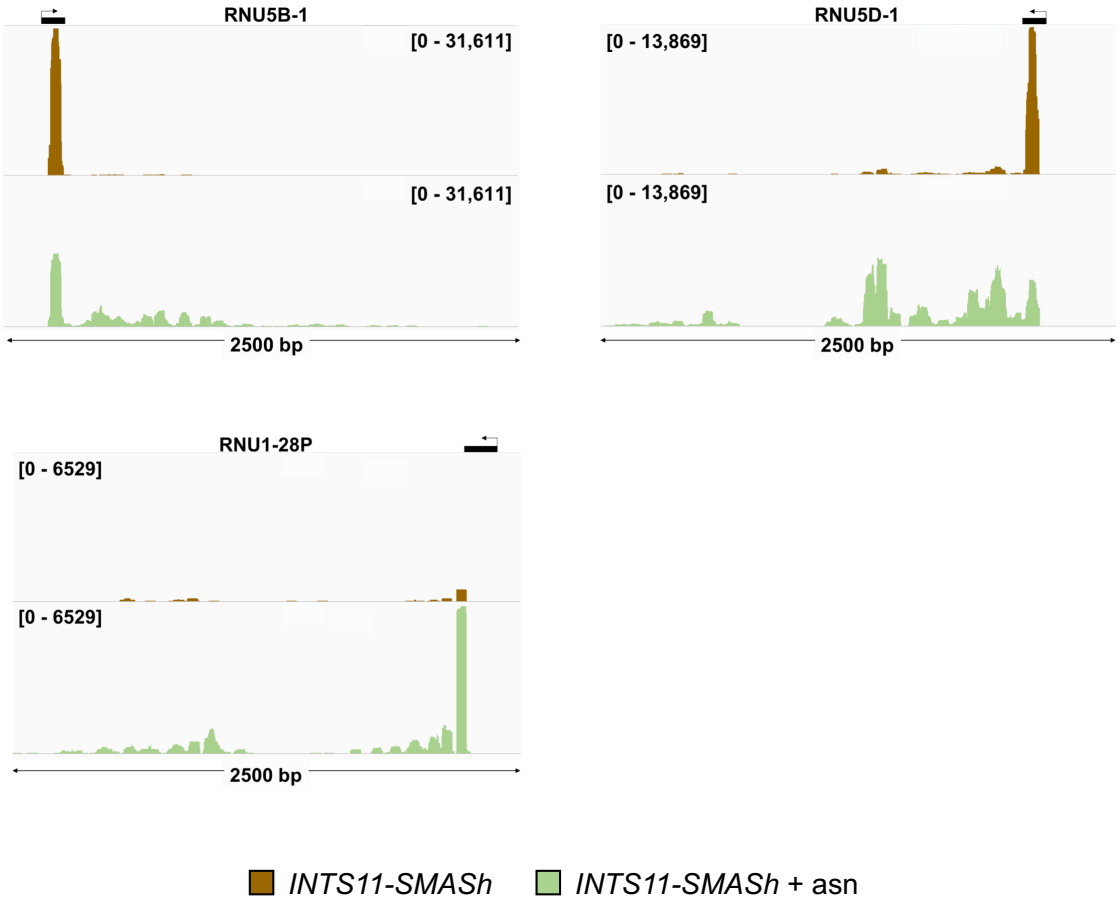

Figure 4.6 *INTS11-SMASh* RPKM coverage tracks for snRNAs

RPKM coverage tracks for snRNAs RNU5B-1, RNU5D-1 and RNU1-28P. Upon depletion of INTS11 there is a 3' extension of all three snRNAs that stops by approximately 2 Kb downstream of the snRNA TES. The numbers in brackets show the average RPKM normalised read count range.

4.3 Depletion of the largest Integrator subunit does not prevent snRNA termination

The current findings of snRNA termination after INTS11 depletion were slightly unexpected, due to previous reports of the close coupling between snRNA 3' end processing and termination. Following this I decided to investigate whether disrupting Integrator complex function as a whole, instead of only the endonuclease subunit, would inhibit snRNA termination. INTS1 is the largest subunit of the Integrator with 2190 amino acids (Baillat et al, 2005; Baillat et al, 2015). A knockout mice model of INTS1 has been shown to have growth arrest in early blastocyst stage embryos and apoptotic cell death (Hata and Nakayama, 2007). Additionally, INTS1 has been suggested to function as a scaffold protein for Integrator assembly and therefore disruption of INTS1 in human cells results in a loss of Integrator complex function (Hata and Nakayama, 2007; Baillat et al, 2005).

INTS1 was depleted in *HCT116:TIR1* cells using *INTS1* siRNA. To check levels of *INTS1* depletion by RNAi a qRT-PCR was conducted showing an average reduction of 84% (Figure 4.7B). As before, RNA levels downstream of three snRNAs (RNU1-1, RNU4-2 and RNU1-28P) were investigated in *HCT116:TIR1* cells that had been treated with either control siRNA or *INTS1* siRNA (Figure 4.7A). Different snRNAs were used to those in Figure 4.4, to allow investigation of RNA levels downstream of the TES at set intervals, that was previously not possible due to non-specific primers. Depletion of *INTS1* produced a similar effect to conditional depletion of INTS11 in *INTS11-SMASh* cells. There was a significant accumulation of unprocessed snRNA following INTS1 depletion, thus showing RNAi depleted *INTS1* levels were sufficient to cause Integrator dysfunction. As seen with INTS11 depletion, snRNA readthrough was relatively short with levels returning to background by 1 – 3 Kb downstream of the snRNA TES. This result shows that disrupting Integrator formation causes production of unprocessed extended snRNAs that are still capable of termination within a window close to the TES. It is possible the small amount of INTS1 remaining after INTS1 siRNA depletion may be sufficient for extended snRNA termination. However, the presence of detectable snRNA readthrough demonstrates that Integrator function has been significantly impaired by INTS1 depletion.

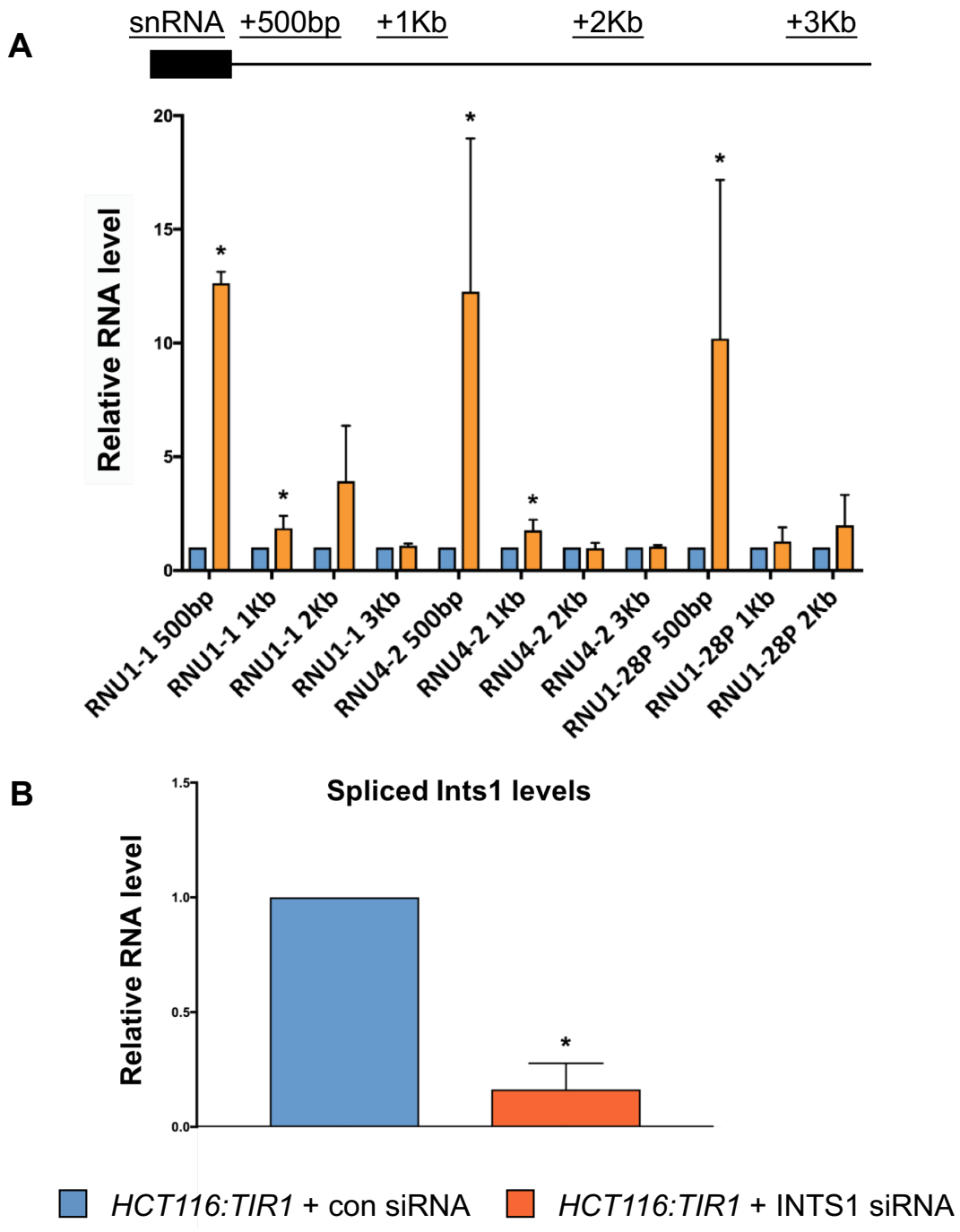

Figure 4.7 RNA levels downstream of snRNAs after INTS1 depletion

All levels were normalised to β actin. * denotes $p < 0.05$, error bars show standard deviation. Data is the mean of three independent experiments with samples run in triplicate. A) qRT-PCR detection of RNA levels downstream of three snRNAs: RNU1-1, RNU4-2 and RNU1-28P. Conducted in *HCT116:TIR1* cells treated with control siRNA or INTS1 siRNA. Quantitation of RNA is expressed as fold change relative to non-depleted *HCT116:TIR1* cells. B) qRT-PCR detection of INTS1 levels in *HCT116:TIR1* cells treated with control siRNA and INTS1 siRNA.

4.4 Depletion of INTS11 causes a further reduction in snRNA precursor transcript levels after inhibition of transcription

I next wanted to investigate the impact of INTS11 on snRNA transcript turnover. To do this, *INTS11-SMASh* cells were grown in the presence of actinomycin D for 0 minutes or 15 minutes and treated or not with asunaprevir (Figure 4.8). Actinomycin D acts as a transcription inhibitor by intercalating into GC rich DNA sequences and preventing RNA polymerase elongation (Trask and Muller, 1988). This process is fast and acts on all three RNA polymerases, as well as causing hyperphosphorylation of the Pol II CTD (Cassé et al, 1999). In untreated *INTS11-SMASh* cells, inhibition of transcription caused a reduction in the three uncleaved snRNA precursor transcripts of RNU5A-1, RNU4-1, RNU1-1, as measured by qRT-PCR. These snRNAs were investigated due to relevant primers to detect precursor transcripts already being available in the laboratory. The relative RNA concentration decreased to between 46 – 59% of RNA levels at 0 minutes of Actinomycin D treatment. This is expected as normal turnover of RNA occurs whilst the production of new transcripts is inhibited, overall resulting in a reduction of transcript levels. In comparison, when INTS11 was depleted a further decline in all three snRNA precursor transcripts was observed. This resulted in a mean RNA concentration that was reduced to between 25 – 36 % of levels at 0 minutes of Actinomycin D treatment. This apparent decrease in snRNA precursor transcript levels when INTS11 is depleted could be caused in a couple of ways. Firstly, INTS11 may not be depleted sufficiently to completely inhibit snRNA processing. Secondly, inhibiting INTS11 cleavage results in unprocessed snRNAs which may have increased efficacy for degradation by the exosome, for example. Finally, INTS11 depletion may cause a reduction in transcription of snRNAs.

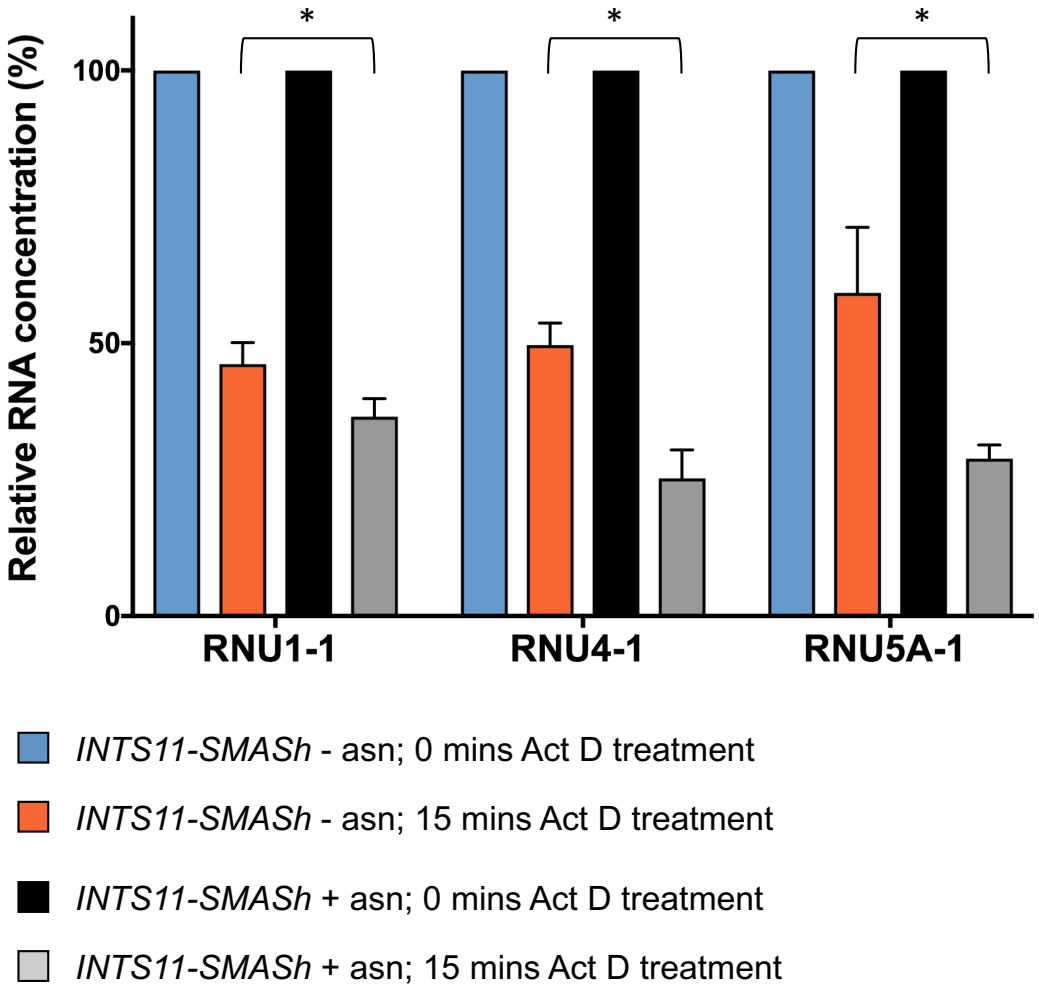

Figure 4.8 Precursor snRNA transcript levels after Actinomycin D treatment in *INTS11-SMASh* cells

qRT-PCR detection of RNA concentrations for three uncleaved precursor snRNAs (RNU5A-1, RNU4-1 and RNU1-1) in *INTS11-SMASh* cells treated or not with asunaprevir (asn), with 0 minutes or 15 minutes of Actinomycin D (Act D) treatment. Primers spanning the snRNA TES and immediately downstream of the TES were used. Quantitation is expressed as a percentage relative to 0 minutes of Actinomycin D treatment in both asunaprevir treated and untreated conditions. All levels were normalised to β actin. * denotes $p < 0.05$, error bars show standard deviation. Data is the mean of three independent experiments with samples run in triplicate each time.

4.5 Effect of DIS3 depletion on snRNA transcription

To assess whether the exosome is responsible for degradation of snRNA precursors, I decided to investigate the effects of DIS3 depletion on snRNAs. This would potentially explain snRNA reduction when transcription was inhibited, even when INTS11 is not present (Figure 4.8), Labno et al (2016) previously reported an accumulation of longer snRNA transcripts when both the endonuclease and exonuclease activity of DIS3 had been abolished by mutations in the PIN and RNB domains respectively. They hypothesised these transcripts were readthrough snRNAs that had not been cleaved by the Integrator and therefore extended downstream of the TES. In addition, DIS3 was found to degrade mature snRNA and the extended snRNA transcripts, with a slight increase in levels of mature snRNA transcripts observed with catalytically dead DIS3. In contrast, Szczepinska et al (2015) found little to no accumulation of snRNAs in either DIS3 PIN, RNB or both domain mutants suggesting DIS3 may not be part of the main pathway for snRNA degradation. Using qRT-PCR and *DIS3-AID* RNA-Seq data I aimed to examine these contrasting findings in more detail.

4.5.1 DIS3 depletion also produces extended snRNAs

Firstly, RNA-Seq data of *DIS3-AID* cells was analysed to visualise the effects on three individual snRNAs: RNU5B-1, RNU5D-1 and RNU1-28P (Figure 4.9A). These snRNAs were investigated as they had been used previously to validate the INTS11-SMASh cell line, show INTS11 depletion effects snRNA processing and confirm the observed extension of snRNAs (Figure 4.4 and 4.6) In all three examples there was an observable extension of the snRNA past the TES upon DIS3 depletion. There also appeared to be a slight increase in the amount of reads over the gene body. An explanation for this is that depletion of DIS3 may prevent degradation of snRNAs and result in their accumulation. Importantly, all three snRNAs showed extension that terminated before 1500 bp downstream of the TES, similar to our findings with INTS11 and INTS1 knockdown.

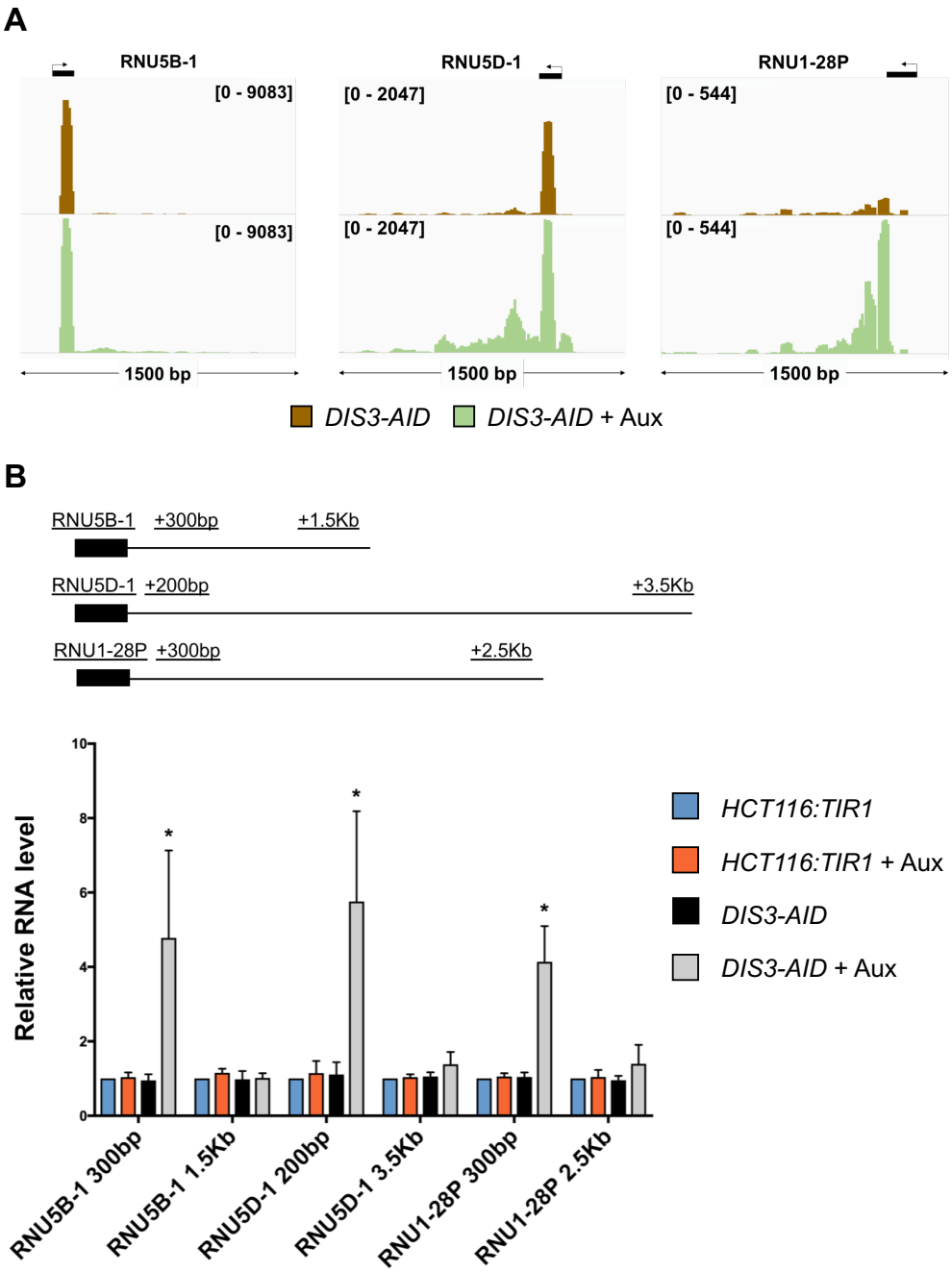

Figure 4.9 RNA levels downstream of snRNAs in *DIS3-AID* cells

A) RPKM coverage tracks for snRNAs RNU5B-1, RNU5D-1 and RNU1-28P in *DIS3-AID* cells treated or not with auxin. The numbers in brackets show the average RPKM normalised read count range. B) qRT-PCR detection of RNA levels downstream of three snRNAs: RNU5B-1, RNU5D-1 and RNU1-28P. Conducted in *HCT116:TIR1* and *DIS3-AID* cells with and without auxin treatment. Quantitation of RNA is expressed as fold change relative to untreated *HCT116:TIR1* cells. All levels were normalised to β actin. * denotes $p < 0.05$, error bars show standard deviation. Data is the mean of three independent experiments with samples run in triplicate each time.

For validation, we measured RNA levels downstream of these snRNAs by qRT-PCR (Figure 4.9B). Similar to the RPKM coverage tracks, extended snRNAs were detected upon DIS3 loss and these snRNAs terminated within 1.5 – 3.5 Kb downstream of the snRNA TES.

I next wanted to determine whether DIS3 depletion had a global effect on snRNAs or whether the effects observed were specific to the three snRNAs investigated. A snRNA metagene plot was conducted using the same list of 95 snRNAs as for the *INTS11-SMASh* metagene plot (Figure 4.5). For this metagene an inclusion window of 100 bp upstream of the snRNA TSS and 1000 bp downstream of the TES was used (Figure 4.10 and 4.11). There were no apparent differences upstream of the TSS upon DIS3 depletion, however there was an increased number of reads immediately downstream of the TES showing extension of snRNAs. This extension was not as long as seen previously upon INTS11 depletion (approximately 1 – 2 Kb), instead an increase in reads was observed up to approximately 500 bp downstream of the TES before returning to background levels. Additionally, there was a slight increase in reads over the snRNA gene body as similarly seen in RPKM coverage tracks of individual snRNAs (Figure 4.9). Overall these findings suggest DIS3 has a role in snRNA transcription. It is possible that DIS3 specifically degrades misprocessed or prematurely terminated snRNA transcripts, however the exact mechanism is yet unclear.

4.5.2 Depletion of DIS3 causes an accumulation of snRNA precursor transcripts when transcription is inhibited

To determine if DIS3 dysfunction had an effect on transcription of snRNA precursors, Actinomycin D was used to inhibit transcription in the same way as described previously with *INTS11-SMASh* cells using the same snRNA precursor primers (Figure 4.8). *DIS3-AID* cells, treated or not with auxin, underwent 0 minutes or 15 minutes of Actinomycin D treatment (Figure 4.12). As expected, inhibition of transcription in untreated *DIS3-AID* cells caused a reduction in RNU1-1, RNU4-1 and RNU5A-1 uncleaved snRNA precursors.

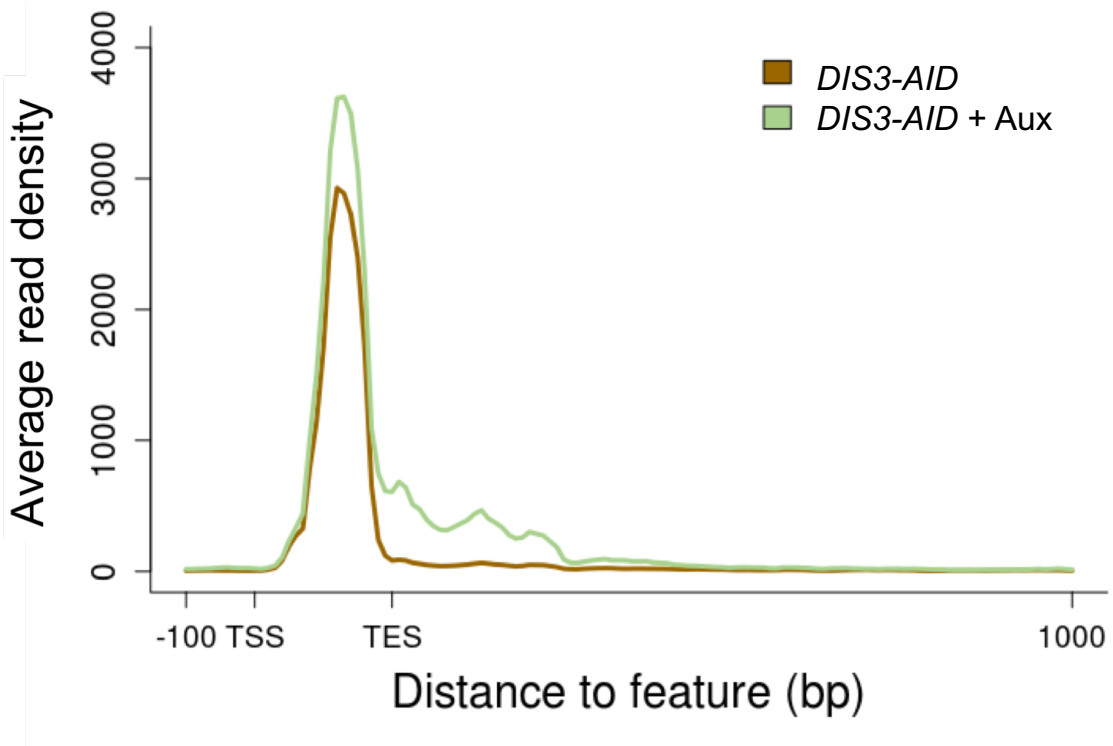

Figure 4.10 *DIS3-AID* snRNA metagene plot

Metagene coverage plot for 95 snRNAs from RNA-Seq data of *DIS3-AID* cells treated or not with auxin. Inclusion window contains 100 bp upstream of the snRNA TSS and 1000 bp downstream of the TES, with a gene body scaled to 200 bp (n = 95). Figure represents one biological replicate, a second replicate is shown in Figure 4.11.

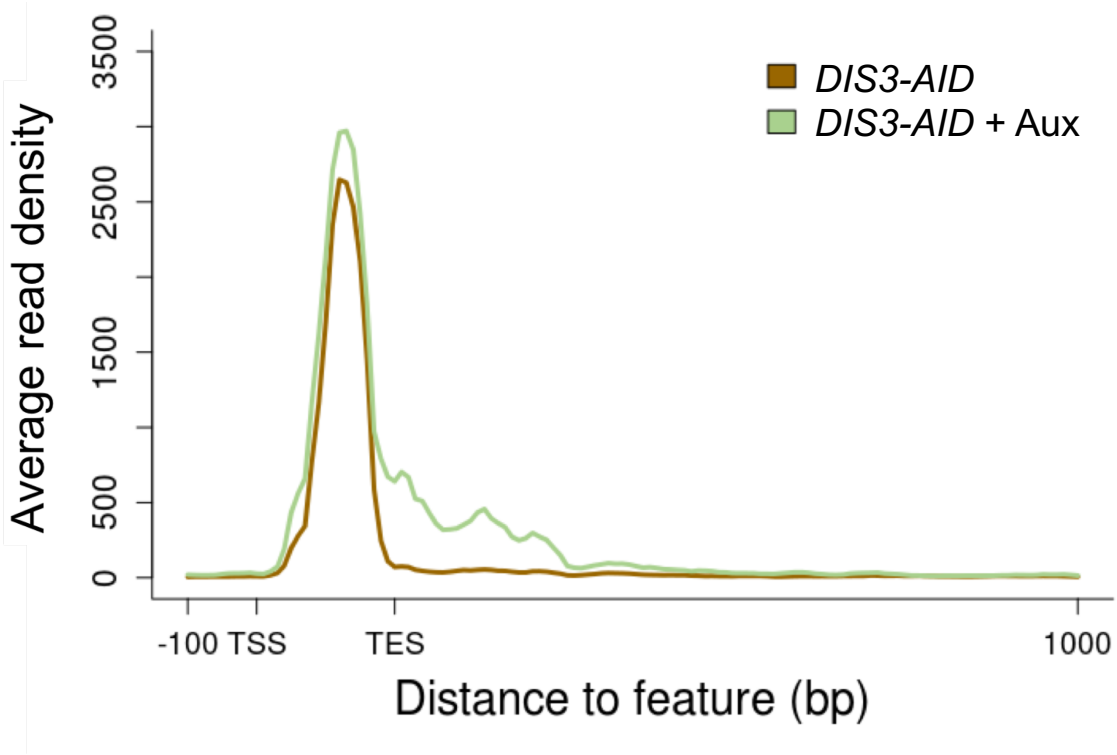

Figure 4.11 Second replicate of DIS3-AID snRNA metagene plot

Second biological replicate of a metagene coverage plot for 95 snRNAs from RNA-Seq data in DIS3-AID cells treated or not with auxin. Inclusion window contains 100 bp upstream of the snRNA TSS and 1000 bp downstream of the TES, with a gene body scaled to 200 bp (n = 95).

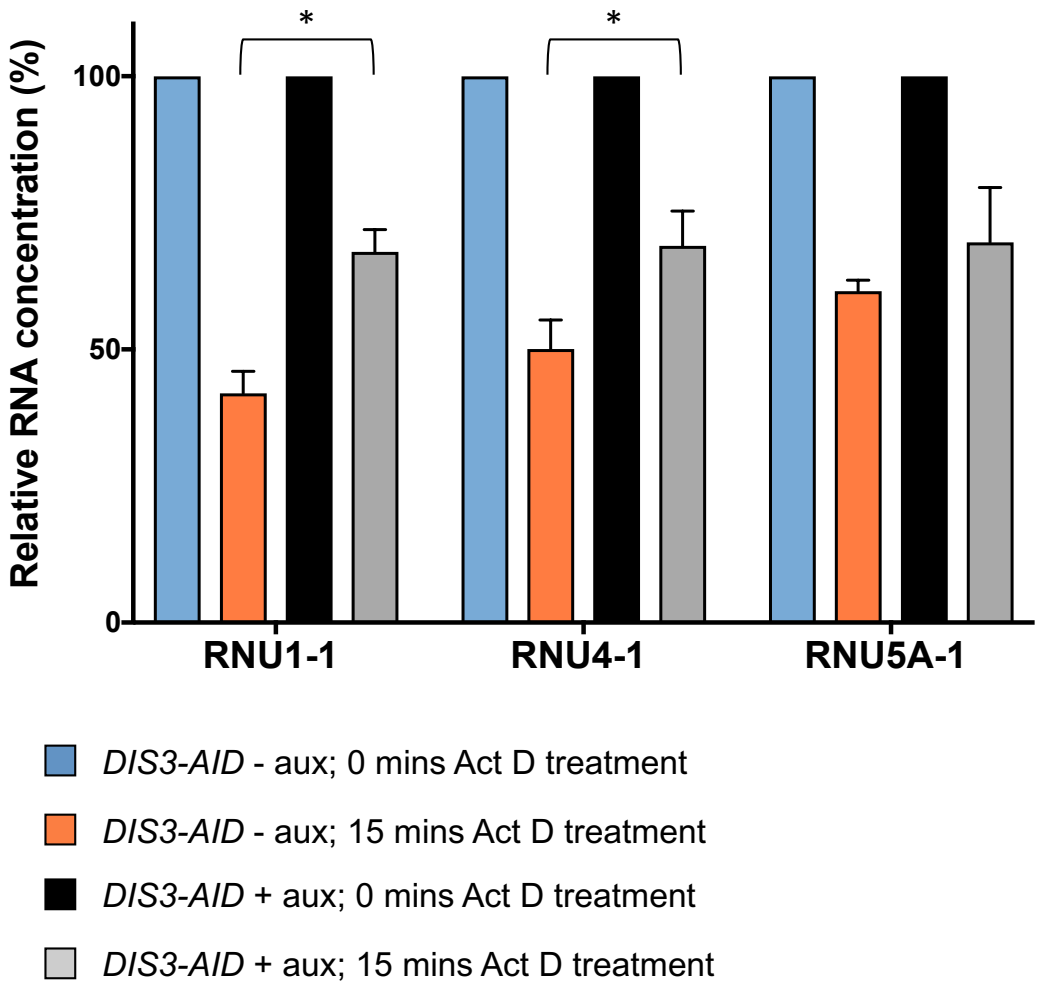

Figure 4.12 snRNA precursor levels after Actinomycin D treatment in *DIS3-AID* cells

qRT-PCR detection of RNA concentrations for three uncleaved precursor snRNAs (RNU5A-1, RNU4-1 and RNU1-1) in *DIS3-AID* cells treated or not with auxin and with 0 minutes or 15 minutes of Actinomycin D treatment. Primers spanning the snRNA TES and immediately downstream of the TES were used. Quantitation is expressed as a percentage relative to 0 minutes of Actinomycin D treatment in both auxin treated and untreated conditions. All levels were normalised to β actin. * denotes $p < 0.05$, error bars show standard deviation. Data is the mean of three independent experiments with samples run in triplicate each time.

On the other hand, depletion of DIS3-AID caused a reduction in snRNA precursor transcript levels which was much less pronounced than when DIS3 was present. This difference was significant for both RNU1-1 and RNU4-1. This could be explained by the normal function of DIS3 degrading snRNA precursor transcripts, causing a reduction in transcript levels when transcription is inhibited. However, upon loss of DIS3 the snRNA precursor transcripts are no longer degraded and instead accumulate. These findings support a major role of DIS3 in the metabolism of snRNA precursors.

