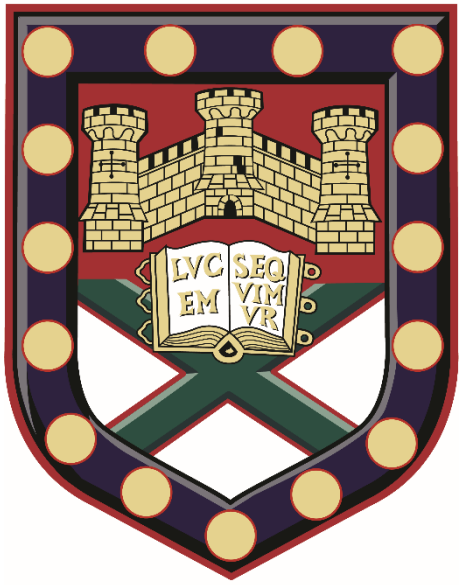
The Mechanisms of Transcription Termination by RNA Polymerase II



Submitted by

JOSHUA DAVID EATON

to the University of Exeter as a thesis for the degree of

DOCTOR OF PHILOSOPHY IN BIOLOGICAL SCIENCES

November 2020

I certify that all material in this thesis which is not my own work has been identified and that any material that has previously been submitted and approved for the award of a degree by this or any other University has been acknowledged.

Josh Eaton

Supervisor: Dr Steven West

Joshua David Eaton

Acknowledgements

Firstly, I would like to thank, Steve, my mentor these past few years who has been a great fountain of inspiration, veracity, and optimism especially at those times when Covid let the doubters surface, but we pulled through… Liverpool Premier League Champions, 2020! YNWA!

I would also like to thank everyone I have had the chance of working alongside. To Chris who is always a source of much amusement, Lee, Laura F., Ryan, Francesca, and Oksana who have been of great help, support and made my time in the group.

Thank you to Laura R., Denzel, Nelly and the others who were always keen to go to the pub/end-of-the-month drinks.

To my good friends who I met in Sheffield, whose long train journeys are always worth the food and drinks if not the company, Lorna, Oli, Hannah, and Ryan.

To Sam and Ruby who have been of source of constant friendship throughout this long ride. Thanks for trusting me with the house on your round the world tour, nothing burnt down, and I only set myself alight once (by accident of course).

Finally, to Niko and Stuart who allowed me to take my first steps in a lab, probably would not have done a PhD without it.

Thank you all, it has been a journey, Josh.

Abstract

RNA polymerase II (Pol II) catalyses the transcription of many RNA classes including protein-coding, small nuclear RNA (snRNAs) and some non-coding RNA classes within eukaryotes. Its journey is ended by the cessation of transcription and the dissolution of the DNA-bound complex in a process known as transcriptional termination. Given the diverse array of transcript classes, several termination mechanisms can induce Pol II termination. One of the most studied is polyadenylation signal (PAS)-dependent termination and occurs at the ends of most protein-coding as well as some other transcript classes. Two long-standing models have been used to explain the role of the PAS in termination and their relevance has been a topic of much debate. The allosteric model suggests Pol II undergoes conformational changes after the PAS to instigate termination, while the torpedo model suggests degradation of the downstream RNA product of PAS-cleavage is important to instigate termination. Here, rapid depletion cell lines are employed to describe the widespread dependence of protein-coding transcript termination on the XRN2 torpedo and the CPSF73 PAS-cleaving endonuclease. CPSF73 depletion leads to profound "run-away" transcription, whereas XRN2 depletion results in a more limited read-through. XRN2 targets Pol II complexes that have undergone slowing or pausing in a protein phosphatase 1 (PP1)-mediated process occurring downstream of the PAS. Additionally, XRN2 can degrade RNA and cause torpedo termination from some other PAS-independent cleavage events. However, this is not a universal process with XRN2 dispensable at snRNA and histone transcripts. Together these results suggest a unified allosteric/torpedo mechanism at protein-coding transcripts, where PAS cleavage precedes a PP1-dependent slowing of Pol II. This facilitates the degradation of downstream RNA by XRN2 and thereby instigates transcriptional termination.

Table of Contents

List of Figures

List of Tables

Abbreviations

AID | Auxin-inducible degron

Bp | Base pair

CA | Calyculin A

CDK | Cyclin-dependent kinase

ChIP | Chromatin immunoprecipitation

CLIP | Cross-linking and immunoprecipitation

CoTC | Co-transcription cleavage element

CPA | Cleavage and polyadenylation

CPSF | Cleavage and polyadenylation specificity factor

CSTF | Cleavage Stimulation Factor

CTD | C-terminal domain of Rbp1, which is the largest subunit, of Pol II

DHFR | Dihydrofolate reductase

DNA | Deoxyribonucleic acid

Dox | Doxycycline

dpCoA | dephosphorylated-CoA

DRB | Dichloro-1-β-D-ribofuranosylbenzimidazole

dsDNA | Double-stranded DNA

DSIF | DRB-sensitive inducing factor

E. coli | *Escherichia coli*

FAD | flavin adenine dinucleotide

FBS | Foetal bovine serum

GB | Gene Body

gDNA | genomic DNA

GEO | Gene Expression Omnibus

HCC | Histone cleavage complex

HDR | Homology-directed repair

IP | Immunoprecipitation

KD | Knock-down

KO | Knock-out

lncRNAs | Long non-coding RNAs

mascRNA | *MALAT1*-associated small cytoplasmic RNA

mini-AID | Minimal auxin-inducible degron domain

MNase | Micrococcal nuclease

mRNA | Messenger RNA

NAD | nicotinamide adenine dinucleotide

NELF | Negative elongation factor

NITC | Nonsense induced transcriptional compensation

NMD | Nonsense-mediated decay

NRD | complex Nrd1-Nab3-Sen1 complex

nt | Nucleotide

MT | Mutant

PABPN1 | Nuclear poly(A) binding protein 1

PAS | Polyadenylation Sequence

PCPA | Premature cleavage and polyadenylation

PIC | Pre-initiation complex

PNUTs | PP1 nuclear targeting subunit

Pol I | RNA Polymerase I

Pol II | RNA Polymerase II

Pol III | RNA Polymerase III

PP1 | Protein phosphatase 1

PROMPTs | Promoter upstream transcripts

PTC | Premature termination codon

P-TEFb | Positive transcription elongation factor b

qPCR | Quantitative polymerase chain reaction

qRT-PCR | Quantitative reverse transcription and polymerase chain reaction

RNA | Ribonucleic acid

RNAi | RNA interference

RNAP | Bacterial RNA polymerase

RTI | Read-through index

Rut | Rho utilisation sequence

SB | Sleeping Beauty (transposon system)

S. py | *Streptococcus pyogenes*

Ser2 | Serine 2 of the heptad repeat (Y**S**PTSPS) within the Pol II CTD

Ser5 | Serine 5 of the heptad repeat (YSPT**S**PS) within the Pol II CTD

Ser7 | Serine 7 of the heptad repeat (YSPTSP**S**) within the Pol II CTD

smFRET | Single molecule Förster resonance energy transfer

SNR | Single nucleotide resolution

snRNAs | Small nuclear RNA

snRNP | Small nuclear ribonucleoprotein

ssRNA | Single-stranded RNA

TBP | TATA box-binding protein

TCR | Transcription coupled repair

TES | Transcription end site

Thr4 | Threonine 4 of the heptad repeat (YSP**T**SPS) within the Pol II CTD

TSS | Transcription start site

TU | Transcription Unit

Tyr1 | Tyrosine 1 of the heptad repeat (**Y**SPTSPS) within the Pol II CTD

UTR | Untranslated region

WB | Western blot

WT | Wild-type

xrRNA | XRN-resistant RNA

YSPTSPS | Consensus heptad repeat of the Pol II CTD

δRZ | hepatitis δ ribozyme

σ-factor | Sigma factor

Chapter 1

1.Introduction

Our understanding of eukaryotic gene expression has developed greatly over the past three decades, from the discovery that polyadenylation signal (PAS) sequences at the end of genes couple RNA 3'-end processing and transcriptional termination, to the crystal structures revealing initiating RNA Polymerase II (Pol II) and the subsequent explosion of transcriptomic sequencing methods that have given unparalleled genome-wide insights such as bidirectional transcription from promoters (Proudfoot and Brownlee, 1976; Connelly and Manley, 1988; Cramer et al., 2000; Preker et al., 2008; Seila et al., 2008). However, many salient questions persist especially surrounding the final stage of the transcription cycle, termination, where the conformational details of polymerase dissolution from the DNA template remain one of the most enigmatic aspects of transcription. This is largely due to the non-trivial and fundamental challenges of isolating RNA polymerase complexes immediately before and after dissociation from the DNA template. Such a process will likely require detailed mechanistic understanding before isolation of the Pol II complexes containing the components with posttranscriptional modifications that immediately precede template dissolution can occur for structural studies. This thesis focuses on the understanding of this last stage of transcription, specifically the transcriptional termination mechanisms of Pol II within humans.

Pol II is one of three DNA-dependent RNA polymerases within the human cell nuclei and it catalyses the transcription of a diverse array of transcript classes including protein-coding messenger RNAs (mRNAs), small nuclear RNAs (snRNAs) and long non-coding RNAs (lncRNAs). There are thought to be several Pol II termination mechanisms for these transcripts with termination able to occur at almost any position whether that be prematurely or for the generation of mature transcripts (Kamieniarz-Gdula and Proudfoot, 2019; Mendoza-Figueroa et al., 2020; Proudfoot, 2016). The most studied of these is PAS-dependent termination at the 3'-end of protein-coding transcripts. This is critical for the completion of a functional protein-coding mRNA by

ensuring PAS cleavage is coupled to polyadenylation and timely termination without transcriptional interference of downstream transcripts *in cis* (Greger et al., 2000; Shearwin et al., 2005). The aims, set out at the end of this chapter, are to examine the contributions that PAS cleavage, 5'→3' exonuclease degradation of the downstream PAS cleavage product and allosteric phosphatase-mediated remodelling have on Pol II termination genome-wide. The data presented here argue for PAS-dependent termination being a ubiquitous termination pathway for protein-coding transcripts. This process depends on the cleavage endonuclease CPSF73 and likely its RNA endonuclease activity without which termination readily fails to occur. Additionally, RNA degradation of the cleaved 5'-end at the PAS by the exonuclease XRN2 accelerates termination but upon XRN2 depletion termination still occurs and Pol II accumulates. The Pol II peak after the PAS, which is enriched by XRN2 depletion, is because of a reduced elongation rate which is mediated by PP1 phosphatase activity. However, before a more detailed discussion of the questions this thesis seeks to address and the rationale, first comes an introduction into DNA-dependent transcription.

1.1 An Overview of Transcription

Transcription is the synthesis of single-stranded RNA (ssRNA) complementary to a starting nucleic acid template. In all domains of life, from prokaryotes to higher eukaryotes, DNA-dependent RNA polymerases catalyse the synthesis of RNA using an organism's double-stranded DNA (dsDNA) genome as template. This nascent RNA is always synthesised in a 5'→3' direction by the assembly of complementary ribonucleotides triphosphates upon one of the DNA strands, termed template strand, within the polymerase active site (**Fig 1.1a**). The remaining displaced DNA strand is referred to as the coding (or non-template) strand because the deoxyribonucleotides within it are analogous to the newly created RNA with only thymine within DNA substituted for uracil. The genomic region where a transcription event occurs is termed a transcription unit (TU), starting at the site of RNA polymerase initiation until the position of termination. Transcription is the core mechanism underpinning gene expression, where genes are described as a functional and heritable unit within DNA required to produce an RNA and protein. In bacteria, some genes are found clustered within one TU, called an operon, and this leads to the

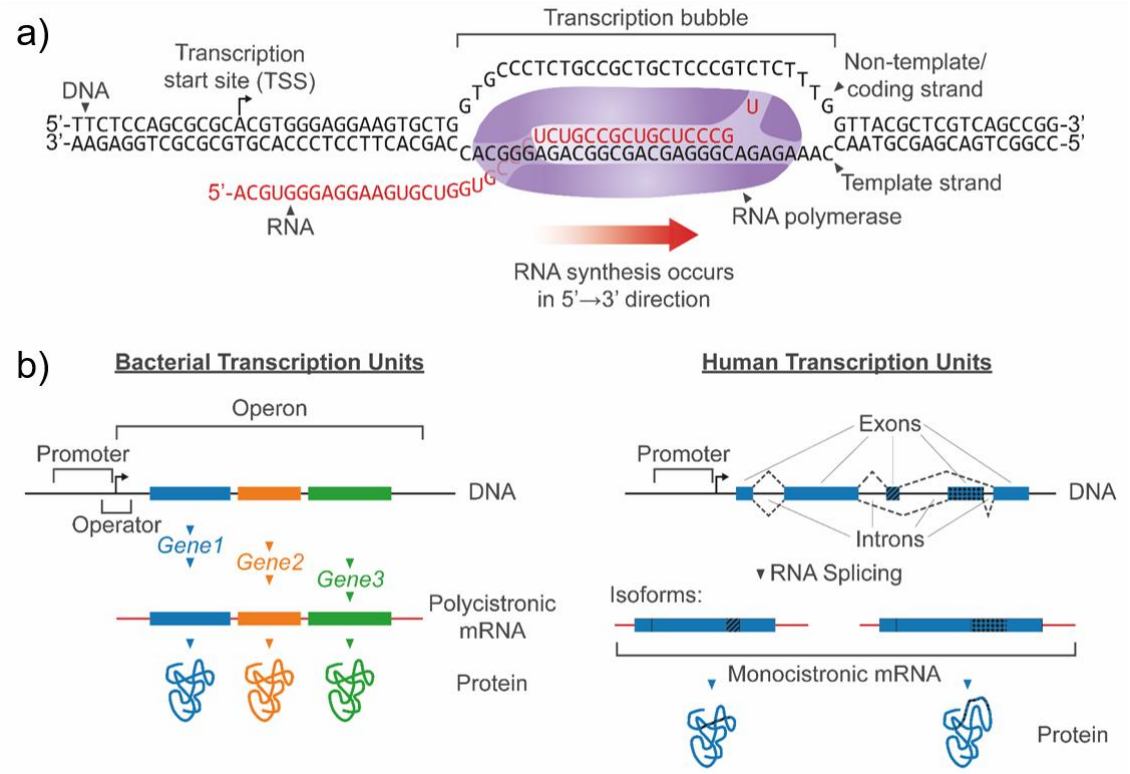


Figure 1.1 | Transcription by DNA-dependent RNA polymerases. **a**) Schematic of transcription by DNA-dependent RNA polymerase (shown in purple). RNA transcription occurs in a 5'→3' direction. **b**) A comparison of transcription unit (TU) organisation between bacteria and humans for protein-coding transcripts. In bacteria, multiple genes can be expressed using the same polycistronic mRNA. Whereas, in humans each protein-coding TU expresses a monocistronic mRNA (containing one protein-coding gene), however, this can generate multiple transcript/protein isoforms by processes which include alternative splicing leading to differential exon inclusion and alternative PAS selection leading to different 3' untranslated regions (UTR).

production of polycistronic RNA that encodes multiple polypeptides (**Fig 1.1b**). This differs in eukaryotes were protein-coding transcripts are monocistronic, which means each transcript only encodes one polypeptide. However, this does not mean all the RNAs produced within higher eukaryotes at any one time (transcriptome) can be ascribed a function, even though, it certainly remains a cell's aim to transcribe functional RNAs whether they be protein-coding for the production of polypeptides or non-coding for other uses such as in splicing and translation (Graur et al., 2013; Palazzo and Koonin, 2020). Additionally, transcription can occur from RNA templates during host infection with certain types of RNA viruses. Consequently, these enzymes are classed as RNA-dependent RNA polymerases. Some RNA viruses also possess the ability to synthesis DNA from RNA during viral replication, but this is a different process known as reverse transcription. The replication of RNA viruses is not discussed further.

The transcription cycle has three phases: initiation, elongation, and termination (Shandilya and Roberts, 2012). Transcription initiation is the assembly of RNA polymerase and cofactors at DNA sequences referred to as promoter regions. First transcription factors bind at promoters which in turn recruit cofactors that trigger chromatin opening and facilitate RNA polymerase complex assembly. This allows transcription bubble formation where the DNA template strand is unhybridized and free to bind complementary ribonucleotides for assembly of the growing RNA within the polymerase active site. After initiation, the RNA polymerase is released into processive elongation where ribonucleotide triphosphates complementary to the template strand are readily incorporated into the growing RNA strand in a 5'→3' direction. This is followed by transcriptional termination, which is the cessation of incorporating ribonucleotides and dissolution of the ternary complex of bound RNA, DNA, and protein. Although these three phases of transcription occur in all domains of life, bacteria (and prokaryotes enlarge) generally have fewer protein complex components and lack some regulatory complexity compared to archaea and eukaryotes i.e. lacking RNA splicing and introns, having shorter length genes and the core RNA polymerase enzyme consisting of 5 protein subunits in *Escherichia coli* (*E. coli)* compared with 12 or more for human isoforms (**Fig 1.2**; Werner, 2007). Despite the bacterial and eukaryote differences, common

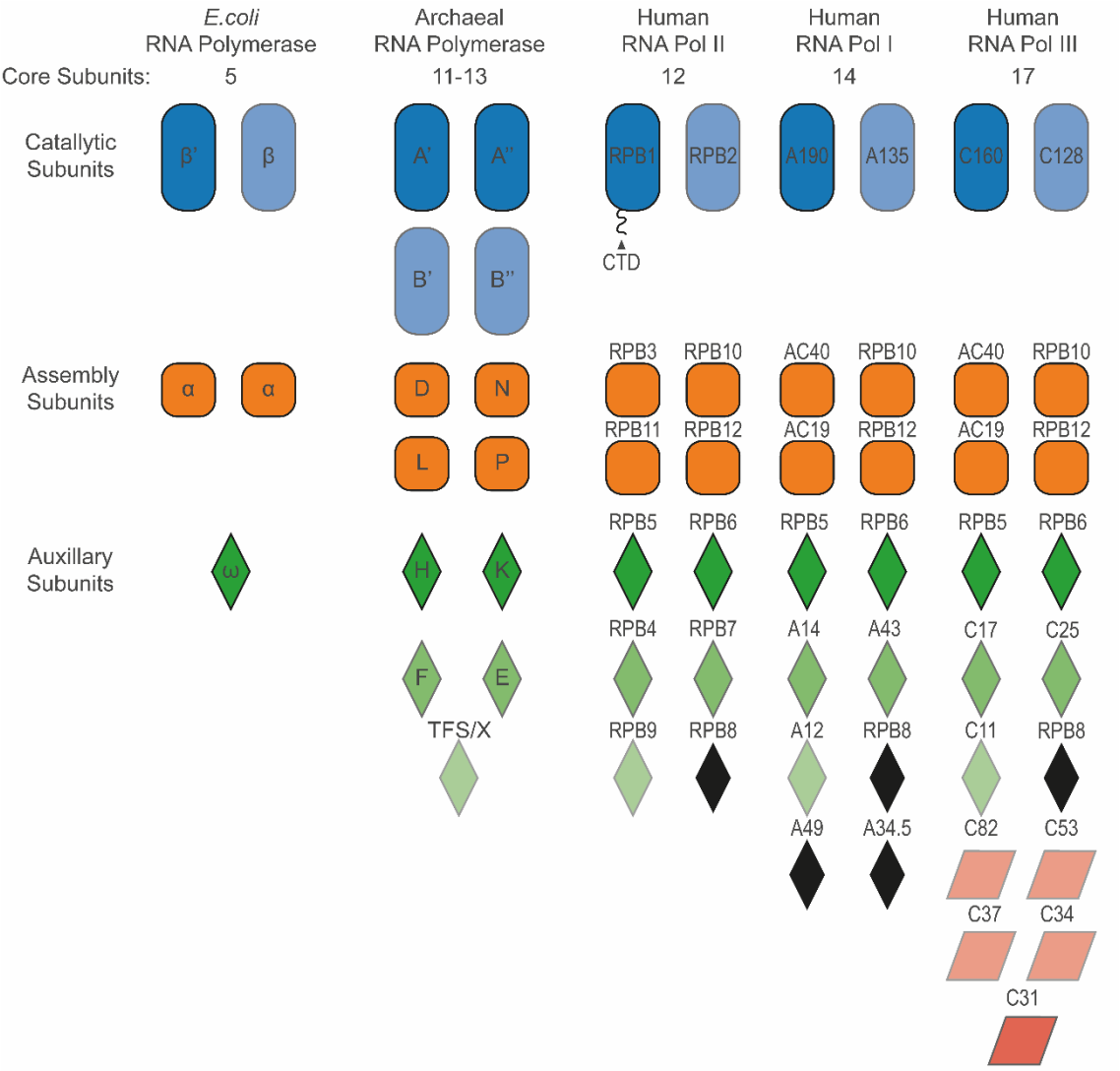


Figure 1.2 | Comparison of RNA polymerase domain architecture across the three domains of life. The subunit comparison of the architectures of the RNA polymerases present in *E. coli*, Archaea and humans. The polymerases are shown in order of increasing subunit complexity (left to right) and coloured to show related subunits. The disordered c-terminal domain (CTD) of the largest Pol II subunit, RPB1, is shown as a tail. [Figure adapted from Werner (2007)].

parallels can be drawn between the mechanisms of initiation, elongation, and termination (Porrua et al., 2016).

1.1.1 Bacterial Transcription Initiation

In bacteria, tight control of gene expression is required to respond to external stimuli and environmental changes. One mechanism for responding to such changes is by modulating transcription initiation of collected groups of related genes with similar promoters. The 5-subunit core (α, α, β, β', ω) of bacterial RNA polymerase (RNAP – not to be confused with the eukaryotic Pol II abbreviation) requires binding of a sigma (σ) factor, forming a holoenzyme, to be recruited to DNA promoters. Therefore σ-factors act as transcription initiation factors. Different σ-factors recognise different promoter sequence motifs. A major σ-factor responsible for transcription of many *E. coli* housekeeping genes and genes involved in exponential growth is σ^{70} (also known as RpoD), which binds the TTGACA and TATAAT consensus sequences -35 bp and -10 bp upstream of the transcription start site (TSS) respectively (Campbell et al., 2002; Murakami and Durst, 2003). Other contacts are made with DNA by σ^{70} at the extended -10 element and discriminator region and by RNAP α-subunits C-terminal domain at the UP element (**Fig 1.3a**; Murakami et al., 2002; Sanderson et al., 2003).

Whilst changes in specific σ-factor availability directly regulates initiation by controlling RNAP holoenzyme formation for target promoters, other methods for regulation of transcription initiation include activator or repressor proteins binding operators. Operator sequences are regions of DNA, usually in proximity to promoters, recognised and bound by transcription regulators. One example is IclR which represses expression of the *aceBAK* operon, which encodes genes used in acetate metabolism, and its own gene *iclR* (Gui et al., 1996). On its own gene, IclR simply uses steric hindrance to compete with RNAP for binding at the promoter region because the operator sequence overlaps the promoter at -29 and +9 bp relative to the TSS (**Fig 1.3b**). This leads to autogenous regulation of IclR repressor expression. However, transcriptional repression of the aceBAK operon not only use steric hindrance but has a dual mechanism involving two operator binding sites located between -125 to -99 and -52 to -19 bp relative to the TSS (Yamamoto and Ishihama, 2003). The promoter-proximal IclR binding site, at -52 to -19 bp, overlaps the promoter and

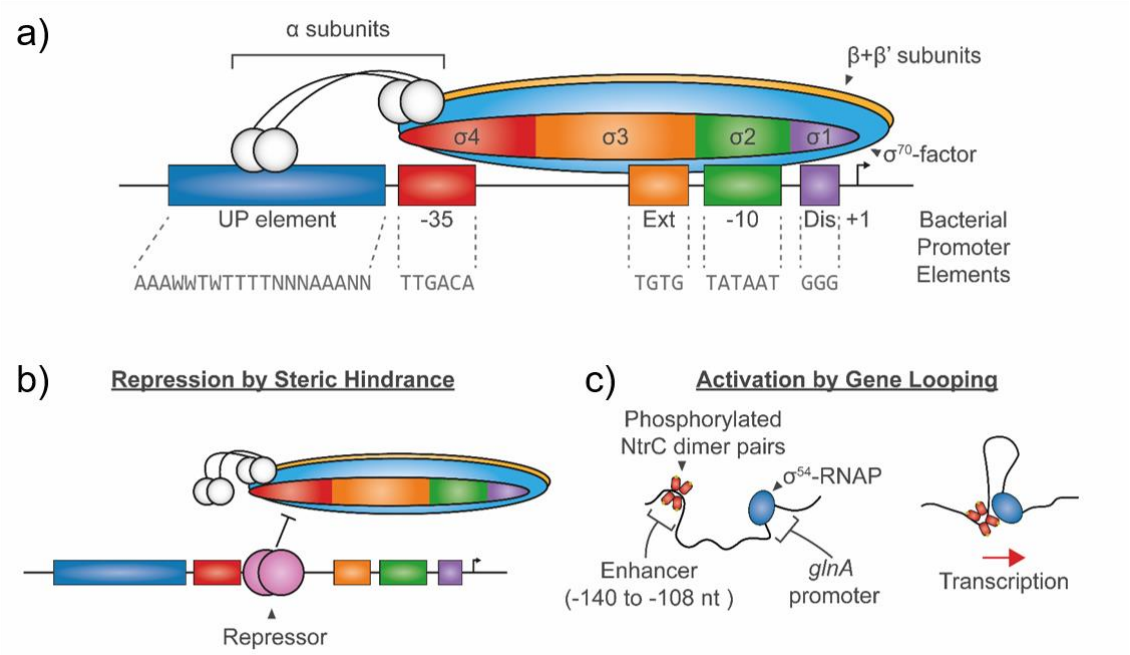


Figure 1.3 | Overview of bacterial RNA polymerase (RNAP) transcriptional initiation. a) Schematic of a bacterial σ^{70} promoter highlighting consensus sequences and the domains of σ^{70}-RNAP important for their recognition. The domains of σ^{70} are labelled (σ1-σ4). [N=A/T/C/G; W=A/T] **b)** Repression of transcription initiation often occurs by steric hinderance with a repressor (shown in pink) binding at or near promoter sequences and occluding RNAP binding. **c)** Regulation of transcriptional initiation can also be mediated by DNA looping. A schematic of DNA looping mediated activation is shown for the *glnA* gene with a pair of phosphorylated NtrC protein dimers binding to an upstream enhancer sequence (between -140 to -108). This leads to gene looping with σ^{54}-RNAP stimulating promoter DNA opening and initiation. Yellow dots represent phosphorylation on NtrC. [Figure parts **a**,**b** are adapted from Browning and Busby (2016)]

thus competes with RNAP for promoter binding in a classical steric hindrance method. Additionally, IclR binding at the distal site (-125 to -99 bp) interacts with RNAP and shifts the α-subunit binding site upstream of the UP element further repressing transcription.

Other mechanisms for regulation of bacterial transcription initiation include DNA looping of sequences upstream of the promoter. One of the best-studied examples of gene activation by DNA looping of an upstream enhancer sequence in *E. coli* is at σ^{54} promoters (Danson et al., 2019). Genes under the expression of σ^{54}-promoters are associated with nitrogen metabolism and RNAP-σ^{54} binding occurs at consensus sequences CTGGNA and TTGCA that are -24 bp and -12 bp upstream of the TSS respectively (**Fig 1.3c**) (Burrows et al., 2003; Doucleff et al., 2007; Sasse-Dwight and Gralla, 1988). Unlike for other σ-factors, RNAP-σ^{54} binding to promoters is incapable of initiating transcription alone. Transcription initiation requires activation by DNA looping with an enhancer-binding activator protein, which is bound approximately -100 to -150 bp upstream of the TSS, in an ATPase-dependent manner (Popham et al., 1989; Buck et al., 1986; Bush and Dixon, 2012). Structures of the RNAP-σ^{54}, DNA and an ATP analogue reveal that σ^{54} occupies the β and β' DNA-binding cleft of RNAP and upon activation on interaction with an upstream enhancer-binding protein leads to a conformational change that opens RNAP DNA cleft for DNA binding, promotes DNA strand separation and formation of a transcription bubble (Bose et al., 2008; Yang et al., 2015; Glyde et al., 2017). In addition to activation, DNA looping can also cause transcriptional repression and one such example of this is at the *gal* operon. The *gal* operon contains genes involved in galactose metabolism and under higher concentrations of D-galactose the sugar binds allosterically to GalR repressor and decreases GalR affinity for operator sites. There are two operator binding sites for the GalR repressor protein, which are situated upstream and downstream of the promoter, and once each operator is bound by a GalR dimer both tetramerise to form a DNA loop that prevents initiation (Semsey et al., 2004; Swint-Kruse and Matthews, 2009). In all, bacteria use a variety of mechanism to regulate transcription initiation and whilst eukaryotes have more components and complexity many similarities can be draw through its mechanisms.

1.1.2 Bacterial Transitions Into and Out of Processive Transcription Elongation

The regulation of transcription initiation (described above) plays an important role in allowing a bacterium to modulate RNAP recruitment to different promoters. As each promoter can only assemble the transcription bubble of a single polymerase at any given moment the question arises as to whether it is the recruitment of the RNAP holoenzyme to promoters or the transition of RNAP from initiation into productive elongation which is the rate-limiting step. In support of the latter, analysis by single-molecule Förster resonance energy transfer (smFRET) reveals RNAP+σ^{70} stalls after synthesis of a 6 nt RNA for approximately 20s before productive transcription elongation ensues, abortive termination and RNA release, or a long-lived RNA bound RNAP complex remains (Duchi et al., 2016; Dulin et al., 2018). Structural studies of the initiating RNAP+σ^{70} suggest the reason for such a stall is due to σ^{70} occupying the RNA exit channel such that only 6 nt can be synthesised until σ^{70} undergoes a conformational change or dissociates to vacate the exit channel (Zhang et al., 2012; Zuo and Steitz, 2015).

Additionally, RNAP elongation speed and pausing can be modulated by one of the only evolutionary conserved general transcription factors present in all domains of life, NusG in bacteria or Spt5 in archaea and eukaryotes (Tomar and Artsimovitch, 2013). NusG consists of two domains, an NGN domain that binds the β' clamp helices near the central cleft of RNAP and the KOW domain that interacts with other transcription factors. Interestingly, NusG plays converse roles in different bacterial species being responsible for promoting transcription elongation and suppressing pausing within *E. coli*, whereas NusG can stimulate sequence-specific pausing of RNAP in *Bacillus subtilis* (Sevostyanova et al., 2011; Turtola and Belogurov, 2016; Yakhnin et al., 2019).

1.1.3 Bacterial Transcription Termination

In bacteria several termination mechanisms exist to cease RNAP transcription including Rho-dependent, intrinsic, Mfd-dependent and RNaseJ1 (*Bacillus subtilis*) termination. The Rho protein is an ATP-dependent hexameric helicase responsible for transcriptional termination of ~20 % mRNA transcription events (Peters et al., 2009). Rho-dependent termination occurs after RNAP transcription past a pyrimidine-rich Rho utilisation (*rut*) RNA sequence is recognised by a Rho hexamer (Bogden et al., 1999). Rho surveys for rut

sequences in an open-ring conformation and on binding changes to a closed-ring conformation in an ATP-dependent process trapping the RNA (Thomsen and Berger, 2009; Thomsen et al., 2016). The Rho helicase then translocates along the RNA in a 5'→3' direction, which utilizes ATP in the process, and upon capture of RNAP transcription termination ensues (Brennan et al., 1987; Schwartz et al., 2007). Interestingly, NusG can promote Rho-dependent termination by stimulating ring closure on weaker *rut* sequences and binding of antitermination factors to NusG can inhibit Rho interactions with NusG and thereby regulate Rho-dependent termination (Lawson et al., 2018). Upon capture of RNAP by Rho the disassembly mechanism is thought to involve Rho invasion into the main channel leading to RNA:DNA hybrid melting within the active centre that is followed by complex dissociation (Epshtein et al., 2010).

In a similar 'chasing' mechanism to Rho, the Mfd DNA translocase is responsible for displacement of stalled RNAP at DNA lesions targeted by transcription couple repair (TCR) (Adebali et al., 2017; Selby and Sancar, 1993). Mfd termination plays an important role in minimising collisions between stalled RNAP and replisomes, which can be a major cause of genome instability (Dutta et al., 2011; Pomerantz and O'Donnell, 2010). The DNA translocase activity is ATP-dependent, traversing DNA slowly (7 bp/s) over short ~200 bp distances, and upon Mfd capture of RNAP can either rescue transcription or induces termination depending on the severity of pause (Le et al., 2018).

Another termination mechanism initially observed in *Bascillus subtilis* involves RNaseJ1 cleaving nascent RNA, degrading RNA with 5'→3' exonuclease activity and promoting termination of stalled RNAP complexes (Šiková et al., 2020). This torpedo mechanism displays strong parallels to termination of protein-coding genes in eukaryotes where RNA is cleaved before being degraded by the torpedo exonucleases Rat1/XRN2 in yeast and humans respectively (see section 1.1.7). Instead, RNaseJ1 is a member of the metallo-β-lactamase family and possesses endoribonuclease activity *in vitro* in addition to exonuclease activity (Even et al., 2005). Thus, RNase J1 possesses a dual endo-/exo-ribonuclease role and is closely related to the human CPSF73 endoribonuclease responsible for generating the exonuclease entry site for torpedo termination at the 3'-ends of transcripts. Recently, related metallo-β-lactamases have been identified in Archaea that promote transcriptional

termination and they also appear to possess dual endo-/exo-ribonuclease activities (Sanders et al., 2020; Yue et al., 2019). Thus, it appears that torpedo termination is universally conserved throughout all domains of life. Mechanistically (at least) RNaseJ1 and Rho-dependent termination share similarities with both using RNA as a guide to catch RNAP and instigate termination.

In a very different mechanism, RNAP can undergo intrinsic (also referred to as rho-independent) termination upon transcription past a termination signal. An intrinsic termination signal consisting of a GC-rich inverted repeat sequence followed by at least four thymidine residues (Brendel et al., 1986). Once transcribed this leads to RNAP pausing over a stretch of rU:dA hybrids and RNA hairpin formation. The rU:dA hybrids have an extremely low thermodynamic stability compared with other hybrids and a stretch of four or more induces RNAP pausing (Martin and Tinoco, 1980; Gusarov and Nudler, 1999; Nudler and Gottesman, 2002). This pause allows hairpin formation of upstream inverted repeat sequences, which invades the RNA exit channel, inducing RNA:DNA hybrid melting and dissociation of the complex through allostery (Epshtein et al., 2007). Intrinsic termination can occur prematurely to regulate the expression of a transcript by a process called transcription attenuation. Within *E. coli*, the *trp* operon is one well studied example of this where concentrations of tryptophan (or $tRNA^{trp}$ specifically) change the co-transcriptional translation efficiency of the early *trp* leader mRNA that in turn influences mRNA stem loop formation and RNAP termination (Yanofsky et al., 1981). When the concentrations of $tRNA^{trp}$ are sufficiently high, translation occurs unimpeded into a region that prevents an early stem loop from forming and means a downstream stem loop forms near a U-rich pause attenuator site (at ~140 nt) inducing RNAP termination (**Fig 1.4**). During low concentrations of $tRNA^{trp}$, translation stalls and allows the early stem loop to form preventing stem loop formation within the attenuator sequence meaning RNAP can continue to transcribe the long *trp* operon. The process of intrinsic termination is conserved within eukaryotes, with RNA polymerase III (Pol III) terminating at stretches of four or more thymidine residues.

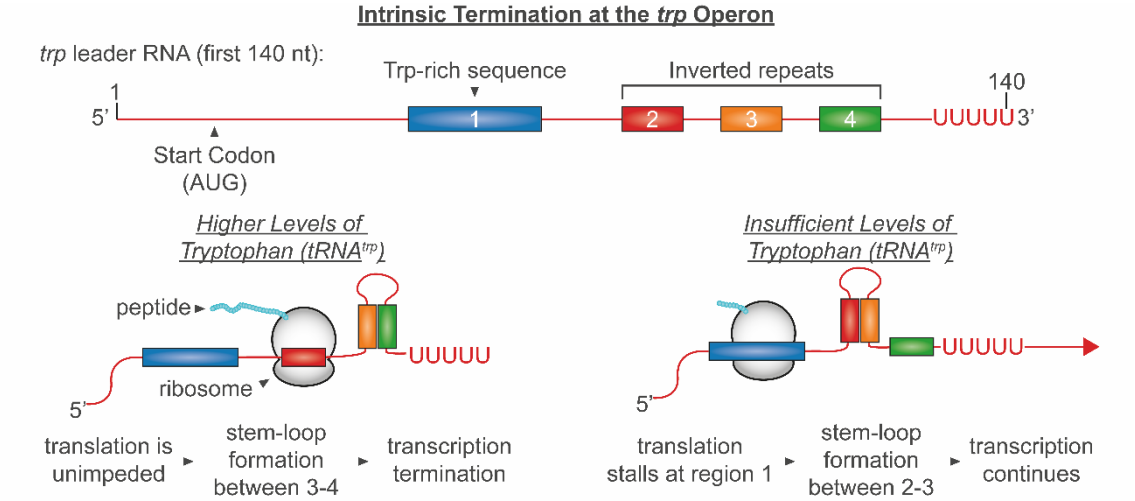


Figure 1.4 | Intrinsic termination and transcriptional control at the *trp* operon by attenuation. Top) Schematic of the first 140 nt of the *trp* operon or the *trp* leader RNA. **Bottom-Left)** Under higher cellular levels of $tRNA^{trp}$, co-transcriptional translation occurs unimpeded through region 1 and into region 2. This causes stem-loop formation between regions 3 and 4 (orange and green) upstream of a run of five uridines. This forms an attenuation signal that terminates RNAP transcription prematurely. **Bottom-Right)** Under insufficient cellular levels of $tRNA^{trp}$, co-transcriptional translation stalls during the Trp-rich region1. This allows stem-loop formation between regions 2 and 3 (red and orange), thus preventing the latter stem loop and transcription continues into the *trp* operon. Ribosomes are shown in white with a peptide (light-blue) extruding.

1.1.4 Eukaryotic Transcription Initiation

In humans there are three RNA polymerase enzymes within the nucleus, Pol I, Pol II and Pol III, and each have promoters containing different conserved DNA sequence elements. As with bacteria, these polymerases cannot recognise or initiate transcription at promoter sequences themselves and require transcription initiation factors to recognise promoter sequences, recruit the polymerase and promote DNA unwinding and template strand loading into the active site. This complex of RNA polymerase and transcription factors bound at a promoter and ready to initiate is called the pre-initiation complex (PIC).

PIC formation for the three polymerases differs slightly but all involve binding of transcription factors upstream of the TSS (Cramer, 2019). For Pol II, the enzyme of focus in this thesis, this involves TFIID, whose largest subunit is TATA box-binding protein (TBP), binding -25 bp upstream of the TSS at the TATA element with the consensus sequence TATAa/tAa/t (Nikolov and Burley, 1997). The binding of TBP with the minor groove in DNA causes bending of the helix (**Fig 1.5**). Then TFIIB binds TFIID through interactions with its c-terminus and recruits a preformed Pol II-TFIIF complex to the promoter (Sainsbury et al., 2013). Next TFIIE and TFIIH bind the complex forming the PIC. One TFIIH subunit is XPB, a DNA helicase that binds DNA downstream of the promoter and hydrolyses ATP to unwind DNA and thread it into the Pol II active site catalysing promoter opening (Kim et al., 2000; Grünberg et al., 2012). XPB helicase activity can be inhibited by the compound triptolide, which is irreversible as it binds covalently (Titov et al., 2011). Inhibition of initiation by triptolide reveals different promoter susceptibility with most showing clearance of Pol II at promoters and others maintaining a promoter Pol II peak by chromatin immunoprecipitation (ChIP)-seq (Chen et al., 2015b). However, unlike triptolide inhibition, depletion of XPB reveals it can be dispensable for initiation and not always required for promoter melting (Alekseev et al., 2017; Dienemann et al., 2019). The transition from PIC into promoter escape is also regulated by the Mediator complex. The Mediator complex comprises of approximately 35 subunits that form modules including ‘head’, ‘middle’ and ‘tail’ (Plaschka et al., 2015; Nozawa et al., 2017a; Tsai et al., 2017). The conserved ‘head’ and ‘middle’ modules make contacts with Pol II, TFIIB and TFIIH of the PIC. Additionally, Mediator can stimulate the kinase activity of a different TFIIH

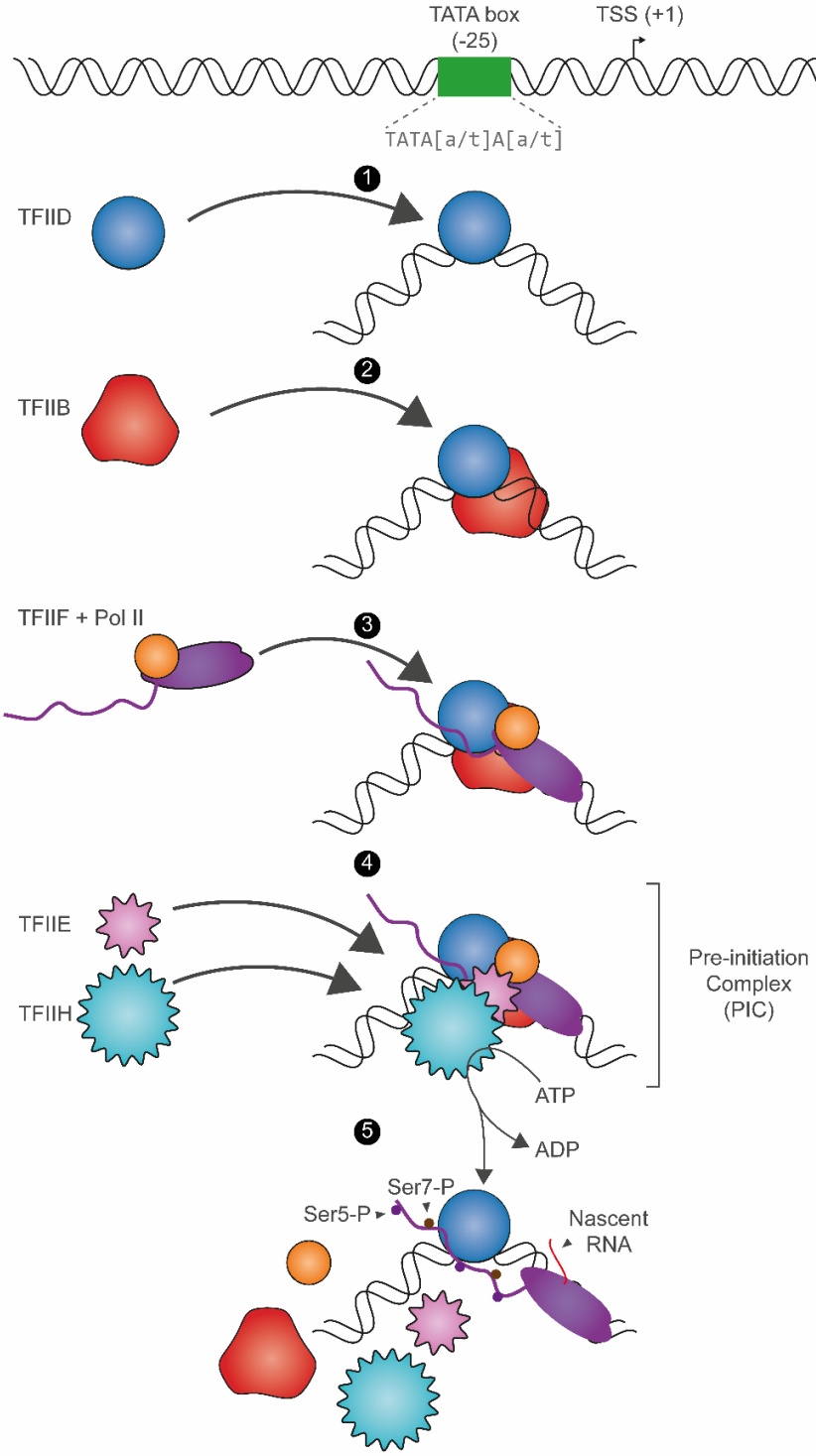


Figure 1.5 | Assembly of the Pol II Pre-initiation Complex (PIC). Schematic of stepwise Pol II PIC assembly showing general class II transcription factors. Initially promoters are recognised by TFIID (contains TATA box-binding protein) inducing a bend in the DNA on minor grove binding (**1**). Then TFIIB binds making contains with the DNA and TFIID (**2**). Next the preformed complex of TFIIF and Pol II core complex binds to the promoter (**3**). Following this TFIIE and then TFIIH also bind to the complex forming the PIC (**4**). Then (XPB subunit) helicase and (CDK7 subunit) kinase activities of TFIIH use ATP hydrolysis to melt the DNA and thread the template strand into Pol II active site and phosphorylate Ser5 and Ser7 of the Pol II CTD (purple tail) allowing promoter escape and transcription factor disassembly (**5**).

subunit, CDK7 (or Kin28 in yeast), which phosphorylates the Pol II C-terminal domain (CTD) of the largest subunit Rpb1 (Kim et al., 1994; Feaver et al., 1994). CDK7 is a dual-specificity CTD kinase that phosphorylates Serine 5 (Ser5), which in turn leads to Mediator dissociation from the PIC promoting promoter escape, and Serine 7 (Ser7), which is important for snRNA processing and Integrator complex recruitment (Akhtar et al., 2009; Egloff et al., 2007; Glover-Cutter et al., 2009; Jeronimo and Robert, 2014; Kim et al., 2009; Wong et al., 2014).

The Pol II CTD, which is part of the largest subunit Rpb1, is a tail-like low complexity extension consisting of repeats of the seven consensus amino acids residues Tyr1-Ser2-Pro3-Thr4-Ser5-Pro6-Ser7 (YSPTSPS) (Eick and Geyer, 2013). The CTD and consensus heptad is conserved in yeast and humans but the former consists of only 26 repeats compared with 52 in humans. Another difference is that the yeast repeats show very little variation and mostly follow the consensus sequence, whereas humans have several deviations from the consensus heptad among the repeats. The CTD heptad repeats are unique to Pol II and not present within Pol I and Pol III homologues. The CTD residues Tyr/Ser/Thr are substrates for kinases and phosphatases that lead to distinct patterns or constellations of phosphorylation which change throughout the transcription cycle and how these patterns dictate the association of factors is termed 'the CTD code' (Komarnitsky et al., 2000; Buratowski et al., 2003; Corden et al., 2016). As mentioned above, Ser5-P and Ser7-P are found enriched at promoters and Ser7-P within the gene bodies of snRNAs also. Another residue, Ser2 is phosphorylated by CDK9, a subunit of the positive transcription elongation factor b (P-TEFb) complex and CDK12 (which also targets Ser5) (Ahn et al., 2004; Boehm et al., 2003; Bowman et al., 2013; Ni et al., 2004; Shim et al., 2002; Tellier et al., 2020). The distribution of Ser2-P increases after the promoter and increases as a function of time rather than Pol II elongation distance (Komarnitsky et al., 2000; Joo et al., 2019). CDK9 has also be shown to phosphorylate Ser5 *in vitro* (Bösken et al., 2014, Czudnochowski et al., 2012). Furthermore, Tyr1 is found phosphorylated at promoters (preferentially in the antisense/ PROMPT direction) and enhancers within metazoans (Descostes et al., 2014; Hsin et al., 2014). However, this differs to yeast where Tyr1-P increases after the TSS and found across gene bodies (Mayer et al., 2012). The phosphorylation of Tyr1 is catalysed by c-Abl1

and c-Abl2 kinases and *in vitro* yeast CTD treatment by c-Abl1/2 enriches Ser2 phosphorylation by P-TEFb (Baskaran et al., 1997; Mayfield et al., 2019). Finally, Thr4-P is found localised at the 3'-ends of protein-coding and histone transcripts and two kinases have been implicated in its phosphorylation, Plk3 and CDK9 (Hintermair et al., 2012; Hsin et al., 2011). Mass spectrometry analysis of neighbouring residues within individual heptads revealed most repeats are only phosphorylated at a single position (Ser2 and Ser5 being the most common), but double-phosphorylated repeats were detected albeit at ~30 fold lower abundances compared to mono-phosphorylated repeats (Schüller et al., 2016). Additionally, given the multiple CTD residues implicated with CDK9 kinase activity (Ser2, Ser5 and Thr4) mass spec was performed on cells treated with CDK9 inhibitor flavopiridol. Interestingly, this led to a reduction of Ser2-P levels in heptads throughout the CTD with little change in Ser5-P and Thr4-P abundances. This does not necessarily rule-out CDK9 involvement at these residues in some capacity, just that their phosphorylated forms are not highly dependent on CDK9 activity alone. Whilst much of the focus has been on CTD phosphorylation (partly due to the availability of specific antibodies) other CTD modifications including Ser/Thr glycosylation, Lys acetylation/methylation (of non-consensus repeat residues) and Pro isomerisation also occur but their implications for transcription (and CTD phosphorylation of neighbouring residues) are less well established (Ali et al., 2019; Dias et al., 2015; Eick and Geyer, 2013; Kelly et al., 1993; Voss et al., 2015).

1.1.5 Pol II Transcription Elongation

After initial transcription, general transcription factor disassociation and promoter escape, Pol II often pauses ~50 bp downstream of the TSS (Nechaev et al., 2010; Nojima et al., 2015; Core and Adelman, 2019). Before this Pol II promoter-proximal pausing, elongation factors associate with Pol II as the nascent RNA emerges from the polymerase. One of these is dichloro-1-β-D-ribofuranosylbenzimidazole (DRB)-sensitive inducing factor (DSIF) formed of a heterodimer of Spt5 and Spt4 which bind Pol II near the RNA exit channel (Bernecky et al., 2017; Vos et al., 2018a). Upon Spt5 binding, it stimulates co-transcriptional capping of the emerging nascent RNA and its deletion in yeast severally reduces global transcription (Wen and Shatkin, 1999; Pei and Shuman, 2002; Shetty et al., 2017). Recruitment of Spt5 to Pol II may be

regulated by MYC which can promote Spt5 binding in a CDK7-dependent manner (Baluapuri et al., 2019). Another factor, negative elongation factor (NELF) binds to the Pol II-Spt5 complex and both DSIF and NELF promote promoter-proximal pausing and restraining Pol II elongation (Lee et al., 2008; Vos et al., 2018a; Wu et al., 2003). These promoter-proximally paused complexes are released by recruitment of P-TEFb whose subunit CDK9 phosphorylates the Pol II CTD, Spt5 and NELF (Fujinaga et al., 2004; Marshall et al., 1996; Yamada et al., 2006). This induces NELF dissociation and binding of Spt6 and the PAF complex forming the activated elongation complex (Vos et al., 2018b). The phosphorylation of Spt5 in humans and yeast is regulated by competing kinase, CDK9 and phosphatase, PP1 activities that form a regulatory circuit that controls the elongation rate of transcribing Pol II (Booth et al., 2018; Parua et al., 2018; Parua et al., 2020).

1.1.6 Pol II Termination Occurs at the Beginning, Middle and End of Genes

A growing view of Pol II transcriptional termination is that it is more than simply a means to an end of the transcription cycle because the process it is not solely restricted to the end of TUs. Indeed, premature termination can occur at almost any position throughout a TU and may have important roles in regulating transcription of full-length transcripts (**Fig 1.6**) (Kamieniarz-Gudla and Proudfoot, 2019; Mendoza-Figueroa et al., 2020). The most studied of the Pol II termination mechanisms is at the 3'-ends of PAS-dependent transcripts. However, other mechanism by the Integrator complex (at the 3'-end of snRNAs and prematurely at a subset of protein-coding transcripts) and processing of early PAS (which is called premature cleavage and polyadenylation or PCPA) are also discussed. It is possible premature termination may quell a significant proportion of Pol II transcription on some transcripts.

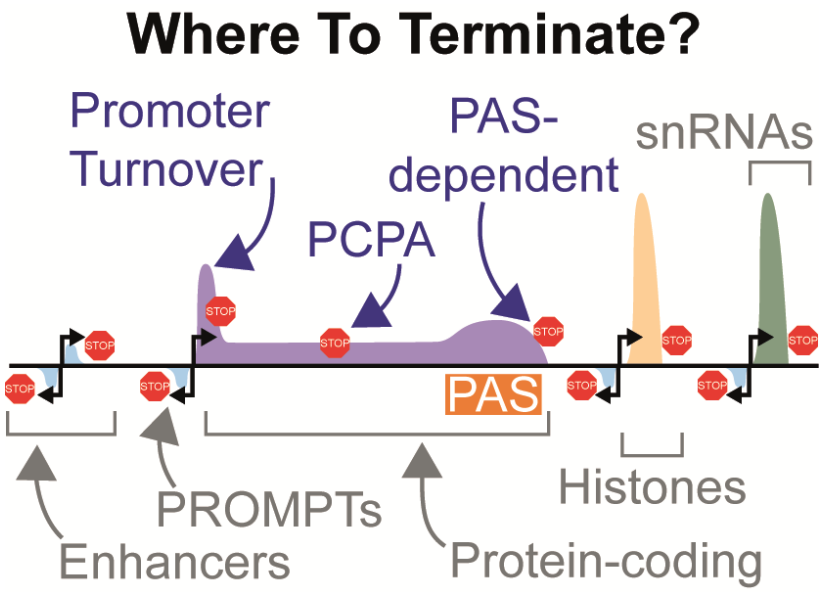


Figure 1.6 | Pol II termination can occur throughout the transcription cycle. Overview of different Pol II transcript classes with positions of transcription termination marked with stop signs. Polymerase occupancies are shown as coloured areas for different transcription classes above and below the lines for sense and antisense transcription, respectively. PAS, polyadenylation signal; PCPA, premature cleavage and polyadenylation; PROMPTs, promoter upstream transcripts; snRNAs, small nuclear RNAs.

1.1.7 Pol II PAS-dependent Termination

Most protein-coding genes contain a PAS at the 3'-end that serves to define the end of the mature transcript and instigate Pol II termination. A PAS is an RNA sequence that includes the canonical hexamer AAUAAA or a close variant flanked by an upstream U-rich and downstream U/GU-rich sequences (Proudfoot, 2011). These sequences are recognised by the cleavage and polyadenylation (CPA) complex, which is a multi-subunit complex formed of the following subcomplexes of cleavage and polyadenylation specificity factor (CPSF), cleavage stimulatory factor (CstF) and cleavage factors I and IIm (Kumar et al., 2019). Upon assembly of the CPA complex, pre-mRNA cleavage occurs between the AAUAAA hexamer and downstream elements at a cleavage site that lacks a known sequence motif but most commonly occurs between the dinucleotides UA or CA in humans (Li and Du, 2013). PAS-recognition by the CPA complex is mediated through components of the CPSF subcomplex, CPSF30 and WDR33, which bind the AAUAAA RNA hexamer (Clerici et al., 2017; Sun et al., 2018; Schönemann et al., 2014; Chan et al., 2014). Whereas RNA cleavage is fulfilled by the CPSF73 endoribonuclease (Mandel et al., 2006). Mutations of a PAS abolish Pol II termination and cause failure in 3'-end cleavage and polyadenylation, thus highlighting both processes are coupled to correct PAS-processing (Logan et al., 1987). This can be the molecular basis for diseases such as α-thalassaemia, where PAS mutation in one or more of the α-globin gene copies reduces the amount of α-globin produced (Whitelaw and Proudfoot, 1986; Proudfoot, 2011). This finding lead to the proposal of two different models explaining PAS-dependent termination, allosteric/anti-terminator and torpedo models (Logan et al., 1987; Connelly and Manley, 1988; Proudfoot, 1989).

The allosteric/anti-terminator model posits that transcription across a PAS causes a conformational change within and/or disassembly of factors from Pol II that promotes its termination (**Fig 1.7**). One key difference compared to torpedo termination is that there is no obligation for RNA cleavage to occur co-transcriptionally before termination, whereas it is a required prerequisite for the torpedo model. Antitermination factors are described as being able to suppress Pol II termination until transcription of a particular PAS sequence. Recently two antitermination factors, SCAF4 and SCAF8, were described that bind Pol II mutually exclusively. Gene knock-out (KO) of both factors leads to premature

CPA (PCPA) at previously silenced PAS sequences within the gene body (GB) (Gregersen et al., 2019). Premature PAS usage and PCPA can be exacerbated under a number of conditions including depletion of U1 small nuclear RNA (U1 snRNA) by morpholino, cyclin-dependent kinase (CDK)12 or nuclear poly(A) binding protein (PABPN1) (Kaida et al., 2019; Dubbury et al., 2018; Jenal et al., 2012). In addition, the levels of the CPA factor PCF11 also influences early PAS usage (Kamieniarz-Gdula et al., 2019). The allosteric model also suggests conformational changes within Pol II can aid termination. Pol II transcription of PAS containing templates *in vitro*, which used purified extracts, was able to undergo some termination in the absence of transcript cleavage (Zhang et a., 2015). Moreover, observations of transcription complexes by electron microscopy, known as Miller spreads, rarely capture co-transcriptional cleavage at PAS (Osheim et al., 2002). However, as all observations are carried out under unperturbed conditions rapid 5'→3' exonuclease degradation of the downstream degradation product could explain this. Likewise, allosteric termination before the PAS seems to be a rare event because PAS mutation causes accumulation of transcripts at the locus (Custodio et al., 1999). Notably, eukaryotic Pol III, Nrd1-Nab3-Sen1 (NRD) complex with Pol II in yeast, and intrinsic termination of bacterial RNAP (see above) establish a conserved president for RNA polymerase termination via cleavage independent mechanisms in all domains of life (Zenkin, 2014; Nielsen et al., 2013; Porrua and Libri, 2015).

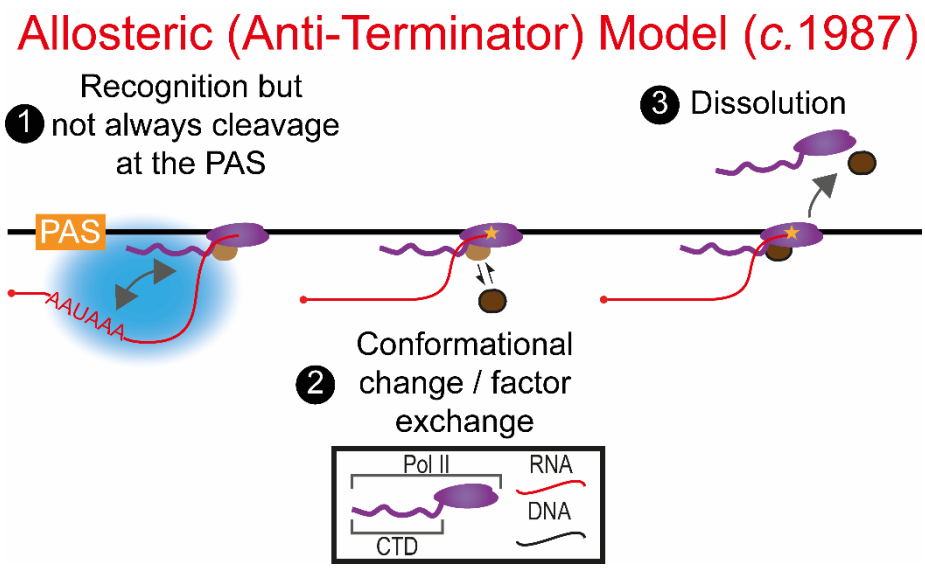


Figure 1.7 | Allosteric/Anti-Terminator Model for Transcriptional Termination. Pol II core enzyme is shown (in purple) with the disordered RPB1 CTD drawn as a tail. DNA and RNA are drawn in black and red respectively. The recognition of the PAS and communication to the transcribing polymerase is shown by a down headed arrow (point 1). The recruitment or dissociation of factors after the PAS are shown by brown ovals with the conformational change within the Pol II active centre denoted as a yellow asterisk (point 2). This then leads to dissolution of the ternary complex (of DNA, RNA and protein) and ultimately transcriptional termination (point 3).

Alternatively, the torpedo model proposes a 5'→3' exonuclease degrades the downstream RNA product of co-transcriptional cleavage and facilitates termination on reaching Pol II (**Fig 1.8**). The generation of an XRN2 substrate (5'-phosphorylated RNA end) can occur via cleavage of a PAS, by another co-transcriptional cleavage (CoTC) processes, or through 5'-end decapping (Brannan et al., 2012; Davidson et al., 2012; Dhir et al., 2015; West et al., 2008). Observations that CPA factors are needed for Pol II termination at 3'-ends of genes supports the torpedo model (Birse et al., 1998; Nojima et al., 2015). The nuclear 5'→3' exonuclease was identified as XRN2 in humans and Rat1 in yeast with read-through transcription occurring on their depletion (Kim et al., 2004; West et al., 2004). Subsequent experiments further widened the requirement of Rat1 in termination of most yeast protein-coding genes, whereas XRN2 RNA interference (RNAi) had little effects on protein-coding genes beyond those from plasmid reporters (Baejen et al., 2017; Brannan et al., 2012; Nojima et al., 2015). Likewise, little effect was observed by Pol II ChIP upon overexpression of a catalytically inactive XRN2 mutant (XRN2-MT) alone (Fong et al., 2015). This XRN2-MT contains a single amino acid substitution (D235A) within its active site at a position that when mutated in the homologous exonuclease Xrn1 abolishes nuclease activity while preserving RNA binding through a conserved PO_4 binding pocket (Jinek et al., 2011). Therefore, the prediction is overexpressing XRN2-MT would create a dominant-negative scenario where the mutant competes with endogenous XRN2 by sequestering 5' RNA substrates. Interestingly, when cells were jointly treated with RNAi of endogenous XRN2 and XRN2-MT overexpression, Pol II ChIP reveals a widespread accumulation of polymerase after the PAS genes (Fong et al., 2015). The pursuit of Pol II by XRN2 predicts that increasing or decreasing the speed of either will affect the position of termination. Indeed, changes in Pol II elongation speed impact termination with faster Pol II mutants evading termination for longer than slower versions. In yeast, Rat1 mutation does not prevent degradation of the co-transcriptional RNA products of PAS cleavage even though Pol II termination is delayed because of the redundant activity from the related Xrn1 5'→3' exonuclease (Luo et al., 2006). One of the questions examined later within this thesis is whether redundant exonuclease activity within humans may explain the disparity between the results by Pol II ChIP (Fong et al., 2015) and Pol II-associated RNA (Nojima et al., 2015) as the latter

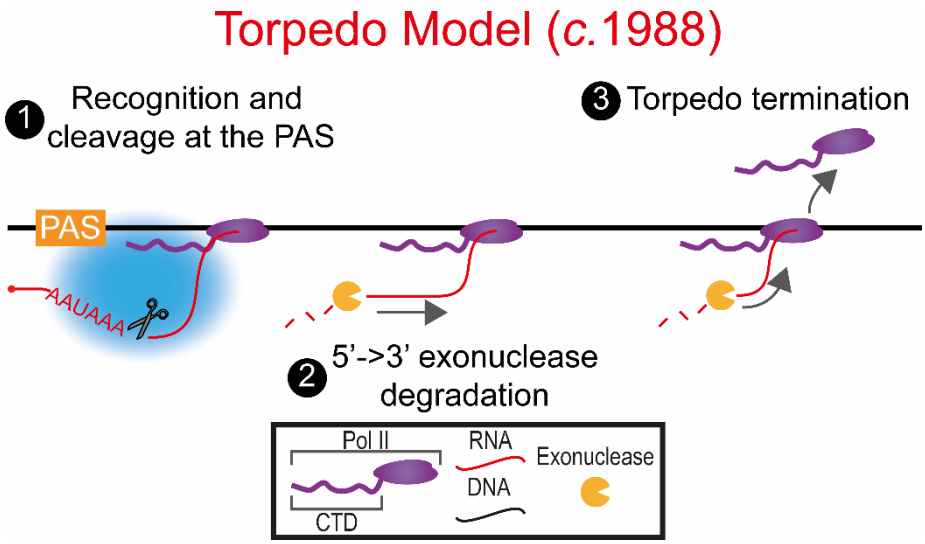


Figure 1.8 | Torpedo Model for Transcriptional Termination. The torpedo model posits that after recognition of a PAS, RNA cleavage occurs (point 1). The downstream product of RNA cleavage then provides a 5'-end entry site for a 5'→3' exonuclease that degrades the nascent transcript as a "molecular torpedo" chasing towards Pol II (point 2). Once the exonuclease catches Pol II, it induces complex dissolution and transcriptional termination (point 3).

method requires an RNA substrate to capture Pol II position.

As PAS-dependent termination of Pol II transcription is important to ensure correct processing of a complete and cogent mRNA, timely Pol II dissolution to prevent transcriptional interference and prevent disorganisation of 3D genome structure which can impact genome stability, termination might not solely depend on one mechanism (Proudfoot, 2016; Heinz et al., 2018; Nozawa et al., 2017b). In budding yeast, a failsafe termination pathway occurs after PAS cleavage in the absence of Rat1 where Nrd1 of the NRD complex facilitates Pol II termination (Rondón et al., 2009). A similar XRN2-independent termination mechanism may occur in humans and could explain why Pol II accumulation does not propagate further downstream upon dominant-negative XRN2-MT express and Pol II ChIP (investigated in chapter 4). Such a mechanism may mirror the termination mechanism of promoter upstream transcripts (PROMPTs) that accumulate upon depletion of the 3'→5' exosome complex suggesting a lack of XRN2 involvement (Preker et al., 2008).

1.1.8 Pol II Termination on Other Transcript Classes

Replication-dependent histone transcripts differ from the broader class of protein-coding transcripts as they do not employ a PAS. However, histone 3'-ends still undergo RNA cleavage by the same endoribonuclease, CPSF73, which is present in both the histone cleavage complex (HCC) and the CPA complex in addition to CPSF100 and Symplekin (Sun et al., 2020; Marzluff and Koreski, 2017). The HCC is recruited to histone pre-mRNA by U7 snRNA and manipulation of the histone processing signal sequence causes a transcription termination defect in an analogous result as for PAS-containing transcripts (Chodchoy et al., 1991). Under dominant-negative XRN2-MT conditions, Pol II termination is mildly affected at histone transcripts indicating a similar torpedo mechanism (Fong et al., 2015). However, the dominant-negative approach of overexpressing XRN2-MT could also inhibit other 5'→3' exonucleases from accessing the downstream 5'-end of histone cleavage. The endonuclease, CPSF73, is proposed as both an endonuclease and exonuclease that has dual roles for histone pre-mRNA processing (Yang et al., 2009; Yang et al., 2020). CPSF73 is a member of the metallo-β-lactamase family of enzymes and a homologous protein in bacteria, RNaseJ, has dual endo- and exonuclease activities (Richards and Belasco, 2011). Interestingly, degradation of the

downstream histone cleavage product *in vitro* can occur from both a 5'-hydroxyl and 5'-phosphate with activity unaffected by divalent cation chelators, which is in agreement with CPSF73 structures showing a high affinity for a coordinated Zn^{2+} metal ion (Yang et al., 2009; Kolev et al., 2008; Mandel et al., 2006). Whereas, the exonuclease XRN2 only degrades 5'-phosphate RNAs and contains Mg^{2+} or Mn^{2+} cations within its active centre (Stevens and Maupin, 1987; Xiang et al., 2009). Moreover, evolutionary president exists for a dual functioning enzyme with archaeal RNA polymerase terminated by FttA, a CPSF73 homologue, that possesses both endo- and exonuclease activities (Sanders et al., 2020). Recently, another metallo-β-lactamase endonuclease in humans, MBLAC1, has been shown to process histone pre-mRNA *in vitro* and *in vivo* creating the possibility that different histone transcripts and isoforms may be processed by multiple pathways (Pettinati et al.,2018).

Another transcript class is uridine-rich snRNAs whose 3'-ends are processed by the Integrator complex (containing core subunits IntS1-14) with the IntS9 and IntS11 subunits sharing sequence homology with CPSF100 and CPSF73 respectively and IntS11 possessing endonuclease activity (Baillat et al., 2005; Dominski et al., 2005). snRNA transcripts contain a 3' box element which is the site of Integrator processing and depletion of Integrator subunits causes Pol II termination defects (O'Reilly et al., 2014; Egloff et al., 2010). These Integrator-dependent Pol II termination functions are linked to IntS11 catalytic activity (Tatomer et al., 2019; Elrod et al., 2019). The XRN2 effects at snRNAs are extremely minimal which could mean RNA cleavage can promote transcription termination independent of a 5'→3' activity (Fong et al., 2015). Another possibility is that snRNAs termination occurs allosterically in which case the RNA products of this would be predicted to generate an exposed 3'-end that would be degraded by 3'→5' exonuclease activity of the exosome. In support of such possibility, a protein-RNA binding site study using cross-linking and immunoprecipitation (CLIP) with an expressed catalytic mutant subunit of the exosome, DIS3, found some snRNA read-through products are bound by the exosome (Szczepińska et al., 2015). It remains unclear if intrinsic termination at snRNAs operates as an alternative or together alongside Integrator cleavage but observations on U2 snRNA reveal transcription carries on beyond the 3' box element so the latter is a distinct possibility at least for some transcripts (Cuello et al., 1999). Other factors connected with snRNA

termination include NELF and ARS2 (Yamamoto et al., 2014; Hallais et al., 2013). Additionally, the CPA components PCF11 and SSU72 are also implicated with PCF11 and are able to terminate Pol II in the absence exonuclease activity in fruit flies (O'Reilly et al., 2014; Zhang and Gilmour, 2006).

1.1.9 Pol II Premature Termination

After loading of initiating Pol II at human promoters, complexes transition from the PIC into transcription but pause soon after (Core and Adelman, 2019). If signalled these stalled Pol II can undergo pause-release by dissociation of NELF, PAF1 binding and phosphorylation several components by CDK9 (a subunit of P-TEFb), whose targets include Pol II CTD, Spt6 and Spt5 (Vos et al., 2018b; Booth et al., 2018; Parua et al., 2020). However, recent studies have highlighted an alternative fate at promoter-proximal locations via rapid Pol II turnover (Krebs et al., 2017; Steurer et al., 2018; Erickson et al., 2018; Nilson et al., 2017). Whilst these termination mechanisms have not yet been fully established, the Integrator complex is thought to play a role in attenuating some of these transcripts because its depletion leads to increased transcription into the gene bodies of a large subset of protein-coding transcripts (Tatomer et al., 2019; Elrod et al., 2019; Stadelmayer et al., 2014; Gardini et al., 2014). The endonuclease activity of IntS11 is required for this activity with overexpression of a catalytic inactive point mutant (E203Q) unable to restore transcription attenuation at promoters (Tatomer et al., 2019; Elrod et al., 2019). It remains unlikely but overexpression of a related metallo-β-lactamase point mutant, CPSF73 H73A, does not fully reconstitute complex formation to endogenous levels so it should be considered a possibility for IntS11 also (Kolev et al., 2008). The liberated 5' RNA of IntS11 endonuclease cleavage may be an entry site for a 5'→3' exonuclease but XRN2 involvement is yet to be demonstrated, akin to snRNAs (Davidson et al., 2019; Elrod et al., 2019). A recent study has revealed Integrator recruits protein phosphatase 2A (PP2a) to promoter-proximal Integrator target transcripts through interactions by the IntS8 subunit and that this association is required for transcriptional attenuation at these sites (Huang et al., 2020). Interestingly, the PP2a phosphatase targets include Pol II CTD and Spt5 that are known substrates of the CDK9 kinase whose activity promotes promoter-release. This suggests a kinase-phosphatase switch could

play an important role in regulating Pol II elongation speed and thus the fate of early paused transcripts through triggering termination or elongation release pathways. Integrator is also involved in enhancer RNA, PROMPT and lncRNA transcription with Integrator depletion leading increased GB transcription (Elrod et al., 2019; Lai et al., 2015; Nojima et al., 2018a). Recruitment of Integrator to these transcripts depends on Spt6 and depletion of Spt6, likewise, increases transcription of these transcripts (Nojima et al., 2018a).