4.6 Depletion of INTS1 and DIS3 together has an accumulative effect on snRNA processing

As both loss of Integrator and DIS3 function causes accumulation of extended snRNAs, I investigated the effects of eliminating both. Using *DIS3-AID* cells with auxin and a siRNA to INTS1 allowed depletion of DIS3 and INTS1 simultaneously. RNA levels downstream of RNU4-2, RNU1-28P and RNU1-1 snRNAs, shown previously to extend upon INTS1 depletion (Figure 4.7), were measured using *DIS3-AID* cells with a non-targeting siRNA as a control (Figure 4.13). As now expected, loss of DIS3 caused accumulation of extended snRNAs.

Similarly to previously demonstrated, loss of INTS1 by siRNA produced a readthrough effect on all three snRNAs in *DIS3-AID* cells. Readthrough RNA concentrations decreased to control levels by 1 – 2 Kb downstream of the TES. When *DIS3-AID* cells were treated with both auxin and INTS1 siRNA there was an enhanced accumulation of extended snRNAs compared to INTS1 depletion alone. Depletion of INTS1 causes loss of Integrator function, meaning snRNAs are no longer cleaved and explains the observed readthrough effect. As DIS3 loss has a cumulative effect on INTS1 depletion, it may be that DIS3 can degrade these extended snRNAs and that loss of DIS3 results in their further accumulation. This supports findings by Labno et al (2016) who suggested DIS3 degrades both mature snRNA and extended snRNA transcripts. Interestingly although DIS3 and INTS1 depletion alone resulted in extended snRNAs, an accumulative effect was not observed for extended snRNA transcript length with termination occurring at around 1 – 2 Kb. This shows that independent termination pathways are present for extended snRNAs.

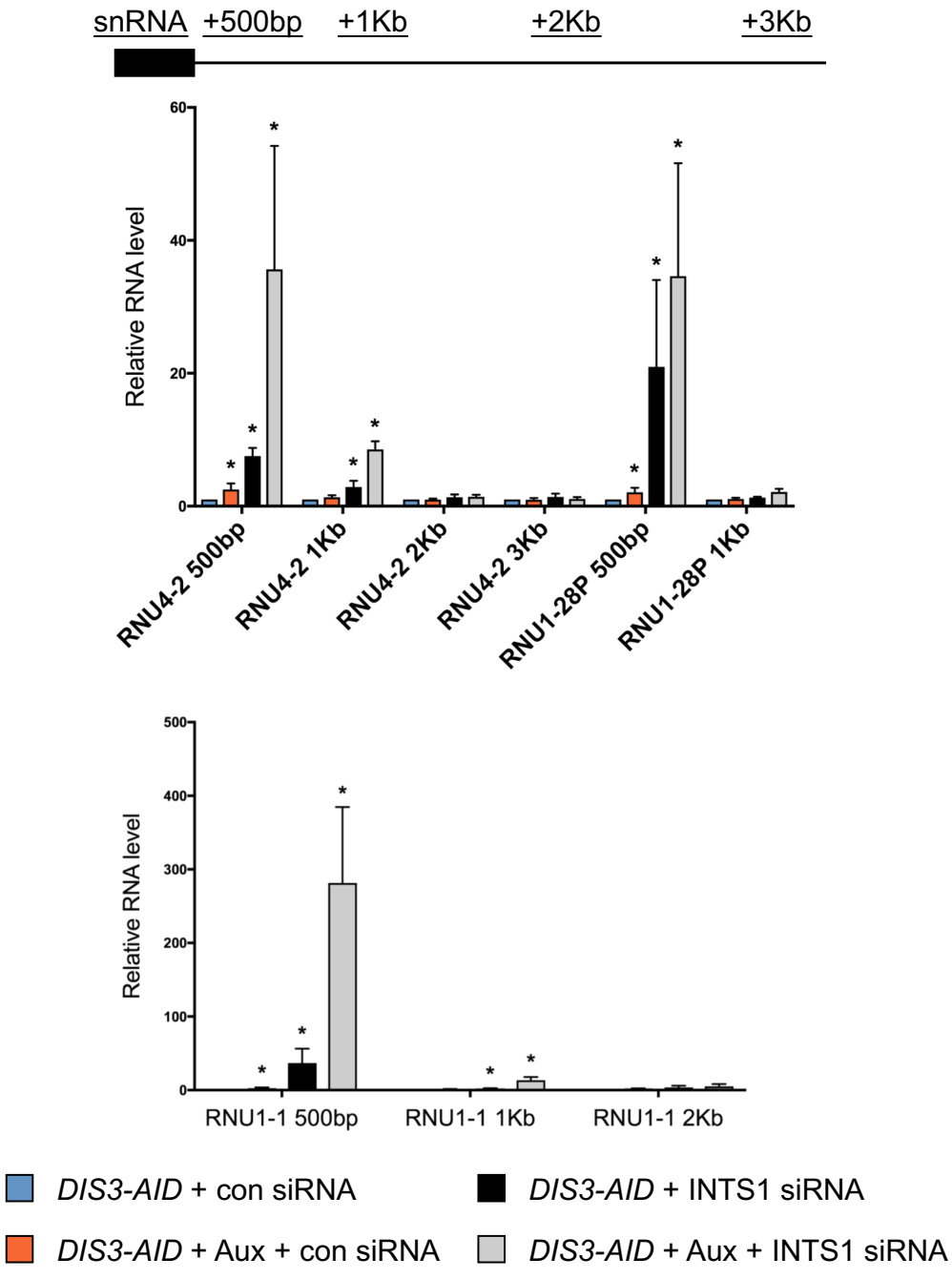

Figure 4.13 RNA levels downstream of snRNAs after INTS1 siRNA depletion in *DIS3-AID* cells

qRT-PCR detection of RNA levels downstream of three snRNAs: RNU1-1, RNU4-2 and RNU1-28P. Conducted in *DIS3-AID* cells treated or not with auxin and either control siRNA or INTS1 siRNA. Quantitation of RNA is expressed as fold change relative to non-depleted *DIS3-AID* cells with control siRNA. All levels were normalised to β actin. * denotes $p < 0.05$, error bars show standard deviation. Data is the mean of three independent experiments with samples run in triplicate each time.

4.7 Termination of extended snRNAs is likely to occur without cleavage

The results discussed so far suggest that snRNAs are still terminated in the absence of Integrator function. However, it is unknown whether this termination is caused by a downstream cleavage event by another endonuclease or through independent RNA Pol II dissociation from the genome. If the extended snRNAs were cleaved downstream of their TES, it is plausible that RNA Pol II would continue to extend downstream before dissociating and therefore create a small RNA fragment that may not have been detected in our data so far due to rapid degradation. XRN2 is the major 5' – 3' exoribonuclease in the nucleus and may be responsible for degradation of such a transcript, in a way that is similar to that suggested in the torpedo model of mRNA termination.

XRN2-AID cells had been previously generated in our lab, as described in Eaton et al (2018) and were used to investigate RNA levels around the 3' end of extended snRNAs that had shown extension upon Ints1 and DIS3 depletion, RNU1-1, RNU4-2 and RNU1-28P (Figure 4.14). XRN2 was depleted or not by 2 hours addition of auxin and cells were either treated with control siRNA or INTS1 siRNA. No significant differences were observed upon XRN2 depletion alone. As described in other cell lines, INTS1 siRNA treatment caused an accumulation of extended snRNAs that terminated before 2 Kb – 3 Kb downstream of the snRNA TES. These extended snRNAs were similarly detected upon depletion of both INTS1 and XRN2, however there were no significant differences in their levels between INTS1 depletion alone and both XRN2 and INTS1 depletion together. In addition there was no accumulation of RNA at 3 Kb downstream of the extended snRNAs and in the case of RNU4-2 at 2Kb downstream, therefore suggesting XRN2 does not degrade a transcript formed by Pol II extension after extended snRNA cleavage. From this I conclude that independently of a cleavage event, Pol II dissociation between 2 – 3 Kb downstream of snRNAs occurs when Integrator function is impaired and is the probable cause of extended snRNA termination.

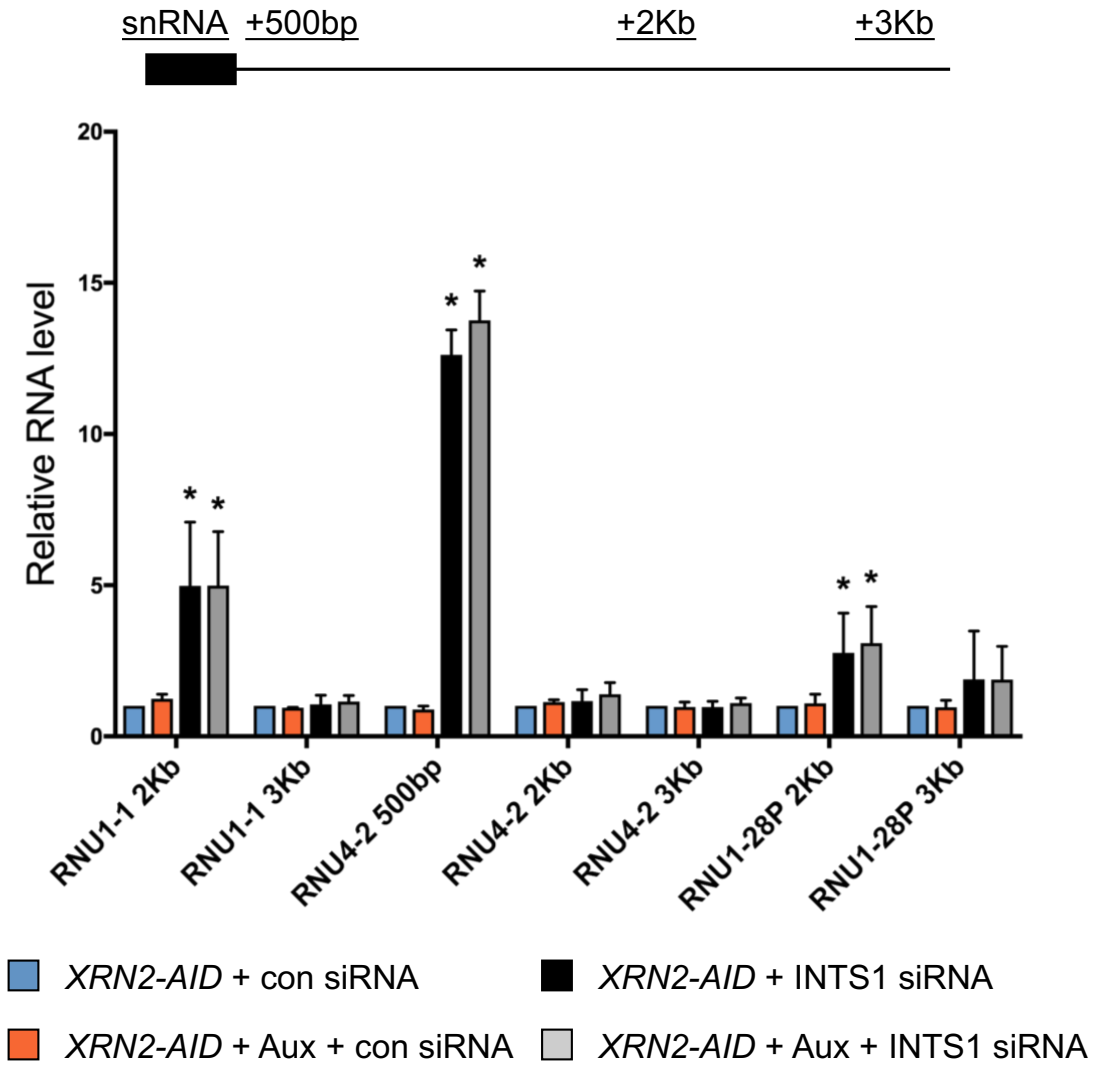

Figure 4.14 RNA levels downstream of snRNAs after INTS1 siRNA depletion in *XRN2-AID* cells

qRT-PCR detection of RNA levels downstream of three snRNAs: RNU1-1, RNU4-2 and RNU1-28P. Conducted in *XRN2-AID* cells treated or not with auxin and either control siRNA or INTS1 siRNA. Quantitation of RNA is expressed as fold change relative to non-depleted *XRN2-AID* cells with control siRNA. All levels were normalised to β actin. * denotes $p < 0.05$, error bars show standard deviation. Data is the mean of three independent experiments with samples run in triplicate each time.

4.8 Summary

In this chapter I have used RNA-seq analysis and qRT-PCR validation to emphasise the importance of endonuclease functions in snRNA transcription. Firstly, it was shown that depletion of the Integrator causes global extension of snRNAs downstream of their TES, with the metagene plot suggesting snRNA readthrough is still terminated by 2 Kb (Figure 4.5). INTS1 and INTS11 depletion results in dysfunction of the Integrator by disrupting proper Integrator formation or through loss of Integrator endonuclease activity, respectively. Both of these effects resulted in loss of Integrator cleavage at the 3' end of snRNAs upon recognition of the 3' box and therefore resulted in extended snRNA transcripts. Furthermore, neither INTS11 nor INTS1 depletion prevented termination of readthrough snRNAs. Instead it was observed that extended snRNAs caused by Integrator dysfunction are terminated within a window of 1 – 3 Kb downstream of the TES (Figure 4.6 and Figure 4.7). Inhibition of transcription showed that INTS11 depletion caused a reduction in the levels of snRNA precursors (Figure 4.8). Potentially these findings could be the result of increased degradation efficiency of extended snRNAs or an overall reduction in transcription of snRNAs upon INTS11 depletion.

Secondly, I showed that DIS3 depletion was sufficient to cause extension of snRNAs and DIS3 plays a role in degradation of snRNA precursors / extended snRNA transcripts. Interestingly, DIS3 dependent snRNA extension did not continue further than 1 – 3 Kb downstream of the snRNA TES, similar to snRNA readthrough observed upon Integrator dysfunction (Figure 4.9 and Figure 4.10). Whether DIS3 depletion also effects snRNA processing is unclear. In addition, the findings observed upon transcription inhibition suggested that DIS3 depletion does not affect the levels of snRNA transcription (Figure 4.12). Accumulation of snRNA precursor transcripts occurred upon DIS3 depletion as in normal conditions DIS3 would degrade snRNAs. I hypothesise that DIS3 also degrades extended snRNAs rather than having a function in their processing. This is supported by the accumulative effect seen when both DIS3 and Integrator function are impaired, compared to either DIS3 depletion or INTS1 depletion alone (Figure 4.13). INTS1 depletion resulted in accumulation of extended snRNAs through loss of their 3' end cleavage, which was intensified by DIS3 depletion causing defective degradation of extended snRNAs and resulting in

their accumulation. DIS3 degradation of extended snRNAs has been reported previously by Labno et al (2016).

Thirdly, another prominent finding throughout this chapter was that extended snRNAs are terminated relatively close to the snRNA TES. As mentioned previously this termination could be induced by Integrator independent transcript cleavage. It was suggested that Pol II would continue transcription slightly downstream of a cleavage event, as seen with protein-coding genes, and therefore a short transcript would be produced. This transcript would likely be rapidly degraded by a 5' – 3' exonuclease like XRN2. However, upon XRN2 depletion there was no observed significant differences in RNA levels downstream of the extended snRNAs (Figure 4.14). Therefore for a downstream cleavage event to occur, an endonuclease other than the Integrator and a different 5' – 3' exonuclease would potentially be required. As XRN2 is the major exonuclease in the nucleus, it appears termination of extended snRNAs is more likely a result of Pol II dissociation without cleavage. The lack of an XRN2 effect in this data supports the work of Eaton et al (2018), who found no role for XRN2 in snRNA termination.

Overall I have highlighted the function of both the Integrator and exosome in snRNA metabolism, whilst also exploring the association between snRNA 3' cleavage and termination. Although I have shown Integrator cleavage of snRNA is not necessary for transcription termination, it is possible that cleavage promotes more efficient termination and therefore disruption of the Integrator causes termination delay. In the following chapter I will investigate the role of CPSF73, an endonuclease that is also known for its role in cleavage at the 3' end of some genes, in particular protein coding mRNA.

5. Results Chapter 3: The role of the endonuclease CPSF73 in processing of protein-coding genes and transcription of snRNAs

Transcription has been most studied at protein-coding genes and CPSF73 is known to have a major role in mRNA processing. CPSF73 recognises the AAUAAA hexamer sequence of the mRNA PAS and co-transcriptionally cleaves the mRNA 18 – 30 nts downstream of the PAS. This releases the nascent RNA and allows binding of polyadenylation factors (Proudfoot et al, 2011; Ryan et al, 2004). In the torpedo model of transcription termination at protein coding genes, it is CPSF73 cleavage that enables transcription termination of mRNA. In this model it is believed that cleavage at a PAS site is required for mRNA transcription termination, with Pol II pausing after PAS transcription to enhance transcription termination efficiency (Fusby et al, 2016; Eaton et al, 2018). In the allosteric model of transcription termination, it is thought that transcription of the PAS results in a conformational change in the Pol II elongation complex leading to termination. This model is supported by data showing cleavage is not required for termination (Osheim et al, 1999; Osheim et al, 2002; Zhang et al, 2015a). Similar to the Integrator endonuclease activity at snRNAs, CPSF73 is responsible for 3' end cleavage of mRNA. Therefore, I wanted to investigate whether CPSF73 depletion would cause a processing defect on protein coding genes, like that seen with extended snRNAs upon INTS11 and INTS1 depletion. Additionally, these findings could potentially support or dispute the torpedo termination model. For these experiments, CPSF73 was genetically modified in HCT116 cells to bring it under inducible control.

5.1 Production of the *CPSF73-AID* cell line

To investigate the role of the endonuclease CPSF73, an inducible CPSF73 knockdown cell line had been previously produced by Steven West. The aim was to generate *CPSF73-AID* cells using *HCT116:TIR1* parent cells and the same CRISPR/Cas9 protocol as for *DIS3-AID* cell production. However, this method yielded no cell colonies. It was hypothesised that the constitutively active TIR1 expression in *HCT116:TIR1* cells might have an effect on the levels of tagged CPSF73. Therefore, HCT116 cells with inducible TIR1 expression under

a tet promoter were instead generated (*HCT116:TIR1*^{tet}) and then utilised for *CPSF73* genetic modification with the AID-tag to produce *CPSF73-AID* cells. Putting TIR1 under the control of a tet promoter allowed TIR1 inducible expression by addition of doxycycline, a synthetic tetracycline alternative, to cell media. After which addition of auxin should be able to deplete CPSF73 in *CPSF73-AID* cells.

5.1.1 Full depletion of AID-tagged CPSF73 is dependent on TIR1 expression

To ensure the *CPSF73-AID* cell line was capable of inducible CPSF73 depletion, a western blot was conducted. Using *HCT116:TIR1* cells as a control, the levels of CPSF73 were analysed in *CPSF73-AID* cells treated or not with doxycycline (dox) for 16 hours, auxin for 2 hours or both (Figure 5.1). In untreated conditions *CPSF73-AID* cells showed similar levels of tagged CPSF73, as shown by the higher band, to endogenous levels of CPSF73 in *HCT116:TIR1* cells. An endogenous CPSF73 band was not present in *CPSF73-AID* cells, suggesting both alleles of *CPSF73* had been successfully tagged with the AID. Importantly, addition of auxin alone did not have an effect on tagged CPSF73 levels showing that TIR1 expression is required for CPSF73 depletion in this cell line. Auxin and doxycycline treatment together resulted in a near complete depletion of CPSF73. All further studies requiring CPSF73 depletion were then conducted by 16 hours doxycycline treatment and 2 hours auxin treatment in *CPSF73-AID* cells.

Interestingly, doxycycline treatment alone in *CPSF73-AID* cells caused a reduction in tagged CPSF73 levels. Others have reported auxin-independent depletion of AID tagged proteins in human, chicken and yeast cells (Zasadzinska et al, 2018; Nishimura and Fukagawa, 2017; Morawska and Ulrich, 2013). This finding gives support to the hypothesis that TIR1 expression affects tagged CPSF73 levels and helps explain why generation of *CPSF73-AID* cells in a parental *HCT116:TIR1* background was unsuccessful. In support of this, Mendoza-Ochoa et al (2019) found that in yeast, auxin independent depletion of the tagged protein could be caused by high levels of TIR1 expression. Proteasome-mediated AID tagged protein degradation in the absence of auxin was also reported by Sathyan et al (2019).

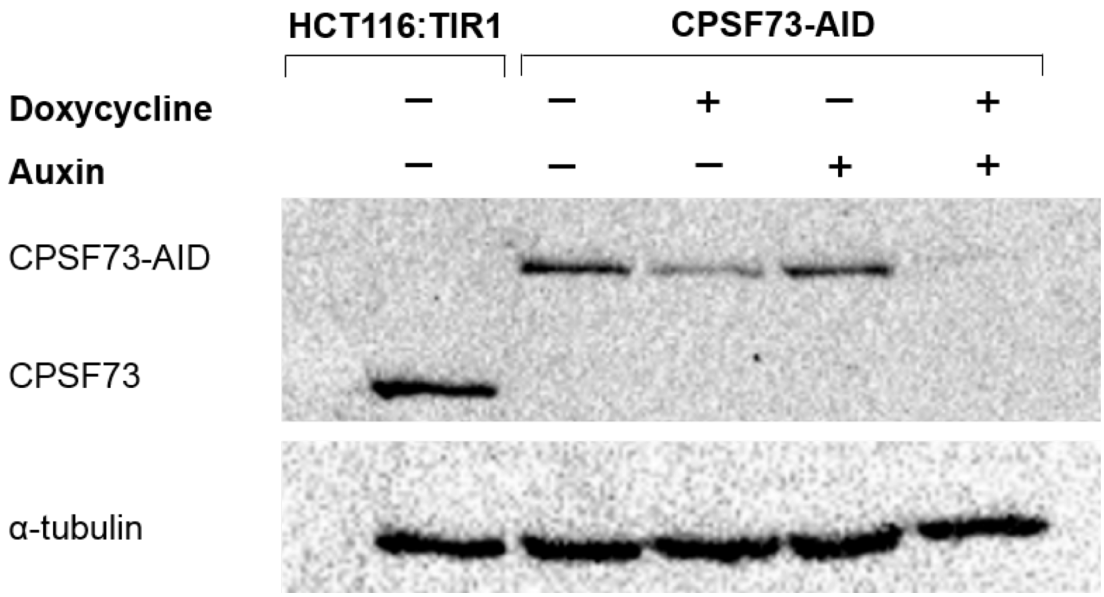

Figure 5.1 Western blot of CPSF73

Western blot showing the levels of endogenous CPSF73 in *HCT116:TIR1* cells (lower band) and AID-tagged CPSF73 in *CPSF73-AID* cells (higher band), treated or not for 16 hours with doxycycline to induce TIR1 expression, 2 hours auxin treatment or both. Alpha tubulin was used as a loading control. Full depletion of tagged CPSF73 requires TIR1 expression.

To overcome this issue they expressed an auxin response transcription factor (ARF), which in plants binds to AID in the absence of auxin (Figure 1.3). Expression of ARF rescued constitutive degradation of AID tagged proteins and increased the rate of degradation upon auxin addition.

5.2 CPSF73 depletion causes extensive readthrough of protein coding mRNA.

As it is known that CPSF73 is responsible for cleavage of mRNA, I investigated the effects of CPSF73 depletion on transcription and termination of protein-coding genes. RNA-Seq was conducted on CPSF73-AID cells, using single-end 50 bp reads. To generate RNA libraries, nascent nuclear RNA was extracted from cells after 16 hours of doxycycline and 2 hours auxin treatment, or no treatment. For analysis, reads were aligned to the genome after filtering. RPKM normalisation coverage plots were used to visualise read changes throughout the genome. Table 5.1 shows the RNA-Seq sequencing depth and coverage for both replicates of CPSF73-AID cells treated or not with auxin.

Table 5.1 RNA-Seq statistical information for CPSF73-AID cells

Merged replicate sequence libraries	CPSF73-AID without auxin	CPSF73-AID with auxin
Sequencing depth over all exons	28.6	28.5
Sequencing coverage over all exons	3.3	2.6

Sequencing depth over all exons = (Total number of mapped reads * average read length (bp)) / total length of merged exons

Sequencing coverage over all exons = (Total number of mapped reads to exons * average read length (bp)) / total length of all exons

To produce metagene plots, all annotated genes were filtered by expression and low or unexpressed genes were discarded. An inclusion window around each gene consisting of 3 Kb upstream of the TSS and 7 Kb downstream of the TES was utilised (Figure 5.2A). Any genes that overlapped other genes due to this inclusion window were removed from the analysis, to ensure a minimal false-positive discovery of CPSF73 depletion effects. Therefore 4702 genes were included in the metagene plot analysis, showing the average transcription profile of these genes.

From Figure 5.2A there appeared to be no effect on transcription levels upstream of the TSS when CPSF73 was depleted. On the other hand a major accumulation of reads extending downstream of the TES, that did not terminate before 7 Kb, were observed specifically upon CPSF73 depletion. This finding shows CPSF73 is required for protein-coding mRNA cleavage and that CPSF73 depletion results in extended mRNA transcripts due to continuation of Pol II transcription. In addition, a decrease in the average read density over the gene body is observed upon CPSF73 depletion. This could suggest that CPSF73 depletion has an effect on Pol II occupancy at protein-coding genes. In fact, Eaton et al (2018) found a general reduction in transcription upon CPSF73 loss and in support Mapendano et al (2010) reported an impairment in transcription with PAS mutations or polyadenylation factor depletion. One explanation for this is that upon CPSF73 depletion, Pol II will not dissociate from the genome and instead continues transcribing. This results in less recycled Pol II and therefore a reduction in transcription of the gene (Mapendano et al, 2010)

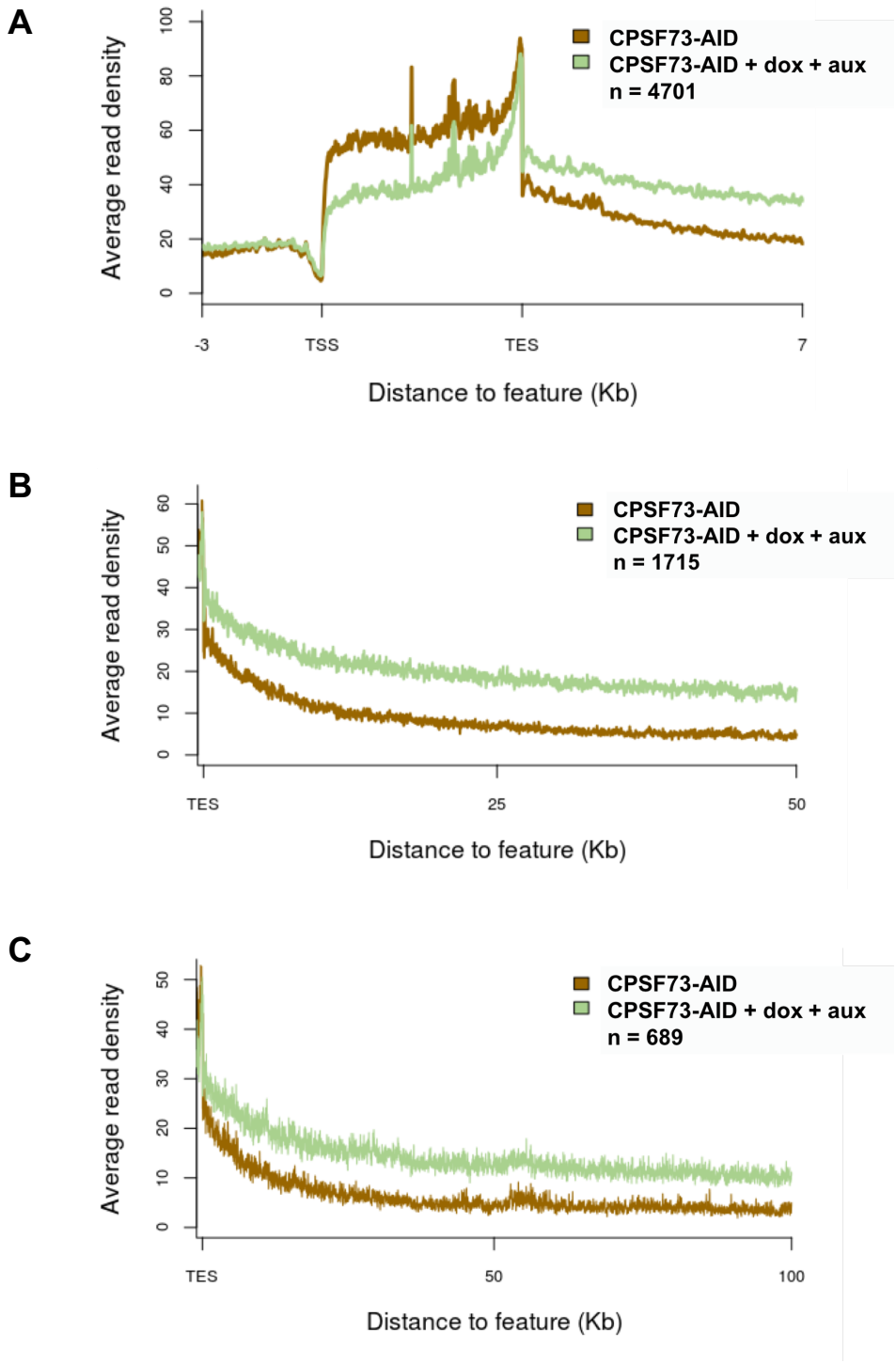

Figure 5.2 Metagene profiles of protein coding genes in *CPSF73-AID* cells

Metagene profile plots of non-overlapping protein-coding genes in *CPSF73-AID* cells with or without doxycycline and auxin treatment. A) Metagene of 4701 protein-coding genes. Inclusion window is 3 Kb upstream of the TSS and 7 Kb downstream of the TES, with the gene body scaled to 5Kb. B) Metagene of 1715 genes, showing 50 Kb downstream of the TES. C) Metagene of 689 genes, showing 100 Kb downstream of the TES. All represent one biological replicate; a second replicate is shown in Figure 5.3.

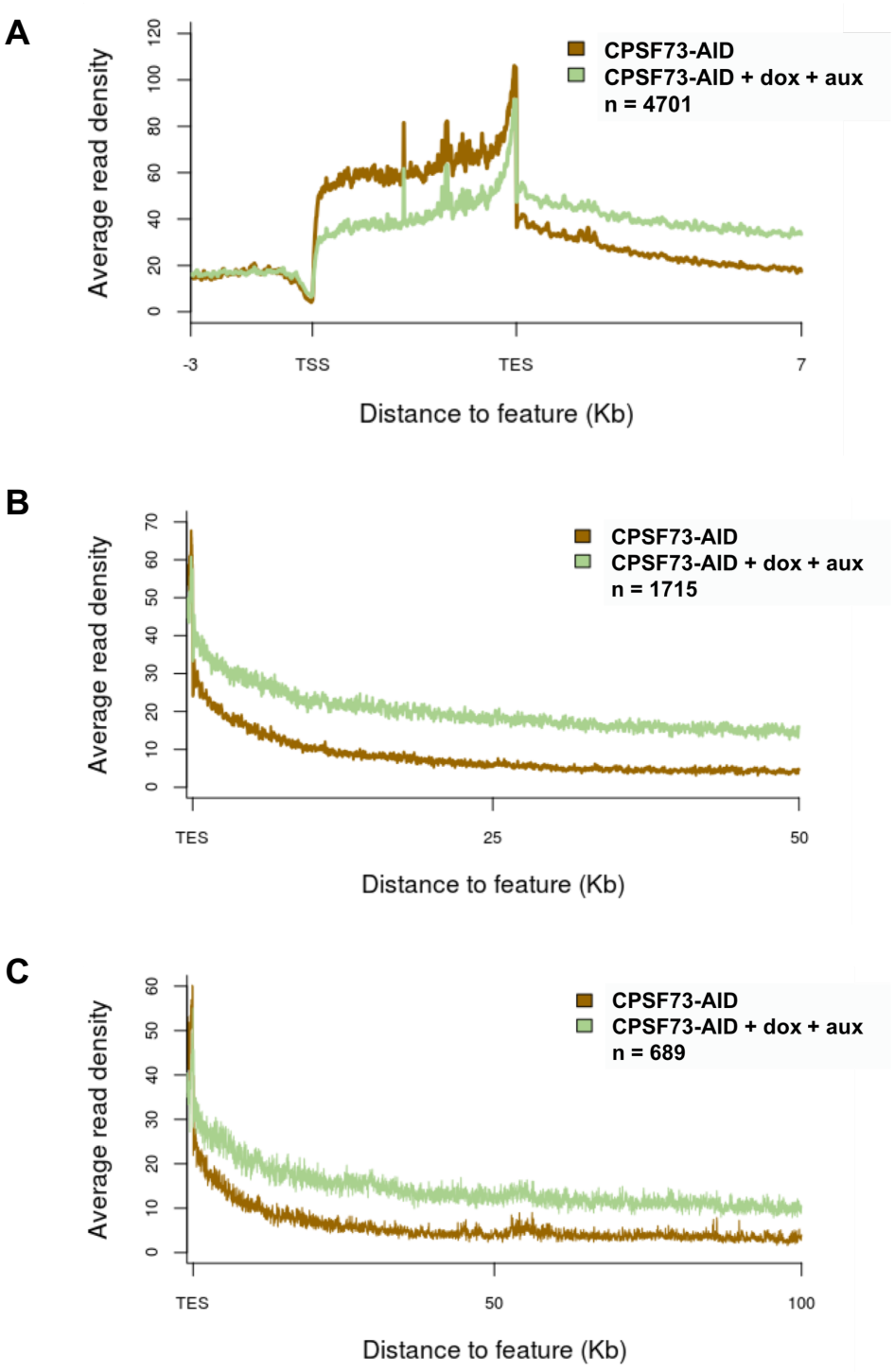

Figure 5.3 Second replicate of protein-coding gene metagene profiles in *CPSF73-AID* cells

Second biological replicate for metagene profile plots of non-overlapping protein-coding genes in *CPSF73-AID* cells with or without doxycycline and auxin treatment. A) Metagene of 4701 protein-coding genes. Inclusion window is 3 Kb upstream of the TSS and 7 Kb downstream of the TES, with the gene body scaled to 5Kb. B) Metagene of 1715 genes, showing 50 Kb downstream of the TES. C) Metagene of 689 genes, showing 100 Kb downstream of the TES.