Why some Pol II short transcripts are targets for Integrator and others not can depend on promoter identity with Integrators role at locations close to promoters reinforced by observations that repositioning the Integrator 3' box to downstream locations on snRNAs decreases cleavage efficiency (Hernandez and Weiner, 1986; Ramamurthy et al., 1996; Tatomer et al., 2019). A possible reason that XRN2 has not yet be implicated in degrading Integrator RNA 3' cleavage products is that it is excluded (or not recruited) to these promoter-proximal locations. In agreement with this, phosphorylation sites on XRN2 are known substrates for CDK9 and one, Thr439-P, is thought to promote XRN2 activity (Sanso et al., 2016). Likewise, in *C. elegans* Pol II promoters determine a transcripts susceptibility to XRN2 termination, which may indicate an evolutionary precedent for the differing XRN2 sensitivities to RNA endonuclease cleavage events by Integrator and CPA complexes in humans (Miki et al., 2017). After promoter-proximal pausing, Pol II must also overcome a strong barrier at the first (+1) nucleosome with its relative positioning effecting transcription elongation and pausing (Jimeno-González et al., 2015; Mayer et al., 2015; Weber et al., 2014). The +1 nucleosome pause coincides with site of premature termination by PCPA at premature PAS and these products are substrates of the exosome (Chiu et al., 2018). In the sense direction, U1 small nuclear ribonucleoprotein (snRNP) suppresses PCPA sites but has little effect on PASs within upstream PROMPTs (Almada et al., 2013; Ntini et al., 2013). This is likely because U1 snRNP, which binds 5' splice sites, is known to suppress processing of "nearby" premature PAS signals in a mechanism called telescripting (Ashe et al., 2000; Kaida et al., 2010).

Another factor implicated in premature termination at promoters and enhancers is WDR82, which associates in complexes with SET1A/B and PNUTs/PP1 recruiting them to initiating Pol II (Austenaa et al., 2015). Depletion

of any three of these factors (WDR82, SET1A/B or PNUTs) leads to a transcriptional termination defect. SET1A/B is a H3K4 trimethyltransferase and its recruitment to early Pol II suggest the involvement of histone methylation in promoting early termination. In yeast, deletion of the homologue, SET1, worsens the termination defects of NRD complex substrates (Terzi et al., 2011). Additionally, a WDR82 and protein phosphatase 1 (PP1) containing complex is guided by the PP1 regulatory subunit PP1 nuclear targeting subunit (PNUTs) (Lee et al., 2010; Nuland et al., 2013). Also, PP1 and PNUTs are found to copurify with the CPA complexes with other CPA complex components not restricted to 3'-ends but also found at promoters (Shi et al., 2009; Nojima et al., 2015). Likewise, homologues of PP1, Glc7 in budding yeast and Dis2 in fission yeast, have also been shown to be important for PROMPT transcription termination (Kecman et al., 2018; Nedea et al., 2008; Parua et al., 2018; Schreieck et al., 2014). Substrates of PP1 phosphatase activity are known to include Spt5 whose dephosphorylation induces slowing of Pol II elongation (Cortazar et al., 2019; Kecman et al., 2018; Parua et al., 2018; Parua et al., 2020).

1.2 Investigating Termination of Transcription by Pol II Utilising Rapid Depletion Cell Lines

As stated in section 1.1, transcriptional termination is a fundamental cellular process that has conserved components and mechanistic similarities throughout the three domains of life. Where termination occurs within a TU and its efficiency has profound implications for gene expression, disease and responding to cellular stimuli. For example, a failure in timely termination is a common cellular consequence of viral and osmotic stress (Erickson et al., 2018; Rutkowski et al., 2015; Vilborg et al., 2015). Furthermore, variations of termination efficiency by changes to PAS sequences can be the molecular basis of diseases, such as in thalassaemia and hereditary thrombophilia. Therefore, transcriptional termination warrants study to develop our understanding of the fundamental process and to expose new perspectives that might have translational potential. Here the focus is on PAS-dependent transcriptional termination of Pol II within humans which occurs at the 3'-ends of most protein-coding transcripts. There have been two competing models proposed, the allosteric/anti-terminator model and the torpedo model (see

section 1.1.7). Evidence in support of each has continued to grow over the last three decades and led to an ongoing debate (Libri, 2015).

At the outset of this work, salient questions remained over the involvement of an exonuclease at the end of PAS-dependent transcripts which is a requirement of the torpedo model. The nuclear exonuclease proposed, Rat1 in yeast and XRN2 in humans, were identified using single gene reporter assays (Kim et al., 2004; West et al., 2004). The susceptibility of the downstream product of PAS cleavage to exonuclease degradation includes most protein-coding genes in yeast (Kim et al., 2004; Baejen et al., 2017). However, generalising the effects of XRN2 degradation of the downstream product of PAS cleavage has been less clear in humans. The knock-down (KD) of XRN2 does not lead to an accumulation of read-through RNAs at the 3'-ends (Brannan et al., 2012; Nojima et al., 2015). Similarly, the overexpression of catalytically inactive XRN2-MT alone has little effect at 3'-ends (Fong et al., 2015). Interestingly, joint overexpression of inactive XRN2-MT and KD of endogenous XRN2 does yield accumulation of Pol II at the 3' -ends of protein-coding transcripts (Fong et al., 2015). The question then arrives is this disparity due to sufficiently remaining endogenous XRN2 present after RNAi treatment or XRN2-MT overexpression alone that masks the expected accumulation of RNA and Pol II as envisaged by the torpedo model. This is the focus of chapter 3.

Another possible explanation is that a second exonuclease may have some redundant activity for the downstream product of co-transcriptional PAS cleavage. This is not unexpected because in yeast mutation of Rat1 and overexpression of NLS tagged XRN1 suppressed accumulation of downstream RNA but did not rescue Pol II termination (Luo et al., 2006). The possibility of redundant exonucleases is explored within chapter 4.

Thirdly a key point of contention is whether Pol II termination can occur independently of co-transcriptional RNA cleavage and if so at what frequency does it occur downstream of a PAS because such a process is only compatible with the allosteric model. Some limited termination has been observed in the absence of RNA cleavage on DNA templates *in vitro* using purified nuclear extracts (Zhang et al., 2015). Conversely, KD treatment of the CPA complex components in cells including the endonuclease CPSF73 or other components of the CSTF leads to read-through transcription downstream of PAS-dependent

transcripts and this accumulation is specific for Pol II-associated RNA where the CTD is phosphorylated at the Ser2 position (Nojima et al., 2015). However, this termination defect does eventually return to background levels downstream of these transcripts, which may indicate a secondary cleavage-independent termination pathway. Such a failsafe termination pathway occurs in yeast with the NRD complex but no equivalent homologues for Nrd1 and Nab3 component have been identified in humans (Rondón et al., 2009). Alternatively, incomplete KD of CPA components could mean termination occurring at more distal sites occurs because of delayed cleavage rather than under conditions were cleavage has been abolished. Investigations with the CPA complex endonuclease CPSF73 are detailed in chapter 5.

Chapter 2

2. Materials and Methods

2.1 Antibodies

Table 2.1 | Antibodies.

Antibody [clone]	Host species	Clonality	Manufacturer	Identifier
XRN2	Rabbit	Polyclonal	Bethyl Laboratories	A301-101A
SF3B155	Rabbit	Polyclonal	Abcam	ab39578
Rabbit IgG – conjugated HRP	Goat	Polyclonal	Cell Signalling Technology	7074
Flag [M2]	Mouse	Monoclonal	Sigma	F3165
MYC [9E10]	Mouse	Monoclonal	Sigma	M4439
Mouse IgG – conjugated HRP	Rabbit	Polyclonal	Abcam	ab97046
RNA Pol II Total CTD*	Mouse	Monoclonal	MBL Technologies	MABI0601 or CMA601
Dom3Z/DXO	Rabbit	Polyclonal	Millipore	ABE1306
RNA Pol II Total CTD [8WG16]†	Mouse	Monoclonal	Abcam	ab817
CPSF73	Rabbit	Polyclonal	Bethyl Laboratories	A301-090A
Tubulin	Mouse	Monoclonal	Abcam	ab7291
PP1α (PPP1CA)	Rabbit	Polyclonal	Bethyl Laboratories	A300-904A
PP1β (PPP1CB)	Rabbit	Polyclonal	Bethyl Laboratories	A300-905A
EXOSC10	Mouse	Monoclonal	Santa Cruz Biotechnology	Sc-374595-X

*This product has since been discontinued. †8WG16 is no longer available from Abcam but is available from other suppliers.

2.2 Plasmids

Table 2.2 | Plasmids.

Plasmid Name	Description	Bacterial resistance marker	Source
pX330	Contains human codon optimised CRISPR/Cas9 from *Streptococcus pyogenes* (*S. py*) and two BbsI restriction sites for sgRNA annealing sequence to be introduction upstream of a sgRNA scaffold.	Ampicillin	Addgene #42230
pBABE TIR1-9myc	Contains osTIR1 (from *Oryza sativa*) an E3 ubiquitin ligase and a c-terminal 9x myc tag.	Ampicillin	Addgene# 64945
pSBbi-Bla	A sleeping beauty transposon system plasmid for the constitutive expression of a gene to be cloned between two SfiI restriction sites and the expression of a blasticidin resistance gene.	Ampicillin	Addgene #60526
pSBbi-Pur	A sleeping beauty transposon system plasmid for the constitutive expression of a gene to be cloned between two SfiI restriction sites and the expression of a puromycin resistance gene.	Ampicillin	Addgene# 60523
pCMV(CAT) T7-SB100	A sleeping beauty transposon system plasmid containing the transposase gene encoding the enzyme required to mediate integration of the transposon cassette.	Chloram-phenicol	Addgene #34879
pMK243 (Tet-OsTIR1-PURO)	A HDR-template containing *AAVS1* homology arms flanking a puromycin resistance gene, *TIR1* under the control of a doxycycline (Dox)/tetracycline-inducible promoter and a reverse tetracycline-controlled transactivator, which binds to the TIR1 promoter in the presence of Dox/tetracycline to induce transcrption.	Ampicillin	Addgene #72835
AAVS1 T2 CRISPR plasmid	A pX330-derived plasmid (Addgene #42230) containing the AAVS1 targeting sgRNA.	Ampicillin	Addgene #72833
eSpCas9(1.1)_No_FLAG_ATP1	A pX330-derived plasmid for the tandem expression of two sgRNAs with one targeting an *ATP1A1* exon and the other awaiting	Ampicillin	Addgene #86612

A1_G2 _Dual_sgRN A	cloning between two BbsI restriction sites. Plasmid also expresses an enhanced specificity *S. py* Cas9 (1.1) without an N-terminal FLAG tag.		
eSpCas9(1.1)_No_ FLAG_ATP1 A1_G3 _Dual_sgRN A	A pX330-derived plasmid for the tandem expression of two sgRNAs with one targeting an *ATP1A1* intron and the other awaiting cloning between two BbsI restriction sites. Plasmid also expresses an enhanced specificity *S. py* Cas9 (1.1) without an N-terminal FLAG tag. This plasmid is designed for use in HDR applications.	Ampicillin	Addgene #86613
ATP1A1_pla smid_ donor_RD	A HDR-template containing *ATP1A1* homology arms with mutations causing amino acid substitutions Q118R and N129D, which confer cellular resistance to ouabain.	Ampicillin	Addgene #86551

2.3 Cloning

2.3.1 Recipes for agar plates, media and antibiotic stocks.

LB consists of 10 $g \cdot L^{-1}$ NaCl, 10 $g \cdot L^{-1}$ Tryptone, and 5 $g \cdot L^{-1}$ Yeast extract. For LB-agar plates 15 $g \cdot L^{-1}$ agar is added to LB and the mixture supplemented with an antibiotic if required. Antibiotics stocks are 100 $mg \cdot mL^{-1}$ ampicillin (Sigma #A9518) in 70 % ethanol and 34 $mg \cdot mL^{-1}$ chloramphenicol (Sigma #C0378) in 100 % ethanol. When antibiotics were used to prepare LB-agar plates or LB cultures they were diluted to final concentrations of 100 $\mu g \cdot mL^{-1}$ for ampicillin and 25 $\mu g \cdot mL^{-1}$ for chloramphenicol.

2.3.2 Preparing chemically competent bacterial cells

On occasions purchased NEB 5-alpha Competent E. coli (NEB #C2987H) were used for plasmid cloning and are a DH5α-derivative with the following genotype:

fhuA2 Δ(argF-lacZ)U169 phoA glnV44 Φ80 Δ(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17

However, most transformations used self-prepared chemically competent cells made in batches. First, a 10 mL LB starter culture was inoculated with a

single clone of NEB 5-alpha cells (same as above), which had been previously streaked on an LB agar plate, and incubated overnight at 37 °C on an orbital shaker set to 200 rpm. The following day 2.5 mL of this culture was used to inoculate a 250 mL LB flask that was incubated at 37 °C and 200 rpm until an optical density at 600 nm (OD_{600}) of 0.45-0.55 (~5-6 h). Then the culture was chilled for 2 h on ice and centrifuged at 4000 rpm for 20 min 4 °C. The cell pellet was resuspended in 12 mL of chilled salt solution (100 mM $CaCl_2$, 70 mM $MnCl_2$, 40 mM Na Acetate pH 5.5). The solution was split into four tubes and topped up to 50 mL with salt solution. The cells were incubated on ice in the salt solution for 45 min before centrifuging at 3500 rpm for 10 min 4 °C. The pellets were resuspended in 25 mL salt solution + 15 % (v/v) glycerol (Fisher #10021083). Finally, 50-60 µL aliquots of the bacterial suspension were prepared and snap-frozen in liquid nitrogen before storing at -80 °C for later use.

2.3.3 Bacterial transformations, colony screens and plasmid preparations

Firstly an aliquot of competent cells is thawed on ice for 10 min. To the tube 10 µL of a T4 ligation, 2-4 µL of a Gibson assembly reaction or 10 ng of purified plasmid DNA was added and mixed by flicking a few times before being placed back on ice for a further 20 min. The cells then underwent heat shock at 42 °C for 60 s and quickly placed back on ice for 2 min. The cells were resuspended with 200 µL SOC media (NEB #B9020S), plated onto antibiotic selective LB-agar plates and incubated overnight at 37 °C. For transformations of ligation reactions, colonies were screened by colony PCR. This involves touching a colony with a pipette tip, scraping the pipette tip inside an empty PCR tube and then placed in another tube containing ~50 µL LB. Then a 25 µL PCR reaction was set up in each of the scratched tubes with TAQ DNA polymerase (NEB #M0273) and the products were screened by agarose gel electrophoresis (AGTC #ADG1) using a gel prepared with a 1 in 20000 dilution of Midori Green Advance (#MG04) stain. Colonies that yield a positive screen result by a band of the correct size were taken forward and inoculated in 10 mL LB with antibiotic overnight at 37 °C 200 rpm. After growth mini-prep plasmids purification kit (Qiagen #27106) was carried out according to the manufacturer's instructions. Then plasmid concentrations were determined by NanoDrop and

sent for Sanger sequencing with primers that span the target site to confirm correct cloning.

2.4 Tissue Culture

The human colorectal carcinoma HCT116 cell line and modified derivatives were maintained in Dulbecco's Modified Eagle's Medium-high glucose (Sigma #D6429), 10% foetal bovine serum (FBS; Gibco #10500064), 1x penicillin/streptomycin (Sigma #P4333) at 37 °C and 5 % CO_2. Cells were split with 1x phosphate buffer saline (PBS, Sigma #P4417) and Trypsin-EDTA (Sigma #T4174) for seeding into flasks or dishes (Greiner Bio-One CELLSTAR). For freezing, trypsinised cells were centrifuged slowly at 300xG for 5 min and then resuspended in freezing media (90 % FBS; 10% DMSO Sigma #D2650) before aliquoting into cryovials. Cryovials were frozen in a controlled rate freezing container (with a cooling rate of approximately -1 °C·min^{-1}) placed at -80 °C. For long-term storage cryovials were moved to -150 °C. Cells were thawed quickly by warming cryovials in a 37 °C water bath. Once thawed, cells were transferred to 15 mL tubes containing 10 mL of warm media and pelleted at 300 xG for 5 min. The supernatant was discarded and the pellet resuspended in warm media before seeding appropriately.

2.5 Creation of the homogenous genome-edited cell lines

Table 2.3 | Cell lines.

Designation	Description	Drug resistance (locus)	Additional information
HCT116	Human colorectal carcinoma cell line.	N/A	Authenticated using the Cancer Cell Line Authentication Server.
TIR1	HCT116 cells (as above) modified to constitutively express TIR1 using the sleeping beauty (SB) transposon system.	Blasticidin (SB transposon) .	Eaton et al. (2018).
XRN2-AID	*TIR1* cells (as above) further modified using CRISPR/Cas9 mediated HDR to insert a C-terminal in-frame cassette at the	Blasticidin (SB transposon)	Eaton et al. (2018).

	XRN2 locus containing a (3x)mAID tag, P2A cleavage site, either Hygromycin or Neomycin resistance marker and followed by a stop codon.	; Hygromycin (*XRN2* locus); Neomycin (*XRN2* locus).	
XRN2-AID XRN2-MT(D235A)	*XRN2-AID* cells (as above) further modified using SB transposon system to introduce a constitutively expressed catalytically inactive single point mutant (D235A) of XRN2 using a puromycin resistance marker.	Blasticidin (SB transposon) ; Hygromycin (*XRN2* locus); Neomycin (*XRN2* locus); Puromycin (SB transposon) .	Eaton et al. (2018).
DXO-KO	HCT116 cells (as above) further modified using CRISPR/Cas9 mediated NHEJ to introduce indels within *DXO* and *ATP1A1.* Cells were then enriched for desired edits using an ouabain co-selection strategy (Agudelo et al., 2017).	Ouabain (*ATP1A1* locus).	N/A
DXO-KO XRN2-AID	*XRN2-AID* cells (as above) further modified using CRISPR/Cas9 mediated NHEJ to introduce indels within *DXO* and *ATP1A1.* Cells were then enriched for desired edits using an ouabain co-selection strategy (Agudelo et al., 2017).	Blasticidin (SB transposon) ; Hygromycin (*XRN2* locus); Neomycin (*XRN2* locus); Ouabain	N/A

		(*ATP1A1* locus).	
inducible TIR1	HCT116 cells (as above) further modified using CRISPR/Cas9 mediated HDR to insert a cassette at the *AAVS1* safe-harbour locus containing doxycycline-inducible *TIR1* and constitutively expressed puromycin resistance marker (Natsume et al., 2016).	Puromycin (*AAVS1* locus).	Eaton et al. (2020).
CPSF73-AID	Inducible *TIR1* cells (as above) further modified using CRISPR/Cas9 mediated HDR to insert a C-terminal in-frame cassette at the *CPSF73* locus containing a full-length AID tag, P2A cleavage site, either Hygromycin or Neomycin resistance marker and followed by a stop codon.	Puromycin (*AAVS1* locus); Hygromycin (*CPSF73* locus); Neomycin (*CPSF73* locus).	Eaton et al. (2020).
XRN2-AID RBM3 (δRZ[WT])	*XRN2-AID* cells (as above) further modified using CRISPR/Cas9 mediated HDR to introduce the wild-type hepatitis δ-ribozyme sequence downstream of the *RBM3* PAS and introduce the *ATP1A1* ouabain resistant substitutions Q118R and N129D. Cells were then enriched for desired edits using an ouabain co-selection strategy (Agudelo et al., 2017).	Blasticidin (SB transposon); Hygromycin (*XRN2* locus); Neomycin (*XRN2* locus); Ouabain (*ATP1A1* locus).	Eaton et al. (2020).
XRN2-AID RBM3 (δRZ[MT])	*XRN2-AID* cells (as above) further modified using CRISPR/Cas9 mediated HDR to introduce the single point mutant hepatitis δ-ribozyme sequence downstream of the *RBM3* PAS and introduce the *ATP1A1* ouabain resistant substitutions Q118R and N129D. Cells	Blasticidin (SB transposon); Hygromycin (*XRN2* locus);	Eaton et al. (2020).

	were then enriched for desired edits using an ouabain co-selection strategy (Agudelo et al., 2017).	Neomycin (*XRN2* locus); Ouabain (*ATP1A1* locus).	
CPSF73-AID RBM3 (δRZ[WT])	*CPSF73-AID* cells (as above) further modified using CRISPR/Cas9 mediated HDR to introduce the wild-type hepatitis δ-ribozyme sequence downstream of the *RBM3* PAS and introduce the *ATP1A1* ouabain resistant substitutions Q118R and N129D. Cells were then enriched for desired edits using an ouabain co-selection strategy (Agudelo et al., 2017).	Puromycin (*AAVS1* locus); Hygromycin (*CPSF73* locus); Neomycin (*CPSF73* locus); Ouabain (*ATP1A1* locus).	Eaton et al. (2020).
CPSF73-AID RBM3 (δRZ[MT])	*CPSF73-AID* cells (as above) further modified using CRISPR/Cas9 mediated HDR to introduce the single point mutant hepatitis δ-ribozyme sequence downstream of the *RBM3* PAS and introduce the *ATP1A1* ouabain resistant substitutions Q118R and N129D. Cells were then enriched for desired edits using an ouabain co-selection strategy (Agudelo et al., 2017).	Puromycin (*AAVS1* locus); Hygromycin (*CPSF73* locus); Neomycin (*CPSF73* locus); Ouabain (*ATP1A1* locus).	Eaton et al. (2020).
DIS3-AID	*TIR1* cells (as above) further modified using CRISPR/Cas9 mediated HDR to insert a C-terminal in-frame cassette at the *DIS3* locus containing a full-length AID tag, P2A cleavage site, either Hygromycin or Neomycin resistance marker and followed by a stop codon.	Blasticidin (SB transposon); Hygromycin (*DIS3* locus); Neomycin	Davidson et al. (2019).

		(*DIS3* locus).	
DIS3-AID MORF4L2(xrRNA)	*DIS3-AID* cells (as above) further modified using CRISPR/Cas9 mediated HDR to introduce the West Nile virus XRN-resistant RNA (xrRNA) sequence downstream of the *MORF4L2* PAS and introduce the *ATP1A1* ouabain resistant substitutions Q118R and N129D. Cells were then enriched for desired edits using an ouabain co-selection strategy (Agudelo et al., 2017).	Blasticidin (SB transposon) ; Hygromycin (*DIS3* locus); Neomycin (*DIS3* locus); Ouabain (*ATP1A1* locus).	Eaton et al. (2020).
XRN2-AID MORF4L2(xrRNA)	*XRN2-AID* cells (as above) further modified using CRISPR/Cas9 mediated HDR to introduce the West Nile virus XRN-resistant RNA (xrRNA) sequence downstream of the *MORF4L2* PAS and introduce the *ATP1A1* ouabain resistant substitutions Q118R and N129D. Cells were then enriched for desired edits using an ouabain co-selection strategy (Agudelo et al., 2017).	Blasticidin (SB transposon) ; Hygromycin (*XRN2* locus); Neomycin (*XRN2* locus); Ouabain (*ATP1A1* locus).	Eaton et al. (2020).

2.5.1 Annealing sgRNA primers and ligation into BbsI cut px330-derived plasmids

First, 2 µg of the px330-derived plasmid used for sgRNA cloning was cut with BbsI (NEB #R0539) restriction enzyme using NEBuffer 2.1 in a 50 µL reaction volume for 2 h at 37 °C. The linearised plasmid product was purified by phenol/chloroform extraction and ethanol precipitation. This involves adding 50 µL H_2O and then 100 µL (1:1) phenol/chloroform mix (basic phenol Sigma #P4557). The mixture was shaken and then centrifuged at 16,000 xG for 10

min. The top aqueous phase (~100 µL) was taken to a tube containing 1 µL glycogen, 10 µL Na Acetate pH 5.5 when 250 µL 100 % ethanol was added, mixed and then centrifuged at 16,000 xG for 10 min 4 °C. The pellet was washed with 70 % ethanol, resuspended in 20 µL H_2O and then quantified by NanoDrop (ND-2000).

For the insert, forward and reverse sgRNA primers with vector-compatible sticky ends were added to oligo annealing buffer (10 mM Tris pH 7.5, 50 mM NaCl) to a final concentration of 5 µM each in a 50 µL volume. To anneal primers the solution was first denatured at 95 °C for 5 min and allowed to cool slowly to RT. Next, 1 µL of a 1 in 30 dilution of the annealed primers was added along with 50 ng BbsI cut vector to a 20 µL T4 DNA ligase reaction (NEB #M0202S) and incubated at RT for 1 h. Finally, 10 µL was transformed into chemically competent DH5α *E. coli* cells, as described in 2.3.3. The colonies were screened by PCR (see section 2.3.3) using the reverse sgRNA cloning primer and either 'CRISPR_sgRNA_scr_F' for px330 (Addgene #42230) or 'Dual_sgRNA_scr_F' for dual sgRNA ouabain plasmids (Addgene #86612 & #86613). Clones with the correct sized band were inoculated in LB, had plasmid mini-preps prepared (Qiagen #27106) and Sanger sequencing performed using either the 'CRISPR_sgRNA_scr_F' or 'Dual_sgRNA_scr_F' primers (for single and dual sgRNA plasmids, respectively, as above) to confirm correct cloning.

2.5.2 T7 endonuclease surveyor assay

The T7 endonuclease surveyor assay for screening sgRNAs targeting RBM3 and MORF4L2 3'-end loci used an ouabain enrichment strategy to co-select for transfected cells with ouabain-resistance conferring indels at the ATP1A1 locus. The screened sgRNA was cloned into the ATP1A1-G2 dual sgRNA plasmid (see section 2.2). First, a 25 % confluent 6-well plate was used to transfect 2 µg of the sgRNA+Cas9 containing plasmid per well using 4 µL jetPRIME reagent (Polyplus transfection) according to the manufacturer's protocol. Then 24 h post-transfection the media was replenished, after a further 24 h later ouabain was added to 0.5 µM in media. Ouabain selection was maintained for 10 days with media changed every 3 days. After selection adhered cells were washed in 1x PBS and harvested into pellets by centrifuging at 500 xG for 5 min. Then genomic DNA was isolated using QuickExtract DNA

Extract solution (EpiCentre). Briefly, this involves resuspending the cell pellet in x10 its corresponding volume with the QuickExtract solution, incubating for 8 min at 65 °C and then for 2 min at 98 °C with vortexing for 15 s in between steps. Using 1uL of this QuickExtract as genomic DNA (gDNA) template a 25 µL Q5 PCR reaction (NEB #M0491S) was prepared per condition with primers that span the sgRNA target site and generate an expected fragment size of ~300-500 bp. The PCR product was purified by DNA phenol/chloroform ethanol precipitation and resuspended in 25 µL H_2O. Then for each PCR template, a +/- enzyme condition was prepared to contain 10 µL of the DNA template, 2 µL T7E1 reaction buffer, and 6 µL H_2O. These templates were incubated for 5 min at 95 °C, cooled -2 °C·s^{-1} until 85 °C, and cooled at -0.1 °C·s^{-1} until 4 °C when tubes were placed on ice. Finally, 2 µL T7E1 (NEB #M0302S) or H_2O was added to each +/- condition respectively and incubated at 37 °C for 1 h. The resulting products were run on a 2% agarose gel stained with Midori Green and imaged using UV illumination (Bio-Rad ChemiDoc XRS+). A comparison of band intensity between T7E1 treated and untreated conditions were used to estimate target site cleavage for each sgRNA tested.

2.5.3 Generation of auxin-inducible degron (AID) system cell lines

The generation of *XRN2-AID* and *CPSF73-AID* cell lines involved CRISPR/Cas9 mediated homology-directed repair (HDR) to insert an in-frame AID cassette. This cassette, whose sequences can be found in section 2.12.5, consisted of a degron (*3xminiAID* or full-length *AID* respectively), P2A cleavage site, Hygromycin or Neomycin resistance marker and followed by a stop codon. These sequences were surrounded on either side by symmetric (~500 bp) homology arms and cloned within a pUC18 vector (repair donor). Cells were seeded to ~40 % confluency in a 6-well plate and transfected with 3 µg total DNA (1 µg of each plasmid) consisting of px330 (sgRNA & Cas9; Addgene #42230) plasmid, pUC repair donor with Hygromycin resistance marker and pUC repair donor with Neomycin marker using 4 µL jetPRIME reagent (Polyplus transfection) and 200 µL JetPrime buffer per well. Media was replaced 24 h post-transfection and 48 h post-transfection cells were split into a 100 mm dish containing 150 µg·mL^{-1} Hygromycin B (Invitrogen #10687010) and 800 µg·mL^{-1} Neomycin (G418 Sigma#A1720). After ~10 days under antibiotic selection single colonies were transferred to single wells of a 24 well plate before screening by PCR and western blotting. Genomic insertions were confirmed by

Sanger sequencing. *DIS3-AID* cells where generated similarly to *XRN2-AID* cells using the constitutively-expressed (sleeping beauty) *TIR1* cells as the parental cell line, however, the degron tag is the full-length AID (*IAA17*) sequence (Davidson et al., 2019).

2.5.4 Generation of *DXO-KO* cell lines using ouabain co-selection

The generation of *DXO-KO* and *DXO-KO XRN2-AID* cells involved CRISPR/Cas9 mediated non-homologous end joining (NHEJ) to introduce indels at DXO and ATP1A1 and co-selection strategy using ouabain (Agudelo et al., 2017). Cells were seeded ~40 % confluency in a 6-well plate and transfected with 2ug of dual sgRNA (DXO and ATP1A1-guide2 sgRNAs) & Cas9 plasmid (Addgene #86612). Otherwise, the transfection was carried out to the manufacturer's protocol with 4 µL jetPRIME reagent (Polyplus transfection) and 200 µL JetPrime buffer per well. The media was replaced 24 h post-transfection and 48 h post-transfection cells were split into a 100 mm dish containing 0.5 µM ouabain and maintain for ~10 days. After single colonies were transferred to separate wells in a 24 well plate and then screened by PCR and western blotting. The indels were determined by cloning the DXO sgRNA target region into a pUC18 vector and performing Sanger sequencing of multiple clones from the transformed plate to confirm the presence of frame-shift mutations within both DXO alleles.

2.5.5 Genomic insertion of xrRNA, δRZ and MALAT1 3'-end sequences

The insertions of δRZ, xrRNA and MALAT1 3'-end sequences downstream of *RBM3* and *MORF4L2* were carried out using a similar ouabain co-selection strategy as for DXO-KO generation (see the above section 2.5.4; Agudelo et al., 2017). However, these cell lines utilised CRISPR/Cas9 mediated HDR to insert sequences instead. Cells were seeded ~40 % confluency in a 6-well plate and transfected with 3 µg (1 µg each plasmid) of dual sgRNA (DXO and ATP1A1-guide3 sgRNAs) & Cas9 plasmid (Addgene #86613), an ATP1A1 repair donor (Addgene #86551) and a pUC repair donor containing the insertion sequence. The transfection used 4 µL jetPRIME reagent (Polyplus transfection) and 200 µL JetPrime buffer per well. Media was replaced 24 h post-transfection and 48 h post-transfection cells were split into a 100 mm dish containing 0.5 µM ouabain and maintain for ~10 days. After, single colonies were transferred to separate wells in a 24 well plate and then

screened by PCR and western blotting. Finally, a genomic PCR fragment produced using at least one primer annealing outside the homology arms of the repair plasmid was Sanger Sequenced to confirm the inserted sequence.

2.6 Mammalian native elongating transcript sequencing (mNET-seq)

2.6.1 mNET-seq library preparation

mNET-seq was carried out with the following modifications to the Nojima et al. (2016) protocol. Two 145 mm dishes were used per condition and seeded to give a 70-80 % confluency for the beginning of the experiment. For depletion conditions, auxin was added to culture media to a final concentration of 500 µM for 2 h prior to cell harvest. The MNase digestion of chromatin pellets occurred for 90 s and was then quenched with EGTA. An immunoprecipitation (IP) of the solubilised Pol II-associated complexes was incubated for 1 h at 4 °C using 3.75 µg RNA Pol II CTD (MABI0601) antibody and 50 µL M-280 sheep anti-Mouse IgG Dynabeads™ (Invitrogen #11202D) per 145 mm dish. After washing, pooling of duplicate tubes and 5' T4 PNK (NEB #M0236S) treatment the RNA was eluted off the Dynabeads with 300 µL of a modified RNA Lysis buffer consisting of 1:1:1 ratio of ZYMO RNA lysis buffer:100% ethanol:NET2-buffer (ZYMO Quick-RNA Microprep Kit #R1050). To isolate RNA fragments < 200 nt the eluates were passed through a ZYMO spin column and the flow-through collected. To this 300 µL 100% ethanol was added before being bound to a second column and washed with ethanol using the standard ZYMO Quick-RNA protocol. These short RNA fragments were eluted in 8 µL H_2O and 2 µL of this used to assess the size distribution and quantity on a TapeStation 2200 (Agilent) with a Hi-Sensitivity RNA ScreenTape. The remaining 6 µL for each sample were made into libraries using NEBNext™ Small RNA Library (#E7330S) according to the manufacturer's protocol. After PCR amplification, the cDNA constructs were purified using a Zymo DNA clean and concentrator-5 kit (#RD4003) before loading onto a 5% PAGE-TBE gel (with Ficoll-400 loading dye) and run at 250V for ~25 min. The gel was then stained for 40 min with 1x SYBR gold (Invitrogen #S11494) in TBE buffer. The gel was visualised on a blue-light box to allow 140-220nt band excision and then cDNA libraries were purified from gel slices as described in Nojima et al. (2016). The final libraries were quality checked on a D1000 DNA ScreenTape and qPCR-determined Illumina library adapter concentrations used for pooling. Pooled libraries were

sequenced on an Illumina HiSeq 2500 platform with a 50 bp paired-end Rapid Run flow cell by Exeter Sequencing Service.

2.6.2 mNET-seq bioinformatic data processing

For mammalian native elongating transcript sequencing 50 bp paired-end reads were assessed for quality using FASTQC. Adapters were removed from reads using Trim Galore! v0.4.4 (wrapper tool for Cutadapt v.1.15). Reads were aligned to Ensembl human GRCH38.p10 release 91 with Hisat2 v2.0.5 using known splice sites extracted from Gencode release 27. The paired-end parameter '--fr' to select concordantly mapped pairs was specified. mNET-seq single nucleotide resolution files were generated with the 3' most nucleotide of fragments (reverse complement of the 5' most nucleotide of the second read displayed with the strandedness of the first read) for concordantly mapped reads with the samflag pairs 99/147 and 83/163. Then bigWig files were generated for each strand using bamCoverage from Deeptools v3.0.1. Specifically, the settings used were the 'BPM' option (equivalent to TPM for RNA-seq data) over single bp windows across the whole genome.

2.7 Chromatin-associated RNA sequencing (chrRNA-seq)

2.7.1 chrRNA-seq Library preparation

Cell pellets were collected from one 100 mm dish per condition and then resuspended in hypotonic lysis buffer to collect nuclei (HLB; 10 mM Tris at pH 7.5, 10 mM NaCl, 2.5 mM MgCl2, 0.5 % NP40). From this, nuclei were isolated by underlaying the lysate with HLB+10 % sucrose and centrifuged at 500 xG for 5 min. The pelleted nuclei were resuspended in 100 µL of NUN1 (20 mM Tris-HCl at pH 7.9, 75 mM NaCl, 0.5 mM EDTA, 50 % glycerol, 0.85 mM DTT) and transferred to a new tube. Then 1 mL of NUN2 (20 mM Tris-HCl at pH 7.9, 75 mM NaCl, 0.5 mM EDTA, 50 % glycerol, 0.85 mM DTT) a denaturing lysis buffer was added and the mix incubated on ice for 20 min with regular vortex throughout. An insoluble chromatin pellet was gathered by centrifuging at 16000 xG for 10 min and RNA isolated from this using 1 mL TRIzol (Invitrogen #15596026). To each tube, 200 µL chloroform was added, the sample shaken and centrifuges at 16000 xG for 15 min. The top aqueous phase (~500 µL) was transferred to a new tube with 1 µL glycogen carrier (Roche #10901393001) and then 500 µL Isopropanol was added before centrifuging at 16000 xG for 10 min 4 °C. Following this, the pellet was washed with 75 % ethanol and

centrifuged again before the removal of ethanol. Then the RNA was resuspended in nuclease free water at briefly placed at 50 °C for 1-2 min to aid resuspension. The RNA was checked using a TapeStation 2200 to confirm RNA integrity with a RIN score, which is based on the 18S and 26S rRNA peak abundance. Next rRNA was depleted using Ribo-Zero (Illumina) and the library constructed using a Tru-Seq Stranded Total RNA kit (Illumina #20020597) according to the manufacturer's protocol. Finally, these cDNA libraries were pooled and sequenced on an Illumina HiSeq 2500 platform with a 50 bp single-end Rapid Run flow cell by Exeter Sequencing Service.

2.7.2 chrRNA-seq bioinformatic data processing

For chromatin-associated RNA-seq raw 50 bp single-end reads were assessed for quality using FASTQC. Adapters were removed from reads using Trim Galore! v0.4.4 (wrapper tool for Cutadapt v.1.15). Reads shorter than 20bp were discarded. Alignment of chromatin-associated RNA-seq reads to the Ensembl human GRCH38.p10 release 91 with Hisat2 v2.0.5 using known splice sites extracted from Gencode release 27. The primary alignments for single-end mapped reads (i.e. excluding SAM flag 260) were extracted using SAMtools v1.4.1 and converted to BAM files. Then bigWig files were generated for each strand using bamCoverage from Deeptools v3.0.1. Specifically, the settings used were the 'BPM' option (equivalent to TPM for RNA-seq data) over single bp windows across the whole genome.

2.8 Deposited Gene Expression Omnibus (GEO) data

Deposited GEO datasets can be found at https://www.ncbi.nlm.nih.gov/geo/ with the following accession numbers:

- mNET-seq of XRN2-AID cells (GSE109003).
- Chromatin-associated RNA-seq of CPSF73-*AID* cells (GSE137727).

2.9 Quantitative reverse transcription and PCR (qRT-PCR)

An ~80% confluent well of a 24-well plate was used for each condition. Media was removed and adhered cells were washed with 1x PBS. Then total RNA was isolated using 0.5 mL·well^{-1} of Tri-Reagent/TRIzol (Invitrogen #15596026). To each tube, 100 µL chloroform was added, the sample shaken and centrifuges at 16000 xG for 15 min. The top aqueous phase (~250 µL) was transferred to a new tube with 1 µL glycogen carrier (Roche #10901393001) and then 250 µL Isopropanol was added before centrifuging at 16000 xG for 10

min 4 °C. Following this, the pellet was washed with 75 % ethanol. Then the RNA was resuspended in nuclease free water at briefly placed at 50 °C for 2-3 min to aid resuspension. The RNA was then DNase treated to remove residual traces for 1h at 37 °C with Turbo-DNase (Invitrogen #AM2238) in a 100 µL volume with 2 µL RNase inhibitor present (NEB #M0314S). The DNase-treated RNA was then purified using phenol/chloroform extraction, which involves the following steps interspersed with 16000 xG centrifugation for 10 min: mixing with 100 µL of (1:1) phenol/chloroform (acidic phenol Sigma #P4682 and chloroform); removal of aqueous phase into a new tube containing 1 µL glycogen+10 µL Na Acetate pH 5.5 + 250 µL 100 % ethanol and; washing the pellet in 75 % ethanol. The RNA pellet was resuspended in 30 µL of nuclease free-water and briefly placed at 50 °C for 2-3 min to aid resuspension. The purified DNase-treated RNA was then quantified using a NanoDrop (ND-2000) and 1 µg was annealed to random hexamers by incubating for 5 min at 70 °C before snap-cooling on ice. Each sample was then reverse transcribed with protoscript II RT (NEB #M0368S) according to manufacturer's protocol in a 20 µL reaction volume using following incubation steps: 5 min at 25 °C, 1 h at 42 °C and 20 min at 70 °C. The resulting cDNA was then diluted with 30 µL H_2O. Finally, 1 µL of the 50 µL diluted cDNA was used in each 8 µL total reaction mix containing Luna qPCR Master Mix (NEB #M3003S) with amplicon primers. Samples were then analysed on a Qiagen Rotorgene Q fitted with a 72-well rotor. For each amplicon comparative quantitation is used to show the relative fold change compared to a control condition and then samples are normalised to relative fold changes of a spliced *ACTB* amplicon to control for variations in input RNA.

2.10 Chromatin Immunoprecipitation and quantitative PCR (ChIP-qPCR)

Cells were seeded for a 70 % confluency with one 100 mm dish used per condition. Dishes were rinsed in 1x PBS before cross-linked in 0.5 % formaldehyde PBS for 10 min and quenched in 125 mM glycine PBS for a further 5 min. Cells were pelleted at 500 xG 4 min 4 °C and then resuspended in 400 µL RIPA ChIP buffer (50 mM Tris-HCl pH 8, 150 mM NaCl, 1 % NP40, 0.5 % DOC, 0.1 % SDS, 5 mM EDTA pH 8). The resuspended lysate was then sonicated in a Bioruptor Plus (Diagenode) for 10 cycles of 30 s on, 30 s off at 4

°C on high setting. The samples were then centrifuged at 16000 xG for 10 min 4 °C and after the supernatant transferred to a new tube. The supernatant was then split into two tubes with a further 10 % of the volume kept in a third tube for input. Then 40 µL of M-280 sheep anti-Mouse IgG Dynabeads™ (Invitrogen #11202D) that had been pre-incubated for ~2 h with either 3 µg antibody or mock-treated (in RIPA ChIP with 1x protease inhibitor (Roche #11836170001). The IP was placed on a rotating wheel for 3h at 4 °C and after were first rinsed in RIPA ChIP, then in high-salt ChIP-wash buffer (100 mM Tris-HCl pH 8.0, 500 mM NaCl, 1 % NP40, 1 % DOC), followed by two longer washes with ChIP-wash with 5 min rotating on the wheel at 4 °C and finally rinsed once more in RIPA ChIP buffer. Complexes were eluted off beads with 500 µL elution buffer (0.1 M $NaHCO_3$, 1 % SDS) per sample for 30 min on a rotating wheel and the supernatant transferred to a new tube. Cross-links were reversed overnight by adding 25 µL 5M NaCl and placing tubes at 65 °C. The DNA was then purified by (1:1) phenol/chloroform extraction (basic phenol Sigma #P4557 and chloroform) and ethanol precipitation with pellets resuspended in 100 µL H_2O. Then 1 µL of each sample was used per 8 µL qPCR total reaction containing Luna qPCR Master Mix (NEB #M3003S). Antibody signal was first calculated as a percentage of input signal and then normalised relative to the Pol II IP signal of an upstream amplicon.

2.11 Western blotting

Cells grown in a 6-well plate were washed with 1xPBS, scraped in 1xPBS, transferred to a microcentrifuge tube and spun at 500 xG for 5 min. Cell pellets were lysed in 10-fold the pellet volume of RIPA buffer (50 mM Tris-HCl pH 7.4, 150 mM NaCl, 1 % NP-40, 1 % DOC, 0.1 % SDS, 1 mM EDTA), vortexed 2-3 times during a 20 min incubation on ice. Then cell lysates were centrifuged at 16,000 xG for 10 min and the supernatant transferred to a new tube. To this 4xSDS-loading buffer (8 % SDS, 40 % glycerol, 250 mM Tris-HCl pH 6.8, 0.006 % Bromophenol-blue and add 50 µL β-mercaptoethanol to 500 µL of the mix prior to use) was added before loading onto a discontinuous gels with 1mm depth mini-gel glass plates (Bio-Rad). The gel was run (in 1x running buffer of 25 mM Tris; 192 mM glycine, 0.1 % SDS) for approximately 1 h at 25 mA. Next, protein from the gel was transferred onto a nitrocellulose membrane by semi-dry transfer using Bio-Rad Trans-Blot Turbo (with 1x transfer buffer of

48 mM Tris, 39 mM Glycine, 20% Methanol, 1.39 mM SDS) and the efficacy checked by ponceau staining (Sigma #P7170). The membrane was blocked in 5 % (w/v) non-fat milk powder PBST (1xPBS and 0.1 % TWEEN-20) for 1 h and probed with primary antibody, which unless otherwise stated was with 1:1000 dilution in 5 % milk-PBST for 1.5 h. After primary antibody probing the membrane was washed 3 times for 5 min in PBST and probed with 1:2500 dilution of an α-species IgG secondary antibody conjugated to horse radish peroxidase in 5 % milk-PBST for 30 min. Afterwards the membrane was again washed 3 times and incubated for 1 min in a 1:1 ECL reaction mix [Solution1 = 100 mM Tris-HCl pH8.5, 2.5 mM Luminol (Sigma #123072), 400 µM p-Coumaric Acid (Sigma #C9008); Solution 2 = 100 mM Tris-HCl pH8.5, 0.02 % H_2O_2 (Sigma #H1009)] before chemiluminescence image acquisition (Bio-Rad ChemiDoc XRS+).

2.12 Nucleic acid sequences

2.12.1 Primers for qPCR

ACTB spliced: catccgcaaagacctgtacg/cctgcttgctgatccacatc

ACTB UCPA: gcttttggtctccctggga/ctgcactctgggtaaggaca

ACTB ds1.7kb: ccaaccagatgtgttccgtg/caagaccaccaccacaatcg

ACTB ds6.3kb: aggaggcaatgctggagaat/gtacctgggaactctgcact

ACTB ds9.3kb: cagggaagacgtgctaggaa/tcctttctcctctgctcagc

MORF4L2 Ex4: tcttgaaccagctctcccag/tactgccaccatctccgttt

MORF4L2 UCPA: gtagccacggttttctggaaa/ accagtaacatgaaaggcacac

MORF4L2 ds200: tgttactggttggtattctggt/ tttgagtcccatttatttgctgg

MORF4L2 ds600: accccagtgacctcatttagt/ acacccgccaaattcatgtt

MORF4L2 ds2.7kb: agcatgctagtgggaaatcc/ggatctcctcaggctttggt

MORF4L2 ds4.2kb: ccccatgacattcagtgcct/tgcttccgtaccaatccaca

MORF4L2 ds8.5kb: gccaaggacacacagctaag/ tccttttcagagagccagga

MYC ds5kb: tggaagaggagccaaaggag/ggaagctgcgttcatgtgat

MYC ds7.6kb: gaacccctctttccctccaa/ccccaaagctaccacaggat

PPP1CA (PP1α) spliced: accccgagaacttcttcctg/gatgggcaggcagttgaag

PPP1CB (PP1β) spliced: tggtggaatgatgagtgtgga/cacctttcttcggcggatt

RBM3 UCPA: tgctgtgaaagagtatattcgt/gtctgccttgtttcttggctcc

RBM3 ds1.1kb: gaatcaggcatttacaggactggc/agcgcatgcccaattacctttac

RBM3 ds1.9kb: gtcgctcccatgtacaacac/actgctatgagaaggtgggc

RBM3 ds8.5kb: ccattgtggtcagaaaggctcttg/tggaccccaccaatgcatgatata

RBM3 ds11kb: gggcagtaaacccctctagagttc/ggttgtggtgatagcctgcattac

YTHDF3 ds10kb: acaaaaggacagcagaggga/agcctcttcttatgccaccc

YTHDF3 ds20kb: agcagctgtctagacccaag/gtagcaacgcctttccagag

2.12.2 Primers for screening sgRNAs

CRISPR_sgRNA_scr_F: gagagataattggaattaatttgact

Dual_sgRNA_scr_F: ttttacggttcctggccttttg

2.12.3 sgRNAs target sites

XRN2: AGGGATATCCCAGAGAAGGA

CPSF73: GGCTGCACAGAGACTGTACG

DXO: GTGCTGGCTCCTGGAACACCG

RBM3: GTTATCTATGATAACTAGCA

MORF4L2: TCCCTGAGTTGCCACCAGAG

2.12.4 siRNA sequences

Negative Control 1: Thermo Fisher Silencer Select siRNA #4390843

PP1α: Thermo Fisher Silencer Select siRNA #s10930

PP1β: Thermo Fisher Silencer Select siRNA #s10935

XRN2: Thermo Fisher Silencer Select siRNA #s22412

2.12.5 Sequences used for genomic insertion

3xflag 3xmini-aid

GGGGGTGGCAGCGGCGACTACAAAGATCACGACGGAGACTATAAAGATCACGACATCGATTATAAAGATG
ACGACGATAAAGGTTCCGGTAAGGAAAAGAGCGCTTGCCCGAAGGATCCCGCAAAGCCCCCTGCTAAGGC
TCAGGTGGTCGGTTGGCCACCTGTACGATCCTATCGAAAGAATGTCATGGTATCTTGCCAGAAGTCTTCC
GGTGGTCCAGAGGCCGCTGCATTCGTAAAGGTTAGCATGGATGGTGCCCCTTATCTCCGGAAGATAGACT
TGAGGATGTATAAGGGCGGCGGTAGCGGTGGTGGAAAAGAGAAATCCGCTTGCCCCAAGGATCCAGCAAA
ACCTCCGGCCAAGGCTCAAGTGGTGGGTTGGCCCCCAGTAAGGTCTTACCGCAAAAACGTCATGGTCAGC
TGTCAAAAAAGTTCCGGCGGTCCAGAAGCAGCAGCATTCGTAAAAGTCTCCATGGATGGGGCCCCCTATC
TCAGAAAAATAGACCTGAGGATGTATAAAGGTGGCGGATCAGGTGGGAAGGAGAAGTCCGCCTGCCCGAA
GGACCCGGCCAAGCCACCGGCGAAAGCGCAAGTGGTAGGTTGGCCTCCAGTTAGGAGCTATCGGAAAAAT
GTTATGGTGAGTTGCCAGAAATCATCTGGAGGACCTGAAGCGGCTGCGTTTGTAAAGGTCTCTATGGACG
GTGCGCCGTATTTGCGCAAGATCGATCTTAGAATGTATAAG

P2A

GGATCAGGGGCCACTAACTTTTCCCTGCTGAAGCAGGCCGGAGACGTGGAGGAGAACCCCGGGCCC

Neomycin resistance gene

ATGCCTGTAATTTCTACCCAGACTGGACGGGCCATGATTGAGCAAGACGGGCTCCACGCTGGCAGCCCCG
CAGCTTGGGTCGAGCGACTGTTCGGGTACGATTGGGCACAGCAGACAATAGGGTGCAGCGATGCCGCCGT
CTTCCGGCTCAGCGCGCAAGGCCGGCCTGTCCTGTTTGTTAAAACCGATCTGAGCGGGGCCCTGAACGAA
CTGCAGGATGAGGCGGCTAGACTTAGCTGGCTTGCGACCACCGGAGTGCCGTGTGCTGCCGTTCTGGACG
TCGTAACAGAGGCGGGAAGGGATTGGCTGCTGCTCGGGGAGGTCCCTGGCCAAGATTTGTTGTCCTCCCA

CCTGGCACCTGCAGAGAAGGTAAGCATCATGGCAGATGCCATGCGCAGGCTGCACACCCTGGATCCCGCC
ACGTGTCCTTTCGACCACCAGGCCAAGCACCGAATTGAGAGGGCCAGGACACGCATGGAGGCCGGCCTGG
TGGATCAGGACGATCTTGACGAGGAACATCAGGGCCTCGCCCCAGCGGAGCTCTTTGCTCGGCTGAAAGC
TAGAATGCCTGATGGTGAAGATCTCGTCGTGACCCACGGAGATGCCTGCCTGCCCAACATCATGGTAGAA
AACGGACGCTTCTCTGGCTTTATCGATTGTGGCCGGCTTGGAGTTGCTGATAGATATCAGGACATTGCAC
TCGCGACAAGAGACATTGCCGAGGAACTCGGTGGTGAATGGGCAGACCGGTTCCTGGTGCTGTACGGGAT
CGCTGCCCCTGACTCACAGAGGATCGCATTTTACAGGTTGCTGGACGAATTTTTTTAA

Hygromycin resistance gene

ATGAAAAAGCCTGAACTCACCGCGACGTCTGTCGAGAAGTTTCTGATCGAAAAGTTCGACAGCGTCTCCG
ACCTGATGCAGCTCTCGGAGGGCGAAGAATCTCGTGCTTTCAGCTTCGATGTAGGAGGGCGTGGATATGT
CCTGCGGGTAAATAGCTGCGCCGATGGTTTCTACAAAGATCGTTATGTTTATCGGCACTTTGCATCGGCC
GCGCTCCCGATTCCGGAAGTGCTTGACATTGGGGAGTTCAGCGAGAGCCTGACCTATTGCATCTCCCGCC
GTGCACAGGGTGTCACGTTGCAAGACCTGCCTGAAACCGAACTGCCCGCTGTTCTTCAGCCGGTCGCGGA
GGCTATGGATGCGATCGCTGCGGCCGATCTTAGCCAGACGAGCGGGTTCGGCCCATTCGGACCGCAAGGA
ATCGGTCAATACACTACATGGCGTGATTTCATATGCGCGATTGCTGATCCCCATGTGTATCACTGGCAAA
CTGTGATGGACGACACCGTCAGTGCGTCCGTCGCGCAGGCTCTCGATGAGCTGATGCTTTGGGCCGAGGA
CTGCCCCGAAGTCCGGCACCTCGTGCACGCGGATTTCGGCTCCAACAATGTCCTGACGGACAATGGCCGC
ATAACAGCGGTCATTGACTGGAGCGAGGCGATGTTCGGGGATTCCCAATACGAGGTCGCCAACATCTTCT
TCTGGAGGCCGTGGTTGGCTTGTATGGAGCAGCAGACGCGCTACTTCGAGCGGAGGCATCCGGAGCTTGC
AGGATCGCCACGCCTCCGGGCGTATATGCTCCGCATTGGTCTTGACCAACTCTATCAGAGCTTGGTTGAC
GGCAATTTCGATGATGCAGCTTGGGCGCAGGGTCGATGCGACGCAATCGTCCGATCCGGAGCCGGGACTG
TCGGGCGTACACAAATCGCCCGCAGAAGCGCGGCCGTCTGGACCGATGGCTGTGTAGAAGTACTCGCCGA
TAGTGGAAACCGACGCCCCAGCACTCGTCCGAGGGCAAAGGAATAG

SV40 poly(A) signal

AACTTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCACAAATAAAGCAT
TTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTTA

XRN2 5' homology arm

GTGTAATAAGTCTAAATTGATGTGGGTATCTTACCACAAAGTGACTTGAATTACTACTGCTAGGACAGTG
AGAAAATTGAGAACCACTGTCTGTACATGTTGTTTACACAGAACACTTTAGTTATTTGTGTGCATTTGTG
ATTGTTAAGGTTTTTTGTTTTATTTTTCAGTAATAGCATTTGTGCTAGCCTCCAACTTTGCAACAAGTCT
GTATTAAAGCTCTGGATCAAAGCACCTTTTATGGGGCCTTTCCATGTGCTGTACCTTTAACACATACTCA
GTTTCCTTATGATGTGTTTTTCCATAGAGGTTTAAAGTTAACTGACTTGCAGGAGTATCGGTCCAGAAAA
TAAACTCTTTCTTTTGTTTATTTTCAGGGATATCCCAGAGAAGGAAGAAAATACCCTTTGCCACCACCCT
CAGGAAGATACAATTGGAAT

XRN2 3' homology arm

GCTTTTGTAAAGCTTTCCCAAATCCTTTCATCATTCTACAGTTTTATGCTATTTGTGGAAAGATTTCTTT
CTCAAGTAGTAGTTTTTAATAAAACTACAGTACTTTGTGTATTTCTTTTAACTGTGTATATTTCTACTGA
TCTGATCTCACTGTTTATGTTGCTTTCCAAAGATGTATGTTGCATAATACAGTGGATCTGAATTTATTAT
TGCTTATAAAACACATTTGATGGAATAGGAGTACTGGTTTTTCATAATGGTTAAAAATGAAACCAGCTGT
GGATTTCAAAACACAGTGTATTCTAGATCATCTAAGATCCATGCTGATTTTTATTGCACAAGAATTAGGT
TTGAACTCGAGCTGGAACCTCAGCAAACTAGAGTATAT

Codon optimised AID (IAA17)

GGTAGTGGCATGATGGGTAGTGTGGAGCTGAACCTGCGCGAGACCGAGCTGTGCTTGGGACTGCCTGGCG
GCGATACGGTTGCACCCGTTACCGGGAACAAGAGGGGCTTCAGCGAGACAGTGGATCTCAAGCTGAATCT
GAACAACGAACCTGCAAATAAAGAGGGAAGCACCACTCATGACGTAGTGACATTCGATAGTAAAGAGAAA
TCTGCTTGCCCGAAGGATCCAGCTAAGCCCCCGGCCAAGGCCCAGGTGGTGGGATGGCCCCCGGTGCGCT

CCTACCGCAAAAACGTGATGGTATCATGCCAGAAAAGCAGCGGGGGGCCCGAAGCCGCCGCTTTTGTTAA
AGTGTCAATGGACGGGGCTCCATACCTGAGGAAGATCGATCTCCGGATGTACAAGTCTTACGATGAACTG
AGCAACGCGCTTTCAAACATGTTCTCATCTTTCACCATGGGAAAGCATGGGGGCGAAGAAGGAATGATTG
ACTTCATGAATGAGAGAAAACTGATGGATCTCGTCAATTCTTGGGACTACGTGCCTTCATACGAGGATAA
GGATGGAGATTGGATGCTGGTAGGAGACGTGCCTTGGCCCATGTTCGTGGACACTTGCAAAAGGCTCAGA
CTGATGAAGGGTAGCGATGCCATCGGCTTGGCACCCCGCGCGATGGAGAAGTGTAAATCTAGGGCC

CPSF73 5' homology arm

CCACATCCATTCCTTGCCAAGTATCATTTACTAGATCAAACTGTGGGCTTTGATGTAAATGTAGTTTACT
AGACTTTCCCCAGTCTTTCACCCCAGCCTCAAGTCATCACTAATTAGGACCGTGCTGCTGTCAGGAAGCA
CTGCACGCCCACAAGTGTGTAGGGCGGCCGTTCTGTTTCATGGTAATCAGTCCCACCATGACCTCTGCA
CACACAGATGATGTTCTTTTTTTTAGTTTGAGACCCGGTCTCGCAGTGCCGCCCAGGCTGGAGTGCAGT
GGTGCAGTCACAGCTCACTGCAGCCTCAACCTTCCCGGCTCAGTGATCCTCCCACCTCAGACTCTTATCT
GGGACCACAGGCACACGCCACCACAGCTGGCTAATTTTTTATGAGATGATGGTTTTTTTAAAGAGTATT
CATTTATCTTCTATATAATCATTATAGACTTAATTCTAACAGTCTTGTTTGTGCCTCACTTTCAGACTGTAG
AATGTGAAGAGGGAAGTGAAGACGATGAATCCCTCCGAGAAATGGTGGAGCTGGCTGCACAGAGACT
GTACGAAGCCCTGACGCCAGTTCAC

CPSF73 3' homology arm

GACTGTGCCTGTATATGAACTTTGAAAAAATACTTGACTCTACTTTTGTTACCTAAAATAAAATGCATTC
GTTTCTCTGGGGGAGCCTGTTTACTTTTAATGTCAAATGGCCTTTATTTCAACAGCCTGAATACTGCTAA
ATTGCTAATTAATTTGTCCATTATTCTAGAACTAACTACTAGATCAACTGCCCATTATTTTAGAATTTTG
GATTCTTCTTCCAGGCATGTATGTGCAGCTCCCATTGAAACCATCAAGATCTGCCGATAGCAACCGCTGC
TGGTTACCCTCTCCTCTGGGGTAACCAATTTGAGTTAATAATAAGGATTCTAAGTTGCACTTGAATCTTT
TCTGTCTTCATCTCCACTGCTGCTGTTCGAGTCCAAGTCTACTCTCCCCTCTGAATTCCTGCAACCACCT
CCATCTCCTCCCCTATAGCTGATTCCTGGAACAGACCTGGCCTC

xrRNA

GTACTTCGAAATGTCATCCTCTGTCTGACACTGAACGTAATCCAGACGCGTAAGTCAGGCCGGAAAATTC
CCGCCACCGGAAGTTGAGTAGACGGTGCTGCCTGCGACTCAACCCCAGGAGGACTGGGTGAACAAAGCTG
CGAAGTGATCCATGTAAGCCCTCAGAACCGTCTCGGAAAGAGGACCCCACATGTTGTAGCTTCAAGGCCC
AATGTCAGACCACGCCATGGCGTGCCACTCTGCGGAGAGTGCAGTCTGCGACAGTGCCCCAGGAGGACTG
GGTGAGGATCCTACCTACAAACGGCACGAGCATCAGCC

δRZ[WT]

AGGGCGGCATGGTCCCAGCCTCCTCGCTGGCGCCGCCTGGGCAACATGCTTCGGCATGGCGAATGGGACC
AAA

δRZ[MT]

AGGGCGGCATGGTCCCAGCCTCCTCGCTGGCGCCGCCTGGGCAACATGCTTCGGCATGG<u>T</u>GAATGGGACC
AAA

The inactivating single point mutation is underlined.

MALAT1 3'-end

GGCCATGCAGGCCAATGCTCTTCAGTAGGGTCATGAAGGTTTTTCTTTTCCTGAGAAAACAACACGTATT
GTTTTCTCAGGTTTTGCTTTTTGGCCTTTTTCTAGCTTAAAAAAAAAAAAAAGCAAAAGATGCTGGTGGTT
GGCACTCCTGGTTTCCAGGACGGGGTTCAAATCCCTGCGGCGTCTTTGCTTGGCCCTGAAGGCC

Chapter 3

3. Rapid depletion of XRN2 reveals widespread co-transcriptional degradation of the downstream products of PAS cleavage

Declarations: Steven West (S.W.) generated and validated the XRN2-AID HCT116 cell line.

At the outset of this thesis, one of the first aims was to ascertain the role of XRN2 (if any) in transcriptional termination. A major question was why KD of XRN2 has a little general effect on termination while competitive inhibition of 5'→3' activity does (Nojima et al., 2015; Fong et al., 2015). Also, some *in vitro* studies have cast doubt as to whether exonucleases are capable of terminating Pol II complexes and co-transcriptionally degrade the downstream PAS cleavage products if any (Dengl and Cramer, 2009; Pearson and Moore, 2013; Zhang et al., 2015). Conversely, other *in vitro* studies have observed co-transcriptional exonuclease RNA degradation and dissolution of assembled Pol II complexes (Park et al., 2015). To investigate these questions, a new approach was sought to selectively degrade endogenous XRN2 more rapidly than RNAi. This would test whether a lack of RNA accumulation in KD experiments was due to incomplete depletion and whether the effects of competitive inhibition could be assigned to XRN2 function. If such a system were paired with nascent transcription methods, it could then further dissect any potential involvements in co-transcriptional contexts.

3.1 Using the auxin-inducible degron (AID) system to rapidly deplete XRN2

The auxin-inducible degron (AID) system was chosen to investigate XRN2 function because it offers rapid depletion times (from 1 h upon treatment with auxin) (Nishimura et al., 2009; Natsume et al., 2016). These experimental time frames are shorter than those used for RNAi (~24-72 h) and even some other degron tagging methods such as DHFR (> 6 h when carried out to a

comparable depletion level) and SMASH (~24h) (Sheridan and Bentley, 2016; Chung et al., 2015). The increased speed of depletion reduces the risk of secondary or indirect effects developing and changes to gene expression, such as the upregulation of redundant or compensatory proteins, having consequence. Secondly, the AID system commonly provides a level of depletion that is rarely achievable by RNAi, where the relative abundance of the target protein is reduced to undetectable levels when compared to control samples by western blotting. The human colorectal carcinoma cell line, HCT116, was chosen because it has a near-diploid karyotype, whereas many other immortalised carcinoma cell lines are highly aneuploid (Knutsen et al., 2010). This reduces the risk of multiple copies of the target gene being present, which should facilitate the homozygous tagging of all alleles. Indeed, HCT116 cells have previously been successfully employed with the AID system and have a high rate of recombination that may lend itself to this CRISPR editing strategy (**Fig 3.1a**), which relies on HDR with a co-transfected construct containing homologous regions to each side of the sgRNA target site (Natsume et al., 2016).

It should be noted, there are limitations to using cancer cell lines as models because many studies have highlighted their substantial transcriptomic and genomic differences that fail to accurately recapitulate the disease complexity of the primary tumour biopsy cells from which they once derived (Chen et al., 2015a; Gillet et al., 2011; Liu et al., 2019). However, this thesis does not focus or seek to comment on colorectal carcinoma. Additionally, as the AID system requires multiple rounds of genome editing involving expansion from a single cell after each round and some of the nascent transcriptome techniques planned require large amounts of input (10^7 cells per condition for mNET-seq), cancer cell lines offer a compromise due to their relatively fast growth rate and immortalisation offering longer culture durations.

A C-terminal tagging approach was used for XRN2 because the entire cassette (consisting of AID-P2A-HygroR/NeoR-PAS) is large and it was feared a large first exon may reduce the levels of protein expression or interfere with splicing as many mammalian genes have short first exons (Bieberstein et al., 2012). Also, many of the XRN2 active site residues are near the N-terminus and tagging here may risk adversely affecting its activity (Xiang et al., 2009).

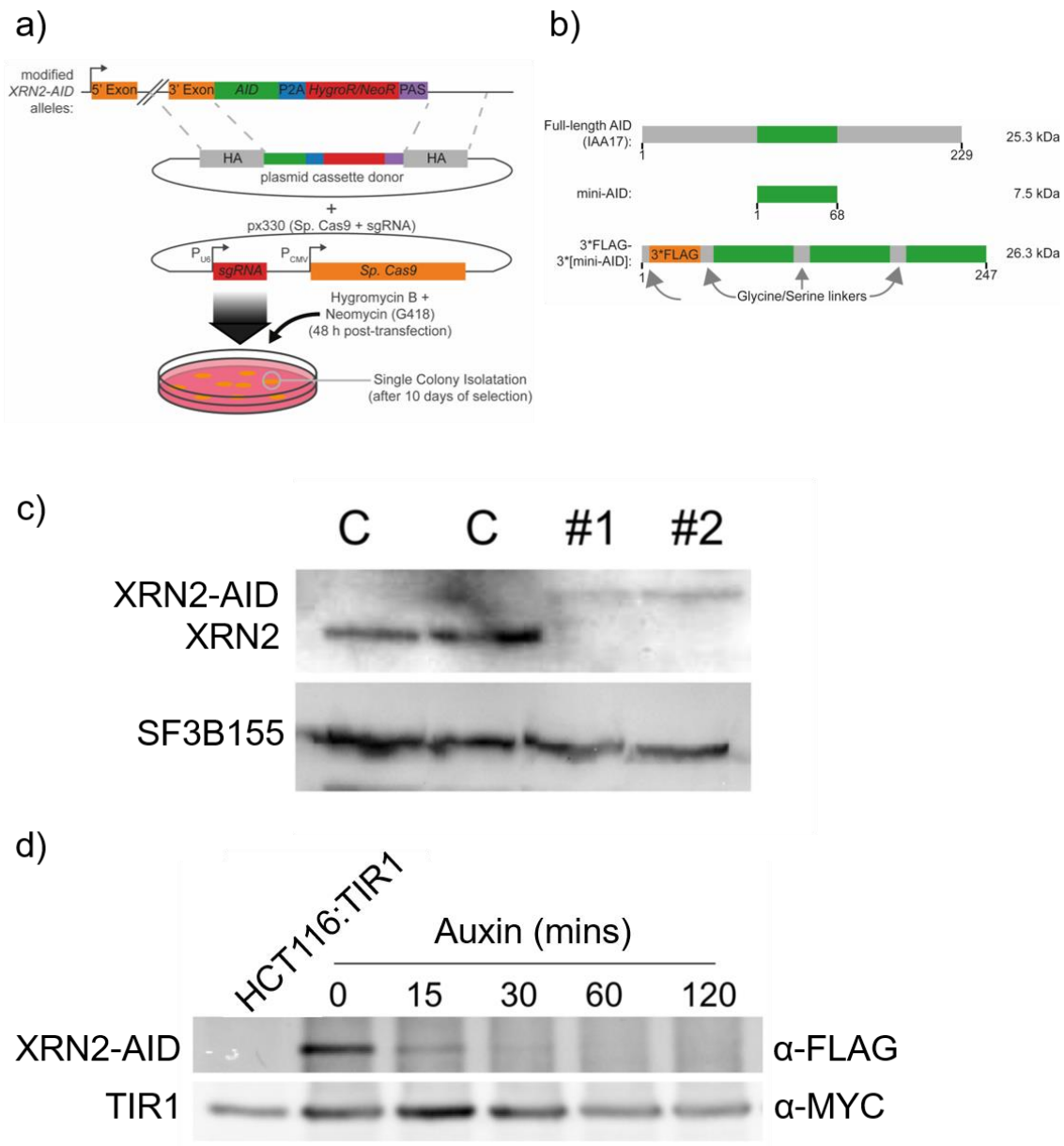


Figure 3.1 | Rapid and complete depletion of XRN2 with an auxin inducible degron (AID) tagged cell line. **a**) Schematic of XRN2-AID HDR plasmid repair constructs which are co-transfected with the CRISPR/Cas9 plasmid (Sp. = humanised S. pyogenes). Both HygroR and NeoR containing plasmids are used to enrich for biallelic insertions. **b)** Domain structure of degrons including the three tandem mini-AID tag used for XRN2-AID. **c**) WB showing a ~27 kDa increase in migration of XRN2 among two isolated drug resistant clones suggesting mini-AID tag incorporation (C = control untagged sample; *#n* = clone *n* that is grown after single cell selection and is resistant to hygromycin and neomycin/G418). SF3B155 is used as a loading control. **d**) WB showing the time-dependent depletion of XRN2-AID by auxin addition. The parental cell line HCT116, modified with ectopically expressed TIR1, was used as a control. TIR1 is used as a loading control. The WBs in **c** and **d** were made by S.W. and reproduced from Eaton et al. (2018).

The AID tag used here contains three tandem Flag epitopes followed by three tandem minimal AID domains (mini-AID) from IAA17 protein of *Arabidopsis thaliana* with each feature joined by glycine serine linkers (**Fig 3.1b**; Kubota et al., 2013). This three tandem mini-AID is suggested to provide more efficient degradation than a single full-length AID because of observations that the growth defects, when tagged to essential proteins in budding yeast, are more severe. The XRN2-3x[mini-AID] (*XRN2-AID* herein) cell line, which was generated and validated by Steven West, had isolated clones screened for genome integration via PCR and then further validated by western blot. These protein extracts were probed an anti-XRN2 antibody and the two modified clones tested both showed an increase in band migration, which was consistent with the 26.3 kDa degron tag size, suggesting the successful homozygous tagging of all alleles (**Fig 3.1c**). However, XRN2-AID band protein abundance is reduced to 10-20 % of endogenous untagged XRN2 in the absence of auxin. Some other groups, using the AID system, have similarly noticed auxin-independent protein depletion for some tagged proteins suggesting such a result is not uncommon (Morawska and Ulrich, 2013; Nishimura and Fukagawa, 2017; Zasadzińska et al., 2018). Regardless of this after 1 h of auxin addition, the cell line still exhibited auxin-induced depletion of XRN2-AID to undetectable levels relative to the already reduced amount within untreated *XRN2-AID* cells (**Fig 3.1d**).

3.2 Using mNET-seq to generate single-nucleotide resolution transcription profiles upon XRN2 loss

To assess the effects of XRN2-AID loss upon transcription mNET-seq was chosen as a genome-wide approach to gain high-resolution mapping of nascent Pol II transcription complexes (Nojima et al., 2015). This method enables the mapping of Pol II at single-nucleotide resolution providing many insights unavailable from other methods. The library protocol involves selectively isolating Pol II-associated RNA and building a short-read library such that the terminal residue of the RNA fragment will correspond to the last ribonucleotide incorporated by Pol II (**Fig 3.2a**). In capturing the 3' terminal nucleotide it provides a stochastic snapshot of Pol II transcription positions. The method involves nuclei and subsequent insoluble chromatin pellet purification from a dish of cells. Then a short period of enzymatic fragmentation

is carried out by micrococcal nuclease (MNase) to solubilise chromatin-associated complexes. Pol II-associated chromatin is immunoprecipitated with a specific antibody before purification of its associated RNA. In this case, an antibody was chosen that is predicted to target all forms of Pol II, recognising the heptad repeat of the CTD. Importantly, the endonucleolytic cleavage of single and double-stranded RNA and DNA by MNase produces 5' hydroxyl and 3' phosphates (Cuatrecasas et al., 1967). Whereas, RNA fragments that have inaccessible 3'-ends from within the active site of Pol II will have the 3' hydroxyl preserved. This allows chemical isolation with a truncated RNA ligase that ligates 5' adenylated adapters to 3' hydroxyl ends of ssRNA in the absence of ATP (Viollet et al., 2011). Thus, only these 3' hydroxyl species will have the 3' adapter ligated, be a substrate for PCR amplification and be able to bind the Illumina sequencing flow cell.

The bioinformatic processing strategy for mNET-seq involves trimming, mapping, generation of single nucleotide resolution (snr) alignments, read normalisation and plotting of mNET-seq profiles (**Fig 3.2b**). To ensure capture of the 3' most nucleotide, which can usually be considered as the Pol II position, paired-end sequencing is carried out on the libraries. This allows the generation of mNET-seq snr profiles by converting concordantly mapped read pairs into a 1bp read with the complementarity of the forward read at the 5' position of the reverse read. Unlike Pol II ChIP, mNET-seq offers superior resolution, only captures Pol II with an associated RNA and, identifies the directionality of transcription events.

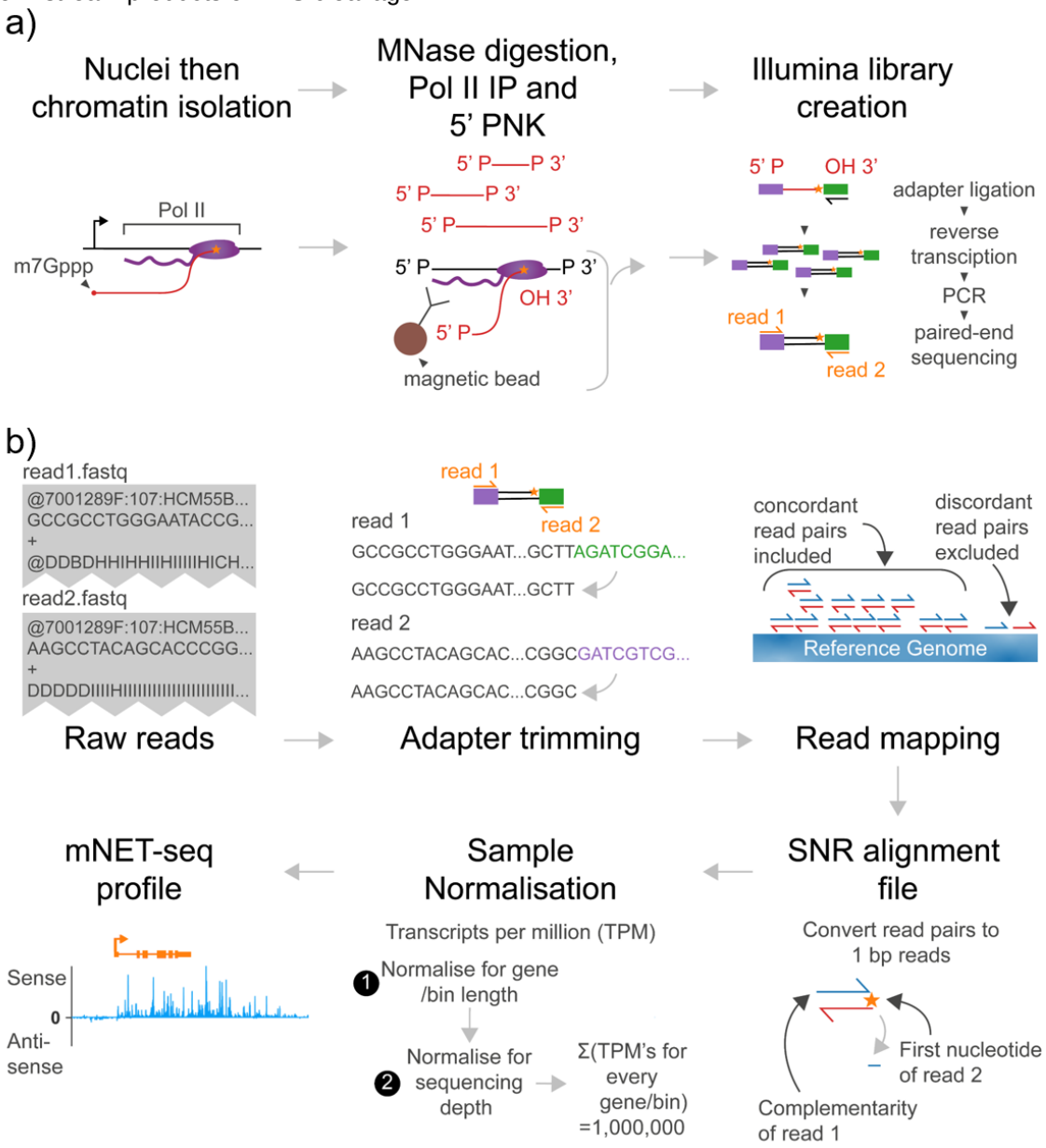


Figure 3.2 | Using mNET-seq to generate signal nucleotide resolution (SNR) transcription profiles for XRN2 depletion. **a**) The mNET-seq library protocol schematic for isolation of nascent RNA from transcribing Pol II complexes and preparation for Illumina sequencing. This used an antibody (MBL CMA601) specific to the total CTD heptad repeats of Pol II regardless of residue phosphorylation status [adapted from Nojima et al., (2015)]. **b**) Bioinformatic processing strategy for trimming, mapping and, normalisation of raw reads into mNET-seq SNR profiles. The orange star indicates the position of the 3' nucleotide captured from within the Pol II active site.

3.3 Quality assessment of mNET-seq sequencing files

In total six samples were sequenced. Two of these were from control cell lines for comparison to the *XRN2-AID* edited cell line, one being the unmodified parental HCT116 cells. The other contained *TIR1*, the required E3 ligase component, integrated within the *AAVS1* safe-harbour locus and under the control of a doxycycline promoter (Natsume et al., 2016). This is unlike the *XRN2-AID* cells, where *TIR1* is integrated at unknown sites by the sleeping beauty transposon system, thus the *TIR1* cells will not be an exact control for the same level of TIR1 expression. However, it still provides some assurance that the presence of TIR1 alone should not directly affect the transcription pattern within HCT116 cells. These mNET-seq samples were then quality assessed before processing. The Phred quality scores of the average base calls for the raw reads were high for all samples and >30 throughout the 50 bp reads captured (**Fig 3.3a**). A Phred score of 30 indicates a bp has a 1 in 1000 probability of having been incorrectly recorded by the sequencing machine. The adapter content of the mNET-seq libraries is seen to increase from ~20bp because these are short fragment libraries that have been sequenced for 50 bp cycles (**Fig 3.3b**). Some of the variability between samples probably arises due to slight differences in gel excision purification of library batches. The composition of the first nucleotide of read 1 and the first nucleotide of read 2 are shown (**Fig 3.3c,d**). The former indicates the nucleotide directly after the MNase cleavage site. MNase is known to have a strong bias for cleavage upstream of an A or T nucleotides (Dingwall et al., 1981; Allan et al., 2012). This enrichment bias is seen in all samples except for *XRN2-AID* replicate 2, which likely indicates a poorer digestion reaction. The nucleotide composition of the first bp of read 2 (the complement of Pol II active site residue) is more evenly distributed with a less visible or consistent bias present. After removal of the adapters, a sample of 100,000 raw reads was screened for the presence of contaminants. For all samples, > 80 % of reads mapped to the human genome (**Fig 3.3e**). Some reads did map to the mouse and CHO cell genomes, but these were classed as reads that matched ‘multiple genomes’. As very few reads are unique to either the mouse or CHO cell genomes, these multiply mapped reads are unlikely to originate from these sources and probably derive from conserved eukaryotic sequences that are also found in humans. This is more likely to occur for mNET-seq libraries given the small fragment sizes used.

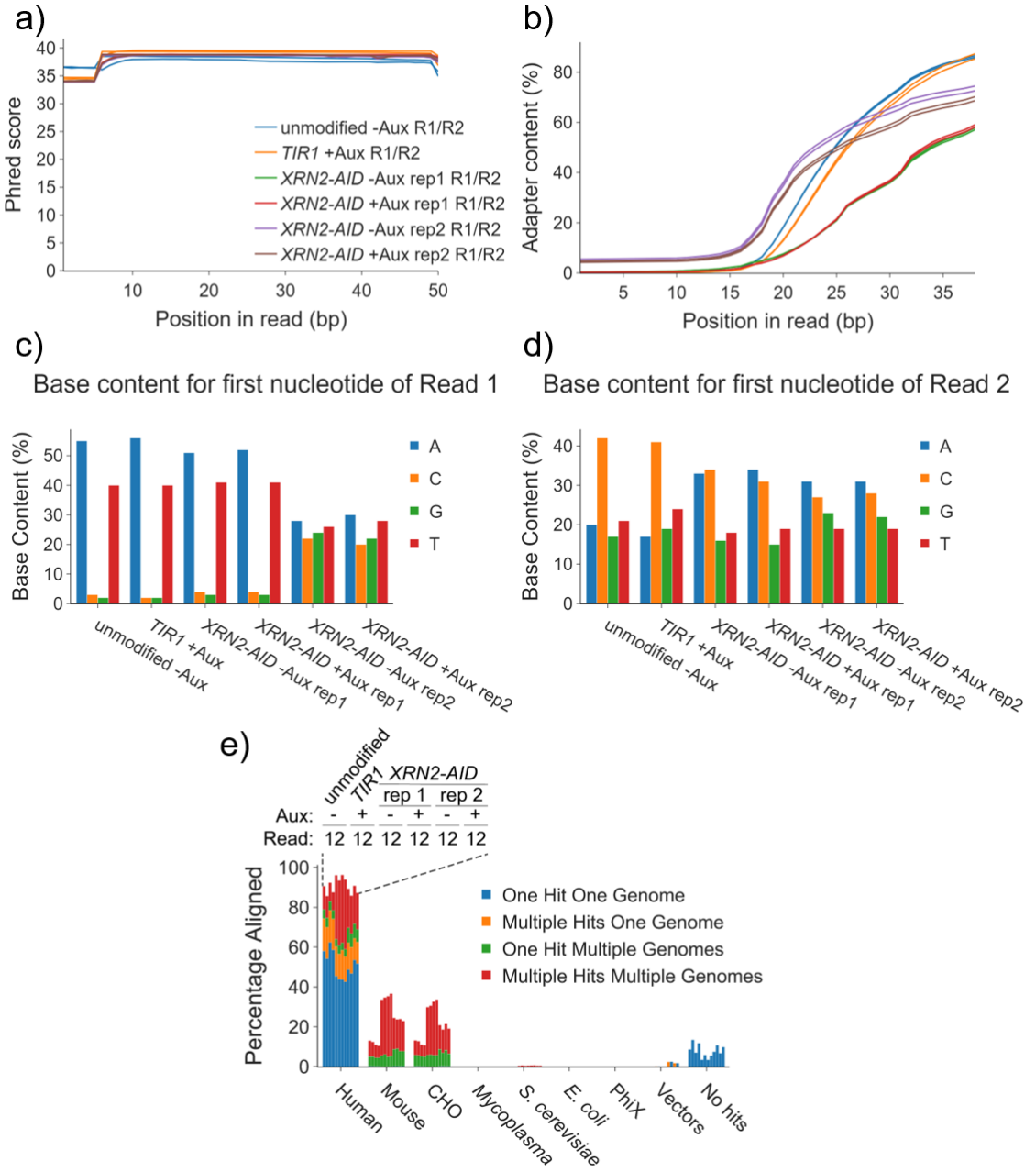


Figure 3.3 | Quality assessment graphs for control and *XRN2-AID* mNET-seq samples. **a)** The average Phread quality score for each nucleotide position within the raw reads. **b)** Adapter content (%) for each base pair within the raw reads. **a,b)** The key applies to both. 'R1' = read 1 and 'R2' = read 2. Data generated from FastQC. **c,d**) The percentage nucleotide composition of the 5' nt of the RNA fragment (first nt of read 1, **c**) and the complementary nt at the 3' end (first nt of read 2, **d**). **e)** Adapter trimmed reads were screened for contamination with 'fastq_screen'. A sample of 100000 reads, taken evenly throughout the read file, were aligned to various datasets of commonly used lab species and reagents. The key for the grouped bars is shown above human. HCT116 cells were either unmodified, expressing *TIR1*, or expressing *TIR1* along with biallelic AID tagging of XRN2 in the presence and absence of Aux (Aux = auxin; nt = nucleotide). The control *TIR1* cell line, which is not a direct parent to the *XRN2-AID* cells, is integrated within the AAVS1 locus under the control of a doxycycline-inducible promoter (dox was added 24 h before harvest).

The mapping of reads and filtering for concordant read pairs was done by HISAT2 and the proportions are shown in **Table 3.1**. After processing, the control mNET-seq conditions were compared to XRN2-AID untreated with auxin to confirm that the transcription profiles and termination regions remain unchanged after the genome-editing process. Of the four protein-coding genes examples, shown in **Fig 3.4**, all profiles largely correlate with the signal reducing to background at similar positions after the 3' annotated ends of the genes. Interestingly, *XRN2-AID* cells (in the absence of auxin) can maintain efficient Pol II transcription termination at the reduced protein abundance mentioned previously. This is also true for XRN2-AID replicate 2 minus auxin, although, the slightly reduced GB signal seen in examples could be reflective of reduced coverage depth caused by the impaired MNase digestion. This should not interfere with the direct comparison of replicate 2 samples however because both treated and untreated samples were prepared at the same time and have the same reduced MNase cleavage signature (**Fig 3.3c**).

3.4 XRN2-AID depletion leads to widespread read-through at protein-coding genes

Next, to investigate the involvement of XRN2 in degrading the downstream product of PAS cleavage and the wider role in PAS-dependent termination, protein-coding genes were examined upon XRN2-AID depletion. Protein-coding genes require a PAS to undergo transcriptional termination so offer a powerful insight into PAS-dependent termination (Whitelaw and Proudfoot, 1986; Connelly and Manley, 1988). It is important to consider the presence of neighbouring TUs close to or overlapping the 3'-ends of protein-coding TUs as reads from these transcription events can be aberrantly considered to have originated from the TUs to which they are closely spaced. Also, the effects of transcriptional interference by Pol II collisions may change the profile and the outcome of where RNA accumulation (if there is any) occurs in relation to the transcript in question. Six examples of protein-coding TUs whose 3'-ends are clear from such neighbouring interference all show RNA accumulation beyond the PAS (**Fig 3.5**). For some TUs (like *MYC* and *TBL1XR1*) the accumulation of RNA occurring beyond the PAS is mainly within the same footprint of where transcription occurs in the absence of auxin (**Fig 3.5a,b**). However, in other cases (like *ACTB*, *RPL30*, *YTHDC2* and *EEF1A1*) the RNA accumulates not

Table 3.1 | Proportion of reads pairs that map concordantly to human genome for control and *XRN2-AID* mNET-seq samples using HISAT2.

Samples	Total reads	Read pairs	Concordantly mapped read pairs	% of read pairs that are concordant
Unmodified –Aux	181465993	78366207	59616293	76.07
***TIR1* +Aux**	167947689	73500152	58133575	82.43
***XRN2-AID* rep1 –Aux**	213073958	72408380	51527384	71.16
***XRN2-AID* rep1 +Aux**	199467883	63867518	45168884	70.72
***XRN2-AID* rep2 –Aux**	127341950	50324673	35756473	71.05
***XRN2-AID* rep2 +Aux**	225398876	90692141	67468275	74.39

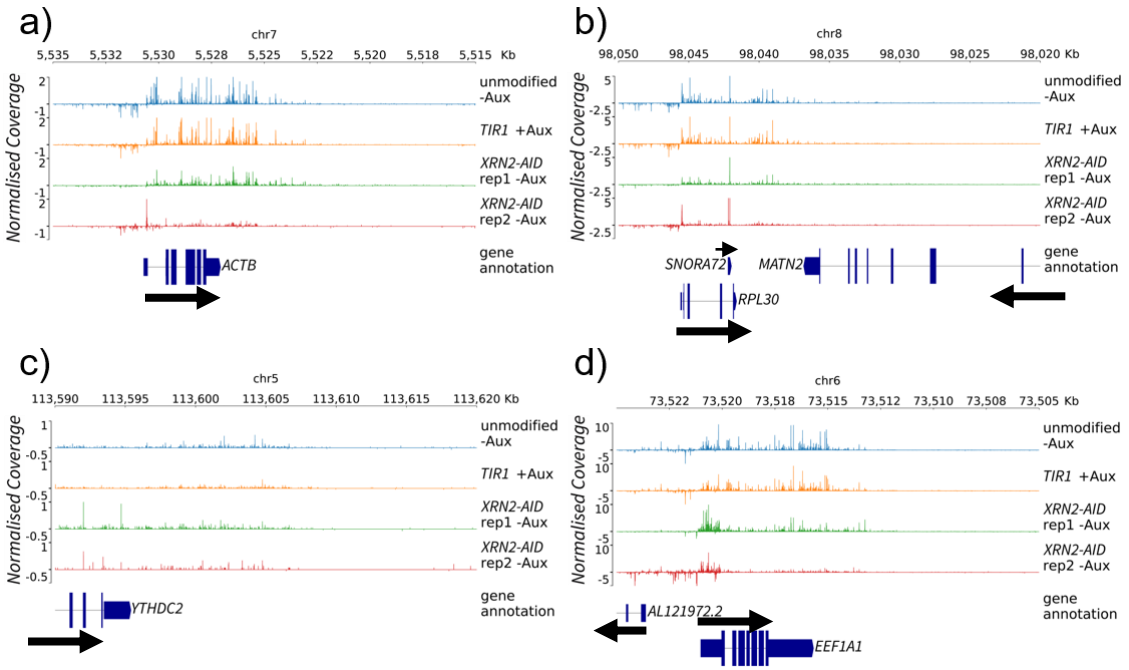


Figure 3.4 | Conserved termination regions between parental and genome edited cell lines under control conditions. Comparing the transcription termination regions, on a handful of protein-coding genes, between the unmodified HCT116 parental cells (blue), *TIR1* cells with auxin treatment (orange) and *XRN2-AID TIR1* edited cells without auxin treatment for two replicates (green and red respectively). The mNET-seq transcription signal reduces to background beyond the annotated genes at similar positions in all four examples and the GB profiles are also similar. The average signal for the second replicate of *XRN2-AID* untreated is consistently lower than the other samples. Positive and negative peaks correspond to sense and antisense transcription, respectively. Within the annotation track coding exons are shown by larger width blue bar with a line corresponding to intronic regions for a primary transcript isoform. Black arrows indicate transcript directionality and normalised coverage for chromosomal snapshots equals single base pair bin size transcripts per million (TPM).

only within the same transcription footprint as untreated cells but extends beyond this point and into downstream intergenic regions (**Fig 3.5c-f**). The accumulation of RNA for *TBL1XR1* is spread over hundreds of kb but itself is over 150 kb long, whereas the signal beyond the *MYC* PAS covers a ~5 kb range. Likewise, read-through transcription varies but less dramatically extending over a ~5 kb region for *ACTB* and ~10 kb for *RPL30*. One possibility is this extended read-through could be revealing positions where Pol II normally terminates. If so, it is unclear why the polymerases within these termination regions would only be visible on some genes and not others. Whilst the distance of read-through transcription and RNA accumulation varies between TUs, the distance for each TU is reproduced between replicate samples indicating context or loci-specific reasoning for this variation. Such conservation of loci-specific (or transcript-specific) termination characteristics is not unexpected because Fong et al. (2015) showed that although 3'-end Pol II pause occurs at different distances after the PAS, the position of the peaks are unchanged in multiple cell lines. Additionally, the longer read-through could be because of differences in transcription speeds with an increase of just 220 nt/min shifting Pol II into downstream sequences (Fong et al., 2015). One similarity all examples share is that the RNA signal within auxin conditions always reduces towards the background levels of the control sample gradually (when no expressed neighbouring TU's occur downstream). This suggests transcription still ceases and does so stochastically with XRN2 accelerating termination. Perhaps such punctuality is desirable to avoid transcriptional interference of neighbouring TUs or ensure efficient Pol II recycling.

When exploring these findings more generally a metagene profile, which averages many overlaid protein-coding genes, confirms the widespread nature of XRN2 degradation of PAS cleavage products (**Fig 3.6a**). To avoid overlapping signal from neighbouring transcription interfering with the interpretation, isolated protein-coding transcripts were examined. First 1501 TUs were identified whose primary transcript isoform had no neighbouring transcript annotation within -5kb to +20kb up and downstream, respectively, regardless of whether such neighbouring was expressed under untreated conditions. Then once ranked for expression the top 40 % (n=600) were used within the metagene profile. Like most gene profile examples shown above, RNA accumulates further downstream in the auxin treated sample with a

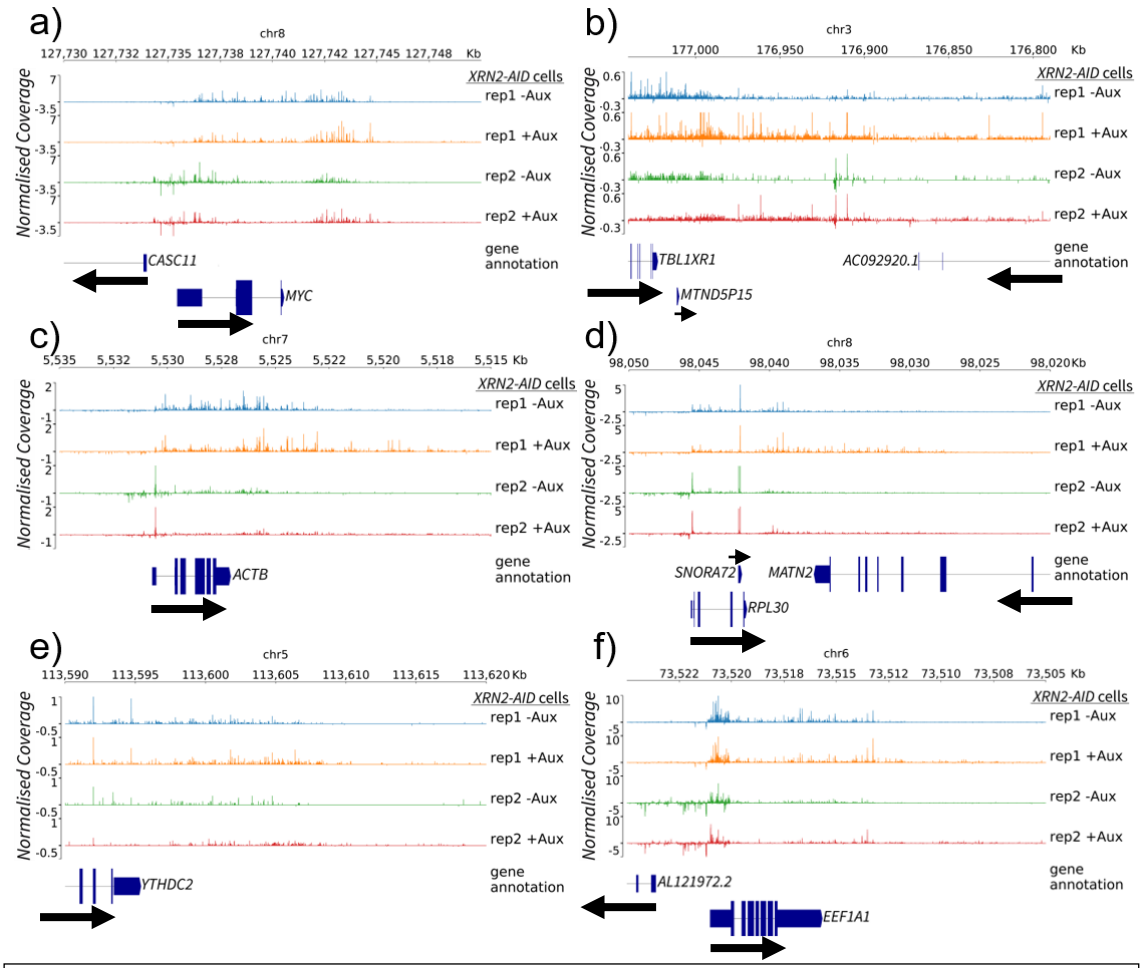


Figure 3.5 | *XRN2-AID* loss causes read-through of various magnitudes and lengths at protein-coding genes. Six chromosomal snapshots of protein-coding genes are shown (**a-f**) and represent common transcriptional termination defects upon XRN2-AID depletion. The accumulation of transcription varies, being present over the same termination regions in -Aux condition in a 1kb window (e.g. *MYC* in **a**) or spread over hundreds of kb (e.g. *TBL1XR1* in **b**). In examples **c-f** XRN2-AID depletion not only causes increased transcription at the 3'-end but the signal extends further downstream to regions not previously transcriptionally active. Black arrows indicate transcript directionality and normalised coverage for chromosomal snapshots equals single base pair bin size TPM.

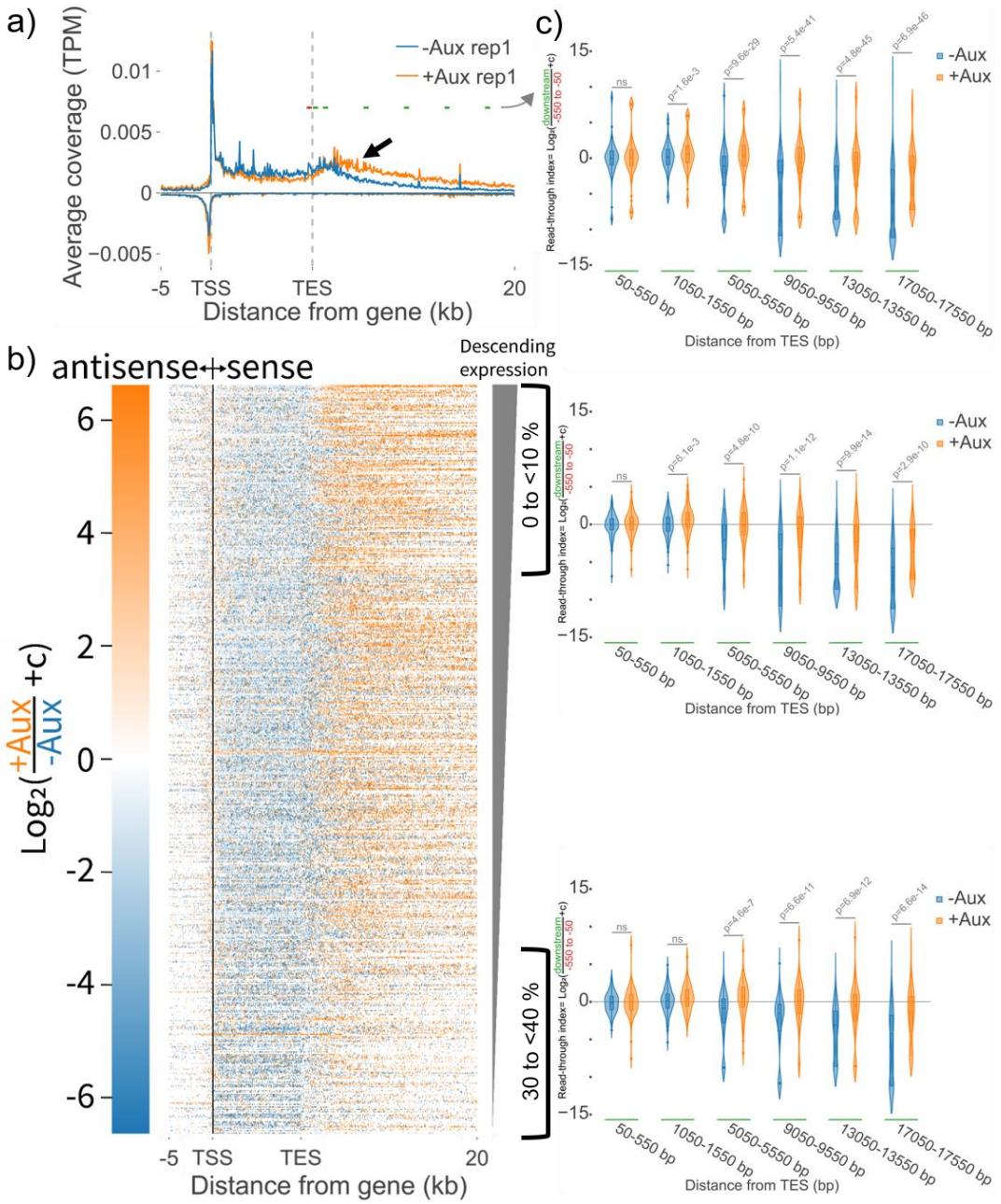


Figure 3.6 | Genome-wide effects of *XRN2-AID* depletion on protein-coding genes. Isolated protein-coding primary isoform transcripts where selected whose annotation has no neighbouring annotation between -5kb and 20kb up- and downstream respectively (see methods). Of these 1501 protein-coding transcripts, the top 40 % expressed examples (n=600) were taken forward for metagene analysis. The GB region is scaled to 10 kb and a 50 bp bin size is used. **a**) Metagene profile on the 40 % highest expressed protein-coding transcripts. Sense and antisense coverage are displayed as positive and negative values. Black bars correspond to read-through index (RTI) regions used in **c**. **b**) Heatmap showing the log_2 fold change between normalised TPM coverage of (depleted / untreated) XRN2-AID replicate 1 single protein-coding transcripts used within metagene **a**. The log_2 fold change left and right of the TSS is calculated using antisense and sense coverage, respectively. **c**) Violin-plots for RTI of 600 protein-coding genes. p-values were calculated by a two-sided Mann-Whitney test and adjusted using the Holm method. The two violin plots adjacent to the heatmap use a subset of 150 genes.

delayed hump visible after the PAS (see arrow). This transcription then extends into the 3' intergenic flanking regions but tends toward background levels by 20kb beyond the PAS. Whereas at the promoter, there is a far less pronounced effect with the promoter peak for both the sense and antisense (PROMPTs) peaks largely matching but with a small increase upon XRN2-AID depletion. For PROMPTs transcripts more generally, there is no appreciable difference with the transcription profile between the two conditions. This reaffirms the PROMPTs shown in the individual gene examples above (**Fig 3.5 a,c,d,f**; see asterisks). These observations are not biased by a few strongly expressed examples as a heatmap showing the log_2 fold change between auxin treated and untreated conditions confirms the widespread read-through and lack of antisense PROMPT effect (**Fig 3.5b**). Another observation is that the GB signal is slightly lower (light blue within the heatmap) when XRN2-AID is depleted. This only appears as a minor reduction within the individual gene snapshots and the metagene profile. Whilst the heatmap demonstrates the widespread occurrence of the RNA accumulation downstream of the PAS it does not compare this to the relative level of expression within the upstream transcript. For example, the RNA accumulation at 15-20kb beyond the annotated end of the transcript or transcription end site (TES) appears more pronounced for the top 10% of genes than the less well-expressed 30-40% but, this may be because the former's signal started at greater level above background transcriptional noise. To address this a read-through index (RTI) was calculated as a ratio of average signal downstream of the TES to average signal -550 to -50 bp upstream of the TES (adapted from Nojima et al., 2015). The RTI is shown for increasing distances downstream of the transcript for the top 40 %, top 10% and 30-40% (**Fig 3.6c**). Interestingly, both subsets (0-10% and 30-40%) have similar violin shapes and heights at ~13 and 17kb upon auxin treatment. Therefore, this suggests the average read-through distance (relative to GB signal) does not depend on the level of gene expression.

The proportion of protein-coding genes that demonstrate an XRN2-dependent termination defect is often queried. Whilst this is easier to determine for protein-coding genes with no neighbouring TUs downstream it is somewhat more complicated where TUs overlap, occur in tightly packed clusters or within regions that are pervasively transcribed in both directions under untreated

conditions. The graphs presented above show that XRN2 is ubiquitously employed in the transcriptional termination of isolated protein-coding genes. Therefore, the likely interpretation is that XRN2 also targets other protein-coding genes in more crowded settings, but the effects may be less visible because of multiple transcription events from neighbouring TUs. However, at present it cannot be ruled out that some of the protein-coding genes within these confounded loci may not involve XRN2 within transcriptional termination.

3.5 Investigating XRN2-AID depletion at smaller gene classes

Although replication-dependent histones are not polyadenylated they do share some of the same 3'-end processing components with other protein-coding genes including CPSF73, CPSF100 and Symplekin (Sullivan et al., 2009). Therefore, CPSF73 cleavage has the potential to provide a similar entry site for XRN2 at the 3'-ends of histone pre-mRNA as it does at the PAS of protein-coding genes. Replication-dependent histones, as their name suggests, are expressed during the S-phase of the cell cycle as DNA is replicated, but due to the high level of histone transcription during this period, nascent RNA is easily detectable from asynchronous cells, even though such signal originates from a proportion of cells. As expected, histone gene loci are abundantly detected in the mNET-seq data (**Fig 3.7**). On examination for an XRN2-dependent termination defect, little to no read-through transcription was visible. At the histone cluster on chromosome 6, an overview reveals tightly segmented TUs that appear unchanged across conditions (**Fig 3.7a**). Zoomed in snapshots of two sections of this cluster shows the associated-Pol II RNA signal remains unchanged across the entire TU to the point of termination upon XRN2-AID loss (**Fig 3.7b,c**). Similarly, little to no differences between untreated and treated samples of the same replicate were detected at TUs of a second histone cluster with closely matching profiles of RNA signal recovered throughout the gene bodies (**Fig 3.7d,e**). This indicates that not all CPSF73 cleavage events lead to XRN2 degradation of 3' flanking RNA and Poll II termination. However, a small termination delay has been seen on histone genes by Pol II ChIP upon dominant-negative overexpression of XRN2-MT and additional RNAi of endogenous XRN2 (Fong et al., 2015). This difference could in part be because overexpression of XRN2-MT has the potential to inhibit other 5' to 3' exonuclease. Thus, an alternative exonuclease could be involved in degrading

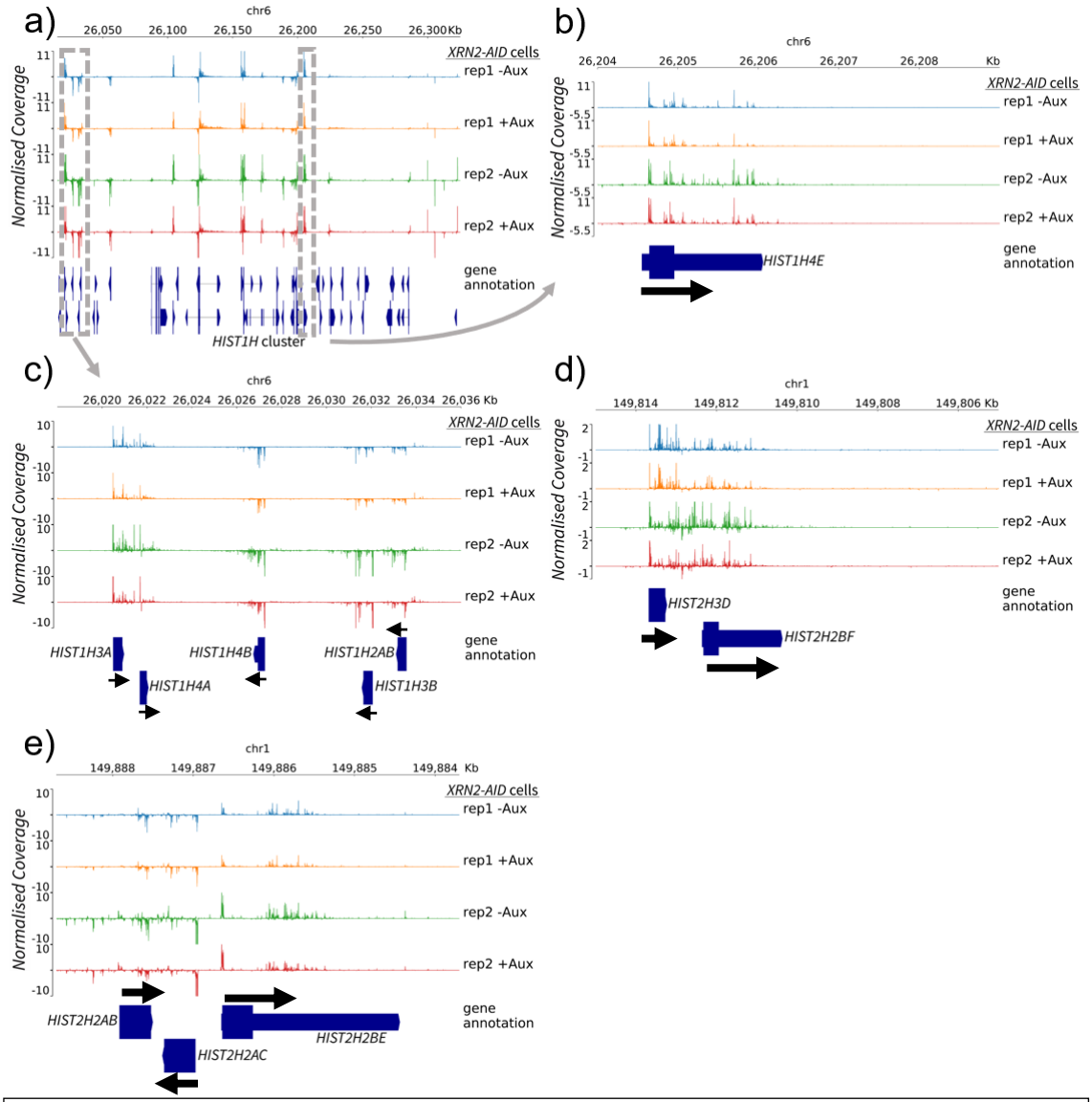


Figure 3.7 | Replication-dependent histone genes have little/no XRN2-dependence on termination. **a)** Overview of the HIST1H cluster of replication-dependent histone genes. Grey boxes and arrows indicate the zoomed-in view shown in neighbour parts. **b,c)** Zoomed in regions of HIST1H genes in a showing in better clarity the lack of read-through and termination defect. d,c) Chromosomal snapshot of histone genes but from the HIST2H cluster. Black arrows indicate transcript directionality and normalised coverage for chromosomal snapshots equals single base pair bin size TPM.

3’ flanking RNA of histone pre-mRNA.

Another class of TU that also undergo 3’-end cleavage are snRNAs, but this is carried out by the integrator complex (Baillat et al., 2005). The nuclease component of the integrator complex, INTS11, is also a metallo-β lactamase family member like CPSF73. These short ~200 bp snRNAs associate with proteins to form snRNPs of the spliceosome complex and function to remove intronic sequences by splicing for pre-mRNA. Within the XRN2-AID mNET-seq, abundant peaks can be seen aligning to snRNA genes (**Fig 3.8a,b**). The mNET-seq protocol, which isolates Pol II complexes via IP, is known to co-precipitate co-transcriptional complexes including the spliceosome, of which the spliceosome is enriched with the Ser5P CTD isoform of Pol II (Nojima et al., 2018b). This co-precipitation of spliceosomal snRNPs which contain mature snRNA is likely to be the source of some of these abundant peaks (**Fig 3.8e**) and similar observations have been seen by other methods that detect co-precipitated 3’ hydroxyl RNAs (Churchman and Weissman, 2011). When scaled-up these snRNA loci reveal the nascent transcriptional signal derived from Pol II and XRN2-AID depletion has little to no effect (**Fig 3.8c,d**). This is reproduced in a metagene profile of snRNAs, although, because far fewer snRNA TU are present within the genome (*vs* protein-coding) only 83 expressed examples were averaged (**Fig 3.8f**). Like for histones, this is different from the small delay in termination observed by Pol II ChIP upon overexpression of XRN2-MT (Fong et al., 2015).

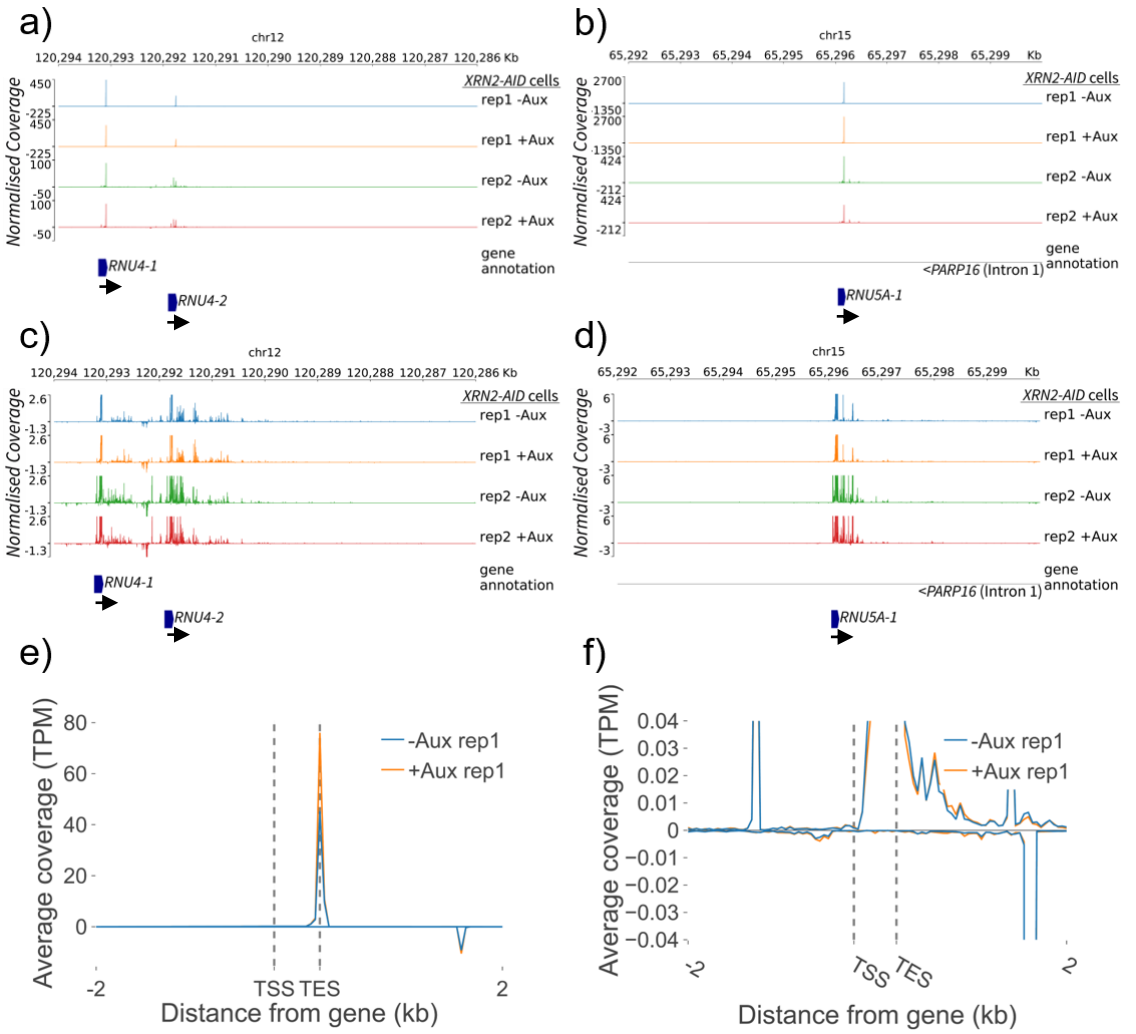


Figure 3.8 | Lack of XRN2-dependent termination defects for other small non-coding TUs. a,b) Examples of snRNA genes showing the enriched co-precipitated cleavage products and mature snRNAs. **c,d**) A scaled in view of the snRNAs (in **a,b**) to reveal the nascent transcription signal that occurs well below the abundance of co-precipitated product. **e,f**) A metagene of snRNAs (n=83) showing the average coverage +/- 2kb up and downstream. The transcript body is scaled to 500 bp and a 50 bp bin size is used for averaging. A scaled-up version is shown on the right to reveal the transcription signal. Black arrows indicate transcript directionality and normalised coverage for chromosomal snapshots equals single base pair bin size TPM.

3.6 XRN2 degrades downstream products at some non-PAS cleavage sites

In addition to those by CPSF73 and integrator, other co-transcriptional cleavage events occur at the 3'-ends of some transcript classes. A small group of non-coding RNAs, which host miRNAs, depend on microprocessor cleavage for termination (Dhir et al., 2015). For this miRNA host gene subclass, RNAi of Drosha or DGCR8 leads to a prolonged read-through transcription. On examining whether XRN2 was involved in degrading the products of downstream microprocessor cleavage similar large peaks, which are likely co-precipitated cleavage intermediates, were seen across the miRNA annotated sites like for snRNAs (**Fig 3.9a,c**). Beyond the terminal annotated miRNA, read-through transcription was extended upon depletion of XRN2-AID (**Fig 3.9b**). The associated Pol II signal extends ~20 kb downstream before returning to background. Moreover, given the siDROSHA and siDGCR8 RNAi by Dhir et al. (2015) caused read-through > 20 kb suggests these TUs do not efficiently use a PAS in the absence of microprocessor cleavage. However, as the RNAi experiments used Hela cells it is an assumption that this also applies within our HCT116 cells. Thus, two non-mutually exclusive possibilities exist firstly that XRN2 is degrading the cleavage products of microprocessor to accelerate Pol II termination and/or secondly the read-through signal derives from downstream PAS cleavage. Recent POINT-5 seq has identified XRN2-sensitive PAS cleavage sites but the extent to which termination of *MIR17HG* depends on PAS cleavage within HCT116 cells is unknown (Sousa-Luis et al., 2020). Another example of non-coding TUs with a different 3'-end cleavage mechanism are *MALAT1* and *NEAT1* (Wilusz et al., 2008). These lncRNA transcripts have a *MALAT1*-associated small cytoplasmic RNA (mascRNA) at their 3'-end, which is small 61 nt tRNA-like structured sequence. Two endogenous RNases, P and Z, excise the mascRNA by cleaving up and downstream, respectively. Upon XRN2-AID depletion at these genes transcription read-through extends beyond the ends of the gene (**Fig 3.9d,e**). The data here supports the view of XRN2 co-transcriptionally degrading the downstream cleavage products from multiple mechanisms other than just PAS cleavage (Fong et al., 2015).

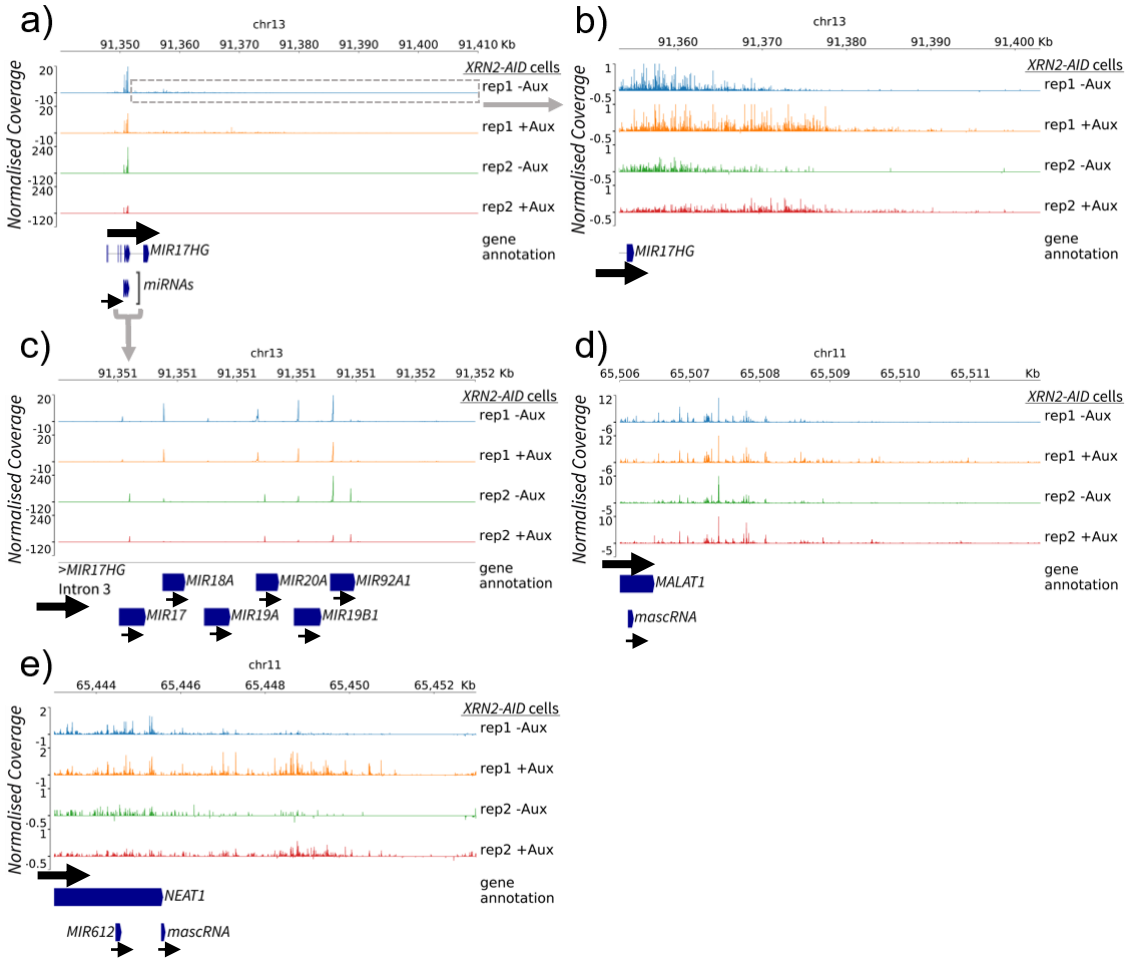


Figure 3.9 | XRN2-AID accelerates termination at some non-PAS cleavage sites. a) A chromosomal snapshot of *MIR17HG*, a miRNA host gene. **b**) The termination region after the annotated microRNAs (in **a**) reveals an XRN2-AID read-through defect where microprocessor cleavage likely provides XRN2 with the entry site. **c**) A zoomed-in view of the six microRNAs found within intron 3 of *MIR17HG* shows the enrichment by coprecipitation of cleavage intermediates and products. **d,e**) The termination regions of two lincRNA genes that have a unique mascRNA structured tRNA-like element at the 3' ends. This mascRNA, which is cleaved by RNAase P and Z, likely provides the entry site for XRN2 which reveals a read-through defect upon XRN2-AID depletion. Black arrows indicate transcript directionality and normalised coverage for chromosomal snapshots equals

3.7 Discussion

The data presented here for XRN2-AID depletion reveals the widespread role for 5'→3' degradation of downstream cleavage products of PAS cleavage and some other cleavage events as suggested by the torpedo model. As mNET-seq records Pol II-associated RNA, it strongly indicates that such degradation occurs co-transcriptionally and takes part within a primary termination mechanism. The reduced levels of XRN2-AID in the absence of auxin, when compared to endogenous levels of XRN2, is a caveat of the XRN2-AID cell line. However, transcription termination still occurs in similar positions downstream of the PAS when compared to untagged parental cells and read-through transcription depends upon auxin addition to modified cells (**Fig 3.4,3.5**). In hindsight, the unexpected finding that such a small amount of XRN2 is sufficient for normal co-transcriptional degradation and termination likely means remaining amounts of XRN2 after RNAi results in the minimal effect of XRN2 downstream of the PAS-dependent transcripts (Nojima et al., 2015). This could be due to (or at least exacerbated by) the processive nature of XRN2 catalysis where once XRN2 begins degradation of an exposed 5' RNA it continues to sequentially degrade nucleotides without dissociation (Lasater and Eichler, 1987). In a broader context, these experiments generalise the human findings of XRN2 which was originally described with single-gene studies on transfected plasmids (West et al., 2004). Additionally, this work largely recapitulates the findings from dominant-negative studies and confirms that XRN2 is the responsible nuclease, which has been blocked by competitive inhibition of 5'→3' activity, at PAS-dependent transcripts (Fong et al., 2015). In yeast, the XRN2 orthologue Rat1 is also widely responsible for degradation downstream of the PAS and Pol II termination (Kim et al., 2004; Baejen et al., 2017). This demonstrates a conserved role for a 5'→3' exonuclease activity co-transcriptionally pursuing Pol II across eukaryotes.

Even though low levels of XRN2-AID are sufficient for normal termination and life of the cell, why it is at such a low level is of interest. As stated previously, such observations have been noted by other labs. One possibility is that TIR1 may have some recognition of degrons in the absence of auxin. Recently, newer adaptations of the AID system for use in eukaryotic/mammalian models have sought to address auxin-independent

degradation by adding in an additional component of the plant pathway that binds to the TIR1 E3 ubiquitin ligase in the absence of auxin (Li et al., 2019; Sathyan et al., 2019). Both groups found a reduction in auxin-independent degradation of tagged proteins using this modified approach. Future experiments to incorporate one of these components into *XRN2-AID* cells may resolve this issue and potentially improve the rate of degradation. Another development uses Auxinole, a TIR1 inhibitor, which binds to the auxin binding pocket and blocks TIR1 association with AID tags (Yesbolatova et al., 2019). Such a method may prevent basal degradation of AID tagged proteins in the absence of auxin and rescue XRN2-AID protein levels but will compete with auxin for binding so removal and higher auxin concentrations may be necessary to avoid a reduced degradation rate. Another possibility is that the AID tag itself may have a destabilising effect on the attached protein. In this scenario, the modifications developed above may not resolve the reduced levels of XRN2-AID as they improve TIR1 auxin-dependence into a more 'switch-like' behaviour by targeting TIR1 rather than the AID tag.

The variety of read-through lengths and magnitudes of RNA accumulation observed at PAS-dependent transcripts but with reproducibility across replicates may indicate a transcript/loci-specific context to this variability. A broad termination motif that causes Pol II slowing (at somewhat degenerate sequences) downstream of C/G stretches and shortly followed by A/T stretches may explain such read-through variability (Schwalb et al., 2016). Such a motif would result in a stretch of dA:rU or dT:rA duplexes within the polymerase and constructs encoding dA:rU stretches placed 600bp downstream of a PAS confirmed a reduction in Pol II. Such dA:rU stretches have a very low thermodynamic stability and may be responsible for limiting ultimate read-through distance upon XRN2-AID depletion by facilitating termination (Martin et al., 1980). The ability for poly(T) tracts encoding dA:rU duplex to facilitate in termination has precedent among other RNA polymerases e.g. Pol III (Nielsen et al., 2013). Another reasoning behind various transcript-specific read-through distances could be Pol II elongation speed. It has been shown that when recognition of the PAS is inhibited, by expression of the influenza A protein NS1a, Pol II continues to transcribe hundreds of kilobases downstream of protein-coding genes and in so doing changes the 3D genome architecture

(Bauer et al., 2018; Heinz et al., 2018). Notably, elongating Pol II has the potential to transcribe through intergenic heterochromatin sequences and displace cohesin-mediated CTCF loops leaving them in its wake. Therefore, it is less likely such heterochromatin protein bound complexes would affect read-through distances but, instead the speed of Pol II elongation could influence the distance past the PAS Pol II reaches before recognition, cleavage and pausing occurs. Likewise, under such a scenario the strength of the PAS may play a role in the 'timing' of cleavage by changes in the affinity particular sequences have with different CPA complex members, which on average may lead to a delay and more transcription past the PAS before recognition, cleavage and Pol II pausing occurs.

As stated above, the lack of XRN2 effect seen upon auxin treatment at histone and snRNAs contrasts with the albeit small effects by Pol II ChIP upon dominant-negative treatment (Fong et al., 2015). One potential explanation is that XRN2-MT competitively inhibits all 5'→3' activity including a redundant exonuclease. For instance, at histones CPSF73 is thought to have both an endonuclease and 5'→3' exonuclease activities and XRN2-MT overexpression could interfere with the latter activity (Yang et al., 2020). However, at snRNAs a 5'→3' exonuclease activity has not been implicated on the downstream products of Integrator cleavage and Integrator uncleaved snRNA precursors are substrates of DIS3 3'→5' degradation (Davidson et al., 2020). Therefore, if the subtle differences at snRNAs between Pol II ChIP upon XRN2-MT overexpression and mNET-seq upon XRN2-AID depletion is due to a different 5'→3' nuclease activity it could only be active between the integrator cleavage site and intrinsic termination site and have competed with DIS3 degradation. A salient finding from the XRN2-AID mNET-seq data is that not all RNA cleavage events are sensitive to XRN2 degradation (i.e. CPSF73 at histones transcripts and Integrator at snRNA transcripts). The reasons behind this remain unclear but one possible explanation is regulation of XRN2 activity or recruitment to the locus. A Thr439-P phosphorylation site on XRN2, which a substrate of CDK9, is thought to promote XRN2 activity (Sanso et al., 2016). Interestingly, while CDK9 inhibition by DRB does not prevent Pol II termination of snRNAs it does cause accumulation of RNA downstream of the Integrator processing site (Medlin et al., 2003).

Lastly, XRN2-AID depletion always results in the inevitable cessation of Pol II transcription, albeit at more distal locations on some transcripts. Why the read-through transcription is only "partial" and does not result in profound transcription beyond protein-coding transcripts is of focus in the next chapter. Several explanations exist that could explain this including the presence of an auxiliary (or fail-safe) termination pathway, incomplete XRN2-AID depletion leaves trace amounts, a redundant 5'→3' exonuclease complements XRN2 degradation or pausing prevents Pol II elongation. Given that cessation of Pol II transcription still occurs upon XRN2-AID depletion and the number of clear examples of transcriptional interference are limited, which is mainly because TUs would need to be very close (< 10kb the approximate read-through distance), it is curious as to why XRN2 is essential across almost all human tissues and cell types, as determined by multiple KO screens (Lenoir et al., 2018). Whilst Pol II recycling may underlie this, the additional roles XRN2 has in maturation of rRNA or nuclear RNA turnover (and ribonucleotide salvage) could also be pertinent reasons for its essentiality.

Chapter 4

4. Molecular dissections of the 5'→3' torpedo activity downstream of PAS cleavage

Declarations: The XRN2-AID XRN2-MT(D235A) HCT116 cells were generated and validated by S.W. The DIS3-AID cells were generated and validated by Laura Francis (L.F.). For the DXO-KO and DXO-KO XRN2-AID cells S.W. performed the initial transfections and ouabain selection but validation and sequencing of indels was performed by J.E. All other ouabain derived cells (δRZ and xrRNA) were generated and validated by J.E.

Within chapter 3, the results showed that depletion of XRN2 causes a downstream shift in the position where Pol II termination normally occurs on protein-coding transcripts. Specifically, this is the accumulation of Pol II-associated RNA within the flanking sequences downstream of these transcripts as detected by mNET-seq. However, the accumulation of RNA does not appear to propagate progressively downstream with the length and magnitude of read-through transcription varying in severity for different transcripts. Therefore, as transcription termination only appears delayed several possibilities arise that might explain why. Firstly, is the depletion of XRN2-AID by the AID system only partial with some remaining trace amounts responsible for the finite read-through generated. Secondly, are redundant 5'→3' exonucleases supplementing the XRN2 torpedo activity. Finally, is a torpedo- or exonuclease-independent mechanism responsible for the lack of progressive read-through. This chapter aims to investigate these questions with a focus on the 5'→3' torpedo activity downstream of PAS cleavage sites. The techniques employed here include genetic editing by ectopic insertion downstream of the protein-coding transcripts with viral sequences, which occlude exonuclease activity by cleavage or RNA secondary structure mechanisms, and indel-inducing KO of a known nuclear 5'→3' exonuclease, DXO.

4.1 Strategies to nullify any potential trace XRN2 remaining after AID depletion

At protein-coding transcripts, given that RNAi of XRN2 leads to no widespread read-through, joint RNAi and dominant-negative overexpression of XRN2-MT causes short extensions of Pol II downstream of transcripts and XRN2-AID depletion causes a more distal yet still finite read-through. This suggests the possibility that trace XRN2 could still remain after auxin-depletion. One strategy to further deplete any possible remaining XRN2 is to initially use RNAi followed up with auxin-induced depletion within *XRN2-AID* cells. As RNAi and AID depletion are alternative mechanisms leading to protein depletion (mRNA degradation and protein degradation respectively) the combined effort could result in a greater depletion level than auxin depletion alone. After 48 h of RNAi, *XRN2-AID* cells were treated with and without auxin for 1 h before relative RNA fold change for amplicons were examined by qRT-PCR (**Fig 4.1a**). Whilst the mean fold change is higher for control *vs.* the joint-treated (siXRN2 +Aux) samples compared to the RNAi (siXRN2 -Aux) or auxin (siCtrl +Aux) treated alone it was not significantly different (two-tailed student's t-test, $p>0.05$) at 2kb and 6kb downstream of *RPL30*, a protein-coding transcript. Unusually, the RNAi alone condition causes RNA read-through (on *RPL30* at 2kb and 6kb downstream of the PAS) that is approximately equal to the auxin treated condition (orange *vs.* green bars). The disparity with the previously described lack of widespread XRN2 read-through by RNAi is probably explained by *XRN2-AID* cells having an already reduced level of XRN2-AID compared to unmodified HCT116 parental cells (**see Fig 3.1c**); this is roughly 20 % of wild-type levels. Additionally, auxin treatment was only performed for 1 h rather than lasting for 2 h with the mNET-seq. This may explain the smaller fold changes observed for auxin only treatment seen here (compare blue *vs.* orange bars to mNET-seq in **Fig 3.5d**).

As an alternative strategy to exclude trace XRN2 (if any exists), the 5'-end of the downstream RNA product of PAS cleavage was occluded. This will not only have the effect of inhibiting trace XRN2-AID but also inhibit any other potential 5'→3' exonuclease that may possess some redundant activity. To achieve this *XRN2-AID* cells were further modified by insertion of a constitutively expressed catalytically inactive *XRN2-MT* using a second sleeping

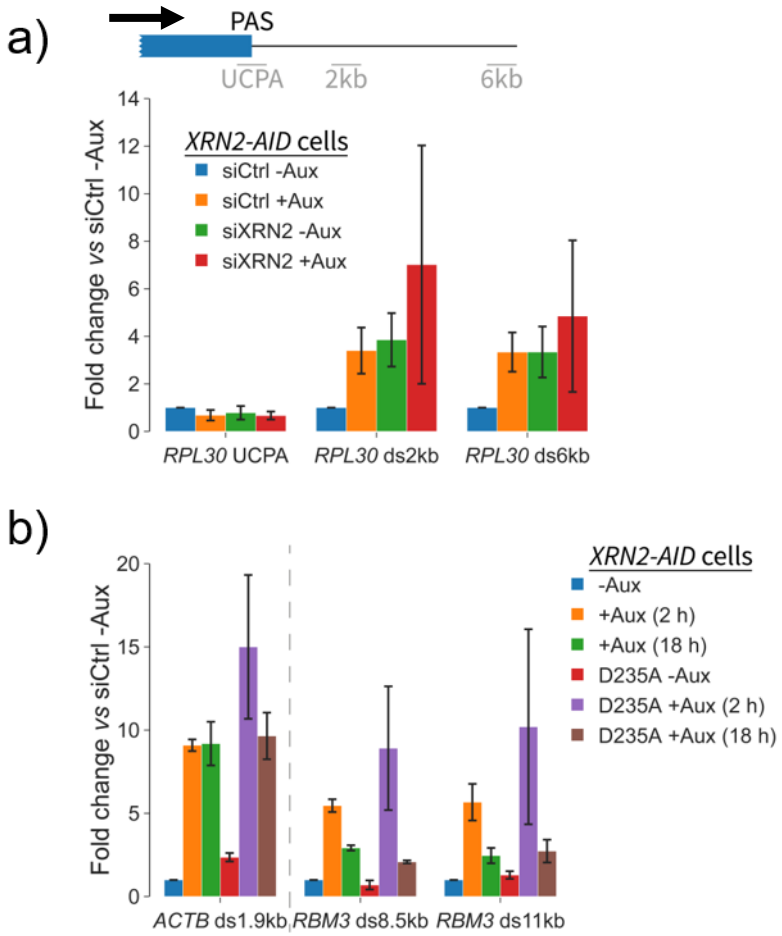


Figure 4.1 | Investigating trace XRN2-AID that remains after auxin treatment (if it exists) to exclude it as the cause of finite transcription read-through. a) qRT-PCR of total cell RNA from XRN2-AID cells transfected with control or XRN2 targeting siRNAs (48 h) followed by treatment with auxin or not (1 h). Error bars are s.e.m and n=3. **b**) qRT-PCR of total cell RNA from XRN2-AID cells with and without constitutively expressed catalytically inactive XRN2 (D235A) and treated with auxin (2 or 18 h) or ethanol mock (18 h). n = 2 and error bars show data point range. Black arrows indicate transcript directionality.

beauty transposon cassette that differs from the *TIR1* cassette by containing a puromycin resistance marker instead of blasticidin (Kowarz et al., 2015). These *XRN2-AID XRN2-MT* cells (herein called *D235A* cells) have the same catalytically inactivating amino acid substitution, D235A, used by Fong et al. (2015) that is predicted to preserve the PO_4 RNA binding pocket (Jinek et al., 2011). The relative fold changes of flanking RNA downstream of *ACTB* and *RBM3* transcripts was determined by qRT-PCR from *XRN2-AID* and *D235A* cells. Cells were either untreated, treated with auxin for 2 h or treated for 18 h (**Fig 4.1b**). In the *D235A* cells, the prolonged depletion for 18 h after auxin treatment reduced the RNA accumulation compared to after 2 h (red & purple *vs.* red & brown). Whereas, in *XRN2-AID* cells the prolonged depletion (18 h) differed little from a 2 h depletion at the *ACTB* ds1.9kb amplicon, but did reduce the increase in flanking RNA at RBM3 amplicons ds8.5kb and ds11kb compared to a shorter auxin depletion (blue & orange *vs.* blue & green). This reduction in the accumulated flanking RNA from 2 h to 18 h after auxin treatment may reflect the presence of a compensatory mechanism that arises after prolonged acute XRN2-AID depletion. Additionally, the effect of constitutively expressing XRN2-MT in *D235A* cells *vs. XRN2-AID* cells may lead to a minor accumulation of RNA after 2h of treatment (orange *vs.* purple). However, it should be stressed this qRT-PCR experiment is only based on two biological replicates so interpretation should be treated with caution.

4.2 Evaluation of redundancy in 5'→3' exonuclease activities between DXO and XRN2

It remains possible that another 5'→3' exonuclease could act redundantly with XRN2. In yeast, Rat1 (XRN2 homologue) has its exoribonuclease activity stimulated by the binding partner Rai1 (Xue et al., 2000). This Rat1-Rai1 heterodimer greatly improves Rat1 degradation efficiency *in vitro* for RNA containing secondary structures (Xiang et al., 2009). Additionally, Rai1 possesses RNA pyrophosphohydrolase activity towards 5'-triphosphorylated (uncapped) RNAs releasing pyrophosphate and decapping endonuclease activity towards the unmethylated guanylate triphosphate cap structure releasing GpppN (Jiao et al., 2010; Xiang et al., 2009). These Rai1 activities produce a 5'-phosphorylated RNA that is a potential substrate for rat1 degradation and so it has been suggested the heterodimer is involved in an

RNA surveillance pathway for incompletely capped RNAs (Xiang et al., 2009). In a different yeast, *Kluyveromyces lactis* contains a protein with homology to Rai1 called Dxo1 that has an additional 5'→3' exoribonuclease activity but no triphosphate pyrophosphohydrolase activity (Chang et al., 2012). Dxo1 exoribonuclease activity is distributive unlike the processive nature of Rat1/XRN2 meaning it dissociates after each catalytic event (i.e. removal of a nucleotide). In humans, the homologue of Rai1/Dxo1 is called DXO (formerly DOM3Z) and possesses pyrophosphohydrolase, decapping and distributive 5'→3' exoribonuclease activities (Jiao et al., 2013). The decapping activity has specificity for incomplete and noncanonical capped RNAs including unmethylated guanylate (GpppN), nicotinamide adenine dinucleotide (NAD), flavin adenine dinucleotide (FAD) and dephospho-CoA (dpCoA) (Jiao et al., 2017; Doamekpor et al., 2020). DXO also has activity towards the mature methylated guanylate (m7GpppN) cap but this is 6-fold lower efficiency than that for NAD capped substrates (Jiao et al., 2017). Unlike Rai1, DXO in humans does not interact with XRN2 indicating a diverged role (Xiang et al., 2009). Together these results question whether DXO shares some redundancy with XRN2 and can use its exoribonuclease activity to degrade the downstream product of PAS cleavage.