As mRNA readthrough caused by CPSF73 depletion was still present at 7 Kb downstream of the TES, metagene plots with increased inclusion windows were generated with the aim to observe the length of extension. A metagene plot showing 50 Kb (Figure 5.2B) and 100 Kb (Figure 5.2C) downstream of the TES were generated, ensuring any overlapping genes were then removed. This resulted in the analysis of 1715 and 689 protein-coding genes for each metagene plot, respectively. Analysis of 1715 protein-coding genes showed that upon CPSF73 depletion there is an accumulation of extended transcripts that have readthrough of at least 50 Kb. Although the 100 Kb metagene plot had a slightly reduced average read density, showing less RNA transcripts extended to this length, there was still an obvious increase in the amount of extended transcripts upon CPSF73 depletion. The increased average read density did not reduce to levels comparable to those observed in the presence of CPSF73, showing that CPSF73-dependent readthrough can extend further than 100 Kb for some transcripts. Overall this data shows it is unlikely that Pol II terminates on these genes without CPSF73.

5.2.1 Unprocessed mRNAs can show more than 400 Kb readthrough

It is clear from the metagene plot analysis that CPSF73 depletion causes readthrough that can extend to 100 Kb and beyond, however it was still unclear when and if this extension would terminate. Using the list of 689 non-overlapping genes at 100 Kb, genes were randomly selected for further visualisation by RPKM coverage tracks. Three of these genes, AGTPBP1, GPD2 and NEK7 are shown in Figure 5.4. The pink area in this figure highlights the readthrough caused by CPSF73 loss. For all three genes extension was observed beyond approximately 400 Kb downstream of the TES. It is possible that with an increased auxin treatment time (> 2 hours) and therefore longer depletion of CPSF73, this extension could continue further than 400 Kb. Overall these findings suggest that Pol II transcription termination of protein-coding genes is tightly coupled to mRNA cleavage by CPSF73. As mRNA extension can be observed for thousands of base pairs it could be argued that CPSF73 is necessary for termination. Therefore giving further support to the XRN2 model / torpedo model for transcription termination.

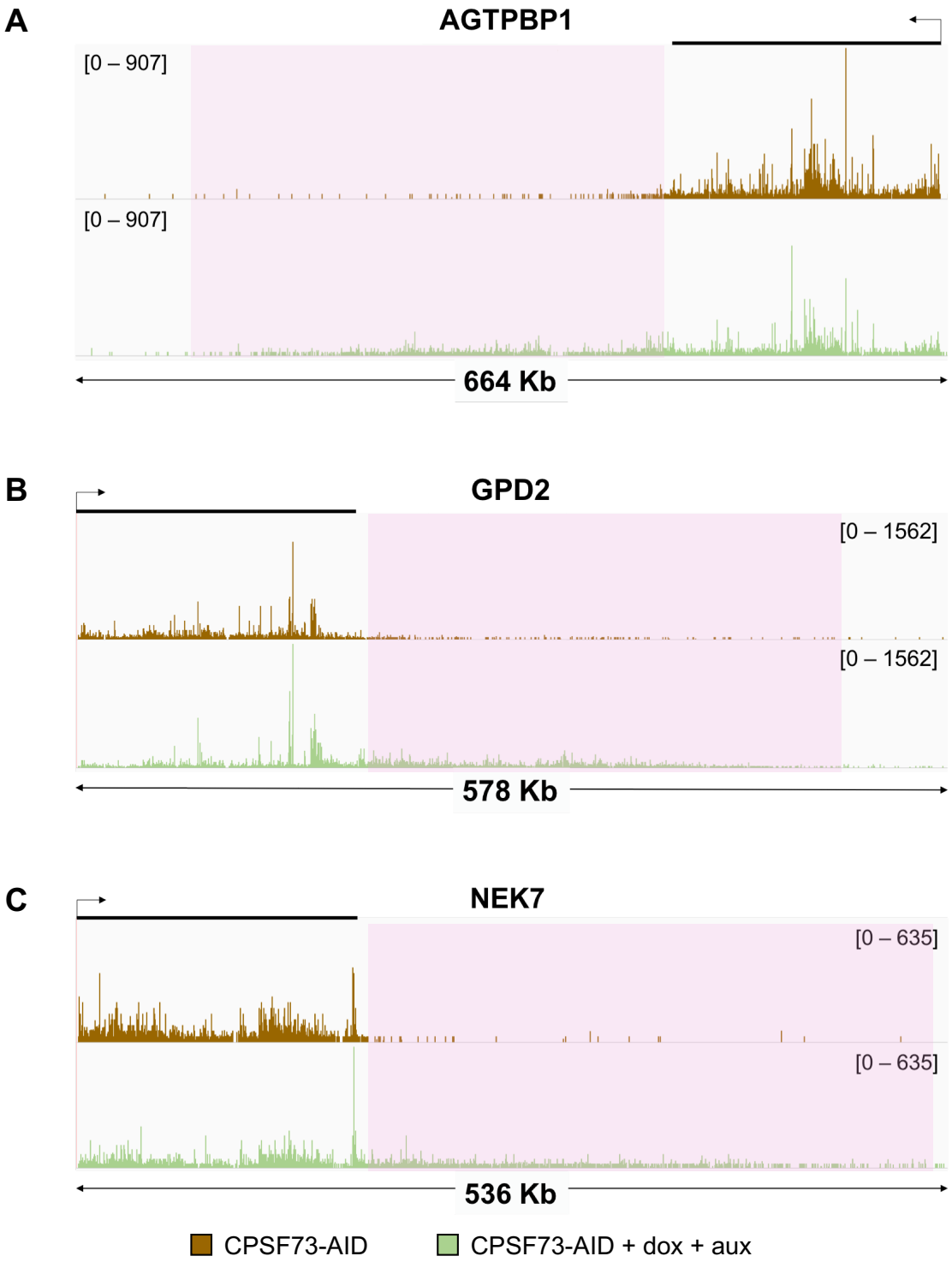

Figure 5.4 RPKM coverage tracks of extended mRNAs in *CPSF73-AID* cells

RPKM normalised coverage tracks showing three protein-coding genes, AGTPBP1, GPD2 and NEK7, from *CPSF73-AID* cells treated or not with doxycycline and auxin. The pink box area highlights the extension readthrough caused by CPSF73 depletion. The numbers in brackets show the average RPKM normalised read count range. Figure represents one biological replicate, a second is shown in Figure 5.5.

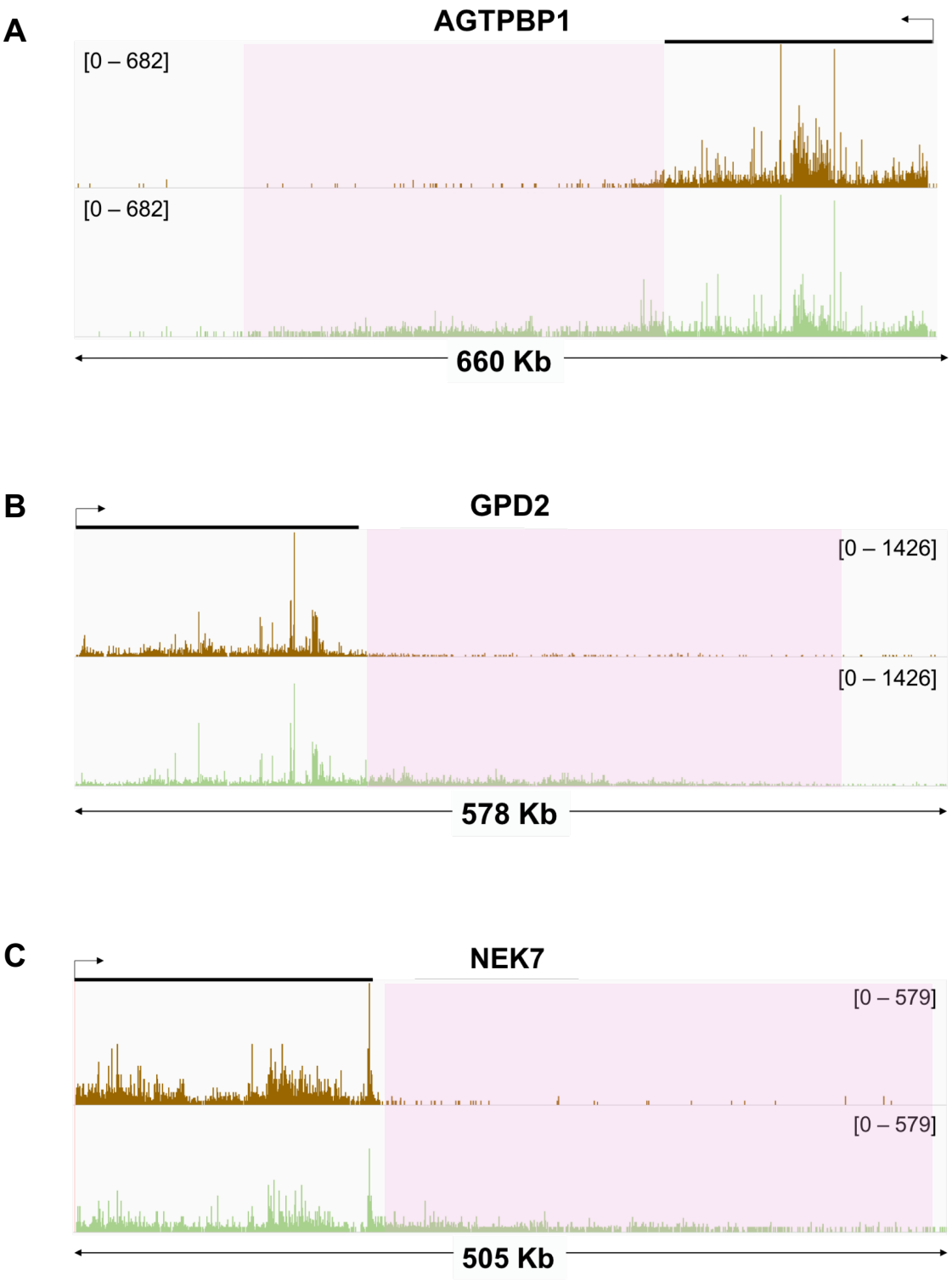

Figure 5.5 Second replicate RPKM coverage tracks of extended mRNAs in *CPSF73-AID* cells

Second biological replicate for RPKM normalised coverage tracks showing three protein-coding genes, AGTPBP1, GPD2 and NEK7, from *CPSF73-AID* cells treated or not with doxycycline and auxin. The pink box area highlights the extension readthrough caused by CPSF73 depletion. The numbers in brackets show the average RPKM normalised read count range.

5.2.2 mRNA readthrough can extend into neighbouring genes

CPSF73-dependent readthrough occurred at an extensive number of protein-coding genes, with extension often passing through other expressed genes. In these types of situations and from the RNA-Seq data alone, it was difficult to differentiate from which gene the readthrough originated from or whether increased reads over a gene were caused by readthrough into that gene or transcription upregulation. Instead I investigated readthrough into non-expressed neighbouring genes (Figure 5.6). In Figure 5.6B, CPSF73 depletion caused readthrough of a protein-coding gene, *SPIN1*, that extended approximately 300 Kb downstream of the TES. This readthrough extended into another previously non-expressed protein-coding gene, *NXNL2*, as well as a long intergenic non-coding RNA (lincRNA), *AL592486.1*.

Figure 5.6C shows a similar example, with readthrough from either *HIF1A* or the long non-coding RNA *AL137129* extending into *SYT16* which was not expressed in *CPSF73-AID* cells under no treatment conditions. In Figure 5.6A readthrough that may occur from multiple genes leads to the upregulation of reads over lincRNA *AL162414* and protein-coding gene *MUSK*. Overall it is clear that readthrough does not easily terminate and instead can continue into neighbouring genes, passing through their PAS sites and continuing onwards. This is not just true for expressed or non-expressed protein-coding genes, but also for readthrough extending into lncRNAs and potentially causing their upregulation. CPSF73 depletion causes issues with gene expression and regulation, which is likely to lead to unviable cells. This would explain why we were unable to obtain tagged-CPSF73 in a constitutively TIR1 expressing cell background (*HCT116:TIR1* cells). As shown previously, TIR1 expression causes a reduction in tagged CPSF73 levels (Figure 5.1). If these reduced levels were sufficient to cause readthrough as shown here, gene expression would be highly altered and would potentially affect cell survival.

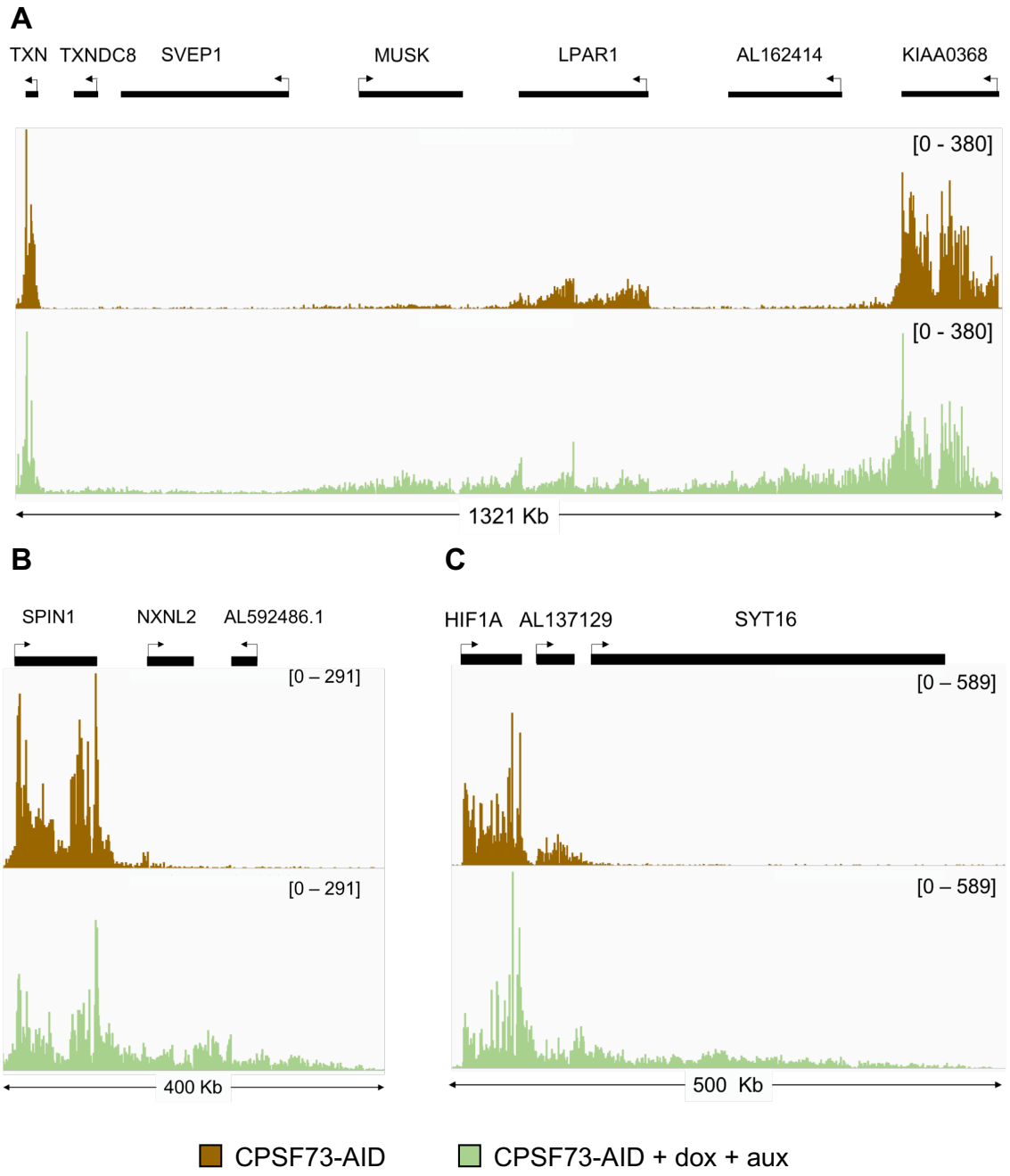

Figure 5.6 RPKM coverage tracks showing CPSF73 depletion dependent readthrough into neighbouring genes

RPKM normalised coverage tracks from *CPSF73-AID* cells treated or not with doxycycline and auxin. A, B and C each show different genes with readthrough of various length extending into neighbouring genes that weren't expressed under normal conditions. The numbers in brackets show the average RPKM normalised read count range. Figure represents one biological replicate, a second biological replicate is shown in Figure 5.7.

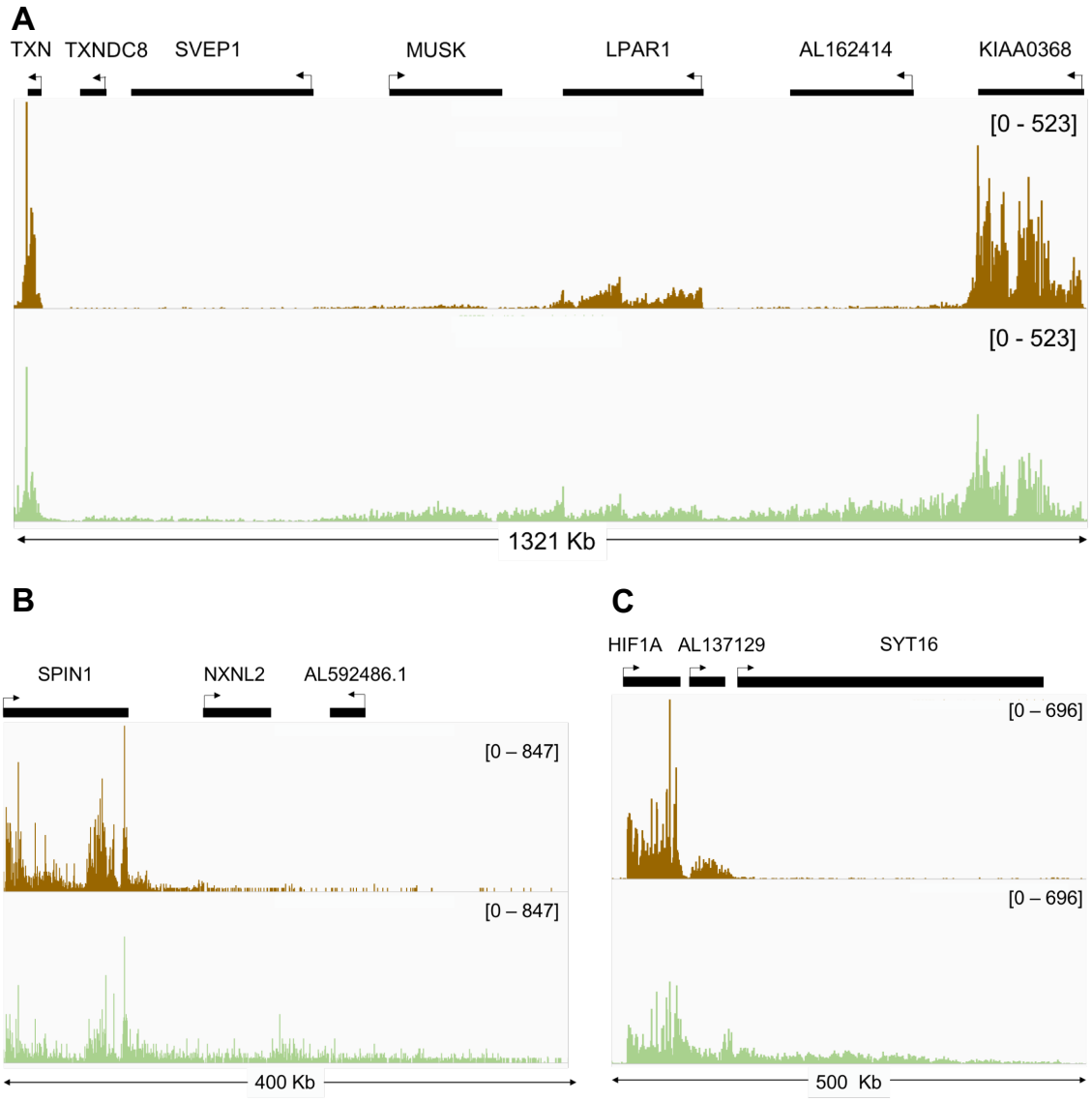

Figure 5.7 Second replicate RPKM coverage tracks showing CPSF73 depletion dependent readthrough into neighbouring genes

Second biological replicate for RPKM normalised coverage tracks from *CPSF73-AID* cells treated or not with doxycycline and auxin. A, B and C each show different genes with readthrough of various length extending into neighbouring genes that weren't expressed under normal conditions. The numbers in brackets show the average RPKM normalised read count range.

5.3 CPSF73 does not appear to play a role in snRNA processing

Although previous studies have shown the endonuclease component of the Integrator is responsible for cleavage of snRNAs, due to the homology between CPSF73 and INTS11 I wanted to investigate if CPSF73 had any effect on snRNA transcription. Using the same list of snRNAs to create the INTS11-SMASh snRNA metagene plot (Figure 4.5), a metagene plot was generated for *CPSF73-AID* cells (Figure 5.8A). From the metagene there were no observable differences, suggesting depletion of CPSF73 has no effect, at any level, on snRNA transcription and processing. To further confirm this finding I investigated some individual snRNAs, including RNU5B-1 and RNU5D-1, as shown in Figure 5.8B. RPKM coverage plots supported the finding that CPSF73 does not play a role in snRNA maturation, as no differences were observed upon CPSF73 depletion. These results were not unexpected as no role for CPSF73 in snRNA processing has been found previously, except in plants (Liu et al, 2016). Additionally, although the fission yeast CPSF73 homolog, YSH1, was found to bind to snRNAs it was not necessary for snRNA transcription termination, unlike the XRN2 homolog, DHP1 (Larochelle et al, 2018). However, in humans CPSF73 is known to interact with mature snRNAs, i.e. u7 snRNP, to regulate 3' end processing of RDH pre-mRNA (Yang et al, 2013).

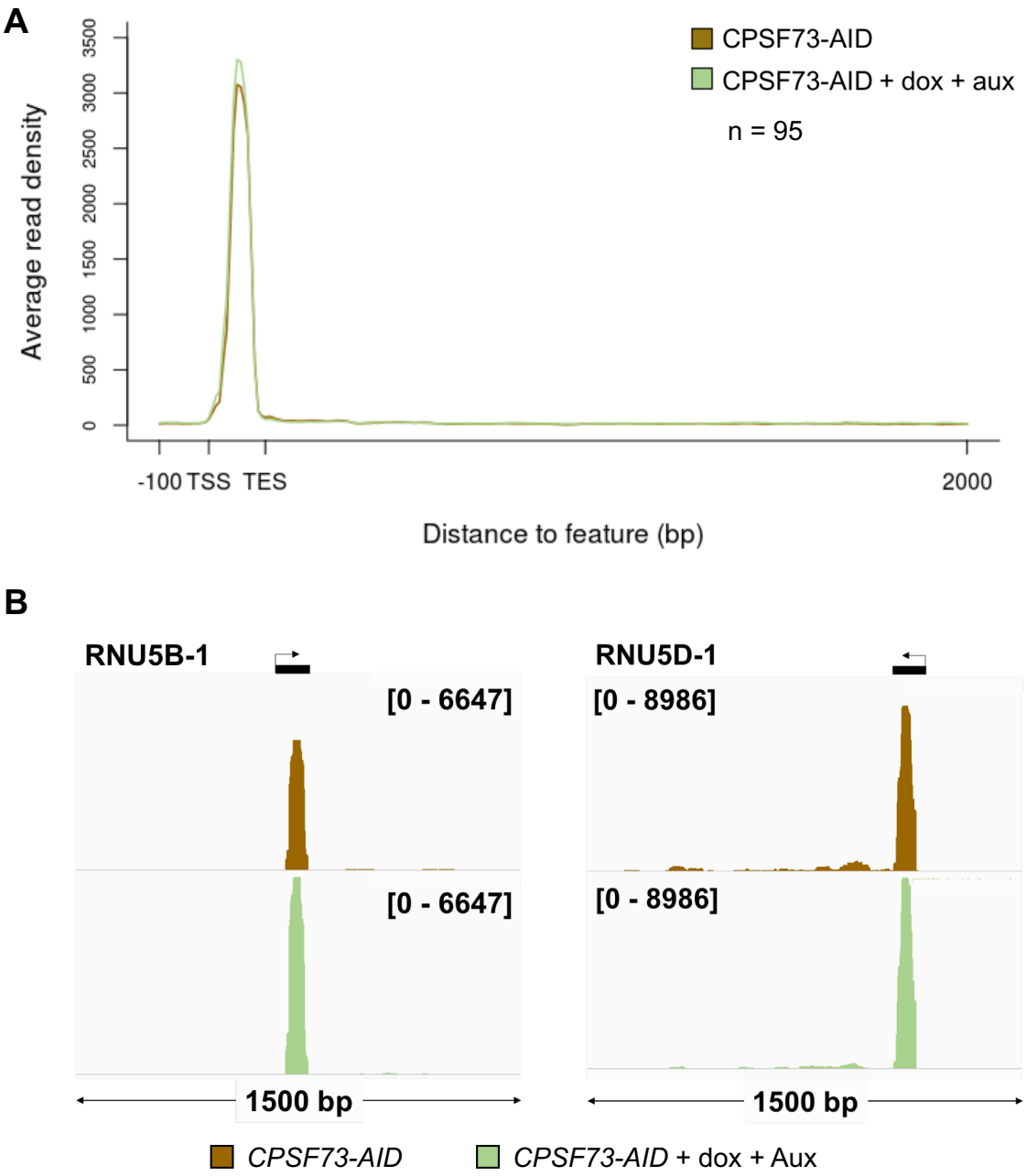

Figure 5.8 *CPSF73-AID* snRNA metaplot and snRNA RPKM coverage tracks

One biological replicate represented; see Figure 5.9 for an additional replicate. A) Metagene coverage plot for 95 snRNAs in *CPSF73-AID* cells, with an inclusion window 100 bp upstream of the TSS and 2000 bp downstream of the TES. The gene body was scaled to 100 bp. B) RPKM coverage tracks for snRNAs RNU5B-1 and RNU5D-1 in *CPSF73-AID* cells treated or not with doxycycline and auxin. The numbers in brackets show the average RPKM normalised read count range.

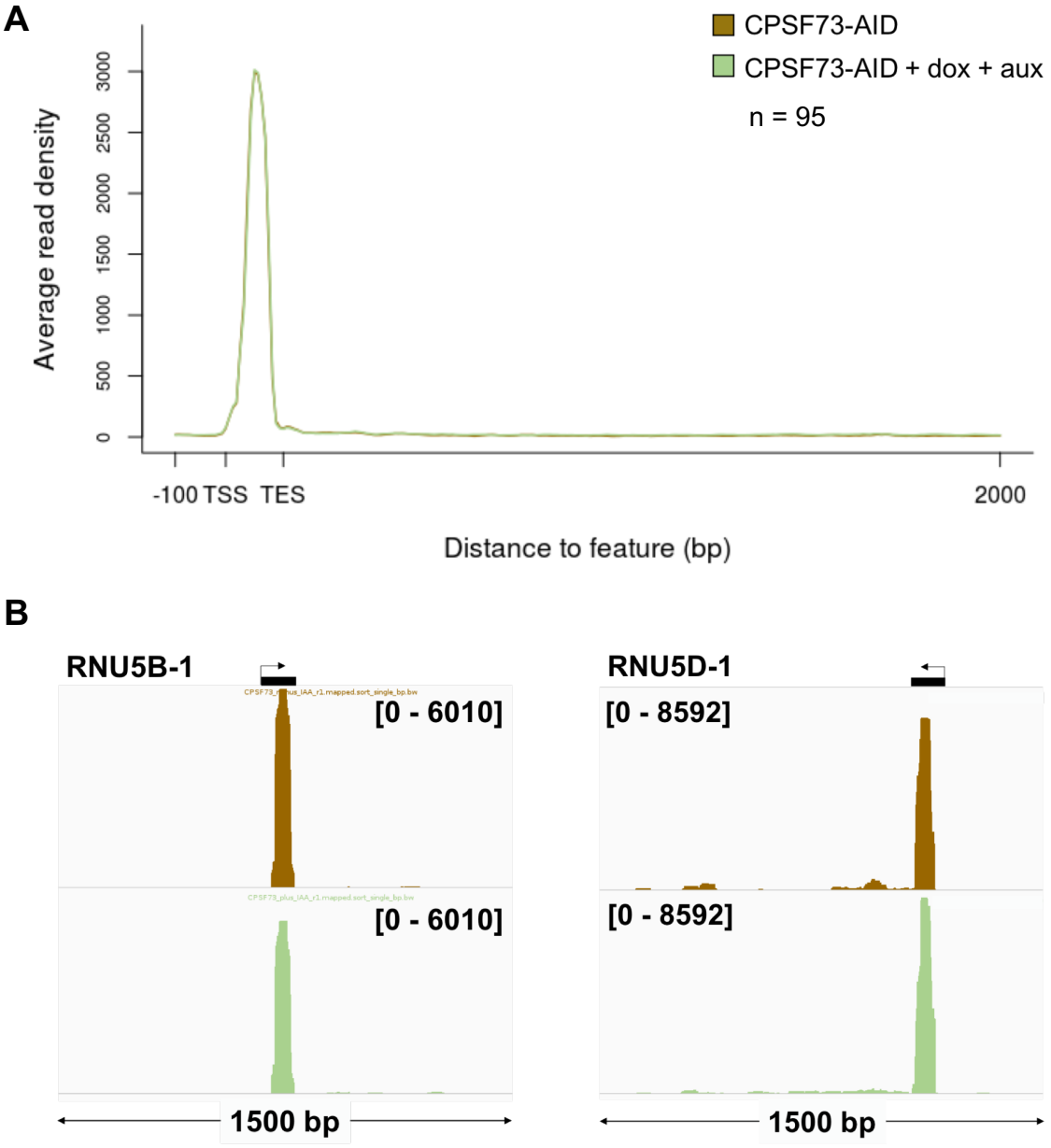

Figure 5.9 Second replicate *CPSF73-AID* snRNA metaplot and snRNA RPKM coverage tracks

Second biological replicate. A) Metagene coverage plot for 95 snRNAs in CPSF73-AID cells, with an inclusion window 100 bp upstream of the TSS and 2000 bp downstream of the TES. B) RPKM coverage tracks for snRNAs RNU5B-1 and RNU5D-1 in CPSF73-AID cells. The numbers in brackets show the average RPKM normalised read count range.

5.4 Summary

RNA-seq data from *CPSF73-AID* cells was able to demonstrate the major role of CPSF73 in protein-coding mRNA processing. Upon CPSF73 depletion numerous protein-coding genes show aberrant transcription termination resulting in Pol II readthrough (Figure 5.2). In some cases this readthrough was able to extend past 400 Kb downstream of the TES, showing impairment of Pol II dissociation from the genome when mRNA cleavage is impaired (Figure 5.4). Additionally, readthrough was not perturbed by neighbouring genes, with extension causing an accumulation of reads in genes that were previously lowly or not expressed (Figure 5.6). The global and major readthrough effect at protein coding genes when CPSF73 cleavage is inhibited, demonstrates the close relationship between 3' end cleavage of mRNAs and their termination. These findings are in line with previous work from the West laboratory (Eaton et al, 2018), giving support to the torpedo model of transcription termination and disputing such studies that suggested cleavage was not essential for termination (Osheim et al, 1999; Osheim et al, 2002; Zhang et al, 2015a). Additionally, in this work no function for CPSF73 was found in snRNA transcription (Figure 5.8).

Overall these findings are similar to those of Eaton et al (2018) who generated a conditional CPSF73 depletion cell line in HCT116 cells by tagging the C terminus of CPSF73 with a *Echerichia coli* DHFR-based degron. With this system, withdrawal of trimethoprim from cell media caused depletion of CPSF73. Western blot confirmed near complete depletion of tagged CPSF73 after 10 hours, which is slower than the AID system utilised in this work. As the work of Eaton et al (2018) was conducted in my lab, we aimed to produce a cell line capable of a quicker depletion of CPSF73 than the DHFR system. In support of the findings within this work, Eaton et al (2018) found a significant reduction in PAS cleavage at *MYC* and *ACTB* genes by RT-qPCR. As I conducted RNA-Seq on the CPSF73-AID cells I was able to show this defect in PAS cleavage was more widespread. The main difference in this work compared to Eaton et al (2018) was the use of CPSF73-AID cells to investigate the more immediate effects of CPSF73 depletion and the ability to analyse thousands of genes by conducting RNA-Seq.

Eaton et al (2018) performed ChIP on CPSF73-DHFR cells. They found that loss of CPSF73 caused a general reduction in transcription and extensive

transcription readthrough at *MYC* and *ACTB,* which also supports the findings in this chapter. Additionally, Eaton et al (2018) was able to show that a CPSF73 active site mutant could not support efficient transcriptional termination of *MYC* or *ACTB*, which was something not investigated within this work.

In the next chapter the function of CPSF73 is explored further, by investigating its role in the transcription and processing of RDHs, alongside DIS3 and the Integrator.

6. Results Chapter 4: Endonuclease function in replication dependent histone transcription and processing

Replication-dependent histones (RDHs) aid in packaging of newly synthesised DNA and are often found in clusters. They are transcribed by Pol II, are not polyadenylated and have a unique processing pathway (Figure 1.1). RDHs contain a 3' stem-loop and 5' purine-rich histone downstream element (HDE). Cleavage of RDH pre-mRNA occurs between the stem loop and HDE by the histone cleavage complex (HCC) which includes CPSF73, CPSF100, CstF64 and Symplekin (Marzluff and Koreski, 2017). Recruitment of the HCC requires binding of U7 snRNP to the HDE as well as the stem-loop binding protein (SLBP) aiding in stabilisation of U7 snRNP on the RDH pre-mRNA, potentially through interactions with FLASH (Skrajna et al, 2017). Similar to spliceosomal snRNPs, U7 snRNP contains a binding site for a Sm ring. The core U7 snRNP consists of U7 snRNA bound to five Sm proteins found in spliceosomal snRNAs. However, it also contains Lsm10 and Lsm11 proteins which replace SmD1 and SmD2. It is Lsm11 binding to FLASH that creates a docking platform for the HCC (Yang et al, 2013; Burch et al, 2011).

RDH pre-mRNA 3' end cleavage occurs rapidly upon transcription of the processing signal and unlike polyadenylated genes, where Pol II occupancy continues 4 – 6 Kb downstream of the TES, transcription terminates shortly after as shown by a quick drop in Pol II occupancy after the RDH TES (Anamika et al, 2012). CPSF73, as part of the HCC, is believed to be the main endonuclease responsible for RDH pre-mRNA cleavage and has been shown to be cross-linked to the RDH pre-mRNA cleavage site (Dominski et al, 2005). This endonuclease also has a major role in cleavage / polyadenylation of protein-coding mRNA (Mandel et al, 2006).