As DXO is not essential a gene KO cell line can be created, which is the strategy that has previously been successfully employed in identifying DXO NAD^+-capped substrates as mRNAs and small nucleolar RNAs (snoRNA) (Jiao et all., 2017). To generate the KO cell line a cotargeting strategy was used where the targeting at a gene of interest (GOI) and another gene, *ATP1A1*, allows for enrichment of the desired modification by co-selection (Agudelo et al., 2017). The method works by mutating residues in the surface loop of an essential Na^+/K^+ active transporter, ATP1A1, responsible for electrochemical gradient homeostasis (**Fig 4.2a**). Some of the mutations within the surface loop give rise to mutants that still function as a Na^+/K^+ transporter but resistant to inhibition from the cardiotonic steroid ouabain (Laursen et al., 2015; Ogawa et al., 2009; Croyle et al., 1997). Specifically, when Q118 and N129 are substituted with positively charged residues high levels of cellular ouabain resistance are observed upon overexpression of the mutant (Croyle et al., 1997). The protocol uses a single CRISPR/Cas9 plasmid that expresses eCas9 and two sgRNA with one targeting the GOI and the other ATP1A1 surface loop.

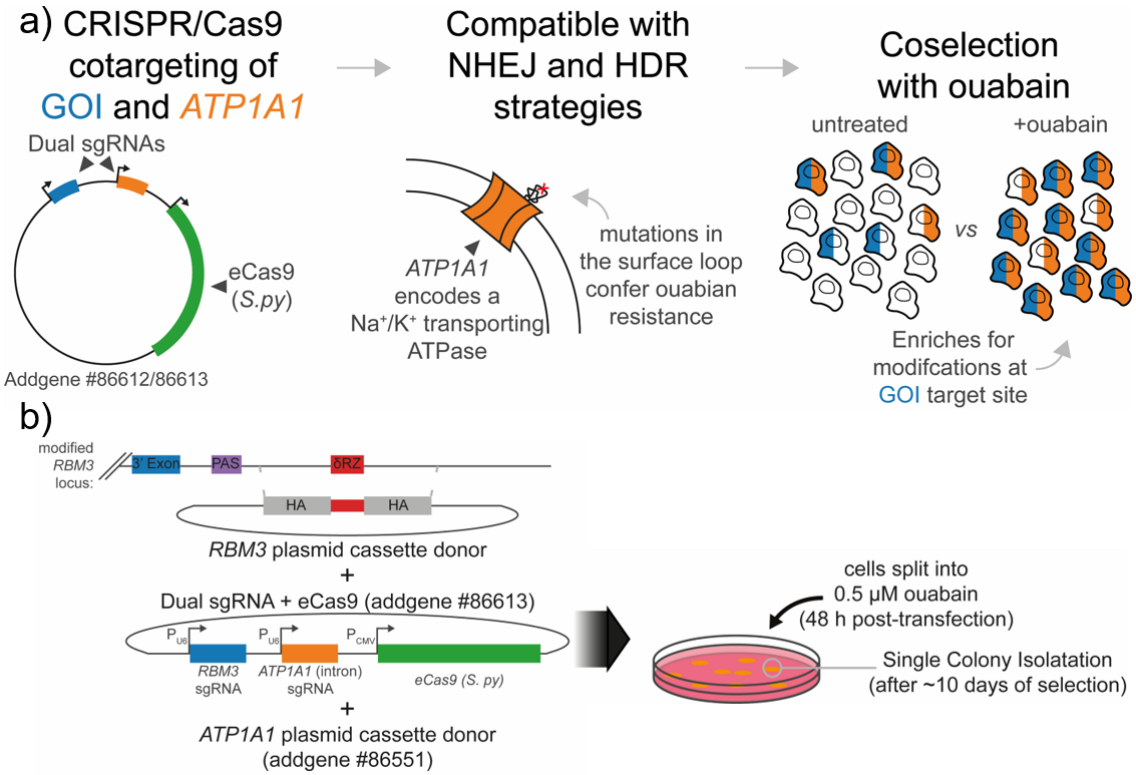


Figure 4.2 | Using a cotargeting strategy compatible with both NHEJ and HDR genome editing. a) Schematic of ouabain co-selection strategy for enriching CRISPR/Cas9 modifications at a gene of interest (GOI) target as described in Agudelo et al. (2017). **b)** An overview of the HDR protocol. Three plasmids, the *ATP1A1* donor, the GOI donor and the CRISPR/Cas9 with sgRNA cassettes, were transfected into cells. After 48 h the cells were split into a dish containing 0.5 μM ouabain. This selection media was replaced every few days (for ~10 days in total) until single colonies could be isolated. The NHEJ/indel formation strategy is similar except for that an exonic ATP1A1 sgRNA is used (Addgene #86612) and the two template donor plasmids are omitted.

After selection of ouabain-resistant colonies the proportion of clones that also possess a mutation at the GOI is highly enriched and vastly reduces the number of colonies requiring screening before a clone with a desired edit is found. Ouabain co-selection is compatible with enriching for both the generation of indels by non-homologous end joining (NHEJ) and mutation of sequences by homology-directed repair (HDR). For NHEJ a sgRNA targeting the exonic surface loop residues of *ATP1A1* is used, whereas for HDR the sgRNA targets the downstream *ATP1A1* intron and a co-transfected donor plasmid contains the Q118R and N129D ouabain-resistant mutations for repair. The protocol involves transfection of cells with plasmid(s), including the Cas9 + dual sgRNA, for 48 h (**Fig 4.2b**). Then cells are split into 0.5 μM ouabain containing media, which was replenished every ~3 days and after ~10 days separate cell colonies were isolated for screening. To generate a *DXO-KO* cell line the NHEJ double-strand DNA break repair pathway is desired as it is error-prone and commonly introduces indel mutations, some of which will cause a frameshift that prevents gene expression as a premature termination codon (PTC) will be utilised during translation and either trigger mRNA degradation by nonsense-mediated decay (NMD) or result in the production of a severely truncated protein. Within cells the DNA repair pathway chosen varies depending on cell cycle stage, target locus, nuclease and cell type, but within an asynchronous population it is more frequently the error-prone NHEJ pathway that occurs after Cas9 cleavage rather than HDR (Miyaoka et al., 2016).

The *DXO-KO* colonies were generated within the parental HCT116 and the previously modified *XRN2-AID* cell lines. The sgRNA used targeted the first coding exon (exon 2) of *DXO*. After selection successfully modified homozygous clones for each cell line were identified that had lost the presence of DXO by western blot (**Fig 4.3a**). To confirm successful editing, genomic DNA (gDNA) was isolated from the cell lines to amplify the DXO sgRNA site with primers flanking the locus and the produced amplicons were cloned into a plasmid vector. Then multiple plasmid colonies were sent for Sanger sequencing to determine the sequences of indels present within each allele. The indels generated within the *DXO-KO* and *DXO-KO XRN2-AID* cell lines all varied but all cause a frameshift (**Fig 4.3b**).

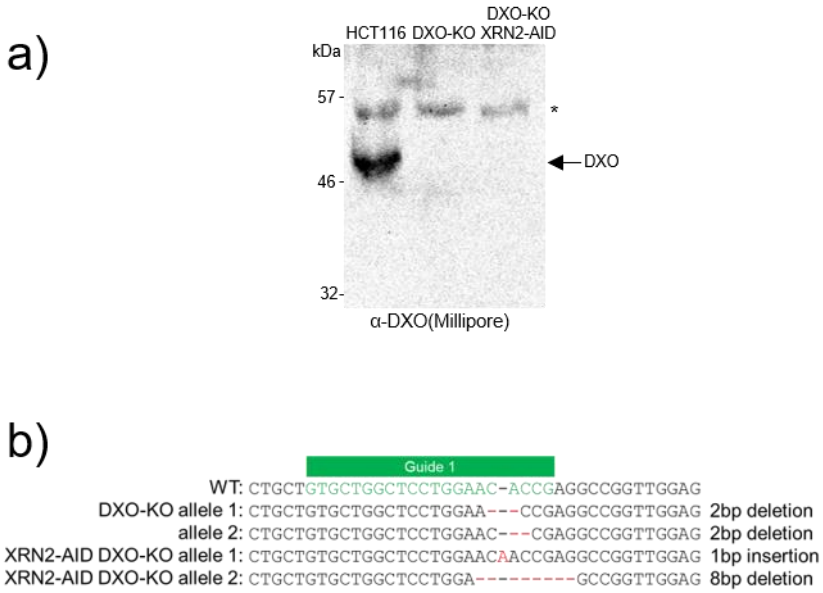


Figure 4.3 | Generation and validation of *DXO-KO* cell lines using ouabain co-selection. a) WB validation of the *DXO-KO* and *DXO-KO XRN2-AID* cell lines. Asterisk denotes a non-specific band used as a loading control**. b)** Cloning and Sanger sequencing of DXO-KO indels within the positive clones. The DXO sgRNA target sequence is shown in green. Red dashes indicate deletions and red nucleotides indicate insertions.

To assess for redundancy between DXO and XRN2, mNET-seq was performed on HCT116 and *DXO-KO* cells in the absence of auxin and *XRN2-AID DXO-KO* cells in the presence and absence of auxin using the same protocol as described in section 3.2. The HCT116 parental data set is the same one used as a control reference within chapter 3. After sequencing, the raw reads were quality assessed before data processing. The PHRED quality scores of the average base calls for the raw reads were > 30 throughout the entire length of the 50 bp paired reads captured (**Fig 4.4a**). Like the XRN2-AID mNET-seq datasets, the adapter content here increases from ~20bp reflecting the short fragment libraries generated and the 50 bp sequencing cycles (**Fig 4.4b**). The composition of the first nucleotide from both read pairs of each sample are shown in **Fig 4.4c,d**. As mentioned previously, MNase has a strong bias for cleaving upstream of A and T nucleotides and this is reflected in all of these samples indicating successful digestion of the chromatin pellet during the library protocol. Next reads were trimmed of adapter sequences and screened for the presence of contaminants using Fastqsceen. In all samples, >80% of the reads mapped to the human genome (**Fig 4.4e**). As previously for the XRN2-AID samples some reads did map to the mouse and CHO cell genomes but these were classed as mapping to 'multiple genomes'. Because there are very few reads mapping uniquely to these contaminant references this suggests that these multiply mapped reads probably derived from conserved eukaryotic sequences.

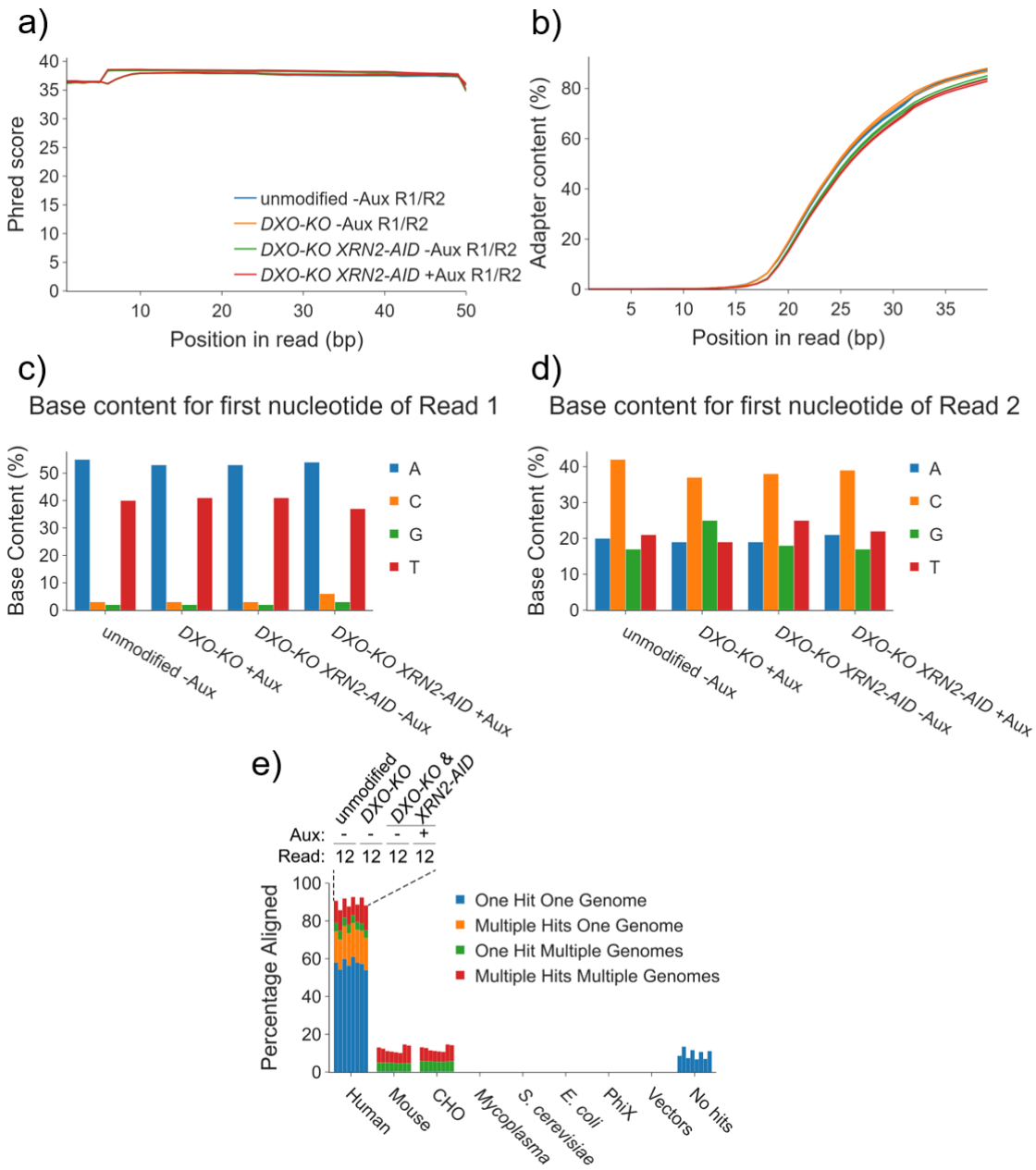


Figure 4.4 | Quality assessment graphs for DXO-KO mNET-seq samples. **a)** The average Phread quality score for each nucleotide position within the raw reads. **b)** Adapter content (%) for each base pair within the raw reads. **a,b)** The key applies to both. 'R1' means read 1 and 'R2' = read 2. Data generated from FastQC. **c,d)** The percentage nucleotide composition of the 5' nt of the RNA fragment (first nt of read 1, **'c'** and the complementary nt at the 3' end (first nt of read 2, **'d'**. **e)** Adapter trimmed reads were screened for contamination with 'fastq_screen'. A sample of 100000 reads, taken evenly throughout the read file, were aligned to various datasets of commonly used lab species and reagents. The key for the grouped bars is shown above human. HCT116 cells were either unmodified, with DXO-KO or DXO-KO and expressing TIR1 along with biallelic AID tagging of XRN2 in the presence and absence of Aux (Aux = auxin; nt = nucleotide).

The 3'-end of protein-coding transcripts were analysed by total-CTD Pol II mNET-seq to assess the impact of DXO-KO on degrading the downstream product of PAS cleavage with its distributive exoribonuclease activity. Of the four example protein-coding transcripts shown all have no or very minor changes in the DXO-KO (orange) mNET-seq profiles compared to the HCT116 (blue) parental profile (**Fig 4.5a-d**). Both the mNET-seq profile and the final position of termination are unchanged in each trace. Likewise, the profile is also unchanged for the *DXO-KO XRN2-AID* cells untreated with auxin (green *vs.* blue). However, when *DXO-KO XRN2-AID* cells are treated with auxin for 2 h transcription read-through is observed downstream of these transcripts. For reference, the *XRN2-AID* rep 1 mNET-seq samples untreated and treated with auxin (2 h) from chapter 3 are shown in purple and brown, respectively. When comparing *XRN2-AID* cells either unmodified or modified with DXO-KO and untreated or treated with auxin (green *vs.* purple and red *vs.* brown) the profiles and position of final termination correlate. This suggests DXO exoribonuclease activity does not functionally complement XRN2-AID depleted cells by co-transcriptionally degrading the products of PAS cleavage for the majority of Pol II species (as captured by the total-CTD antibody). However, as mNET-seq is specific for Pol II-associated transcripts with a protected/inaccessible 3'-end it does not rule out a role for DXO in RNA metabolism pathways that act post-transcriptionally or pathways that act on a subset or underrepresented species of RNA/Pol II. Likewise, when mammalian DXO targets NAD capped RNAs these transcripts are a small proportion and have at most 6 % NAD capped RNAs relative to total mRNA which increases to a maximum of 11% of total mRNA within a *DXO-KO* cell line (Jiao et al., 2017). The human data differs from that in budding yeast where a Rai1 deletion strain stabilised RNA downstream of transcripts (Kim et al., 2004).

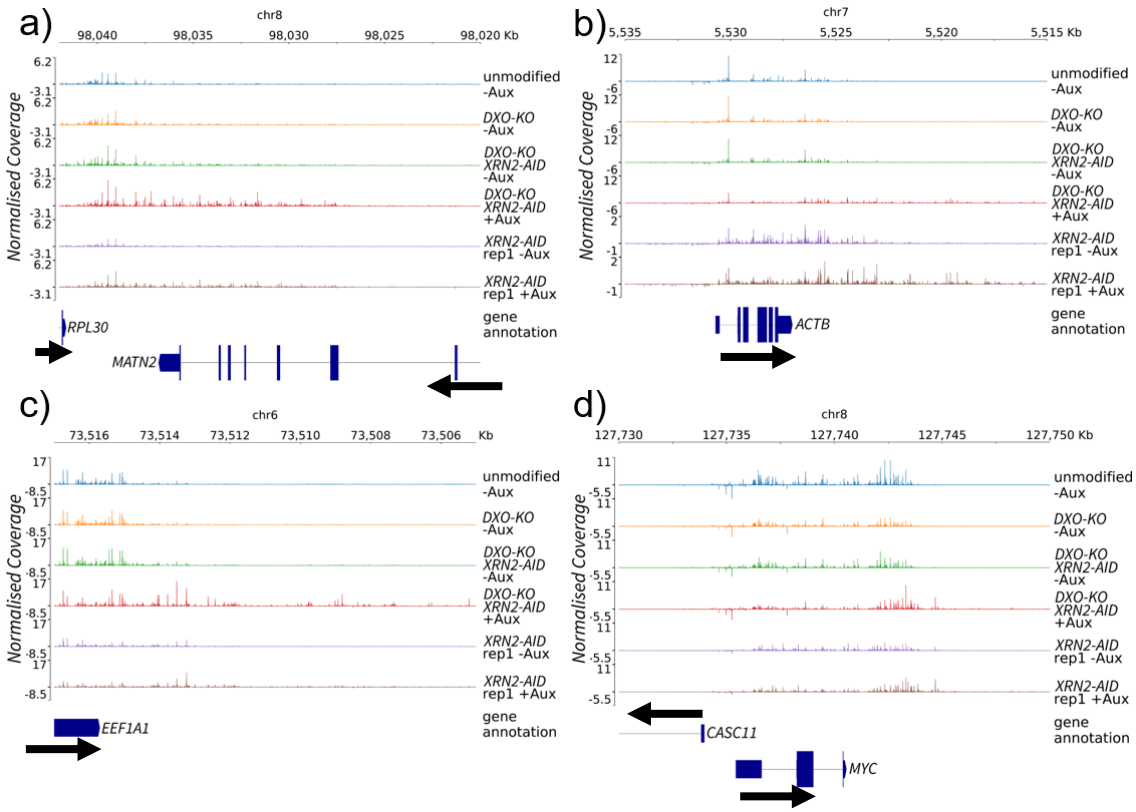


Figure 4.5 | *DXO-KO* cells show no redundancy between DXO and XRN2 activities at the 3'-end of protein-coding transcripts. a-d) Total Pol II mNET-seq snr snapshots are shown of four protein-coding transcripts**,** *RPL30*, *ACTB*, *EEF1A1* and *MYC* (**a-d**, respectively). Samples were either untreated (-Aux) or treated with auxin for 2 h (+Aux). XRN2-AID rep1 samples shown in purple and brown traces are reproduced from chapter 3 to allow comparison. The transcription profiles only change within the XRN2-AID samples treated with auxin. Black arrows indicate transcript directionality and normalised coverage for chromosomal snapshots equals

4.3 Insertion of viral sequences to occlude 5'→3' torpedo activity on single transcripts

So far strategies to further remove trace XRN2 (if any), occlude the 3' PAS cleavage product from other exonucleases and remove DXO (a known nuclear exonuclease) have failed to accumulate downstream RNA above that already observed after auxin depletion of XRN2-AID alone. These strategies suggest that neither trace levels of XRN2 nor DXO explain the lack of long read-through transcription seen when XRN2-AID is depleted. Nevertheless, they do not rule out the existence of uncharacterised 5'→3' exonucleases that can act when XRN2 is absent. Because of this, a further strategy was designed to inhibit 5'→3' exonucleases in a manner agnostic to their identity.

An experiment was envisaged where insertion of RNA sequences inserted downstream of a PAS would allow interference of 5'→3' exonucleases on a single transcript within the *XRN2-AID* cell line without global impairment of XRN2 while untreated. The hepatitis δ ribozyme (δRZ) sequence is an efficiently self-cleaving RNA that generates RNA products with 5'-hydroxyl and a cyclic 2',3'-monophosphate (Sharmeen et al., 1988). The downstream product is resistant to exonuclease degradation because a phosphorylated 5'-end is required by many exonucleases for degradation (including XRN2), whereas the upstream product can be rapidly degraded from the 3'-end (Stevens and Maupin et al., 1987; Muniz et al., 2015). The δRZ also has an inactivating single-point mutant (δRZ[MT]) that abolishes cleavage activity (Fong et al., 2009). In essence, cleavage by this ribozyme produces a 5' end that will not be degraded by a 5'→3' exonuclease, regardless of its identity.

To obviate the need for a selection marker at the insertion site, which would have the undesired effect of changing the local transcription dynamics surround the δRZ and so prevent useful interpretation, a HDR ouabain co-selection strategy was used (see description in above section 4.2, **Fig 4.2b**). This means only the 73 bp δRZ sequence will be inserted downstream of a target protein-coding transcript's PAS (with other edits occurring at *ATP1A1* to allow ouabain-resistant co-selection). To further simplify the editing strategy a protein-coding transcript on the X-chromosome was sought so that only one allele requires editing because HCT116 cells are male. The transcript, *RBM3*, was chosen because it is well expressed, has no other expressed transcripts

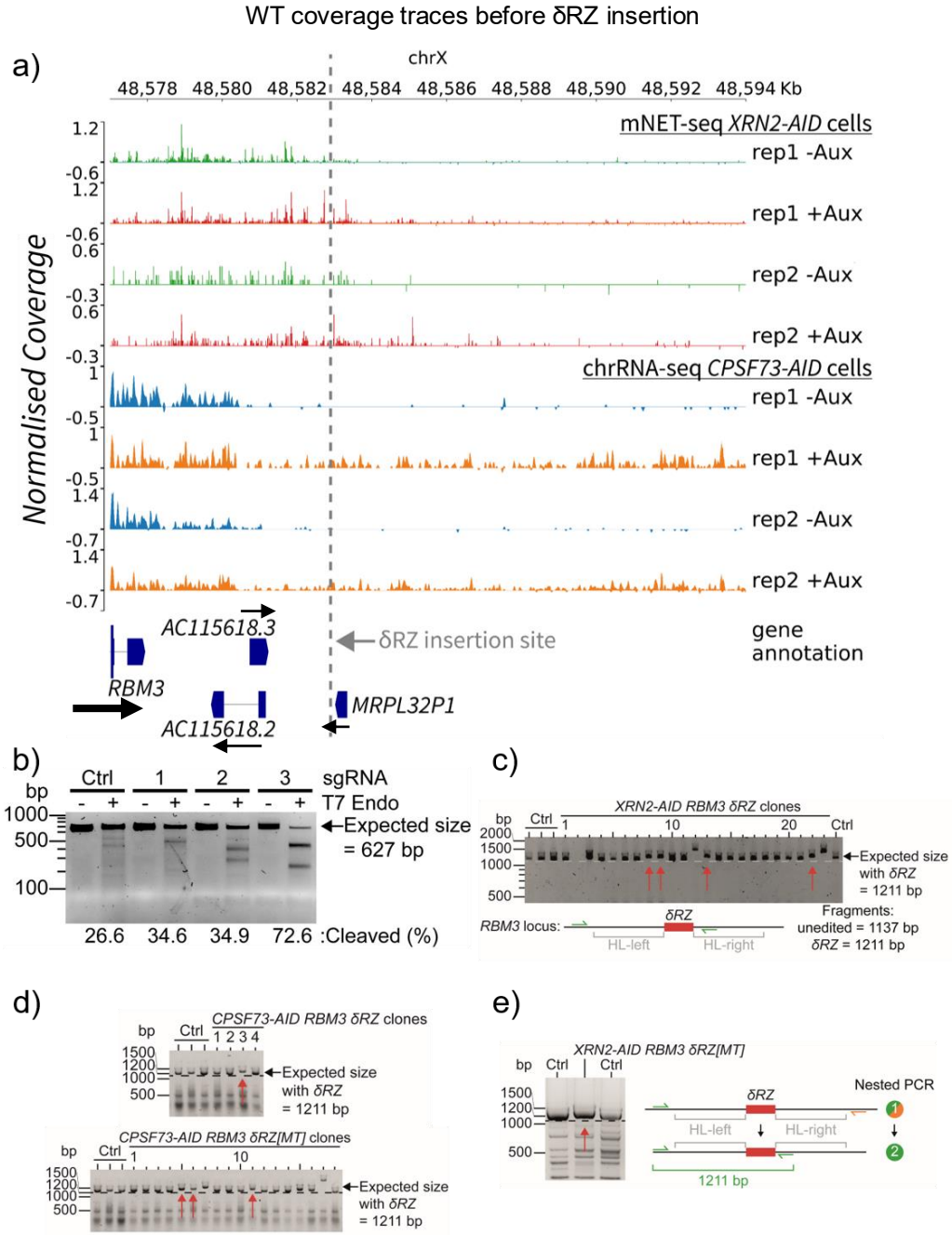


Figure 4.6 | Generation and validation of δRZ insertion downstream of *RBM3* within *XRN2-AID* and *CPSF73-AID* cell lines. a) Genome snapshot of *RBM3* flanking region showing the typical XRN2 and CPSF73 dependent transcription read-through of a protein-coding transcript. The δRZ insertion site is displayed as a dashed line. **b)** Surveyor assay of three sgRNAs targeting a region downstream of RBM3 PAS. The expected PCR fragment is 627 bp with primers flanking the RBM3 sgRNA targeting sites and the product was digested or not with T7 Endonuclease. The cleaved percentage was determined by comparing the intensity of the 627 bp band in the T7 endo treated to the untreated condition. The control sgRNA condition used the unedited ouabain cotargeting plasmid (Addgene #86612). Cells were split into 0.5 μM ouabain 48 h after transfection and left for 10 days to enrich for successfully transfected cells. **c,d,e)** A genomic DNA PCR screen of RBM3 δRZ insertion clones within *XRN2-AID* **(c+e)** and *CPSF73-AID* **(d)** cells. A single PCR was performed for δRZ[WT] (**c**) into XRN2-AID cells and δRZ[WT/MT] into CPSF73-AID cells (**d**). Whereas a nested PCR was performed for δRZ[MT] into XRN2-AID RBM3 cells. Dashed lines are drawn from the migration front of control PCR products and red arrows highlight cell clones that have the correct insertion size. Black arrows indicate transcript directionality and normalised coverage for chromosomal snapshots equals single base pair bin size TPM.

immediately downstream within HCT116 cells and displayed typical PAS-dependent termination traits. For instance, within the *XRN2-AID* mNET-seq dataset, read-through transcription and accumulation of RNA is observed across a range of ~10kb downstream of *RBM3* upon depletion of XRN2-AID by auxin (**Fig 4.6a**). Similarly, Pol II termination at *RBM3* depends on PAS processing as seen using a *CPSF73-AID* cell line (this cell line is the focus of the next chapter but is displayed here as a control). The insertion site for δRZ sequence is shown as a dashed line over the WT sequencing data obtained before the insertion occurred (**Fig 4.6a**). It is ~4 kb downstream of the PAS and is positioned at the frontier of the XRN2-AID mNET-seq signal in untreated cells. The reasoning for this site is detailed as follows.

When XRN2-AID is depleted, transcription read-through occurs and more signal is detected downstream of the WT (or untreated) termination region. The positioning of the ribozyme at this WT frontier allows assessment of whether the region of extended signal reveals the true Pol II termination site downstream of RBM3 that under WT conditions is masked by rapid degradation from XRN2 (or other exonucleases). There is a possibility that, rather than Pol II terminating stochastically across the region downstream of the PAS, most of the polymerases transcribe past this frontier but when they do the majority have undergone PAS cleavage and rapid exonuclease degradation, thus becoming invisible to mNET-seq and only revealed by XRN2-AID depletion. If upon loss of XRN2-AID most of the Pol II complexes transcribe past the δRZ then the signal downstream of it will be stabilised by the 5'-hydroxyl-end but notably the upstream fragment will be rapidly degraded from the cyclic 2',3'-monophosphate in a 3'→5' direction (Sharmeen et al., 1988; Muniz et al., 2015). Alternatively, if upon loss of XRN2 these upstream regions instead accumulate in the δRZ cells and the downstream regions only show a modest stabilisation then this region of extended signal arises from only a fraction of the *RBM3* transcribing polymerases.

To select the exact insertion site within this region a surveyor assay was performed with three candidate sgRNAs to determine the one with the best targeting efficiency. A surveyor assay reveals the sgRNA that induces the most indel formation by NHEJ and therefore highest cleavage percentage by T7 endonuclease I (T7 Endo I). The highest indel forming sgRNA is then used as

an estimate for the sgRNA with the best target site cleavage efficiency (**Fig 4.6b**). The sgRNAs were cloned into the NHEJ + eCas9 ouabain plasmid and then transfected into cells. Cells were selected with ouabain to improve the efficiency of the surveyor assay by selecting cells that had been successfully transfected with the plasmid and undergone editing at the *ATP1A1* transcript. Of the three sgRNAs tested, guide 3 was selected as it was markedly better at inducing indels. The δRZ wild-type [WT] and single-point mutant [MT] sequences were then inserted into XRN2-AID and CPSF73-AID cell lines. After single colonies were isolated they were initially screened by PCR using primers flanking the insertion site with at least one primer annealing outside of the homology arms to exclude the possibility of donor plasmid contamination (**Fig 4.6c,d,e**). Positive clones that showed an increase of ~74 bp (highlighted with red arrows) had the PCR products sequenced to confirm correct insertions.

Relative RNA fold changes were examined by qRT-PCR on total RNA from δRZ[WT/MT] modified cell lines. The *CPSF73-AID* cells modified with δRZ[WT/MT] are shown here as a control but are discussed within the future section 5.5 (**Fig 4.7a**). When auxin is added to *CPSF73-AID* cells it results in most Pol II complexes transcribing beyond the δRZ insertion site (positive control). This allows confirmation that the δRZ[WT] sequence does cleave within the context of the RBM3 sequence because loss of CPSF73 results in upstream amplicons (at UCPA and ds1.1kb) failing to accumulate in δRZ[WT] modified cells, whereas in unmodified and δRZ[MT] modified CPSF73-AID cells there is a significant accumulation (red *vs.* brown, $p<0.05$). Notably, the CPSF73-AID data demonstrates the expected trend if most of the polymerases transcribe past the δRZ insertion site with degraded upstream fragments and stabilised downstream fragments in the absence of XRN2.

Within *XRN2-AID* cells modified with δRZ[WT] a downstream amplicon at ds8.5kb shows mild RNA accumulation even in the presence of XRN2 compared to the δRZ[MT] (green *vs.* purple, $p<0.05$) (**Fig 4.7b**). This is likely due to a small proportion of polymerases that transcribes past the δRZ insertion site in untreated/non-depleted conditions because it is also present in untreated *CPSF73-AID* δRZ[WT] cells. Additionally, when XRN2 is lost the presence of the δRZ[WT] does marginally improved the RNA accumulation at ds8.5kb and ds11kb amplicons downstream of the insertion site but this increase is not

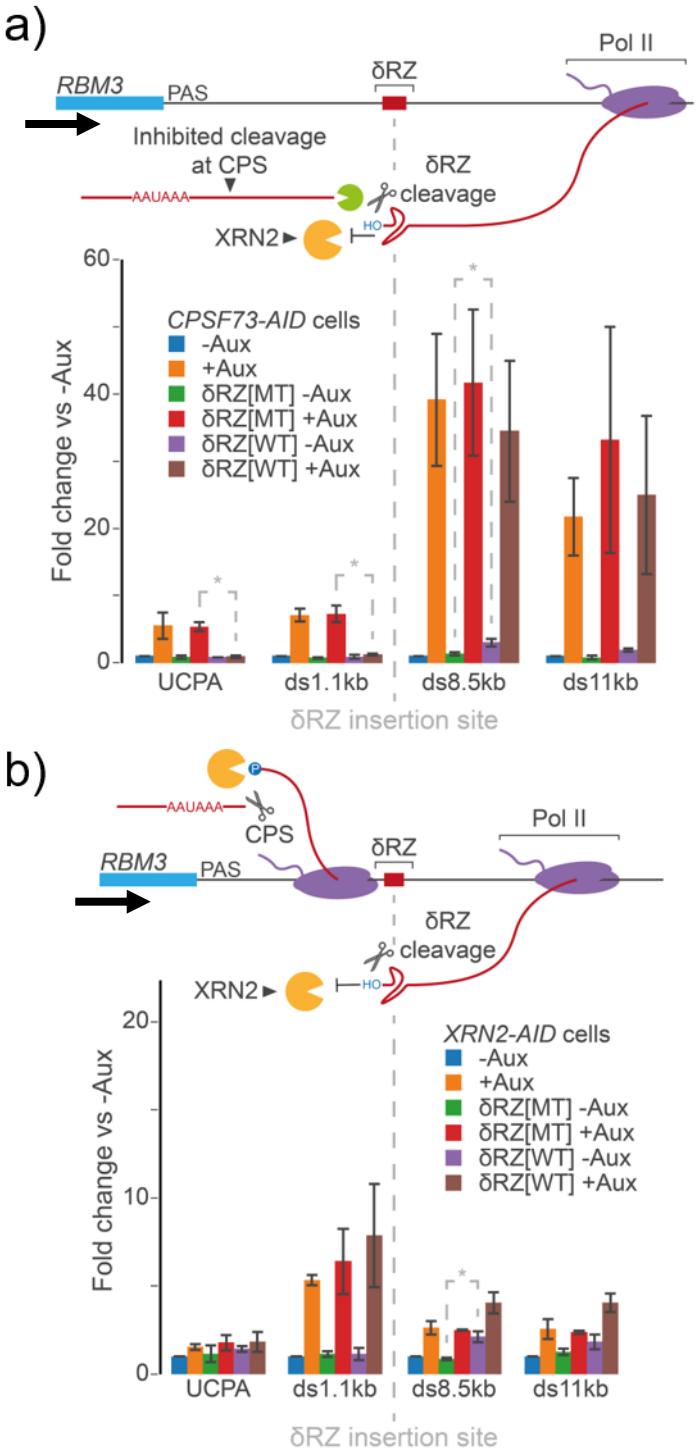


Figure 4.7 | XRN2 depletion leads to the accumulation of Pol II complexes that are stalled or slowed. qRT-PCR analysis of *CPSF73-AID* (**a**) or *XRN2-AID* (**b**) cells unmodified and modified downstream of *RBM3* by insertion of δRZ with [WT] or inactivating point mutant (MT). **a)** Fold changes of total cell RNA from *CPSF73-AID* cells normalised to spliced *ACTB* and relative to '-Aux'. Cells are treated with neither or both Dox (18 h) followed by auxin (3 h) and labelled '-Aux' or '+Aux' respectively (n=3; error bars=sem). **b)** Fold changes of total cell RNA from *XRN2-AID* cells normalised to spliced *ACTB* and relative to '-Aux'. Cells are treated or not with auxin (2 h; n=3; error bars=sem). CPS=cleavage and polyadenylation site (or site of PAS cleavage). Asterisk (*) indicates two-tailed student's t-test where $p<0.05$. Black arrows indicate transcript

significant (red *vs.* brown). This mirrors the small but non-significant increases at downstream amplicons when XRN2-AID cells treated with auxin were additionally treated with siXRN2 or expressed XRN2-MT (**see Fig 4.1**). This indicates that whilst there may be some trace 5'→3' activity it does not promote efficient and generalised Pol II termination after XRN2-AID depletion by auxin alone and has little impact on the final frontier of Pol II. Lastly, the upstream amplicon, ds1.1kb, is stabilised in all *XRN2-AID* cells upon loss of XRN2, which is unlike CPSF73 loss in the δRZ[WT] modified *CPSF73-AID* cells where the ribozyme cleavage prevents this accumulation. This suggests that in *XRN2-AID* cells most Pol II does not transcribe past the δRZ insertion site when XRN2 is lost and therefore the accumulated Pol II targeted by XRN2 is likely stalled or slowed.

Given that Pol II complexes are stalled or poorly elongating at the 3'-ends of transcripts in the absence of XRN2 the question remains whether they terminate by an XRN2-independent mechanism. From the results above if such a mechanism does occur it is unlikely to require 5'→3' exonuclease activity. However, a reasonable suggestion is that such a process may release RNA upon Pol II dissolution, exposing a 3'-end that would provide an entry site for a 3'→5' exoribonuclease degradation. In humans, DIS3 is a catalytic component of the exosome complex which degrades many short non-coding RNAs such as PROMPTs and enhancers with 3'→5' exoribonuclease activity (Preker et al., 2008). Exosome depletion may have a synthetic lethal relationship with XRN2 depletion and cause further accumulation of downstream RNAs. To test this, we sought to introduce another viral sequence, which occludes 5'→3' degradation, downstream of the PAS of *MORF4L2*, a protein-coding transcript on the X-chromosome, using a similar ouabain co-selection strategy as used above. The Pol II termination characteristics of *MORF4L2* appear typical of protein-coding transcripts (and like *RBM3*) to the extent it has limited read-through upon XRN2 loss and longer read-through upon CPSF73 loss (**Fig 4.8a**). Here two tandem XRN-resistant RNAs (xrRNA) derived from West Nile virus are inserted as it folds into a secondary structure that impairs 5'→3' degradation and has previously been employed to investigate the directionality of RNA decay (Chapman et al., 2014; Horvathova et al., 2017; Voigt et al., 2019). The xrRNA was introduced into *XRN2-AID* and *DIS3-AID* cells, of which

the latter is an AID tagged cell line of the major catalytic component of the exosome with rapid depletion leading to accumulation of PROMPT and enhancer RNAs (Davidson et al., 2019). The *DIS3-AID* cells were generated and validated by L.F. and this is documented in Davidson et al. (2019). Four candidate sgRNAs targeting the downstream flank of *MORF4L2* (~400 bp beyond the PAS) were screened to estimate the sgRNA with the best targeting efficiency (**Fig 4.8b**). Of those tested, guide 4 produced the most indels so was selected. Cells were co-transfected with the dual sgRNA+ eCas9 plasmid and two template donor plasmids using the same protocol as described in **Fig 4.2b**. After 48 h cells were placed in ouabain containing selection media. Then isolated colonies were screened using PCR and fragments with an increased migration size of ~318 bp (highlighted with red arrows) were sent for sequencing to confirm the correct insertion of the xrRNA (**Fig 4.8c,d**).

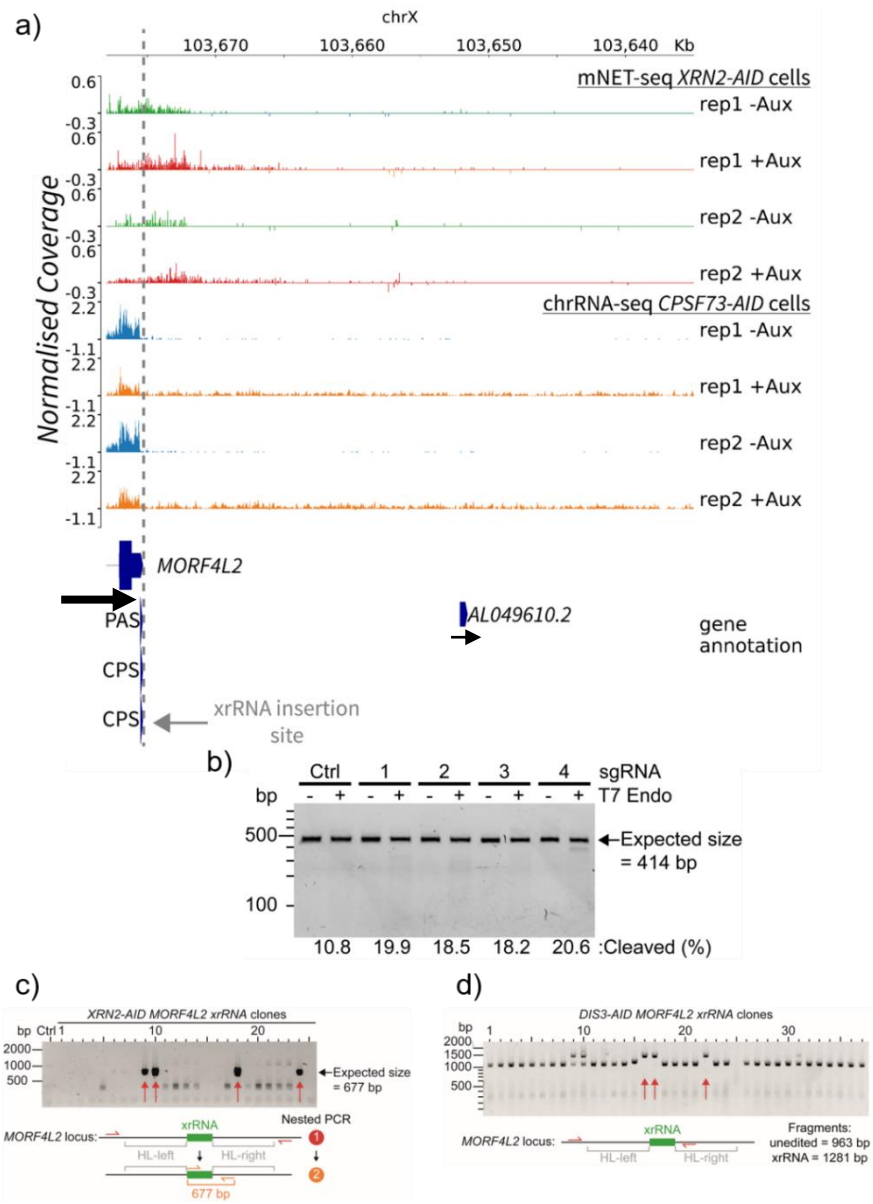


Figure 4.8 | Generation and validation of xrRNA insertion downstream of *MORF4L2* within *XRN2-AID* and *DIS3-AID* cell lines. a) Genome snapshot of *MORF4L2* flanking region showing the typical XRN2 and CPSF73 dependent transcription read-through of a protein-coding transcript. The xrRNA insertion site is displayed as a dashed line. **b)** Surveyor assay (as in Fig 4.3) to screen of sgRNAs targeting a region downstream of *MORF4L2* PAS. The expected PCR fragment is 414 bp with primers flanking the RBM3 sgRNA targeting sites and the product was digested or not with T7 Endonuclease. **c)** A genomic DNA nested PCR screen of XRN2-AID MORF4L2 xrRNA insertion clones. The first PCR used primers (red) that annealed outside of the homology arms and the second PCR used this template with primers (orange) annealing to the xrRNA and downstream homology arm. The expected fragment size is 677 bp for correct insertion of the xrRNA sequence. **d)** A genomic DNA PCR screen of DIS3-AID MORF4L2 xrRNA insertion clones. The PCR used one primer that annealed outside of the homology arms and another within the downstream homology arm. The expected fragment size is 1281 bp for correct insertion of the xrRNA sequence. Red arrows highlight positive clones. Black arrows indicate transcript directionality and normalised coverage for chromosomal snapshots equals single base pair bin size TPM.

Within the *DIS3-AID* xrRNA-modified cells, XRN2 will be unable to degrade past ~400 bp downstream of the *MORF4L2* PAS as it will encounter the xrRNA sequence. The xrRNA will not only block XRN2 but other potential 5'→3' exonuclease, such as nuclear localised XRN1 or CPSF73 exonuclease activity (Luo et al., 2006; Yang et al., 2020). The use of rapid depletion DIS3-AID cells will allow investigation of whether Pol II dissolution occurs in the absence of XRN2 by assessing whether such a process exposes an RNA 3'-end from the active site that could be degraded by DIS3 (**Fig 4.9a**, see question mark). qRT-PCR was performed on total RNA extracted from unmodified and xrRNA-modified *DIS3-AID* cells untreated or treated with auxin (2 h). At amplicons upstream of the xrRNA there is little difference in fold changes between all conditions (**Fig 4.9b**). However, at positions downstream of the insertion site xrRNA-modified cells lead to accumulated RNA in both untreated and auxin treated conditions. The depletion of DIS3-AID does not affect the accumulation at ds600bp amplicon and suggests marginal 3'→5' degradation occurs and/or inefficient XRN2-independent Pol II dissolution and termination (green *vs.* red at ds600bp).

In addition, qRT-PCR was performed on XRN2-AID unmodified and xrRNA-modified cells. As expected at a position upstream (ds200bp) of the insertion site on *MORF4L2* (but downstream of the PAS), RNA accumulates only when XRN2 is depleted by auxin treatment (**Fig 4.9c**). Whereas, immediately downstream of the xrRNA at the ds600bp amplicon, RNA is stabilised with xrRNA-modified cells even in the presence of XRN2, as was the case above in *DIS3-AID* modified cells. This demonstrates that the xrRNA is functioning correctly to inhibit 5'→3' exonucleases. However, at more distal positions downstream of the ds600bp amplicon the accumulation of RNA in modified cells was not significantly greater than in unmodified cells when treated with auxin (orange *vs.* red at positions >ds2.7kb). This strongly implies that 5'→3' exonucleases (other than XRN2) do not readily degrade the downstream products of PAS-cleavage and are not involved in a general auxiliary Pol II termination mechanism in the absence of XRN2.

Interestingly, the xrRNA modified XRN2-AID cells allow examination of a core suggestion of the torpedo model. The model implies that capture of Pol II by an exonuclease rather than simply RNA degradation is critical to instigating

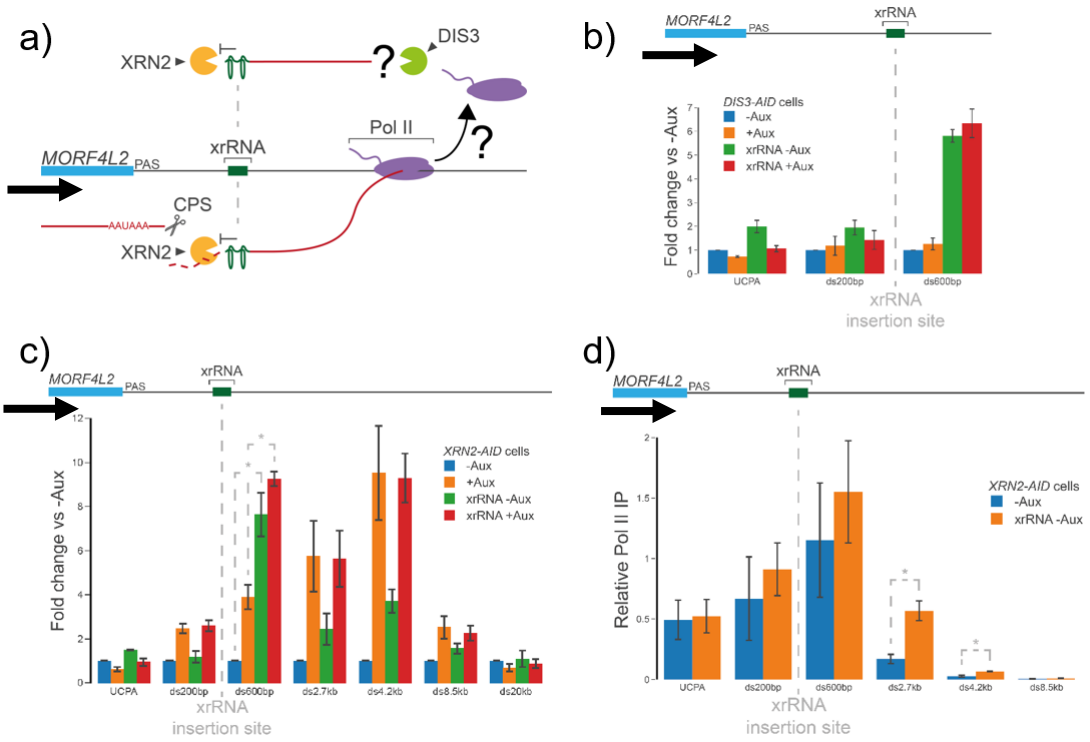


Figure 4.9 | DIS3 is not ubiquitously involved in a 5'→3' exonuclease-independent termination mechanism. **a)** Schematic showing the downstream flanking region of *MORF4L2* locus with the inserted xrRNA sequence (dark green). XRN2-independent termination possibilities are shown with question marks; one being whether Pol II dissolution occurs in the absence of XRN2 degradation, and the other is if the subsequent released RNA is degraded by the 3'→5' exonuclease DIS3. **b)** qRT-PCR of total RNA from *DIS3-AID* cells unmodified or modified with xrRNA insertion downstream of *MORF4L2* and treated or not with auxin (2 h). The RNA fold change is shown relative to unmodified DIS3-AID cells not treated with auxin and normalised for RNA input using a spliced ACTB amplicon. n=3 and error bars are s.e.m. **c)** qRT-PCR of total RNA from *XRN2-AID* cells unmodified or modified with xrRNA insertion downstream of MORF4L2 and treated or not with auxin (2 h). The RNA fold change is shown relative to unmodified DIS3-AID cells not treated with auxin and normalised for RNA input using a spliced ACTB amplicon. n=3 and error bars are s.e.m. Asterisks indicates a $p<0.05$ at the ds600bp amplicon. **d)** Pol II ChIP of XRN2-AID cells unmodified or modified with xrRNA insertion downstream of *MORF4L2* and not treated with auxin. Asterisks indicate a $p<0.05$ at the ds2.7kb and ds4.2kb amplicons. Black arrows indicate transcript directionality.

termination. As xrRNA occludes downstream RNA degradation by XRN2, which has already begun degrading the RNA from PAS cleavage site, a termination defect would indicate prolonged chasing and capture of Pol II is required. A Pol II ChIP was carried out on unmodified and xrRNA-modified MORF4L2 cells that were untreated (with XRN2 present) (**Fig 4.9d**). The amplicons downstream of the xrRNA insertion site show significant increase in Pol II accumulation (at ds2.7kb and ds4.3kb, $p<0.05$) and therefore demonstrating that degradation of a few hundred nt alone is insufficient and strongly implies that XRN2 must come within sufficient contact of Pol II to instigate termination.

4.4 Discussion

This chapter examines the 5'→3' exonuclease "torpedo" activities downstream of protein-coding transcripts. The mNET-seq data from DXO-KO cell lines demonstrate that DXO is not involved in degrading the downstream product of PAS-cleavage and does not affect the read-through generated downstream of the PAS after XRN2-AID depletion (see **Fig 4.5**). The only other known human 5'→3' exoribonucleases that could potentially have redundancy with XRN2 are XRN1, which when overexpressed with a nuclear-localising tag in yeast degrades flanking RNA but does not induce Pol II termination, and a secondary role of CPSF73 exoribonuclease activity (Luo et al., 2006; Yang et al., 2020). However, these are unlikely to be involved downstream of protein-coding transcripts, because insertion of the xrRNA sequence into the flank of *MORF4L2*, which blocks 5'→3' exoribonucleases, does not induce read-through at more distal positions (see ds8.5kb **Fig 4.9c,d**). Thus, termination or stalling appears only delayed.

One caveat of using a genetic KO approach compared to protein depletions is that the gene may be susceptible to nonsense induced transcriptional compensation (NITC) which can mask its true impact (El-Brolosy et al., 2019; Ma et al., 2019). Most genetic KO methods use the introduction of frameshift indel mutations early on in a transcript leading to a PTC which result in either mRNA degradation by NMD or the production of severely truncated proteins. NITC is dependent on a PTC that induces NMD and leads to the upregulation of compensatory genes. Whilst much of the mechanism of NITC is still to be uncovered it requires the COMPASS complex, which is responsible

for H3K4me3 deposition, and the upregulated genes have increased H3K4me3 at their promoters. However, it seems unlikely that *DXO-KO* cells induce NITC because these cells have previously been used to identify accumulated NAD^+-capped RNAs (Jiao et al., 2017). Also, NAD+-capped RNA represents only a subset of total RNA and of the transcripts presented in Jiao et al. (2017) the one with the greatest proportion of NAD+-capped RNA relative to total RNA represented 6 % of total RNA in control cells and increased to 11% of total RNA in *DXO-KO* cells. This small fraction of total RNA which are NAD+-capped species may explain the lack of observable difference between control cells. Additionally, the overexpression of a catalytic point mutant E234A that abolishes exonuclease activity (which negates the NITC issue) similarly leads to no discernible change in the accumulation of flanking RNA (Jiao et al., 2013; data not shown).

One interesting finding from the insertion of the xrRNA 400 bp downstream of *MORF4L2* is that degradation of the first 400 nt by XRN2 (i.e. prior to the xrRNA) is insufficient to trigger termination and continued degradation of XRN2 to a position close to Pol II is needed to induce termination (see **Fig 4.9d**). The exact details surrounding how XRN2 stimulates Pol II complex dissolution from DNA remains enigmatic. *In vitro* XRN2 and Rat1 terminate Pol II in an ATP-dependent process without the need for additional components (Park et al., 2015). However, they are unable to terminate *E. coli* RNAP indicating a specific interaction is likely required. Likewise, *E. coli* or *Bacillus subtilis* RNAP is terminated more efficiently by RNaseJ1 exonuclease activity than by XRN1 (Šiková et al., 2020). Torpedo termination in eukaryotes may share similar mechanistic similarities with the interaction between XRN2 and Pol II mirroring that of Rho and RNAP, which is thought to induce conformation changes and RNA:DNA hybrid melting that stimulate disassembly (Epshtein et al., 2010).

The δRZ insertion shows that the most accumulated Pol II complexes arising upon XRN2 depletion have poor elongation capacity because most do not transcribe far downstream with the read-through signal resulting from only a small proportion of total polymerases. As these Pol II complexes are stalled or slowed it suggests such an activity serves to facilitate Pol II capture by XRN2. Correspondingly, such slowing has been observed over termination regions at

motifs consisting of a stretch of low thermodynamic stability hybrids following an upstream C/G-rich segment (Schwalb et al., 2016). Given this it brings into question if the extended read-through transcription seen with Pol II mutants with a relatively faster elongation rate is due to kinetic competition with XRN2 or whether it represents a further position a polymerase reaches before pausing is onset (Fong et al., 2015). Pausing over termination regions is not a guaranteed outcome because during situations of osmotic or viral stress transcription can continue many 100s kb downstream of genes so the process must be a regulated (Bauer et al., 2018; Erickson et al., 2018; Rutkowski et al., 2015; Vilborg et al., 2015). At what stage of transcription this pausing is triggered and how it is mediated is the focus of the next chapter.

Chapter 5

5. PAS-dependent CPSF73 cleavage triggers a joint allosteric and torpedo termination mechanism of Pol II

Declarations: CPSF73-AID cells were generated and chrRNA-seq library creation (excluding bioinformatic analysis) was performed by S.W. The western blot validating CPSF73-AID depletion was created by L.F. and the western blot showing PP1 RNAi efficiency was created by S.W.

So far, chapter 3 has demonstrated XRN2 involvement at the 3'-ends of PAS-dependent transcripts, miRNA cluster host genes (e.g. MIR17HG) and some PAS-independent non-coding transcripts (e.g. MALAT1/NEAT1) by use of a rapid depletion AID cell line. However, this effect appeared incomplete with an accumulation of Pol II downstream of transcripts only extending on average ~10kb. Then, chapter 4 further analysed the 5'→3' exonuclease torpedo activity and showed that trace XRN2 (if it exists after auxin-mediated depletion) or potential other 5'→3' exonucleases were not responsible for the limited read-through distance. As such, the arrest or slowing of Pol II that is nevertheless unable to terminate might adequately explain the limited read-through transcription seen in chapter 3. Importantly, the theoretical intrinsic termination products of these accumulated Pol II at protein-coding transcripts (if they exist) are not abundant substrates for DIS3 in the absence of 5'→3' exonuclease degradation arguing against widespread exosome degradation of an intrinsic termination failsafe process. This does not rule out a secondary failsafe mechanism at 3'-ends of protein-coding transcripts but does suggest that if such a process exists it is unlikely to widely require the exosome to degrade downstream flanking RNAs or to occur by a defined mechanism on every gene. Now, in this chapter the focus changes to the process of RNA-cleavage.

Given that co-transcriptional 5'→3' exonuclease activity of XRN2 is widespread at 3' flanking positions downstream of PAS it would, therefore, be

assumed that PAS-dependent RNA-cleavage, an essential prerequisite for exonuclease degradation, must also be required. However, earlier observations argued the opposite. EM micrographs of Pol II transcription on single genes, known as Miller spreads, and other RT-PCR techniques showed little RNA cleavage after the PAS under equilibrium conditions (Osheim et al., 1999; Osheim et al., 2002; Baurén et al., 1998). This has often been cited as evidence in support of direct allosteric termination in the absence of PAS cleavage. However, an equally likely alternative is that rapid exonuclease degradation means RNA-cleavage escapes detection in unperturbed conditions. In retrospect, the results in the chapters above and those by Fong et al. (2015) demonstrating XRN2 torpedo activity argue in favour of the latter interpretation. However, within *in vitro* purified systems, Pol II transcription across a PAS results in some limited termination in the absence of RNA-cleavage (Zhang et al., 2015). Whereas, *in vivo* analysis shows RNAi of CPSF73, the CPA complex endoribonuclease, and other CPA components causes widespread read-through transcription at protein-coding genes, yet this is usually over short distances of < 10 kb (Nojima et al., 2015). Thus, one question was whether these read-through polymerases with uncleaved PAS RNA (by RNAi of CPSF73) share the same characteristics as those that accumulate upon XRN2-AID depletion, i.e. was this limited read-through no longer because of a secondary cleavage-independent process. Another possibility is that (as for RNAi of XRN2) RNAi of CPSF73 leads to a partial depletion and therefore the limited read-through, in this case, equates with delayed cleavage.

5.1 Generation and sequencing of a CPSF73-AID rapid-depletion cell line

To determine whether the limited Pol II read-through upon CPSF73 KD was because of delayed cleavage from incomplete depletion or an auxiliary cleavage-independent mechanism, a *CPSF73-AID* cell line was sought. Such a cell line, like the *XRN2-AID* cells, would have the advantage of rapid and possibly more complete depletion over RNAi. Choosing to deplete the endonuclease CPSF73 instead of other CPA complex members means that some PAS recognition by the “polyadenylation module” of CPSF components (consisting of CPSF160, CPSF30, WDR33, Fip1) may still occur but RNA cleavage will not take place (Mandel et al., 2006; Schönemann et al., 2014). A full-length AID domain was used because it was feared that the three-tandem

mini-AID tag within the XRN2-AID cells might have been destabilising and responsible for XRN2-AID's lower expression level compared to endogenous untagged. Initial attempts to make a CPSF73-AID cell line were unsuccessful until TIR1 was placed under the inducible expression of doxycycline (Dox) with one cassette integrated at the AAVS1 safe-harbour locus (**Fig 5.1a;** Natsume et al., 2016). This is in contrast with XRN2-AID cells where multiple copies of a constitutively expressed TIR1 cassette are inserted by the sleeping beauty transposon system (Kowarz et al., 2015). This suggests partial recognition of CPSF73-AID by TIR1 in the absence of auxin may lead to a lethal scenario. Indeed, a western blot showing depletion of homozygous-tagged CPSF73-AID cells treated with Dox (18 h) and then auxin (3 h) also shows partial depletion after treatment with Dox (18 h) alone (**Fig 5.1b**). As mentioned in chapter 3, the newer AID system modifications by additional integration of a TIR1 binding component (such as Arf16b) or use of Auxinole may help in this scenario to reduce auxin-independent protein degradation (Li et al., 2019; Sathyan et al., 2019; Yesbolatova et al., 2019).

To gain insight into nascent transcripts that may be affected by CPSF73-AID depletion chromatin-associated RNA was sequenced (chrRNA-seq) as these species are highly enriched by the method (**Fig 5.2**). This approach has been used widely in transcriptional studies and captures transcripts produced from all three human polymerases (Wuarin and Schibler, 1994; West et al., 2008; Nojima et al., 2018a; Kamieniarz-Gdula et al., 2019). Briefly, this method begins by isolating whole nuclei before lysis and collection of an insoluble chromatin-pellet in a denaturing UREA buffer, which mirrors the beginning of the mNET-seq protocol. The RNA is then extracted from the chromatin-pellet using Trizol before short-read library generation using a random hexamer RT-PCR amplification approach (as detailed in section 2.7). Single-end sequencing was carried-out and bioinformatic processing involves adaptor trimming read mapping and read count normalisation. Although rRNA removal was performed as part of the library preparation using the Ribozero kit further bioinformatic removal of rRNA has not been performed but may be beneficial. An independent meta-analysis shows that some chrRNA-seq samples benefit from such an approach but for the samples used here the rRNA depletion has worked well and only a small proportion of rRNA reads remain (Tellier and Murphy, 2020).

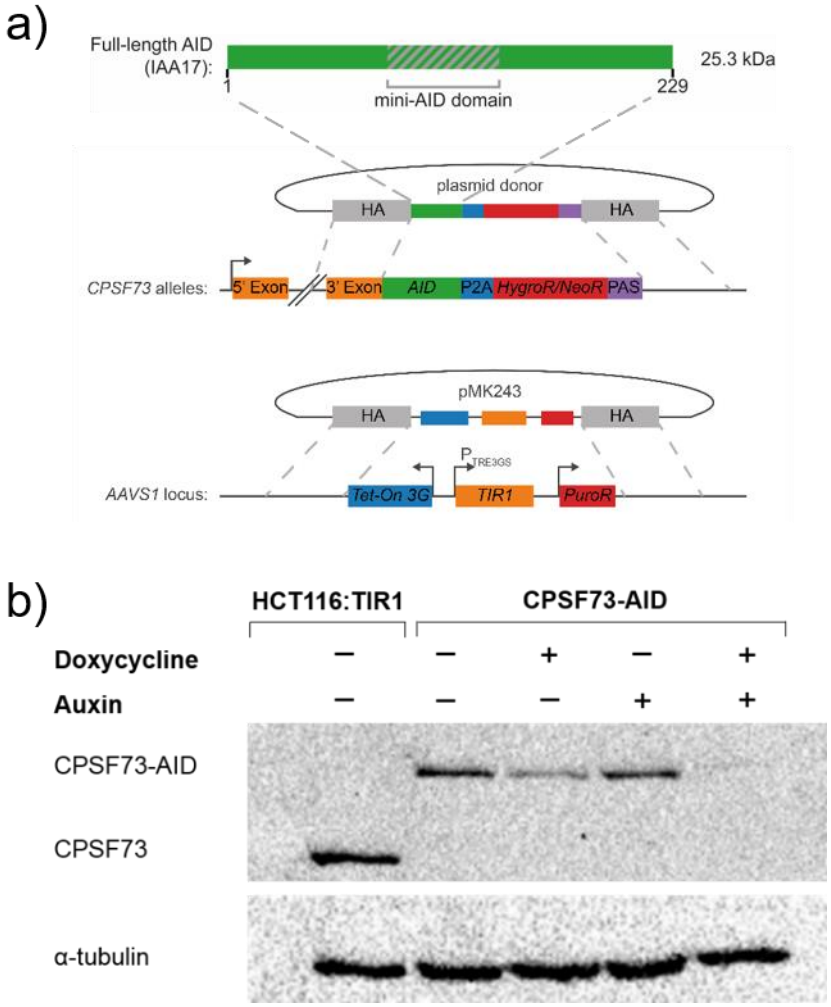


Figure 5.1 | A CPSF73-AID cell line for rapid protein degradation. **a**) [Top] Schematic of CPSF73-AID HDR plasmid repair constructs, shown with full AID tag domain structure above. Both hygroR and NeoR containing plasmids are co-transfected to enrich for biallelic insertions. [Bottom] TIR1 HDR construct for integration into the AAVS1 "safe-harbour" locus and under the inducible expression of Dox. Both rounds of HDR required co-transfection with the CRISPR/Cas9 plasmid (Sp. = humanised S. pyogenes) containing the appropriate sgRNA. **b**) WB of parental HCT116 cells, which is modified with AAVS1 Dox-inducible TIR1 cassette, and CPSF73-AID cells. The cells are untreated or treated with Dox (18 h), Auxin (18 h) and Dox (18 h) followed by auxin (3 h). The robust depletion of CPSF73-AID requires both Dox and auxin treatments. α-tubulin is used as a loading control. The WB was made by L.F. and reproduced from Eaton et al. (2020).

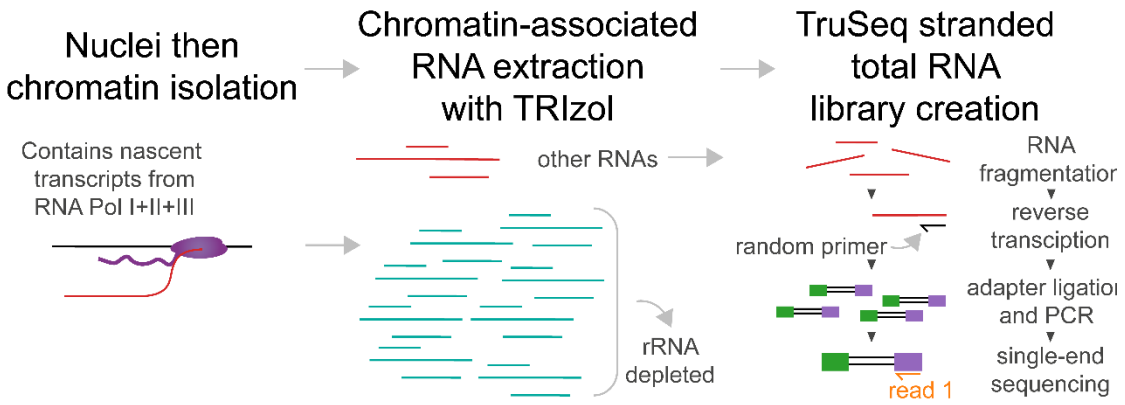


Figure 5.2 | Schematic of chrRNA-seq library preparation. **a**) An overview illustrating the main protocol steps in chrRNA-seq. The isolation of nuclei in hypotonic lysis buffer and sucrose cushion followed by insoluble chromatin pellet extraction from this using denaturing buffer mirrors the beginning of the mNET-seq protocol (left). Then, RNA is purified from the insoluble chromatin pellet and will therefore contain any transcript associated with chromatin included nascent transcripts from all three human RNA polymerases. The rRNA transcripts are then depleted to enrich for other transcripts (middle). The remaining RNA is used to create a cDNA library with the TruSeq stranded total RNA kit, which utilises a random primer for the reverse transcription step to randomised short-read fragments created (right).

A total of 4 samples were sequenced for 50 bp read length consisting of two replicates of control ethanol treated and Dox (18 h) followed by Auxin (3 h) treated. The average Phread quality scores by read position are all >30 and so demonstrates sufficiently good read quality from the Illumina HiSeq machine (**Fig 5.3a**). Raw reads then underwent adapter trimming with most trimmed sequences being short stretches of <5 bp (**Fig 5.3b**). This left trimmed reads with the majority > 45 bp in length and many remained full length (**Fig 5.3c**). A sample of these reads were then screened for contamination by alignment to various organism genomes including some common sources of contamination e.g. *E.coli* and *Mycoplasma* (**Fig 5.3d**). The majority of these reads aligned once or more only to the human reference genome with only a small proportion aligning to multiple genomes of the humans, mouse and CHO cell. These multiple genome reads are likely to originate from conserved eukaryotic sequences because very few reads aligned uniquely to the mouse and CHO cell genomes. All reads were then mapped to the human genome using HISAT2 with >90% aligning at least once (**Table 5.1**).

5.2 CPSF73-AID loss causes profound read-through transcription from protein-coding genes

After the loss of CPSF73-AID, the chrRNA-seq reveals a profound change in the transcription landscape with a lack of transcript expression definition. A 5 megabase zoomed-out snapshot on chromosome 1 shows that in control samples coverage is seen in distinct units covering the footprint of transcript annotation, on both sense and antisense strands. In contrast upon CPSF73-AID loss transcription continues to read-through and bleeds into neighbouring TUs (**Fig 5.4a**). When this effect is viewed on “isolated” protein-coding transcripts with no neighbouring transcripts downstream, read-through transcription can be seen extending hundreds of kb beyond the WT termination region (**Fig 5.4b,c**). The length of this read-through is similar to what might be expected from estimates the transcription elongation rate. These estimates vary for different transcripts but are on average ~2 kb·min^{-1} (Singh and Padgett, 2009; Fuchs et al., 2014). Cleavage will begin to be impaired from the beginning of the 3 h window of auxin addition but only more completely blocked towards the end. Therefore, by this estimation transcription over a 2-3 h period should extend to (120-180 mins *

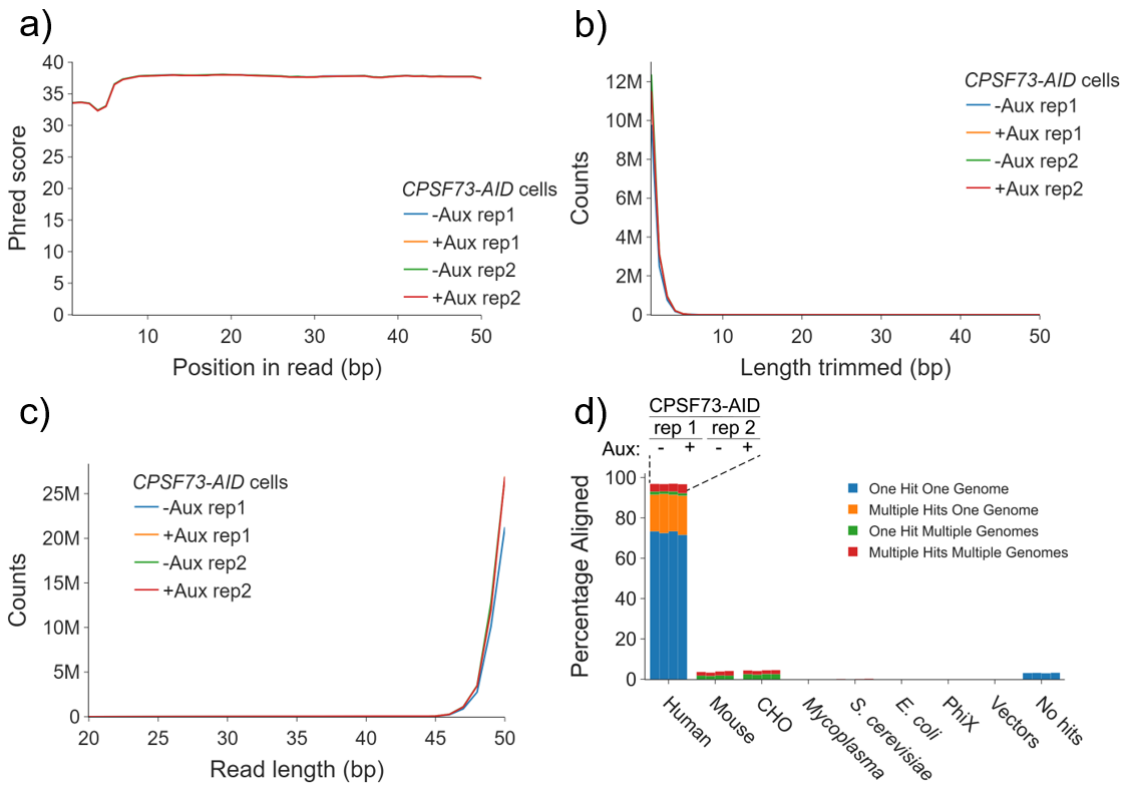


Figure 5.3 | Quality assessment graphs for CPSF73-AID chrRNA-seq samples. a) The average Phread quality score for each nucleotide position within the raw reads. **b)** The number of nucleotides removed from the 3'-end of raw reads by adaptor and quality trimming. **c**) The read length (bp) of reads after adaptor and quality trimming. **d**) Adapter trimmed reads were screened for contamination with 'fastq_screen'. A sample of 100000 reads, taken evenly throughout the read file, were aligned to various datasets of commonly used lab species and reagents. The key for the grouped bars is shown above human. The -Aux/+Aux conditions correspond to ethanol/Dox (18 h) treatment followed by ethanol/auxin (3 h) treatment, respectively.