As U7 snRNA plays a major role in RDH pre-mRNA processing, it appears likely that disruption of Integrator function could indirectly affect RDH transcription through decreased levels of processed U7 snRNA. Interestingly, the Integrator has also been suggested to have a direct role in RDH processing. Skaar et al (2015) found the Integrator binds to the 3' end of RDH genes and knockdown of Integrator subunit 3 (INTS3) caused accumulation of unprocessed polyadenylated RDH transcripts. However, in Drosophila experiments no link

between Integrator dysfunction and histone pre-mRNA processing or cleavage and polyadenylation were found. In comparison, knockdown of CPSF73, CPSF100, Symplekin or SLBP was sufficient to affect RDH RNA processing (Ezzeddine et al, 2011). Therefore it is still not known what role, if any, the Integrator may have in RDH transcription. Unfortunately, RDH transcripts were not fully detectable in my INTS11 RNA-Seq dataset meaning I was unable to draw conclusions on the direct effect of Integrator depletion.

So far I have shown DIS3 is responsible for degradation of a multitude of transcripts and therefore it is possible DIS3 may also play a role in RDH mRNA or RDH precursor degradation. In support, Mullen and Marzluff (2008) found that disrupting exosome function caused a reduction in histone mRNA degradation. In a follow-up study, Slevin et al (2014) elucidated a pathway for histone degradation. Firstly a 3' – 5' exonuclease, 3'hExo, binds to SLBP and degrades the histone mRNA into the stem-loop. This forms a degradation intermediate with SLBP still bound. Upon removal of SLBP which would otherwise block further degradation, the exosome is able to degrade the histone mRNA. In addition, depletion of DIS3 showed readthrough histone transcripts produced by CstF64 knockdown were degradation targets of the exosome (Romeo et al, 2014). Thus, DIS3 as part of the exosome could be crucial for RDH mRNA and / or RDH precursor degradation. In this chapter I investigate the role of endonucleases on RDHs.

6.1 CPSF73 depletion doesn't affect RDH pre-mRNA processing

As mentioned in the previous chapter, *CPSF73-AID* cells underwent RNA-Seq to elucidate direct substrates of CPSF73. Using this data, I firstly wanted to elucidate the effect of CPSF73 knockdown on RDH mRNA, due to its major function as part of the HCC. RDH genes often cluster on the genome and therefore a cluster of 5 RDHs (HIST1H3A, HIST1H4A, HIST1H4B, HIST1H3B, HIST1H2AB) could be easily visualised together (Figure 6.1 and 6.2). Interestingly, no differences were observed in these five RDH transcripts upon CPSF73 depletion. CPSF73 may function at a specific subset of RDHs, therefore I further analysed another cluster of RDH genes. RPKM coverage tracks were used to visualise HIST1H4D, HIST1H1PS1, HIST1H3D, HIST1H2AD,

HIST1H2BF and HIST1H4E (Figure 6.3 and 6.4). Again, no apparent differences were observed.

One possible explanation for observing no effect on RDH transcription, may be that another endonuclease is sufficient for RDH pre-mRNA cleavage upon CPSF73 depletion. For example, MBLAC1, which when depleted has been shown to cause accumulation of unprocessed RDH transcripts (Pettinati et al, 2018). Another explanation is that remaining levels of CPSF73 after doxycycline and auxin treatment are sufficient for RDH pre-mRNA processing. However, due to the absence of a visible band on the western blot when cells were treated with doxycycline and auxin, I am confident that this is unlikely (Figure 5.1). In addition, the massive effects I observed of CPSF73 depletion on mRNA genes suggest a sufficient depletion of CPSF73 (Chapter 5).

Pettinati et al (2018) observed an approximate readthrough of 200 bp on multiple, but not all, RDH genes when CPSF73 was depleted by RNAi. To ensure this small readthrough effect had not been visually overlooked when analysing the RDH genes in cluster, I investigated several genes individually. Three of these genes (HIST1H3B, HIST1H4B and HIST1H2BC), which were shown to have a major readthrough effect upon MBLAC1 depletion and similar effect upon CPSF73 depletion in Pettinati et al (2018), are shown in Figure 6.5. In contrast to the findings of Pettinati et al (2018), no readthrough effect was observed on these RDH mRNAs when CPSF73 was depleted in our *CPSF73-AID* cells.

The differences between these two works may be due to the methodologies used. For example, Pettinati et al (2018) used siRNA in HeLa cells for depletion of CPSF73 compared to my use of the AID system in HCT116 cells. In addition, their study utilised cells synchronised in early S-phase during which RDH genes are rapidly transcribed and they specifically analysed chromatin associated RNA, in comparison to nuclear RNA extracted for my investigations. Therefore, readthrough RDH transcripts may not have been detected in this work due to their rapid turnover at the end of S phase of the cell cycle (Marzluff et al, 2008). However, defective processing of RDH pre-mRNA has been shown to cause their aberrant polyadenylation by use of a downstream PAS (Kari et al, 2013; Romeo et al, 2014). These polyadenylated transcripts are stable throughout the cell cycle and therefore more likely to be detected (Levine et al, 1987).

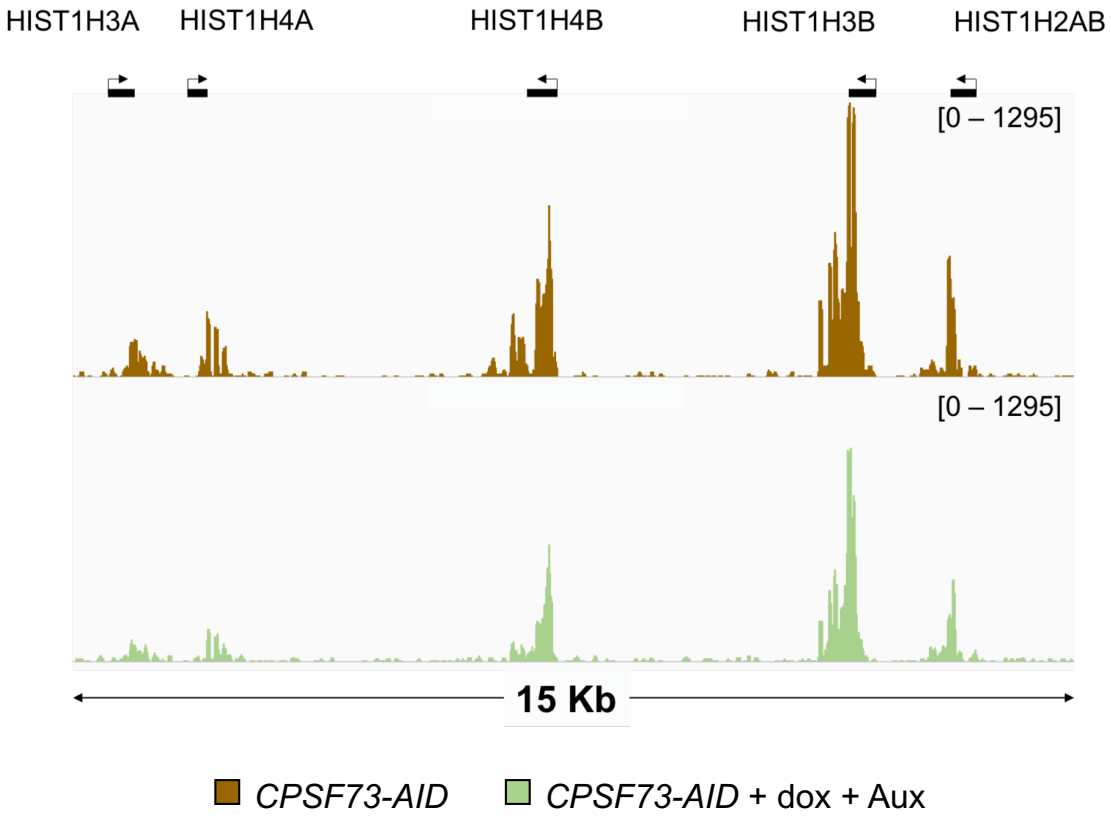

Figure 6.1 *CPSF73-AID* RPKM coverage track of a RDH gene cluster

RPKM normalised coverage track showing five RDH genes (HIST1H3A, HIST1H4A, HIST1H4B, HIST1H3B and HIST1H2AB) in *CPSF73-AID* cells with or without doxycycline and auxin treatment. No apparent differences are visualised. The numbers in brackets show the average RPKM normalised read count range. Figure is representative of one biological replicate, a second biological replicate is shown in Figure 6.2.

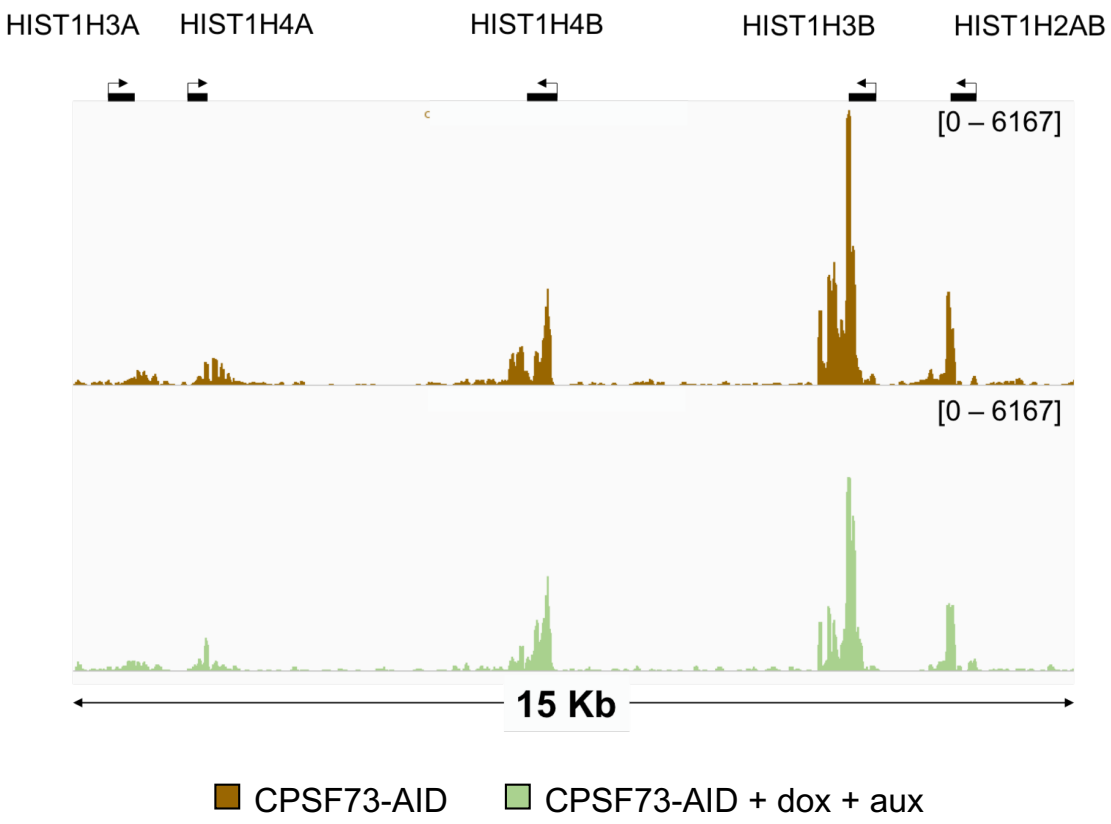

Figure 6.2 Second replicate *CPSF73-AID* RPKM coverage track of a RDH gene cluster

Second biological replicate for RPKM normalised coverage track showing five RDHs (HIST1H3A, HIST1H4A, HIST1H4B, HIST1H3B and HIST1H2AB) in *CPSF73-AID* cells with or without doxycycline and auxin treatment. No apparent differences are visualised. The numbers in brackets show the average RPKM normalised read count range.

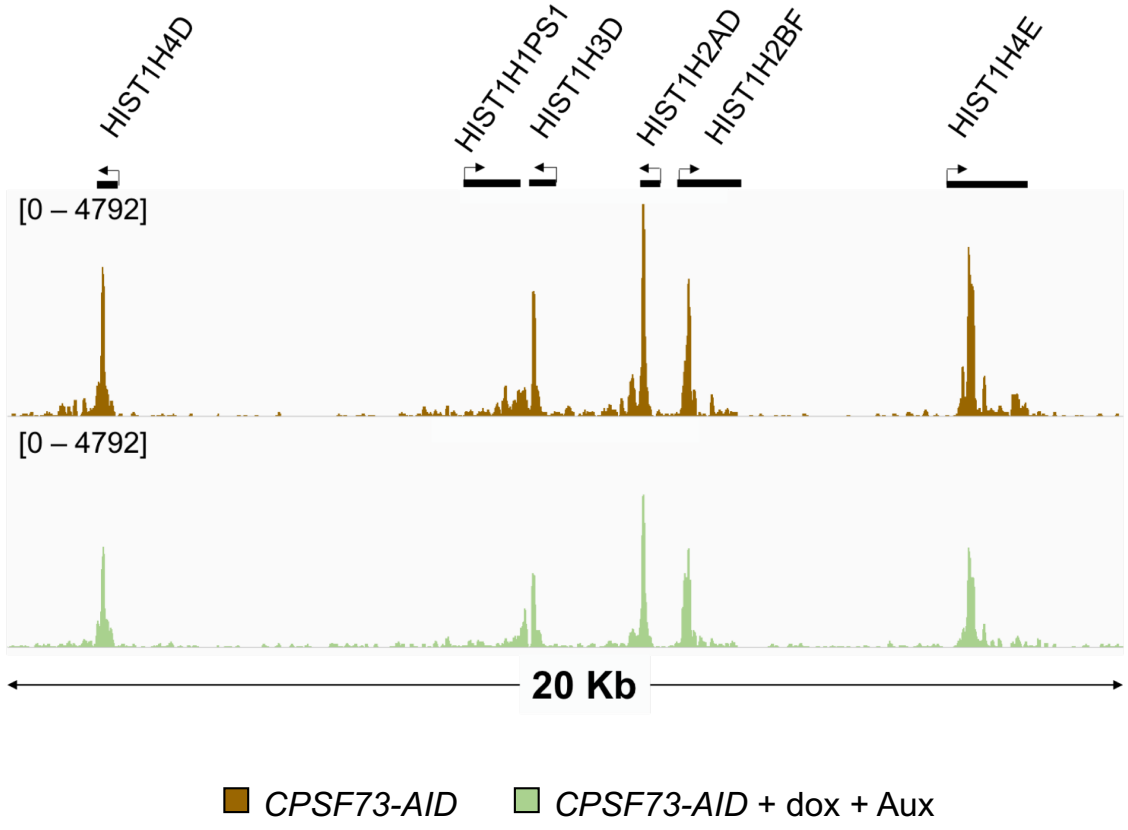

Figure 6.3 *CPSF73-AID* RPKM coverage track of a second RDH gene cluster

RPKM normalised coverage track showing six RDH genese (HIST1H4D, HIST1H1PS1, HIST1H3D, HIST1H2AD, HIST1H2BF, HIST1H4E) in *CPSF73-AID* cells with or without doxycycline and auxin treatment. No apparent differences are visualised. The numbers in brackets show the average RPKM normalised read count range. Figure is representative of one biological replicate, a second biological replicate is shown in Figure 6.4.

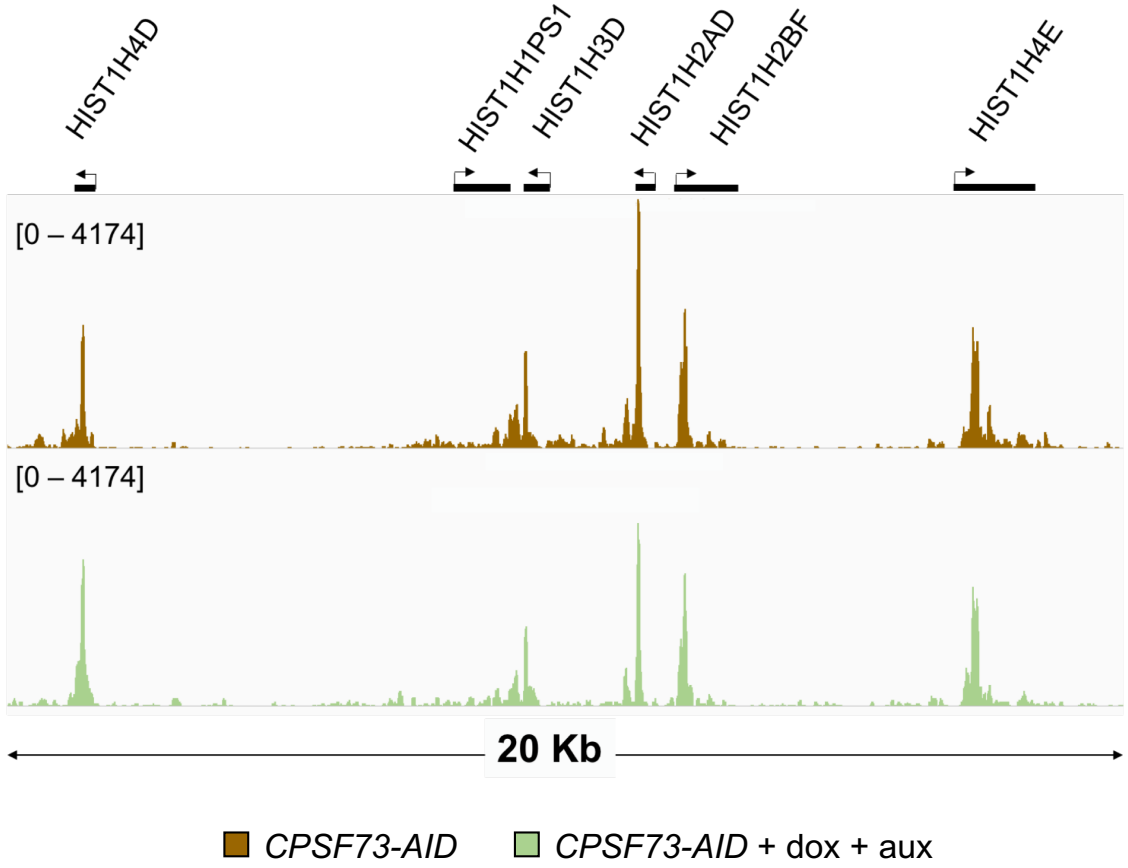

Figure 6.4 Second replicate *CPSF73-AID* RPKM coverage track of a second RDH gene cluster

Second biological replicate for RPKM normalised coverage track showing six RDHs (HIST1H4D, HIST1H1PS1, HIST1H3D, HIST1H2AD, HIST1H2BF, HIST1H4E) in *CPSF73-AID* cells with or without doxycycline and auxin treatment. No apparent differences are visualised. The numbers in brackets show the average RPKM normalised read count range.

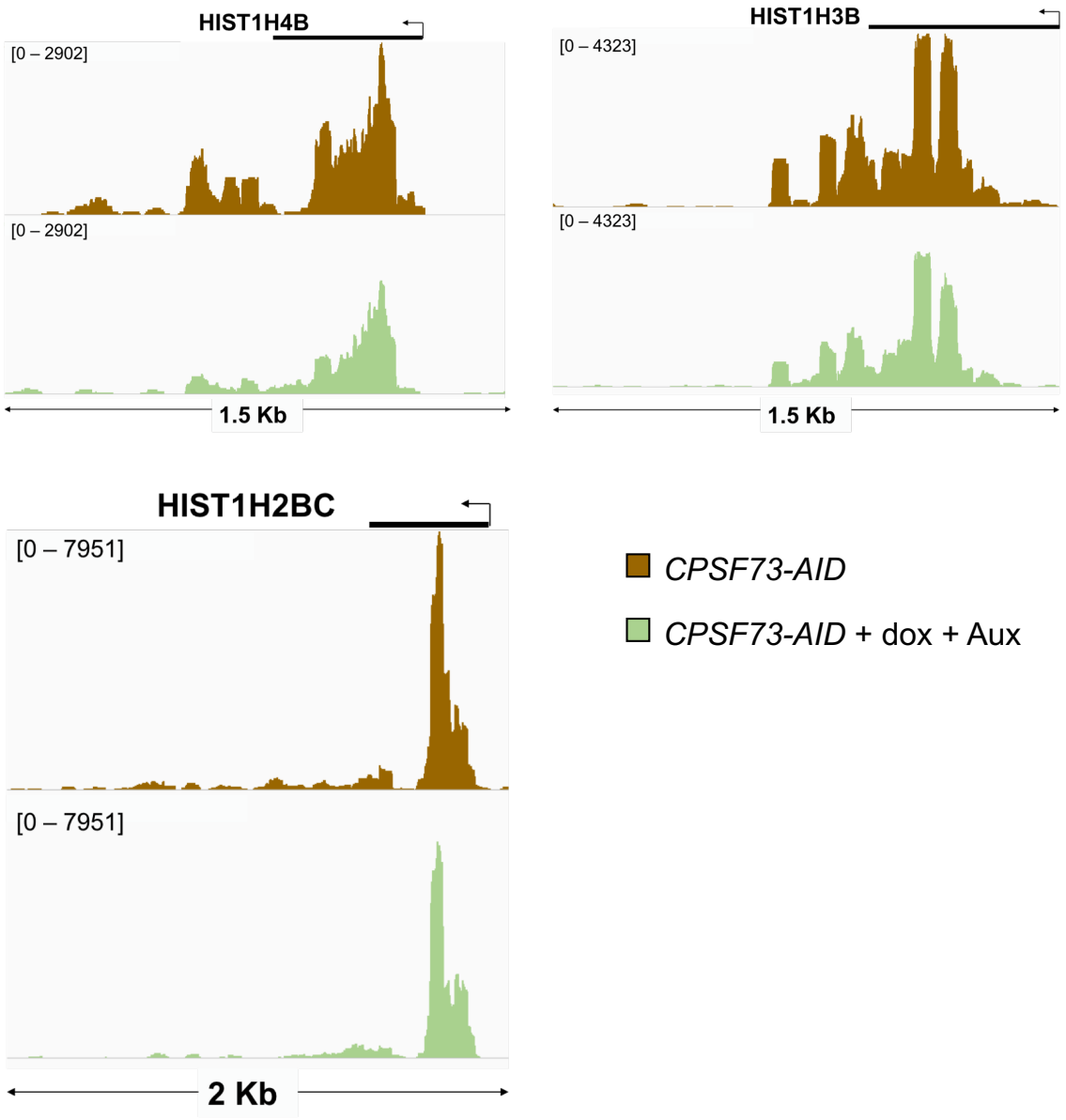

Figure 6.5 RPKM coverage tracks of individual RDHs in *CPSF73-AID* cells

A closer visualisation of RPKM normalised coverage tracks for HIST1H4B, HIST1H3B and HIST1H2BC in *CPSF73-AID* cells treated or not with doxycycline and auxin. No apparent differences are visualised. The numbers in brackets show the average RPKM normalised read count range. Figure represents one biological replicate, a second replicate is represented in Figure 6.6.

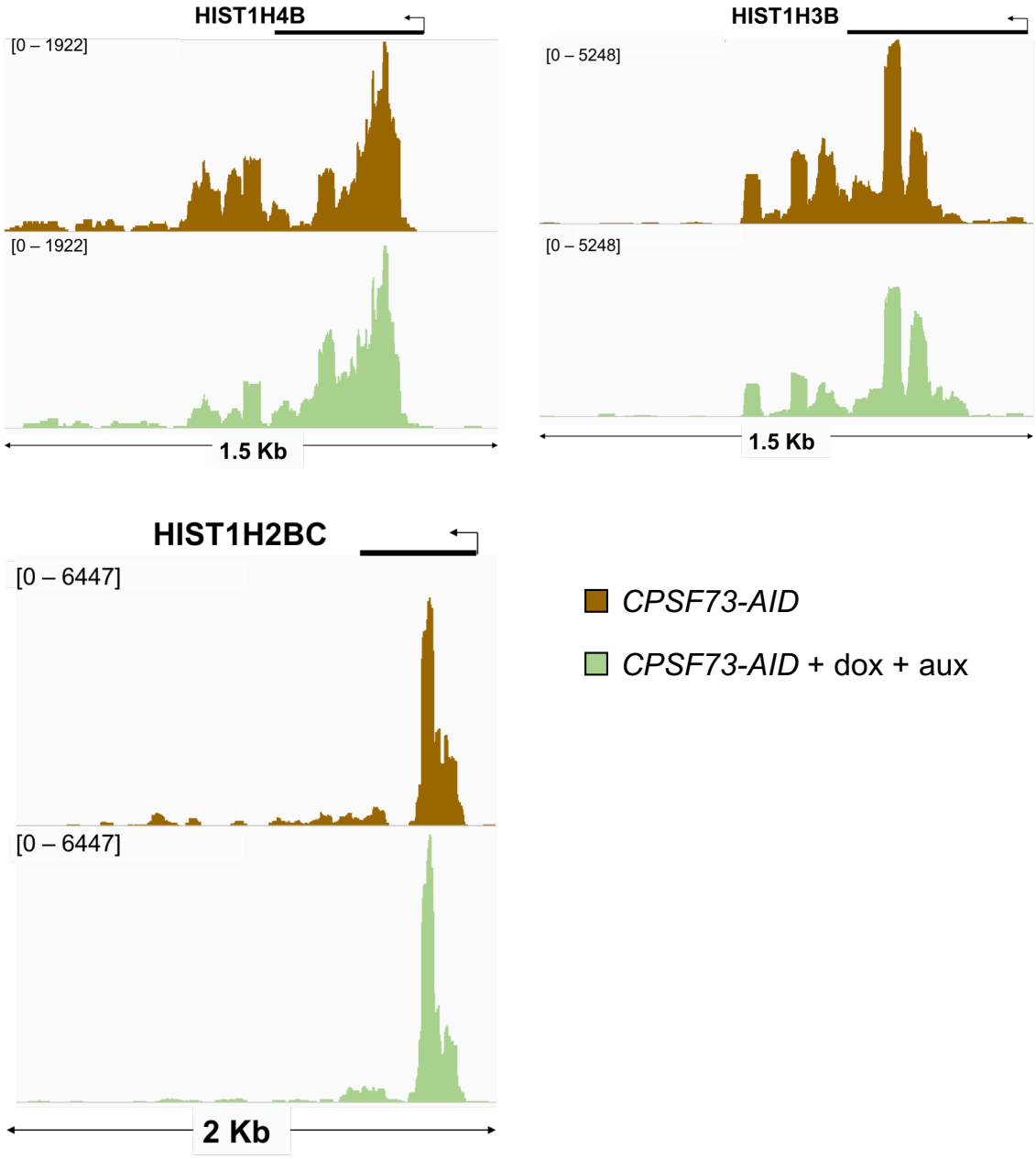

Figure 6.6 Second replicate RPKM coverage tracks of individual RDHs in *CPSF73-AID* cells

Second biological replicate of RPKM normalised coverage tracks for HIST1H4B, HIST1H3B and HIST1H2BC in *CPSF73-AID* cells treated or not with doxycycline and auxin. No apparent differences are visualised. The numbers in brackets show the average RPKM normalised read count range.

6.2 DIS3 depletion causes accumulation of RDH PROMPTs

As previously mentioned, the exosome has been found to degrade mature RDH transcripts. However, whether EXOSC10, DIS3 or both subunits provide this degradation activity is unclear. Therefore, I analysed the same cluster of histones as in Figure 6.1 using *DIS3-AID* RNA-Seq data (Figure 6.7 and 6.8). DIS3 depletion caused an increase in reads upstream of the TSS of some RDHs, including HIST1H4B and HIST1H2AB as highlighted by the arrows. These reads show PROMPT accumulation in the opposite transcription direction to the associated RDH gene, as confirmed by split strand visualisation (Figure 6.9). This demonstrates that PROMPTs can also derive from RDH genes and accumulate upon DIS3 depletion. Aside from PROMPT accumulation, no other changes in read levels were observed. Therefore, DIS3 may not be responsible for RDH mRNA degradation as an accumulation of mature RDH transcripts might have been expected if their degradation had been inhibited. Although DIS3 depletion appears to show no effects on RDH mRNA degradation, the exosome may still play a role. Instead, the other catalytic subunit EXOSC10 may be responsible for RDH degradation or may show redundancy to DIS3. Alternatively, 1 hour of DIS3 depletion may not have been sufficient for RDH transcript accumulation. S-phase, when RDH genes are rapidly transcribed, lasts for approximately 8 hours and RDHs are quickly degraded afterwards (Hahn et al, 2009; Harris et al, 1991). Therefore, effects on RDH transcripts observed upon DIS3 depletion may be dependent on the cell cycle phase.

6.3 Preventing U7 snRNA binding to the HDE of RDH genes causes defective RDH processing

The role of the Integrator in RDH processing is currently unclear, although both a direct and indirect effect have been postulated (Skaar et al, 2015; Ezzeddine et al, 2011). As the Integrator is responsible for proper processing of snRNAs as shown in the previous chapter, Integrator dysfunction would affect U7 snRNA processing and therefore could indirectly cause misprocessing of RDH transcripts. To specifically investigate the effects of Integrator dysfunction on RDHs through U7 snRNA misprocessing, I used an antisense morpholino oligonucleotide (AMO) to U7 snRNA.

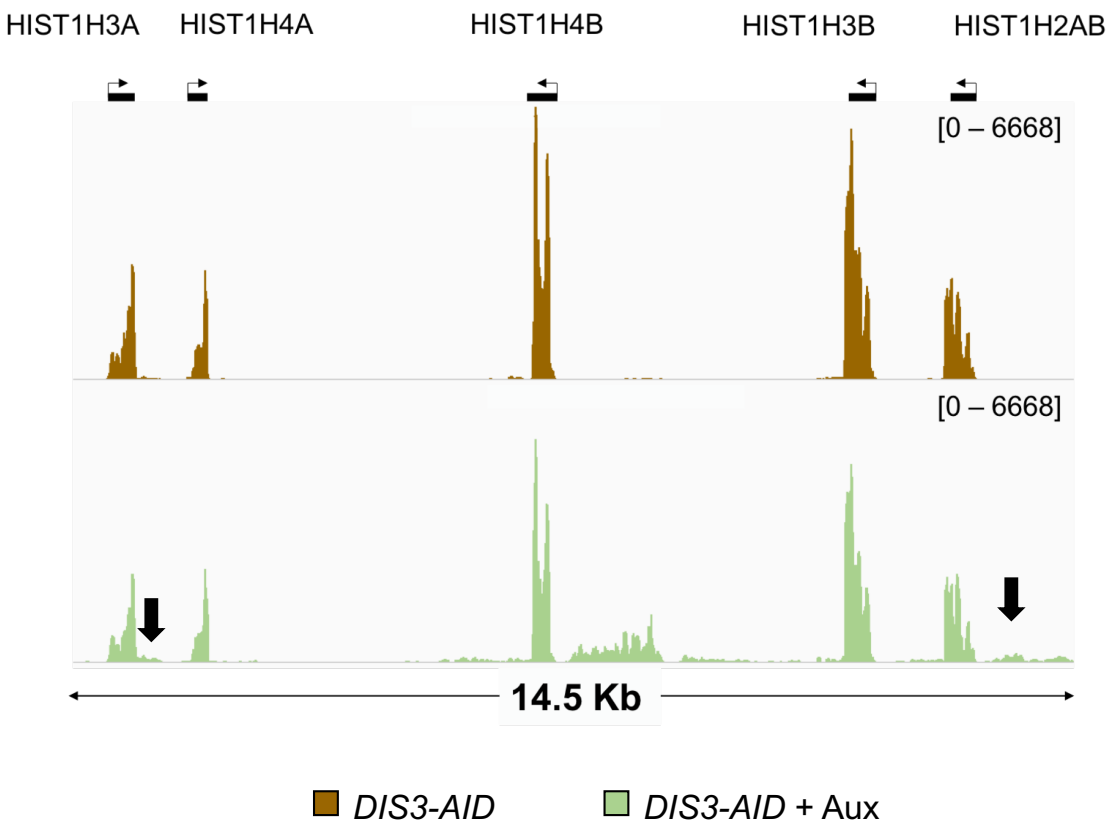

Figure 6.7 *DIS3-AID* RPKM coverage track of a RDH gene cluster

RPKM normalised coverage track showing five RDH genes (HIST1H3A, HIST1H4A, HIST1H4B, HIST1H3B and HIST1H2AB) in *DIS3-AID* cells with or without auxin treatment. The black arrows highlight reads corresponding to PROMPT transcription in the opposing direction to transcription of the associated RDH gene. The numbers in brackets show the average RPKM normalised read count range. Figure represents one biological replicate, a second biological replicate can be found in Figure 6.8.

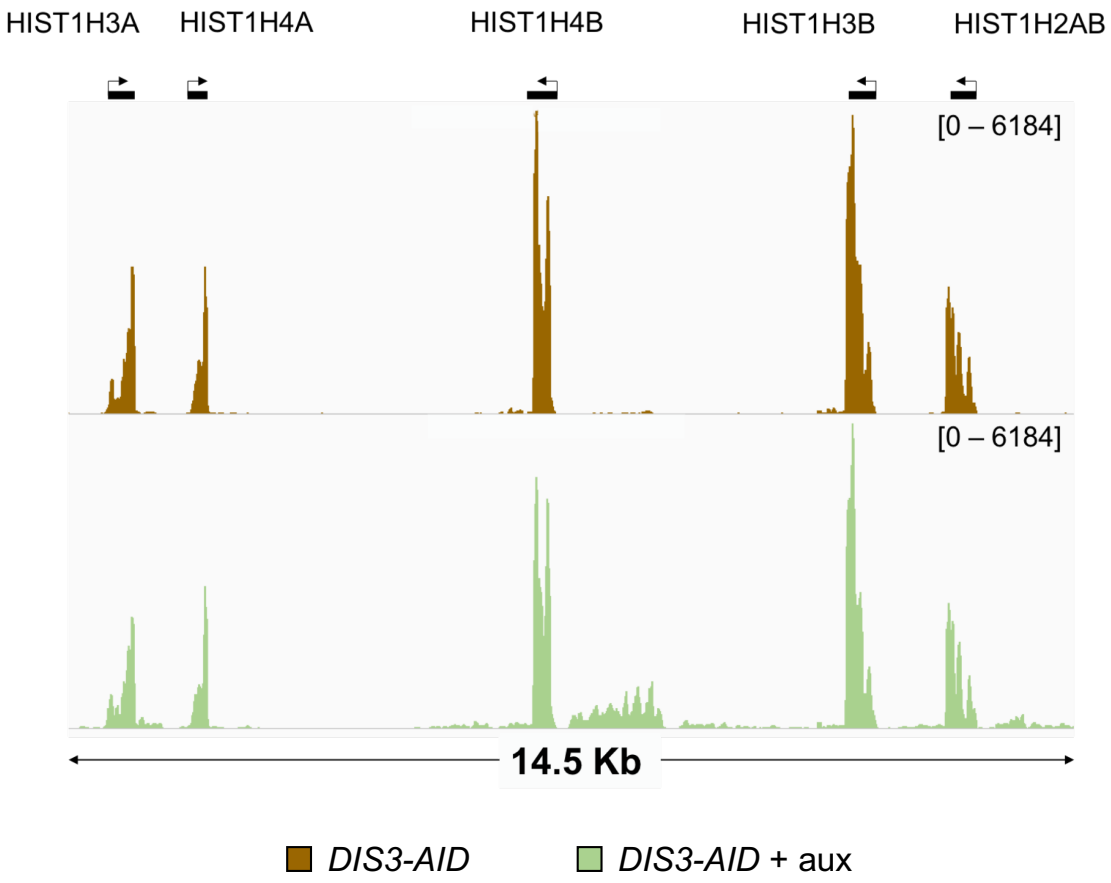

Figure 6.8 Second replicate *DIS3-AID* RPKM coverage track of a RDH gene cluster

Second biological replicate of RPKM normalised coverage track showing five RDHs (HIST1H3A, HIST1H4A, HIST1H4B, HIST1H3B and HIST1H2AB) in *DIS3-AID* cells with or without auxin treatment. The numbers in brackets show the average RPKM normalised read count range.