Table 5.1 | Proportion of reads that map to human genome at least once for *CPSF73-AID* chrRNA-seq samples using HISAT2.

Samples	Total reads	Reads aligned to multiple sites	Reads aligned to unique sites	% of reads with a primary alignment
***CPSF73-AID* rep1 –Aux**	35291655	3495333	28791242	91.49
***CPSF73-AID* rep1 +Aux**	43863568	4216593	35798844	91.23
***CPSF73-AID* rep2 –Aux**	44634863	4390710	36505556	91.62
***CPSF73-AID* rep2 +Aux**	43772247	4304269	35335186	90.56

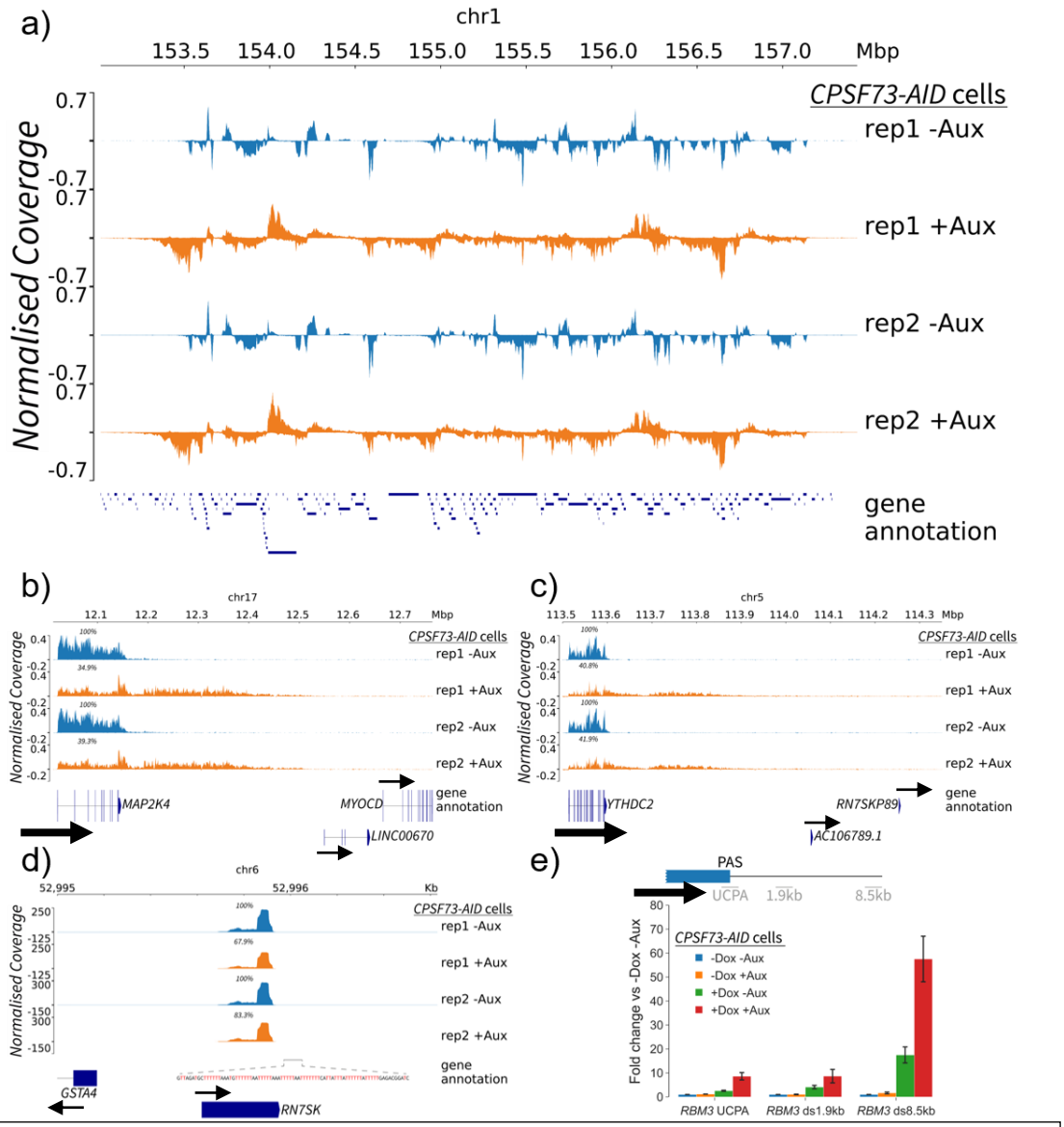


Figure 5.4 | CPSF73-AID loss causes profound read-through at protein-coding transcript. chrRNA-seq for CPSF73-AID depletion. The -Aux/+Aux conditions mean ethanol mock treatment or Dox (18 h) and then auxin (3 h) treatment. Each snapshot is drawn as 1000 bins of averaged signal. **a)** A large zoomed-out 5 megabase (Mbp) snapshot from chr1 shows a cluster of tightly packed transcription units. Blue bars for gene annotation represent annotated transcription units with splicing annotation omitted. **b,c)** A chromosomal snapshot of two protein-coding genes, MAP2K4 and YTHDC2. CPSF73-AID depletion leads to extended read-through of hundreds of kilobases from the transcription unit and a drop in expression level within the GB. Within gene annotation exonic and intronic annotations are shown with as a blue bar and a horizontal line, respectively. Only a single primary transcript isoform is shown for each transcription unit. **d)** *RN7SK*, A Pol III transcript, is shown with a downstream homopolymeric T-rich nucleotide sequence annotated (T's are coloured red). **b-d)** The % mean GB signal relative to the '-Aux' replicate is displayed above each sample. **e)** qRT-PCR of total cell RNA from CPSF73-AID cells treated with EtOH/Dox (18h) followed by EtOH/Aux (3h). Relative RNA levels downstream of *RBM3* protein-coding gene are given normalised to spliced *ACTB* (n=3; error bars=sem).

2 kb/min =) 240–360 kb. This is the case for both protein-coding transcripts in **Fig 5.4b,c** and therefore suggests that termination does not generally occur without PAS processing. This read-through transcription is of a much greater magnitude than that observed by RNAi of CPSF73 (Nojima et al., 2015). Such a difference may be due to incomplete depletion by RNAi which may delay RNA cleavage at the canonical PAS or occur at PAS sequences located downstream.

Additionally, CPSF73-AID depletion causes a decrease in GB signal for protein-coding transcripts, with 34.9-41.9 % remaining mean expression for *MAP2K4* and *YTHDC2* (**Fig 5.4b,c**). One possibly erroneous cause for this might be that because the traces are normalised (using TPM) for library size and there is a lot more signal over previously silent intergenic regions in the +Aux conditions this may equate to lower GB values as the sum of all bins is fixed at 1 million for each condition. However, Pol III transcripts, which intrinsically terminate at runs of T's and are CPSF73-independent, do not show as large a decrease in GB signal. This argues against a global underrepresentation of transcription signal in the '+Aux' condition being the cause (**Fig 5.4d**). The mean GB signal in the '+Aux' conditions for *RN7SK* is 67.9 & 83.3 % of the corresponding '-Aux' replicate. When the GB reductions in *MAP2K4* and *YTHDC2* are normalised against the smaller GB reductions from the *RN7SK* Pol III transcript it results in a >2-fold reduction in GB signal for these protein-coding transcripts. Similarly, Pol II ChIP-PCR of the slower depletion cell line, *CPSF73-DHFR*, shows a reduction in Pol II within the GB of *MYC* and *ACTB* protein-coding genes (Eaton et al., 2018). Such a decrease in GB signal is not unexpected as KD of another CPA complex protein, PCF11, results in a ~2-fold reduction in GB Pol II ChIP signal (Mapendano et al., 2010). Also, CPSF73 and other CPA complex components have been shown to bind RNA from promoters through the GB to 3'-ends of transcripts (Martin et al., 2012; Nojima et al., 2015). With the extended transcription into intergenic regions that occurs upon depletion of CPSF73, these Pol II-specific factors could be quenched through binding excess read-through RNA. It is possible that such a reduction of available Pol-II factors downregulates new Pol II transcription. The read-through transcription from protein-coding transcripts can be reproduced by qRT-PCR where total RNA is normalised to the predominantly cytoplasmic spliced β-actin transcript (**Fig 5.4e**). The read-

through is most pronounced after Dox (18 h) followed by auxin (3 h) treatment, but was also detectable after treatment with Dox alone. However, the sum of the fold-changes from auxin or Dox only treatments does not equate to a large accumulation as seen after dual treatment. This read-through correlates with the partial depletion of CPSF73-AID after Dox only treatment, as seen by western blotting suggesting that the read-through does indeed depend on the loss CPSF73-AID rather than indirect chemical effects (**Fig 5.1b**). These findings are a general effect for isolated protein-coding transcripts with the profound read-through extending >100 kb as shown by metagene profile (**Fig 5.5**). Notably, no read-through was present in the upstream antisense PROMPT direction and specific for the 3'-end termination of protein-coding transcripts. This suggests CPSF73 and the CPA complex cleavage is dispensable for PROMPT termination, whose transcripts accumulate on DIS3-AID depletion, so might signify an intrinsic or allosteric termination mechanism on these transcripts (Davidson et al., 2019).

This pattern of "run-away" transcription read-through and a decrease in GB signal can also be found at more clustered loci where multiple transcription events occur. For instance, the antisense read-through from *ERRFI1* can be seen extending and the GB signal decreasing upon CPSF73-AID depletion (**Fig 5.6a left**). Additionally, a lesser expressed convergent protein-coding transcript, *PARK7,* is within *ERRFI1* antisense read-through region and upon CPSF73-AID loss very little sense transcription is observed across the annotated transcript or downstream (**Fig 5.6a right**). This is an example of transcriptional interference where the transcription from one transcript negatively impacts the expression of a second occurring *in cis* (Greger and Proudfoot, 1998; Shearwin et al., 2005). These convergent Pol II collisions are commonplace throughout the genome with depletion of CPSF73-AID. A slightly different scenario is where the read-through extends into a transcript transcribed in tandem (in the same direction). One example is *WDR44* read-through extends into *IL13RA1* (**Fig 5.6 b**). The decrease in GB expression within *IL13RA1*, upon CPSF73-AID loss, is not as dramatic as for *PARK7* and is more in line with the relative fold-change observed across GBs of other isolated transcripts. One reason for this may be that *WDR44* and *IL13RA1* are much more closely matched in expression level, whereas there is a larger difference

in *ERRFI1* and *PARK7* expression level in untreated conditions. At very tightly clustered loci, like that on chr14 with alternating sense and antisense transcription (on PRPF39, FPBP3, FANCM and MIS18BP1), CPSF73-AID loss results in the decrease signal magnitude and loss of transcript definition with both sense and antisense coverage present across all transcripts (**Fig 5.6c**). The transcriptional 'haze', which is like the expression of open chromatin, shows the acute consequences of improper transcription cleavage and termination and likely means loss of encoding complete and cogent messages from these transcripts.

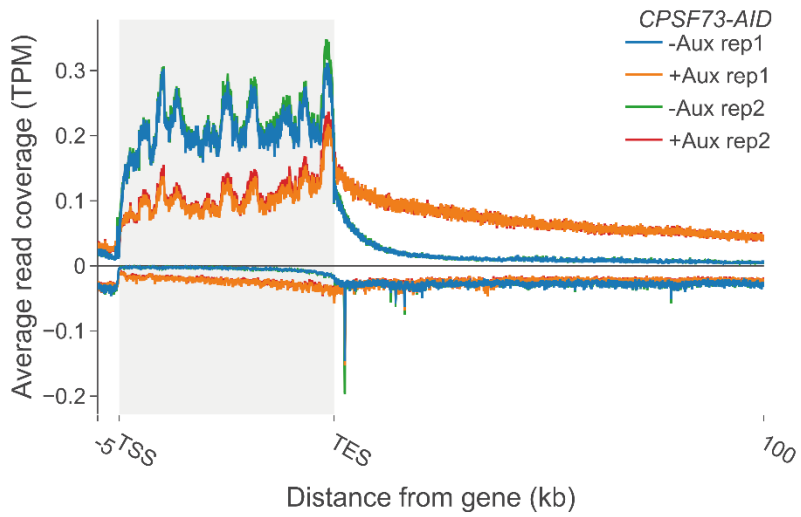


Figure 5.5 | The runaway read-through at protein-coding transcripts upon CPSF73-AID loss is general and does not impact PROMPTs. A metagene profile of chrRNA-seq for CPSF73-AID depletion at expressed isolated protein-coding transcripts. The averaged signal 100 kb downstream of the annotated end of the transcript (TES) and the 5kb upstream of each transcript is shown. N=785 transcripts. The selection criteria for an isolated protein-coding transcript was one that had no other expressed transcript present within the -Aux condition of the -5 kb to +100kb range. The -Aux/+Aux conditions mean ethanol mock treatment or Dox (18 h) and then auxin (3 h) treatment. The highlighted grey region between TSS and TES represents the GB segment which is scaled to 50 kb.

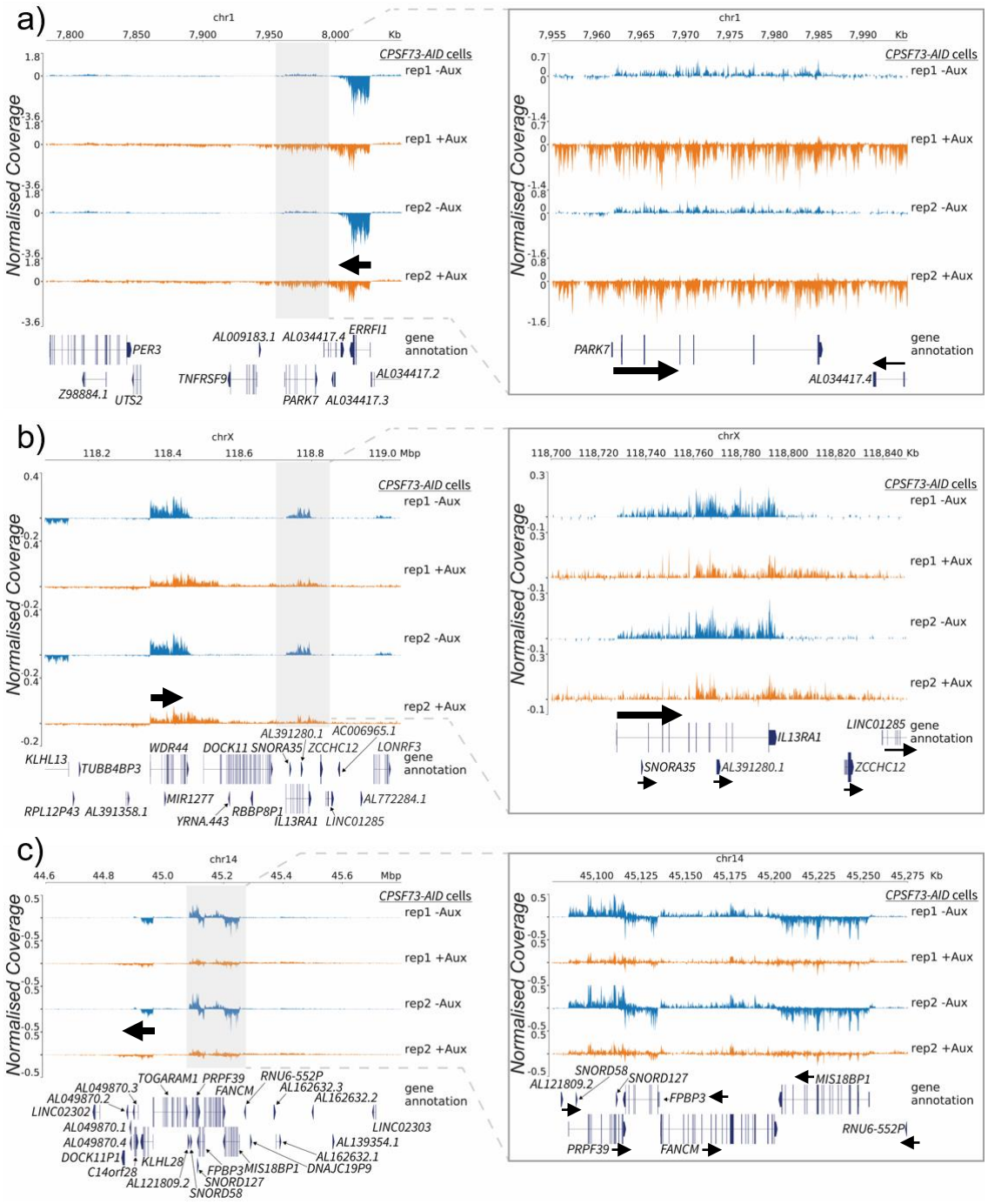


Figure 5.6 | CPSF73-AID dependent read-through effects on neighbouring transcription units. a) Antisense expression from ERRFI1 causes read-through of hundreds of kb when CPSF73-AID is depleted (left-panel). This read-through extends through a downstream convergent protein-coding gene PARK7 shown as a zoomed-in snapshot (right-panel). **b)** CPSF73-AID loss causes WDR44 read-through transcription into downstream protein-coding gene IL13RA1, which also transcribes in the same sense direction. **c)** A tight clustering of protein-coding transcript expression in alternating sense and antisense directions (from PRPF39, FPBP3, FANCM and MIS18BP1) is shown in -Aux. The definition of expression is lost upon CPSF73-AID depletion with both sense and antisense signal covering these zoomed-in transcripts (right-panel).

5.3 Effects of CPSF73-AID loss at short transcripts

As mentioned previously, replication-dependent histones are a different class of protein-coding transcript that do not use PAS or produce 3' polyadenylated mRNA. Instead, histone mRNAs end with a 3' stem-loop formed after processing of a histone dependent element (HDE) in a U7 snRNA-dependent manner (Chodchoy et al., 1991; Marzluff and Koreki, 2017). The HCC shares three components CPSF100, Symplekin and the endonuclease CPSF73 with that of the PAS processing CPA complex (Dominski et al., 2005; Kolev and Steitz, 2005; Kolev et al., 2008; Sullivan et al., 2009). However, the downstream cleavage product for histones showed little effect upon XRN2-AID depletion, whereas protein-coding genes produced read-through transcription (**Fig 3.5&3.7**). This difference on XRN2 dependence may be because CPSF73 (like some other related metallo β-lactamase enzymes) can have dual endo- and exo-ribonuclease activities and the exonuclease activity may only be active within the HCC (Yang et al., 2009; Yang et al., 2020). Another possibility is that a newly discovered RNA endonuclease, MBLAC1, from the same metallo β-lactamase family has been shown to have histone pre-mRNA processing activity and so could constitute a separate XRN2-independent termination pathway (Pettinati et al., 2018). Therefore, histone loci within the CPSF73-AID chrRNA-seq dataset were examined to investigate the dependence of histone pre-mRNA on CPSF73-AID. Unfortunately, many histone transcripts within the chrRNA-seq data are partially obscured by confounding read-through from neighbouring protein-coding transcripts. Upstream of the HIST1H cluster TRIM38, a protein-coding transcript, reads into many histone transcripts that then confounds interpretation (**Fig 5.7a**). Likewise, *LINC00869* upstream and *SF3B4* downstream of the HIST2H cluster have read-through upon CPSF73-AID loss that completely obscures the locus (**Fig 5.7c**). The clearest example available is *HIST1H3E* were expression is relatively high in -Aux control condition and upstream read-through in +Aux condition is lower than the transcription interference for many other histone transcripts (**Fig 5.7b**). Upon CPSF73-AID loss read-through can be seen extending ~70kb downstream into a previously silent region. To determine if this read-through is from HIST1H3E transcription events a time-course of shorter auxin depletion times may reveal extending read-through before upstream read-through reaches HIST1H3E or

inhibition of transcription using triptolide/ flavopiridol could be used to chase away

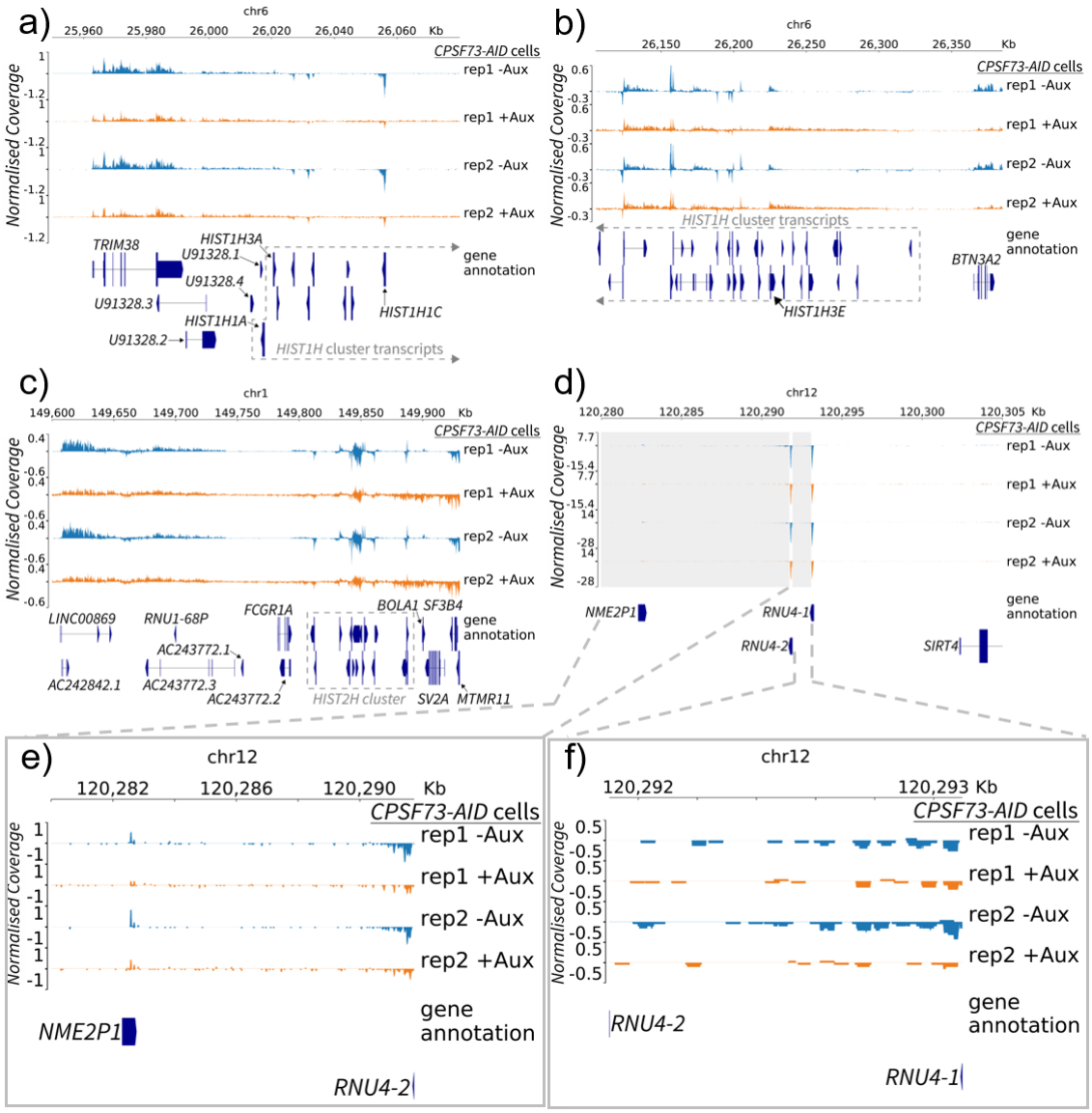


Figure 5.7 | CPSF73-AID depletion effects at histone and snRNA transcription units. Chromosomal snapshots from CPSF73-AID chrRNA-seq. **a,b)** Snapshots of the left and right ends of the HIST1H histone cluster on chr6. **c)** Snapshot of the HIST2H histone cluster on chr1. **d)** snRNAs RNU4-1 and RNU4-2 show little difference in the abundance of chromatin-associated transcript after CPSF73-AID loss. **e,f)** A zoomed-in snapshot of the downstream signal from RNU4-2 and RNU4-1, respectively.

transcription to see if the signal reduces from the promoter of HIST1H3E or one upstream. However, both experiments would likely be noisy, may generate an uncertain result and would only give insight into this one histone transcript that may not be typical of the rest.

Another short transcript class, snRNAs, also have little XRN2-AID dependent effect (**Fig 3.8**) but have been shown to have a small increase in downstream Pol II by ChIP upon dominant-negative XRN2-MT overexpression (Fong et al., 2015). However, 3'-end cleavage/processing of snRNAs occurs via a different metallo β-lactamase endonuclease, INTS11 of the integrator complex (O'Reilly et al., 2014). Accordingly, CPSF73-AID depletion leads to little change in *RNU4-1* and *RNU4-2* abundant GB signal (likely from mature transcript) or any difference in downstream transcription signal (**Fig5.7d-f**).

5.4 Effects of CPSF73-AID loss at long non-coding transcripts

Two lncRNA transcripts that have a unique termination mechanism are *MALAT1* and *NEAT1*. They have a mascRNA tRNA-like element at their 3'-end that is excised by RNase P/Z cleavage (Wilusz et al., 2008). Previously in section 3.6, results demonstrated that the downstream product of this RNase P/Z cleavage was co-transcriptionally degraded by XRN2 because XRN2-AID loss increased downstream signal by mNET-seq. Correspondingly, CPSF73-AID loss shows no profound read-through downstream of these transcripts by chrRNA-seq (**Fig 5.8a**). This shows CPSF73 is dispensable for their termination and likely means XRN2 can additionally degrade the products of a non-CPA cleavage events to stimulate termination. At other lncRNA transcripts, such as TUG1 and NORAD, CPSF73-AID loss causes a profound read-through at these transcripts highlighting their termination is dependent on CPSF73 (**Fig 5.8b,c**). This differs from observations by CPSF73 RNAi where no defect is present even though such KD is sufficient to cause read-through on protein-coding transcripts (Schlackow et al., 2017). One possible explanation for this disparity might be due to the efficacy of depletion. That is because the defect observed at protein-coding transcripts by CPSF73-AID RNAi produces extended read-through that eventually stops within ~10 kb, whereas CPSF73-AID depletion causes >100 kb read-through suggesting PAS cleavage is only delayed by incomplete depletion of CPSF73 in the former. It is possible TUG1 has a particularly efficient PAS that means its termination is unaffected by RNAi

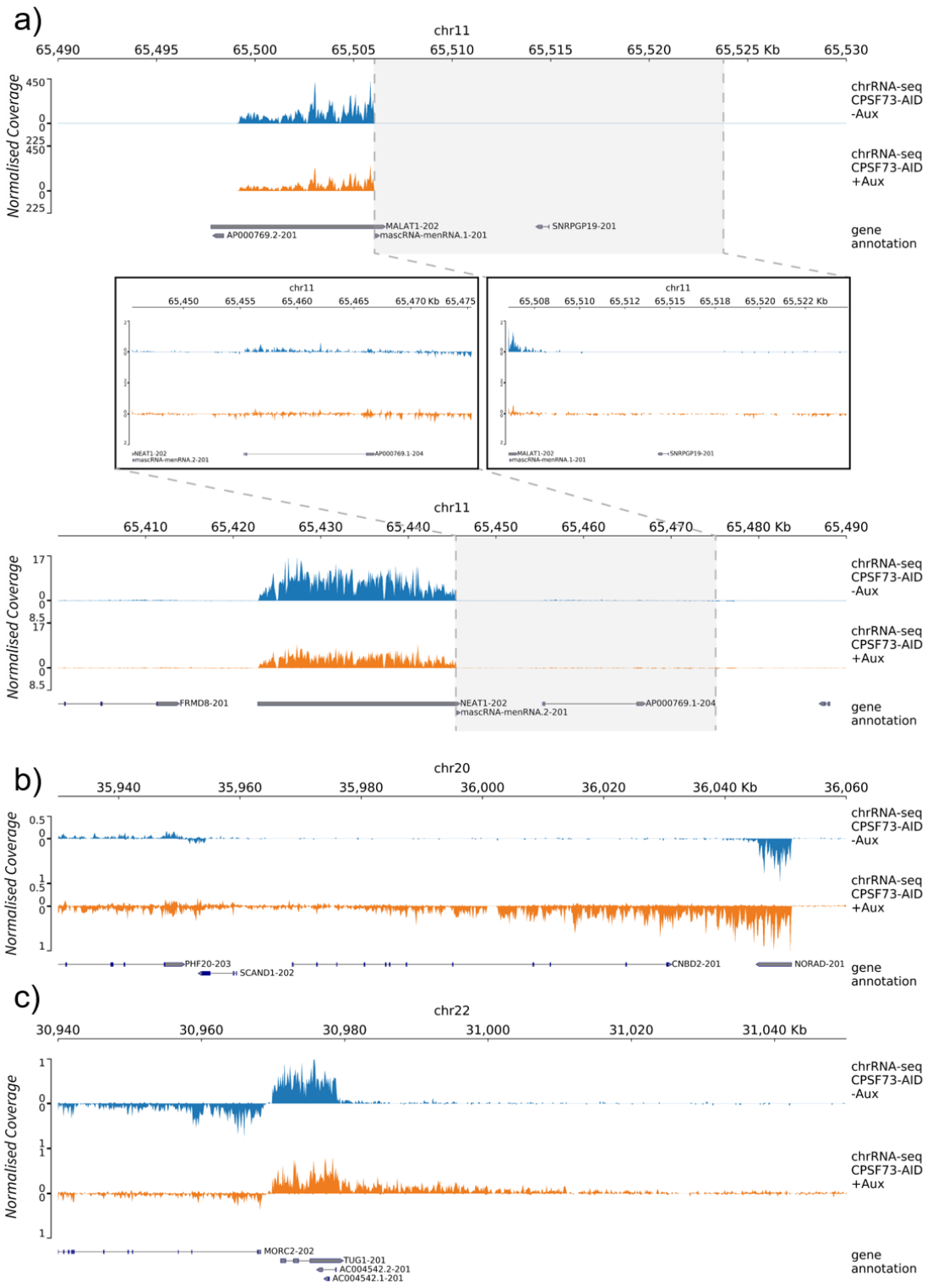


Figure 5.8 | CPSF73-AID depletion effects at non-coding RNAs. Chromosomal snapshots from CPSF73-AID chrRNA-seq. **a)** Snapshots of *MALAT1* and *NEAT1* non-coding RNAs whose transcripts end with a mascRNA element which is excised by RNase P/Z cleavage. Inset boxes show a zoomed-in segment downstream to enrich the nascent transcription signal. **b,c)** Snapshots of NORAD and TUG1 transcripts, respectively. These are two long non-coding transcripts and are dependent on CPSF73 for termination because profound read-through is observed on CPSF73-AID.

of CPSF73. Similarly, the read-through caused by CPSF73-AID depletion on TUG1 is shorter in magnitude than other transcripts but still visible at ~80 kb downstream (**Fig 5.8c**). Another possibility is that TUG1 termination is partially redundant with an as yet unidentified auxiliary pathway, but given the current data this seems unlikely.

5.5 Co-transcriptional RNA-cleavage per se does not always ensure Pol II transcription termination

So far, the chrRNA-seq shows CPSF73-AID loss causes a failure of PAS-dependent transcripts to undergo transcriptional termination. Whereas, XRN2-AID loss causes shorter read-through transcription and a pile-up of Pol II downstream of PAS-dependent transcripts. This pile-up of Pol II reveals XRN2 is part of the main termination pathway and these polymerases are already committed to a slowed elongation rate. Lots of the Pol II transcript classes discussed throughout (including PAS-dependent, snRNAs, histones, microRNA host genes and MALAT1/NEAT1) utilise an RNA-cleavage event during 3'-end processing. Furthermore, RNA cleavage (by RNaseH) of nascent transcripts causes Pol II arrest *in vitro* (Újvári et al., 2002). Therefore, does co-transcriptional RNA-cleavage itself commit Pol II to a slowed elongation and termination fate? To investigate this, the self-cleaving δRZ sequence which is inserted downstream of a PAS of *RBM3* within the *CPSF73-AID* cell line was re-examined (previously introduced in section 4.3). This cell line allows endogenous PAS cleavage to be substituted for a rapid artificial cleavage event that does not support XRN2 degradation (and subsequent torpedo termination) due to the 5'-hydroxl end generated. Interestingly, in this setting δRZ cleavage does not rescue Pol II termination when CPSF73-AID is depleted by auxin addition (orange/red vs brown at 8.5kb and 11kb; **Fig 5.9b**). As previously mentioned, it is known that the δRZ can cleave within the RBM3 3' flank because upstream degradation occurs from the 3'-end cyclic 2',3'-monophosphate after auxin addition within the CPSF73-AID cells presumably by DIS3 (orange/red vs brown at UCPA and 1.1kb). Thus, it appears RNA-cleavage of a nascent transcript in isolation is insufficient to induce elongation slowing after PAS transcription.

It was then determined whether an XRN2 compatible CPA-independent cleavage event could support transcription termination in the absence of PAS

cleavage. That is, is it enough for liberation of a 5'-phosphorylated RNA downstream of protein-coding transcripts to provide an entry site for XRN2 to instigate transcription termination? The non-coding transcripts *MALAT1* and *NEAT1*, which terminate in a XRN2-dependent (see **Fig 3.9d,e**) and CPSF73-independent (**Fig 5.8a**) manner, undergo RNase P and RNase Z cleavage at a mascRNA "tRNA-like" structure (Wilusz et al., 2008). The *MALAT1* termination sequence (including triple helix and mascRNA sequences) has previously been inserted downstream of an intronless *β-globin* cassette and was successfully cleaved *in vivo* producing a stabilised transcript (Brown et al., 2012). Therefore, the *MALAT1* termination sequence was inserted ~400 bp downstream of a protein-coding transcript PAS in *CPSF73-AID* cells. The same ouabain co-selection HDR CRISPR/Cas9 strategy used for insertion of the δRZ and xrRNA sequences was repeated to insert the *MALAT1* element. The *MORF4L2* transcript was chosen as it is on the x chromosome (so monoallelic in male HCT116 cells) and shows a typical protein-coding phenotype with XRN2-AID depletion causing Pol II accumulation extending < 10 kb and CPSF73-AID depletion causing prolonged read-through of over >100 kb (**Fig 5.10a**). In the absence of auxin qRT-PCR shows the *MALAT1* sequence causes a small increase in upstream amplicons (with no such increase at downstream positions) indicating that RNaseP cleaves because correct folding of the triple helix sequence, which confers the stabilisation effects, depends upon the triple helix being terminally situated at the 3'-end of the transcript (blue vs green at UCPA and 200bp; **Fig 5.10b**; Brown et al., 2012). Secondly, this accumulation of upstream amplicons (in '*MALAT1* -Aux') also indicates that *MALAT1* sequence processing occurs faster than PAS processing because XRN2 has not completely masked the accumulation of these transcripts. Also, downstream RNA read-through that accumulates on CPSF73-AID depletion is abolished by insertion of the *MALAT1* sequence (orange vs red at 2.7kb, 4.2kb, 8.5kb; **Fig 5.10b**). The decrease not only causes a reduction in RNA but also rescues Pol II termination as seen by Pol II ChIP (orange *vs.* red at 2.7kb, 4.2kb, 8.5kb) (**Fig 5.10c**). This suggests that the liberation of a 5'-phosphate during 3'-end RNA-cleavage is key to instigating termination at PAS-dependent transcripts by providing an entry site for XRN2 to act as the molecular torpedo. However, it remains unclear whether such a repurposed cleavage event can

induce Pol II elongation slowing as revealed by XRN2-AID depletion at PAS-dependent transcripts.

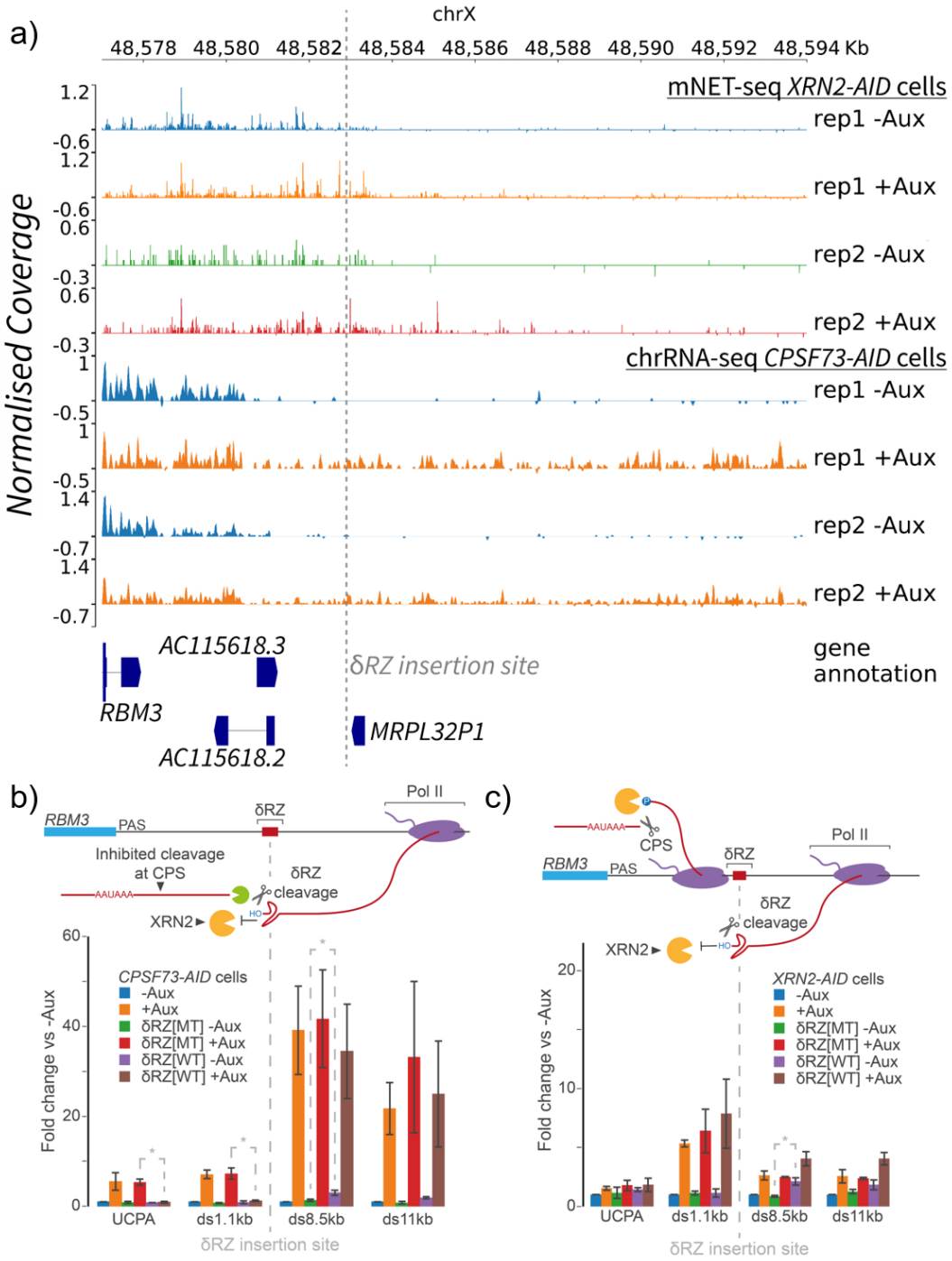


Figure 5.9 | Ribozyme cleavage does not rescue termination when CPSF73 is lost. Reproduced from Figure 4.7 but described in the context of cleavage implications. **a)** Chromosomal snapshot showing *XRN2-AID* mNET-seq and *CPSF73-AID* chrRNA-seq auxin-dependent read-through downstream of *RBM3*. **b,c)** Analysis of *CPSF73-AID* (**b**) or *XRN2-AID* (**c**) cells unmodified and modified downstream of *RBM3* by insertion of a hepatitis δ ribozyme (δRZ) with WT or inactivating point mutant (MT). **b)** qRT-PCR of total cell RNA normalised from *CPSF73-AID* cells to spliced *ACTB* shown as a relative fold change to '-Aux'. Cells are treated with neither or both Dox (18 h) followed by auxin (3 h) and labelled '-Aux' or '+Aux' respectively (n=3; error bars=sem). **c)** qRT-PCR of total cell RNA from *XRN2-AID* cells normalised to spliced *ACTB* shown as a relative fold change to '-Aux'. Cells are treated or not with auxin (2 h; n=3; error

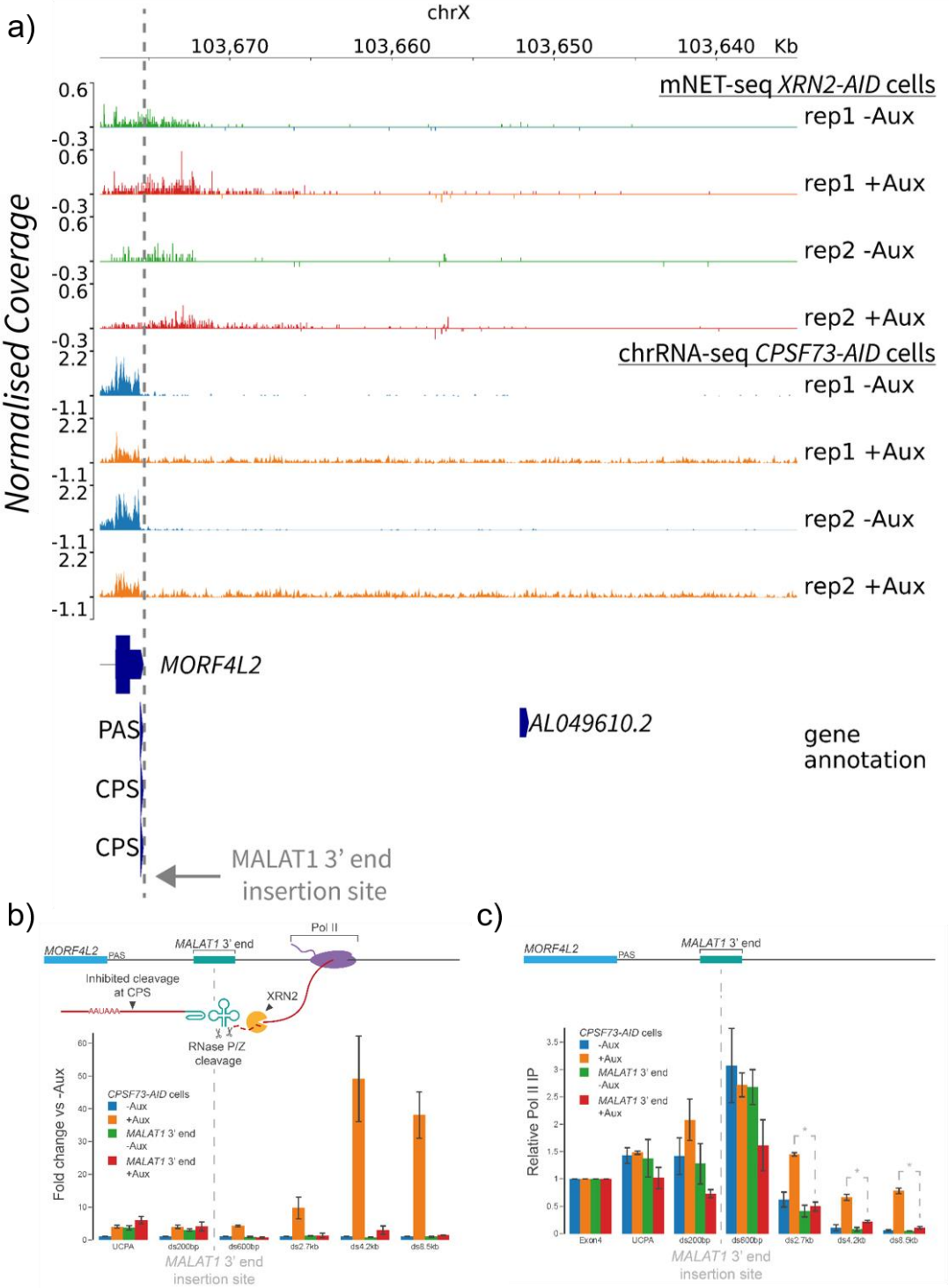


Figure 5.10 | *MALAT1* 3'-end insertion downstream of a PAS rescues termination in *CPSF73-AID* cells. a) Chromosomal snapshot showing XRN2-AID mNET-seq and CPSF73-AID chrRNA-seq auxin-dependent read-through downstream of *MORF4L2*. **b,c)** Analysis of CPSF73-AID cells unmodified and modified downstream of *MORF4L2* by insertion of the *MALAT1* 3'-end sequence. Cells are treated with neither or both Dox (18 h) followed by auxin (3 h) and labelled '-Aux' or '+Aux' respectively. **b)** qRT-PCR of total cell RNA normalised to spliced *ACTB* shown as a relative fold change to '-Aux' (n=3, error bars=sem). **c)** Pol II ChIP-PCR (8WG16) of *MORF4L2* with % of input normalised to GB signal of the exon4 amplicon (n=3, error bars=sem).

5.6 PP1 phosphatase activity assists Pol II termination by XRN2

Whilst the MALAT1 element showed an exogenously-directed RNA cleavage event can replace PAS cleavage and effectively restore Pol II termination some other artificially directed RNA cleavage events, namely the δRZ cleavage, are incapable of restoring Pol II termination in the absence of PAS cleavage. Thus, the difference between the long read-through seen by CPSF73-AID loss and the Pol II arrested downstream of PAS-dependent genes seen by XRN2-AID loss cannot then be fully explained by any RNA cleavage in isolation and is likely linked to a downstream activity triggered by the CPA complex upon cleavage.

Studies in *S. pombe* have revealed the phosphatase, Dis2, causes a slowing in the Pol II elongation rate by dephosphorylation of Spt5 and Pol II CTD (Kecman et al., 2018; Parua et al., 2018). Other studies in *S. cerevisiae* show the Dis2 orthologue, Glc7, is responsible for phosphorylation-based switch at transcript 3'-ends leading to transcription termination (Nedea et al., 2008; Schreieck et al., 2014). It was therefore wondered whether the polymerases that accumulate upon XRN2-AID depletion do so because of a similar phosphorylation-based switch method. In humans, mass spectrometry of the CPA complex identified isoforms orthologous to Dis2 and Glc7, as PP1α and PP1β (Shi et al., 2009). Using *XRN2-AID* cells, KD of the PP1 isoforms in conditions α, β and α+β leads to a corresponding reduction in both the mRNA and protein (**Fig 5.11a**). The effect of PP1 isoform KD with and without XRN2-AID depletion on transcription read-through was analysed using qRT-PCR of total cellular RNA (**Fig 5.11b**). PP1 KD in the presence of XRN2 did not result in extended read-through. In contrast, when combined with XRN2-AID depletion, read-through extends further than XRN2-AID depletion alone at *ACTB*, *MYC* and *YTHDF3*. The effect varies in magnitude: being largest for joint α+β KD at *YTHDF3* and varies in the contribution from each isoforms as can be seen for *MYC*, which is almost solely reliant on PP1α KD. Whereas, for *YTHDF3* there is an effect with KD of either isoform that increases upon joint KD. This suggests that transcripts may undergo phosphatase-mediated pausing at different rates and with a different reliance on PP1 isoforms.

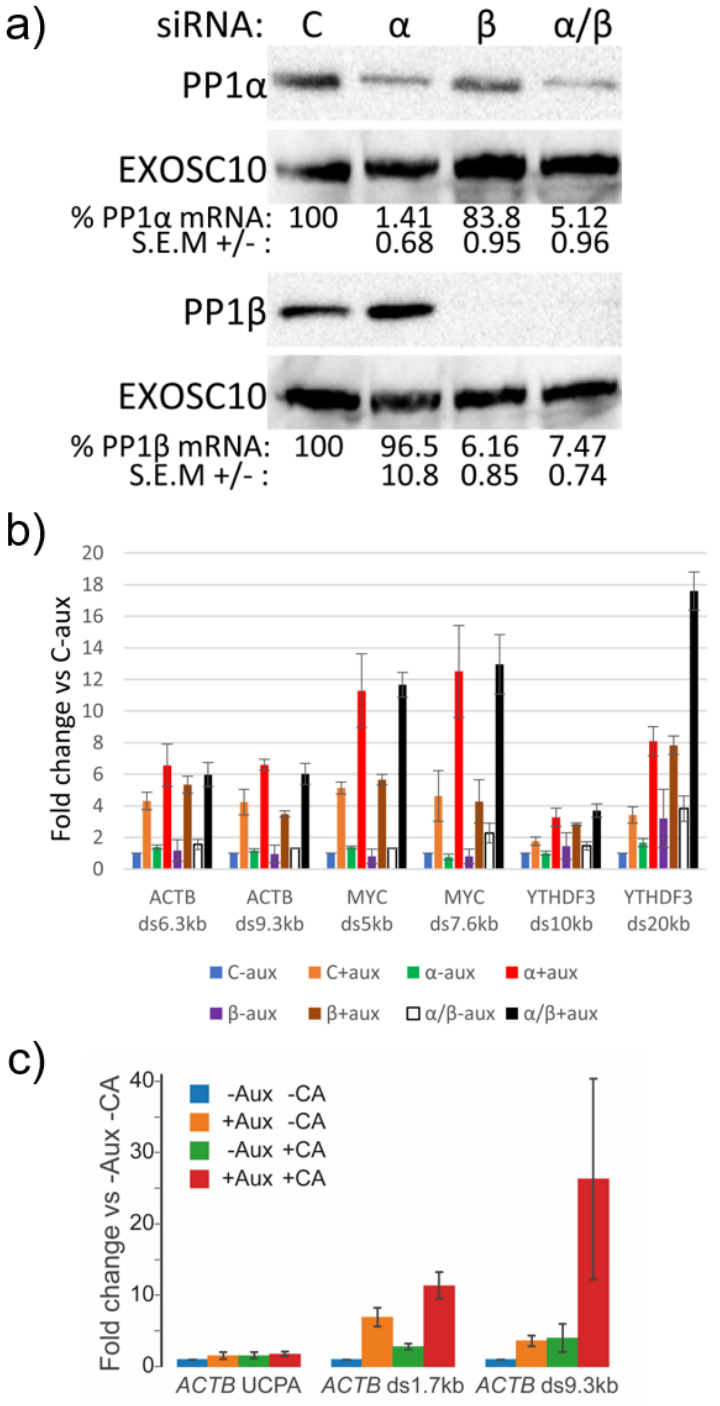


Figure 5.11 | The piling-up of Pol II after XRN2-AID depletion is dependent on PP1 phosphatase activity. **a)** WBs of *XRN2-AID* cells transfected twice with siRNAs targeting control (C), PP1α, PP1β and PP1α+PP1β (24 h + 48 h). EXOSC10 is probed as a loading control. qRT-PCR of spliced mRNA levels (n=3) is normalised to spliced *ACTB* and shown as a relative % of control siRNA treated. WB and qRT-PCR performed by S.W. and reproduced from Eaton et al. (2020). **b)** qRT-PCR of *XRN2-AID* cells treated with C, PP1α, PP1β and PP1α+PP1β siRNAs (72 h) followed by auxin treatment or not (2 h). The fold change shown relative to 'C-Aux' treated and normalised to spliced ACTB (n=3; error bars=sem). **c)** qRT-PCR of *XRN2-AID* cells treated with mock, auxin, Calyculin A (CA) or both with all treatments occurring for 1 h. The fold change is shown relative to mock treated and normalised to spliced ACTB (n=3; error bars=sem).

One disadvantage of using RNAi is that the reduction in PP1 protein levels may cause reduced CPA complex stability that would impair cleavage and explain the extended read-through. However, whilst unknown in humans this argument is does not to apply in budding yeast where Glc7 depletion does not affect CPF complex stability (orthologous to CPA complex in humans; Schreieck et al., 2014). To exclude the possibility that PP1 loss causes disruption of associated complexes, a PP1-phophatase inhibitor was used as an alternative strategy. Calyculin A (CA) is a potent protein phosphatase inhibitor known to target both PP1 and PP2A (Resjö et al., 1999). qRT-PCR was performed on *XRN2-AID* cells untreated or treated with auxin, CA, auxin and CA (**Fig 5.11c**). All treatments including auxin were performed for the shorter time of 1 h because longer CA treatment affected cell morphology. The treatment with both auxin and CA extends read-through transcription at *ACTB* further than auxin or CA treatment alone. Taken together the PP1 siRNA and inhibitor experiments with previous studies in yeast suggest that PP1 mediates a Pol II elongation slow down after the PAS of polymerases that are then targeted by XRN2. The elongation rate slow down likely functions to expedite Pol II capture by XRN2 and facilitate termination.

5.7 Discussion

The results above show the profound effect depleting CPSF73-AID has on transcription termination. Its magnitude is commonly hundreds of kilobases for well-expressed isolated protein-coding transcripts (see **Fig 5.4&5.5**). This is approximately equal to the distance that Pol II travels during the 3 hr period of depletion, based on estimates of its elongation speed (Singh and Padgett, 2009; Fuchs et al., 2014). This suggests a total failure of Pol II termination. Similar results have been observed on a handful of protein-coding transcripts when others in the lab tagged CPSF73 with dihydrofolate reductase (DHFR), a different degron system that takes longer (~10 h vs. 1-2 h) to deplete sufficiently (Eaton et al., 2018). Interestingly, in the CPSF73-DHFR cell line, overexpression of a catalytically inactive point mutant (H73A) termed CPSF73-MT, did not rescue prolonged read-through suggesting it is truly CPSF73 endonuclease activity that is responsible for triggering termination and not solely PAS recognition by CPA complex binding. However, this point mutant, H73A, only partially reconstitutes complex formation to endogenous levels,

which should be considered as a caveat (Kolev et al., 2008). As the overexpression of CPSF73-MT did not even partially diminish the read-through distance or magnitude in the CPSF73-DHFR cell line this seems unlikely.

In chapter 4 the insertion of the δRZ sequence downstream of the RBM3 PAS highlighted that XRN2 targets Pol II complexes that are stalled or slowed (section 4.3). The loss of XRN2 in this context leads to accumulation of Pol II both downstream and upstream of the δRZ insertion site. However, δRZ cleavage produces an upstream product that is a substrate for rapid degradation (Muniz et al., 2015) and this was not observed here but was in CPSF73-AID δRZ modified cells. Given δRZ cleavage within CPSF73-AID cells does not prevent profound read-through it suggests that PAS cleavage must precede pausing onset, and this is corroborated by PP1/PNUTs coprecipitation with other CPA complex components (Shi et al., 2009).

Further to this, insertion of the MALAT1 mascRNA element downstream of MOR4L2 PAS does rescue profound read-through on depletion of CPSF73-AID indicating that some cleavage events generating a 5'-phosphate can induce efficient termination via an XRN2 pathway. Others in the lab and other groups have recently demonstrated that cleavage of nascent RNA by antisense oligo directed RNaseH-NLS is capable to inducing premature termination via XRN2 degradation (Eaton et al., 2020; Lai et al., 2020; Lee and Mendell, 2020). As RNase H cleavage generates a downstream 5'-phosphate this likewise confirms its requirement of torpedo termination in eukaryotes. Whether these artificially directed cleavage events can induce Pol II pausing in an analogous manner to CPA cleavage remains to be determined. However, RNaseH cleavage of nascent transcripts does cause Pol II arrest *in vitro* (Újvári et al., 2002).

The molecular process of Pol II pausing in yeast is mediated by Dis2 phosphatase activity dephosphorylating specific residues of Spt5, a component of DSIF (Parua et al., 2018; Kecman et al., 2018). Here data demonstrates that the human orthologue, PP1, mediates this elongation slow down with small molecule inhibition or PP1 RNAi leading to extended read-through auxin treated XRN2-AID cells (see **Fig 5.11**). A recent paper by Cortazar et al. (2019) demonstrated that in humans cells, PP1 dephosphorylates residues on Spt5, their removal correlating with pausing downstream of the PAS. PAS mutation is unable to induce pausing until transcription extends past downstream cryptic

PAS sequences. The hyperphosphorylation of Spt5 occurs at elongation release by CDK9 subunit of P-TEFb. Thus, in this model kinase-phosphatase networks compete to regulate the elongation processivity of Pol II through phosphorylation. This phosphorylation-induced conformational changes observed are akin to an allosteric transition as predicted by torpedo model.

It is worth mentioning that Xrn2 and Pol II are predicted to degrade and transcribe, respectively, at similar rates (~2kb/min) (Hoek et al., 2019). It is therefore advantageous to slow Pol II in order to favour its capture by XRN2. Although dephosphorylation of Spt5 goes some way to explaining how this occurs, it may also be the case that other relevant PP1 targets remain to be discovered. Nevertheless, the work in this thesis unifies the early allosteric and torpedo models into a single mechanism. This provides a clear conceptual framework to understand how transcription terminates on protein-coding genes (**Fig 5.12**).

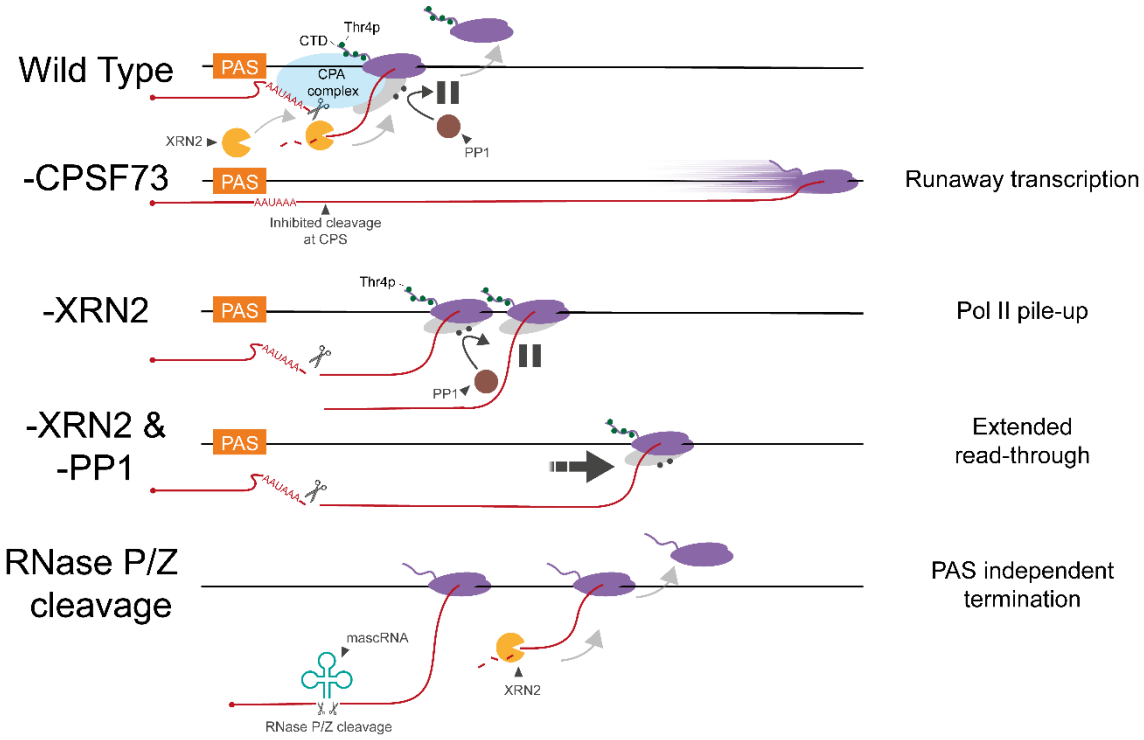


Figure 5.12 | Model of Pol II transcriptional termination at the 3'-ends of genes. (Wild Type) Pol II transcription across the PAS causes recognition and cleavage by the CPA complex. This induces Pol II pausing, which is mediated by PNUTs/PP1 phosphatase activity on Spt5 residues and other targets. This stalled complex is a target for torpedo termination via processive XRN2 degradation of the nascent RNA. This model involves both allosteric changes to the Pol II complex and a torpedo termination mechanism. (-CPSF73) Cleavage is inhibited and profound "runaway" transcription ensues. (-XRN2) The loss of XRN2 degradation causes the accumulation or pile-up of stalled complexes. (-XRN2 & -PP1) The loss of both XRN2 and PP1 causes extended read-through into distal regions (compared to -XRN2 alone). (RNaseP/Z) The cleavage and excision of a mascRNA element by RNase P/Z provides a PAS-independent entry site for XRN2 degradation and torpedo termination.

5.8 Conclusions

- XRN2 terminates transcription on most protein-coding genes and at a number of non-coding loci.
- It does so by targeting the 5' phosphate following transcript cleavage (most commonly via CPSF73 downstream of the PAS).
- It functions by chasing Pol II down and this process is facilitated by mechanisms that slow down polymerases after the poly(A) site has been processed.
- Polymerase slowing is facilitated by PP1 phosphatase activity.
- At protein-coding genes, all of the above require the activity of CPSF73, the nuclease that cuts downstream of the PAS. Without CPSF73 Pol II engages in runaway read-through transcription.

5.8.1 Future works

The data presented here provides strong evidence for the widespread role of XRN2 in degrading the downstream products of PAS cleavage. However, one of the most enigmatic puzzles surrounding termination is what happens to Pol II when it is "caught" by XRN2. *In vitro* systems that can recapitulate termination would be useful to investigate this. Additionally, many of the initiating and elongating Pol II complexes have been resolved to excellent resolution by cryogenic electron microscopy and such an approach may lend itself to capturing the XRN2 complex interaction.

Another key area of interest is the mechanisms underpinning phosphatase-mediated elongation slowing. Several phosphatases and kinases have now been implicated in elongation control at different transcript classes. At PAS-dependent transcripts, the main focus should be to identify these PP1 substrates as there are likely more than solely Spt5. One approach may be to carryout phospho-proteomics following inhibition of PP1/PNUTs. Also, it remains unclear how changes to phosphorylation sites provides control over elongation speed and whether this is mediated via conformational changes, associated factor exchanges or both.

References

Adebali, O., Chiou, Y.-Y., Hu, J., Sancar, A., and Selby, C.P. (2017). Genome-wide transcription-coupled repair in Escherichia coli is mediated by the Mfd translocase. Proc Natl Acad Sci USA 114, E2116–E2125.

Agudelo, D., Duringer, A., Bozoyan, L., Huard, C.C., Carter, S., Loehr, J., Synodinou, D., Drouin, M., Salsman, J., Dellaire, G., et al. (2017). Marker-free coselection for CRISPR-driven genome editing in human cells. Nature Methods 14, 615–620.

Ahn, S.H., Kim, M., and Buratowski, S. (2004). Phosphorylation of serine 2 within the RNA polymerase II C-terminal domain couples transcription and 3' end processing. Mol Cell 13, 67–76.

Akhtar, Md.S., Heidemann, M., Tietjen, J.R., Zhang, D.W., Chapman, R.D., Eick, D., and Ansari, A.Z. (2009). TFIIH Kinase Places Bivalent Marks on the Carboxy-Terminal Domain of RNA Polymerase II. Molecular Cell 34, 387–393.

Alekseev, S., Nagy, Z., Sandoz, J., Weiss, A., Egly, J.-M., May, N.L., and Coin, F. (2017). Transcription without XPB Establishes a Unified Helicase-Independent Mechanism of Promoter Opening in Eukaryotic Gene Expression. Molecular Cell 65, 504-514.e4.

Ali, I., Ruiz, D.G., Ni, Z., Johnson, J.R., Zhang, H., Li, P.-C., Khalid, M.M., Conrad, R.J., Guo, X., Min, J., et al. (2019). Crosstalk between RNA Pol II C-Terminal Domain Acetylation and Phosphorylation via RPRD Proteins. Molecular Cell 74, 1164-1174.e4.

Allan, J., Fraser, R.M., Owen-Hughes, T., and Keszenman-Pereyra, D. (2012). Micrococcal Nuclease Does Not Substantially Bias Nucleosome Mapping. Journal of Molecular Biology 417, 152–164.

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

Ashe, M.P., Furger, A., and Proudfoot, N.J. (2000). Stem-loop 1 of the U1 snRNP plays a critical role in the suppression of HIV-1 polyadenylation. RNA 6, 170–177.

Austenaa, L.M.I., Barozzi, I., Simonatto, M., Masella, S., Della Chiara, G., Ghisletti, S., Curina, A., de Wit, E., Bouwman, B.A.M., de Pretis, S., et al. (2015). Transcription of Mammalian cis-Regulatory Elements Is Restrained by Actively Enforced Early Termination. Molecular Cell 60, 460–474.

Baejen, C., Andreani, J., Torkler, P., Battaglia, S., Schwalb, B., Lidschreiber, M., Maier, K.C., Boltendahl, A., Rus, P., Esslinger, S., et al. (2017). Genome-wide Analysis of RNA Polymerase II Termination at Protein-Coding Genes. Molecular Cell 66, 38-49.e6.

Baillat, D., Hakimi, M.-A., Näär, A.M., Shilatifard, A., Cooch, N., and Shiekhattar, R. (2005). Integrator, a Multiprotein Mediator of Small Nuclear RNA Processing, Associates with the C-Terminal Repeat of RNA Polymerase II. Cell 123, 265–276.

Baluapuri, A., Hofstetter, J., Dudvarski Stankovic, N., Endres, T., Bhandare, P., Vos, S.M., Adhikari, B., Schwarz, J.D., Narain, A., Vogt, M., et al. (2019). MYC Recruits SPT5 to RNA Polymerase II to Promote Processive Transcription Elongation. Molecular Cell 74, 674-687.e11.

Baskaran, R., Chiang, G.G., Mysliwiec, T., Kruh, G.D., and Wang, J.Y.J. (1997). Tyrosine Phosphorylation of RNA Polymerase II Carboxyl-terminal Domain by the Abl-related Gene Product. J. Biol. Chem. 272, 18905–18909.

Bauer, D.L.V., Tellier, M., Martínez-Alonso, M., Nojima, T., Proudfoot, N.J., Murphy, S., and Fodor, E. (2018). Influenza Virus Mounts a Two-Pronged Attack on Host RNA Polymerase II Transcription. Cell Reports 23, 2119-2129.e3.

Baurén, G., Belikov, S., and Wieslander, L. (1998). Transcriptional termination in the Balbiani ring 1 gene is closely coupled to 3′-end formation and excision of the 3′-terminal intron. Genes Dev 12, 2759–2769.

Bernecky, C., Plitzko, J.M., and Cramer, P. (2017). Structure of a transcribing RNA polymerase II–DSIF complex reveals a multidentate DNA–RNA clamp. Nature Structural & Molecular Biology 24, 809–815.

References

Bieberstein, N.I., Carrillo Oesterreich, F., Straube, K., and Neugebauer, K.M. (2012). First Exon Length Controls Active Chromatin Signatures and Transcription. Cell Reports 2, 62–68.

Birse, C.E., Minvielle-Sebastia, L., Lee, B.A., Keller, W., and Proudfoot, N.J. (1998). Coupling Termination of Transcription to Messenger RNA Maturation in Yeast. Science 280, 298–301.

Boehm, A.K., Saunders, A., Werner, J., and Lis, J.T. (2003). Transcription Factor and Polymerase Recruitment, Modification, and Movement on dhsp70 In Vivo in the Minutes following Heat Shock. Molecular and Cellular Biology 23, 7628–7637.

Bogden, C.E., Fass, D., Bergman, N., Nichols, M.D., and Berger, J.M. (1999). The Structural Basis for Terminator Recognition by the Rho Transcription Termination Factor. Molecular Cell 3, 487–493.

Booth, G.T., Parua, P.K., Sansó, M., Fisher, R.P., and Lis, J.T. (2018). Cdk9 regulates a promoter-proximal checkpoint to modulate RNA polymerase II elongation rate in fission yeast. Nature Communications 9, 1–10.

Bose, D., Pape, T., Burrows, P.C., Rappas, M., Wigneshweraraj, S.R., Buck, M., and Zhang, X. (2008). Organization of an Activator-Bound RNA Polymerase Holoenzyme. Molecular Cell 32, 337–346.

Bösken, C.A., Farnung, L., Hintermair, C., Merzel Schachter, M., Vogel-Bachmayr, K., Blazek, D., Anand, K., Fisher, R.P., Eick, D., and Geyer, M. (2014). The structure and substrate specificity of human Cdk12/Cyclin K. Nature Communications 5, 3505.

Bowman, E.A., Bowman, C.R., Ahn, J.H., and Kelly, W.G. (2013). Phosphorylation of RNA polymerase II is independent of P-TEFb in the C. elegans germline. Development 140, 3703–3713.

Brannan, K., Kim, H., Erickson, B., Glover-Cutter, K., Kim, S., Fong, N., Kiemele, L., Hansen, K., Davis, R., Lykke-Andersen, J., et al. (2012). mRNA Decapping Factors and the Exonuclease Xrn2 Function in Widespread Premature Termination of RNA Polymerase II Transcription. Molecular Cell 46, 311–324.