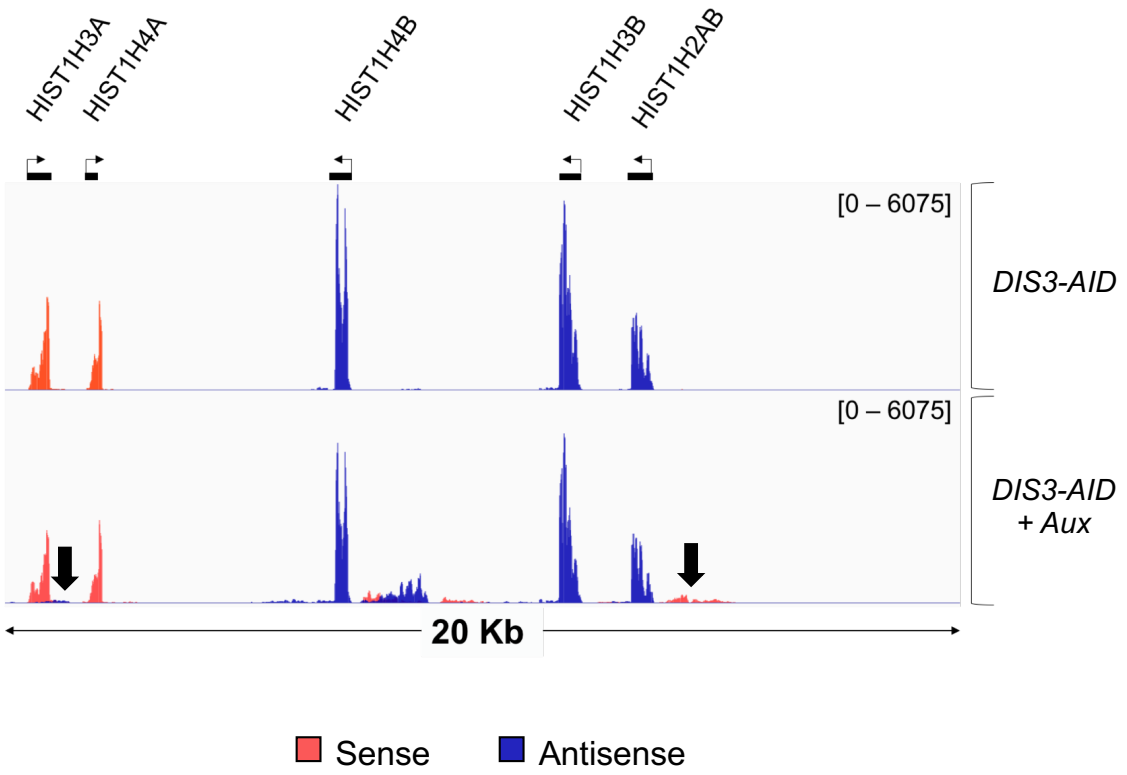

Figure 6.9 *DIS3-AID* RPKM split strand coverage track of a RDH gene cluster

RPKM normalised coverage track showing five RDHs (HIST1H3A, HIST1H4A, HIST1H4B, HIST1H3B and HIST1H2AB) in *DIS3-AID* cells (first replicate) with or without auxin treatment. Strands have been separated, with the sense strand in red and antisense strand in blue. The black arrows highlight reads corresponding to PROMPT transcription in the opposing direction to transcription of the associated RDH gene. The numbers in brackets show the average RPKM normalised read count range.

This allowed easy occlusion of U7 snRNP, thus inhibiting its function, in various cell lines and further analysis of the effects.

6.3.1 Occlusion of U7 snRNP causes extension of RDHs

After using a AMO to bind to and occlude U7 snRNP in *HCT116:TIR1* cells, I first examined the effects on RDH pre-mRNA processing. To do this, qRT-PCR was conducted to measure uncleaved and downstream RNA levels of two RDH genes, HIST1H4H and HIST1H3B (Figure 6.10A and 6.10B). Blocking of U7 snRNP binding caused a strong accumulation of uncleaved HIST1H4H and HIST1H3B in comparison to *HCT116:TIR1* cells transfected with a control AMO. Therefore, preventing U7 snRNP binding to the HDE causes impaired cleavage of RDH pre-mRNA. Lsm11, a subunit of U7 snRNP, normally forms a docking platform with FLASH for the HCC. The defective RDH cleavage upon U7 snRNA inhibition is therefore likely caused by impaired recruitment of the HCC (Yang et al, 2013; Burch et al, 2011).

To determine how far unprocessed RDH transcripts would extend past the TES when U7 snRNA was bound by AMO, RNA levels downstream of the TES were measured by qRT-PCR (Figure 6.10A and 6.10B). For HIST1H4H an increase in reads 150 bp downstream of the TES was observed, with a decline to background levels by 1 Kb. For HIST1H3B the observed readthrough was longer, with a significant increase in RNA levels at 2 Kb downstream. However, there were no significant differences by 3 Kb past the TES. These findings suggest that abrogating RDH pre-mRNA cleavage produces extended RDH transcripts which terminate relatively close to the TES and shows their extension is not finite.

Interestingly, this extension of RDH mRNA is similar to the extension of snRNAs observed upon INTS11 or INTS1 depletion (Figure 4.5 and 4.7). Both RDH and snRNA misprocessed transcripts show extension that is not finite, with termination occurring by 3 Kb downstream of the TES. Therefore, correct processing may not be required for termination of these extended transcripts.

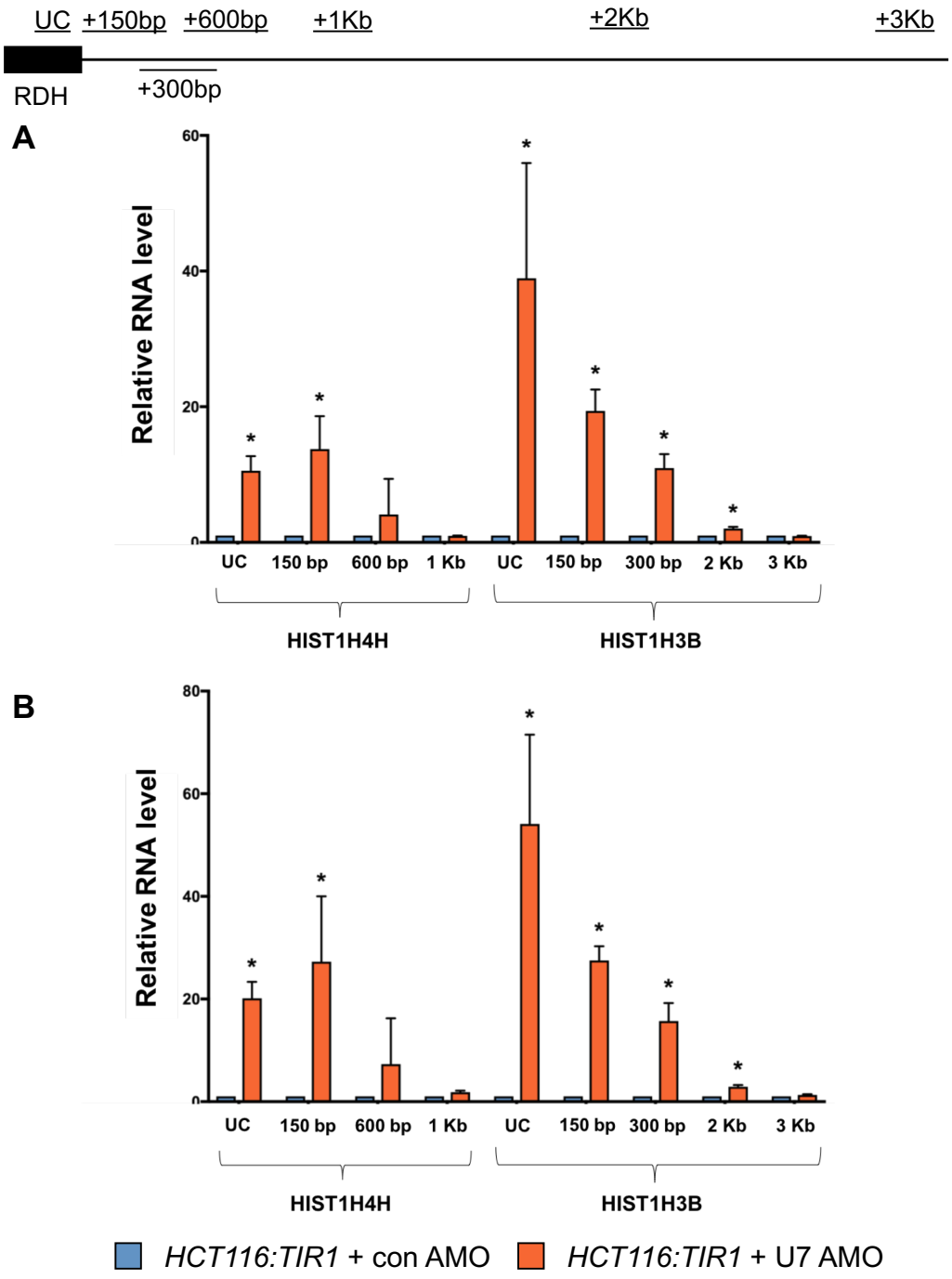

Figure 6.10 RNA levels downstream of RDHs after U7 snRNP depletion

qRT-PCR detection of RNA levels downstream of HIST1H4H or HIST1H3B in *HCT116:TIR1* cells treated with either control antisense morpholino oligonucleotide (AMO) or a U7 snRNA AMO. Levels of uncleaved (UC) histones were measured using primers homologous to the upstream (forward primer) and downstream (reverse primer) region of the TES. Quantitation of RNA is expressed as fold change relative to *HCT116:TIR1* cells with control AMO. Standard deviation is plotted by error bars, * denotes a p value < 0.05. A) Normalised to β actin. B) Normalised to gene body.

6.3.2 No significant differences in Pol II occupancy were found on RDH genes after blocking U7 snRNP binding

To determine whether using a U7 snRNP AMO affected RDH transcription rates, Pol II occupancy was analysed downstream of HIST1H4H and HIST1H3B genes by chromatin immunoprecipitation (ChIP) (Figure 6.11). This experiment was conducted in *HCT116:TIR1* cells electroporated with either a control AMO or U7 snRNA AMO. For HIST1H3B there were no significant differences in Pol II occupancy, with levels decreasing to near zero by 2 Kb. However, HIST1H4H showed a significant increase of Pol II occupancy of uncleaved transcripts when U7 snRNA binding was obstructed. This significant increase did not continue downstream of the TES although levels remained higher in the U7 snRNA AMO condition compared to controls. At 1 Kb downstream of the TES, Pol II occupancy had decreased to low levels when either U7 snRNA AMO was present or absent. These findings suggest that Pol II occupancy is unchanged on RDH genes when U7 snRNA binding is inhibited, although there was a slight increase over uncleaved HIST1H4H transcripts. Further experimental analysis would therefore be required at other RDH genes to determine the full effects of U7 snRNA on Pol II occupancy at RDH genes.

6.3.3 DIS3 depletion has no effect on RDH processing

DIS3-AID RNA-Seq data suggested DIS3 depletion had no effect on RDH transcription (Figure 6.7). To validate this finding, qRT-PCR was used to determine levels of two RDH transcripts and RNA levels downstream of their TES (Figure 6.12). Primers designed over the gene body (gb) of both HIST1H3B and HIST1H4H showed no differences in transcript levels of these histones upon DIS3 depletion. In fact, for HIST1H3B levels appeared to decrease although the result was not significant. This corroborates the RNA-Seq findings and suggests RDH mRNA may not be a substrate of DIS3 or that another nuclease shows redundancy when DIS3 is depleted. Additionally, DIS3 depletion had no effect on the levels of uncleaved RDHs, showing DIS3 does not affect RDH processing mechanisms. In further support, RNA levels downstream of the TES were not significantly different.

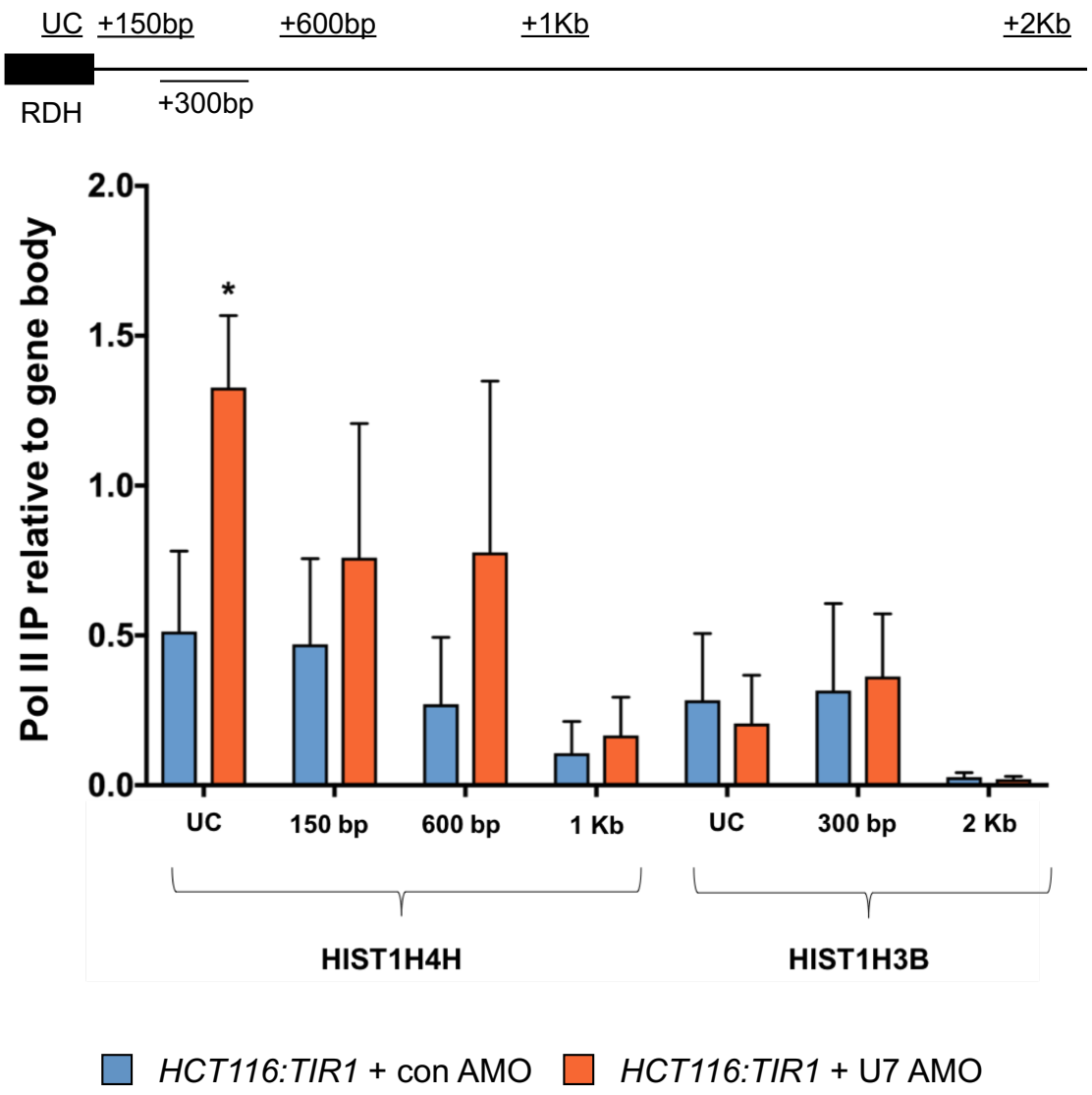

Figure 6.11 ChIP of RDHs in *HCT116:TIR1* cells with U7 snRNP depletion

ChIP results measuring the relative levels of Pol II occupation to gene body levels, of uncleaved (UC) and downstream regions of HIST1H4H and HIST1H3B RDH genes. Conducted in *HCT116:TIR1* cells treated with either a control AMO or a U7 snRNP AMO. Error bars plot standard deviation and * denotes a p value < 0.05. Data is the mean of three independent experiments with samples run in triplicate each time.

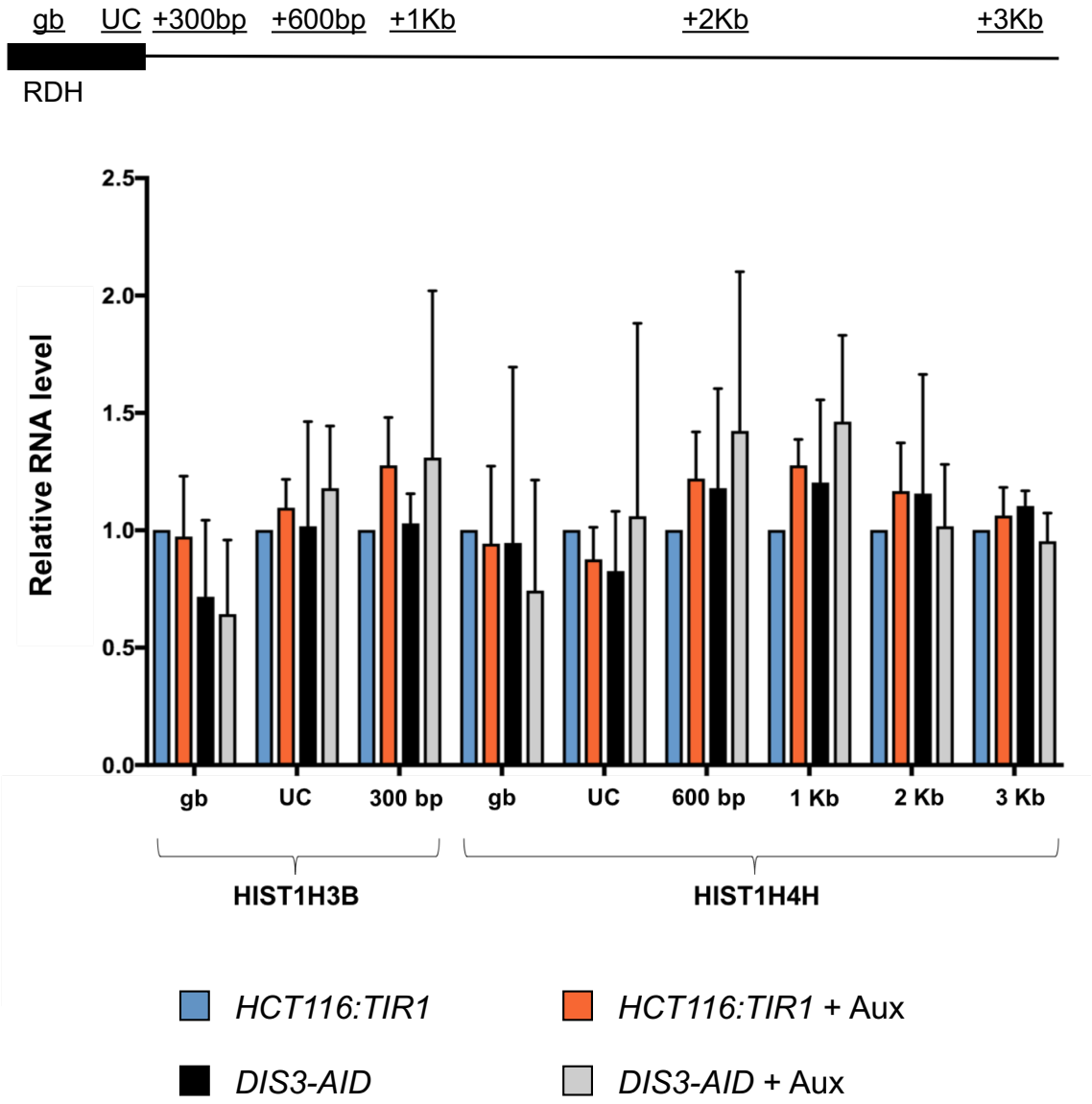

Figure 6.12 RNA levels downstream of RDHs in *DIS3-AID* cells

qRT-PCR detection of RNA levels over the gene body (gb), uncleaved (UC) or downstream regions of HIST1H3B and HIST1H4H. Conducted in *HCT116:TIR1* cells and *DIS3-AID* cells treated or not with auxin. Quantitation of RNA is expressed as fold change relative to *HCT116:TIR1* cells. All levels are normalised to β actin. Error bars plot standard deviation. Data is the mean of three independent experiments with samples run in triplicate each time.

6.3.4 DIS3 depletion and U7 snRNP occlusion has a cumulative effect on extended RDHs

Extended RDH transcripts have been previously reported in a CstF64 knockdown cell model (Romeo et al, 2014). These transcripts were also found to be degraded by the exosome. Therefore, I used *DIS3-AID* cells to determine if DIS3 could be responsible for degradation of extended RDHs produced by U7 snRNA occlusion. *DIS3-AID* cells were electroporated with either control or U7 snRNA AMO, before treatment or not with auxin for 2 hours. RNA levels downstream of the TES for HIST1H4H and HIST1H3B were determined by qRT-PCR (Figure 6.13A and 6.13B).

As seen previously, depletion of DIS3 by auxin addition had no effect on RDH mRNA processing. Similar to the results in *HCT116:TIR1* cells, U7 snRNA AMO in *DIS3-AID* cells produced an increase in uncleaved histones which showed extension that terminated by 1 – 2 Kb downstream of the TES. Interestingly, when U7 snRNA binding was inhibited and DIS3 depleted together a cumulative effect could be seen as an even bigger increase in uncleaved and extended RDH transcripts. The extension length of these transcripts was only slightly, if at all, increased by DIS3 depletion. Due to this and having observed no effect of DIS3 on RDH processing previously in this work, it is likely that the accumulation of extended RDH transcripts is due to loss of their degradation by DIS3.

From the data it is currently unclear whether DIS3 degrades mature RDH RNA. However, from these findings it can be concluded that DIS3 is able to degrade unprocessed extended RDHs. It is possible that DIS3 degrades both mature and extended transcripts and that another nuclease shows redundancy for mature RDH degradation or that 1 hour of DIS3 depletion is not enough for mature RDH accumulation. Alternatively, DIS3 may only degrade the extended RDH transcripts and this discrepancy may be related to polyadenylation. Mature RDHs are not polyadenylated whereas it has been shown that unprocessed RDHs become polyadenylated by use of downstream PASs (Kari et al, 2013; Romeo et al, 2014; Sullivan et al, 2009). It is not understood how DIS3 recognises targets for degradation, but polyadenylation may be involved in some way.

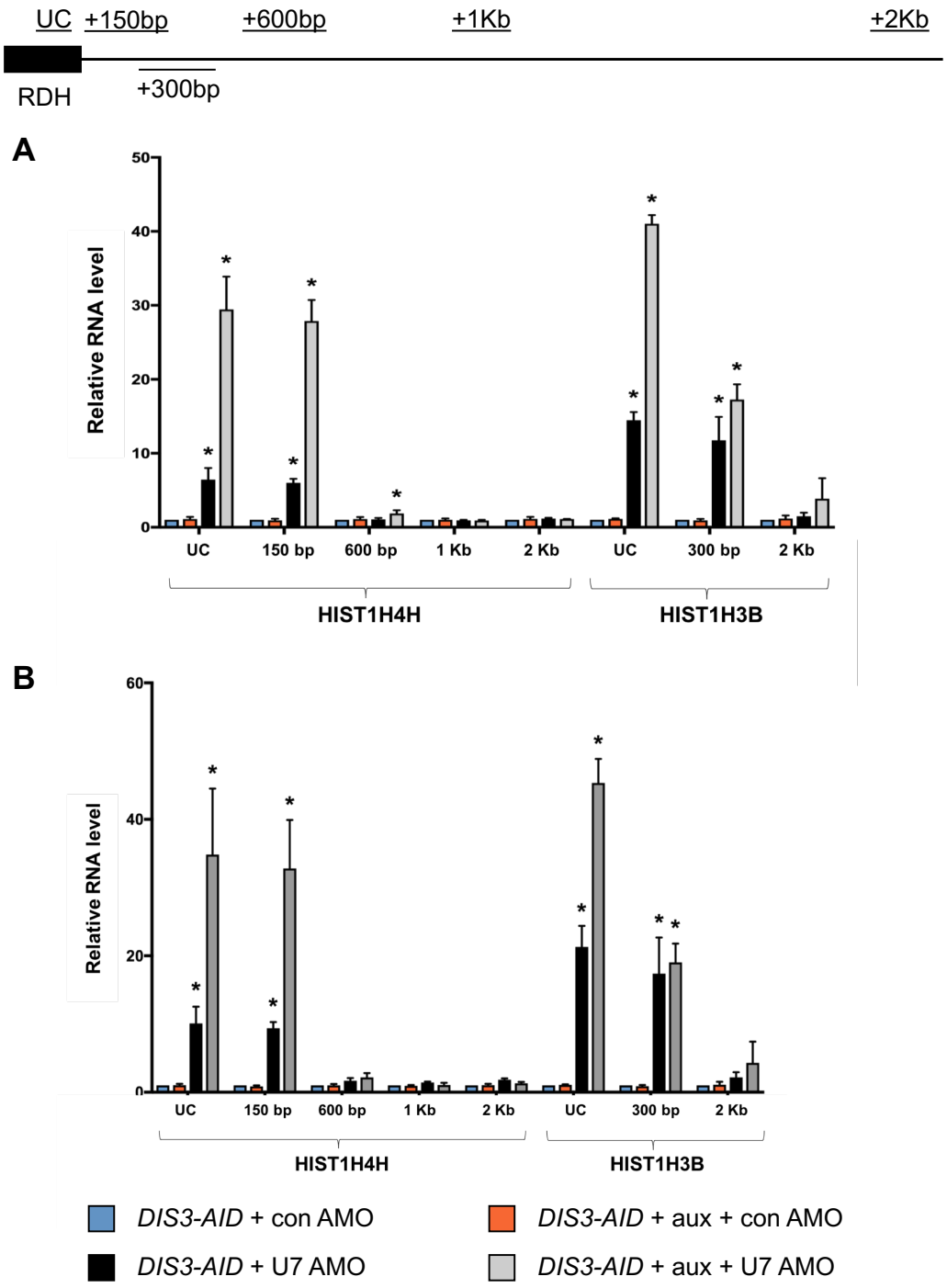

Figure 6.13 RNA levels downstream of RDHs with DIS3 depletion and U7 snRNP occlusion

qRT-PCR detection of uncleaved (UC) and extended HIST1H4H and HIST1H3B transcripts in *DIS3-AID* cells electroporated with either control AMO or U7 snRNA AMO and treated or not with auxin. Quantitation of RNA is expressed as fold change relative to non-depleted Dis3-AID cells with control AMO. * denotes $p < 0.05$, error bars plot standard deviation. Data is the mean of three independent experiments with samples run in triplicate each time. A) Normalised to β actin. B) Normalised to gene body.

The accumulative effect produced by DIS3 depletion was also observed for snRNAs, which showed increased levels of unprocessed transcripts upon DIS3 and INTS1 depletion (Figure 4.13). In similarity to RDHs, DIS3 depletion did not alter the length of these extended snRNA transcripts. This demonstrates the major role of DIS3 in degrading misprocessed transcripts from a variety of genes.

6.4 Summary

Firstly, in this chapter I have shown that conditional depletion of CPSF73 in HCT116 cells appears to have no transcriptional effect on RDH genes as inferred from RNA-Seq analysis. Two clusters of RDH transcripts were analysed, to help rule out the possibility of CPSF73 acting on only a subset of RDHs (Figure 6.1 and 6.3). As it had been expected that CPSF73 depletion would cause misprocessing and readthrough of RDH transcripts, it is possible another endonuclease may show redundancy. One candidate for this is the endonuclease MBLAC1, which was shown to selectively target RDH pre-mRNA for processing and abrogate cell cycle progression upon its depletion (Pettinati et al, 2018). From the same study it was shown that either MBLAC1 or CPSF73 depletion caused readthrough at numerous RDH genes of approximately 200 bp in length. Therefore, I analysed three of the same genes used in the study that had shown “major” misprocessing. However, no effect was observed in this work (Figure 6.5). These differences may be due to variances in methodology, such as using unsynchronised cells as RDHs are produced and degraded during S phase (Marzluff et al, 2008). However, U7 snRNP AMO experiments described here were also conducted on unsynchronised cells and showed extension of RDH transcripts, suggesting that the lack of CPSF73 effect is unlikely to be caused by cell phase synchronisation differences. Additionally I cannot rule out the possibility of an incomplete depletion of CPSF73 in our cell line, although the western blot data and strong effect on mRNA transcripts would argue against this.

Secondly, DIS3 depletion caused accumulation of RDH associated PROMPTs but had no effect on transcript levels or processing of RDH pre-mRNA (Figure 6.7 and 6.12). This finding suggests DIS3 may not be responsible for RDH mRNA degradation but doesn’t rule out the exosome entirely, as EXOSC10

may play a role. However a reason for no changes in RDH transcript levels may be due to the small depletion time of Dis3 at only 1 hour, as levels of RDHs are tightly regulated by the cell cycle.

On the other hand, inhibition of U7 snRNA binding to RDH genes with an AMO was able to disrupt the normal transcription of RDHs. U7 snRNP is responsible for recruitment of the HCC. Interestingly, as the Integrator is responsible for correct production of U7 snRNA, Integrator dysfunction may have an indirect effect on RDHs. Unfortunately, RDH transcripts were not fully detectable in the Ints11 RNA-Seq data and therefore I was unable to look into these effects more directly. In this work, use of a U7 snRNA AMO resulted in accumulation of extended RDH transcripts (Figure 6.10). This misprocessing of RDH pre-mRNA did not have finite extension, with extended transcripts terminating by approximately 2 Kb downstream of the TES. However, Pol II occupancy levels appeared unchanged on two RDH transcripts (Figure 6.11). One RDH transcript, HIST1H4H, did show a significant increase in Pol II occupancy over the TES. This significant change did not continue further downstream although levels remained higher in the U7 snRNA AMO condition compared to controls. Therefore, the investigation of Pol II occupancy at further RDH transcripts is required to elucidate the full effects of U7 snRNA occlusion on transcription levels of RDHs.

Finally, although DIS3 reduction appeared to have no effect on RDH transcription, it was able to have an accumulative effect with U7 snRNA AMO on RDH extended transcript levels, with little effect on extension length (Figure 6.13). Although it is not clear whether DIS3 can degrade mature RDH transcripts, as I showed DIS3 had no effect on RDH pre-mRNA processing, this data suggests DIS3 is responsible for degrading unprocessed RDH transcripts. If DIS3 specifically degrades extended RDHs, a possible mechanism for differentiating between them and properly processed RDHs may be attributed to differences in polyadenylation.

From the data in chapters 2 and 4, there are a few similarities between RDH and snRNA misprocessed transcripts. For example, both RDHs and snRNAs do not appear to require cleavage for termination of extended transcripts (Figure 6.10 and 4.5). Additionally, DIS3 depletion causes a further increase in levels of RDH and snRNA extended transcripts, with either occlusion of U7

snRNP or Integrator dysfunction respectively (Figure 6.13 and 4.13). However, DIS3 depletion alone was capable of causing an increase in misprocessed snRNAs whereas there was no apparent effect on RDH transcripts. It is possible that misprocessing commonly occurs at snRNAs and DIS3 is responsible for degrading such transcripts. Therefore upon DIS3 depletion, misprocessed transcripts accumulate. Whereas processing of RDH pre-mRNA could be more tightly regulated or controlled, reducing the number of misprocessed transcripts in normal conditions. 14 different subunits have been found in the Integrator complex and RNAi depletion of nearly any subunit was found to disrupt snRNA processing in Drosophila (Ezzedine et al, 2011). This suggests Integrator subunit interactions are highly sensitive to disruption and could explain common misprocessing of snRNAs. Alternatively, DIS3 may have a role in snRNA processing although there is no current evidence for this.

7. Discussion

Endonucleases play an important role in the maintenance of the RNA environment, as well as the processing and termination of different RNA species. In particular, work has shown that the catalytic subunit of the exosome, DIS3, is important for degradation of a multitude of RNA species including those that currently don't have a known role and may be a bi-product of efficient transcription, such as PROMPTs (Mitchell et al, 2014; Preker et al 2011). Another endonuclease, CPSF73, is known to have an important role in 3' end cleavage of protein-coding mRNA and has been linked to efficient transcriptional termination (Proudfoot et al, 2011; Fusby et al, 2016; Eaton et al, 2018). Additionally, the catalytic subunit of the Integrator, INTS11, has an important role in 3' end cleavage of snRNAs known to form the spliceosome (Baillat et al, 2005). However a lot is still unknown about the substrates of these endonucleases, as well as their exact roles and contributions to maintaining the RNA environment.

Previous work to elucidate endonuclease roles in transcriptional termination and RNA metabolism have been limited by the available methodology such as RNAi, which can be slow, have indirect effects and incomplete levels of gene downregulation. In this work I have utilised the AID and SMASh systems to generate cell lines capable of conditional, rapid and specific target protein depletion of either DIS3, INTS11 or CPSF73. I was then able to investigate the roles of the target endonuclease by analysing RNA effects after protein depletion, through RNA-Seq. In so doing, this work has provided a broad view of the immediate substrates for these three endonucleases in human cells. Additionally, the results from RNA-Seq of CPSF73-AID cells has provided further support for the importance of CPSF73 cleavage of protein-coding mRNA in transcription termination, giving support to the torpedo model of transcriptional termination.