Brendel, V., Hamm, G.H., and Trifonov, E.N. (1986). Terminators of Transcription with RNA Polymerase from Escherichia coli: What They Look Like and How to Find Them. Journal of Biomolecular Structure and Dynamics 3, 705–723.

Brennan, C.A., Dombroski, A.J., and Platt, T. (1987). Transcription termination factor rho is an RNA-DNA helicase. Cell 48, 945–952.

Brown, J.A., Valenstein, M.L., Yario, T.A., Tycowski, K.T., and Steitz, J.A. (2012). Formation of triple-helical structures by the 3′-end sequences of MALAT1 and MENβ noncoding RNAs. PNAS 109, 19202–19207.

Browning, D.F., and Busby, S.J.W. (2016). Local and global regulation of transcription initiation in bacteria. Nature Reviews Microbiology 14, 638–650.

Buck, M., Miller, S., Drummond, M., and Dixon, R. (1986). Upstream activator sequences are present in the promoters of nitrogen fixation genes. Nature 320, 374–378.

Buratowski, S. (2003). The CTD code. Nature Structural & Molecular Biology 10, 679–680.

Burrows, P.C., Severinov, K., Ishihama, A., Buck, M., and Wigneshweraraj, S.R. (2003). Mapping σ54-RNA Polymerase Interactions at the –24 Consensus Promoter Element. J. Biol. Chem. 278, 29728–29743.

Bush, M., and Dixon, R. (2012). The Role of Bacterial Enhancer Binding Proteins as Specialized Activators of σ54-Dependent Transcription. Microbiol. Mol. Biol. Rev. 76, 497–529.

Campbell, E.A., Muzzin, O., Chlenov, M., Sun, J.L., Olson, C.A., Weinman, O., Trester-Zedlitz, M.L., and Darst, S.A. (2002). Structure of the bacterial RNA polymerase promoter specificity sigma subunit. Mol Cell 9, 527–539.

Chan, S.L., Huppertz, I., Yao, C., Weng, L., Moresco, J.J., Yates, J.R., Ule, J., Manley, J.L., and Shi, Y. (2014). CPSF30 and Wdr33 directly bind to AAUAAA in mammalian mRNA 3′ processing. Genes Dev. 28, 2370–2380.

Chang, J.H., Jiao, X., Chiba, K., Oh, C., Martin, C.E., Kiledjian, M., and Tong, L. (2012). Dxo1 is a new type of eukaryotic enzyme with both decapping and 5′-3′ exoribonuclease activity. Nature Structural & Molecular Biology 19, 1011–1017.

References

Chapman, E.G., Costantino, D.A., Rabe, J.L., Moon, S.L., Wilusz, J., Nix, J.C., and Kieft, J.S. (2014). The Structural Basis of Pathogenic Subgenomic Flavivirus RNA (sfRNA) Production. Science 344, 307–310.

Chen, B., Sirota, M., Fan-Minogue, H., Hadley, D., and Butte, A.J. (2015a). Relating hepatocellular carcinoma tumor samples and cell lines using gene expression data in translational research. BMC Medical Genomics. 8, S5.

Chen, F., Gao, X., and Shilatifard, A. (2015b). Stably paused genes revealed through inhibition of transcription initiation by the TFIIH inhibitor triptolide. Genes Dev. 29, 39–47.

Chiu, A.C., Suzuki, H.I., Wu, X., Mahat, D.B., Kriz, A.J., and Sharp, P.A. (2018). Transcriptional Pause Sites Delineate Stable Nucleosome-Associated Premature Polyadenylation Suppressed by U1 snRNP. Molecular Cell 69, 648-663.e7.

Chodchoy, N., Pandey, N.B., and Marzluff, W.F. (1991). An intact histone 3'-processing site is required for transcription termination in a mouse histone H2a gene. Molecular and Cellular Biology 11, 497–509.

Chung, H.K., Jacobs, C.L., Huo, Y., Yang, J., Krumm, S.A., Plemper, R.K., Tsien, R.Y., and Lin, M.Z. (2015). Tunable and reversible drug control of protein production via a self-excising degron. Nature Chemical Biology 11, 713–720.

Churchman, L.S., and Weissman, J.S. (2011). Nascent transcript sequencing visualizes transcription at nucleotide resolution. Nature 469, 368–373.

Clerici, M., Faini, M., Aebersold, R., and Jinek, M. (2017). Structural insights into the assembly and polyA signal recognition mechanism of the human CPSF complex. ELife 6, e33111.

Connelly, S., and Manley, J.L. (1988). A functional mRNA polyadenylation signal is required for transcription termination by RNA polymerase II. Genes Dev. 2, 440–452.

Corden, J.L. (2016). Pol II CTD Code Light. Molecular Cell 61, 183–184.

Core, L., and Adelman, K. (2019). Promoter-proximal pausing of RNA polymerase II: a nexus of gene regulation. Genes Dev. 33, 960–982.

Cortazar, M.A., Sheridan, R.M., Erickson, B., Fong, N., Glover-Cutter, K., Brannan, K., and Bentley, D.L. (2019). Control of RNA Pol II Speed by PNUTS-PP1 and Spt5 Dephosphorylation Facilitates Termination by a “Sitting Duck Torpedo” Mechanism. Molecular Cell 76, 896-908.e4.

Cramer, P. (2019). Organization and regulation of gene transcription. Nature 573, 45–54.

Cramer, P., Bushnell, D.A., Fu, J., Gnatt, A.L., Maier-Davis, B., Thompson, N.E., Burgess, R.R., Edwards, A.M., David, P.R., and Kornberg, R.D. (2000). Architecture of RNA polymerase II and implications for the transcription mechanism. Science 288, 640–649.

Croyle, M.L., Woo, A.L., and Lingrel, J.B. (1997). Extensive Random Mutagenesis Analysis of the Na+/K+-ATPase α Subunit Identifies Known and Previously Unidentified Amino Acid Residues that Alter Ouabain Sensitivity Implications for Ouabain Binding. European Journal of Biochemistry 248, 488–495.

Cuatrecasas, P., Fuchs, S., and Anfinsen, C.B. (1967). Catalytic Properties and Specificity of the Extracellular Nuclease of Staphylococcus aureus. J. Biol. Chem. 242, 1541–1547.

Cuello, P., Boyd, D.C., Dye, M.J., Proudfoot, N.J., and Murphy, S. (1999). Transcription of the human U2 snRNA genes continues beyond the 3’ box in vivo. EMBO J 18, 2867–2877.

Custódio, N., Carmo-Fonseca, M., Geraghty, F., Pereira, H.S., Grosveld, F., and Antoniou, M. (1999). Inefficient processing impairs release of RNA from the site of transcription. EMBO J 18, 2855–2866.

Czudnochowski, N., Bösken, C.A., and Geyer, M. (2012). Serine-7 but not serine-5 phosphorylation primes RNA polymerase II CTD for P-TEFb recognition. Nature Communications 3, 842.

Danson, A.E., Jovanovic, M., Buck, M., and Zhang, X. (2019). Mechanisms of σ54-Dependent Transcription Initiation and Regulation. Journal of Molecular Biology 431, 3960–3974.

References

Davidson, L., Kerr, A., and West, S. (2012). Co-transcriptional degradation of aberrant pre-mRNA by Xrn2. The EMBO Journal 31, 2566–2578.

Davidson, L., Francis, L., Cordiner, R.A., Eaton, J.D., Estell, C., Macias, S., Cáceres, J.F., and West, S. (2019). Rapid Depletion of DIS3, EXOSC10, or XRN2 Reveals the Immediate Impact of Exoribonucleolysis on Nuclear RNA Metabolism and Transcriptional Control. Cell Reports 26, 2779-2791.e5.

Davidson, L., Francis, L., Eaton, J.D., and West, S. (2020). Integrator-Dependent and Allosteric/Intrinsic Mechanisms Ensure Efficient Termination of snRNA Transcription. Cell Reports 33, 108319.

Dengl, S., and Cramer, P. (2009). Torpedo Nuclease Rat1 Is Insufficient to Terminate RNA Polymerase II in Vitro. J. Biol. Chem. 284, 21270–21279.

Descostes, N., Heidemann, M., Spinelli, L., Schüller, R., Maqbool, M.A., Fenouil, R., Koch, F., Innocenti, C., Gut, M., Gut, I., et al. (2014). Tyrosine phosphorylation of RNA polymerase II CTD is associated with antisense promoter transcription and active enhancers in mammalian cells. ELife 3, e02105.

Dhir, A., Dhir, S., Proudfoot, N.J., and Jopling, C.L. (2015). Microprocessor mediates transcriptional termination of long noncoding RNA transcripts hosting microRNAs. Nat Struct Mol Biol 22, 319–327.

Dias, J.D., Rito, T., Torlai Triglia, E., Kukalev, A., Ferrai, C., Chotalia, M., Brookes, E., Kimura, H., and Pombo, A. (2015). Methylation of RNA polymerase II non-consensus Lysine residues marks early transcription in mammalian cells. ELife 4, e11215.

Dienemann, C., Schwalb, B., Schilbach, S., and Cramer, P. (2019). Promoter Distortion and Opening in the RNA Polymerase II Cleft. Molecular Cell 73, 97-106.e4.

Dingwall, C., Lomonossoff, G.P., and Laskey, R.A. (1981). High sequence specificity of micrococcal nuclease. Nucleic Acids Res 9, 2659–2674.

Doamekpor, S.K., Grudzien-Nogalska, E., Mlynarska-Cieslak, A., Kowalska, J., Kiledjian, M., and Tong, L. (2020). DXO/Rai1 enzymes remove 5′-end FAD and dephospho-CoA caps on RNAs. Nucleic Acids Res 48, 6136–6148.

Dominski, Z., Yang, X., Purdy, M., Wagner, E.J., and Marzluff, W.F. (2005). A CPSF-73 Homologue Is Required for Cell Cycle Progression but Not Cell Growth and Interacts with a Protein Having Features of CPSF-100. Molecular and Cellular Biology 25, 1489–1500.

Doucleff, M., Pelton, J.G., Lee, P.S., Nixon, B.T., and Wemmer, D.E. (2007). Structural Basis of DNA Recognition by the Alternative Sigma-factor, σ54. Journal of Molecular Biology 369, 1070–1078.

Dubbury, S.J., Boutz, P.L., and Sharp, P.A. (2018). CDK12 regulates DNA repair genes by suppressing intronic polyadenylation. Nature 564, 141–145.

Duchi, D., Bauer, D.L.V., Fernandez, L., Evans, G., Robb, N., Hwang, L.C., Gryte, K., Tomescu, A., Zawadzki, P., Morichaud, Z., et al. (2016). RNA Polymerase Pausing during Initial Transcription. Molecular Cell 63, 939–950.

Dulin, D., Bauer, D.L.V., Malinen, A.M., Bakermans, J.J.W., Kaller, M., Morichaud, Z., Petushkov, I., Depken, M., Brodolin, K., Kulbachinskiy, A., et al. (2018). Pausing controls branching between productive and non-productive pathways during initial transcription in bacteria. Nature Communications 9, 1478.

Dutta, D., Shatalin, K., Epshtein, V., Gottesman, M.E., and Nudler, E. (2011). Linking RNA Polymerase Backtracking to Genome Instability in E. coli. Cell 146, 533–543.

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

Eaton, J.D., Francis, L., Davidson, L., and West, S. (2020). A unified allosteric/torpedo mechanism for transcriptional termination on human protein-coding genes. Genes Dev. 34, 132–145.

Egloff, S., O'Reilly, D., Chapman, R.D., Taylor, A., Tanzhaus, K., Pitts, L., Eick, D., and Murphy, S. (2007). Serine-7 of the RNA Polymerase II CTD Is Specifically Required for snRNA Gene Expression. Science 318, 1777–1779.

References

Egloff, S., Szczepaniak, S.A., Dienstbier, M., Taylor, A., Knight, S., and Murphy, S. (2010). The Integrator Complex Recognizes a New Double Mark on the RNA Polymerase II Carboxyl-terminal Domain. J. Biol. Chem. 285, 20564–20569.

Eick, D., and Geyer, M. (2013). The RNA Polymerase II Carboxy-Terminal Domain (CTD) Code. Chem. Rev. 113, 8456–8490.

El-Brolosy, M.A., Kontarakis, Z., Rossi, A., Kuenne, C., Günther, S., Fukuda, N., Kikhi, K., Boezio, G.L.M., Takacs, C.M., Lai, S.-L., et al. (2019). Genetic compensation triggered by mutant mRNA degradation. Nature 568, 193–197.

Elrod, N.D., Henriques, T., Huang, K.-L., Tatomer, D.C., Wilusz, J.E., Wagner, E.J., and Adelman, K. (2019). The Integrator Complex Attenuates Promoter-Proximal Transcription at Protein-Coding Genes. Molecular Cell 76, 738-752.e7.

Epshtein, V., Cardinale, C.J., Ruckenstein, A.E., Borukhov, S., and Nudler, E. (2007). An Allosteric Path to Transcription Termination. Molecular Cell 28, 991–1001.

Epshtein, V., Dutta, D., Wade, J., and Nudler, E. (2010). An allosteric mechanism of Rho-dependent transcription termination. Nature 463, 245–249.

Erickson, B., Sheridan, R.M., Cortazar, M., and Bentley, D.L. (2018). Dynamic turnover of paused Pol II complexes at human promoters. Genes Dev. 32, 1215–1225.

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

Feaver, W.J., Svejstrup, J.Q., Henry, N.L., and Kornberg, R.D. (1994). Relationship of CDK-activating kinase and RNA polymerase II CTD kinase TFIIH/TFIIK. Cell 79, 1103–1109.

Fong, N., Ohman, M., and Bentley, D.L. (2009). Fast ribozyme cleavage releases transcripts from RNA polymerase II and aborts co-transcriptional pre-mRNA processing. Nat. Struct. Mol. Biol. 16, 916–922.

Fong, N., Brannan, K., Erickson, B., Kim, H., Cortazar, M.A., Sheridan, R.M., Nguyen, T., Karp, S., and Bentley, D.L. (2015). Effects of Transcription

Elongation Rate and Xrn2 Exonuclease Activity on RNA Polymerase II Termination Suggest Widespread Kinetic Competition. Molecular Cell 60, 256–267.

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

Fujinaga, K., Irwin, D., Huang, Y., Taube, R., Kurosu, T., and Peterlin, B.M. (2004). Dynamics of Human Immunodeficiency Virus Transcription: P-TEFb Phosphorylates RD and Dissociates Negative Effectors from the Transactivation Response Element. Molecular and Cellular Biology 24, 787–795.

Gardini, A., Baillat, D., Cesaroni, M., Hu, D., Marinis, J.M., Wagner, E.J., Lazar, M.A., Shilatifard, A., and Shiekhattar, R. (2014). Integrator Regulates Transcriptional Initiation and Pause Release following Activation. Molecular Cell 56, 128–139.

Gillet, J.P., Calcagno, A.M., Varma, S., Marino, M., Green, L.J., Vora, M. I., Patel, C., Orina, J.N., Eliseeva, T.A., Singal, V., et al. (2011). Redefining the relevance of established cancer cell lines to the study of mechanisms of clinical anti-cancer drug resistance. Proceedings of the National Academy of Sciences 108, 18708–18713.

Glover-Cutter, K., Larochelle, S., Erickson, B., Zhang, C., Shokat, K., Fisher, R.P., and Bentley, D.L. (2009). TFIIH-Associated Cdk7 Kinase Functions in Phosphorylation of C-Terminal Domain Ser7 Residues, Promoter-Proximal Pausing, and Termination by RNA Polymerase II. Molecular and Cellular Biology 29, 5455–5464.

Glyde, R., Ye, F., Darbari, V.C., Zhang, N., Buck, M., and Zhang, X. (2017). Structures of RNA Polymerase Closed and Intermediate Complexes Reveal Mechanisms of DNA Opening and Transcription Initiation. Molecular Cell 67, 106-116.e4.

Graur, D., Zheng, Y., Price, N., Azevedo, R.B.R., Zufall, R.A., and Elhaik, E. (2013). On the Immortality of Television Sets: “Function” in the Human Genome According to the Evolution-Free Gospel of ENCODE. Genome Biol Evol 5, 578–590.

References

Greger, I.H., and Proudfoot, N.J. (1998). Poly(A) signals control both transcriptional termination and initiation between the tandem GAL10 and GAL7 genes of Saccharomyces cerevisiae. The EMBO Journal 17, 4771–4779.

Greger, I.H., Aranda, A., and Proudfoot, N. (2000). Balancing transcriptional interference and initiation on the GAL7 promoter of Saccharomyces cerevisiae. PNAS 97, 8415–8420.

Gregersen, L.H., Mitter, R., Ugalde, A.P., Nojima, T., Proudfoot, N.J., Agami, R., Stewart, A., and Svejstrup, J.Q. (2019). SCAF4 and SCAF8, mRNA Anti-Terminator Proteins. Cell 177, 1797-1813.e18.

Grünberg, S., Warfield, L., and Hahn, S. (2012). Architecture of the RNA polymerase II preinitiation complex and mechanism of ATP-dependent promoter opening. Nat Struct Mol Biol 19, 788–796.

Gui, L., Sunnarborg, A., Pan, B., and LaPorte, D.C. (1996). Autoregulation of iclR, the gene encoding the repressor of the glyoxylate bypass operon. J Bacteriol 178, 321–324.

Gusarov, I., and Nudler, E. (1999). The Mechanism of Intrinsic Transcription Termination. Molecular Cell 3, 495–504.

Hallais, M., Pontvianne, F., Andersen, P.R., Clerici, M., Lener, D., Benbahouche, N.E.H., Gostan, T., Vandermoere, F., Robert, M.-C., Cusack, S., et al. (2013). CBC–ARS2 stimulates 3′-end maturation of multiple RNA families and favors cap-proximal processing. Nat Struct Mol Biol 20, 1358–1366.

Heinz, S., Texari, L., Hayes, M.G.B., Urbanowski, M., Chang, M.W., Givarkes, N., Rialdi, A., White, K.M., Albrecht, R.A., Pache, L., et al. (2018). Transcription Elongation Can Affect Genome 3D Structure. Cell 174, 1522-1536.e22.

Hernandez, N., and Weiner, A.M. (1986). Formation of the 3′ end of U1 snRNA requires compatible snRNA promoter elements. Cell 47, 249–258.

Hintermair, C., Heidemann, M., Koch, F., Descostes, N., Gut, M., Gut, I., Fenouil, R., Ferrier, P., Flatley, A., Kremmer, E., et al. (2012). Threonine-4 of mammalian RNA polymerase II CTD is targeted by Polo-like kinase 3 and required for transcriptional elongation. The EMBO Journal 31, 2784–2797.

Hoek, T.A., Khuperkar, D., Lindeboom, R.G.H., Sonneveld, S., Verghagen, B.M.P., Boersma, S., Vermeulen. M., and Tanenbaum, M.E. (2019). Single-Molecule Imaging Uncovers Rules Governing Nonsense-Mediated mRNA Decay. Molecular Cell 75, 324-339.

Horvathova, I., Voigt, F., Kotrys, A.V., Zhan, Y., Artus-Revel, C.G., Eglinger, J., Stadler, M.B., Giorgetti, L., and Chao, J.A. (2017). The Dynamics of mRNA Turnover Revealed by Single-Molecule Imaging in Single Cells. Molecular Cell 68, 615-625.e9.

Hsin, J.-P., Sheth, A., and Manley, J.L. (2011). RNAP II CTD Phosphorylated on Threonine-4 Is Required for Histone mRNA 3′ End Processing. Science 334, 683–686.

Hsin, J.-P., Li, W., Hoque, M., Tian, B., and Manley, J.L. (2014). RNAP II CTD tyrosine 1 performs diverse functions in vertebrate cells. ELife 3, e02112.

Huang, K.-L., Jee, D., Stein, C.B., Elrod, N.D., Henriques, T., Mascibroda, L.G., Baillat, D., Russell, W.K., Adelman, K., and Wagner, E.J. (2020). Integrator Recruits Protein Phosphatase 2A to Prevent Pause Release and Facilitate Transcription Termination. Molecular Cell S1097276520305797.

Jenal, M., Elkon, R., Loayza-Puch, F., van Haaften, G., Kühn, U., Menzies, F.M., Vrielink, J.A.F.O., Bos, A.J., Drost, J., Rooijers, K., et al. (2012). The Poly(A)-Binding Protein Nuclear 1 Suppresses Alternative Cleavage and Polyadenylation Sites. Cell 149, 538–553.

Jeronimo, C., and Robert, F. (2014). Kin28 regulates the transient association of Mediator with core promoters. Nature Structural & Molecular Biology 21, 449–455.

Jiao, X., Xiang, S., Oh, C., Martin, C.E., Tong, L., and Kiledjian, M. (2010). Identification of a quality-control mechanism for mRNA 5′-end capping. Nature 467, 608–611.

Jiao, X., Chang, J.H., Kilic, T., Tong, L., and Kiledjian, M. (2013). A Mammalian Pre-mRNA 5′ End Capping Quality Control Mechanism and an Unexpected Link of Capping to Pre-mRNA Processing. Molecular Cell 50, 104–115.

References

Jiao, X., Doamekpor, S.K., Bird, J.G., Nickels, B.E., Tong, L., Hart, R.P., and Kiledjian, M. (2017). 5' End Nicotinamide Adenine Dinucleotide Cap in Human Cells Promotes RNA Decay through DXO-Mediated deNADding. Cell 168, 1015-1027.e10.

Jimeno-González, S., Ceballos-Chávez, M., and Reyes, J.C. (2015). A positioned +1 nucleosome enhances promoter-proximal pausing. Nucleic Acids Res 43, 3068–3078.

Jinek, M., Coyle, S.M., and Doudna, J.A. (2011). Coupled 5′ Nucleotide Recognition and Processivity in Xrn1-Mediated mRNA Decay. Molecular Cell 41, 600–608.

Joo, Y.J., Ficarro, S.B., Chun, Y., Marto, J.A., and Buratowski, S. (2019). In vitro analysis of RNA polymerase II elongation complex dynamics. Genes Dev. 33, 578–589.

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

Kamieniarz-Gdula, K., and Proudfoot, N.J. (2019). Transcriptional Control by Premature Termination: A Forgotten Mechanism. Trends in Genetics 35, 553–564.

Kamieniarz-Gdula, K., Gdula, M.R., Panser, K., Nojima, T., Monks, J., Wiśniewski, J.R., Riepsaame, J., Brockdorff, N., Pauli, A., and Proudfoot, N.J. (2019). Selective Roles of Vertebrate PCF11 in Premature and Full-Length Transcript Termination. Molecular Cell 74, 158-172.e9.

Kecman, T., Kuś, K., Heo, D.-H., Duckett, K., Birot, A., Liberatori, S., Mohammed, S., Geis-Asteggiante, L., Robinson, C.V., and Vasiljeva, L. (2018). Elongation/Termination Factor Exchange Mediated by PP1 Phosphatase Orchestrates Transcription Termination. Cell Reports 25, 259-269.e5.

Kelly, W.G., Dahmus, M.E., and Hart, G.W. (1993). RNA polymerase II is a glycoprotein. Modification of the COOH-terminal domain by O-GlcNAc. J Biol Chem 268, 10416–10424.

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

Kim, M., Suh, H., Cho, E.-J., and Buratowski, S. (2009). Phosphorylation of the Yeast Rpb1 C-terminal Domain at Serines 2, 5, and 7. J. Biol. Chem. 284, 26421–26426.

Kim, T.K., Ebright, R.H., and Reinberg, D. (2000). Mechanism of ATP-dependent promoter melting by transcription factor IIH. Science 288, 1418–1422.

Kim, Y.-J., Björklund, S., Li, Y., Sayre, M.H., and Kornberg, R.D. (1994). A multiprotein mediator of transcriptional activation and its interaction with the C-terminal repeat domain of RNA polymerase II. Cell 77, 599–608.

Knutsen, T., Padilla-Nash, H.M., Wangsa, D., Barenboim-Stapleton, L., Camps, J., McNeil, N., Difilippantonio, M.J., and Ried, T. (2010). Definitive molecular cytogenetic characterization of 15 colorectal cancer cell lines. Genes, Chromosomes and Cancer 49, 204–223.

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

Kolev, N.G., Yario, T.A., Benson, E., and Steitz, J.A. (2008). Conserved motifs in both CPSF73 and CPSF100 are required to assemble the active endonuclease for histone mRNA 3′-end maturation. EMBO Reports 9, 1013–1018.

Komarnitsky, P., Cho, E.-J., and Buratowski, S. (2000). Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. Genes Dev. 14, 2452–2460.

Kowarz, E., Löscher, D., and Marschalek, R. (2015). Optimized Sleeping Beauty transposons rapidly generate stable transgenic cell lines. Biotechnology Journal 10, 647–653.

References

Krebs, A.R., Imanci, D., Hoerner, L., Gaidatzis, D., Burger, L., and Schübeler, D. (2017). Genome-wide Single-Molecule Footprinting Reveals High RNA Polymerase II Turnover at Paused Promoters. Molecular Cell 67, 411-422.e4.

Kubota, T., Nishimura, K., Kanemaki, M.T., and Donaldson, A.D. (2013). The Elg1 Replication Factor C-like Complex Functions in PCNA Unloading during DNA Replication. Molecular Cell 50, 273–280.

Kumar, A., Clerici, M., Muckenfuss, L.M., Passmore, L.A., and Jinek, M. (2019). Mechanistic insights into mRNA 3′-end processing. Current Opinion in Structural Biology 59, 143–150.

Lai, F., Damle, S.S., Ling, K.K., and Rigo, F. (2020). Directed RNase H Cleavage of Nascent Transcripts Causes Transcription Termination. Molecular Cell.

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

Lasater, L.S., and Eichler, D.C. (1984). Isolation and properties of a single-strand 5'----3' exoribonuclease from Ehrlich ascites tumor cell nucleoli. Biochemistry 23, 4367–4373.

Laursen, M., Gregersen, J.L., Yatime, L., Nissen, P., and Fedosova, N.U. (2015). Structures and characterization of digoxin- and bufalin-bound Na+,K+-ATPase compared with the ouabain-bound complex. PNAS 112, 1755–1760.

Lawson, M.R., Ma, W., Bellecourt, M.J., Artsimovitch, I., Martin, A., Landick, R., Schulten, K., and Berger, J.M. (2018). Mechanism for the Regulated Control of Bacterial Transcription Termination by a Universal Adaptor Protein. Molecular Cell 71, 911-922.e4.

Le, T.T., Yang, Y., Tan, C., Suhanovsky, M.M., Fulbright, R.M., Inman, J.T., Li, M., Lee, J., Perelman, S., Roberts, J.W., et al. (2018). Mfd Dynamically Regulates Transcription via a Release and Catch-Up Mechanism. Cell 172, 344-357.e15.

Lee, J.-S., and Mendell, J.T. (2020). Antisense-Mediated Transcript Knockdown Triggers Premature Transcription Termination. Molecular Cell.s

Lee, C., Li, X., Hechmer, A., Eisen, M., Biggin, M.D., Venters, B.J., Jiang, C., Li, J., Pugh, B.F., and Gilmour, D.S. (2008). NELF and GAGA Factor Are Linked to Promoter-Proximal Pausing at Many Genes in Drosophila. MCB 28, 3290–3300.

Lee, J.-H., You, J., Dobrota, E., and Skalnik, D.G. (2010). Identification and Characterization of a Novel Human PP1 Phosphatase Complex. J. Biol. Chem. 285, 24466–24476.

Lenoir, W.F., Lim, T.L., and Hart, T. (2018). PICKLES: the database of pooled in-vitro CRISPR knockout library essentiality screens. Nucleic Acids Res 46, D776–D780.

Li, X.-Q., and Du, D. (2013). RNA Polyadenylation Sites on the Genomes of Microorganisms, Animals, and Plants. PLOS ONE 8, e79511.

Li, S., Prasanna, X., Salo, V.T., Vattulainen, I., and Ikonen, E. (2019). An efficient auxin-inducible degron system with low basal degradation in human cells. Nature Methods 16, 866–869.

Libri, D. (2015). Endless Quarrels at the End of Genes. Molecular Cell 60, 192–194.

Liu, K., Newbury, P.A., Glicksberg, B.S., Zeng, W.Z.D., Paithankar, S., Andrechek, E.R., and Chen, B. (2019). Evaluating cell lines as models for metastatic breast cancer through integrative analysis of genomic data. Nature Communications 10, 2138.

Logan, J., Falck-Pedersen, E., Darnell, J.E., and Shenk, T. (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. PNAS 84, 8306–8310.

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

Ma, Z., Zhu, P., Shi, H., Guo, L., Zhang, Q., Chen, Y., Chen, S., Zhang, Z., Peng, J., and Chen, J. (2019). PTC-bearing mRNA elicits a genetic compensation response via Upf3a and COMPASS components. Nature 568, 259–263.

References

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

Mapendano, C.K., Lykke-Andersen, S., Kjems, J., Bertrand, E., and Jensen, T.H. (2010). Crosstalk between mRNA 3′ End Processing and Transcription Initiation. Molecular Cell 40, 410–422.

Marshall, N.F., Peng, J., Xie, Z., and Price, D.H. (1996). Control of RNA Polymerase II Elongation Potential by a Novel Carboxyl-terminal Domain Kinase. J. Biol. Chem. 271, 27176–27183.

Martin, F.H., and Tinoco, I. (1980). DNA-RNA hybrid duplexes containing oligo(dA:rU) sequences are exceptionally unstable and may facilitate termination of transcription. Nucleic Acids Res 8, 2295–2300.

Martin, G., Gruber, A.R., Keller, W., and Zavolan, M. (2012). Genome-wide Analysis of Pre-mRNA 3′ End Processing Reveals a Decisive Role of Human Cleavage Factor I in the Regulation of 3′ UTR Length. Cell Reports 1, 753–763.

Marzluff, W.F., and Koreski, K.P. (2017). Birth and Death of Histone mRNAs. Trends in Genetics 33, 745–759.

Mayer, A., Heidemann, M., Lidschreiber, M., Schreieck, A., Sun, M., Hintermair, C., Kremmer, E., Eick, D., and Cramer, P. (2012). CTD Tyrosine Phosphorylation Impairs Termination Factor Recruitment to RNA Polymerase II. Science 336, 1723–1725.

Mayer, A., di Iulio, J., Maleri, S., Eser, U., Vierstra, J., Reynolds, A., Sandstrom, R., Stamatoyannopoulos, J.A., and Churchman, L.S. (2015). Native Elongating Transcript Sequencing Reveals Human Transcriptional Activity at Nucleotide Resolution. Cell 161, 541–554.

Mayfield, J.E., Irani, S., Escobar, E.E., Zhang, Z., Burkholder, N.T., Robinson, M.R., Mehaffey, M.R., Sipe, S.N., Yang, W., Prescott, N.A., et al. (2019). Tyr1 phosphorylation promotes phosphorylation of Ser2 on the C-terminal domain of eukaryotic RNA polymerase II by P-TEFb. ELife 8, e48725.

Medlin, J.E., Uguen, P., Taylor, A., Bentley, D.L., and Murphy, S. (2003). The C-terminal domain of pol II and a DRB-sensitive kinase are required for 3′ processing of U2 snRNA. The EMBO Journal 22, 925–934.

Mendoza-Figueroa, M.S., Tatomer, D.C., and Wilusz, J.E. (2020). The Integrator Complex in Transcription and Development. Trends in Biochemical Sciences S0968000420301778.

Miki, T.S., Carl, S.H., and Großhans, H. (2017). Two distinct transcription termination modes dictated by promoters. Genes Dev. 31, 1870–1879.

Miyaoka, Y., Berman, J.R., Cooper, S.B., Mayerl, S.J., Chan, A.H., Zhang, B., Karlin-Neumann, G.A., and Conklin, B.R. (2016). Systematic quantification of HDR and NHEJ reveals effects of locus, nuclease, and cell type on genome-editing. Scientific Reports 6, 23549.

Muniz, L., Davidson, L., and West, S. (2015). Poly(A) Polymerase and the Nuclear Poly(A) Binding Protein, PABPN1, Coordinate the Splicing and Degradation of a Subset of Human Pre-mRNAs. Molecular and Cellular Biology 35, 2218–2230.

Murakami, K.S., and Darst, S.A. (2003). Bacterial RNA polymerases: the wholo story. Current Opinion in Structural Biology 13, 31–39.

Murakami, K.S., Masuda, S., Campbell, E.A., Muzzin, O., and Darst, S.A. (2002). Structural Basis of Transcription Initiation: An RNA Polymerase Holoenzyme-DNA Complex. Science 296, 1285–1290.

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

Nechaev, S., Fargo, D.C., Santos, G. dos, Liu, L., Gao, Y., and Adelman, K. (2010). Global Analysis of Short RNAs Reveals Widespread Promoter-Proximal Stalling and Arrest of Pol II in Drosophila. Science 327, 335–338.

Nedea, E., Nalbant, D., Xia, D., Theoharis, N.T., Suter, B., Richardson, C.J., Tatchell, K., Kislinger, T., Greenblatt, J.F., and Nagy, P.L. (2008). The Glc7 Phosphatase Subunit of the Cleavage and Polyadenylation Factor Is Essential for Transcription Termination on snoRNA Genes. Molecular Cell 29, 577–587.

References

Ni, Z., Schwartz, B.E., Werner, J., Suarez, J.-R., and Lis, J.T. (2004). Coordination of Transcription, RNA Processing, and Surveillance by P-TEFb Kinase on Heat Shock Genes. Molecular Cell 13, 55–65.

Nielsen, S., Yuzenkova, Y., and Zenkin, N. (2013). Mechanism of Eukaryotic RNA Polymerase III Transcription Termination. Science 340, 1577–1580.

Nikolov, D.B., and Burley, S.K. (1997). RNA polymerase II transcription initiation: A structural view. PNAS 94, 15–22.

Nilson, K.A., Lawson, C.K., Mullen, N.J., Ball, C.B., Spector, B.M., Meier, J.L., and Price, D.H. (2017). Oxidative stress rapidly stabilizes promoter-proximal paused Pol II across the human genome. Nucleic Acids Res 45, 11088–11105.

Nishimura, K., and Fukagawa, T. (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, 253–260.

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

Nojima, T., Gomes, T., Grosso, A.R.F., Kimura, H., Dye, M.J., Dhir, S., Carmo-Fonseca, M., and Proudfoot, N.J. (2015). Mammalian NET-Seq Reveals Genome-wide Nascent Transcription Coupled to RNA Processing. Cell 161, 526–540.

Nojima, T., Gomes, T., Carmo-Fonseca, M., and Proudfoot, N.J. (2016). Mammalian NET-seq analysis defines nascent RNA profiles and associated RNA processing genome-wide. Nature Protocols 11, 413–428.

Nojima, T., Tellier, M., Foxwell, J., Ribeiro de Almeida, C., Tan-Wong, S.M., Dhir, S., Dujardin, G., Dhir, A., Murphy, S., and Proudfoot, N.J. (2018a). Deregulated Expression of Mammalian lncRNA through Loss of SPT6 Induces R-Loop Formation, Replication Stress, and Cellular Senescence. Molecular Cell 72, 970-984.e7.

Nojima, T., Rebelo, K., Gomes, T., Grosso, A.R., Proudfoot, N.J., and Carmo-Fonseca, M. (2018b). RNA Polymerase II Phosphorylated on CTD Serine 5

Interacts with the Spliceosome during Co-transcriptional Splicing. Molecular Cell 72, 369-379.e4.

Nozawa, K., Schneider, T.R., and Cramer, P. (2017a). Core Mediator structure at 3.4 Å extends model of transcription initiation complex. Nature 545, 248–251.

Nozawa, R.-S., Boteva, L., Soares, D.C., Naughton, C., Dun, A.R., Buckle, A., Ramsahoye, B., Bruton, P.C., Saleeb, R.S., Arnedo, M., et al. (2017b). SAF-A Regulates Interphase Chromosome Structure through Oligomerization with Chromatin-Associated RNAs. Cell 169, 1214-1227.e18.

Ntini, E., Järvelin, A.I., Bornholdt, J., Chen, Y., Boyd, M., Jørgensen, M., Andersson, R., Hoof, I., Schein, A., Andersen, P.R., et al. (2013). Polyadenylation site–induced decay of upstream transcripts enforces promoter directionality. Nat Struct Mol Biol 20, 923–928.

Nudler, E., and Gottesman, M.E. (2002). Transcription termination and anti-termination in E. coli. Genes to Cells 7, 755–768.

Nuland, R. van, Smits, A.H., Pallaki, P., Jansen, P.W.T.C., Vermeulen, M., and Timmers, H.T.M. (2013). Quantitative Dissection and Stoichiometry Determination of the Human SET1/MLL Histone Methyltransferase Complexes. Molecular and Cellular Biology 33, 2067–2077.

Ogawa, H., Shinoda, T., Cornelius, F., and Toyoshima, C. (2009). Crystal structure of the sodium-potassium pump (Na+,K+-ATPase) with bound potassium and ouabain. PNAS 106, 13742–13747.

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

Osheim, Y.N., Proudfoot, N.J., and Beyer, A.L. (1999). EM Visualization of Transcription by RNA Polymerase II: Downstream Termination Requires a Poly(A) Signal but Not Transcript Cleavage. Molecular Cell 3, 379–387.

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

Palazzo, A.F., and Koonin, E.V. (2020). Functional Long Non-coding RNAs Evolve from Junk Transcripts. Cell.

Park, J., Kang, M., and Kim, M. (2015). Unraveling the mechanistic features of RNA polymerase II termination by the 5′-3′ exoribonuclease Rat1. Nucleic Acids Res 43, 2625–2637.

Parua, P.K., Booth, G.T., Sansó, M., Benjamin, B., Tanny, J.C., Lis, J.T., and Fisher, R.P. (2018). A Cdk9–PP1 switch regulates the elongation–termination transition of RNA polymerase II. Nature 558, 460–464.

Parua, P.K., Kalan, S., Benjamin, B., Sansó, M., and Fisher, R.P. (2020). Distinct Cdk9-phosphatase switches act at the beginning and end of elongation by RNA polymerase II. Nature Communications 11, 4338.

Pearson, E.L., and Moore, C.L. (2013). Dismantling Promoter-driven RNA Polymerase II Transcription Complexes in Vitro by the Termination Factor Rat1. J. Biol. Chem. 288, 19750–19759.

Pei, Y., and Shuman, S. (2002). Interactions between Fission Yeast mRNA Capping Enzymes and Elongation Factor Spt5. J. Biol. Chem. 277, 19639–19648.

Peters, J.M., Mooney, R.A., Kuan, P.F., Rowland, J.L., Keleş, S., and Landick, R. (2009). Rho directs widespread termination of intragenic and stable RNA transcription. PNAS 106, 15406–15411.

Pettinati, I., Grzechnik, P., Ribeiro de Almeida, C., Brem, J., McDonough, M.A., Dhir, S., Proudfoot, N.J., and Schofield, C.J. (2018). Biosynthesis of histone messenger RNA employs a specific 3’ end endonuclease. ELife 7, e39865.

Plaschka, C., Larivière, L., Wenzeck, L., Seizl, M., Hemann, M., Tegunov, D., Petrotchenko, E.V., Borchers, C.H., Baumeister, W., Herzog, F., et al. (2015). Architecture of the RNA polymerase II–Mediator core initiation complex. Nature 518, 376–380.

Pomerantz, R.T., and O’Donnell, M. (2010). Direct Restart of a Replication Fork Stalled by a Head-On RNA Polymerase. Science 327, 590–592.

Popham, D.L., Szeto, D., Keener, J., and Kustu, S. (1989). Function of a bacterial activator protein that binds to transcriptional enhancers. Science 243, 629–635.

Porrua, O., and Libri, D. (2015). Transcription termination and the control of the transcriptome: why, where and how to stop. Nat Rev Mol Cell Biol 16, 190–202.

Porrua, O., Boudvillain, M., and Libri, D. (2016). Transcription Termination: Variations on Common Themes. Trends in Genetics 32, 508–522.

Preker, P., Nielsen, J., Kammler, S., Lykke-Andersen, S., Christensen, M.S., Mapendano, C.K., Schierup, M.H., and Jensen, T.H. (2008). RNA Exosome Depletion Reveals Transcription Upstream of Active Human Promoters. Science 322, 1851–1854.

Proudfoot, N.J. (1989). How RNA polymerase II terminates transcription in higher eukaryotes. Trends in Biochemical Sciences 14, 105–110.

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

Proudfoot, N.J. (2016). Transcriptional termination in mammals: Stopping the RNA polymerase II juggernaut. Science 352.

Proudfoot, N.J., and Brownlee, G.G. (1976). 3′ Non-coding region sequences in eukaryotic messenger RNA. Nature 263, 211–214.

Ramamurthy, L., Ingledue, T.C., Pilch, D.R., Kay, B.K., and Marzluff, W.F. (1996). Increasing the Distance Between the snRNA Promoter and the 3′ Box Decreases the Efficiency of snRNA 3-End Formation. Nucleic Acids Res 24, 4525–4534.

Resjö, S., Oknianska, A., Zolnierowicz, S., Manganiello, V., and Degerman, E. (1999). Phosphorylation and activation of phosphodiesterase type 3B (PDE3B) in adipocytes in response to serine/threonine phosphatase inhibitors: deactivation of PDE3B in vitro by protein phosphatase type 2A. Biochem. J. 341 (Pt 3), 839–845.

Richards, J., and Belasco, J.G. (2011). Ribonuclease J: How to Lead a Double Life. Structure 19, 1201–1203.

Rondón, A.G., Mischo, H.E., Kawauchi, J., and Proudfoot, N.J. (2009). Fail-Safe Transcriptional Termination for Protein-Coding Genes in S. cerevisiae. Molecular Cell 36, 88–98.

Rutkowski, A.J., Erhard, F., L'Hernault, A., Bonfert, T., Schilhabel, M., Crump, C., Rosenstiel, P., Efstathiou, S., Zimmer, R., Friedel, C.C., et al. (2015). Widespread disruption of host transcription termination in HSV-1 infection. Nature Communications 6, 1–15.

Sainsbury, S., Niesser, J., and Cramer, P. (2013). Structure and function of the initially transcribing RNA polymerase II-TFIIB complex. Nature 493, 437–440.

Sanders, T.J., Wenck, B.R., Selan, J.N., Barker, M.P., Trimmer, S.A., Walker, J.E., and Santangelo, T.J. (2020). FttA is a CPSF73 homologue that terminates transcription in Archaea. Nat Microbiol.

Sanderson, A., Mitchell, J.E., Minchin, S.D., and Busby, S.J.W. (2003). Substitutions in the Escherichia coli RNA polymerase σ70 factor that affect recognition of extended −10 elements at promoters. FEBS Letters 544, 199–205.

Sansó, M., Levin, R.S., Lipp, J.J., Wang, V.Y.-F., Greifenberg, A.K., Quezada, E.M., Ali, A., Ghosh, A., Larochelle, S., Rana, T.M., et al. (2016). P-TEFb regulation of transcription termination factor Xrn2 revealed by a chemical genetic screen for Cdk9 substrates. Genes Dev 30, 117–131.

Sasse-Dwight, S., and Gralla, J.D. (1988). Probing the Escherichia coli glnALG upstream activation mechanism in vivo. Proceedings of the National Academy of Sciences 85, 8934–8938.

Sathyan, K.M., McKenna, B.D., Anderson, W.D., Duarte, F.M., Core, L., and Guertin, M.J. (2019). An improved auxin-inducible degron system preserves native protein levels and enables rapid and specific protein depletion. Genes Dev. 33, 1441–1455.

Schlackow, M., Nojima, T., Gomes, T., Dhir, A., Carmo-Fonseca, M., and Proudfoot, N.J. (2017). Distinctive Patterns of Transcription and RNA Processing for Human lincRNAs. Molecular Cell 65, 25–38.

Schönemann, L., Kühn, U., Martin, G., Schäfer, P., Gruber, A.R., Keller, W., Zavolan, M., and Wahle, E. (2014). Reconstitution of CPSF active in polyadenylation: recognition of the polyadenylation signal by WDR33. Genes Dev. 28, 2381–2393.

Schreieck, A., Easter, A.D., Etzold, S., Wiederhold, K., Lidschreiber, M., Cramer, P., and Passmore, L.A. (2014). RNA polymerase II termination involves C-terminal-domain tyrosine dephosphorylation by CPF subunit Glc7. Nature Structural & Molecular Biology 21, 175–179.

Schüller, R., Forné, I., Straub, T., Schreieck, A., Texier, Y., Shah, N., Decker, T.-M., Cramer, P., Imhof, A., and Eick, D. (2016). Heptad-Specific Phosphorylation of RNA Polymerase II CTD. Molecular Cell 61, 305–314.

Schwalb, B., Michel, M., Zacher, B., Frühauf, K., Demel, C., Tresch, A., Gagneur, J., and Cramer, P. (2016). TT-seq maps the human transient transcriptome. Science 352, 1225–1228.

Schwartz, A., Margeat, E., Rahmouni, A.R., and Boudvillain, M. (2007). Transcription Termination Factor Rho Can Displace Streptavidin from Biotinylated RNA. J. Biol. Chem. 282, 31469–31476.

Seila, A.C., Calabrese, J.M., Levine, S.S., Yeo, G.W., Rahl, P.B., Flynn, R.A., Young, R.A., and Sharp, P.A. (2008). Divergent Transcription from Active Promoters. Science 322, 1849–1851.

Selby, C.P., and Sancar, A. (1993). Molecular mechanism of transcription-repair coupling. Science 260, 53–58.

Semsey, S., Tolstorukov, M.Y., Virnik, K., Zhurkin, V.B., and Adhya, S. (2004). DNA trajectory in the Gal repressosome. Genes Dev. 18, 1898–1907.

Sevostyanova, A., Belogurov, G.A., Mooney, R.A., Landick, R., and Artsimovitch, I. (2011). The β Subunit Gate Loop Is Required for RNA Polymerase Modification by RfaH and NusG. Molecular Cell 43, 253–262.

Shandilya, J., and Roberts, S.G.E. (2012). The transcription cycle in eukaryotes: From productive initiation to RNA polymerase II recycling. Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms 1819, 391–400.

Sharmeen, L., Kuo, M.Y., Dinter-Gottlieb, G., and Taylor, J. (1988). Antigenomic RNA of human hepatitis delta virus can undergo self-cleavage. Journal of Virology 62, 2674–2679.

Shearwin, K.E., Callen, B.P., and Egan, J.B. (2005). Transcriptional interference – a crash course. Trends in Genetics 21, 339–345.

Sheridan, R.M., and Bentley, D.L. (2016). Selectable one-step PCR-mediated integration of a degron for rapid depletion of endogenous human proteins. BioTechniques 60, 69–74.

Shetty, A., Kallgren, S.P., Demel, C., Maier, K.C., Spatt, D., Alver, B.H., Cramer, P., Park, P.J., and Winston, F. (2017). Spt5 Plays Vital Roles in the Control of Sense and Antisense Transcription Elongation. Molecular Cell 66, 77-88.e5.

Shi, Y., Di Giammartino, D.C., Taylor, D., Sarkeshik, A., Rice, W.J., Yates, J.R., Frank, J., and Manley, J.L. (2009). Molecular Architecture of the Human Pre-mRNA 3′ Processing Complex. Molecular Cell 33, 365–376.

Shim, E.Y., Walker, A.K., Shi, Y., and Blackwell, T.K. (2002). CDK-9/cyclin T (P-TEFb) is required in two postinitiation pathways for transcription in the C. elegans embryo. Genes Dev. 16, 2135–2146.

Šiková, M., Wiedermannová, J., Převorovský, M., Barvík, I., Sudzinová, P., Kofroňová, O., Benada, O., Šanderová, H., Condon, C., and Krásný, L. (2020). The torpedo effect in Bacillus subtilis: RNase J1 resolves stalled transcription complexes. The EMBO Journal 39, e102500.

Singh, J., and Padgett, R.A. (2009). Rates of in situ transcription and splicing in large human genes. Nature Structural & Molecular Biology 16, 1128–1133.

Sousa-Luis, R., Dujardin, G., Zukher, I., Kimura, H., Carmo-Fonseca, M., Proudfoot, N.J., and Nojima, T. (2020). Point Technology Illuminates the Processing of Polymerase-Associated Intact Nascent Transcripts. BioRxiv 2020.11.09.374108.

Stadelmayer, B., Micas, G., Gamot, A., Martin, P., Malirat, N., Koval, S., Raffel, R., Sobhian, B., Severac, D., Rialle, S., et al. (2014). Integrator complex

regulates NELF-mediated RNA polymerase II pause/release and processivity at coding genes. Nature Communications 5, 5531.

Steurer, B., Janssens, R.C., Geverts, B., Geijer, M.E., Wienholz, F., Theil, A.F., Chang, J., Dealy, S., Pothof, J., Cappellen, W.A. van, et al. (2018). Live-cell analysis of endogenous GFP-RPB1 uncovers rapid turnover of initiating and promoter-paused RNA Polymerase II. PNAS 115, E4368–E4376.

Stevens, A., and Maupin, M.K. (1987). A 5′–3′ exoribonuclease of human placental nuclei: purification and substrate specificity. Nucleic Acids Res 15, 695–708.

Sullivan, K.D., Steiniger, M., and Marzluff, W.F. (2009). A Core Complex of CPSF73, CPSF100, and Symplekin May Form Two Different Cleavage Factors for Processing of Poly(A) and Histone mRNAs. Molecular Cell 34, 322–332.

Sun, Y., Zhang, Y., Hamilton, K., Manley, J.L., Shi, Y., Walz, T., and Tong, L. (2018). Molecular basis for the recognition of the human AAUAAA polyadenylation signal. Proc. Natl. Acad. Sci. U.S.A. 115, E1419–E1428.

Sun, Y., Zhang, Y., Aik, W.S., Yang, X.-C., Marzluff, W.F., Walz, T., Dominski, Z., and Tong, L. (2020). Structure of an active human histone pre-mRNA 3′-end processing machinery. Science 367, 700–703.

Swint-Kruse, L., and Matthews, K.S. (2009). Allostery in the LacI/GalR family: variations on a theme. Current Opinion in Microbiology 12, 129–137.

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

Tatomer, D.C., Elrod, N.D., Liang, D., Xiao, M.-S., Jiang, J.Z., Jonathan, M., Huang, K.-L., Wagner, E.J., Cherry, S., and Wilusz, J.E. (2019). The Integrator complex cleaves nascent mRNAs to attenuate transcription. Genes Dev. 33, 1525–1538.

Tellier, M., and Murphy, S. (2020). Incomplete removal of ribosomal RNA can affect chromatin RNA-seq data analysis. Transcription 0, 1–6.

Tellier, M., Zaborowska, J., Caizzi, L., Mohammad, E., Velychko, T., Schwalb, B., Ferrer-Vicens, I., Blears, D., Nojima, T., Cramer, P., et al. (2020). CDK12 globally stimulates RNA polymerase II transcription elongation and carboxyl-terminal domain phosphorylation. Nucleic Acids Res 48, 7712–7727.

Terzi, N., Churchman, L.S., Vasiljeva, L., Weissman, J., and Buratowski, S. (2011). H3K4 Trimethylation by Set1 Promotes Efficient Termination by the Nrd1-Nab3-Sen1 Pathway. Molecular and Cellular Biology 31, 3569–3583.

Thomsen, N.D., and Berger, J.M. (2009). Running in Reverse: The Structural Basis for Translocation Polarity in Hexameric Helicases. Cell 139, 523–534.

Thomsen, N.D., Lawson, M.R., Witkowsky, L.B., Qu, S., and Berger, J.M. (2016). Molecular mechanisms of substrate-controlled ring dynamics and substepping in a nucleic acid-dependent hexameric motor. Proc Natl Acad Sci USA 113, E7691–E7700.

Titov, D.V., Gilman, B., He, Q.-L., Bhat, S., Low, W.-K., Dang, Y., Smeaton, M., Demain, A.L., Miller, P.S., Kugel, J.F., et al. (2011). XPB, a subunit of TFIIH, is a target of the natural product triptolide. Nature Chemical Biology 7, 182–188.

Tomar, S.K., and Artsimovitch, I. (2013). NusG-Spt5 Proteins—Universal Tools for Transcription Modification and Communication. Chem. Rev. 113, 8604–8619.

Tsai, K.-L., Yu, X., Gopalan, S., Chao, T.-C., Zhang, Y., Florens, L., Washburn, M.P., Murakami, K., Conaway, R.C., Conaway, J.W., et al. (2017). Mediator structure and rearrangements required for holoenzyme formation. Nature 544, 196–201.

Turtola, M., and Belogurov, G.A. (2016). NusG inhibits RNA polymerase backtracking by stabilizing the minimal transcription bubble. ELife 5, e18096.

Újvári, A., Pal, M., and Luse, D.S. (2002). RNA Polymerase II Transcription Complexes May Become Arrested If the Nascent RNA Is Shortened to Less than 50 Nucleotides. J. Biol. Chem. 277, 32527–32537.

Vilborg, A., Passarelli, M.C., Yario, T.A., Tycowski, K.T., and Steitz, J.A. (2015). Widespread Inducible Transcription Downstream of Human Genes. Molecular Cell 59, 449–461.

Viollet, S., Fuchs, R.T., Munafo, D.B., Zhuang, F., and Robb, G.B. (2011). T4 RNA ligase 2 truncated active site mutants: improved tools for RNA analysis. BMC Biotechnol 11, 72.

Voigt, F., Gerbracht, J.V., Boehm, V., Horvathova, I., Eglinger, J., Chao, J.A., and Gehring, N.H. (2019). Detection and quantification of RNA decay intermediates using XRN1-resistant reporter transcripts. Nature Protocols 14, 1603–1633.

Vos, S.M., Farnung, L., Urlaub, H., and Cramer, P. (2018a). Structure of paused transcription complex Pol II–DSIF–NELF. Nature 560, 601–606.

Vos, S.M., Farnung, L., Boehning, M., Wigge, C., Linden, A., Urlaub, H., and Cramer, P. (2018b). Structure of activated transcription complex Pol II–DSIF–PAF–SPT6. Nature 560, 607–612.

Voss, K., Forné, I., Descostes, N., Hintermair, C., Schüller, R., Maqbool, M.A., Heidemann, M., Flatley, A., Imhof, A., Gut, M., et al. (2015). Site-specific methylation and acetylation of lysine residues in the C-terminal domain (CTD) of RNA polymerase II. Transcription 6, 91–101.

Weber, C.M., Ramachandran, S., and Henikoff, S. (2014). Nucleosomes Are Context-Specific, H2A.Z-Modulated Barriers to RNA Polymerase. Molecular Cell 53, 819–830.

Wen, Y., and Shatkin, A.J. (1999). Transcription elongation factor hSPT5 stimulates mRNA capping. Genes Dev. 13, 1774–1779.

Werner, F. (2007). Structure and function of archaeal RNA polymerases. Molecular Microbiology 65, 1395–1404.

West, S., Proudfoot, N.J., and Dye, M.J. (2008). Molecular Dissection of Mammalian RNA Polymerase II Transcriptional Termination. Molecular Cell 29, 600–610.

Whitelaw, E., and Proudfoot, N. (1986). Alpha-thalassaemia caused by a poly(A) site mutation reveals that transcriptional termination is linked to 3′ end processing in the human alpha 2 globin gene. The EMBO Journal 5, 2915–2922.

Wilusz, J.E., Freier, S.M., and Spector, D.L. (2008). 3′ End Processing of a Long Nuclear-Retained Noncoding RNA Yields a tRNA-like Cytoplasmic RNA. Cell 135, 919–932.

Wong, K.H., Jin, Y., and Struhl, K. (2014). TFIIH Phosphorylation of the Pol II CTD Stimulates Mediator Dissociation from the Preinitiation Complex and Promoter Escape. Molecular Cell 54, 601–612.

Wu, C.-H., Yamaguchi, Y., Benjamin, L.R., Horvat-Gordon, M., Washinsky, J., Enerly, E., Larsson, J., Lambertsson, A., Handa, H., and Gilmour, D. (2003). NELF and DSIF cause promoter proximal pausing on the hsp70 promoter in Drosophila. Genes Dev. 17, 1402–1414.

Wuarin, J., and Schibler, U. (1994). Physical isolation of nascent RNA chains transcribed by RNA polymerase II: evidence for cotranscriptional splicing. Molecular and Cellular Biology 14, 7219–7225.

Xiang, S., Cooper-Morgan, A., Jiao, X., Kiledjian, M., Manley, J.L., and Tong, L. (2009). Structure and function of the 5'-->3' exoribonuclease Rat1 and its activating partner Rai1. Nature 458, 784–788.

Xue, Y., Bai, X., Lee, I., Kallstrom, G., Ho, J., Brown, J., Stevens, A., and Johnson, A.W. (2000). Saccharomyces cerevisiae RAI1 (YGL246c) Is Homologous to Human DOM3Z and Encodes a Protein That Binds the Nuclear Exoribonuclease Rat1p. Molecular and Cellular Biology 20, 4006–4015.

Yakhnin, H., Yakhnin, A.V., Mouery, B.L., Mandell, Z.F., Karbasiafshar, C., Kashlev, M., and Babitzke, P. (2019). NusG-Dependent RNA Polymerase Pausing and Tylosin-Dependent Ribosome Stalling Are Required for Tylosin Resistance by Inducing 23S rRNA Methylation in Bacillus subtilis. MBio 10.

Yamada, T., Yamaguchi, Y., Inukai, N., Okamoto, S., Mura, T., and Handa, H. (2006). P-TEFb-Mediated Phosphorylation of hSpt5 C-Terminal Repeats Is Critical for Processive Transcription Elongation. Molecular Cell 21, 227–237.

Yamamoto, K., and Ishihama, A. (2003). Two different modes of transcription repression of the Escherichia coli acetate operon by IclR. Molecular Microbiology 47, 183–194.

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

Yang, X., Sullivan, K.D., Marzluff, W.F., and Dominski, Z. (2009). Studies of the 5′ Exonuclease and Endonuclease Activities of CPSF-73 in Histone Pre-mRNA Processing. Molecular and Cellular Biology 29, 31–42.

Yang, X., Sun, Y., Aik, W.S., Marzluff, W.F., Tong, L., and Dominski, Z. (2020). Studies with recombinant U7 snRNP demonstrate that CPSF73 is both an endonuclease and a 5'-3' exonuclease. RNA rna.076273.120.

Yang, Y., Darbari, V.C., Zhang, N., Lu, D., Glyde, R., Wang, Y.-P., Winkelman, J.T., Gourse, R.L., Murakami, K.S., Buck, M., et al. (2015). Structures of the RNA polymerase-σ54 reveal new and conserved regulatory strategies. Science 349, 882–885.

Yanofsky, C. (1981). Attenuation in the control of expression of bacterial operons. Nature 289, 751–758.

Yesbolatova, A., Natsume, T., Hayashi, K., and Kanemaki, M.T. (2019). Generation of conditional auxin-inducible degron (AID) cells and tight control of degron-fused proteins using the degradation inhibitor auxinole. Methods 164–165, 73–80.

Yue, L., Li, J., Zhang, B., Qi, L., Zhao, F., Li, L., and Dong, X. (2019). aCPSF1 controlled archaeal transcription termination: a prototypical eukaryotic model. BioRxiv 843821.

Zasadzińska, E., Huang, J., Bailey, A.O., Guo, L.Y., Lee, N.S., Srivastava, S., Wong, K.A., French, B.T., Black, B.E., and Foltz, D.R. (2018). Inheritance of CENP-A Nucleosomes during DNA Replication Requires HJURP. Dev Cell 47, 348-362.e7.

Zenkin, N. (2014). Ancient RNA stems that terminate transcription. RNA Biology 11, 295–297.

References

Zhang, Z., and Gilmour, D.S. (2006). Pcf11 Is a Termination Factor in Drosophila that Dismantles the Elongation Complex by Bridging the CTD of RNA Polymerase II to the Nascent Transcript. Molecular Cell 21, 65–74.

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

Zhang, Y., Feng, Y., Chatterjee, S., Tuske, S., Ho, M.X., Arnold, E., and Ebright, R.H. (2012). Structural Basis of Transcription Initiation. Science 338, 1076–1080.

Zuo, Y., and Steitz, T.A. (2015). Crystal Structures of the E. coli Transcription Initiation Complexes with a Complete Bubble. Molecular Cell 58, 534–540.

Appendix

Publications Pertaining to This Thesis

In chronological order:

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

Eaton, J.D., and West, S. (2018). An end in sight? Xrn2 and transcriptional termination by RNA polymerase II. Transcription 9, 321–326.

Davidson, L., Francis, L., Cordiner, R.A., **Eaton, J.D.**, Estell, C., Macias, S., Cáceres, J.F., and West, S. (2019). Rapid Depletion of DIS3, EXOSC10, or XRN2 Reveals the Immediate Impact of Exoribonucleolysis on Nuclear RNA Metabolism and Transcriptional Control. Cell Reports 26, 2779-2791.e5.

Eaton, J.D., Francis, L., Davidson, L., and West, S. (2020). A unified allosteric/torpedo mechanism for transcriptional termination on human protein-coding genes. Genes Dev. 34, 132–145.

Eaton, J.D., and West, S. (2020). Termination of Transcription by RNA Polymerase II: BOOM! Trends in Genetics 36, 664–675.

Davidson, L., Francis, L., **Eaton, J.D.**, and West, S. (2020). Integrator-Dependent and Allosteric/Intrinsic Mechanisms Ensure Efficient Termination of snRNA Transcription. Cell Reports 33, 108319.

Highlighted Publications

i. Xrn2 accelerates termination by RNA polymerase II, which is underpinned by CPSF73 activity

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

Xrn2 accelerates termination by RNA polymerase II, which is underpinned by CPSF73 activity

Joshua D. Eaton,[1,6] Lee Davidson,[1,2,6] David L.V. Bauer,[3] Toyoaki Natsume,[4,5] Masato T. Kanemaki,[4,5] and Steven West[1]

[1]The Living Systems Institute, University of Exeter, Exeter EX4 4QD, United Kingdom; [2]Department of Molecular Biology and Biotechnology, University of Sheffield, Sheffield S10 2TN, United Kingdom; [3]Sir William Dunn School of Pathology, University of Oxford, Oxford OX1 3RE, United Kingdom; [4]Division of Molecular Cell Engineering, National Institute of Genetics, Research Organization of Information and Systems (ROIS), Mishima, Shizuoka 411-8540, Japan; [5]Department of Genetics, Graduate University for Advanced Studies (SOKENDAI), Mishima, Shizuoka 411-8540, Japan

Termination is a ubiquitous phase in every transcription cycle but is incompletely understood and a subject of debate. We used gene editing as a new approach to address its mechanism through engineered conditional depletion of the 5′ → 3′ exonuclease Xrn2 or the polyadenylation signal (PAS) endonuclease CPSF73 (cleavage and polyadenylation specificity factor 73). The ability to rapidly control Xrn2 reveals a clear and general role for it in cotranscriptional degradation of 3′ flanking region RNA and transcriptional termination. This defect is characterized genome-wide at high resolution using mammalian native elongating transcript sequencing (mNET-seq). An Xrn2 effect on termination requires prior RNA cleavage, and we provide evidence for this by showing that catalytically inactive CPSF73 cannot restore termination to cells lacking functional CPSF73. Notably, Xrn2 plays no significant role in either Histone or small nuclear RNA (snRNA) gene termination even though both RNA classes undergo 3′ end cleavage. In sum, efficient termination on most protein-coding genes involves CPSF73-mediated RNA cleavage and cotranscriptional degradation of polymerase-associated RNA by Xrn2. However, as CPSF73 loss caused more extensive readthrough transcription than Xrn2 elimination, it likely plays a more underpinning role in termination.

[*Keywords*: Xrn2; transcriptional termination; CPSF73; torpedo; allosteric; RNA polymerase II]

Supplemental material is available for this article.

Received October 23, 2017; revised version accepted January 5, 2018.

Transcriptional termination can be defined as the cessation of RNA polymerization and dissolution of the ternary complex of RNA polymerase II (Pol II), DNA, and RNA. Termination is a biologically important process, as it prevents transcriptional interference of genes and ensures that polymerases are available for new rounds of gene expression. Despite the fact that all transcription ends this way, it is perhaps the least understood phase in the cycle. A polyadenylation signal (PAS) is a prerequisite for termination, and mutations within it were shown decades ago to cause extended transcriptional readthrough (Whitelaw and Proudfoot 1986; Connelly and Manley 1988). Two models, the allosteric and torpedo, have since framed efforts to understand PAS-dependent termination (Porrua and Libri 2015; Proudfoot 2016). In the allosteric mechanism, transcription of a PAS causes a change in Pol II structure or alters the composition of the elongation complex to promote termination. In the torpedo model, RNA cleavage generates a Pol II-associated RNA substrate for 5′ → 3′ degradation that triggers termination by pursuing and catching the polymerase (Connelly and Manley 1988; Proudfoot 1989). Multiple studies provide support for both models, with the actual mechanism likely to incorporate aspects of each. However, their relative contributions are debated due to different results obtained in a variety of experimental systems (Libri 2015).

Early support for the torpedo model came from observations that depletion of the nuclear 5′ → 3′ exonuclease Xrn2 caused termination defects on transfected plasmids (West et al. 2004). Its homolog, Rat1, was simultaneously found to promote termination more widely in budding yeast (Kim et al. 2004), with recent transcriptome-wide analysis supporting this finding (Baejen et al. 2017). The broader role of Xrn2 in human cells has been less clear. RNAi of Xrn2 showed no general function in termination

[6]**These authors contributed equally to this work.**
Corresponding author: s.west@exeter.ac.uk
Article published online ahead of print. Article and publication date are online at http://www.genesdev.org/cgi/doi/10.1101/gad.308528.117. Freely available online through the *Genes & Development* Open Access option.

at the 3′ ends of protein-coding genes (Nojima et al. 2015), but a significant effect was later observed upon concurrent expression of catalytically dead Xrn2 (Fong et al. 2015). It is likely that the inactive protein binds Xrn2 substrates and blocks their degradation by the diminished levels of endogenous Xrn2. As such, RNAi may not always reveal the complete set of functions for some proteins.

Rat1 was shown to promote the recruitment of some polyadenylation factors to budding yeast genes and so may sometimes affect termination indirectly through impacting PAS function (Luo et al. 2006). In this instance, cotranscriptional degradation of PAS-cleaved RNA was insufficient to cause termination on some genes, highlighting the possibility that RNA degradation may not always release polymerase (Luo et al. 2006). Even so, catalytically inactive Rat1 does not support termination on other yeast genes, and Rat1, Xrn1, and Xrn2 can all dissociate Pol II from DNA in purified systems (Kim et al. 2004; Park et al. 2015). In *Caenorhabditis elegans*, Xrn2 depletion does not affect termination on the majority of protein-coding genes, suggesting that the torpedo mechanism is less widely used in that organism (Miki et al. 2017).

To understand the extent to which the allosteric and torpedo models explain the termination mechanism, it is important to distinguish the role of PAS recognition from PAS cleavage, which is difficult to do in vivo. A human PAS is recognized by several multisubunit complexes that bind to its AAUAAA hexamer and downstream G/U-rich motif (Proudfoot 2012). AAUAAA is recognized by the CPSF30 and WDR33 subunits of cleavage and polyadenylation specificity factor (CPSF), with endonuclease activity provided by CPSF73 (Mandel et al. 2006; Shi et al. 2009; Chan et al. 2014; Schonemann et al. 2014). Although CPSF73 was identified as the nuclease over a decade ago (Mandel et al. 2006), its function in termination is not fully characterized. This issue has been tackled using in vitro systems competent for transcription and RNA processing, which revealed that a PAS can promote termination in the absence of cleavage (Zhang et al. 2015). While highlighting the capacity of PAS recognition to affect Pol II activity, it is unknown whether this mechanism promotes termination in cells.

Therefore, several aspects of termination in human cells are incompletely understood, especially in terms of their generality, and understanding of the process has lagged behind that of other model organisms. It is not known whether Xrn2 degrades PAS-cleaved RNA generally or whether this process is cotranscriptional, as was envisaged in the torpedo model. Possible effects of Xrn2 on PAS cleavage are also not established in a global manner. It is also unclear whether PAS cleavage is required for termination or whether polymerase release can be promoted by cleavage-independent factors, which is an issue that has an impact on the applicability of current models.

As RNAi approaches take days and since protein depletion is often incomplete, we adopted gene editing to engineer conditional depletion of Xrn2 or CPSF73 on faster time scales. This was used to show that Xrn2 degrades the 3′ product of PAS cleavage cotranscriptionally and promotes efficient termination genome-wide, which we mapped transcriptome-wide at high resolution. Importantly, we show that CPSF73 activity is required for efficient termination, confirming a primary mechanism in which PAS cleavage precedes degradation of polymerase-associated RNA. However, CPSF73 elimination causes stronger termination defects than the loss of Xrn2, suggesting that it might promote termination by additional mechanisms when the primary process fails.

Results

An auxin-inducible degron (AID) system for rapid Xrn2 depletion

To set up a system for rapid elimination of Xrn2, CRISPR/Cas9 was used to tag *XRN2* with an AID (Fig. 1A,B). AID-tagged proteins are degraded upon addition of indole-3-acetic acid (referred to here as auxin [IAA]) in a manner dependent on plant Tir1 protein (Nishimura et al. 2009; Natsume et al. 2016). HCT116 cells were chosen for this experiment due to their diploid nature. Cells expressing Tir1 were subjected to CRISPR/Cas9 genome editing using repair templates that incorporated three tandem mini-AID degrons and hygromycin or neomycin selection markers (Kubota et al. 2013; Natsume et al. 2016). Selection markers were separated from the tag by a P2A sequence that was cleaved during translation (Kim et al. 2011). Transfection of these two constructs together with an *XRN2*-specific guide RNA expressing Cas9 plasmid yielded multiple resistant colonies, and homozygous modification was demonstrated by PCR (Fig. 1C).

Western blotting confirmed homozygous targeting in two selected positive clones, shown by the higher-molecular-weight Xrn2 and the absence of any signal at the size expected for native Xrn2 (Fig. 1D). It is notable that Xrn2-AID is present at lower levels than endogenous Xrn2, suggesting a destabilizing effect of the tag. Even so, *XRN2-AID* cells showed no growth defects (Supplemental Fig. 1A). Further RNA analyses performed throughout this study also showed that RNA degradation functions are virtually unimpaired in *XRN2-AID* cells.

To test Xrn2-AID depletion, Western blotting was performed over a time course of auxin addition (Fig. 1E). Xrn2-AID was detected through the Flag epitope present within the AID tag, with specificity shown by a lack of signal in unmodified HCT116 cells. Importantly, Xrn2-AID levels are reduced within 30 min of auxin treatment and were virtually undetectable after 1 h. As such, this system allows rapid and conditional depletion of Xrn2. The addition of auxin to the culture medium of *XRN2-AID* cells completely prevented cell colony formation, showing that Xrn2 is an essential protein (Supplemental Fig. 1B).