7.1 Rapid and conditional protein depletion

The auxin inducible degron system was first discovered in plants, whereby it is used to mediate gene expression to regulate plant growth and development (Dharmasiri et al, 2005). The AID system utilises the plant specific F-box protein, TIR1, which forms a E3 ubiquitin ligase complex with SCF. As the SCF complex is also expressed in non-plant species, through the expression of TIR1 in human

cells, the AID system can be employed for ubiquitin-mediated proteome degradation and as a genomic manipulation tool (Nishimura et al, 2009; Holland et al, 2012). Due to developments in CRISPR/Cas9 technologies, it is now possible to integrate the AID tag into human target genes and thus allow their conditional protein depletion (Natsume et al, 2016).

In this thesis I have shown that incorporation of an AID tag to a gene of interest, in human cells expressing TIR1 protein, results in a rapid, specific and inducible depletion of the target protein. Protein depletion occurs upon auxin addition and utilising this system I was able to significantly decrease DIS3 levels in *DIS3-AID* cells after 1 hour of auxin treatment (Figure 3.3) (Davidson et al, 2019). Importantly, only AID-tagged protein levels are affected by auxin addition and this requires TIR1 expression, as shown by auxin having no effect on untagged protein levels nor when TIR1 is not expressed (Figure 3.3 and 5.1). Furthermore, auxin does not affect cell survival of untagged cells and AID-tagging the DIS3 protein also had no effect on cell survival, although a slower growth phenotype was observed (Figure 3.5).

There are several reasons why the AID system may be more beneficial that RNAi. Firstly, as previously mentioned, the AID system has a faster rate of protein depletion than RNAi methods. Secondly, RNAi methods are known to have limitations due to off-target effects. Using connectivity maps, Smith et al (2017) found that RNAi protein depletion resulted in stronger and more pervasive off-targets than generally appreciated, whereas off-target effects from CRISPR methods were negligible. Additionally, RNAi methods have been found to produce false negative results or reduced phenotypes due to incomplete depletion of the target protein (Eaton et al, 2018). Finally, as shown by RNA-Seq on my *DIS3-AID* cells, the AID system was capable of elucidating more RNA targets than previous RNAi experiments (Szczepinska et al, 2015). Overall the AID system may be able to enhance our knowledge of specific gene targets and functions, through easier investigation of immediate depletion effects, that have so far been unnoticed by RNAi.

7.2 DIS3 is responsible for degradation of a multitude of RNA transcripts

DIS3 is a major catalytic component of the exosome, consisting of both exonuclease and endonuclease activity (Schaeffer et al, 2009; Schneider et al, 2009). EXOSC10 is the other catalytic subunit of the exosome and it has been proposed that both EXOSC10 and DIS3 degrade separate RNA transcripts. For example, DIS3 has been shown to degrade numerous unstructured RNA transcripts whereas EXOSC10 may specifically degrade smaller RNAs including pre-rRNA and snoRNAs (Szczepinska et al, 2015; Januszyk et al, 2011). As part of the exosome, these two proteins may also show redundancy for each other and as such it has been difficult to elucidate and categorise specific substrates for either nuclease. In an attempt to overcome such issues, I utilised the AID system to enable rapid depletion of DIS3. This allowed investigation of the immediate effects of DIS3 loss, potentially before activation of redundancy pathways that could be mediated by EXOSC10.

Through RNA-Seq analysis, I was able to identify accumulation of numerous RNA targets upon DIS3 depletion (Chapter 3). These included PROMPT RNAs that derived from bidirectional transcription at protein-coding gene promoters and RDH promoters (Figure 3.6 and 6.7). Levels of short RNAs derived from premature transcription termination also accumulated upon DIS3 loss. Additionally, an upregulation of *de novo* transcripts from intergenic transcriptome regions where bidirectional transcription occurs and were similar to eRNAs, was observed. This experiment detected more DIS3 dependent accumulation of potential eRNAs and novel transcripts than previously reported (Szczepinska et al, 2015). The diverse range of DIS3 sensitive RNA substrates shown throughout this work supports previous suggestions that DIS3 is responsible for the majority of RNA degradation by the exosome (Dziembowski et al, 2007; Szczepinska et al, 2015).

Interestingly, DIS3 may not be responsible for RDH mRNA degradation. DIS3 depletion had no effect on mature RDH levels but instead had a cumulative effect with U7 snRNA occlusion on extended RDH transcript levels (Figure 6.12 and 6.13). Therefore, DIS3 may specifically degrade misprocessed RDH transcripts. These findings do not necessarily mean that the exosome is not involved in mature RDH degradation however. Instead, EXOSC10 may be the nuclease responsible for exosome mediated RDH mRNA degradation. Andersen

et al (2013) found mature RDH transcripts accumulate upon depletion of a core exosome subunit, RRP40. Similarly, depletion of another core exosome subunit, RRP41 or EXOSC10 slowed histone mRNA degradation (Mullen and Marzluff, 2008). Furthermore, data from Slevin et al (2014) suggests mature RDHs are degraded in two phases. Firstly, degradation into the stem loop by the 3' – 5' exonuclease 3'hExo, resulting in the formation of a degradation intermediate. Secondly, the intermediate is degraded by the exosome containing EXOSC10. These findings suggest the exosome does have a role in degradation of mature RDHs and therefore EXOSC10 may be responsible instead of DIS3.

7.3 The role of DIS3 in snRNA transcription and degradation

Through RNA-Seq analysis and qRT-PCR I was able to show that DIS3 plays a role in snRNA metabolism. Depletion of DIS3 alone resulted in an accumulation of extended snRNAs (Figure 4.9 and 4.10). An explanation for this would be a role of DIS3 in snRNA 3' end processing; however I believe this is unlikely. There are no previous reports showing evidence of a role for DIS3 in 3' end snRNA processing, except for in budding yeast (Allmang et al, 1999). Additionally, the average snRNA extension effect observed upon DIS3 depletion (approximately 500 bp downstream) is not as pronounced as when INTS11 is depleted (approximately 1 Kb downstream) (Figure 4.10 and Figure 4.5). Although there is an increased accumulation of extended snRNA transcripts upon DIS3 and INTS1 depletion together, compared to either alone, the length of extension is unaltered (Figure 4.13). Instead I hypothesise DIS3 has a major role in degradation of snRNA precursor or misprocessed transcripts. This is supported by the data showing a greater accumulation of precursor snRNAs upon DIS3 depletion when transcription is inhibited by actinomycin D, compared to the presence of DIS3 (Figure 4.12). Depleting DIS3 causes a reduction in snRNA precursor degradation, resulting in their increased levels.

This does not fully explain why extended snRNAs are apparent upon DIS3 depletion. I propose that the Integrator is not fully efficient at cleaving snRNA at their TES and that due to this, extended snRNAs are commonly generated. DIS3 is responsible for degradation of these misprocessed transcripts, hence why they aren't normally visible in the cell and their accumulation is observed upon DIS3

depletion. Rapid degradation of misprocessed snRNAs is important to prevent them entering the normal snRNA biogenesis pathways and sequestering mechanisms from correctly processed snRNAs. It is possible that the Integrator may still cleave extended snRNAs upon DIS3 depletion, just further downstream than at mature snRNAs. Whereas upon INTS11 depletion, the Integrator cannot cleave the snRNA and so readthrough is longer. This might explain why the observed extended transcripts are on average shorter upon DIS3 depletion than INTS11 depletion.

In support of this hypothesis is the work by Labno et al (2016), who investigated the role of DIS3L2. DIS3L2 has a similar sequence to DIS3, however it lacks the PIN domain, is not known to be a subunit of a macromolecular complex and localises to the cytoplasm where it degrades RNA in an exosome-independent manner (Lubas et al, 2013). Labno et al (2016) generated HEK293T cell lines expressing shRNAs that were capable of silencing endogenous DIS3L2 and either had inducible expression of WT DIS3L2 or a catalytically dead mutant. Upon DIS3L2 dysfunction there was an accumulation of cytoplasmic extended snRNAs, without a change in mature snRNA levels. As DIS3L2 is not present in the nucleus, it is likely that it has a major role as a surveillance pathway for cytoplasmic misprocessed precursors. This and the presence of extended snRNAs in the cytoplasm without Integrator dysfunction, gives support to misprocessing of snRNAs being a common occurrence. If snRNA readthrough is frequent, the cell may have adapted mechanisms to ensure termination of misprocessed snRNAs and might help explain why termination still occurs close to the TES upon Integrator subunit depletion.

7.4 How does DIS3 recognise target substrates for degradation?

How DIS3 recognises specific RNA substrates and prevents accumulation of aberrant mRNAs and potentially toxic protein products, is currently unclear. This is an important question when considering transcripts that undergo the same processing steps and have a similar structure to mature RNAs, for example PROMPTs which have a 5’ cap and are polyadenylated (Preker et al, 2011). As previously described in Chapter 1, human cells contain a NEXT complex that has been shown to promote degradation of PROMPTs and 3’ extended RNAs and a

PAXT complex that promotes degradation of long poly(A) tailed transcripts (Lubas et al, 2011; Tseng et al, 2015; Hrossova et al, 2015; Meola et al, 2016). Both of these complexes contain the RNA helicase MTR4, that acts as a complex scaffold and can also associate with EXOSC10 (Lubas et al, 2011; Meola et al 2016). Therefore, through MTR4 mediated exosome interaction with the PAXT or NEXT complex, the exosome may be targeted to specific RNA substrates. Specifically, DIS3 may be targeted to prematurely terminated transcripts by PAXT or NEXT, which associate with ARS2 (Anderson et al, 2013; Meola et al, 2016). ARS2 further associates with the CBC (to form CBCA) to recruit protein complexes involved in 3' end processing, maturation and degradation (Gruber et al, 2009; Hallais et al, 2013; Andersen et al, 2013). Iasillo et al (2017) found pervasive transcript turnover was supported by ARS2 function and that ARS2 also had a role in termination downstream of short snRNAs, RDHs, PROMPTs and eRNAs. Depletion of CBCA components ARS2 and CBP80, resulted in accumulation of 3' extended RDH transcripts and PROMPTs. Polyadenylated, longer replication-independent histone gene levels were not significantly altered by this depletion (Andersen et al, 2013). Andersen et al (2013) also demonstrated a physical link between the NEXT complex and CBCA and therefore suggested a link from the cap to the exosome. These findings demonstrate that exosome interactions with associated accessory factors may mediate exosome target specificity.

In addition, the work presented in this thesis suggested DIS3 may specifically degrade extended RDH transcripts and not mature RDHs (Figure 6.7, 6.12 and 6.13). As mature RDHs are not polyadenylated but extended RDH transcripts are, it is possible that hyperadenylation may induce DIS3-mediated decay (Narita et al, 2007; Romeo et al, 2014). Misprocessing of snRNAs has also been shown to result in their polyadenylation (Skaar et al, 2015; Yamamoto et al, 2014). Interestingly, Bresson and Conrad (2013) suggested that the nuclear poly(A) binding protein promotes hyperadenylation and decay of unstable transcripts. In support of our findings, Romeo et al (2014) observed exosome mediated degradation of extended polyadenylated RDH transcripts. As many histone genes contain PASs downstream of their cleavage sites, they suggested that polyadenylation of misprocessed RDH transcripts might be a mechanism to prevent readthrough into neighbouring RDH genes. Due to the closely clustered

location of RDH genes, readthrough could generate deleterious fusion proteins. Polyadenylation may therefore be a signal for misprocessed transcript degradation.

It has also been shown that PROMPT poly(A) signals are functional but linked to degradation (Ntini et al, 2013). There is a higher amount of PASs located upstream of protein-coding gene promoters than downstream, allowing efficient Pol II progression along the protein coding gene and helping to prevent bidirectional transcription (Ntini et al, 2013). Interestingly mRNA-like PASs are also frequently found downstream of snRNA and RDH genes (Almada et al, 2013). It is possible that polyadenylation of small transcripts such as RDHs, snRNAs and PROMPTs, which all showed DIS3 sensitivity in this work, results in their termination, whereas polyadenylation of longer transcripts, such as mRNAs, increases stability. Ntini et al (2013) found that promoter-proximal PASs more efficiently couple to exosome mediated degradation than RNAs with longer transcription units. Additionally, Hallais et al (2013) suggested that cap-proximal PASs processed by CBCA lead to RNA degradation through recruitment of NEXT and the exosome. As PROMPTs, snRNAs and RDH transcripts are all relatively short, this might explain how their polyadenylation could link to their efficient degradation. As previously mentioned, ARS2 may help couple the exosome to target transcripts and ARS2 binding was found to be enriched at terminator regions of RDH and snRNA genes (Andersen et al, 2013).

7.5 snRNA cleavage by the Integrator and transcription termination

It is known that the Integrator has a pivotal role in 3' end processing of snRNA and that the endonuclease subunit, INTS11, is thought to be responsible for snRNA cleavage (Baillat et al, 2005; Ezzedine et al, 2011; Dominski et al, 2005; Abrecht and Wagner, 2012). Depletion of Integrator subunits including INTS11 and INTS1 causes accumulation of misprocessed snRNAs and in this work I have provided further support for these findings (Figure 4.5 and 4.7) (Ezzedine et al, 2011; Baillat et al, 2005; Hata and Nakayama, 2007). Although the mechanism of snRNA termination is not fully understood, previous work has suggested a strong link between snRNA processing and efficient transcription termination (Ramamurthy et al, 1996; O'Reilly et al, 2014). However, my results

are contradictory to these findings. Instead I found that extended snRNAs, caused by INTS11 or INTS1 depletion, are still capable of termination (Figure 4.5, 4.6 and 4.7). This termination occurs within 1 – 3 Kb downstream of the snRNA TES, suggesting that integrator cleavage of snRNAs is not tightly linked to snRNA termination. One possible explanation for these apparent differences in my work compared to O'Reilly et al (2014) could be due to methodology.

O'Reilly et al (2014) depleted INTS11 and INTS9 Integrator subunits using RNAi methods in HeLa cells. They then investigated Pol II occupancy levels downstream of the TES for U1 and U2 snRNAs, however they only investigated 0.9 Kb and 1.2 Kb downstream respectively. An increase in Pol II % input was found downstream of the snRNA genes upon Integrator subunit depletion. Comparing these results with my data, I observed increased RNA levels downstream of snRNAs, often up to 1.2 Kb. This would corroborate with an increased Pol II occupancy at these locations, as seen by O'Reilly et al (2014). However, my data suggests termination of extended snRNAs by 1 – 3 Kb downstream of the TES and as such would expect Pol II occupancy to deplete within this window. Therefore it is possible that extended snRNA termination, as shown by reduced Pol II occupancy, may have been observed by O'Reilly et al (2014) if they had investigated further downstream of the snRNA TES. Overall, both of our findings show an extension of snRNA transcription upon defective 3' end snRNA processing, although I have shown termination still occurs. Therefore, Integrator cleavage of snRNAs may promote efficient transcription termination but not be necessary.

7.6 Is there a secondary endonuclease responsible for RDH cleavage?

RDH pre-mRNA is cleaved at the 3' end by the HCC complex, to form mature RDHs. The HCC is composed of multiple proteins including CPSF73, CPSF100 and Symplekin (Yang et al, 2013; Kolev et al, 2005). It is thought that CPSF73 endonuclease activity is responsible for histone processing, with cleavage occurring between the stem loop and HDE regions of the RDH gene (Yang et al, 2013; Dominski et al, 2005; Kolev et al, 2008; Sullivan et al, 2009). However, in this work I unexpectedly observed no effects on RDH pre-mRNA processing when CPSF73 was depleted in the *CPSF73-AID* cell line (Figure 6.1,

6.3 and 6.5). This is in contrast to findings by Pettinati et al (2018) who found at least 27 RDHs were misprocessed upon CPSF73 depletion.

Comparing my data with Pettinati et al (2018) I could directly analyse the same RDHs as in their study. Pettinati et al (2018) found a major effect of CPSF73 depletion on the processing of the following histone genes that I did not see any effect for in my CPSF73-AID cell line: *HIST1H4B, HIST1H3B, HIST1H2BC, HIST1H2BF* and *HIST1H4E* (Figure 6.1 and 6.3). As the observed misprocessing of these RDHs was approximately a 200 bp extension, I further analysed three histones (*HIST1H4B, HIST1H3B* and *HIST1H2BC*) more closely to ensure I had not visually overlooked any effect (Figure 6.5). However, I was still not able to detect any misprocessing effect at these RDHs upon CPSF73 depletion.

The discrepancies between these two works could be due to a number of reasons. Firstly the methodology used, with Pettinati et al utilising RNAi methods to deplete CPSF73 in HeLa cells synchronised in early S-phase. Readthrough RDH transcripts may not have been detected in my work due to the use of unsynchronised cells and the rapid turnover of RDHs at the end of S phase (Marzluff et al, 2008). However, the major effect on RDH pre-mRNA processing observed with occlusion of U7 snRNA in the same cells would argue against this. Additionally, misprocessing of RDHs results in polyadenylated transcripts that are stable throughout the cell cycle and thus making their detection easier (Kari et al, 2013; Romeo et al, 2014; Levine et al, 1987). Another possible reason is CPSF73 was not fully depleted upon doxycycline and auxin addition, however the depletion appeared near complete by western blot and was sufficient for major aberrant processing of mRNA (Figure 5.1 and 5.2).

From my results, CPSF73 may not be the main endonuclease for RDH 3' end cleavage, although many studies have shown otherwise (Yang et al, 2009; Dominski et al, 2005; Yang et al, 2013; Sullivan et al, 2009; Kolev et al, 2008). Therefore it may be more likely another protein shows redundancy upon CPSF73 depletion or that CPSF73 and another endonuclease act at individual sets of RDHs. A potential candidate for this is MBLAC1, an endoribonuclease that has a similar MBL domain to CPSF73 but with distinctive structural features, including the absence of a β-CASP domain found in CPSF73 (Pettinati et al, 2018). These differences could reflect specific substrate recognition. Depletion of MBLAC1 by

CRISPR/Cas9 techniques or siRNA resulted in readthrough RDH transcripts, showing a similar profile of effects to CPSF73 depletion (Pettinati et al, 2018).

7.7 Endonuclease depletion results in extended RNA transcripts that terminate at different lengths, depending on the RNA species.

Throughout this work I have shown that extended transcripts are generated when processing is disrupted or accumulate upon dysfunction of degradation pathways. Specifically I have shown there is an accumulation of extended snRNA transcripts upon DIS3 or INTS11 depletion; accumulation of extended RDH transcripts upon U7 snRNA occlusion, which is exacerbated by DIS3 depletion; and continuous readthrough of protein-coding genes upon CPSF73 depletion. This extension of snRNAs, RDHs and mRNAs is due to disruption in their processing, specifically the 3' end cleavage of these transcripts. Although misprocessed snRNA and RDH transcripts extend, they terminate relatively close to their TES; whereas misprocessed mRNA transcripts show readthrough that can continue for > 400 Kb (Figure 4.5, 6.10 and 5.4). It may be that 3' end cleavage and processing is tightly coupled to termination of mRNA transcripts, but not as much with snRNAs or RDH mRNA.

For RDHs, it has been shown that misprocessing results in polyadenylation of extended transcripts (Narita et al, 2007; Romeo et al, 2014). In this work, CPSF73 depletion did not appear to affect processing of RDHs, however preventing U7 snRNA binding to the HDE caused RDH mRNA extension. As previously described another endonuclease may be responsible for RDH pre-mRNA cleavage: MBLAC1 (Pettinati et al, 2018). Either way, U7 snRNA-dependent extended RDH transcripts terminated closely to the TES and this may be due to the presence of a downstream PAS site. As CPSF73 is still present it may cleave extended RDH transcripts at downstream PASs, resulting in both polyadenylation and termination. PASs have been found downstream of both snRNAs and RDH genes with misprocessed snRNAs also showing polyadenylation (Almada et al, 2013; Skaar et al, 2015; Yamamoto et al, 2014). Therefore, snRNAs may undergo a similar process where disrupting their canonical Integrator mediated cleavage instead causes CPSF73-mediated cleavage at a downstream PAS and termination. It is also possible a nuclease

other than CPSF73 enacts at these downstream PAS sites. Overall, this would explain the short extension of both RDHs and snRNAs observed throughout this work. Whereas long mRNA extension cannot be terminated at downstream PASs due to the loss of CPSF73, hence extension continues until RNA Pol II dissociates from the genome.

The fact that mRNAs show profuse extension upon CPSF73 depletion strongly suggests that PAS cleavage and CPSF73 are detrimental for termination. This supports the XRN2 torpedo model of transcription termination and previous work from the West lab (Eaton et al, 2018). Interestingly, although snRNAs are cleaved by the Integrator, resulting in an entry site for a 5' – 3' exonuclease, there is currently no evidence that XRN2 has a role in snRNA transcription termination (O'Reilly et al, 2014; Eaton et al, 2018). Likewise, RDH pre-mRNA cleavage provides possible entry for XRN2, but no role for XRN2 in RDH termination has been observed (Eaton et al, 2018).

Another reason why extended snRNAs may terminate close to their TES is due to the higher number of nucleosomes present 1 Kb downstream of the snRNA coding region (Egloff et al, 2009; O'Reilly et al, 2014). Nucleosomes can prevent efficient transcription and have been shown as early mRNA transcription quality checkpoints, causing premature termination through cleavage at cryptic PASs close to the TSS (Chui et al, 2018). This mechanism may be responsible for the production of prematurely terminated transcripts that accumulated upon DIS3 depletion (Figure 3.6 and 3.8). Alternatively, other work has shown intronic cryptic PASs can prematurely terminate mRNA transcription (Kaida et al, 2010; Berg et al, 2012). The nucleosome-free region in mRNA is only approximately 100 nts from the TSS, whereas it spans the entire snRNA transcription unit (Schones et al, 2008; Segal et al, 2006; Egloff et al, 2009). As described in the introduction, NELF promotes Pol II pausing at promoter-proximal mRNA sites and NELF phosphorylation, along with CTD and DSIF phosphorylation, is required for the transition to productive elongation (Ping and Rana, 2001; Peterlin and Price, 2006; Kwak and Lis, 2013). In contrast, at snRNA genes NELF is recruited to Pol II as it reaches the nucleosome dense area, coinciding with the end of the transcription unit. Depletion of NELF causes extension of snRNAs past the TES, although these transcripts still terminate (Egloff et al, 2009; O'Reilly et al, 2014). NELF mediated Pol II pausing on snRNAs and / or nucleosomes acting as a

barrier to transcription may promote snRNA transcription termination over continued elongation. This process of termination may not require Integrator-mediated snRNA cleavage and therefore may also be sufficient for termination of extended snRNA transcripts.

7.8 Future work and limitations

This work provided a clearer role for three individual ribonucleases, DIS3, INTS11 and CPSF73, through the analysis of genome-wide substrate effects upon their depletion. In terms of DIS3 and CPSF73, as little as 2 hours of depletion was sufficient for significant observable substrate perturbation in processing, termination and degradation events. This demonstrates both the effectiveness of the AID system to study functional genomics and the essential role of these endonucleases. However, the findings within this study have left some unanswered questions and generated new ones.

One issue of this study is that of protein redundancy, in particular when discussing DIS3 function in the exosome complex. As DIS3 is not the only active nuclease in the exosome complex, it is difficult to determine whether no effect upon DIS3 depletion translates into a non-exosome mediated degradation pathway. Specifically, an accumulation of mature RDH transcripts was not observed upon DIS3 depletion. However I cannot rule out the possibility that EXOSC10 and, therefore the exosome, plays a role. In addition, it has been shown that the exosome can exist in different isoforms, with either DIS3 binding, EXOSC10 binding or both together (Lykke-Andersen et al, 2011). Therefore, further work is required to elucidate the specific composition of the exosome when degrading different classes of RNA substrates. This would potentially aid in determining degradation pathways for specific RNAs.

In this discussion I have speculated on how extended transcripts may still terminate when their 3' end processing pathways are disrupted. A potential mechanism I described was that of downstream PAS cleavage by CPSF73 in RDH and snRNA extended precursors. It would therefore be interesting to further investigate this through depletion of CPSF73 coupled with the depletion of necessary proteins for 3' end cleavage of snRNAs and RDHs, for example INTS11 and U7 snRNA respectively. If a longer extension of transcripts was

observed, this may suggest PAS cleavage is responsible for misprocessed transcript termination.

Although I was able to look at indirect effects of Integrator dysfunction through the occlusion of U7 snRNA, it would be interesting to determine if direct depletion of Integrator subunits affected RDH pre-mRNA processing. In particular due to previous reports of Integrator subunit binding to the 3' end of RDHs and accumulation of misprocessed RDH transcripts upon INTS1 depletion (Skaar et al, 2015). Unfortunately, this was something not addressed in this work due to a lack of RDH mRNA expression in the INTS11-SMASh RNA-Seq dataset.

As CPSF73 depletion did not appear to have an effect on RDH pre-mRNA processing in this work, it will be important to further determine the role of CPSF73 in RDH transcription. Whether a protein such as MBLAC1 shows redundancy to CPSF73 or unsynchronised cells caused the lack of an observed effect, may need to be elucidated. Furthermore, whether prolonged periods of CPSF73 depletion could help determine CPSF73 function. Prolonged auxin or asunaprevir treatment may be beneficial for all three endonucleases investigated here, to provide a greater insight into the impact of their depletion over multiple rounds of transcription. This may also uncover potential redundant pathways for endonuclease function, that are not apparent when investigating nascent transcripts upon short protein depletion times.

Although this study has provided insight into the role and substrates of three endonucleases, there are limitations to the methodology and results. Firstly, it is important to note that due to the expense of RNA-Sequencing, I was only able to conduct 2 repeats of RNA-Seq for both DIS3 and CPSF73 depletion, or in the case of INTS11 only 1 repeat. In particular for INTS11 this makes the results less reliable as I am not able to show reproducibility. Additionally, a further repeat would have allowed investigation of statistical significance of findings for DIS3 and CPSF73. However, where possible I have aimed to provide another method to investigate the accuracy and significance of results found by RNA-Seq, often through the use of RT-qPCR. This method has its own caveats, as in this study I only used one housekeeping gene to normalise RT-qPCR results. To improve the reliability of RT-qPCR results, multiple housekeeping genes could have been utilised.

Secondly, in this work there has been a focus on both protein-coding genes and smaller transcripts such as snRNAs and RDHs. I initially aimed to look at the global effects of the target protein depletion on RNA through RNA-Seq and therefore utilised 50 nt reads for the RNA sequencing library preparation. However, this resulted in limitations for detection of smaller RNAs, as they were not always expressed in the RNA-Seq data. To overcome this issue and for further investigation specifically into changes of small RNA expressions I would suggest a small nuclear RNA-Seq method was instead used. Furthermore, due to the original aims of the research, expression levels of RDHs were not enriched for in experiments. This made some of my findings on RDH transcription and termination difficult to differentiate between an actual effect or due to natural fluctuating RDH transcript levels in relation to cell cycle phase. To investigate, in particular, the findings of CPSF73 depletion having no effect on RDH transcription, I would in the future ensure synchronisation of cells in S phase for all experiments. Finally, all research conducted in this work only utilised one cell type, those of a human colorectal carcinoma cell line, HCT116 cells. Therefore it is possible that some or all results are artefacts of this particular cell line. Further work should be undertaken to validate these findings in other human cell lines, before results are determined to be accurate in all human cells.

7.9 Conclusions

Overall, the findings within this work have shown that disruption of cleavage at the 3' end of multiple transcripts results in aberrant extended transcripts of differing lengths. Some of these extended transcripts may not only be produced when cleavage is disrupted but may also occur under normal circumstances. For example, extended snRNA transcripts were observed to accumulate upon depletion of DIS3 (Figure 4.10). Although this could be due to DIS3 having a role in snRNA 3' end cleavage and termination, as previously mentioned it is more likely that DIS3 normally degrades these extended snRNAs. An interesting finding in this study is the length of extension observed for transcripts upon cleavage disruption. It appears disruption of cleavage mechanisms for longer RNA transcripts, such as CPSF73 depletion effects on protein-coding mRNAs, show extension past the TES that continues for thousands of Kb (Figure 5.4). In comparison, shorter transcripts show extension

that terminates relatively close to the TES and does not continue indefinitely i.e. extension of snRNAs and RDHs upon Integrator and U7 snRNA dysfunction (Figure 4.5 and 6.10).

There are a couple of mechanisms that may explain these differences in extension between different RNA species. In summary, it is possible that transcription of shorter transcripts, such as snRNAs, commonly produce extended transcripts due to inefficient cleavage and / or termination and are normally rapidly degraded. Therefore, cells may have innate mechanisms in place to prevent termination extending indefinitely at these transcripts. For example the use of redundant cleavage mechanisms, such as CPSF73 and MBLAC1 at RDHS, or possible feature found downstream of the TES including an increased number of nucleosomes that could impede further Pol II transcription. Furthermore, during transcription of shorter transcripts the transcribing Pol II does not enter an elongation phase and the resulting phosphorylation state of Pol II differs to that of elongating Pol II on protein-coding genes (ref). Due to this, Pol II may be capable to more readily dissociate at shorter transcripts if extension / inefficient transcription termination does occur. On the other hand, Pol II transcribing longer transcripts results in an elongation state that may not be as easily terminated if normal termination mechanisms fail or falter, therefore extension readily occurs and continues.

Finally, this work further supports findings that the exosome, and in particular the DIS3 subunit, is responsible for degradation of a number of RNA substrates including de novo transcripts found in this study. Using conditional protein depletion cell lines I have been able to elucidate specific roles and substrates of DIS3, INTS11 and CPSF73. Although future work is necessary to investigate several unanswered questions, the results throughout this study show the effectiveness of using CRISPR/Cas9 techniques with an AID system for functional genomic studies.

References

Albrecht, T. R., S. P. Shevtsov, Y. Wu, L. G. Mascibroda, N. J. Peart, K. L. Huang, I. A. Sawyer, L. Tong, M. Dundr and E. J. Wagner (2018). "Integrator subunit 4 is a 'Symplekin-like' scaffold that associates with INTS9/11 to form the Integrator cleavage module." Nucleic Acids Res **46**(8): 4241-4255.

Albrecht, T. R. and E. J. Wagner (2012). "snRNA 3' end formation requires heterodimeric association of integrator subunits." Mol Cell Biol **32**(6): 1112-1123.

Allmang, C., J. Kufel, G. Chanfreau, P. Mitchell, E. Petfalski and D. Tollervey (1999). "Functions of the exosome in rRNA, snoRNA and snRNA synthesis." EMBO J **18**(19): 5399-5410.

Almada, A. E., X. Wu, A. J. Kriz, C. B. Burge and P. A. Sharp (2013). "Promoter directionality is controlled by U1 snRNP and polyadenylation signals." Nature **499**(7458): 360-363.

Anamika, K., À. Gyenis, L. Poidevin, O. Poch and L. Tora (2012). "RNA polymerase II pausing downstream of core histone genes is different from genes producing polyadenylated transcripts." PLoS One **7**(6): e38769.

Andersen, P. R., M. Domanski, M. S. Kristiansen, H. Storvall, E. Ntini, C. Verheggen, A. Schein, J. Bunkenborg, I. Poser, M. Hallais, R. Sandberg, A. Hyman, J. LaCava, M. P. Rout, J. S. Andersen, E. Bertrand and T. H. Jensen (2013). "The human cap-binding complex is functionally connected to the nuclear RNA exosome." Nat Struct Mol Biol **20**(12): 1367-1376.

Andersson, R., C. Gebhard, I. Miguel-Escalada, I. Hoof, J. Bornholdt, M. Boyd, Y. Chen, X. Zhao, C. Schmidl, T. Suzuki, E. Ntini, E. Arner, E. Valen, K. Li, L. Schwarzfischer, D. Glatz, J. Raithel, B. Lilje, N. Rapin, F. O. Bagger, M. Jørgensen, P. R. Andersen, N. Bertin, O. Rackham, A. M. Burroughs, J. K. Baillie, Y. Ishizu, Y. Shimizu, E. Furuhata, S. Maeda, Y. Negishi, C. J. Mungall, T. F. Meehan, T. Lassmann, M. Itoh, H. Kawaji, N. Kondo, J. Kawai, A. Lennartsson, C. O. Daub, P. Heutink, D. A. Hume, T. H. Jensen, H. Suzuki, Y. Hayashizaki, F. Müller, A. R. R. Forrest, P. Carninci, M. Rehli and A. Sandelin (2014). "An atlas of active enhancers across human cell types and tissues." Nature **507**(7493): 455-461.

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

Änkö ,M. L (2014). “Regulation of gene expression programmes by serine–arginine rich splicing factors.” Semin. Cell Dev. Biol. **32**:11–21

Baillat, D., M. A. Hakimi, A. M. Näär, A. Shilatifard, N. Cooch and R. Shiekhattar (2005). "Integrator, a multiprotein mediator of small nuclear RNA processing,

associates with the C-terminal repeat of RNA polymerase II." Cell **123**(2): 265-276.

Baillat, D. and E. J. Wagner (2015). "Integrator: surprisingly diverse functions in gene expression." Trends Biochem Sci **40**(5): 257-264.

Banerji, J., S. Rusconi and W. Schaffner (1981). "Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences." Cell **27**(2 Pt 1): 299-308.

Barnett, D. W., E. K. Garrison, A. R. Quinlan, M. P. Strömberg and G. T. Marth (2011). "BamTools: a C++ API and toolkit for analyzing and managing BAM files." Bioinformatics **27**(12): 1691-1692.

Barrangou, R., C. Fremaux, H. Deveau, M. Richards, P. Boyaval, S. Moineau, D. A. Romero and P. Horvath (2007). "CRISPR provides acquired resistance against viruses in prokaryotes." Science **315**(5819): 1709-1712.

Belur, L. R., J. L. Frandsen, A. J. Dupuy, D. H. Ingbar, D. A. Largaespada, P. B. Hackett and R. Scott McIvor (2003). "Gene insertion and long-term expression in lung mediated by the Sleeping Beauty transposon system." Mol Ther **8**(3): 501-507.

Bentley, D. L. (2014). "Coupling mRNA processing with transcription in time and space." Nat Rev Genet **15**(3): 163-175.

Berg, M. G., L. N. Singh, I. Younis, Q. Liu, A. M. Pinto, D. Kaida, Z. Zhang, S. Cho, S. Sherrill-Mix, L. Wan and G. Dreyfuss (2012). "U1 snRNP determines mRNA length and regulates isoform expression." Cell **150**(1): 53-64.

Bibikova, M., M. Golic, K. G. Golic and D. Carroll (2002). "Targeted chromosomal cleavage and mutagenesis in Drosophila using zinc-finger nucleases." Genetics **161**(3): 1169-1175.

Bondeson, D. P. and C. M. Crews (2017). "Targeted Protein Degradation by Small Molecules." Annu Rev Pharmacol Toxicol **57**: 107-123.

Bonneau, F., J. Basquin, J. Ebert, E. Lorentzen and E. Conti (2009). "The yeast exosome functions as a macromolecular cage to channel RNA substrates for degradation." Cell **139**(3): 547-559.

Boutros, M. and J. Ahringer (2008). "The art and design of genetic screens: RNA interference." Nat Rev Genet **9**(7): 554-566.

Brannan, K., H. Kim, B. Erickson, K. Glover-Cutter, S. Kim, N. Fong, L. Kiemele, K. Hansen, R. Davis, J. Lykke-Andersen and D. L. Bentley (2012). "mRNA decapping factors and the exonuclease Xrn2 function in widespread premature termination of RNA polymerase II transcription." Mol Cell **46**(3): 311-324.

Braun, J. E., V. Truffault, A. Boland, E. Huntzinger, C. T. Chang, G. Haas, O. Weichenrieder, M. Coles, E. Izaurralde (2012). “A direct interaction between DCP1 and XRN1 couples mRNA decapping to 5’ exonucleolytic degradation.” Nat Struct Mol Biol **19**(12): 1324-1331.

Bresson, S. M. and N. K. Conrad (2013). "The human nuclear poly(a)-binding protein promotes RNA hyperadenylation and decay." PLoS Genet **9**(10): e1003893.

Burch, B. D., A. C. Godfrey, P. Y. Gasdaska, H. R. Salzler, R. J. Duronio, W. F. Marzluff and Z. Dominski (2011). "Interaction between FLASH and Lsm11 is essential for histone pre-mRNA processing in vivo in Drosophila." RNA **17**(6): 1132-1147.

Callebaut, I., D. Moshous, J. P. Mornon and J. P. de Villartay (2002). "Metallo-beta-lactamase fold within nucleic acids processing enzymes: the beta-CASP family." Nucleic Acids Res **30**(16): 3592-3601.

Camper, S. A., R. J. Albers, J. K. Coward, F. M. Rottman (1984). “Effect of undermethylation on mRNA cytoplasmic appearance and half-life.” Mol. Cell. Biol **4**: 538–543.

Carlile, T.M., M. F. Rojas-Duran, B. Zinshteyn, H. Shin, K. M. Bartoli, W. V. Gilbert (2014). “Pseudouridine profiling reveals regulated mRNA pseudouridylation in yeast and human cells.” Nature **515**: 143–146.

Cassé, C., F. Giannoni, V. T. Nguyen, M. F. Dubois and O. Bensaude (1999). "The transcriptional inhibitors, actinomycin D and alpha-amanitin, activate the HIV-1 promoter and favor phosphorylation of the RNA polymerase II C-terminal domain." J Biol Chem **274**(23): 16097-16106.

Cazalla, D., M. Xie and J. A. Steitz (2011). "A primate herpesvirus uses the integrator complex to generate viral microRNAs." Mol Cell **43**(6): 982-992.

Charkrabarti, S., U. Jayachandran, F. Bonneau, F. Fiorini, C. Basquin, S. Domcke, H. Le Hir, E. Conti (2011). “Molecular mechanisms for the RNA-dependent ATPase activity of Upf1 and its regulation by Upf2.” Mol Cell **41**: 693-703.

Chen, C. Y. A., A. B. Shyu (2011). “Mechanisms of deadenylation-dependent decay.” Wiley Interdiscip Rev RNA **2**:167-183.

Chen, F., S. M. Pruett-Miller, Y. Huang, M. Gjoka, K. Duda, J. Taunton, T. N. Collingwood, M. Frodin and G. D. Davis (2011). "High-frequency genome editing using ssDNA oligonucleotides with zinc-finger nucleases." Nat Methods **8**(9): 753-755.

Chen, N., M. A. Walsh, Y. Liu, R. Parker, H. Song (2005). “Crystal structures of human DcpS in ligand-free and m7GDP-bound forms suggest a dynamic mechanism for scavenger mRNA decapping.” J Mol Biol **347**(4): 707-718.

Chen, Y., A. A. Pai, J. Herudek, M. Lubas, N. Meola, A. I. Järvelin, R. Andersson, V. Pelechano, L. M. Steinmetz, T. H. Jensen and A. Sandelin (2016). "Principles for RNA metabolism and alternative transcription initiation within closely spaced promoters." Nat Genet **48**(9): 984-994.

Chiu, A. C., H. I. Suzuki, X. Wu, D. B. Mahat, A. J. Kriz and P. A. Sharp (2018). "Transcriptional Pause Sites Delineate Stable Nucleosome-Associated Premature Polyadenylation Suppressed by U1 snRNP." Mol Cell **69**(4): 648-663.e647.

Chung, H. K., C. L. Jacobs, Y. Huo, J. Yang, S. A. Krumm, R. K. Plemper, R. Y. Tsien and M. Z. Lin (2015). "Tunable and reversible drug control of protein production via a self-excising degron." Nat Chem Biol **11**(9): 713-720.

Cong, L., F. A. Ran, D. Cox, S. Lin, R. Barretto, N. Habib, P. D. Hsu, X. Wu, W. Jiang, L. A. Marraffini and F. Zhang (2013). "Multiplex genome engineering using CRISPR/Cas systems." Science **339**(6121): 819-823.

Cooper, G. M (2000). "The cell: a molecular approach." Sunderland (MA): Sinaeur Associates. Available from: https://www.ncbi.nlm.nih.Gov/books/NBK9849/

Corden, J. L. (1990). "Tails of RNA polymerase II." Trends Biochem Sci **15**(10): 383-387.

Daou-Chabo, R. and C. Condon (2009). "RNase J1 endonuclease activity as a probe of RNA secondary structure." RNA **15**(7): 1417-1425.

Davidson, L., L. Francis, R. A. Cordiner, J. D. Eaton, C. Estell, S. Macias, J. F. Cáceres and S. West (2019). "Rapid Depletion of DIS3, EXOSC10, or XRN2 Reveals the Immediate Impact of Exoribonucleolysis on Nuclear RNA Metabolism and Transcriptional Control." Cell Rep **26**(10): 2779-2791.e2775.

Davidson, L., A. Kerr and S. West (2012). "Co-transcriptional degradation of aberrant pre-mRNA by Xrn2." EMBO J **31**(11): 2566-2578.

Davidson, L., L. Muniz and S. West (2014). "3' end formation of pre-mRNA and phosphorylation of Ser2 on the RNA polymerase II CTD are reciprocally coupled in human cells." Genes Dev **28**(4): 342-356.

de la Mata, M., C. R. Alonso, S. Kadener, J. P. Fededa, M. Blaustein, F. Pelisch, P. Cramer, D. Bentley and A. R. Kornblihtt (2003). "A slow RNA polymerase II affects alternative splicing in vivo." Mol Cell **12**(2): 525-532.

De Santa, F., I. Barozzi, F. Mietton, S. Ghisletti, S. Polletti, B. K. Tusi, H. Muller, J. Ragoussis, C. L. Wei and G. Natoli (2010). "A large fraction of extragenic RNA pol II transcription sites overlap enhancers." PLoS Biol **8**(5): e1000384.

Desrosiers, R., K. Friderici, F. Rottman (1974). "Identification of methylated nucleosides in messenger RNA from Novikoff hepatoma cells." Proc. Natl. Acad. Sci. USA **71**: 3971–3975.

Desrosiers, R. C., K. H. Friderici, F. M. Rottman (1975). "Characterization of Novikoff hepatoma mRNA methylation and heterogeneity in the methylated 5′ terminus." Biochemistry **14**: 4367–4374.

Dharmasiri, N., S. Dharmasiri and M. Estelle (2005). "The F-box protein TIR1 is an auxin receptor." Nature **435**(7041): 441-445.

Dominissini, D., S. Moshitch-Moshkovitz, S. Schwartz, M. Salmon-Divon, L. Ungar, S. Osenberg, K. Cesarkas, J. Jacob-Hirsch, N. Amariglio, M. Kupiec, et al (2012). “Topology of the human and mouse m6 A RNA methylomes revealed by m6 A-seq.” Nature **485**: 201–206.

Dominski, Z., J. A. Erkmann, X. Yang, R. Sànchez and W. F. Marzluff (2002). "A novel zinc finger protein is associated with U7 snRNP and interacts with the stem-loop binding protein in the histone pre-mRNP to stimulate 3'-end processing." Genes Dev **16**(1): 58-71.

Dominski, Z. and W. F. Marzluff (2007). "Formation of the 3' end of histone mRNA: getting closer to the end." Gene **396**(2): 373-390.

Dominski, Z., X. C. Yang and W. F. Marzluff (2005). "The polyadenylation factor CPSF-73 is involved in histone-pre-mRNA processing." Cell **123**(1): 37-48.

Dziembowski, A., E. Lorentzen, E. Conti and B. Séraphin (2007). "A single subunit, Dis3, is essentially responsible for yeast exosome core activity." Nat Struct Mol Biol **14**(1): 15-22.

Eaton, J. D., L. Davidson, D. L. V. Bauer, T. Natsume, M. T. Kanemaki and S. West (2018). "Xrn2 accelerates termination by RNA polymerase II, which is underpinned by CPSF73 activity." Genes Dev **32**(2): 127-139.

Eberle, A. B., S. Lykke-Andersen, O. Mühlemann, T. H. Jensen (2009). “SMG6 promotes endonucleolytic cleavage of nonsense mRNA in human cells.” Nat Struct Mol Biol **16**: 49-55

Eckmann, C. R., C. Rammelt and E. Wahle (2011). "Control of poly(A) tail length." Wiley Interdiscip Rev RNA **2**(3): 348-361.

Egecioglu, D. E., A. K. Henras and G. F. Chanfreau (2006). "Contributions of Trf4p- and Trf5p-dependent polyadenylation to the processing and degradative functions of the yeast nuclear exosome." RNA **12**(1): 26-32.

Egloff, S., H. Al-Rawaf, D. O'Reilly and S. Murphy (2009). "Chromatin structure is implicated in "late" elongation checkpoints on the U2 snRNA and beta-actin genes." Mol Cell Biol **29**(14): 4002-4013.

Egloff, S., D. O'Reilly, R. D. Chapman, A. Taylor, K. Tanzhaus, L. Pitts, D. Eick and S. Murphy (2007). "Serine-7 of the RNA polymerase II CTD is specifically required for snRNA gene expression." Science **318**(5857): 1777-1779.

Egloff, S., S. A. Szczepaniak, M. Dienstbier, A. Taylor, S. Knight and S. Murphy (2010). "The integrator complex recognizes a new double mark on the RNA polymerase II carboxyl-terminal domain." J Biol Chem **285**(27): 20564-20569.

Egloff, S., J. Zaborowska, C. Laitem, T. Kiss, S. Murphy (2011). “Ser7 phosphorylation of the CTD recruits the RPAP2 Ser5 phosphatase to snRNA genes.” Mol Cell **45**(1): 111-122.

Elbashir, S. M., J. Harborth, W. Lendeckel, A. Yalcin, K. Weber and T. Tuschl (2001). "Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells." Nature **411**(6836): 494-498.

Even, S., O. Pellegrini, L. Zig, V. Labas, J. Vinh, D. Bréchemmier-Baey and H. Putzer (2005). "Ribonucleases J1 and J2: two novel endoribonucleases in B.subtilis with functional homology to E.coli RNase E." Nucleic Acids Res **33**(7): 2141-2152.

Ezzeddine, N., J. Chen, B. Waltenspiel, B. Burch, T. Albrecht, M. Zhuo, W. D. Warren, W. F. Marzluff and E. J. Wagner (2011). "A subset of Drosophila integrator proteins is essential for efficient U7 snRNA and spliceosomal snRNA 3'-end formation." Mol Cell Biol **31**(2): 328-341.

Falk, S., F. Bonneau, J. Ebert, A. Kögel and E. Conti (2017). "Mpp6 Incorporation in the Nuclear Exosome Contributes to RNA Channeling through the Mtr4 Helicase." Cell Rep **20**(10): 2279-2286.

Fay, E. J., S. L. Aron, I. A. Stone, B. M. Waring, R. K. Plemper and R. A. Langlois (2019). "Engineered Small-Molecule Control of Influenza A Virus Replication." J Virol **93**(1).

Finkel, D., Y. Groner (1983). “Methylations of adenosine residues (m6 A) in pre-mRNA are important for formation of late simian virus 40 mRNAs.” Virology **131**: 409–425.

Fiorini, F., D. Bagchi, H. Le Hir, V. Croquette (2015). “Human Upf1 is a highly processive RNA helicase and translocase with RNP remodelling activities.” Nat Commun **6**: 7581

Flynn, R. A., A. E. Almada, J. R. Zamudio and P. A. Sharp (2011). "Antisense RNA polymerase II divergent transcripts are P-TEFb dependent and substrates for the RNA exosome." Proc Natl Acad Sci U S A **108**(26): 10460-10465.

Fusby, B., S. Kim, B. Erickson, H. Kim, M. L. Peterson and D. L. Bentley (2016). "Coordination of RNA Polymerase II Pausing and 3' End Processing Factor Recruitment with Alternative Polyadenylation." Mol Cell Biol **36**(2): 295-303.

Gaj, T., C. A. Gersbach and C. F. Barbas (2013). "ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering." Trends Biotechnol **31**(7): 397-405.

Gardini, A., D. Baillat, M. Cesaroni, D. Hu, J. M. Marinis, E. J. Wagner, M. A. Lazar, A. Shilatifard and R. Shiekhattar (2014). "Integrator regulates transcriptional initiation and pause release following activation." Mol Cell **56**(1): 128-139.

Gerlach, P., J. M. Schuller, F. Bonneau, J. Basquin, P. Reichelt, S. Falk and E. Conti (2018). "Distinct and evolutionary conserved structural features of the human nuclear exosome complex." Elife **7**.

Glover-Cutter, K., S. Kim, J. Espinosa and D. L. Bentley (2008). "RNA polymerase II pauses and associates with pre-mRNA processing factors at both ends of genes." Nat Struct Mol Biol **15**(1): 71-78.

Glover-Cutter, K., S. Larochelle, B. Erickson, C. Zhang, K. Shokat, R. P. Fisher and D. L. Bentley (2009). "TFIIH-associated Cdk7 kinase functions in phosphorylation of C-terminal domain Ser7 residues, promoter-proximal pausing, and termination by RNA polymerase II." Mol Cell Biol **29**(20): 5455-5464.

Gray, W. M., J. C. del Pozo, L. Walker, L. Hobbie, E. Risseeuw, T. Banks, W. L. Crosby, M. Yang, H. Ma and M. Estelle (1999). "Identification of an SCF ubiquitin-ligase complex required for auxin response in Arabidopsis thaliana." Genes Dev **13**(13): 1678-1691.

Gray, W. M., S. Kepinski, D. Rouse, O. Leyser and M. Estelle (2001). "Auxin regulates SCF(TIR1)-dependent degradation of AUX/IAA proteins." Nature **414**(6861): 271-276.

Gromak, N., S. West and N. J. Proudfoot (2006). "Pause sites promote transcriptional termination of mammalian RNA polymerase II." Mol Cell Biol **26**(10): 3986-3996.

Gruber, J. J., D. S. Zatechka, L. R. Sabin, J. Yong, J. J. Lum, M. Kong, W. X. Zong, Z. Zhang, C. K. Lau, J. Rawlings, S. Cherry, J. N. Ihle, G. Dreyfuss and C. B. Thompson (2009). "Ars2 links the nuclear cap-binding complex to RNA interference and cell proliferation." Cell **138**(2): 328-339.

Guan, Y., Q. Zhu, D. Huang, S. Zhao, L. Jan Lo and J. Peng (2015). "An equation to estimate the difference between theoretically predicted and SDS PAGE-displayed molecular weights for an acidic peptide." Sci Rep **5**: 13370.

Guenther, M. G., S. S. Levine, L. A. Boyer, R. Jaenisch and R. A. Young (2007). "A chromatin landmark and transcription initiation at most promoters in human cells." Cell **130**(1): 77-88.

Guiro, J. and S. Murphy (2017). “Regulation of expression of human RNA polymerase II transcribed snRNA genes.” Open Biol. **7**(6).

Gupta, R. M. and K. Musunuru (2014). "Expanding the genetic editing tool kit: ZFNs, TALENs, and CRISPR-Cas9." J Clin Invest **124**(10): 4154-4161.

Hackett, P. B., D. A. Largaespada and L. J. Cooper (2010). "A transposon and transposase system for human application." Mol Ther **18**(4): 674-683.

Hahn, A. T., J. T. Jones and T. Meyer (2009). "Quantitative analysis of cell cycle phase durations and PC12 differentiation using fluorescent biosensors." Cell Cycle **8**(7): 1044-1052.

Haigis, K. M., J. G. Caya, M. Reichelderfer and W. F. Dove (2002). "Intestinal adenomas can develop with a stable karyotype and stable microsatellites." Proc Natl Acad Sci U S A **99**(13): 8927-8931.

Hallais, M., F. Pontvianne, P. R. Andersen, M. Clerici, D. Lener, N. I. H. Benbahouche, T. Gostan, F. Vandermoere, M. C. Robert, S. Cusack, C. Verheggen, T. H. Jensen and E. Bertrand (2013). "CBC-ARS2 stimulates 3'-end maturation of multiple RNA families and favors cap-proximal processing." Nat Struct Mol Biol **20**(12): 1358-1366.

Harris, M. E., R. Böhni, M. H. Schneiderman, L. Ramamurthy, D. Schümperli and W. F. Marzluff (1991). "Regulation of histone mRNA in the unperturbed cell cycle: evidence suggesting control at two posttranscriptional steps." Mol Cell Biol **11**(5): 2416-2424.

Hata, T. and M. Nakayama (2007). "Targeted disruption of the murine large nuclear KIAA1440/Ints1 protein causes growth arrest in early blastocyst stage embryos and eventual apoptotic cell death." Biochim Biophys Acta **1773**(7): 1039-1051.

Hauer, M. H., A. Seeber, V. Singh, R. Thierry, R. Sack, A. Amitai, M. Kryzhanovska, J. Eglinger, D. Holcman, T. Owen-Hughes and S. M. Gasser (2017). "Histone degradation in response to DNA damage enhances chromatin dynamics and recombination rates." Nat Struct Mol Biol **24**(2): 99-107.

Heidemann, M., C. Hintermair, K. Voß and D. Eick (2013). "Dynamic phosphorylation patterns of RNA polymerase II CTD during transcription." Biochim Biophys Acta **1829**(1): 55-62.

Herzel, L., D. S. M. Ottoz, T. Alpert and K. M. Neugebauer (2017). "Splicing and transcription touch base: co-transcriptional spliceosome assembly and function." Nat Rev Mol Cell Biol **18**(10): 637-650.

Holland, A. J., D. Fachinetti, J. S. Han and D. W. Cleveland (2012). "Inducible, reversible system for the rapid and complete degradation of proteins in mammalian cells." Proc Natl Acad Sci U S A **109**(49): E3350-3357.

Horii, T. and I. Hatada (2015). "Genome Editing Using Mammalian Haploid Cells." Int J Mol Sci **16**(10): 23604-23614.

Hou, X., Y. Du, Y. Deng, J. Wu and G. Cao (2015). "Sleeping Beauty transposon system for genetic etiological research and gene therapy of cancers." Cancer Biol Ther **16**(1): 8-16.

Houseley, J. and D. Tollervey (2009). "The many pathways of RNA degradation." Cell **136**(4): 763-776.

Hrossova, D., T. Sikorsky, D. Potesil, M. Bartosovic, J. Pasulka, Z. Zdrahal, R. Stefl and S. Vanacova (2015). "RBM7 subunit of the NEXT complex binds U-rich sequences and targets 3'-end extended forms of snRNAs." Nucleic Acids Res **43**(8): 4236-4248.

Hsin, J. P. and J. L. Manley (2012). "The RNA polymerase II CTD coordinates transcription and RNA processing." Genes Dev **26**(19): 2119-2137.

Hsin, J. P., A. Sheth and J. L. Manley (2011). "RNAP II CTD phosphorylated on threonine-4 is required for histone mRNA 3' end processing." Science **334**(6056): 683-686.

Hsin, J. P., K. Xiang and J. L. Manley (2014). "Function and control of RNA polymerase II C-terminal domain phosphorylation in vertebrate transcription and RNA processing." Mol Cell Biol **34**(13): 2488-2498.

Huntzinger, E., I. Kashima, M. Fauser, J. Sauliére, E. Izaurralde (2008). "SMG6 is the catalytic endonuclease that cleaves mRNAs containing nonsense codons in metazoan." RNA NYN **14**: 2609-2617

Iasillo, C., M. Schmid, Y. Yahia, M. A. Maqbool, N. Descostes, E. Karadoulama, E. Bertrand, J. C. Andrau and T. H. Jensen (2017). "ARS2 is a general suppressor of pervasive transcription." Nucleic Acids Res **45**(17): 10229-10241.

Ivics, Z. and Z. Izsvák (2015). "Sleeping Beauty Transposition." Microbiol Spectr **3**(2): MDNA3-0042-2014.

Izban, M. G. and D. S. Luse (1992). "The RNA polymerase II ternary complex cleaves the nascent transcript in a 3'----5' direction in the presence of elongation factor SII." Genes Dev **6**(7): 1342-1356.

Izsvák, Z. and Z. Ivics (2004). "Sleeping beauty transposition: biology and applications for molecular therapy." Mol Ther **9**(2): 147-156.

Izsvák, Z., Z. Ivics and R. H. Plasterk (2000). "Sleeping Beauty, a wide host-range transposon vector for genetic transformation in vertebrates." J Mol Biol **302**(1): 93-102.

Jackson, A. L., S. R. Bartz, J. Schelter, S. V. Kobayashi, J. Burchard, M. Mao, B. Li, G. Cavet and P. S. Linsley (2003). "Expression profiling reveals off-target gene regulation by RNAi." Nat Biotechnol **21**(6): 635-637.

Jacobs, E. Y., I. Ogiwara and A. M. Weiner (2004). "Role of the C-terminal domain of RNA polymerase II in U2 snRNA transcription and 3' processing." Mol Cell Biol **24**(2): 846-855.

Januszyk, K., Q. Liu and C. D. Lima (2011). "Activities of human RRP6 and structure of the human RRP6 catalytic domain." RNA **17**(8): 1566-1577.

Jenal, M., R. Elkon, F. Loayza-Puch, G. van Haaften, U. Kühn, F. M. Menzies, J. A. Oude Vrielink, A. J. Bos, J. Drost, K. Rooijers, D. C. Rubinsztein and R. Agami (2012). "The poly(A)-binding protein nuclear 1 suppresses alternative cleavage and polyadenylation sites." Cell **149**(3): 538-553.

Jeong, S., (2017). "SR proteins: Binders, Regulators and connectors of RNA." Mol Cell **40**(1): 1-9.

Jimeno-González, S., L. L. Haaning, F. Malagon and T. H. Jensen (2010). "The yeast 5'-3' exonuclease Rat1p functions during transcription elongation by RNA polymerase II." Mol Cell **37**(4): 580-587.

Jinek, M., K. Chylinski, I. Fonfara, M. Hauer, J. A. Doudna and E. Charpentier (2012). "A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity." Science **337**(6096): 816-821.

Jonkers, I. and J. T. Lis (2015). "Getting up to speed with transcription elongation by RNA polymerase II." Nat Rev Mol Cell Biol **16**(3): 167-177.

Kadaba, S., X. Wang and J. T. Anderson (2006). "Nuclear RNA surveillance in Saccharomyces cerevisiae: Trf4p-dependent polyadenylation of nascent hypomethylated tRNA and an aberrant form of 5S rRNA." RNA **12**(3): 508-521.

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

Kari, V., O. Karpiuk, B. Tieg, M. Kriegs, E. Dikomey, H. Krebber, Y. Begus-Nahrmann and S. A. Johnsen (2013). "A subset of histone H2B genes produces polyadenylated mRNAs under a variety of cellular conditions." PLoS One **8**(5): e63745.

Kashima, I., A. Yamashita, N. Izumi, N. Kataoka, R. Morishita, S. Hoshino, M. Ohno, G. Dreyfuss, D. Ohno (2006). “Binding of anovel SMG-1-Upf1-eRF1-eRF3 complex (SURF) to the exon junction complex triggers Upf1 phosphorylation and nonsense-mediated mRNA decay.” Genes Dev **20**(3): 355-367

Katahira, J (2012). “mRNA export and the TREX complex.” BBA **1819**(6): 507-513

Kilchert, C., S. Wittmann and L. Vasiljeva (2016). "The regulation and functions of the nuclear RNA exosome complex." Nat Rev Mol Cell Biol **17**(4): 227-239.

Kim, D., B. Langmead and S. L. Salzberg (2015). "HISAT: a fast spliced aligner with low memory requirements." Nat Methods **12**(4): 357-360.

Kim, H., B. Erickson, W. Luo, D. Seward, J. H. Graber, D. D. Pollock, P. C. Megee and D. L. Bentley (2010a). "Gene-specific RNA polymerase II phosphorylation and the CTD code." Nat Struct Mol Biol **17**(10): 1279-1286.

Kim, M., N. J. Krogan, L. Vasiljeva, O. J. Rando, E. Nedea, J. F. Greenblatt and S. Buratowski (2004). "The yeast Rat1 exonuclease promotes transcription termination by RNA polymerase II." Nature **432**(7016): 517-522.

Kim, T. K., M. Hemberg, J. M. Gray, A. M. Costa, D. M. Bear, J. Wu, D. A. Harmin, M. Laptewicz, K. Barbara-Haley, S. Kuersten, E. Markenscoff-Papadimitriou, D. Kuhl, H. Bito, P. F. Worley, G. Kreiman and M. E. Greenberg (2010b). "Widespread transcription at neuronal activity-regulated enhancers." Nature **465**(7295): 182-187.

Kolev, N. G. and J. A. Steitz (2005). "Symplekin and multiple other polyadenylation factors participate in 3'-end maturation of histone mRNAs." Genes Dev **19**(21): 2583-2592.

Kolev, N. G., T. A. Yario, E. Benson and J. A. Steitz (2008). "Conserved motifs in both CPSF73 and CPSF100 are required to assemble the active endonuclease for histone mRNA 3'-end maturation." EMBO Rep **9**(10): 1013-1018.

Kostrouchova, M., M. Krause, Z. Kostrouch and J. E. Rall (2001). "Nuclear hormone receptor CHR3 is a critical regulator of all four larval molts of the nematode Caenorhabditis elegans." Proc Natl Acad Sci U S A **98**(13): 7360-7365.

Kowarz, E., D. Löscher and R. Marschalek (2015). "Optimized Sleeping Beauty transposons rapidly generate stable transgenic cell lines." Biotechnol J**10**(4): 647-653.

Krishnamurthy, S. and M. Hampsey (2009). "Eukaryotic transcription initiation." Curr Biol **19**(4): R153-156.

Krueger, F. (2012). Trim Galore! A wrapper tool around Cutadapt and FastQC to consistently apply quality and adapter trimming to FastQ files, with some extra functionality for MspI-digested RRBS-type (Reduced Representation Bisufite-Seq) libraries. . Available online at: https://www.bioinformatics.babraham.ac.uk/projects/trim_galore.

Kumakura, N., H. Otsuki, M. Tsuzuki, A. Takeda and Y. Watanabe (2013). "Arabidopsis AtRRP44A is the functional homolog of Rrp44/Dis3, an exosome component, is essential for viability and is required for RNA processing and degradation." PLoS One **8**(11): e79219.

Kwak, H., N. J. Fuda, L. J. Core and J. T. Lis (2013). "Precise maps of RNA polymerase reveal how promoters direct initiation and pausing." Science **339**(6122): 950-953.

Kwak, H. and J. T. Lis (2013). "Control of transcriptional elongation." Annu Rev Genet **47**: 483-508.

Kyriakopoulou, C., P. Larsson, L. Liu, J. Schuster, F. Söderbom, L. A. Kirsebom and A. Virtanen (2006). "U1-like snRNAs lacking complementarity to canonical 5' splice sites." RNA **12**(9): 1603-1611.

Kühn, U., M. Gündel, A. Knoth, Y. Kerwitz, S. Rüdel and E. Wahle (2009). "Poly(A) tail length is controlled by the nuclear poly(A)-binding protein regulating the interaction between poly(A) polymerase and the cleavage and polyadenylation specificity factor." J Biol Chem **284**(34): 22803-22814.

LaCava, J., J. Houseley, C. Saveanu, E. Petfalski, E. Thompson, A. Jacquier and D. Tollervey (2005). "RNA degradation by the exosome is promoted by a nuclear polyadenylation complex." Cell **121**(5): 713-724.

Lai, F., A. Gardini, A. Zhang and R. Shiekhattar (2015). "Integrator mediates the biogenesis of enhancer RNAs." Nature **525**(7569): 399-403.

Laitem, C., J. Zaborowska, M. Tellier, Y. Yamaguchi, Q. Cao, S. Egloff, H. Handa and S. Murphy (2015). "CTCF regulates NELF, DSIF and P-TEFb recruitment during transcription." Transcription **6**(5): 79-90.

Lambrus, B. G., T. C. Moyer and A. J. Holland (2018). "Applying the auxin-inducible degradation system for rapid protein depletion in mammalian cells." Methods Cell Biol **144**: 107-135.

Larochelle, M., M. A. Robert, J. N. Hébert, X. Liu, D. Matteau, S. Rodrigue, B. Tian, P. Jacques and F. Bachand (2018). "Common mechanism of transcription termination at coding and noncoding RNA genes in fission yeast." Nat Commun **9**(1): 4364.

Lavi, U., R. Fernandez-Mun˜ oz, J. E. Darnell (1977). “Content of N-6 methyl adenylic acid in heterogeneous nuclear and messenger RNA of HeLa cells.” Nucleic Acids Res. **4**: 63–69.

Lawrence, M., R. Gentleman and V. Carey (2009). "rtracklayer: an R package for interfacing with genome browsers." Bioinformatics **25**(14): 1841-1842.

Lawrence, M., W. Huber, H. Pagès, P. Aboyoun, M. Carlson, R. Gentleman, M. T. Morgan and V. J. Carey (2013). "Software for computing and annotating genomic ranges." PLoS Comput Biol **9**(8): e1003118.

Lebreton, A., R. Tomecki, A. Dziembowski and B. Séraphin (2008). "Endonucleolytic RNA cleavage by a eukaryotic exosome." Nature **456**(7224): 993-996.

Lee, K., C. C. Hsiung, P. Huang, A. Raj and G. A. Blobel (2015). "Dynamic enhancer-gene body contacts during transcription elongation." Genes Dev**29**(19): 1992-1997.

Le Hir, H., E. Izaurralde, L. E. Maquat, M. J. Moore (2000). “The spliceosome deposits multiple protein 20-24 nucloetides upstream of mRNA exon-exon junctions.” EMBO **19**(24):6860-6869.

Levine, B. J., N. Chodchoy, W. F. Marzluff and A. I. Skoultchi (1987). "Coupling of replication type histone mRNA levels to DNA synthesis requires the stem-loop sequence at the 3' end of the mRNA." Proc Natl Acad Sci U S A **84**(17): 6189-6193.

Li, H. (2011). "A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data." Bioinformatics **27**(21): 2987-2993.

Liao, Y., G. K. Smyth and W. Shi (2014). "featureCounts: an efficient general purpose program for assigning sequence reads to genomic features." Bioinformatics **30**(7): 923-930.

Lin, C., E. R. Smith, H. Takahashi, K. C. Lai, S. Martin-Brown, L. Florens, M. P. Washburn, J. W. Conaway, R. C. Conaway and A. Shilatifard (2010). "AFF4, a component of the ELL/P-TEFb elongation complex and a shared subunit of MLL chimeras, can link transcription elongation to leukemia." Mol Cell **37**(3): 429-437.

Liu, J., Y. Yue, D. Han, X. Wang, Y. Fu, L. Zhang, G. Jia, M. Yu, Z. Lu, X. Deng et al (2014). "A METTL3-METTL14 complex mediates mammalian nuclear RNA N6 -adenosine methylation." Nat. Chem. Biol. **10**: 93–95.

Liu, N., K. I. Zhou, M. Parisien, Q. Dai, L. Diatchenko, T. Pan (2017). "N6 -methyladenosine alters RNA structure to regulate binding of a lowcomplexity protein." Nucleic Acids Res **45**(10): 6051-6063.

Liu, X., W. L. Kraus and X. Bai (2015). "Ready, pause, go: regulation of RNA polymerase II pausing and release by cellular signaling pathways." Trends Biochem Sci **40**(9): 516-525.

Liu, Y., S. Li, Y. Chen, A. N. Kimberlin, E. B. Cahoon and B. Yu (2016). "snRNA 3' End Processing by a CPSF73-Containing Complex Essential for Development in Arabidopsis." PLoS Biol **14**(10): e1002571.

Logan, J., E. Falck-Pedersen, J. E. Darnell and T. Shenk (1987). "A poly(A) addition site and a downstream termination region are required for efficient cessation of transcription by RNA polymerase II in the mouse beta maj-globin gene." Proc Natl Acad Sci U S A **84**(23): 8306-8310.

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

Lorentzen, E., J. Basquin, R. Tomecki, A. Dziembowski and E. Conti (2008). "Structure of the active subunit of the yeast exosome core, Rrp44: diverse modes of substrate recruitment in the RNase II nuclease family." Mol Cell **29**(6): 717-728.

Lubas, M., M. S. Christensen, M. S. Kristiansen, M. Domanski, L. G. Falkenby, S. Lykke-Andersen, J. S. Andersen, A. Dziembowski and T. H. Jensen (2011). "Interaction profiling identifies the human nuclear exosome targeting complex." Mol Cell **43**(4): 624-637.

Lubas, M., C. K. Damgaard, R. Tomecki, D. Cysewski, T. H. Jensen and A. Dziembowski (2013). "Exonuclease hDIS3L2 specifies an exosome-independent 3'-5' degradation pathway of human cytoplasmic mRNA." EMBO J **32**(13): 1855-1868.

Luo, W., A. W. Johnson and D. L. Bentley (2006). "The role of Rat1 in coupling mRNA 3'-end processing to transcription termination: implications for a unified allosteric-torpedo model." Genes Dev **20**(8): 954-965.

Luo, Z., C. Lin and A. Shilatifard (2012). "The super elongation complex (SEC) family in transcriptional control." Nat Rev Mol Cell Biol **13**(9): 543-547.

Lykke-Andersen, S., R. Tomecki, T. H. Jensen and A. Dziembowski (2011). "The eukaryotic RNA exosome: same scaffold but variable catalytic subunits." RNA Biol **8**(1): 61-66.