Xrn2 plays a general role in the degradation of 3′ flanking region RNA

Next, we tested the effect of Xrn2 loss on PAS cleavage and the stability of 3′ flanking region RNA from *MYC* and *ACTB* using quantitative RT–PCR (qRT–PCR). RNA was isolated over the same time course as for the

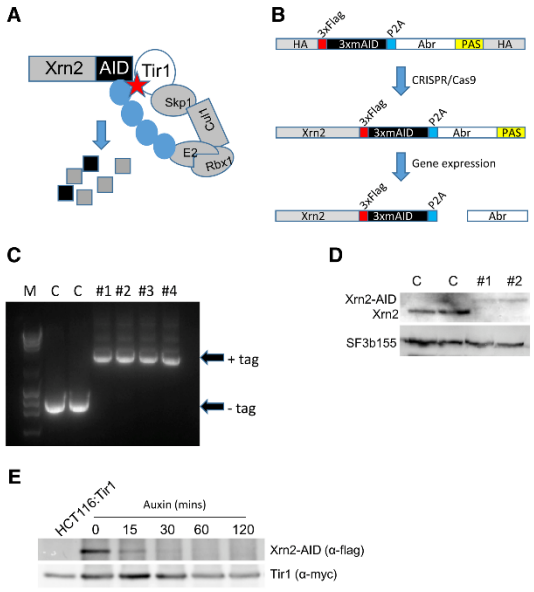


Figure 1. (A) Diagram showing the basis of auxin-dependent depletion of AID-tagged proteins. In the presence of auxin (star), Tir1 facilitates ubiquitination (blue circles) of the AID tag and rapid protein degradation. (*B*) Strategy for AID tagging of Xrn2. Homology arms (HAs) flanked repair cassettes containing 3× miniAID sequences, preceded by a Flag tag and separated from an antibiotic resistance gene (denoted as Abr and either Neo or Hyg) by a P2A cleavage site, with 3′ end processing driven by an SV40 PAS. (*C*) Diagnostic PCR of genomic DNA from antibiotic-resistant cell colonies following CRISPR gene editing. The presence of a tag increases the size of the PCR product compared with the smaller product derived from the unmodified gene. Homozygous modification is shown by the lack of unmodified product in the four drug-resistant colonies (#1–#4). (M) DNA marker. (*D*) Western blot confirmation of Xrn2 tagging. The *top* panel shows Xrn2 in two unmodified cell samples (C) and two gene-edited colonies (#1 and #2). Successful biallelic tagging is shown by the higher-molecular-weight species and the lack of native-sized Xrn2 in CRISPR-modified cells. SF3b155 was probed for as a loading control. (*E*) Time course of auxin addition on *XRN2-AID* cells. Xrn2-AID was detected by anti-Flag, and specificity is shown by the lack of product in Tir1 HCT116 cells, which are not modified at *XRN2*. Tir1 was probed for as a loading control via its myc tag.

Western blot in Figure 1E, and primers were used to detect non-PAS-cleaved (UCPA) RNA or 3′ flanking transcripts (Fig. 2A). An accumulation of 3′ flanking region RNA was seen for both genes by 30 min of auxin treatment. An even greater effect was seen after 60 min that was maintained (but not enhanced) after 120 min. In contrast, Xrn2-AID loss had no obvious effect on PAS cleavage, as no accumulation of UCPA species was observed for either gene at any time point. This experiment shows that in these two cases, Xrn2 degrades RNA beyond the PAS without affecting PAS cleavage. The latter conclusion is further supported by observations that Xrn2-AID loss has no impact on the recruitment of the polyadenylation factor Pcf11 to *ACTB* (Supplemental Fig. 2A). Importantly, 3′ flanking region RNA was stabilized only in the combined presence of the AID tag, Tir1, and auxin, showing that no individual factor indirectly causes the effect (Supplemental Fig. 2B). These findings are unlikely to result from secondary effects due to the speed of Xrn2-AID depletion, especially by comparison with RNAi, with the near-complete elimination of Xrn2-AID revealing function without overexpression of the inactive protein.

We then sought to test the generality of the effects seen on Xrn2-AID loss using nuclear RNA sequencing (RNA-seq) carried out on *XRN2-AID* cells treated with auxin or untreated. We also performed this analysis on a HCT116 cell line that was unmodified at *XRN2* and grown in the absence of auxin. Analysis of individual gene tracks confirmed the effect on *MYC* and *ACTB*, where an enhanced signal beyond their PASs was observed upon Xrn2-AID elimination (Fig. 2B). Further examples of Xrn2 effects are shown for *E2F6* and *RPL30* (Fig. 2C). *XRN2-AID* cells grown in the absence of auxin gave slightly elevated levels of 3′ flanking RNA as compared with cells unmodified at *XRN2*, suggesting that Xrn2-AID can carry out almost all 3′ flanking RNA degradation. Interestingly, strong effects of Xrn2 depletion were seen downstream from where Drosha cleaves microRNA (miRNA) precursors (Supplemental Fig. S3A,B), showing other ways of Xrn2 substrate generation.

Metagene plots were then generated for protein-coding genes that were separated from any reads within 3 kb of their transcription start site (TSS) and 7 kb of the PAS (denoted as TES [transcript end site]). This left 4701 genes for analysis and revealed a clear enhancement of 3′ flanking region RNA upon auxin treatment of *XRN2-AID* cells (Fig. 2D). Xrn2-AID samples obtained in the absence of auxin showed slightly raised levels of 3′ flanking region RNA compared with the cell line unmodified at *XRN2*, arguing that reduced levels of Xrn2-AID do not cause significant readthrough defects. Metagene plots generated from an independent biological replicate showed a similar result (Supplemental Fig. 3C). We note that Xrn2-AID loss is associated with a slight reduction in reads upstream of the PAS, potentially reflecting mildly reduced gene expression that might be caused by Pol II recycling defects. Finally, closer analysis of the TES (PAS) region showed that read counts at this position are similar in all samples (Fig. 2E; Supplemental Fig. 3D). This again suggests that major PAS cleavage defects are not widespread following Xrn2 loss, which is consistent with the analysis of *MYC* and *ACTB* shown above.

Xrn2 degrades 3′ flanking RNA cotranscriptionally and promotes termination

The validity of the torpedo model of termination depends on cotranscriptional degradation of 3′ flanking region RNA taking place (Connelly and Manley 1988; Proudfoot 1989), but this has not been shown for Xrn2. To address this, we immunoprecipitated Pol II-associated RNA following cross-linking of *XRN2-AID* cells treated with

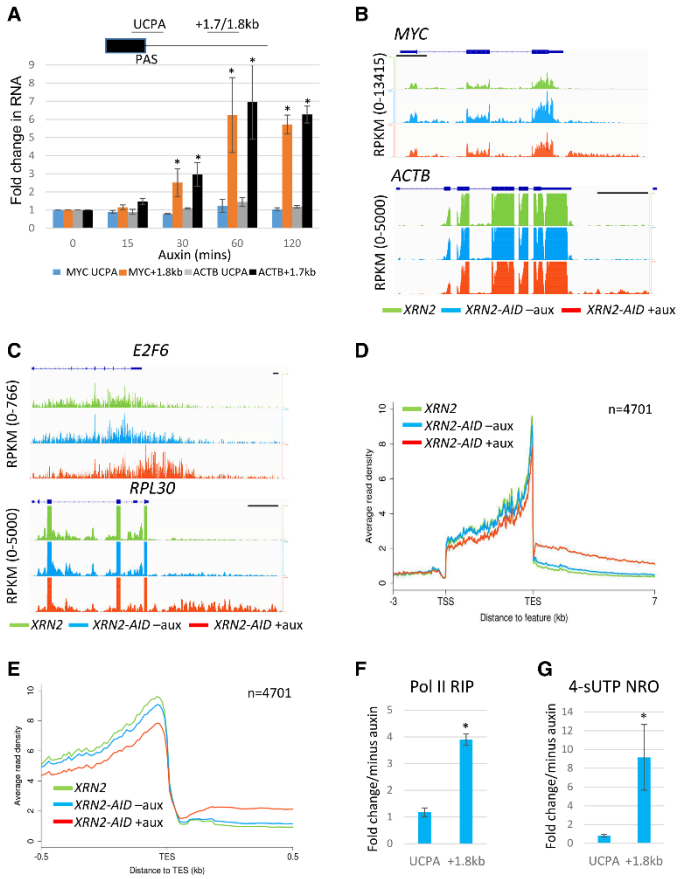


Figure 2. (*A*) qRT–PCR analysis of UCPA and 3′ flanking RNA from *MYC* and *ACTB* genes from total RNA during a time course of auxin addition. Values are plotted relative to those obtained at t0 after normalization to unspliced RNA levels from the respective genes. The diagram depicts the positions of UCPA amplicons and 3′ flank amplicons for both genes (+1.7 kb for *ACTB* and +1.8 kb for *MYC*). Asterisks denote $P < 0.05$ for changes relative to t0 in the absence of auxin. (*B*) Nuclear RNA sequencing (RNA-seq) traces of *MYC* and *ACTB* genes in samples obtained from *XRN2* unmodified cells and *XRN2-AID* cells treated with auxin for 1 h or untreated. The *Y*-axis shows RPKM (reads per kilobase transcript per million mapped reads). Bars, 1 kb. (*C*) As in *B* but showing *E2F6* and *RPL30* genes. (*D*) Metagene plots from nuclear RNA-seq on *XRN2* unmodified cells and *XRN2-AID* cells treated with auxin or untreated. The graph shows the region from 3 kb upstream of the transcription start site (TSS) up to 7 kb beyond the PAS (denoted as transcript end site [TES]). (*E*) A zoomed in view of ±0.5 kb of the TES from the same metagene presented in *D*. (*F*) Pol II RNA immunoprecipitation analysis of UCPA and 3′ flanking (+1.8 kb) RNA from *MYC* in cells depleted of Xrn2-AID (1 h of auxin treatment) or not. Quantitation is shown for +auxin samples relative to –auxin after normalizing to the level of unspliced *MYC* RNA. The asterisk denotes the difference between +auxin and –auxin, where $P < 0.05$. (*G*) 4-thio UTP (4sUTP) nuclear run-on (NRO) analysis of UCPA and 3′ flanking (+1.8 kb) RNA from *MYC* in cells depleted of Xrn2-AID (1 h of auxin treatment)or not . Quantitation is shown for +auxin samples expressed relative to –auxin after normalizing to the level of unspliced *MYC* RNA. The asterisk denotes the difference between +auxin and –auxin, where $P < 0.05$. All error bars show standard deviation from at least three independent experiments.

auxin or untreated and analyzed it by qRT–PCR (Fig. 2F). Levels of UCPA RNA and 3′ flanking region RNA produced from *MYC* were assayed, and Xrn2 loss caused a substantial increase in the latter but not the former. This is consistent with Xrn2 involvement in the cotranscriptional degradation of 3′ flanking region RNA.

As a second measure of cotranscriptional degradation, we isolated nuclei from control or auxin-treated *XRN2-AID* cells and subjected them to nuclear run-on (NRO) analysis in the presence of 4-thio UTP (4sUTP). In this experiment, transcriptionally engaged Pol II was allowed to run on and label the 3′ ends of nascent transcripts in vitro. These were purified via linkage of biotin onto 4sUTP followed by streptavidin capture (see the Materials and Methods) and subjected to qRT–PCR to analyze UCPA and 3′ flanking region transcripts from *MYC* (Fig. 2G). This experiment yielded a result similar to that shown in Figure 2F in that Xrn2 loss increased 3′ flanking region RNA but not UCPA transcripts. The analysis of additional genes confirmed the role of Xrn2 in cotranscriptional degradation of 3′ flanking region RNA (Supplemental Figure 3E,F). Finally, stable integration of wild-type or catalytically inactive (D235A) *XRN2* into *XRN2-AID* cells demonstrated that both RNA degradation and termination defects caused by Xrn2-AID elimination are completely rescued by wild-type Xrn2 but not by D235A (Supplemental Fig. 4). The Xrn2 effects on transcriptional termination therefore require its exoribonuclease function.

Mammalian native elongating transcript sequencing (mNET-seq) reveals a global termination defect on Xrn2 loss

Next, we precisely interrogated the global function of Xrn2 in transcriptional termination using mNET-seq (Nojima et al. 2015). In this method, the position of Pol II is revealed genome-wide at single-nucleotide resolution through its immunoprecipitation and the deep sequencing of RNA extracted from its active site. An antibody was used to capture all forms of Pol II.

MYC and *RPL30* mNET-seq profiles are shown in Figure 3, A and B (*ACTB* in Supplemental Fig. 5A). In cells not treated with auxin, termination occurs downstream from the PAS, where the mNET-seq signal reaches background.

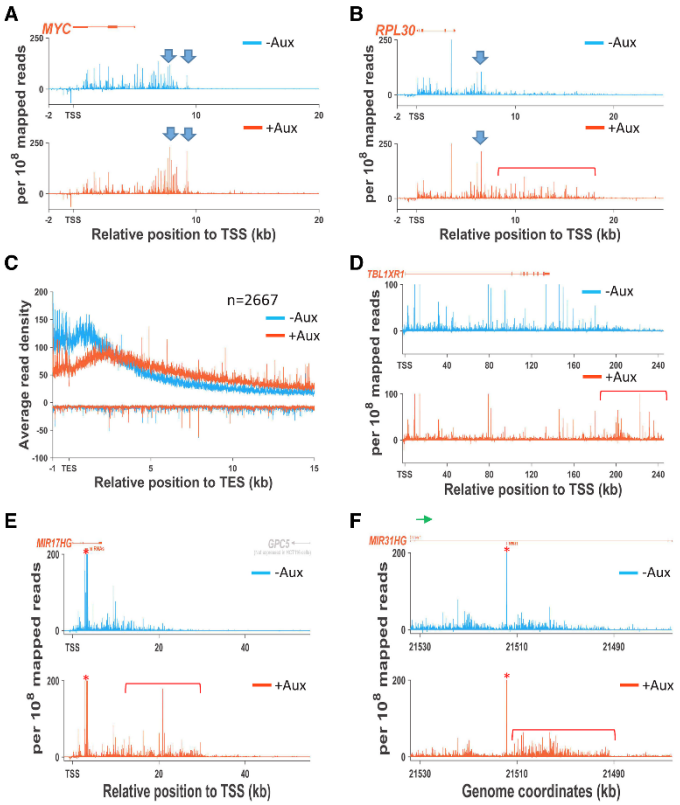


Figure 3. (*A*) *MYC* mNET-seq trace from *XRN2-AID* cells treated (orange) with auxin or untreated (blue) for 2 h. The *X*-axis shows a position relative to the gene TSS in kilobases. Reads are plotted as abundance per 10^8 reads. Blue arrows denote a signal enhanced in the absence of Xrn2. (*B*) As in *A* but for *RPL30*. Additionally, the red bracket marks readthrough upon Xrn2 loss. (*C*) Metagene plot to analyze transcriptional termination on protein-coding genes in mNET-seq data from *XRN2-AID* cells grown with or without auxin. The average read density is shown over positions extending from 1 kb upstream of the TES to 15 kb downstream. The signal less than zero is transcription from the opposite strand, which is at or close to background. (*D*) As in *A* but for *TBL1XR1*. The red bracket denotes the region of extended readthrough. (*E*) As in *A* but for *MIR17HG*. In this case, a red asterisk marks the miRNA cleavage events, and a red bracket marks readthrough. (*F*) As in *E* but for *MIR31HG*. In each diagram, the expressed gene is shown in orange, with nonexpressed genes in gray.

When Xrn2 is eliminated, a clear termination defect is observed, and, due to the high resolution of mNET-seq, it is possible to visualize two manifestations of this. First, where flanking region signal is detected in control cells, it is frequently elevated over the same positions in cells lacking Xrn2. This can be seen in the *MYC* and *RPL30* examples in Figure 3, A and B (blue arrows), and is consistent with polymerase stalling over termination regions facilitating termination by Xrn2. While this provides evidence that Xrn2 might not always have to pursue a still-transcribing Pol II, an additional effect of Xrn2 loss is an enhanced mNET-seq signal beyond where termination takes place in control cells. An example of this is marked by the red bracket on the *RPL30* gene plot in Figure 3B and suggests that normal termination sites can be ignored, with polymerases potentially having escaped pursuit by Xrn2.

We next addressed the generality of Xrn2 function in termination by generating metagene plots from control and auxin-treated cells. We analyzed expressed genes separated from upstream and downstream reads by at least 1 and 15 kb, respectively, which revealed a general transcriptional termination defect upon loss of Xrn2 (Fig. 3C). Interestingly, mNET-seq signal declined even in the absence of Xrn2, suggesting the existence of termination mechanisms that do not depend on it. The metagene plot of a separate biological replicate of this experiment showed the same general termination defect upon Xrn2-AID loss (Supplemental Fig. 5B). Interestingly, some genes were especially sensitive to Xrn2 elimination and showed more extensive readthrough than the genome-wide trend —as exemplified by *TBL1XR1* in Figure 3D. Nuclear RNA-seq analysis confirmed the extended readthrough over *TBL1XR1* (Supplemental Fig. 5C).

PAS cleavage is not the only mechanism to generate RNA 3′ ends. For instance, Drosha processes miRNAs, and a small number of noncoding RNA genes use this mechanism of 3′ end formation (Dhir et al. 2015). We tested whether Xrn2 promoted termination of two examples of these long noncoding primary miRNAs (lnc-pri-miRNAs): *MIR17HG* and *MIR31HG* (Fig. 3E,F). Cotranscriptional miRNA cleavage is visible (Fig. 3E,F, red asterisks) in both cases due to the known capacity of mNET-seq to detect Drosha cleavage products that remain associated with transcribing Pol II (Nojima et al. 2015). For *MIR17HG*, nascent transcription is detected in Xrn2 depleted samples beyond where termination occurs in the control experiment. There is also a higher read density beyond the *MIR31HG* miRNA sequence upon Xrn2 loss, with a noticeable defect emphasized by the reduced read count upstream of the Drosha

cleavage site. This supports the notion that Xrn2 promotes efficient transcriptional termination from multiple cleavage processes, as suggested previously (Fong et al. 2015).

Transcriptional termination on Histone and small nuclear RNA (snRNA) genes is unaffected by Xrn2 loss

Although not polyadenylated, Histone RNAs also use CPSF for 3′ end formation, which could provide an entry site for Xrn2 (Dominski et al. 2005; Kolev et al. 2008), and we were interested in whether this was the case. Figure 4A shows mNET-seq traces of the *HIST1* cluster in *XRN2-AID* cells treated with auxin or untreated. Interestingly, there is no impact of Xrn2 loss on transcriptional termination of any of the genes in this cluster, strongly suggesting that Xrn2 does not play a prominent role in Histone gene termination. This result was confirmed for other examples of Histone genes, and, similarly, RNA-seq showed little to no effect of Xrn2 elimination on 3′ flanking region RNA deriving from these genes (Supplemental Fig. 6). snRNAs also undergo 3′ end cleavage by the integrator complex, and this may also precede Xrn2 activity (Baillat et al. 2005). However, as for Histone genes, our mNET-seq and RNA-seq analyses showed no major role for Xrn2 in their transcriptional termination or in the degradation of their 3′ flanking region transcripts (Fig. 4B; Supplemental Fig. 7). As such, 3′ end cleavage is not always sufficient to promote an Xrn2-dependent termination process.

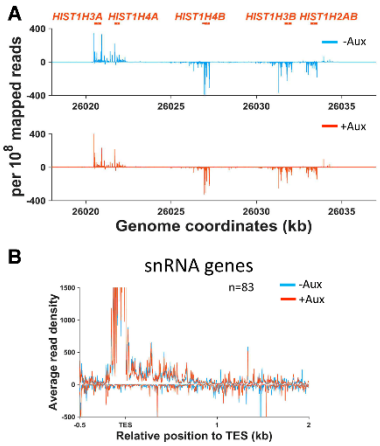


Figure 4. (*A*) mNET-seq profiles over the *HIST1* cluster from *XRN2-AID* cells treated with auxin or untreated. The *Y*-axes show signals per 10^8 mapped reads. It should be noted that reads <0 represent examples of Histone genes expressed on the opposite strand. (*B*) mNET-seq metagene analyses of snRNA genes from *XRN2-AID* cells treated with auxin or untreated. The *Y*-axes show average read density and are scaled to zoom into the termination region where signals are much lower than the snRNA gene body.

Conditional depletion of CPSF73 causes a strong PAS cleavage and termination defect

For Xrn2 to function in termination, RNA cleavage is required, and this presumably occurs most often at the PAS. CPSF73 is the PAS endonuclease in humans, and its depletion by RNAi causes strong termination defects genome-wide, confirming its general function in the process (Nojima et al. 2015). However, depletion of CPSF73 cannot establish whether its catalytic center or physical presence underlies its function in termination. To begin testing this, we generated cells in which the PAS endonuclease CPSF73 could be manipulated in a manner similar to Xrn2-AID. As we were unable to make an AID-tagged version of CPSF73, we tagged its C terminus with an *Escherichia coli* DHFR-based degron using the system used for Xrn2-AID (Iwamoto et al. 2010; Sheridan and Bentley 2016). In this system, cells are grown in the presence of trimethoprim (TMP), the withdrawal of which triggers degradation of the tagged protein. Western blotting confirmed homozygous tagging of *CPSF73* with *DHFR*, as CPSF73-DHFR was seen to migrate at a higher molecular weight than the native protein for which there was no signal in the CRISPR-modified cell line (Fig. 5A). Withdrawal of TMP from the medium promoted near elimination of CPSF73-DHFR after 10 h. This rate of depletion is slower than for Xrn2-AID but more than sevenfold faster than what we used previously for functional depletion of CPSF73 by RNAi (Davidson et al. 2014).

We tested the impact of CPSF73-DHFR elimination on 3′ end processing of *MYC* and *ACTB* transcripts by qRT–PCR of total RNA from *CPSF73-DHFR* cells grown in the presence or absence of TMP (Fig. 5B). For both genes, there was a significant reduction of PAS cleavage, demonstrated by an accumulation of UCPA RNA. Notably, the magnitude of effect (sevenfold to 12-fold) was threefold to fourfold greater than we observed previously by RNAi of CPSF73 (Davidson et al. 2014), highlighting the enhanced effects gained from this system.

To analyze the effect of CPSF73 depletion on termination, Pol II chromatin immunoprecipitation (ChIP) was performed in *CPSF73-DHFR* cells grown in the presence or absence of TMP. Pol II occupancy was monitored downstream from *MYC* and *ACTB* (Fig. 5C,D). In both cases, CPSF73 loss caused a general reduction in transcription, as evidenced by the lower Pol II signal upstream of the PAS (denoted as US). This is consistent with observations that PAS mutations or polyadenylation factor depletion negatively impacts transcription (Mapendano et al. 2010). Despite this, a large termination defect was evident on both genes through the accumulation of Pol II beyond the normal site of termination.

CPSF73 elimination causes more extensive readthrough than loss of Xrn2

We next tested whether CPSF73 and Xrn2 produced differential effects on readthrough transcription. For this, Pol II ChIP was compared for *CPSF73-DHFR* cells ±TMP, on *XRN2-AID* cells, and on *D235A XRN2-AID* cells +auxin

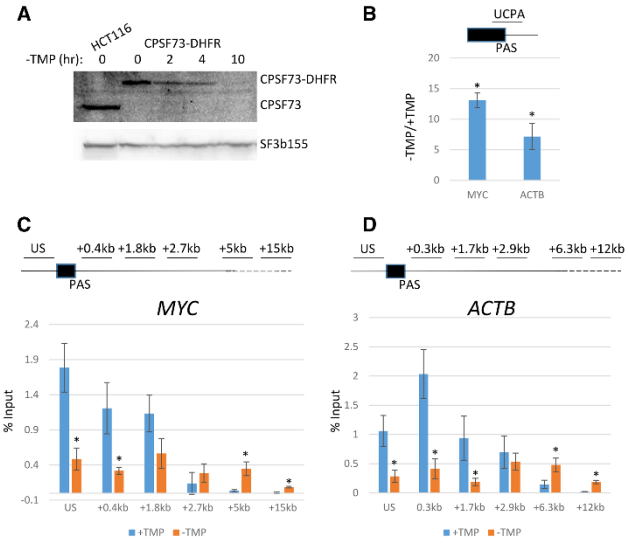


Figure 5. (*A*) Western blot showing successful tagging of *CPSF73* with DHFR and a time course of CPSF73-DHFR depletion in the absence of TMP. The *top* panel shows native CPSF73 in unmodified HCT116 cells and the higher-molecular-weight CPSF73-DHFR in CRISPR-modified cells. CPSF73-DHFR levels are depleted in the absence of TMP. SF3b155 was detected as a loading control. (*B*) qRT–PCR analysis of UCPA RNA from *MYC* or *ACTB* genes in CPSF73-DHFR cells grown in the presence or absence of TMP. Values are expressed relative to those obtained in cells grown in TMP after normalizing to unspliced RNA levels from each gene to account for any effects of transcription. Asterisks denote $P < 0.05$ for differences between +TMP and −TMP. (*C*) Pol II chromatin immunoprecipitation (ChIP) on *MYC* in *CPSF73-DHFR* cells grown in the presence or absence of TMP. Values are expressed as the percentage of input, and asterisks denote differences between +TMP and −TMP samples with $P < 0.05$. (*D*) As in *C* but on *ACTB*. All error bars show standard deviation from at least three independent experiments.

(Fig. 6A,B). *D235A XRN2-AID* cells stably express catalytically inactive Xrn2 that is not sensitive to auxin. When these cells are treated with auxin, 5′ → 3′ degradation of readthrough RNA and termination are more strongly impaired than in auxin-treated *XRN2-AID* cells (Supplemental Fig. 4). Pol II occupancy over extended readthrough regions of *MYC* and *ACTB* was plotted relative to the signal from upstream of the PAS. For both genes, CPSF73 depletion resulted in greater signals over extended positions than elimination of Xrn2 function, suggesting that termination is more adversely effected by loss of CPSF73. qRT–PCR analysis of readthrough RNA over the same positions confirmed this result (Supplemental Fig. 8A). Inhibition of CPSF30 function by influenza NS1A protein (Nemeroff et al. 1998) also caused more extensive transcriptional readthrough than Xrn2, further arguing for a more crucial function of CPSF in promoting termination (Supplemental Fig. 8B,C).

Although auxin-treated *D235A* cells represent the scenario most lacking in 5′ → 3′ degradation of RNA, the smaller effect on termination relative to CPSF73 loss may be due to incomplete Xrn2 depletion or other 5′ → 3′ nucleases acting in its absence. To address this, we analyzed the turnover rate of 3′ flanking region transcripts from *MYC* and *ACTB* in more detail. A time course was used in *XRN2-AID* cells treated with auxin or untreated and in *D235A* cells treated with auxin following transcriptional inhibition by actinomycin D (Act D) (Fig. 6C). In *XRN2-AID* cells not treated with auxin, Act D induced a strong reduction in the level of 3′ flanking region RNA, consistent with rapid degradation. The addition of auxin resulted in greater recovery of 3′ flanking region RNA following Act D treatment that was more pronounced in *D235A* cells treated with auxin. This confirms the role of Xrn2 in their degradation. However, degradation was incompletely blocked by Xrn2 elimination, as ~40%–60% of these transcripts were still degraded after transcriptional inhibition even in auxin-treated *D235A* cells.

The degradation of 3′ flanking region RNA in auxin-treated *D235A* cells could be by alternative 5′ → 3′ exonucleases or from the 3′ end by the exosome. To address this, we treated *D235A* cells with control or human Rrp40 (hRrp40)-specific siRNAs before auxin addition (Fig. 6D; Supplemental Fig. 9A,B). The same experiment was performed on *XRN2-AID* cells not treated with auxin to determine any exclusive effects of hRrp40 depletion. We first tested the effects of these conditions on the levels of *MYC* and *ACTB* 3′ flanking region RNA. hRrp40 depletion alone gave no substantial effect, whereas auxin treatment of *D235A* cells gave the expected strong accumulation. When hRrp40 was depleted from *D235A* cells treated with auxin, there was an accumulation of 3′ flanking region RNA above what was seen upon manipulation of Xrn2 that was most marked for *MYC* transcripts. Therefore, the exosome contributes to readthrough RNA degradation in the absence of Xrn2 function. The level of UCPA transcripts was similar under each of these conditions, arguing that PAS cleavage is unaffected (Supplemental Fig. 9C).

Next, the impact of the exosome on degradation of 3′ flanking RNAs was assessed after 20 min of Act D treatment (Fig. 6E). In the *XRN2-AID* sample treated with control siRNA, Act D treatment caused depletion of *ACTB* and *MYC* flanking transcripts as expected, and hRrp40 depletion gave a similar result. In auxin-treated *D235A* cells, ~40%–60% of 3′ flanking region RNA was again degraded in the absence of Xrn2 function. Importantly, hRrp40 depletion from *D235A* cells grown in auxin essentially blocked degradation, as the level of RNA recovered after transcription inhibition was similar to before Act D addition. This shows that the exosome rather than other 5′ → 3′ exonucleases is responsible for the degradation of

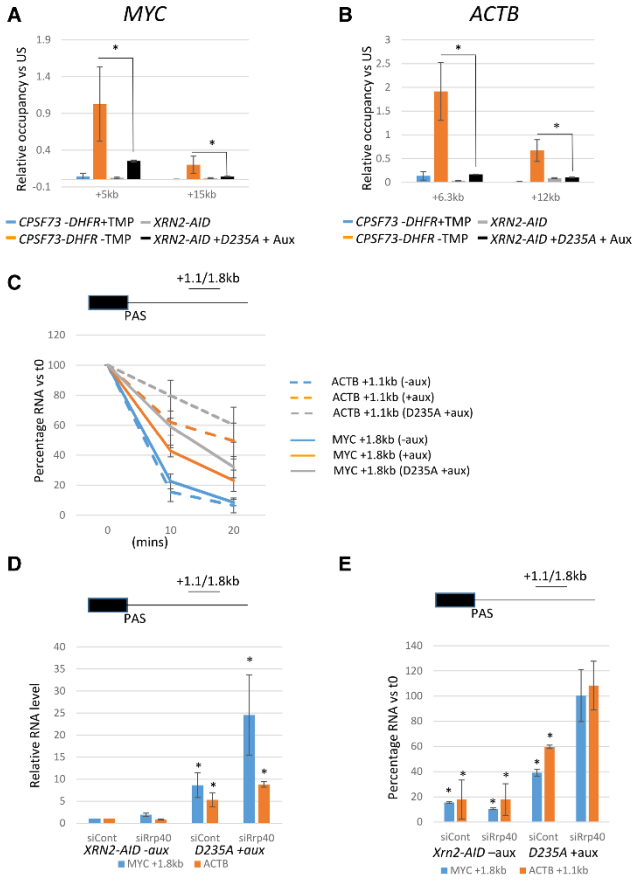


Figure 6. (*A*) Analysis of Pol II occupancy at +5 and +15 kb beyond the *MYC* PAS expressed relative to that upstream of the PAS (US) in *CPSF73-DHFR* cells ±TMP, *XRN2-AID* cells, and *XRN2-AID* + *D235A* cells +auxin (1 h). Asterisks denote $P < 0.05$ between CPSF73-DHFR – TMP and *XRN2-AID* + *D235A* + auxin. (*B*) As in *A* but for 6.3 and 12 kb beyond the *ACTB* PAS. (*C*) qRT–PCR analysis of *ACTB* and *MYC* 3′ flanking region RNA degradation in *XRN2-AID* cells treated with auxin or *D235A* cells treated with auxin (all auxin for 1 h) followed by 10 or 20 min of actinomycin D (Act D) treatment. For each sample set, RNA levels are expressed relative to that recovered at t0. (*D*) qRT–PCR analysis of *ACTB* and *MYC* 3′ flanking region RNA in control or human Rrp40 (hRrp40) siRNA-treated *XRN2-AID* cells or *D235A* cells treated with auxin (all auxin for 1 h). RNA levels are expressed as a fold change relative to those recovered in *XRN2-AID* cells treated with control siRNA following normalization to the level of unspliced *MYC* or *ACTB* transcripts. Asterisks denote $P < 0.05$ versus *XRN2-AID* cells treated with control siRNA. (*E*) qRT–PCR analysis of *ACTB* and *MYC* 3′ flanking region RNA under the conditions used in *D* but after 20 min of Act D treatment. Values are expressed as a percentage of RNA remaining under each condition relative to the amounts recovered in each sample at t0. Asterisks denote $P < 0.05$ versus the 0 time point. All error bars show standard deviation from at least three independent experiments.

RNA that occurs in the absence of functional Xrn2. As such, auxin treatment of D235A cells effectively blocks degradation of 3′ flanking region transcripts from their 5′ ends. A similar result was obtained when transcription was inhibited using flavopiridol (Supplemental Fig. 9D). Act D time course analysis also revealed that CPSF73 elimination prevented turnover of 3′ flanking region RNA, arguing that PAS cleavage is necessary to promote their degradation (Supplemental Fig. 9E). These data argue that the differential effect of Xrn2 and CPSF73 on transcriptional termination is unlikely to be due to an incomplete block of 5′→3′ degradation when Xrn2 is manipulated. As such, they support the existence of additional termination mechanisms that occur in the absence of 5′→3′ degradation.

A CPSF73 active site mutant cannot support efficient transcriptional termination

A primary termination pathway involving Xrn2 predicts a requirement for PAS cleavage. To test whether active CPSF73 is required for termination, we generated plasmids containing either wild-type *CPSF73* or a point-mutated derivative (H73A) shown previously to have diminished nuclease activity (Kolev et al. 2008). The plasmid system was used because repeated attempts to stably integrate H73A into *CPSF73-DHFR* cells failed, potentially because of its deleterious effect. Plasmids also incorporated puromycin selection markers to enrich for transfected cells. Western blotting confirmed similar expression of wild-type and H73A proteins in *CPSF73-DHFR* cells and the expected absence of endogenous-sized CPSF73 in empty vector transfected samples (Fig. 7A).

To test the ability of H73A to function in termination, *CPSF73-DHFR* cells were transfected with empty vector, wild type, or H73A. Transfected cells were then enriched for by puromycin selection before removal (or not) of CPSF-DHFR via 10 h of TMP withdrawal. Chromatin-associated RNA was then isolated to study termination via the extent of nascent RNA transcription, which was assayed by qRT–PCR for *MYC* and *ACTB* genes (Fig. 7B–D). In empty vector transfected cells, TMP withdrawal induced the expected accumulation of UCPA RNA and a strong enhancement of readthrough transcripts extending beyond the PAS. These readthrough defects were substantially suppressed in the absence of TMP by wild-type CPSF73.

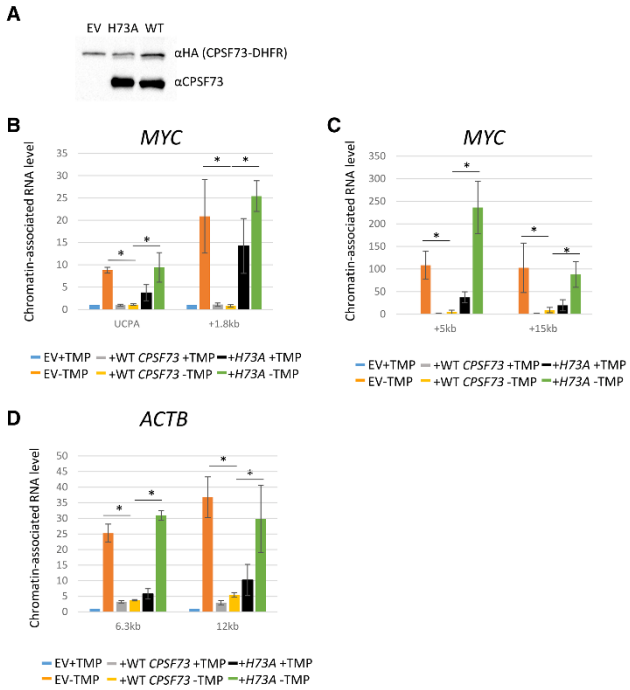


Figure 7. (*A*) Western blotting of *CPSF73-DHFR* cells transfected with H73A CPSF73, wild-type (WT) CPSF73, or empty vector (EV) and probed with anti-HA (to detect CPSF73-DHFR) or anti-CPSF73 (to additionally detect protein derived from transfected constructs). (*B*) qRT–PCR analysis of chromatin-associated RNA isolated from *CPSF-DHFR* cells transfected with empty vector, wild-type, or H73A DHFR ± TMP. Primers were used to detect UCPA Myc RNA or RNA from +1.8 kb beyond the PAS. Values are expressed relative to those in empty vector samples in the presence of TMP after normalizing to unspliced RNA levels. Asterisks display $P < 0.05$ for comparison of the ability or inability of wild-type or H73A CPSF73 to restore termination in relation to the situation lacking CPSF73-DHFR. (*C*) As in *B* but showing signals for +5 and +15 kb beyond the *MYC* PAS. (*D*) As in *B* but detecting RNA from positions +6.3 or +12 kb beyond the *ACTB* PAS. All error bars show standard deviation from at least three independent experiments.

However, H73A expression caused accumulation of readthrough RNA even in the presence of CPSF73-DHFR. This dominant-negative effect indicates that H73A successfully competes with CPSF73-DHFR from PAS cleavage complexes, causing impaired termination. This was confirmed by withdrawal of TMP that showed H73A to be incapable of restoring PAS cleavage or readthrough RNA levels to normal levels. These data strongly suggest that active CPSF73 is required for efficient termination.

Discussion

Our study reveals a clear role for CPSF73 activity and 5′ → 3′ degradation in efficient termination on protein-coding genes as envisioned by the torpedo model. They are most consistent with a primary mechanism in which PAS site cleavage precedes cotranscriptional degradation of Pol II-associated RNA by Xrn2. However, we also observed some termination in situations where 5′ → 3′ degradation of RNA was blocked, arguing for alternative secondary mechanisms. In particular, ablation of CPSF73 activity caused more readthrough than seen on loss of Xrn2, suggesting additional roles for CPSF73 in termination. The observation that miRNA cleavage is capable of promoting Xrn2-dependent termination argues that RNA cleavage may more widely underpin the process beyond protein-coding genes.

Previous reports have reached different conclusions on the role of Xrn2 in termination. Originally, RNAi of Xrn2 caused a termination defect on transfected β-globin plasmids, while a subsequent global analysis found no genome-wide function for Xrn2 in termination at gene 3′ ends using mNET-seq (West et al. 2004; Nojima et al. 2015). An explanation for this came through observations that RNAi of Xrn2 caused termination defects when catalytically inactive Xrn2 was also expressed (Fong et al. 2015). Our results support the view that trace levels of active Xrn2 can provide false negative results in RNAi experiments because Xrn2-AID is virtually eliminated in our system, with its levels likely falling below a critical threshold. Moreover, although Xrn2-AID protein is at substantially reduced levels compared with native Xrn2, this is still sufficient to promote termination, suggesting that a fraction of normal levels supports this function. Finally, expression of inactive Xrn2 in *XRN2-AID* cells has a dominant-negative effect on termination in our system (Supplemental Fig. 4E). These observations may be of importance beyond Xrn2, as they suggest that a degron-based approach can yield a fuller repertoire of functions for some proteins than RNAi alone.

Another finding in our study is that termination is not readily observed in the absence of CPSF73 activity, suggesting that PAS cleavage is required. In vitro experiments suggest that PAS cleavage is not absolutely required for termination (Zhang et al. 2015). However, additional cellular factors may be absent from in vitro systems. Moreover, RNA degradation improved termination in that system, consistent with our finding on the importance of Xrn2 in cells. We do note that H73A CPSF73 has been

shown to immunoprecipitate other CPSF components slightly less efficiently than wild-type CPSF73 (Kolev et al. 2008). This means that the presence of incomplete or unstable CPSF complexes might account for the inability of the H73A mutant to promote termination. If this is true, then it would identify CPSF assembly or activation as providing a crucial function in the process rather than PAS cleavage itself. This would still be an important observation, but we favor PAS cleavage as important for several reasons. First, H73A proved an effective dominant-negative inhibitor of PAS cleavage. Second, partial defects in complex formation might be expected to result in partial termination defects instead of the very strong effect caused by exclusive H73A expression. Moreover, recent results show that polyadenylation factors, exemplified by CstF64, assemble on inactive intronic PASs, but this is insufficient to cause termination unless cleavage is activated by U1 snRNA inhibition (Oh et al. 2017). Finally, the widespread requirement for Xrn2 in efficient termination is most readily explained by PAS cleavage preceding its action.

We also suggest that CPSF73 is required for termination even in the absence of Xrn2. The evidence for this conclusion is that the termination defect is larger upon loss of CPSF73 than when Xrn2 is absent. This could be due to allosteric effects induced by CPSF assembly or activity. Alternatively, such termination could be via the RNA:DNA helicase activity of Senataxin (Skourti-Stathaki et al. 2011), given that its budding yeast homolog, Sen1, can terminate polymerase in purified systems (Porrua and Libri 2013). The exosome may also terminate Pol II by degrading RNA that protrudes from the front of backtracked polymerase (Lemay et al. 2014). Our data argue that these possibilities, including an allosteric mechanism, would require PAS cleavage, given the inability of inactive CPSF73 to support termination. A termination mechanism underpinned by cleavage may also apply following miRNA cleavage. We show an Xrn2 effect on this process; however, the readthrough caused is less than previously observed when miRNA cleavage was prevented by Drosha depletion (Dhir et al. 2015). Drosha depletion caused *MIR17HG* transcription to enter the downstream *GPC5* gene, whereas transcription terminates before this point following Xrn2 loss (Fig. 3E).

While it is difficult to interrogate some molecular details of termination in cells, important principles are consolidated here. In particular, we provide strong evidence that PAS cleavage and cotranscriptional degradation of Pol II-associated RNA are key components of the most efficient termination mechanism. Our results align with predictions of the torpedo model made using highly purified in vitro systems, where it was shown that Xrn2-, Rat1-, and Xrn1-mediated RNA degradation terminates Pol II (Park et al. 2015). In those cases, termination improved when Pol II-associated RNA was longer or when Pol II progression was prevented by nucleotide misincorporation, suggesting that nuclease momentum or polymerase stalling may facilitate the process in cells. Polymerase backtracking over termination regions was inferred from transient transcriptome sequencing (TT-seq) (Schwalb et al. 2016). Moreover, our mNET-seq shows signal accumulation over termination regions that may result from pausing or backtracking. As this signal is often enhanced by loss of Xrn2 (denoted by the blue arrows in Fig. 3), polymerases prone at these sites may be more vulnerable to termination by Xrn2. As we also observed a signal beyond termination sites upon loss of Xrn2, it will be interesting to establish whether this represents polymerases that resume transcription following pausing or those normally terminated by a pause-independent process. In sum, our results provide important details on the termination mechanism in human cells, especially regarding CPSF73 and Xrn2 activities. Our AID system provides a rationale for why RNAi of Xrn2 led to controversy over its role in the process, and our DHFR approach gives strong evidence that PAS cleavage precedes termination.

Materials and methods

Plasmids, primers, and DNA sequences

Primer sequences used for ChIP and qRT–PCR, sequences of repair templates, homology arms, and guide RNA target sites are provided in the Supplemental Material.

Antibodies

The antibodies used were Pol II (CMA601; MBL Technologies), CPSF73 (Abcam, ab72295), CPSF73 for Figure 5A (Bethyl Laboratories, A301-090A), Flag (Sigma, F3165), HA (Roche, 3F10), Xrn2 (Bethyl Laboratories, A301-101), SF3b155 (Abcam, ab39578), Myc (Sigma, 9E10), Pcf11 (Bethyl Laboratories, A303-705 and A303-706), and NS1A (gift from Aldolfo Garcia-Sastre).

Cell culture

HCT116 cells were maintained in DMEM with 10% fetal calf serum. Transfections were with JetPrime (polyplus). For CRISPR, 1 µg of guide RNA plasmid and 1 µg of each repair plasmid were transfected into six-well dishes. Twenty-four hours later, culture medium was changed, and, a further 24 h later, cells were split into a 100-mm dish containing 800 µg/mL neomycin and 150 µg/mL hygromycin. After ~10 d of selection, single colonies were transferred to a 24-well plate and screened by PCR or Western blotting. The presence of repair cassettes at *XRN2* or *CPSF73* was confirmed by Sanger sequencing. An optimized sleeping beauty transposon system (Kowarz et al. 2015) was used to generate Tir1-expressing parental cell lines and cells in which Xrn2 derivatives were stably transfected. A 24-well dish was transfected with 300 ng of sleeping beauty plasmid (derived from pSBbi-puro/pSBbi-blast)) and 100 ng of pCMV(CAT) T7-SB100. Twenty-four hours later, cells were put under selection with 1 µg/mL puromycin or 20 µg/mL blasticidin. For Tir1-expressing cells, single colonies were isolated; for Xrn2 rescue experiments, the entire population was studied. Auxin (Sigma) was added to 500 nM for 60 min unless stated otherwise. TMP (Sigma) was maintained at 20 µM, and, for depletion, cells were grown in medium lacking TMP for 10 h unless stated otherwise. Act D and flavopiridol were used at 5 µg/mL and 1 µM, respectively.

qRT–PCR

Tri reagent (Sigma) was used to isolate total RNA following the manufacturers' guidelines, and RNA was treated with Turbo DNase (Life Technologies) for 1 h. In all cases, reverse transcription of 1 µg of RNA was primed with random hexamers using Protoscript II (New England Biolabs). qPCR was performed using Brilliant III (Agilent Technologies) in a Qiagen Rotorgene instrument. Comparative quantitation was used to establish fold effects.

ChIP and RNA immunoprecipitation

For ChIP, one 100-mm dish of cells was cross-linked for 10 min in 0.5% formaldehyde, and cross-links were quenched in 125 mM glycine for 5 min. Cells were collected (500*g* for 5 min) and resuspended in 400 µL of RIPA buffer (150 mM NaCl, 1% NP40, 0.5% sodium deoxycholate, 0.1% SDS, 50 mM Tris-HCl at pH 8, 5 mM EDTA at pH 8). Samples were sonicated in a Bioruptor sonicator (30 sec on and 30 sec off) 10 times on high setting. Tubes were spun at 13,000 rpm for 10 min. Supernatant was then split into two and added to 30 µL of Dynabeads (Life Technologies) that had been incubated for 2 h with 3 µg of antibody or, as a control, mock-treated. Ten percent of the supernatant was kept for input. Immunoprecipitation was for 2–14 h at 4°C, and beads were then washed twice in RIPA, three times in high-salt wash buffer (500 mM NaCl, 1% NP40, 1% sodium deoxycholate, 100 mM Tris-HCl at pH 8.5), and once in RIPA. Samples were eluted (0.1 M $NaHCO_3$ + 1% SDS), and cross-links were reversed overnight at 65°C. DNA was purified by phenol chloroform extraction and ethanol precipitation. Samples were generally resuspended in 100 µL of water, with 1 µL used per PCR reaction. For RNA immunoprecipitation, cross-links were reversed for 45 min at 65°C. RNA was purified by phenol chloroform extraction and ethanol precipitation followed by DNase treatment and reverse transcription.

Chromatin RNA isolation

Nuclei were isolated from cells by resuspending cell pellets from a 100-mm dish in hypotonic lysis buffer (HLB; 10 mM Tris at pH 7.5, 10 mM NaCl, 2.5 mM $MgCl_2$, 0.5% NP40). This was underlayered with HLB + 10% sucrose and spun at 500*g* for 5 min. Nuclei were resuspended in 100 µL of NUN1 (20 mM Tris-HCl at pH 7.9, 75 mM NaCl, 0.5 mM EDTA, 50% glycerol, 0.85 mM DTT). One milliliter of NUN2 (20 mM HEPES at pH 7.6, 1 mM DTT, 7.5 mM $MgCl_2$, 0.2 mM EDTA. 0.3 M NaCl, 1 M urea, 1% NP40) was added before incubation for 15 min on ice with regular vortexing. Chromatin pellets were isolated by centrifugation at 13,000 rpm in a benchtop centrifuge, and RNA was isolated using Trizol.

4sUTP NRO

Nuclei were isolated as for chromatin-associated RNA. These were resuspended in an equal volume of 2× transcription buffer (40 mM Tris-HCl at pH 7.9, 300 mM KCl, 10 mM $MgCl_2$, 40% glycerol). This was supplemented with rA, C, and G together with 4sUTP (final concentration ~0.1 mM). Following incubation for 15 min at 30°C, RNA was isolated, and biotin linkage and capture were performed as described in Duffy et al. (2015) with some modification. RNA (15–20 µg) was biotinylated in a volume of 250 µL containing 10 mM HEPES (pH 7.5) and 5 µg of MTSEA Biotin-XX (Iris Biotech) dissolved in dimethyl formamide. After incubation in the dark for 90 min, biotinylated RNA was phenol chloroform-extracted and ethanol-precipitated. This was resuspended in RPB (300 mM NaCl, 10 mM Tris at pH 7.5, 5 mM EDTA) and incubated with 150 µL of streptavidin-coated paramagnetic particles (Promega) for 15 min. Beads were washed five times in 100 mM Tris-HCl (pH 7.4), 10 mM EDTA, 1 M NaCl, and 0.1% Tween-20 preheated to 60°C. RNA was eluted in 100 µL of 0.1 M DTT for 15 min at 37°C before final phenol chloroform extraction and ethanol precipitation.

Nuclear RNA-seq

Following 1 h of auxin or mock treatment, nuclei were isolated as for chromatin-associated RNA. Nuclear RNA was extracted using Trizol reagent. rRNA was removed using Ribo-Zero Gold rRNA removal kit (Illumina) according to the user manual. Libraries were prepared using TruSeq stranded total RNA library preparation kit (Illumina) according to the manual and purified using Ampure XP beads (Beckman Coulter). Libraries were screened for fragment size and concentration by Tapestation D1000 (Agilent) and sequenced using HiSeq 2500 (Illumina).

Raw single-end 50-base-pair (bp) reads were screened for sequencing quality using FASTQC, adapter sequences were removed using Trim Galore (wrapper for Cutadapt), and trimmed reads <20 bp were discarded. Reads were aligned to the GRCh38 human genome using Hisat2 (Kim et al. 2015) with splice site annotation from Ensembl. Unmapped and low MAPQ reads were discarded. For metagene analyses, expression levels were calculated for each gene, and genes with low or no expression were removed. A transcriptional window was then applied (TSS – 3 kb and TES + 7 kb). Genes with overlapping reads in this window were discarded (Quinlan and Hall 2010). Metagene profiles were generated using the deeptools suite (Ramirez et al. 2016), with further graphical processing performed in the R environment (http://www.R-project.org). Normalized gene coverage plots were visualized using the Integrated Genome Viewer suite (Robinson et al. 2011).

mNET-seq

A detailed description of the mNET-seq protocol can be found in the study by Nojima et al. (2016). *XRN2-AID* cells were treated for 2 h with auxin or left untreated. NEBNext small RNA libraries were sequenced using HiSeq 2500 (Illumina). Raw 50-bp paired-end sequences had adapter sequences removed using Trim Galore, and resultant reads with a quality of <20 and fragment size of <19 bp were discarded. Reads were aligned using HiSat2 against GRCh38 (Ensemble) with known splice site annotation (Gencode), and concordantly mapped read pairs were selected (Kim et al. 2015).

The mNET-seq traces used single-nucleotide resolution BAM files corresponding to the 3′ end of the RNA fragment (Nojima et al. 2015). For metagene profiles, gene expression was determined by converting raw read counts into transcripts per million (TPM) for each annotated gene (Li and Dewey 2011; Wagner et al. 2012; Liao et al. 2014). For protein-coding metaplots, genes were selected where no other expressed annotated gene overlapped the exclusion range (TES – 1250 bp to TES + 15,250 bp). For each nucleotide across the region, fragments were counted in a 5-bp sliding window and converted to TPM. The normalized metagene profiles represent the average nascent RNA fragment density against relative position from the TES. mNET-seq and RNA-seq data have been deposited with Gene Expression Omnibus (accession no. GSE109003).

Acknowledgments

We thank Karen Moore and Exeter University Sequencing Service for help with RNA-seq and mNET-seq. We thank members of our

laboratory for reading the manuscript. Ervin Fodor is thanked for supporting D.L.V.B. through a program grant from the Medical Research Council. Oksana Gonchar is thanked for assistance with Western blotting. This work was supported by a Wellcome Trust Investigator Award (WT107791/Z/15/Z) and a Lister Institute Research Fellowship held by S.W.

References

Baejen C, Andreani J, Torkler P, Battaglia S, Schwalb B, Lidschreiber M, Maier KC, Boltendahl A, Rus P, Esslinger S, et al. 2017. Genome-wide analysis of RNA polymerase II termination at protein-coding genes. *Mol Cell* **66:** 38–49 e36.

Baillat D, Hakimi MA, Naar AM, Shilatifard A, Cooch N, Shiekhattar R. 2005. Integrator, a multiprotein mediator of small nuclear RNA processing, associates with the C-terminal repeat of RNA polymerase II. *Cell* **123:** 265–276.

Chan SL, Huppertz I, Yao C, Weng L, Moresco JJ, Yates JR III, Ule J, Manley JL, Shi Y. 2014. CPSF30 and Wdr33 directly bind to AAUAAA in mammalian mRNA 3′ processing. *Genes Dev* **28:** 2370–2380.

Connelly S, Manley JL. 1988. A functional mRNA polyadenylation signal is required for transcription termination by RNA polymerase II. *Genes Dev* **2:** 440–452.

Davidson L, Muniz L, West S. 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:** 342–356.

Dhir A, Dhir S, Proudfoot NJ, Jopling CL. 2015. Microprocessor mediates transcriptional termination of long noncoding RNA transcripts hosting microRNAs. *Nat Struct Mol Biol* **22:** 319–327.

Dominski Z, Yang XC, Marzluff WF. 2005. The polyadenylation factor CPSF-73 is involved in histone–pre-mRNA processing. *Cell* **123:** 37–48.

Duffy EE, Rutenberg-Schoenberg M, Stark CD, Kitchen RR, Gerstein MB, Simon MD. 2015. Tracking distinct RNA populations using efficient and reversible covalent chemistry. *Mol Cell* **59:** 858–866.

Fong N, Brannan K, Erickson B, Kim H, Cortazar MA, Sheridan RM, Nguyen T, Karp S, Bentley DL. 2015. Effects of transcription elongation rate and Xrn2 exonuclease activity on RNA polymerase II termination suggest widespread kinetic competition. *Mol Cell* **60:** 256–267.

Iwamoto M, Bjorklund T, Lundberg C, Kirik D, Wandless TJ. 2010. A general chemical method to regulate protein stability in the mammalian central nervous system. *Chem Biol* **17:** 981–988.

Kim M, Krogan NJ, Vasiljeva L, Rando OJ, Nedea E, Greenblatt JF, Buratowski S. 2004. The yeast Rat1 exonuclease promotes transcription termination by RNA polymerase II. *Nature* **432:** 517–522.

Kim JH, Lee SR, Li LH, Park HJ, Park JH, Lee KY, Kim MK, Shin BA, Choi SY. 2011. High cleavage efficiency of a 2A peptide derived from porcine teschovirus-1 in human cell lines, zebrafish and mice. *PLoS One* **6:** e18556.

Kim D, Langmead B, Salzberg SL. 2015. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods* **12:** 357–360.

Kolev NG, Yario TA, Benson E, Steitz JA. 2008. Conserved motifs in both CPSF73 and CPSF100 are required to assemble the active endonuclease for histone mRNA 3′-end maturation. *EMBO Rep* **9:** 1013–1018.

Kowarz E, Loscher D, Marschalek R. 2015. Optimized sleeping beauty transposons rapidly generate stable transgenic cell lines. *Biotechnol J* **10:** 647–653.

Kubota T, Nishimura K, Kanemaki MT, Donaldson AD. 2013. The Elg1 replication factor C-like complex functions in PCNA unloading during DNA replication. *Mol Cell* **50:** 273–280.

Lemay JF, Larochelle M, Marguerat S, Atkinson S, Bahler J, Bachand F. 2014. The RNA exosome promotes transcription termination of backtracked RNA polymerase II. *Nat Struct Mol Biol* **21:** 919–926.

Li B, Dewey CN. 2011. RSEM: accurate transcript quantification from RNA-seq data with or without a reference genome. *BMC Bioinformatics* **12:** 323.

Liao Y, Smyth GK, Shi W. 2014. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30:** 923–930.

Libri D. 2015. Endless quarrels at the end of genes. *Mol Cell* **60:** 192–194.

Luo W, Johnson AW, Bentley DL. 2006. The role of Rat1 in coupling mRNA 3′-end processing to transcription termination: implications for a unified allosteric-torpedo model. *Genes Dev* **20:** 954–965.

Mandel CR, Kaneko S, Zhang H, Gebauer D, Vethantham V, Manley JL, Tong L. 2006. Polyadenylation factor CPSF-73 is the pre-mRNA 3′-end-processing endonuclease. *Nature* **444:** 953–956.

Mapendano CK, Lykke-Andersen S, Kjems J, Bertrand E, Jensen TH. 2010. Crosstalk between mRNA 3′ end processing and transcription initiation. *Mol Cell* **40:** 410–422.

Miki TS, Carl SH, Grosshans H. 2017. Two distinct transcription termination modes dictated by promoters. *Genes Dev* **31:** 1870–1879.

Natsume T, Kiyomitsu T, Saga Y, Kanemaki MT. 2016. Rapid protein depletion in human cells by auxin-inducible degron tagging with short homology donors. *Cell Rep* **15:** 210–218.

Nemeroff ME, Barabino SM, Li Y, Keller W, Krug RM. 1998. Influenza virus NS1 protein interacts with the cellular 30 kDa subunit of CPSF and inhibits 3′end formation of cellular pre-mRNAs. *Mol Cell* **1:** 991–1000.

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

Nojima T, Gomes T, Grosso ARF, Kimura H, Dye MJ, Dhir S, Carmo-Fonseca M, Proudfoot NJ. 2015. Mammalian NET-Seq reveals genome-wide nascent transcription coupled to RNA processing. *Cell* **161:** 526–540.

Nojima T, Gomes T, Carmo-Fonseca M, Proudfoot NJ. 2016. Mammalian NET-seq analysis defines nascent RNA profiles and associated RNA processing genome-wide. *Nat Protoc* **11:** 413–428.

Oh JM, Di C, Venters CC, Guo J, Arai C, So BR, Pinto AM, Zhang Z, Wan L, Younis I, et al. 2017. U1 snRNP telescripting regulates a size-function-stratified human genome. *Nat Struct Mol Biol* **24:** 993–999.

Park J, Kang M, Kim M. 2015. Unraveling the mechanistic features of RNA polymerase II termination by the 5′-3′ exoribonuclease Rat1. *Nucleic Acids Res* **43:** 2625–2637.

Porrua O, Libri D. 2013. A bacterial-like mechanism for transcription termination by the Sen1p helicase in budding yeast. *Nat Struct Mol Biol* **20:** 884–891.

Porrua O, Libri D. 2015. Transcription termination and the control of the transcriptome: why, where and how to stop. *Nat Rev Mol Cell Biol* **16:** 190–202.

Proudfoot NJ. 1989. How RNA polymerase II terminates transcription in higher eukaryotes. *Trends Biochem Sci* **14:** 105–110.

Proudfoot NJ. 2012. Ending the message: poly(A) signals then and now. *Genes Dev* **25:** 1770–1782.

Proudfoot NJ. 2016. Transcriptional termination in mammals: stopping the RNA polymerase II juggernaut. *Science* **352:** aad9926.

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26:** 841–842.

Ramirez F, Ryan DP, Gruning B, Bhardwaj V, Kilpert F, Richter AS, Heyne S, Dundar F, Manke T. 2016. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res* **44:** W160–W165.

Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. 2011. Integrative genomics viewer. *Nat Biotechnol* **29:** 24–26.

Schonemann L, Kuhn U, Martin G, Schafer P, Gruber AR, Keller W, Zavolan M, Wahle E. 2014. Reconstitution of CPSF active in polyadenylation: recognition of the polyadenylation signal by WDR33. *Genes Dev* **28:** 2381–2393.

Schwalb B, Michel M, Zacher B, Fruhauf K, Demel C, Tresch A, Gagneur J, Cramer P. 2016. TT-seq maps the human transient transcriptome. *Science* **352:** 1225–1228.

Sheridan RM, Bentley DL. 2016. Selectable one-step PCR-mediated integration of a degron for rapid depletion of endogenous human proteins. *Biotechniques* **60:** 69–74.

Shi Y, Di Giammartino DC, Taylor D, Sarkeshik A, Rice WJ, Yates JR III, Frank J, Manley JL. 2009. Molecular architecture of the human pre-mRNA 3′ processing complex. *Mol Cell* **33:** 365–376.

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

Wagner GP, Kin K, Lynch VJ. 2012. Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. *Theory Biosci* **131:** 281–285.

West S, Gromak N, Proudfoot NJ. 2004. Human 5′ → 3′ exonuclease Xrn2 promotes transcription termination at co-transcriptional cleavage sites. *Nature* **432:** 522–525.

Whitelaw E, Proudfoot N. 1986. α-Thalassaemia caused by a poly (A) site mutation reveals that transcriptional termination is linked to 3′ end processing in the human α2 globin gene. *EMBO J* **5:** 2915–2922.

Zhang H, Rigo F, Martinson HG. 2015. 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:** 437–448.

ii. An end in sight? Xrn2 and transcriptional termination by RNA polymerase II.

Eaton, J.D., and West, S. (2018). An end in sight? Xrn2 and transcriptional termination by RNA polymerase II. Transcription 9, 321–326.

TRANSCRIPTION
2018, VOL. 9, NO. 5, 321–326
https://doi.org/10.1080/21541264.2018.1498708

Taylor & Francis
Taylor & Francis Group

POINT-OF-VIEW

OPEN ACCESS Check for updates

An end in sight? Xrn2 and transcriptional termination by RNA polymerase II

Joshua D. Eaton and Steven West

The Living Systems Institute, University of Exeter, Exeter, UK

ABSTRACT

Every transcription cycle ends in termination when RNA polymerase dissociates from the DNA. Although conceptually simple, the mechanism has proven somewhat elusive in eukaryotic systems. Gene-editing and high resolution polymerase mapping now offer clarification of important steps preceding transcriptional termination by RNA polymerase II in human cells.

ARTICLE HISTORY
Received 24 May 2018
Revised 26 June 2018
Accepted 27 June 2018

KEYWORDS
Xrn2; RNA polymerase II (Pol II); transcriptional termination; 3′ end processing; cleavage and polyadenylation; RNA degradation

Overview

The most important sequence element for transcriptional termination on protein-coding genes is the polyadenylation signal (PAS), which consists of an AAUAAA hexamer (or a variant thereof) followed by a U or GU rich element [1,2]. The former is bound by the cleavage and polyadenylation specificity factor (CPSF) complex and the latter by cleavage stimulation factors (CstF). The RNA is then cut by CPSF73 [3], whereupon the coding portion is polyadenylated and the 3′ product rapidly degraded. The central role of the PAS in termination provides the premise for two potential mechanisms. One is referred to as the allosteric model and proposes that a PAS-dependent conformational change elicits termination [4]. This might be in RNA polymerase II (Pol II) itself or via an exchange of its associated factors. Another is that RNA cleavage promotes termination by generating a Pol II-associated RNA, which is degraded by a 5′-3′ exonuclease [2,5]. This torpedo model posits that the degrading exonuclease chases down Pol II and somehow signals termination.

The role of a molecular torpedo in the termination process has been controversial. RNA interference (RNAi) of the human 5′-3′ exonuclease, Xrn2, or mutation of its budding yeast equivalent, Rat1, was originally shown to disrupt termination on transfected plasmids and endogenous genes respectively [6,7]. Although the role of Rat1 was recently re-affirmed [8], RNAi of Xrn2 did not reveal a general termination defect at the 3′ end of protein-coding genes [9]. However, termination was subsequently shown to be affected when a catalytically inactive version of Xrn2 is expressed in an RNAi background [10]. The inactive protein may block termination by trace amounts of Xrn2, remaining after RNAi, or prevent redundant 5′-3′ exonucleases from accessing the nascent RNA in the Xrn2-depleted situation [11]. The role of RNA cleavage (a pre-requisite for Xrn2 function) in termination has also been debated. Electron micrographs of Pol II-transcribed genes show few examples of elongating polymerases associated with cleaved RNA [12]. Moreover, PAS cleavage has been argued as dispensable for termination in purified systems [13].

We recently addressed the role of RNA cleavage and degradation in termination by employing gene editing to bring Xrn2 or CPSF73 under inducible control [14]. Xrn2 was tagged with an auxin-inducible degron (AID) and could be eliminated within minutes whereas CPSF73 was combined with an *E. coli* DHFR degron allowing its depletion over the course of several hours. In both cases, this is much faster than the proteins can be depleted by RNAi and it revealed an unambiguous and general termination defect upon Xrn2 elimination. Loss of CPSF73 caused a strong inhibition of PAS cleavage

CONTACT Steven West s.west@exeter.ac.uk

and very extensive read-through transcription. This could not be restored by expressing its inactive derivative confirming its identity as the PAS nuclease and the importance of RNA cleavage to termination. The difference between our results and those obtained using RNAi likely reflect the speed and magnitude of depletion possible with inducible degrons. Our findings support a principal mode of termination in which PAS cleavage precedes degradation by Xrn2 and is summarised in Figure 1. These data clarify salient features of protein-coding gene termination but raise a number of interesting points for discussion.

Gene specific effects of Xrn2 loss on termination

Mammalian native elongating transcript sequencing (mNET-seq) was used to establish the effects of Xrn2 loss on transcription [9,14]. This method provides a picture of Pol II location over the genome at single nucleotide resolution and, unlike run-on based methods, detects polymerases irrespective of their ability to transcribe. These features can be used to identify specific positions where Pol II accumulates under a variety of conditions and very precisely characterise perturbations caused by the loss of factors. We found Xrn2 depletion to induce a termination defect at the majority of protein-coding genes, but observed differences in the appearance and magnitude of effect.

In some cases, Xrn2 elimination caused enhanced signal over regions where there was already evidence of accumulated Pol II in normal cells. This may be due to Xrn2 terminating polymerases that are paused or arrested and is consistent with features that enhance Xrn2-dependent termination *in vitro* [15]. Moreover, transient transcriptome sequencing (TT-seq) identified sites of Pol II termination genome-wide and these correspond to sequences predicted to promote polymerase pausing [16]. Xrn2 loss also often results in Pol II signal beyond where transcription terminates in its presence. This may be where Pol II pausing is infrequent or because weaker pausing has been overcome due to low levels of Xrn2. As such, Xrn2 may pursue a mixture of stalled and transcribing Pol II.

While the majority of protein-coding genes show read-through upon Xrn2 loss there are cases where this is much more extensive – extending tens of kilobases further than normal. Perhaps these are examples of where Pol II does not pause and responds more acutely to Xrn2 loss or these genes may not employ alternative termination mechanisms that could act in the absence of Xrn2 (see below). Other possibilities would include slower 3′ end processing that would delay the production of an Xrn2 entry site or rapid Pol II elongation rates previously shown to extend transcription [10]. A more detailed analysis of primary sequence or chromatin features will be important to understand the basis of these different Xrn2 effects and whether they are variations on the same termination mechanism or represent distinct processes.

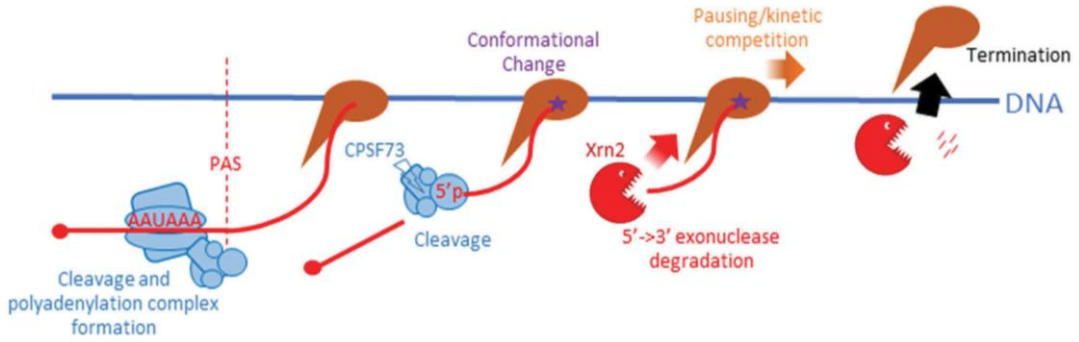


Figure 1. Model for the principle mechanism of Pol II termination in humans. Pol II (brown) transcribes the PAS (AAUAAA in RNA) which is bound by polyadenylation factors (blue shapes) with the RNA cleaved by CPSF73. This process likely induces a change in the elongation complex (star) rendering Pol II more susceptible to termination, which occurs through degradation of the Pol II associated product of PAS cleavage by Xrn2 (red). Xrn2-dependent termination may be augmented by Pol II pausing or arrest downstream of the PAS.

RNA cleavage and Xrn2-independent termination?

It is important to note that, on most genes, mNET-seq and RNA-seq signals eventually drop to near-background levels even when Xrn2 is absent [14]. Analysis of individual genes suggests that this is still the case when inactive Xrn2 is expressed in the absence of Xrn2-AID suggesting that delayed termination even occurs when degradation from the 5′ end is blocked. In contrast, read-through is much more extensive upon loss of CPSF73 with analysis of individual genes showing little evidence of delayed termination. These differences could result from incomplete loss of Xrn2 (though the aforementioned result with inactive Xrn2 would argue against this) or a more effective depletion of CPSF73 compared to Xrn2. Also, CPSF73-DHFR takes longer to deplete than Xrn2-AID, which may have unknown secondary consequences.

Another interpretation of these data is that there is Xrn2-independent termination that still requires CPSF73, which may play a more underpinning role in the process through alternative mechanisms. Interestingly, exosome depletion stabilises some read-through RNA in the absence of Xrn2, which may derive from such processes [14]. The existence of auxiliary termination mechanisms would be reminiscent of fail-safe termination in budding yeast where, for example, compromising Rat1 leads to termination via the Nrd1 pathway [17]. Although the human equivalent of Nrd1 is not described, SETX is similar to the Nrd1-associating RNA:DNA helicase, Sen1, and is proposed to assist Xrn2 in termination [18].

More generally, does the function of CPSF73 in termination depend on its cleavage activity? Experiments using *in vitro* systems have been used to show that PAS-dependent termination can occur without RNA cleavage [13], but it is unclear whether this mechanism applies in cells where multiple processes are coordinated. We found that catalytically inactive CPSF73 does not restore transcriptional termination, when CPSF73 is lost, suggesting a strict requirement for cleavage [14]. Alternatively, inactive CPSF73 may not fully recapitulate the function of the native protein, for example in forming stable or precise interactions with other factors. Although formation of CPSF complexes is partially impaired by inactive CPSF73 this does not readily explain its complete inability to supress transcriptional read-through caused by the absence of CPSF73 [14,19]. Therefore, unless inactive CPSF73 disrupts key molecular contacts outside of its cleavage function, these data support an underpinning role of PAS cleavage in termination. It would generate an entry site for Xrn2 but, if the hypothesis that CPSF73 supports termination in the absence of Xrn2 is true, it may also promote termination by other mechanisms.