Ma, Y., U. Pannicke, K. Schwarz and M. R. Lieber (2002). "Hairpin opening and overhang processing by an Artemis/DNA-dependent protein kinase complex in nonhomologous end joining and V(D)J recombination." Cell **108**(6): 781-794.

Makino, D. L., B. Schuch, E. Stegmann, M. Baumgärtner, C. Basquin and E. Conti (2015). "RNA degradation paths in a 12-subunit nuclear exosome complex." Nature **524**(7563): 54-58.

Malecki, M., S. C. Viegas, T. Cameiro, P. Golik, C. Dressaires, M. G. Ferreira, C. M. Arraiano (2013). "The exoribonuclease DISL2 defines a novel eukaryotic RNA degradation pathway." EMBO **32**(13): 1842-1854.

Mali, P., K. M. Esvelt and G. M. Church (2013a). "Cas9 as a versatile tool for engineering biology." Nat Methods **10**(10): 957-963.

Mali, P., L. Yang, K. M. Esvelt, J. Aach, M. Guell, J. E. DiCarlo, J. E. Norville and G. M. Church (2013b). "RNA-guided human genome engineering via Cas9." Science **339**(6121): 823-826.

Mandel, C. R., S. Kaneko, H. Zhang, D. Gebauer, V. Vethantham, J. L. Manley and L. Tong (2006). "Polyadenylation factor CPSF-73 is the pre-mRNA 3'-end-processing endonuclease." Nature **444**(7121): 953-956.

Mapendano, C. K., S. Lykke-Andersen, J. Kjems, E. Bertrand and T. H. Jensen (2010). "Crosstalk between mRNA 3' end processing and transcription initiation." Mol Cell **40**(3): 410-422.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. . EMBnet J **17 (1): 10-12**.

Martinez-Contreras, R., P. Cloutier, L. Shkreta, J. F. Fisette, T. Revil, B. Chabot (2007). "hnRNP proteins and splicing control." Adv. Exp. Med. Biol **623**:123–47

Marzluff, W. F., P. Gongidi, K. R. Woods, J. Jin and L. J. Maltais (2002). "The human and mouse replication-dependent histone genes." Genomics **80**(5): 487-498.

Marzluff, W. F. and K. P. Koreski (2017). "Birth and Death of Histone mRNAs." Trends Genet **33**(10): 745-759.

Mathy, N., L. Bénard, O. Pellegrini, R. Daou, T. Wen and C. Condon (2007). "5'-to-3' exoribonuclease activity in bacteria: role of RNase J1 in rRNA maturation and 5' stability of mRNA." Cell **129**(4): 681-692.

Matoulka, E., E. Michalova, B. Vojtesek, R. Hrstka (2012). "The role of the 3' untranslated region in post-transcriptional regulation of protein expression in mammalian cells." RNA Biol **9**(5): 563-576.

Mayer, A., M. Lidschreiber, M. Siebert, K. Leike, J. Söding and P. Cramer (2010). "Uniform transitions of the general RNA polymerase II transcription complex." Nat Struct Mol Biol **17**(10): 1272-1278.

Medlin, J., A. Scurry, A. Taylor, F. Zhang, B. M. Peterlin and S. Murphy (2005). "P-TEFb is not an essential elongation factor for the intronless human U2 snRNA and histone H2b genes." EMBO J **24**(23): 4154-4165.

Mendoza-Ochoa, G. I., J. D. Barrass, B. R. Terlouw, I. E. Maudlin, S. de Lucas, E. Sani, V. Aslanzadeh, J. A. E. Reid and J. D. Beggs (2019). "A fast and tuneable auxin-inducible degron for depletion of target proteins in budding yeast." Yeast **36**(1): 75-81.

Meola, N., M. Domanski, E. Karadoulama, Y. Chen, C. Gentil, D. Pultz, K. Vitting-Seerup, S. Lykke-Andersen, J. S. Andersen, A. Sandelin and T. H. Jensen (2016). "Identification of a Nuclear Exosome Decay Pathway for Processed Transcripts." Mol Cell **64**(3): 520-533.

Mitchell, P. (2014). "Exosome substrate targeting: the long and short of it." Biochem Soc Trans **42**(4): 1129-1134.

Mitchell, P., E. Petfalski, A. Shevchenko, M. Mann and D. Tollervey (1997). "The exosome: a conserved eukaryotic RNA processing complex containing multiple 3'-->5' exoribonucleases." Cell **91**(4): 457-466.

Moore, M. J. (2005). “From birth to death: The complex lives of eukaryotic mRNAs.” Science **309**(5740): 1514-518.

Morawska, M. and H. D. Ulrich (2013). "An expanded tool kit for the auxin-inducible degron system in budding yeast." Yeast **30**(9): 341-351.

Mullen, T. E. and W. F. Marzluff (2008). "Degradation of histone mRNA requires oligouridylation followed by decapping and simultaneous degradation of the mRNA both 5' to 3' and 3' to 5'." Genes Dev **22**(1): 50-65.

Mäder, U., L. Zig, J. Kretschmer, G. Homuth and H. Putzer (2008). "mRNA processing by RNases J1 and J2 affects Bacillus subtilis gene expression on a global scale." Mol Microbiol **70**(1): 183-196.

Narita, T., T. M. Yung, J. Yamamoto, Y. Tsuboi, H. Tanabe, K. Tanaka, Y. Yamaguchi and H. Handa (2007). "NELF interacts with CBC and participates in 3' end processing of replication-dependent histone mRNAs." Mol Cell **26**(3): 349-365.

Natsume, T., T. Kiyomitsu, Y. Saga and M. T. Kanemaki (2016). "Rapid Protein Depletion in Human Cells by Auxin-Inducible Degron Tagging with Short Homology Donors." Cell Rep **15**(1): 210-218.

Neil, H., C. Malabat, Y. d'Aubenton-Carafa, Z. Xu, L. M. Steinmetz and A. Jacquier (2009). "Widespread bidirectional promoters are the major source of cryptic transcripts in yeast." Nature **457**(7232): 1038-1042.

Neph, S., M. S. Kuehn, A. P. Reynolds, E. Haugen, R. E. Thurman, A. K. Johnson, E. Rynes, M. T. Maurano, J. Vierstra, S. Thomas, R. Sandstrom, R. Humbert and J. A. Stamatoyannopoulos (2012). "BEDOPS: high-performance genomic feature operations." Bioinformatics **28**(14): 1919-1920.

Ni, Z., B. E. Schwartz, J. Werner, J. R. Suarez and J. T. Lis (2004). "Coordination of transcription, RNA processing, and surveillance by P-TEFb kinase on heat shock genes." Mol Cell **13**(1): 55-65.

Nishimura, K. and T. Fukagawa (2017). "An efficient method to generate conditional knockout cell lines for essential genes by combination of auxin-inducible degron tag and CRISPR/Cas9." Chromosome Res **25**(3-4): 253-260.

Nishimura, K., T. Fukagawa, H. Takisawa, T. Kakimoto and M. Kanemaki (2009). "An auxin-based degron system for the rapid depletion of proteins in nonplant cells." Nat Methods **6**(12): 917-922.

Ntini, E., A. I. Järvelin, J. Bornholdt, Y. Chen, M. Boyd, M. Jørgensen, R. Andersson, I. Hoof, A. Schein, P. R. Andersen, P. K. Andersen, P. Preker, E. Valen, X. Zhao, V. Pelechano, L. M. Steinmetz, A. Sandelin and T. H. Jensen (2013). "Polyadenylation site-induced decay of upstream transcripts enforces promoter directionality." Nat Struct Mol Biol **20**(8): 923-928.

Nudler, E. (2012). "RNA polymerase backtracking in gene regulation and genome instability." Cell **149**(7): 1438-1445.

Nudler, E., A. Mustaev, E. Lukhtanov and A. Goldfarb (1997). "The RNA-DNA hybrid maintains the register of transcription by preventing backtracking of RNA polymerase." Cell **89**(1): 33-41.

Okada-Katsuhata, Y., A. Yamshita, K. Kutsuzawa, N. Izumi, F. Hirahara, S. Ohno (2012). “N- and C-terminal Upf1 phosphorylations create binding platforms for SMG-6 and SMG-5:SMG-7 during NMD.” Nucleic Acids Res **40**(3): 1251-1266.

O'Reilly, D., M. Dienstbier, S. A. Cowley, P. Vazquez, M. Drozdz, S. Taylor, W. S. James and S. Murphy (2013). "Differentially expressed, variant U1 snRNAs regulate gene expression in human cells." Genome Res **23**(2): 281-291.

O'Reilly, D., O. V. Kuznetsova, C. Laitem, J. Zaborowska, M. Dienstbier and S. Murphy (2014). "Human snRNA genes use polyadenylation factors to promote efficient transcription termination." Nucleic Acids Res **42**(1): 264-275.

Ogami, K., Y. Chen and J. L. Manley (2018). "RNA surveillance by the nuclear RNA exosome: mechanisms and significance." Noncoding RNA **4**(1).

Osheim, Y. N., N. J. Proudfoot and A. L. Beyer (1999). "EM visualization of transcription by RNA polymerase II: downstream termination requires a poly(A) signal but not transcript cleavage." Mol Cell **3**(3): 379-387.

Osheim, Y. N., M. L. Sikes and A. L. Beyer (2002). "EM visualization of Pol II genes in Drosophila: most genes terminate without prior 3' end cleavage of nascent transcripts." Chromosoma **111**(1): 1-12.

Perry, R. P., D. E. Kelley (1974). “Existence of methylated messenger RNA in mouse L cells.” Cell **1**: 37–42.

Peterlin, B. M. and D. H. Price (2006). "Controlling the elongation phase of transcription with P-TEFb." Mol Cell **23**(3): 297-305.

Pettinati, I., P. Grzechnik, C. Ribeiro de Almeida, J. Brem, M. A. McDonough, S. Dhir, N. J. Proudfoot and C. J. Schofield (2018). "Biosynthesis of histone messenger RNA employs a specific 3' end endonuclease." Elife **7**.

Piccirillo, C., R. Khanna and M. Kiledjian (2003). "Functional characterization of the mammalian mRNA decapping enzyme hDcp2." RNA **9**(9): 1138-1147.

Pillai, R. S., M. Grimmler, G. Meister, C. L. Will, R. Lührmann, U. Fischer and D. Schümperli (2003). "Unique Sm core structure of U7 snRNPs: assembly by a specialized SMN complex and the role of a new component, Lsm11, in histone RNA processing." Genes Dev **17**(18): 2321-2333.

Pillai, R. S., C. L. Will, R. Lührmann, D. Schümperli and B. Müller (2001). "Purified U7 snRNPs lack the Sm proteins D1 and D2 but contain Lsm10, a new 14 kDa Sm D1-like protein." EMBO J **20**(19): 5470-5479.

Ping, X. L., B. F. Sun, L. Wang, W. Xiao, X. Yang, W. J. Wang, S. Adhikari, Y. Shi, Y. Lv, Y. S. Chen et al. (2014). “Mammalian WTAP is a regulatory subunit of the RNA N6 -methyladenosine methyltransferase.” Cell Res. **24**, 177–189

Ping, Y. H. and T. M. Rana (2001). "DSIF and NELF interact with RNA polymerase II elongation complex and HIV-1 Tat stimulates P-TEFb-mediated phosphorylation of RNA polymerase II and DSIF during transcription elongation." J Biol Chem **276**(16): 12951-12958.

Popp, M. W. L., L. E. Maquat (2013). “Organising principles of mammalian nonsense-mediated mRNA decay.” Annu Rev Genet **47**: 139-165.

Preker, P., K. Almvig, M. S. Christensen, E. Valen, C. K. Mapendano, A. Sandelin and T. H. Jensen (2011). "PROMoter uPstream Transcripts share characteristics with mRNAs and are produced upstream of all three major types of mammalian promoters." Nucleic Acids Res **39**(16): 7179-7193.

Preker, P., J. Nielsen, S. Kammler, S. Lykke-Andersen, M. S. Christensen, C. K. Mapendano, M. H. Schierup and T. H. Jensen (2008). "RNA exosome depletion reveals transcription upstream of active human promoters." Science **322**(5909): 1851-1854.

Proufoot, N. J., A. Furger, M. J. Dye (2002). “Integrating mRNA processing with transcription.” Cell **108**(4): 501-512.

Proudfoot, N. J. (2011). "Ending the message: poly(A) signals then and now." Genes Dev **25**(17): 1770-1782.

Quinlan, A. R. and I. M. Hall (2010). "BEDTools: a flexible suite of utilities for comparing genomic features." Bioinformatics **26**(6): 841-842.

Ramamurthy, L., T. C. Ingledue, D. R. Pilch, B. K. Kay and W. F. Marzluff (1996). "Increasing the distance between the snRNA promoter and the 3' box decreases the efficiency of snRNA 3'-end formation." Nucleic Acids Res **24**(22): 4525-4534.

Ramanathan, A., G. B. Robb and S. H. Chan (2016). "mRNA capping: biological functions and applications." Nucleic Acids Res **44**(16): 7511-7526.

Ramírez, F., D. P. Ryan, B. Grüning, V. Bhardwaj, F. Kilpert, A. S. Richter, S. Heyne, F. Dündar and T. Manke (2016). "deepTools2: a next generation web server for deep-sequencing data analysis." Nucleic Acids Res **44**(W1): W160-165.

Rath, A., M. Glibowicka, V. G. Nadeau, G. Chen and C. M. Deber (2009). "Detergent binding explains anomalous SDS-PAGE migration of membrane proteins." Proc Natl Acad Sci U S A **106**(6): 1760-1765.

Reinberg, D. and R. G. Roeder (1987). "Factors involved in specific transcription by mammalian RNA polymerase II. Transcription factor IIS stimulates elongation of RNA chains." J Biol Chem **262**(7): 3331-3337.

Reis, C. C. and J. L. Campbell (2007). "Contribution of Trf4/5 and the nuclear exosome to genome stability through regulation of histone mRNA levels in Saccharomyces cerevisiae." Genetics **175**(3): 993-1010.

Rienzo, M. and A. Casamassimi (2016). "Integrator complex and transcription regulation: Recent findings and pathophysiology." Biochim Biophys Acta **1859**(10): 1269-1280.

Robinson, J. T., H. Thorvaldsdóttir, W. Winckler, M. Guttman, E. S. Lander, G. Getz and J. P. Mesirov (2011). "Integrative genomics viewer." Nat Biotechnol **29**(1): 24-26.

Robinson, S. R., A. W. Oliver, T. J. Chevassut and S. F. Newbury (2015). "The 3' to 5' Exoribonuclease DIS3: From Structure and Mechanisms to Biological Functions and Role in Human Disease." Biomolecules **5**(3): 1515-1539.

Romeo, V., E. Griesbach and D. Schümperli (2014). "CstF64: cell cycle regulation and functional role in 3' end processing of replication-dependent histone mRNAs." Mol Cell Biol **34**(23): 4272-4284.

Roundtree, I. A., M. E. Evans, T. Pan, C. He (2017). “Dynamic RNA modifications in gene expression regulation.” Cell **169**(7): 1187-1200.

Ryan, K., O. Calvo and J. L. Manley (2004). "Evidence that polyadenylation factor CPSF-73 is the mRNA 3' processing endonuclease." RNA **10**(4): 565-573.

Saldi, T., M. A. Cortazar, R. M. Sheridan and D. L. Bentley (2016). "Coupling of RNA Polymerase II Transcription Elongation with Pre-mRNA Splicing." J Mol Biol **428**(12): 2623-2635.

Sathyan, K. M., B. D. McKenna, W. Anderson , F. M. Duarte , L. J. Core and M. J. Guertin (2019). An improved auxin-inducible degron system preserves native protein levels and enables rapid and specific protein depletion. bioRxiv.

Schaeffer, D., B. Tsanova, A. Barbas, F. P. Reis, E. G. Dastidar, M. Sanchez-Rotunno, C. M. Arraiano and A. van Hoof (2009). "The exosome contains domains with specific endoribonuclease, exoribonuclease and cytoplasmic mRNA decay activities." Nat Struct Mol Biol **16**(1): 56-62.

Schindelin, J., I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, B. Schmid, J. Y. Tinevez, D. J. White, V. Hartenstein, K. Eliceiri, P. Tomancak and A. Cardona (2012). "Fiji: an open-source platform for biological-image analysis." Nat Methods **9**(7): 676-682.

Schmidt, K. and J. S. Butler (2013). "Nuclear RNA surveillance: role of TRAMP in controlling exosome specificity." Wiley Interdiscip Rev RNA **4**(2): 217-231.

Schneider, C., G. Kudla, W. Wlotzka, A. Tuck and D. Tollervey (2012). "Transcriptome-wide analysis of exosome targets." Mol Cell **48**(3): 422-433.

Schneider, C., E. Leung, J. Brown and D. Tollervey (2009). "The N-terminal PIN domain of the exosome subunit Rrp44 harbors endonuclease activity and tethers Rrp44 to the yeast core exosome." Nucleic Acids Res **37**(4): 1127-1140.

Schones, D. E., K. Cui, S. Cuddapah, T. Y. Roh, A. Barski, Z. Wang, G. Wei and K. Zhao (2008). "Dynamic regulation of nucleosome positioning in the human genome." Cell **132**(5): 887-898.

Segal, E., Y. Fondufe-Mittendorf, L. Chen, A. Thåström, Y. Field, I. K. Moore, J. P. Wang and J. Widom (2006). "A genomic code for nucleosome positioning." Nature **442**(7104): 772-778.

Shi, Y., R. A. Mowery, J. Ashley, M. Hentz, A. J. Ramirez, B. Bilgicer, H. Slunt-Brown, D. R. Borchelt and B. F. Shaw (2012). "Abnormal SDS-PAGE migration of cytosolic proteins can identify domains and mechanisms that control surfactant binding." Protein Sci **21**(8): 1197-1209.

Shuman, S. (2001). "Structure, mechanism, and evolution of the mRNA capping apparatus." Prog Nucleic Acid Res Mol Biol **66**: 1-40.

Sigurdsson, S., A. B. Dirac-Svejstrup and J. Q. Svejstrup (2010). "Evidence that transcript cleavage is essential for RNA polymerase II transcription and cell viability." Mol Cell **38**(2): 202-210.

Singh, J. and R. A. Padgett (2009). "Rates of in situ transcription and splicing in large human genes." Nat Struct Mol Biol **16**(11): 1128-1133.

Siwaszek, A., M. Ukleja, A. Dziembowski (2014). “Proteins involved in the degradation of cytoplasmic mRNA in the major eukaryotic model systems.” RNA Biol **11**(9): 1122-1136.

Skaar, J. R., A. L. Ferris, X. Wu, A. Saraf, K. K. Khanna, L. Florens, M. P. Washburn, S. H. Hughes and M. Pagano (2015). "The Integrator complex controls the termination of transcription at diverse classes of gene targets." Cell Res **25**(3): 288-305.

Skaar, J. R., D. J. Richard, A. Saraf, A. Toschi, E. Bolderson, L. Florens, M. P. Washburn, K. K. Khanna and M. Pagano (2009). "INTS3 controls the hSSB1-mediated DNA damage response." J Cell Biol **187**(1): 25-32.

Skipper, K. A., P. R. Andersen, N. Sharma and J. G. Mikkelsen (2013). "DNA transposon-based gene vehicles - scenes from an evolutionary drive." J Biomed Sci **20**: 92.

Skourti-Stathaki, K., N. J. Proudfoot and N. Gromak (2011). "Human senataxin resolves RNA/DNA hybrids formed at transcriptional pause sites to promote Xrn2-dependent termination." Mol Cell **42**(6): 794-805.

Skrajna, A., X. C. Yang, K. Bucholc, J. Zhang, T. M. T. Hall, M. Dadlez, W. F. Marzluff and Z. Dominski (2017). "U7 snRNP is recruited to histone pre-mRNA in a FLASH-dependent manner by two separate regions of the stem-loop binding protein." RNA **23**(6): 938-951.

Slevin, M. K., S. Meaux, J. D. Welch, R. Bigler, P. L. Miliani de Marval, W. Su, R. E. Rhoads, J. F. Prins and W. F. Marzluff (2014). "Deep sequencing shows multiple oligouridylations are required for 3' to 5' degradation of histone mRNAs on polyribosomes." Mol Cell **53**(6): 1020-1030.

Smith, I., P. G. Greenside, T. Natoli, D. L. Lahr, D. Wadden, I. Tirosh, R. Narayan, D. E. Root, T. R. Golub, A. Subramanian and J. G. Doench (2017). "Evaluation of RNAi and CRISPR technologies by large-scale gene expression profiling in the Connectivity Map." PLoS Biol **15**(11): e2003213.

Sontheimer, E. J. and J. A. Steitz (1992). "Three novel functional variants of human U5 small nuclear RNA." Mol Cell Biol **12**(2): 734-746.

Staals, R. H., A. W. Bronkhorst, G. Schilders, S. Slomovic, G. Schuster, A. J. Heck, R. Raijmakers and G. J. Pruijn (2010). "Dis3-like 1: a novel exoribonuclease associated with the human exosome." EMBO J **29**(14): 2358-2367.

Stadelmayer, B., G. Micas, A. Gamot, P. Martin, N. Malirat, S. Koval, R. Raffel, B. Sobhian, D. Severac, S. Rialle, H. Parrinello, O. Cuvier and M. Benkirane (2014). "Integrator complex regulates NELF-mediated RNA polymerase II pause/release and processivity at coding genes." Nat Commun **5**: 5531.

Sullivan, E., C. Santiago, E. D. Parker, Z. Dominski, X. Yang, D. J. Lanzotti, T. C. Ingledue, W. F. Marzluff and R. J. Duronio (2001). "Drosophila stem loop binding protein coordinates accumulation of mature histone mRNA with cell cycle progression." Genes Dev **15**(2): 173-187.

Sullivan, K. D., M. Steiniger and W. F. Marzluff (2009). "A core complex of CPSF73, CPSF100, and Symplekin may form two different cleavage factors for processing of poly(A) and histone mRNAs." Mol Cell **34**(3): 322-332.

Sun, J., H. Pan, C. Lei, B. Yuan, S. J. Nair, C. April, B. Parameswaran, B. Klotzle, J. B. Fan, J. Ruan and R. Li (2011). "Genetic and genomic analyses of RNA polymerase II-pausing factor in regulation of mammalian transcription and cell growth." J Biol Chem **286**(42): 36248-36257.

Suraweera, A., Y. Lim, R. Woods, G. W. Birrell, T. Nasim, O. J. Becherel and M. F. Lavin (2009). "Functional role for senataxin, defective in ataxia oculomotor apraxia type 2, in transcriptional regulation." Hum Mol Genet **18**(18): 3384-3396.

Szczepińska, T., K. Kalisiak, R. Tomecki, A. Labno, L. S. Borowski, T. M. Kulinski, D. Adamska, J. Kosinska and A. Dziembowski (2015). "DIS3 shapes the RNA polymerase II transcriptome in humans by degrading a variety of unwanted transcripts." Genome Res **25**(11): 1622-1633.

Tan, X., L. I. Calderon-Villalobos, M. Sharon, C. Zheng, C. V. Robinson, M. Estelle and N. Zheng (2007). "Mechanism of auxin perception by the TIR1 ubiquitin ligase." Nature **446**(7136): 640-645.

Teale, W. D., I. A. Paponov and K. Palme (2006). "Auxin in action: signalling, transport and the control of plant growth and development." Nat Rev Mol Cell Biol **7**(11): 847-859.

Thiebaut, M., E. Kisseleva-Romanova, M. Rougemaille, J. Boulay and D. Libri (2006). "Transcription termination and nuclear degradation of cryptic unstable transcripts: a role for the nrd1-nab3 pathway in genome surveillance." Mol Cell **23**(6): 853-864.

Tietjen, J. R., D. W. Zhang, J. B. Rodríguez-Molina, B. E. White, M. S. Akhtar, M. Heidemann, X. Li, R. D. Chapman, K. Shokat, S. Keles, D. Eick and A. Z. Ansari (2010). "Chemical-genomic dissection of the CTD code." Nat Struct Mol Biol **17**(9): 1154-1161.

Tomecki, R., M. S. Kristiansen, S. Lykke-Andersen, A. Chlebowski, K. M. Larsen, R. J. Szczesny, K. Drazkowska, A. Pastula, J. S. Andersen, P. P. Stepien, A. Dziembowski and T. H. Jensen (2010). "The human core exosome interacts with differentially localized processive RNases: hDIS3 and hDIS3L." EMBO J **29**(14): 2342-2357.

Trask, D. K. and M. T. Muller (1988). "Stabilization of type I topoisomerase-DNA covalent complexes by actinomycin D." Proc Natl Acad Sci U S A **85**(5): 1417-1421.

Tsao, D. C., N. J. Park, A. Nag and H. G. Martinson (2012). "Prolonged α-amanitin treatment of cells for studying mutated polymerases causes degradation of DSIF160 and other proteins." RNA **18**(2): 222-229.

Tseng, C. K., H. F. Wang, A. M. Burns, M. R. Schroeder, M. Gaspari and P. Baumann (2015). "Human Telomerase RNA Processing and Quality Control." Cell Rep **13**(10): 2232-2243.

Varshney, D., O. Lombardi, G. Schweikert, S. Dunn, O. Suska, V. G. Cowling (2018). “mRNA cap methyltransferase, RNMT-RAM, promotes RNA Pol II-dependent transcription.” Cell Rep **23**(5): 1530-1542.

Wagner, J (2010). "snRNA 3' end formation: the dawn of the Integrator complex." Biochem Soc Trans **38**(4): 1082-1087.

Wagner, E. J. and W. F. Marzluff (2006). "ZFP100, a component of the active U7 snRNP limiting for histone pre-mRNA processing, is required for entry into S phase." Mol Cell Biol **26**(17): 6702-6712.

Wagschal, A., E. Rousset, P. Basavarajaiah, X. Contreras, A. Harwig, S. Laurent-Chabalier, M. Nakamura, X. Chen, K. Zhang, O. Meziane, F. Boyer, H. Parrinello, B. Berkhout, C. Terzian, M. Benkirane and R. Kiernan (2012). "Microprocessor, Setx, Xrn2, and Rrp6 co-operate to induce premature termination of transcription by RNAPII." Cell **150**(6): 1147-1157.

Wang, E. T., R. Sandberg, S. Luo, I. Khrebtukova, L. Zhang, C. Mayr, S. F. Kingsmore, G. P. Schroth and C. B. Burge (2008). "Alternative isoform regulation in human tissue transcriptomes." Nature **456**(7221): 470-476.

Wang, P., K. A. Doxtader, Y. Nam (2016). “Structural basis for cooperative function of Mettl3 and Mettl14 methyltransferases.” Mol. Cell **63**: 306–317.

Wang, Y., J. Liu, B. O. Huang, Y. M. Xu, J. Li, L. F. Huang, J. Lin, J. Zhang, Q. H. Min, W. M. Yang, X. Z. Wang, (2015). “Mechanism of alternative splicing and its regulation.” Biomed Rep. **3**(2): 152-158

Wang, Z. and C. B. Burge (2008). "Splicing regulation: from a parts list of regulatory elements to an integrated splicing code." RNA **14**(5): 802-813.

Ward, A. J. and T. A. Cooper (2010). "The pathobiology of splicing." J Pathol **220**(2): 152-163.

Wasmuth, E. V., K. Januszyk and C. D. Lima (2014). "Structure of an Rrp6-RNA exosome complex bound to poly(A) RNA." Nature **511**(7510): 435-439.

Wasmuth, E. V. and C. D. Lima (2012). "Exo- and endoribonucleolytic activities of yeast cytoplasmic and nuclear RNA exosomes are dependent on the noncatalytic core and central channel." Mol Cell **48**(1): 133-144.

Wei, C. M., A. Gershowitz, B. Moss (1975). “Methylated nucleotides block 5′ terminus of HeLa cell messenger RNA.” Cell **4**: 379–386.

Weick, E. M., M. R. Puno, K. Januszyk, J. C. Zinder, M. A. DiMattia and C. D. Lima (2018). "Helicase-Dependent RNA Decay Illuminated by a Cryo-EM Structure of a Human Nuclear RNA Exosome-MTR4 Complex." Cell **173**(7): 1663-1677.e1621.

West, S., N. Gromak and N. J. Proudfoot (2004). "Human 5' --> 3' exonuclease Xrn2 promotes transcription termination at co-transcriptional cleavage sites." Nature **432**(7016): 522-525.

West, S., N. J. Proudfoot and M. J. Dye (2008). "Molecular dissection of mammalian RNA polymerase II transcriptional termination." Mol Cell **29**(5): 600-610.

Wu, Y., T. R. Albrecht, D. Baillat, E. J. Wagner and L. Tong (2017). "Molecular basis for the interaction between Integrator subunits IntS9 and IntS11 and its functional importance." Proc Natl Acad Sci U S A **114**(17): 4394-4399.

Wyers, F., M. Rougemaille, G. Badis, J. C. Rousselle, M. E. Dufour, J. Boulay, B. Régnault, F. Devaux, A. Namane, B. Séraphin, D. Libri and A. Jacquier (2005). "Cryptic pol II transcripts are degraded by a nuclear quality control pathway involving a new poly(A) polymerase." Cell **121**(5): 725-737.

Xiao, W., S. Adhikari, U. Dahal, Y. S. Chen, Y. J. Hao, B. F. Sun, H. Y. Sun, A. Li, X. L. Ping, W. Y. Lai, et al. (2016). “Nuclear m6 A Reader YTHDC1 Regulates mRNA Splicing.” Mol. Cell **61**: 507–519.

Xie, M., W. Zhang, M. D. Shu, A. Xu, D. A. Lenis, D. DiMaio and J. A. Steitz (2015). "The host Integrator complex acts in transcription-independent maturation of herpesvirus microRNA 3' ends." Genes Dev **29**(14): 1552-1564.

Xu, Z., W. Wei, J. Gagneur, F. Perocchi, S. Clauder-Münster, J. Camblong, E. Guffanti, F. Stutz, W. Huber and L. M. Steinmetz (2009). "Bidirectional promoters generate pervasive transcription in yeast." Nature **457**(7232): 1033-1037.

Yamamoto, J., Y. Hagiwara, K. Chiba, T. Isobe, T. Narita, H. Handa and Y. Yamaguchi (2014). "DSIF and NELF interact with Integrator to specify the correct post-transcriptional fate of snRNA genes." Nat Commun **5**: 4263.

Yamashita, A., T. C. Chang, Y. Yamashita, W. Whu, Z. Zhong, C. Y. Chen, A. B. Shyu (2005). “Concerted action of poly(A) nucleases and decapping enzyme in mammalian mRNA turnover. Nat Struct Mol Biol **12**:1054-1063.

Yan, D., M. Weisshaar, K. Lamb, H. K. Chung, M. Z. Lin and R. K. Plemper (2015). "Replication-Competent Influenza Virus and Respiratory Syncytial Virus Luciferase Reporter Strains Engineered for Co-Infections Identify Antiviral Compounds in Combination Screens." Biochemistry **54**(36): 5589-5604.

Yang, X. C., B. D. Burch, Y. Yan, W. F. Marzluff and Z. Dominski (2009a). "FLASH, a proapoptotic protein involved in activation of caspase-8, is essential for 3' end processing of histone pre-mRNAs." Mol Cell **36**(2): 267-278.

Yang, X. C., I. Sabath, J. Dębski, M. Kaus-Drobek, M. Dadlez, W. F. Marzluff and Z. Dominski (2013). "A complex containing the CPSF73 endonuclease and other polyadenylation factors associates with U7 snRNP and is recruited to histone pre-mRNA for 3'-end processing." Mol Cell Biol **33**(1): 28-37.

Yang, X. C., K. D. Sullivan, W. F. Marzluff and Z. Dominski (2009b). "Studies of the 5' exonuclease and endonuclease activities of CPSF-73 in histone pre-mRNA processing." Mol Cell Biol **29**(1): 31-42.

Yant, S. R., L. Meuse, W. Chiu, Z. Ivics, Z. Izsvak and M. A. Kay (2000). "Somatic integration and long-term transgene expression in normal and haemophilic mice using a DNA transposon system." Nat Genet **25**(1): 35-41.

Yoon, J. H., P. Singh, D. H. Lee, J. Qiu, S. Cai, T. R. O'Connor, Y. Chen, B. Shen and G. P. Pfeifer (2005). "Characterization of the 3' --> 5' exonuclease activity found in human nucleoside diphosphate kinase 1 (NDK1) and several of its homologues." Biochemistry **44**(48): 15774-15786.

Zaborowska, J., S. Egloff, S. Murphy (2016). “The Pol II CTD: new twists in the tail.” Nat. Struct. Mol. Biol. **23**: 771-777

Zasadzińska, E., J. Huang, A. O. Bailey, L. Y. Guo, N. S. Lee, S. Srivastava, K. A. Wong, B. T. French, B. E. Black and D. R. Foltz (2018). "Inheritance of CENP-A Nucleosomes during DNA Replication Requires HJURP." Dev Cell **47**(3): 348-362.e347.

Zhang, H., F. Rigo and H. G. Martinson (2015a). "Poly(A) Signal-Dependent Transcription Termination Occurs through a Conformational Change Mechanism that Does Not Require Cleavage at the Poly(A) Site." Mol Cell **59**(3): 437-448.

Zhang, L., J. D. Ward, Z. Cheng and A. F. Dernburg (2015b). "The auxin-inducible degradation (AID) system enables versatile conditional protein depletion in C. elegans." Development **142**(24): 4374-4384.

Zhang, Y., N. Heidrich, B. J. Ampattu, C. W. Gunderson, H. S. Seifert, C. Schoen, J. Vogel and E. J. Sontheimer (2013). "Processing-independent CRISPR RNAs limit natural transformation in Neisseria meningitidis." Mol Cell **50**(4): 488-503.

Zhou, Q., T. Li and D. H. Price (2012). "RNA polymerase II elongation control." Annu Rev Biochem **81**: 119-143.

Zinder, J. C. and C. D. Lima (2017). "Targeting RNA for processing or destruction by the eukaryotic RNA exosome and its cofactors." Genes Dev **31**(2): 88-100.

Zinder, J. C., E. V. Wasmuth and C. D. Lima (2016). "Nuclear RNA Exosome at 3.1 Å Reveals Substrate Specificities, RNA Paths, and Allosteric Inhibition of Rrp44/Dis3." Mol Cell **64**(4): 734-745.

Łabno, A., Z. Warkocki, T. Kuliński, P. S. Krawczyk, K. Bijata, R. Tomecki and A. Dziembowski (2016). "Perlman syndrome nuclease DIS3L2 controls cytoplasmic non-coding RNAs and provides surveillance pathway for maturing snRNAs." Nucleic Acids Res **44**(21): 10437-10453.