The role of Xrn2 in termination at other Pol II gene classes

Pol II transcribes many different classes of genes, some of which use endonuclease cleavage for 3′ end formation. One example are so-called lnc-pri-miRNA genes that utilise microprocessor cleavage at their 3′ end [20]. Importantly, Xrn2 degrades the 3′ product of this cleavage and loss of Xrn2 causes a termination defect on lnc-pri-miRNA genes suggesting that Xrn2-dependent termination is not restricted to PAS-containing genes [14].

Other transcript classes that undergo 3′ end cleavage include those coding for snRNAs and replication-dependent histones (RDH). RDH genes employ a complex similar to that used for other protein-coding genes with CPSF73 as the endonuclease [21]. The 3′ ends of snRNAs are also formed by endonuclease cleavage, which is performed by the Ints11 subunit of the integrator complex [22]. Interestingly, Xrn2 elimination does not affect read-through at either of these gene classes. Additionally, promoter upstream transcripts (PROMPTs) are largely unaffected by Xrn2 (our unpublished findings). This is despite observations that PROMPTs are relatively rich in PAS's with evidence that these are sometimes functional [23]. Thus, RNA cleavage does not automatically trigger Xrn2-dependent RNA degradation and termination.

These unaffected transcript classes generally derive from shorter genes which may preclude the recruitment of important protein co-factors or dictate a Pol II C-terminal domain (CTD)

modification status that promotes termination differently. For example, the choice of termination pathway in budding yeast is influenced at least in part by the relative densities of Serine 5 and 2 phosphorylation (Ser5p and Ser2p) on the CTD [24]. Ser5p is higher at the beginning of the transcription cycle favouring termination by Nrd1. In contrast, Ser2p is highest at the 3′ end of genes and likely aids Xrn2 recruitment given evidence that Rat1 is part of a complex that is recruited to Pol II by Rtt103 – a factor with some preference towards Ser2p CTD [7].

The short nature of RDH and snRNA genes also makes them susceptible to termination via an Ars2-dependent 3′ end processing pathway that is not active on longer genes (showing some similarity to Nrd1 termination in budding yeast) [25]. Although the details that connect Ars2 with transcript processing are not fully elucidated, this may support different termination modes. Finally, the response of genes to Xrn2 may be specified at the promoter, which is analogous to how different transcript classes, including snRNAs, are only matured when transcription is driven by their own promoter [26]. Indeed, recent work shows the sensitivity of many *C.Elegans* protein-coding genes to Xrn2-dependent termination is determined by promoter identity [27]. Intriguingly, Xrn2 is still recruited to transcription units that are unaffected by its loss, including human RDH and snRNA genes, raising the possibility that an additional feature or factor determines its involvement in termination [10,27].

How does Xrn2 promote termination?

The torpedo model envisages that Xrn2 chases Pol II and then signals termination by a process that is still not understood [2,5]. We have discussed how paused Pol II may constitute a frequent target for Xrn2 and data from purified systems shows that prone polymerases are more effectively terminated by 5′-3′ exonucleases [15]. However, Xrn2 may also have to pursue elongating polymerases as modulating polymerase transcription rates affects the position of termination in cells and extended mNET-seq signal is often observed in its absence [10,14].

RNA degradation may be required to deliver other important termination factors to the polymerase. A relatively small number of proteins are directly capable of inducing Pol II termination and include Xrn2, Pcf11, Sen1 (SETX in humans) and TTF2. Pcf11 causes dissolution of stalled elongation complexes *in vitro* and Sen1 is an RNA:DNA helicase that can also terminate polymerase in purified systems [28,29]. Both of these factors likely require stalled or very slow polymerases to successfully act. This is a feature of Pol II beyond the PAS and may even be promoted by RNA degradation as shortening nascent RNA can impede elongation [16,30]. TTF2 is another strong candidate as it associates with Xrn2 [31]. RNA degradation may also locate Xrn2 in close proximity to Pol II to cause termination more directly by forming specific interactions with Pol II. It is interesting in this regard that Rat1 cannot terminate *E.coli* polymerase whereas Rat and Xrn2 promote Pol II termination in the same highly purified system [15].

In argument of an indirect role for exonucleases in cells, work performed in budding yeast showed that Pol II-associated products of PAS cleavage continued to be degraded in the presence of mutated Rat1, yet the termination defect remained [11]. In this case, degradation was taken over by Xrn1 that localises to the nucleus when Rat1 function is compromised. As Xrn1 can promote termination *in vitro* [15], its inability to do so in cells might be explained by differences in processivity or the absence of interaction partners that enable Xrn2/Rat1 to function in termination. Rat1 may also indirectly affect termination as it is required, in some cases, for the recruitment of 3′ end processing factors to genes, including Pcf11 [11].

In human cells, Xrn2 loss had no obvious impact on Pcf11 recruitment to genes or on PAS cleavage in general [14]. Xrn2 elimination was also sufficient to inhibit co-transcriptional degradation of the 3′ PAS cleavage product and cause a genome-wide termination defect. Thus, any substitution of its degradation function by Xrn1 or other nucleases is probably more limited than in budding yeast. We therefore favour a more exclusive role for Xrn2 in termination and perhaps the biggest question for future study is, still, how and whether this is direct or indirect. Further implementation of inducible protein depletion systems, like AID, will likely shed more light on the termination process by

revealing the contribution(s) of various factors. Ultimately, however, this conundrum will most effectively be tackled using structural biology which has illuminated many of the other phases in the transcription cycle in such amazing detail.

Acknowledgments

We thank members of the lab for reading the manuscript and our funders, The Wellcome Trust and The Lister Institute, for supporting our research. We apologise that not all relevant work could be discussed or directly cited due to space constraints.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the Lister Institute of Preventive Medicine; Wellcome Trust [107791/Z/15/Z].

References

[1] Proudfoot NJ. Ending the message: poly(A) signals then and now. Genes Dev. 2011 Sep 1;25(17):1770–1782. 25/17/1770 [pii]. PubMed PMID: 21896654; PubMed Central PMCID: PMC3175714. eng.

[2] Connelly S, Manley JL. A functional mRNA polyadenylation signal is required for transcription termination by RNA polymerase II. Genes Dev. 1988 Apr;2(4):440–452. PubMed PMID: 2836265.

[3] Mandel CR, Kaneko S, Zhang H, et al. Polyadenylation factor CPSF-73 is the pre-mRNA 3′-end-processing endonuclease. Nature. 2006 Dec 14;444(7121):953–956. nature05363 [pii]. PubMed PMID: 17128255; eng.

[4] Logan J, Falck-Pedersen E, Darnell JE Jr. et al. 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. 1987 Dec;84(23):8306–8310. PubMed PMID: 3479794; PubMed Central PMCID: PMCPMC299531.

[5] Proudfoot NJ. How RNA polymerase II terminates transcription in higher eukaryotes. Trends Biochem Sci. 1989 Mar;14(3):105–110. PubMed PMID: 2658217.

[6] West S, Gromak N, Proudfoot NJ. Human 5′ –> 3′ exonuclease Xrn2 promotes transcription termination at co-transcriptional cleavage sites. Nature. 2004 Nov 25;432(7016):522–525. nature03035 [pii]. PubMed PMID: 15565158; eng.

[7] Kim M, Krogan NJ, Vasiljeva L, et al. The yeast Rat1 exonuclease promotes transcription termination by RNA polymerase II. Nature. 2004 Nov 25;432(7016):517–522. nature03041 [pii]. PubMed PMID: 15565157; eng.

[8] Baejen C, Andreani J, Torkler P, et al. Genome-wide Analysis of RNA Polymerase II Termination at Protein-Coding Genes. Mol Cell. 2017 Apr 6;66(1):38–49 e6. PubMed PMID: 28318822.

[9] Nojima T, Gomes T, Grosso ARF, et al. Mammalian NET-Seq Reveals Genome-wide Nascent Transcription Coupled to RNA Processing. Cell. 2015 Apr 23;161(3):526–540. PubMed PMID: 25910207; PubMed Central PMCID: PMCPMC4410947.

[10] Fong N, Brannan K, Erickson B, et al. Effects of Transcription Elongation Rate and Xrn2 Exonuclease Activity on RNA Polymerase II Termination Suggest Widespread Kinetic Competition. Mol Cell. 2015 Oct 15;60(2):256–267. PubMed PMID: 26474067; PubMed Central PMCID: PMCPMC4654110.

[11] Luo W, Johnson AW, Bentley DL. The role of Rat1 in coupling mRNA 3′-end processing to transcription termination: implications for a unified allosteric-torpedo model. Genes Dev. 2006 Apr 15;20(8):954–965. gad.1409106 [pii]. PubMed PMID: 16598041; PubMed Central PMCID: PMC1472303. eng.

[12] Osheim YN, Sikes ML, Beyer AL. EM visualization of Pol II genes in Drosophila: most genes terminate without prior 3′ end cleavage of nascent transcripts. Chromosoma. 2002 Mar;111(1):1–12. PubMed PMID: 12068918; eng.

[13] Zhang H, Rigo F, Martinson HG. Poly(A) Signal-Dependent Transcription Termination Occurs through a Conformational Change Mechanism that Does Not Require Cleavage at the Poly(A) Site. Mol Cell. 2015 Aug 6;59(3):437–448. PubMed PMID: 26166703.

[14] Eaton JD, Davidson L, Bauer DLV, et al. Xrn2 accelerates termination by RNA polymerase II, which is underpinned by CPSF73 activity. Genes Dev. 2018 Jan 15;32(2):127–139. PubMed PMID: 29432121; PubMed Central PMCID: PMCPMC5830926.

[15] Park J, Kang M, Kim M. Unraveling the mechanistic features of RNA polymerase II termination by the 5′-3′ exoribonuclease Rat1. Nucleic Acids Res. 2015 Mar 11;43(5):2625–2637. PubMed PMID: 25722373; PubMed Central PMCID: PMCPMC4357727.

[16] Schwalb B, Michel M, Zacher B, et al. TT-seq maps the human transient transcriptome. Science. 2016 Jun 3;352(6290):1225–1228. PubMed PMID: 27257258.

[17] Rondon AG, Mischo HE, Kawauchi J, et al. Fail-safe transcriptional termination for protein-coding genes in S. cerevisiae. Mol Cell. 2009 Oct 9;36(1):88–98. PubMed PMID: 19818712; PubMed Central PMCID: PMCPMC2779338.

[18] Skourti-Stathaki K, Proudfoot NJ, Gromak N. Human senataxin resolves RNA/DNA hybrids formed at transcriptional pause sites to promote Xrn2-dependent termination. Mol Cell. 2011 Jun 24;42(6):794–805.

PubMed PMID: 21700224; PubMed Central PMCID: PMCPMC3145960.

[19] Kolev NG, Yario TA, Benson E, et al. Conserved motifs in both CPSF73 and CPSF100 are required to assemble the active endonuclease for histone mRNA 3′-end maturation. EMBO Rep. 2008 Oct;9(10):1013–1018. PubMed PMID: 18688255; PubMed Central PMCID: PMCPMC2572124.

[20] Dhir A, Dhir S, Proudfoot NJ, et al. Microprocessor mediates transcriptional termination of long noncoding RNA transcripts hosting microRNAs. Nat Struct Mol Biol. 2015 Apr;22(4):319–327. PubMed PMID: 25730776; PubMed Central PMCID: PMCPMC4492989.

[21] Dominski Z, Yang XC, Marzluff WF. The polyadenylation factor CPSF-73 is involved in histone-pre-mRNA processing. Cell. 2005 Oct 7;123(1):37–48. PubMed PMID: 16213211.

[22] Baillat D, Hakimi MA, Naar AM, et al. Integrator, a multiprotein mediator of small nuclear RNA processing, associates with the C-terminal repeat of RNA polymerase II. Cell. 2005 Oct 21;123(2):265–276. PubMed PMID: 16239144.

[23] Ntini E, Jarvelin AI, Bornholdt J, et al. Polyadenylation site-induced decay of upstream transcripts enforces promoter directionality. Nat Struct Mol Biol. 2013 Aug;20(8):923–928. PubMed PMID: 23851456.

[24] Gudipati RK, Villa T, Boulay J, et al. Phosphorylation of the RNA polymerase II C-terminal domain dictates transcription termination choice. Nat Struct Mol Biol. 2008 Aug;15(8):786–794. PubMed PMID: 18660821.

[25] Hallais M, Pontvianne F, Andersen PR, et al. CBC-ARS2 stimulates 3′-end maturation of multiple RNA families and favors cap-proximal processing. Nat Struct Mol Biol. 2013 Dec;20(12):1358–1366. PubMed PMID: 24270878.

[26] Hernandez N, Weiner AM. Formation of the 3′ end of U1 snRNA requires compatible snRNA promoter elements. Cell. 1986 Oct 24;47(2):249–258. PubMed PMID: 3768956.

[27] Miki TS, Carl SH, Grosshans H. Two distinct transcription termination modes dictated by promoters. Genes Dev. 2017 Sep 15;31(18):1870–1879. PubMed PMID: 29021241.

[28] Han Z, Libri D, Porrua O. Biochemical characterization of the helicase Sen1 provides new insights into the mechanisms of non-coding transcription termination. Nucleic Acids Res. 2017 Feb 17;45(3):1355–1370. PubMed PMID: 28180347; PubMed Central PMCID: PMCPMC5388409.

[29] Zhang Z, Gilmour DS. Pcf11 is a termination factor in Drosophila that dismantles the elongation complex by bridging the CTD of RNA polymerase II to the nascent transcript. Mol Cell. 2006 Jan 6;21(1):65–74. PubMed PMID: 16387654.

[30] Ujvari A, Pal M, Luse DS. RNA polymerase II transcription complexes may become arrested if the nascent RNA is shortened to less than 50 nucleotides. J Biol Chem. 2002 Sep 6;277(36):32527–32537. PubMed PMID: 12087087.

[31] Brannan K, Kim H, Erickson B, et al. mRNA decapping factors and the exonuclease Xrn2 function in widespread premature termination of RNA polymerase II transcription. Mol Cell. 2012 May 11;46(3):311–324. S1097-2765(12)00213-4 [pii]. PubMed PMID: 22483619; eng.

iii. A unified allosteric/torpedo mechanism for transcriptional termination on human protein-coding genes.

Eaton, J.D., Francis, L., Davidson, L., and West, S. (2020). A unified allosteric/torpedo mechanism for transcriptional termination on human protein-coding genes. Genes Dev. 34, 132–145..

A unified allosteric/torpedo mechanism for transcriptional termination on human protein-coding genes

Joshua D. Eaton, Laura Francis, Lee Davidson, and Steven West

The Living Systems Institute, University of Exeter, Exeter EX4 4QD, United Kingdom

The allosteric and torpedo models have been used for 30 yr to explain how transcription terminates on protein-coding genes. The former invokes termination via conformational changes in the transcription complex and the latter proposes that degradation of the downstream product of poly(A) signal (PAS) processing is important. Here, we describe a single mechanism incorporating features of both models. We show that termination is completely abolished by rapid elimination of CPSF73, which causes very extensive transcriptional readthrough genome-wide. This is because CPSF73 functions upstream of modifications to the elongation complex and provides an entry site for the XRN2 torpedo. Rapid depletion of XRN2 enriches these events that we show are underpinned by protein phosphatase 1 (PP1) activity, the inhibition of which extends readthrough in the absence of XRN2. Our results suggest a combined allosteric/torpedo mechanism, in which PP1-dependent slowing down of polymerases over termination regions facilitates their pursuit/capture by XRN2 following PAS processing.

[*Keywords*: XRN2; transcriptional termination; CPSF73; polyadenylation signal; RNA polymerase II; PP1; antisense oligo]

Supplemental material is available for this article.

Received September 20, 2019; revised version accepted November 18, 2019.

Termination by RNA polymerase II (Pol II) completes the transcription cycle. It serves to recycle Pol II for new rounds of initiation and prevents interference with the transcription of neighboring genes. On protein-coding genes, transcriptional termination is mechanistically connected with 3′ end formation, which involves cleavage and polyadenylation (CPA) (Proudfoot 2011). For CPA, a multiprotein complex is recruited to the polyadenylation signal (PAS), which consists of a hexamer (usually AAUAAA) and a downstream GU (or U)-rich sequence. The 73-kDa component of cleavage and polyadenylation specificity factor (CPSF73) is the endonuclease that cleaves between these two sequences (Mandel et al. 2006). A poly(A) tail is added to the stable upstream cleavage product and the downstream RNA is rapidly degraded. Mutation of the PAS hexamer prevents transcriptional termination—a result that led to two models to explain termination that have been debated ever since (Connelly and Manley 1988; Proudfoot 1989).

The allosteric (or antiterminator) model proposes that transcription of the PAS induces a change in the elongation complex that renders it prone to termination (Logan et al. 1987). This might involve the exchange of associated factors or modification of proteins including Pol II as shown by chromatin immunoprecipitation (ChIP) experiments (Ahn et al. 2004; Kim et al. 2004a). There are also examples of antitermination factors in budding yeast (PC4) and humans (SCAF4 and SCAF8) (Calvo and Manley 2001; Gregersen et al. 2019). Conversely, the 3′ end processing factor, PCF11, can release Pol II from DNA in vitro (Zhang and Gilmour 2006). Allostery may constitute protein modification(s) with phosphorylation-based switches underpinning the 3′ end transition in budding and fission yeast (Schreieck et al. 2014; Kecman et al. 2018; Parua et al. 2018). Specifically dephosphorylation of the Pol II C-terminal domain (CTD) and the SPT5 elongation factor occur after the PAS during an elongation to termination transition. Finally, Pol II conformational changes can result from PAS transcription in purified systems (Zhang et al. 2015).

The torpedo model proposes that the Pol II-associated product of PAS cleavage is degraded by a 5′ → 3′ exonuclease leading to termination (Connelly and Manley 1988; Proudfoot 1989). Evidence for this mechanism was first provided in budding yeast and human cells where mutation or depletion of their respective nuclear 5′ → 3′ exonucleases, Rat1 and XRN2, inhibits termination (Kim et al.

Corresponding author: s.west@exeter.ac.uk
Article published online ahead of print. Article and publication date are online at http://www.genesdev.org/cgi/doi/10.1101/gad.332833.119. Freely available online through the *Genes & Development* Open Access option.

2004b; West et al. 2004). Subsequent studies confirmed the generality of these findings in both organisms (Fong et al. 2015; Baejen et al. 2017; Eaton et al. 2018). An important distinction between the torpedo and allosteric mechanisms is that, in principle, only the former requires PAS cleavage. However, an inactivating point mutation in CPSF73 does not support termination arguing that PAS cleavage cannot readily be bypassed in cells (Eaton et al. 2018).

Given the substantial evidence for both models, it is likely that termination actually employs aspects of each. One possibility is that some genes use an allosteric process with others using XRN2. This might be the case in *C. elegans*, where XRN2 depletion causes termination defects on only a subset of genes (Miki et al. 2017). Another scenario is that allosteric and torpedo processes contribute to a common mode of termination as has been proposed in budding yeast where PCF11 and Rat1 are each required for the others recruitment to genes (Luo et al. 2006). Although human PCF11 is recruited to genes irrespective of XRN2 (Eaton et al. 2018), some PAS-dependent processes could aid XRN2 function. For instance, Pol II pausing downstream from a PAS could facilitate its pursuit by XRN2 especially since pausing enhances termination on reporter plasmids (Gromak et al. 2006).

We investigated the transcriptional termination mechanism in human cells using modified cell lines that allow more rapid depletion of CPSF73 or XRN2 than more commonly used systems. CPSF73 depleted in this manner causes profound readthrough suggesting that its function is essential for termination and that there are no significant fail-safe mechanisms in its absence. CPSF73 is required to slow down Pol II beyond the PAS and for phosphorylation of its CTD on threonine 4 (Thr4p). These slowed-down polymerases pile up beyond the PAS, and their presence, is enriched by rapid depletion of XRN2. Strikingly, longer readthrough can be induced in the absence of XRN2 by inhibiting or depleting protein phosphatase 1 (PP1), identifying a mechanistic basis for the pause. We propose that allosteric events facilitate the pursuit of Pol II and its termination by XRN2 constituting a unified allosteric/torpedo mechanism for transcriptional termination on human protein-coding genes.

Results

CPSF73-associated functions are critical for termination on protein-coding genes

We previously found that rapid elimination of XRN2, via an auxin-inducible degron (AID), causes a general termination defect on protein-coding genes (Eaton et al. 2018). In the same study, CPSF73 elimination via an *E. coli*-derived DHFR degron caused more extended readthrough of selected genes, implying a comparatively more important role for its activities in termination. The DHFR degron requires ~10 h for depletion, whereas XRN2-AID could be depleted faster, so we tagged *CPSF73* with an AID to better-compare CPSF73 and XRN2 functions (Fig.1A; Natsume et al. 2016). AID depletion requires the plant TIR1 protein, which was integrated into HCT116 cells (chosen for their diploid karyotype) to allow its doxycycline (dox)-dependent induction. The Western blot in Figure 1B confirms homozygous tagging of *CPSF73* and that full depletion depends on both dox and auxin. As well as using the same tag, depletion can be achieved in 3 h, providing a more direct comparison to the XRN2-AID system. Note that dox treatment alone induces mild CPSF73 depletion, which may be why we were previously unable to combine CPSF73-AID with constitutive TIR1 expression (Eaton et al. 2018).

To broadly assess the impact of rapid CPSF73 depletion on transcription, we performed RNA-sequencing in mock-treated *CPSF73-AID* cells or the same cells treated with dox and then auxin. Chromatin-associated RNA was sequenced because it is highly enriched in the nascent transcripts that we wished to study. Rapid depletion of CPSF73 caused very obvious and widespread transcriptional readthrough as shown by the chromosome snapshot in Figure 1C. In this 5-Mb view, boundaries of gene transcription are easily observed, but become blurred by profound readthrough following CPSF73 elimination. Zoomed-in tracks of example protein-coding genes (*MAP2K4* and *YTHDC2*) further detail this effect where readthrough is for hundreds of kilobases (Fig. 1D). A metagene plot of transcription across all expressed genes, separated by at least 100 kb, showed that long readthrough is general when CPSF73 is depleted (Fig. 1E). This metagene also shows that CPSF73 loss causes a global reduction in gene body signal, suggesting that its presence positively impacts on transcription.

A termination defect in the absence of CPSF73 is not unexpected; however, the effect here is much greater than previously observed using RNAi (Supplemental Fig. S1A), and more reminiscent of that observed on some genes upon cell stress or viral infection (Vilborg et al. 2015; Heinz et al. 2018). It also reveals a broader range of CPSF73 functions than RNAi depletion, including a role in transcriptional termination of some long noncoding RNAs (Supplemental Fig. S1B,C). The extensive nature of the CPSF73 readthrough also highlights transcriptional interference in *cis*. Figure 1F shows an example whereby readthrough from *ERRFL1* reduces the expression of the convergent *PARK7* gene. Finally, CPSF73 depletion did not affect integrator-dependent snRNA gene termination demonstrating the specificity of these effects (Supplemental Fig. S1D). We conclude that functions of CPSF73 are indispensable for Pol II termination on protein-coding genes.

Predicted RNA products of XRN2-independent termination are not abundant

The very long readthrough seen without CPSF73 contrasts with our previous measurements of Pol II occupancy in the absence of XRN2 (Eaton et al. 2018). This showed a more moderate termination defect, defined as such because readthrough and Pol II signal eventually reduce to background levels even after XRN2 depletion. To compare XRN2 and CPSF73 effects, we analyzed our

previously generated nuclear RNA-seq from *XRN2-AID* cells and newly generated nuclear RNA-seq from C*SF73-AID* cell samples treated or not with auxin (Fig. 2A). Metagene plots were generated for all expressed genes separated from their neighbors by at least 20 kb. Loss of XRN2 shows clear stabilization of readthrough RNA just downstream from the PAS as we previously reported (Eaton et al. 2018), but this effect dissipated by 20 kb. In the absence of CPSF73, there was also strong stabilization of readthrough RNA, but this effect was maintained for the full 20 kb. Thus, there is generally longer readthrough seen when CPSF73 is depleted versus when XRN2 is lost. We focused subsequent efforts on establishing the basis for this difference as a means to understand the termination mechanism.

A reasonable explanation for this difference is XRN2-independent termination that still requires CPSF73 (Eaton and West 2018). This mechanism would release the RNA 3′ end from the terminated polymerase, leaving it vulnerable to 3′ → 5′ degradation by the exosome, which degrades many noncoding RNAs shortly after their synthesis (Preker et al. 2008). To test this, we transfected *XRN2-AID* cells with control siRNAs or siRNAs to deplete the exosome (EXOSC3 and EXOSC10) before treatment or not with auxin to remove XRN2. Chromatin-associated and nucleoplasmic RNA was isolated as we anticipated that termination products might be released from chromatin. To check the validity of this approach, quantitative reverse transcription and PCR (qRT-PCR) was performed to detect a PROMPT transcript upstream of *RBM39*, which is a well-characterized exosome substrate. This was strongly stabilized by exosome depletion, validating siRNA function, with the effect largest in the nucleoplasmic fraction (Supplemental Fig. S2A).

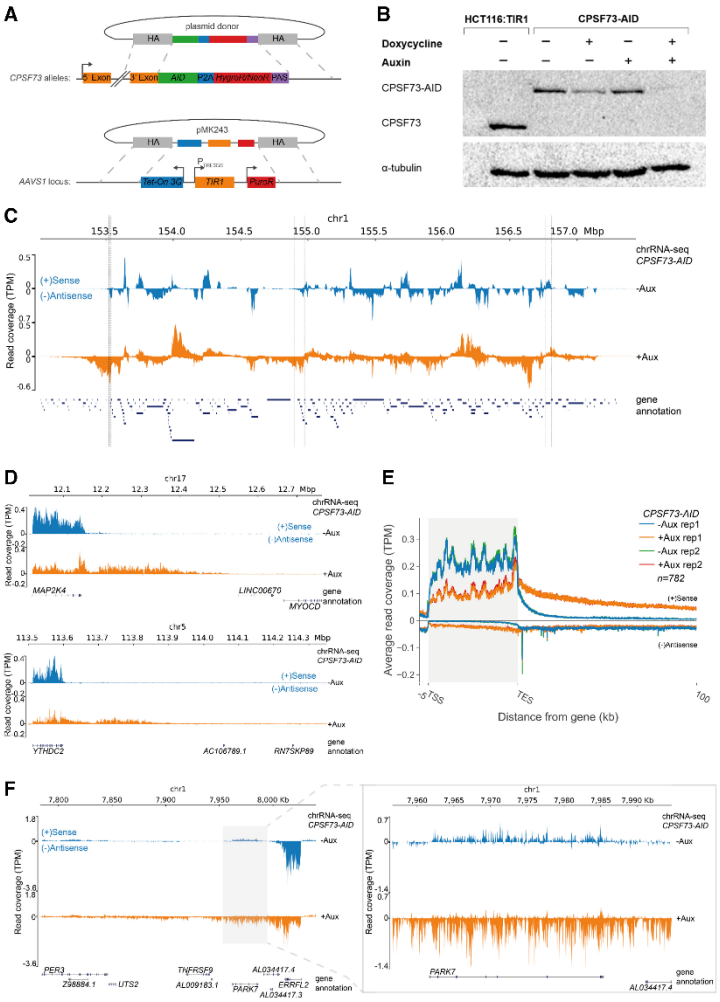


Figure 1. Acute loss of CPSF73 causes profound transcriptional readthrough. (*A*) Diagram depicting the strategy for AID tagging of *CPSF73*. Homology arms (gray) facilitate integration of a cassette containing the AID, separated from Neomycin or Hygromycin resistance genes by a P2A cleavage site and terminated by an SV40 PAS. Dox-inducible *TIR1* is inserted at the *AAVS1* locus. (*B*) Western blot of parental HCT116:TIR1 cells and *CPSF73-AID* cells that were untreated or treated with dox (18 h), auxin (18 h) or dox (18 h) and then auxin (3 h). The *top* panel shows CPSF73 and the *bottom* panel shows the tubulin loading control. (*C*) Chromosomal snapshot from chromatin RNA-seq of *CPSF73-AID* cells treated (orange) or not (blue) with auxin (3 h). Transcription units are seen in control samples (some examples shown with dotted lines). Annotated genes are in blue *below* the snapshot. *Y*-axis scale is transcripts per million (TPM). (*D*) IGV snapshots of *MAP2K4* and *YTHDC2* from the same experiment shown in *C*. (*E*) Metagene plot of expressed genes separated from their neighbors by at least 100 kb in chromatin RNA-seq of *CPSF73-AID* cells treated or not with auxin (3 h). Two biological replicates are plotted. (*F*) Example of transcriptional interference induced in *cis* by CPSF73 loss where readthrough from *ERRFI1* down-regulates the expression of *PARK7*. The *right* panel shows a zoomed in version demonstrating reduced *PARK7* signal coincident readthrough from nearby *ERRFI1*.

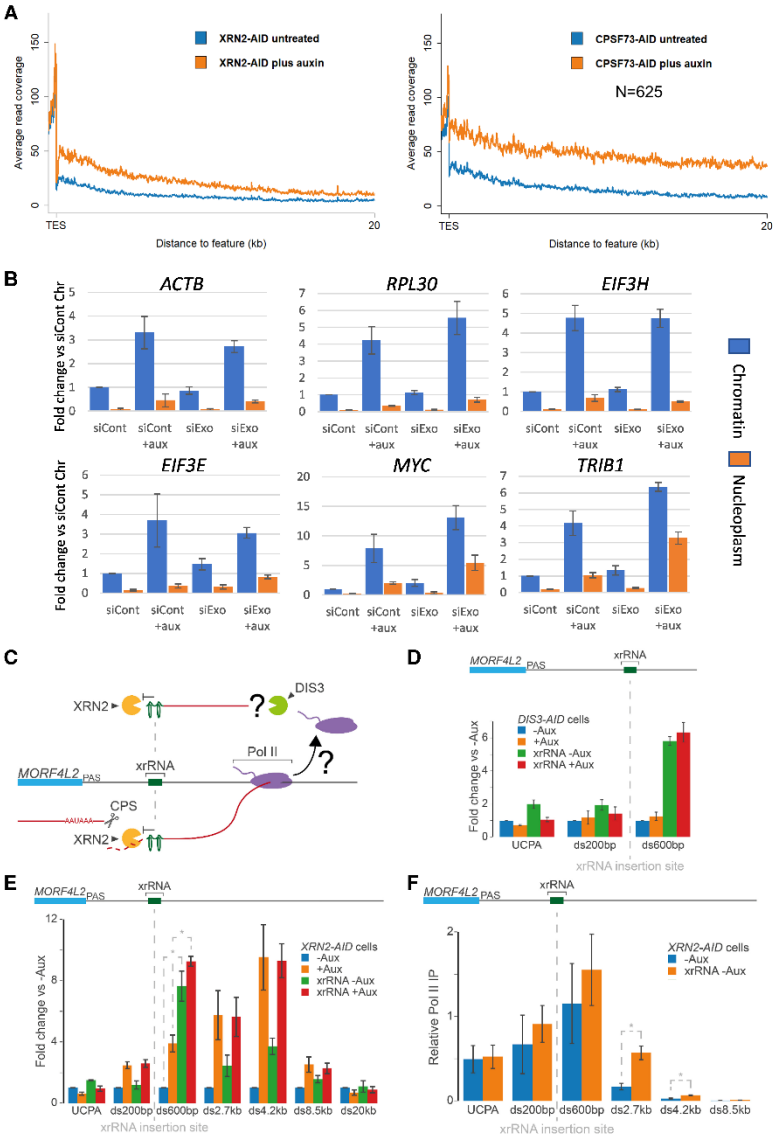


Figure 2. XRN2-independent termination is not readily apparent. (*A*) Metagene plots of expressed genes separated from their neighbors by at least 20 kb from nuclear RNA-seq of *CPSF73-AID* and *XRN2-AID* cells treated or not with auxin. Note that the *n* number is lower here than for the 100-kb window in Figure 1E due to stricter exclusion criteria applied to our previously generated *XRN2-AID* data (see Supplemental Material). (*B*) qRT-PCR analysis of *ACTB*, *MYC*, *RPL30*, *TRIB1*, *EIF3E*, and *EIF3H* 3′ flanking RNA in chromatin-associated and nucleoplasmic RNA from *XRN2-AID* cells transfected with control or exosome siRNAs before treatment or not with auxin (2 h). The graph shows fold change in RNA levels relative to the chromatin fraction of control cells following normalization to spliced ACTB. $n = 3$. Error bars are SEM. (*C*) Schematic illustrating the insertion of xrRNA into the 3′ flank of *MORF4L2* and its predicted effects on transcript degradation. If there is termination in the absence of 5′ → 3′ degradation this should be revealed by acute DIS3 depletion. CPS is cleavage and polyadenylation site. (*D*) qRT-PCR analysis of *MORF4L2* readthrough in unmodified *DIS3-AID* cells and *DIS3-AID* cells modified at *MORF4L2* by xrRNA insertion, then treated or not with auxin (2 h). The graph shows fold change in RNA at each amplicon relative to unmodified *DIS3-AID* cells not treated with auxin after normalizing to spliced ACTB. $n = 3$. Error bars are SEM. UCPA detects transcripts not cleaved at the PAS. (*E*) qRT-PCR analysis of *MORF4L2* readthrough in unmodified *XRN2-AID* cells and *XRN2-AID* cells modified at *MORF4L2* by xrRNA insertion, then treated or not with auxin (2 h). The graph shows fold change in RNA at each amplicon compared with unmodified *XRN2-AID* cells untreated with auxin after normalizing to spliced ACTB. $n = 3$. Error bars are SEM. (*) $P < 0.05$ illustrates the xrRNA effectiveness exemplified at ds600. (*F*) Pol II ChIP analysis on unmodified or xrRNA-modified *MORF4L2* in *XRN2-AID* cells that were not subject to treatment. Graph shows relative Pol II IP after normalizing to the *ACTB* ds1.7 kb amplicon, which is unmodified in both cell lines. $n = 3$. Error bars are SEM. (*) $P < 0.05$.

3′ flanking RNA derived from six protein-coding genes was then analyzed (Fig. 2B). As expected, all were up-regulated by loss of XRN2; however, depletion of the exosome alone had little effect on their levels or distribution. This argues that their degradation is primarily in the 5′ → 3′ direction. Codepletion of XRN2 and the exosome did not enhance RNA levels beyond those seen by eliminating XRN2 alone in most cases. *MYC* and *TRIB1* were two exceptions that showed an additional nucleoplasmic increase, which may indicate some XRN2-independent termination on those genes. However, these examples were still less sensitive to exosome loss than the *RBM39* PROMPT control. In our view, these data do not provide convincing evidence for efficient termination in the absence of XRN2.

These findings imply that degradation of 3′ flanking RNA is commonly unidirectional. To probe this more sensitively, we used CRISPR–Cas9 to edit the *MORF4L2* gene to include an XRN-resistant RNA (xrRNA). Isolated from West Nile virus, the xrRNA forms a structure that impairs 5′ → 3′ exonucleases and can be used to probe the directionality of RNA decay (Fig. 2C; Chapman et al. 2014; Horvathova et al. 2017). *MORF4L2* was chosen because it is well expressed and, being on the X chromosome, requires only one editing event to fully modify. It behaves typically insofar as its readthrough transcription is longer in the absence of CPSF73 versus XRN2 (Supplemental Fig. S2B). We performed this gene editing in our recently described *DIS3-AID* cells in which the major catalytic subunit of the exosome, DIS3, can be rapidly depleted by auxin (Davidson et al. 2019). Total RNA was isolated from xrRNA-modified or unmodified *DIS3-AID* cells treated or not with auxin and qRT-PCR used to assay the levels of *MORF4L2* RNA upstream of and downstream from the xrRNA insertion (Fig. 2D). Similar levels of RNA were obtained upstream of the xrRNA under all conditions; however RNA downstream from the xrRNA was stabilized consistent with impaired 5′ → 3′ degradation. This stabilization was not exacerbated by DIS3 loss again suggesting unidirectional decay and infrequent XRN2-independent transcriptional termination. Importantly, DIS3 elimination strongly stabilized other exosome substrates in the same samples (Supplemental Fig. S2C).

An XRN-resistant RNA does not induce profound readthrough

Another explanation for why readthrough is short in the absence of XRN2 could be the presence of alternative 5′ → 3′ exonucleases or trace levels of XRN2 remaining after auxin treatment. To test this, we used CRISPR–Cas9 to insert the XRN-resistant RNA (xrRNA) downstream from *MORF4L2* in *XRN2-AID* cells to inhibit 5′ → 3′ degradation generally. RNA was isolated from *XRN2-AID* cells modified or unmodified at *MORF4L2* following treatment or not with auxin. qRT-PCR analysis of readthrough transcription showed that RNA between the PAS and xrRNA was stabilized only when XRN2 was depleted as expected (Fig. 2E). Similar to Figure 2D, RNA downstream from the xrRNA was stabilized even in the presence of XRN2 demonstrating that the xrRNA inhibits 5′ → 3′ degradation. Depletion of XRN2 from these cells produced the expected readthrough but its extent (positions beyond ds600) was not significantly greater than in unmodified cells depleted of XRN2. It therefore seems unlikely that alternative 5′ → 3′ exonucleases generally participate in termination.

Efficient pursuit of Pol II by XRN2 is important for termination

The observation that XRN2 initiates degradation at the PAS but is impeded beyond the xrRNA allows a key prediction of the torpedo model to be tested: that Pol II capture, rather than simply degrading RNA, is critical. Specifically, because the xrRNA impedes degradation that has already been initiated at the PAS, it should only result in a termination defect if the continued pursuit of Pol II is important. We performed Pol II ChIP on *MORF4L2* in xrRNA-modified and unmodified *XRN2-AID* cells without depletion of XRN2 (Fig. 2F). The presence of the xrRNA resulted in significantly enhanced Pol II occupancy downstream from its position (2.7 kb and 4.2 kb), strongly suggesting that XRN2 must catch Pol II (or at least get close to it) to terminate it.

Polymerases accumulate over termination regions in the absence of XRN2

Data so far show that CPSF73 loss causes runaway readthrough, which is much less extensive when XRN2 is depleted. This difference is not due to efficient XRN2-independent mechanisms or other 5′ → 3′ exonucleases, active when XRN2 is eliminated. We therefore hypothesized that Pol II accumulating beyond the PAS when XRN2 is eliminated is slowed rather than terminated and that this is why readthrough is shorter than in the absence of CPSF73. To interrogate this, we performed Pol II ChIP on *XRN2-AID* and *CPSF73-AID* cells treated or not with auxin using *MYC* and *ACTB* as model genes (Fig. 3A). Auxin was added for 3 h in both cases to eliminate/minimize any effects of different depletion times. Loss of CPSF73 or XRN2 induces a termination defect on both genes as expected. The effect of CPSF73 loss was greatest at positions furthest downstream from the PAS, whereas the XRN2 effect was more prominent at positions within 2.5 kb of the PAS. This difference is consistent with the idea that polymerases pile up over termination regions in the absence of XRN2 but not when CPSF73 is depleted. This pile up is likely due to slow elongation rather than complete arrest because 3′ flanking region Pol II can still incorporate 4-thiouridine in the absence of XRN2 (Supplemental Fig. S3A).

We next assayed whether the difference in Pol II behavior caused by XRN2 vs CPSF73 depletion is associated with any changes in its modification status. The CTD of the largest Pol II subunit is heavily modified during the transcription cycle (Eick and Geyer 2013). Thr4p is the most obvious candidate for CPSF73-dependent modification as it predominantly occurs after the PAS and its

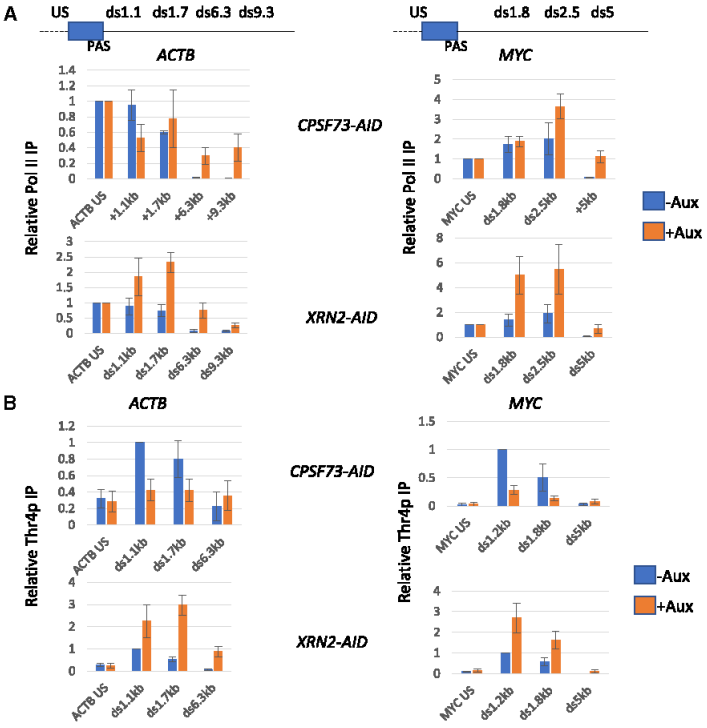


Figure 3. Pol II piles up beyond the PAS when XRN2 is depleted. (*A*) Pol II ChIP analysis on *ACTB* and *MYC* performed in *CPSF73-AID* or *XRN2-AID* cells treated or not with auxin (3 h). Graph shows relative Pol II IP normalized to ACTB US/MYC US in each sample. $n = 3$. Error bars are SEM. (*B*) Thr4p ChIP analysis on *ACTB* and *MYC* performed in *CPSF73-AID* or *XRN2-AID* cells treated or not with auxin (3 h). Graph shows relative Thr4p IP normalized ACTB ds1.1 kb and MYC ds1.2 kb in the respective cell lines not treated with auxin. $n = 3$. Error bars are SEM.

position is shifted downstream when CPSF73 is depleted by RNAi (Schlackow et al. 2017). We assayed Thr4p on *ACTB* and *MYC* in *CPSF73-AID* and *XRN2-AID* cells grown with or without auxin and observed its expected enrichment beyond the PAS in control samples (Fig. 3B). XRN2 loss caused an increase in Thr4p signal downstream from the PAS for both genes similar to the increase in total Pol II occupancy observed in Figure 3A; however, Thr4p is reduced when CPSF73 is eliminated. Thr4p is therefore downstream from CPSF73 activity but upstream of XRN2-dependent transcriptional termination, revealing a chemical difference between Pol II occupying 3′ flanking regions upon CPSF73 versus XRN2 depletion.

Polymerases accumulate over termination regions in the absence of XRN2 but transcribe through them when CPSF73 is eliminated

Reduced Pol II occupancy beyond the PAS is usually interpreted as consequential of transcriptional termination. Our data argue that XRN2 depletion enriches polymerases that can neither elongate well nor terminate. We designed an experiment to test this more directly using a hepatitis δ ribozyme that cleaves very efficiently and can be inactivated by a single mutation (δRZ[WT/MT]) (Fong et al. 2009). RZ cleavage produces an upstream product that is degraded from its 3′ end and a downstream product that cannot be degraded due to a 5′OH (Stevens and Maupin 1987; Muniz et al. 2015). Transcription of the RZ will result in a stable downstream product, the presence of which can be used to report instances of transcription beyond its position. *RBM3* was modified with (δRZ[WT/MT]) in *CPSF73-AID* and *XRN2-AID* cells and was chosen as an alternative to *MORF4L2* (it is also X-chromosomal). The RZ was inserted just downstream from where termination normally occurs and loss of XRN2 induces some transcription beyond the insertion site (Supplemental Fig. S3B). If, in the absence of XRN2, the ultimate position of Pol II occupancy defines an auxiliary termination site, then most Pol II should transcribe the RZ when XRN2 is depleted. If instead most Pol II accumulates upstream of the RZ when XRN2 is eliminated only a fraction will transcribe beyond it. *CPSF73-AID* cells act as a control because its depletion induces profound readthrough beyond the RZ insertion site (Supplemental Fig. S3B).

qRT-PCR was performed to look at readthrough in δRZ[WT/MT]-modified *CPSF73-AID* cells and unmodified *CPSF73-AID* cells treated or not with auxin (Fig. 4A). In unmodified cells, CPSF73 loss induced the expected increase in transcriptional readthrough causing a 25-fold to 50-fold increase in signal at 8.5 and 11 kb beyond the *RBM3* PAS. The result was similar when *RBM3* was modified with δRZ[MT]. With δRZ[WT], there was again strong readthrough when CPSF73 was lost. Importantly, the RZ cleaves in this setting because its upstream product does not accumulate when CPSF73 is depleted from

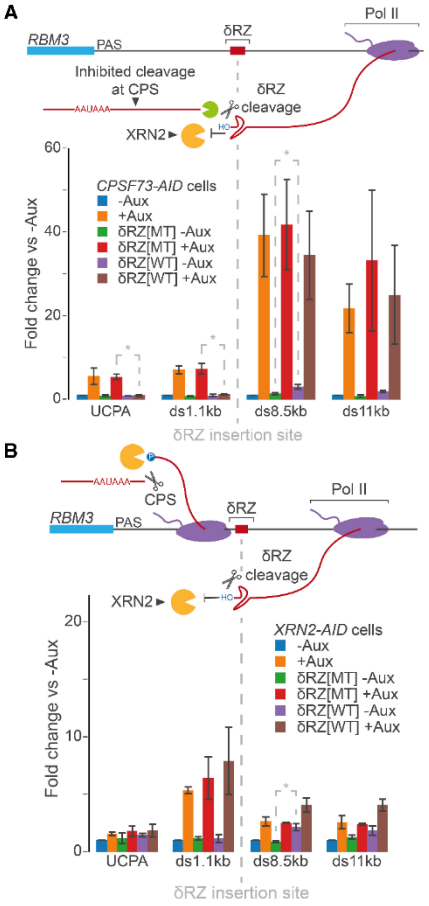


Figure 4. Pol II piles up over termination regions when XRN2 is depleted whereas CPSF73 loss induces runaway transcription. (*A*) qRT-PCR analysis of *RBM3* readthrough in unmodified *CPSF73-AID* cells and *CPSF73-AID* cells modified at *RBM3* by addition of δRZ[WT/MT] and then treated or not with auxin (3 h). The schematic shows the RZ insertion and predicted impact on RNA degradation. The graph shows fold change in RNA over each amplicon relative to unmodified *CPSF73-AID* cells not treated with auxin after normalising to spliced ACTB. $n \geq 3$. Error bars are SEM. (*) $P < 0.05$ to highlight significance of smaller fold effect changes commented on in the main text. UCPA detects transcripts uncleaved at the PAS. (*B*) qRT-PCR analysis of *RBM3* readthrough in unmodified *XRN2-AID* cells and *XRN2-AID* cells modified at *RBM3* by addition of δRZ[WT/MT] and then treated or not with auxin (3 h). The graph shows fold change in RNA at each amplicon relative to unmodified *XRN2-AID* cells after normalizing to spliced ACTB. $n \geq 3$. Error bars are SEM. (*) $P < 0.05$ to highlight significance of smaller fold effect changes commented on in the main text.

δRZ[WT] samples but shows robust up-regulation in the presence of δRZ[MT] (see UCPA and 1.1-kb amplicons). As expected, the 3′ cleavage product cannot be degraded as its accumulation is similar in δRZ[MT] and δRZ[WT] scenarios.

In *XRN2-AID* cells, δRZ[MT]-modification of *RMB3* gives a similar result to unmodified *XRN2-AID* cells in that XRN2 depletion leads to stabilization of 3′ flanking region RNA (Fig. 4B). For the δRZ[WT], there is mild up-regulation of the downstream 8.5-kb amplicon even when XRN2 is present. This is likely due to a small fraction of Pol II transcribing beyond its position as a similarly modest stabilization is also seen here in untreated *CPSF73-AID* cells modified by δRZ[WT]. Compared with loss of CPSF73, XRN2 depletion has much less impact on readthrough beyond the δRZ[WT]. This supports the idea derived from xrRNA-modified *MORF4L2* that alternate means of 5′ → 3′ degradation do not promote efficient termination when XRN2 is depleted. Finally, XRN2 depletion also results in stabilization of RNA upstream of the RZ instead of the destabilization seen without CPSF73, suggesting that comparatively few polymerases go beyond the RZ in its absence. Thus, the frontier of Pol II signal seen when XRN2 is lost is unlikely to represent an auxiliary termination site and is more likely to reflect the forefront of polymerases accumulated upstream of it. Such accumulation of Pol II beyond the PAS upon XRN2 depletion is general, based on analysis of our previously published mNET-seq in cells lacking XRN2 (Supplemental Fig. S3C).

PP1 activity aids termination by XRN2

Recent results in fission yeast implicate PP1 in reducing elongation rate after the PAS through its dephosphorylation of SPT5 and Pol II CTD (Kecman et al. 2018; Parua et al. 2018). We tested whether PP1 activity is involved in the Pol II slowing first using a small molecule inhibitor approach. This has possible advantages over protein depletion: It is more acute, may not disrupt PP1-associated complexes, and is more easily combined with XRN2 depletion to test the relationship between PP1 activity and torpedo termination in the same cells. We assessed the impact of PP1 on XRN2-dependent termination using Pol II ChIP on *ACTB* and *MYC* in *XRN2-AID* cells treated with auxin, the selective PP1 inhibitor, tautomycetin (Choy et al. 2017), or both (Fig. 5A). XRN2 loss showed the expected termination defect on both genes characterized by an accumulation of Pol II downstream from the PAS. PP1 inhibition gave a similar Pol II profile to the untreated control situation with a slight tendency for more readthrough transcription. When XRN2 was depleted from PP1 inhibited cells, the post-PAS accumulation of Pol II that we had characterized as resulting from accumulated slow polymerase was reduced. However, the readthrough at extended positions was maintained. These observations support the idea that PP1 facilitates XRN2-dependent termination by promoting Pol II slowing.

PP1-assisted Pol II slowing might facilitate termination by expediting the capture of polymerases by XRN2, which we showed to be an important part of the mechanism in Figure 2F. To interrogate this, total RNA was isolated

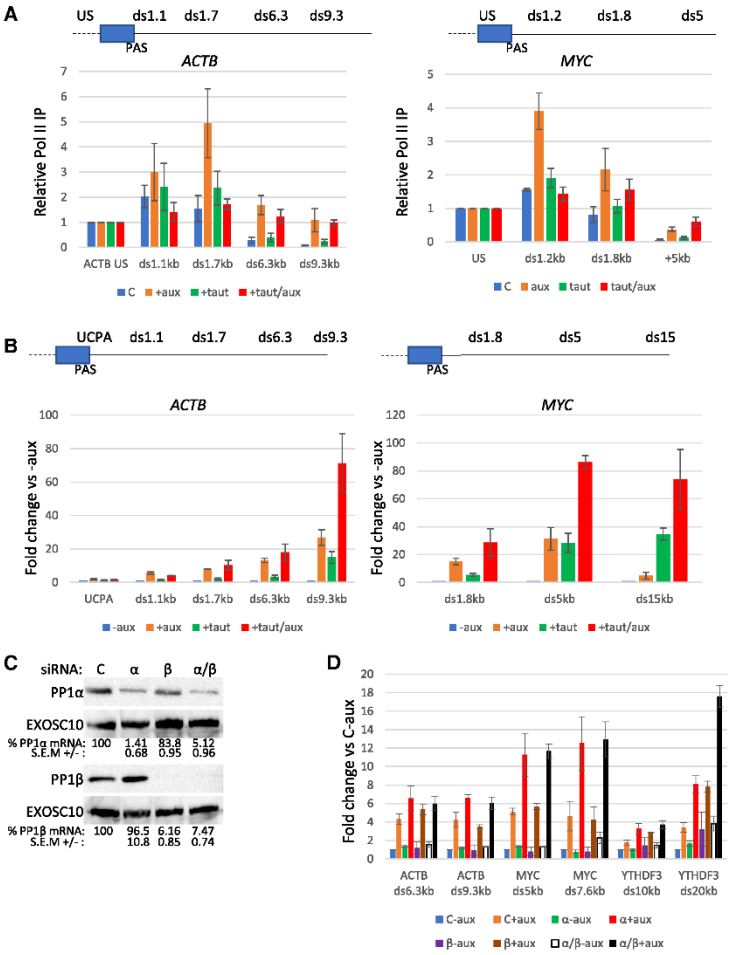


Figure 5. PP1 activity underpins the piling up of Pol II in the absence of XRN2. (*A*) Pol II ChIP on *ACTB* and *MYC* performed in *XRN2-AID* cells that were either untreated or treated with auxin, tautomycetin or both (auxin 2 h, tautomycetin 18 h). Graph shows Pol II IP relative to ACTB US/MYC US for each sample. $n = 3$. Error bars are SEM. (*B*) qRT-PCR analysis of *ACTB* and *MYC* readthrough transcription in *XRN2-AID* cells that were either untreated or treated with auxin, tautomycetin, or both (auxin 2 h, tautomycetin 18 h). Graph shows RNA fold change at each amplicon relative to untreated (-aux) cells following normalization to spliced ACTB. $n = 3$. Error bars are SEM. (*C*) Western blotting of *XRN2-AID* cells treated with control, PP1α, PP1β, or PP1α and β siRNAs. EXOSC10 is shown as a loading control. mRNA levels ($n = 3$) are shown under each blot relative to those in control siRNA treated cells (shown as 100%) after normalization to spliced ACTB transcripts. (*D*) qRT-PCR analysis of *ACTB*, *MYC*, and *YTHDF3* readthrough transcription in *XRN2-AID* cells that were treated with control, PP1α, PP1β, or PP1α and β siRNAs and then with auxin (2 h) or not. Graph shows RNA fold change at each amplicon relative to control siRNA transfected cells not treated with auxin following normalization to spliced ACTB. $n = 3$. Error bars are SEM.

from *XRN2-AID* cells following PP1 inhibition and/or XRN2 depletion and transcriptional readthrough was assayed for *ACTB* and *MYC* by qRT-PCR (Fig. 5B). As expected, XRN2 loss induced readthrough at both genes. Tautomycetin also induced transcriptional readthrough, suggesting that PP1 activity is important for efficient termination. Treatment with tautomycetin and auxin gave the biggest effect and was especially strong at positions furthest beyond the PAS. These effects are larger than in the above ChIP experiments. This is because ChIP signal derives from Pol II occupancy of a short fragment of DNA (due to prior sonication), whereas qRT-PCR signal derives from intact transcripts and reports everything over and beyond a tested amplicon.

Mass spectrometry identified PP1α and β isoforms as components of the 3′ end processing complex (Shi et al. 2009). We therefore specifically tested their involvement in transcriptional readthrough using RNAi to deplete either or both from *XRN2-AID* cells (Fig. 5C). We then used qRT-PCR to assay transcriptional readthrough under these conditions following treatment or not with auxin to deplete XRN2 (Fig. 5D). Depletion of PP1 by itself did not lead to extended readthrough; however, it did so when XRN2 was subsequently depleted. This was the case for both *MYC* and *ACTB* with the effect more subtle for the latter. Analysis of another gene (*YTFDH3*) confirmed that PP1 depletion enhances readthrough when XRN2 is eliminated. An additional protein phosphatase inhibitor, Calyculin A, was used to further confirm the protein phosphatase effect on extended *ACTB* readthrough (Supplemental Fig. S4). The combined interpretation of this ChIP and qRT-PCR is that PP1 is important for Pol II slowing after the PAS and for its efficient capture and termination by XRN2.

Directed RNaseH1 activity bypasses the requirement for CPSF73 in transcriptional termination

The importance of XRN2 in transcriptional termination is shown by the genome-wide termination defect in its

absence. PP1 may also play a direct active role in termination or it could serve to facilitate the XRN2-dependent process by slowing Pol II down. Disentangling these possibilities requires experiments to isolate allosteric and torpedo components from one another. To do this we tested whether alternative endoribonucleolytic cleavage could support termination on a protein-coding gene following CPSF73 loss, as this would potentially separate its RNA cleavage and allosteric functions. As the XRN2-incompatible RZ cleavage does not support termination in the absence of CPSF73 (Fig. 4A), we used RNaseH1, which cuts RNA:DNA hybrids and generates substrates for XRN2 (Hori et al. 2015). To concentrate high levels of RNaseH1 in the nucleus we integrated it into the genome after replacing its mitochondrial localization signal with a nuclear localization signal (to make *NLS-RNASEH1*). Because the requisite puromycin-resistance marker had been used in *CPS73-AID* cells (to introduce TIR1), dox-inducible *NLS-RNASEH1* was integrated into our previously described *CPSF73-DHFR* cell line (Eaton et al. 2018). In those cells, CPSF73 is depleted by removing trimethoprim (TMP) from growth media. The Western blot in Figure 6A confirms the successful introduction dox-inducible *NLS-RNASEH1*.

Next, we designed a GapmeR to the last exon of *ACTB* close to the PAS. GapmeRs are modified antisense oligonucleotides that, when bound to their target transcript, result in cleavage by RNaseH1. ACTB or control GapmeRs were transfected into *NLS-RNASEH1*-expressing *CPSF73-DHFR* cells. We then added dox in the presence or absence of TMP to retain or deplete CPSF73. qRT-PCR, using primers that could only detect uncleaved products, revealed very efficient (>90%) GapmeR-induced cleavage (Fig. 6B). GapmeR-directed cleavage also suppressed the strong readthrough caused by CPSF73 depletion as judged by the reduction in RNA levels beyond the *ACTB* PAS (Fig. 6C). This contrasts with the inability of RZ cleavage to do the same on *RBM3* (Fig. 4A), correlating these effects with the compatibility of cleaved RNA with 5′ → 3′ degradation. Crucially, Pol II ChIP confirmed that the large termination defect caused by CPSF73 depletion was almost completely suppressed by introducing the ACTB GapmeR (Fig. 6D). This experiment shows that CPSF73 itself is not a strict requirement for termination and that an important

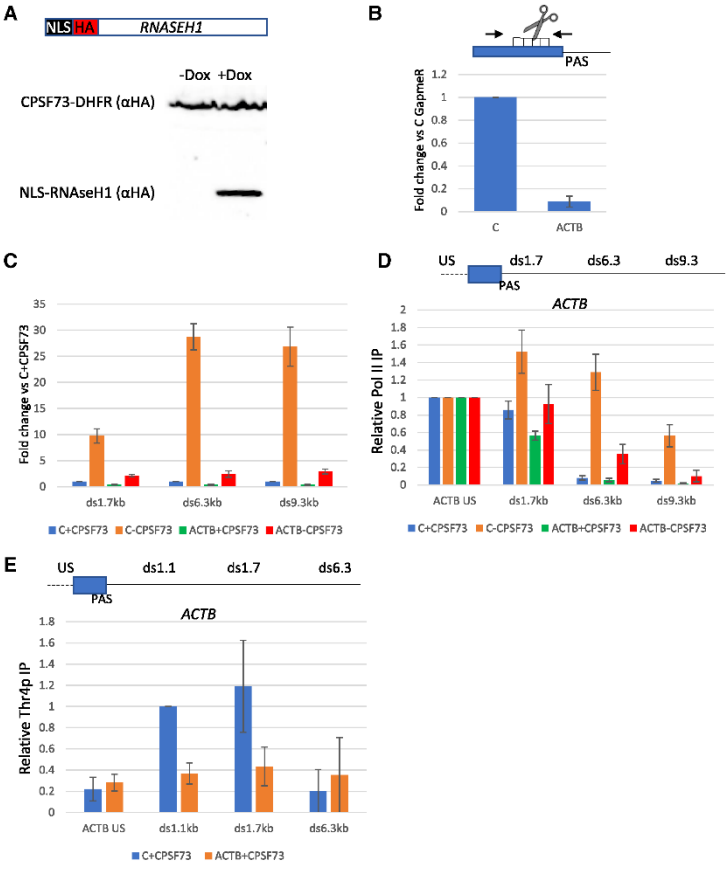


Figure 6. Directed RNaseH1 activity promotes transcriptional termination in the absence of CPSF73. (*A*) Western blot to detect NLS-RNaseH1 introduced into *CPSF73-DHFR* cells. Its expression is dox-dependent and CPSF73 is shown as a loading control. (*B*) qRT-PCR of NLS-*RNASEH1*-modified *CPSF73-DHFR* cells transfected with control or ACTB GapmeRs. Primers flanking the GapmeR targeting site were used to assess cleavage efficiency shown as a fold change in RNA levels relative to control GapmeR following normalization to spliced GAPDH. $n = 4$. Error bars are SEM. (*C*) qRT-PCR of RNaseH1-modified *CPSF73-DHFR* cells transfected with control or ACTB GapmeR and grown in the presence or absence of TMP (12 h). Graph shows fold change in RNA levels compared with control GapmeR transfected cells grown in the presence of TMP (C + CPSF73) following normalization to unspliced ACTB RNA. $n = 3$. Error bars are SEM. (*D*) Pol II ChIP of RNaseH1-modified *CPSF73-DHFR* cells transfected with control or ACTB GapmeRs and grown in the presence or absence of TMP (12 h). Graph shows relative Pol II IP normalized to Pol II occupancy upstream of the PAS (ACTB US) in each condition. $n = 3$. Error bars are SEM. (*E*) Thr4p ChIP of RNaseH1-modified *CPSF73-DHFR* cells transfected with control or ACTB GapmeRs. Graph shows relative Pol II IP normalized to occupancy over the ds1.1-kb amplicon in control GapmeR transfected cells grown in TMP (C + CPSF73). $n = 3$. Error bars are SEM.

part of its normal function in the process is to cleave the transcript.

As RNaseH1 cleavage of *ACTB* RNA promotes termination in the absence of CPSF73, it is likely to do so without associated Pol II slowing. Based on our model, this means that RNaseH1 cleavage might be relatively faster than at the *ACTB* PAS to facilitate Pol II pursuit by XRN2. To test this assumption, we performed Thr4p ChIP in RNaseH1-expressing *CPSF73-DHFR* cells transfected with control or ACTB GapmeRs. Because Thr4p depends on CPSF73 (see Fig. 3B), its presence will indicate prior CPSF73 activity. In control treated cells, Thr4p was highest at 1.1 and 1.7 kb beyond the PAS as expected (Fig. 6E). However, Thr4p did not accumulate at those positions when the ACTB GapmeR was transfected, even though CPSF73 is still present. This suggests that GapmeR-directed cleavage is often faster than at the PAS and provides an explanation for how it promotes efficient termination even in the absence of CPSF73.

RNA cleavage in different locations and by different RNases promotes transcriptional termination without CPSF73

NEAT1 and *MALAT1* noncoding genes terminate via CPSF73-independent yet XRN2-dependent mechanisms (Supplemental Fig. S5), perhaps because their RNAs undergo 3′ end cleavage by RNaseP/Z (Wilusz et al. 2008). We used CRISPR–Cas9 to insert the 3′ end processing element from *MALAT1* into the 3′ flank of *MORF4L2* in *CPSF73-AID* cells. We performed qRT-PCR to assay *MORF4L2* readthrough in modified and unmodified *CPSF73-AID* cells treated or not with auxin (Fig. 7A). While CPSF73 loss induced very strong readthrough at the unmodified *MORF4L2*, this effect was suppressed in the presence of the *MALAT1* 3′ end, strongly suggesting that RNaseP/Z bypasses the need for CPSF73. Interestingly, the MALAT1 3′ end also caused an increase in the level of *MORF4L2* transcripts that were not cleaved at the PAS, even when CPSF73 is present. This argues that RNaseP/Z cleavage often occurs before/without cleavage at the upstream PAS similar to GapmeR-directed RNaseH1 at *ACTB*. ChIP analysis confirmed that the MALAT1 3′ end suppresses the effects of CPSF73 loss on *MORF4L2* termination (Fig. 7B). Thus, alternative (and endogenous) RNases can promote termination on protein-coding genes in the absence of CPSF73.

As these directed RNase experiments occur close to a PAS, there might be other PAS-dependent features that promote termination even without CPSF73. To see whether this is the case, we designed a GapmeR to cleave the second intron of *ETF1* far upstream of its PAS. Importantly, *ETF1* is unlikely to contain cryptic PAS elements since its transcripts do not undergo premature CPA when U1 snRNA is inhibited (Kaida et al. 2010). To test the potential for *ETF1* intron cleavage to promote premature transcriptional termination, we performed qRT-PCR in *NLS-RNASEH1* expressing *CPSF73-DHFR* cells (without depleting CPSF73) transfected with control or ETF1 GapmeRs (Fig. 7C). We also analyzed equivalent samples following splicing inhibition using Pladienolide B (PlaB) in case intron removal confounded pre-mRNA detection. Indeed, PlaB enhanced all three intron-derived species; however, this was almost fully suppressed by the ETF1 GapmeR. This included over the cleavage site (showing efficient cutting) and at downstream position. Pol II ChIP confirmed that the ETF1 GapmeR induced premature termination of transcription and did so in the presence and absence of splicing (Fig. 7D). These experiments indicate that RNA degradation is, at least in principle, sufficient for termination, with allosteric features used to facilitate it in practice. While a PAS-induced modification is not strictly necessary for transcriptional termination, it is probably required for the XRN2-dependent process to be most efficient.

Discussion

We have shown that rapid CPSF73 depletion causes runaway readthrough, which we propose is because processes underpinning transcriptional termination are lost. We identified allosteric and torpedo elements of the transcriptional termination mechanism that were enriched by rapid depletion of XRN2. Our experiments reveal the following mechanism: CPSF73 function (and likely PAS cleavage) promotes Thr4p with PP1 activity playing an important role in slowing Pol II down to facilitate its capture by XRN2. The allosteric events most likely aid XRN2 rather than acting as a separate or mutually exclusive termination pathway because cleavage of primary transcripts promotes termination in the absence of CPSF73 and outside of a PAS context. Consequently, we propose that the long-standing allosteric and torpedo models for transcriptional termination can be unified into a single mechanism with XRN2 likely providing the termination activity (Fig. 7E).

This conclusion followed from our failure to detect efficient/general XRN2-independent termination, although we did observe some evidence for this in two (*MYC* and *TRIB1*) cases. However, 3′-flanking transcripts were always much more sensitive to XRN2 loss than to exosome depletion and remained largely chromatin-associated under all circumstances. While there may be inefficient or stochastic XRN2-independent termination, it does not seem to be linked by a common mechanism. We do not rule out XRN2-independent termination that evades our detection or that occurs on certain genes, but our subsequent finding that PP1 inhibition causes longer readthrough when XRN2 is depleted argues against efficient alternatives.

Our results suggest that endoribonucleolytic cleavage might be an elemental function of CPSF73 in termination. However, polymerases display runaway readthrough in the absence of CPSF73 and are not modified by Thr4p, suggesting that its absence also prevents allosteric events. There are two broad explanations for this: The first is that CPSF73 loss prevents the assembly of other 3′ end processing factors that also promote allostery, and the second is that allosteric events occur downstream from PAS

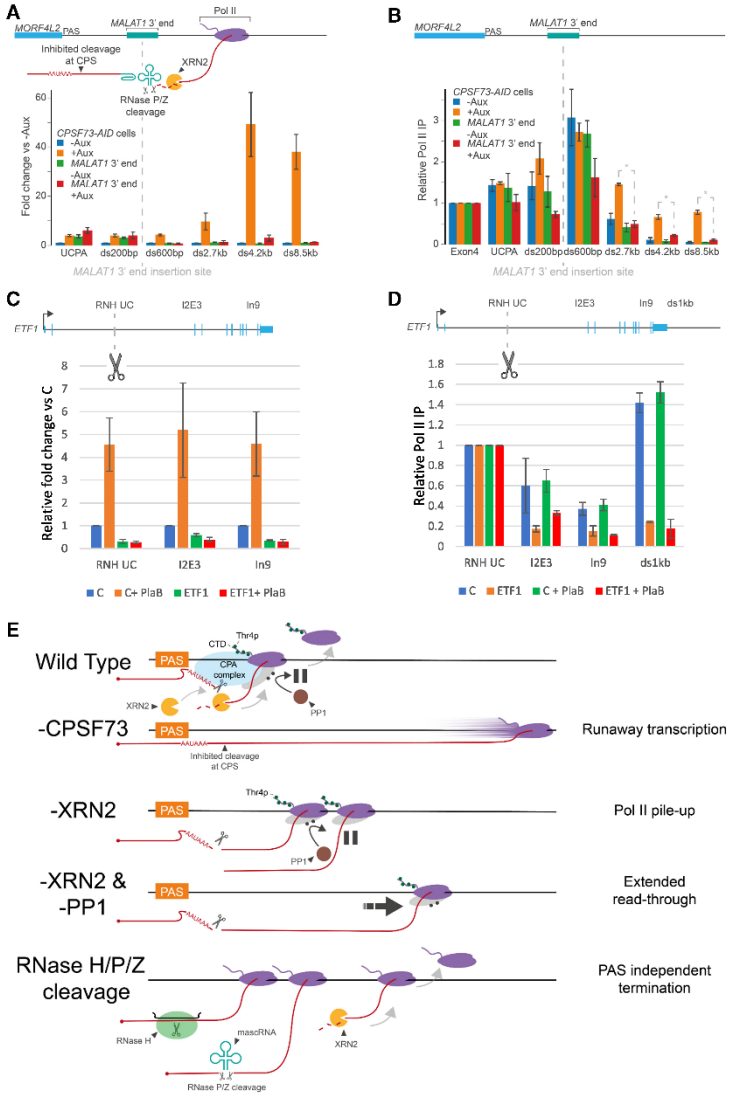


Figure 7. Other RNases promote termination in the absence of CPSF73, which can also occur within gene bodies. (*A*) qRT-PCR analysis of *MORF4L2* readthrough in unmodified *CPSF73-AID* cells and *CPSF73-AID* cells modified at *MORF4L2* by inserting the *MALAT1* 3′ end and then treated or not with auxin (3 h). The graph shows fold change in RNA at each amplicon relative to unmodified CPSF73-*AID* cells not treated with auxin after normalizing to spliced ACTB. $n = 3$. Error bars are SEM. UCPA detects transcripts not cleaved at the PAS. (*B*) Pol II ChIP analysis of *MORF4L2* in unmodified *CPSF73-AID* cells and CPSF73-*AID* cells modified at *MORF4L2* by inserting the *MALAT1* 3′ end and then treated or not with auxin (3 h). The graph shows relative Pol II IP normalized to occupancy over *MORF4L2* exon 4 in each case. $n = 3$. Error bars are SEM. (*C*) qRT-PCR analysis of *ETF1* GapmeR cleavage effects in RNaseH1-modified *CPSF73-DHFR* cells transfected with control or ETF1 GapmeRs, grown in the presence of TMP before treatment or not with PlaB (3 h). The graph shows fold change in RNA at each amplicon relative to control GapmeR-treated cells grown without PlaB after normalizing to spliced ACTB. $n = 3$. Error bars are SEM. GapmeR cleavage site is illustrated (scissors) in the schematic. (*D*) Pol II ChIP of RNaseH1-modified *CPSF73-DHFR* cells transfected with control or ETF1 GapmeRs and grown in the presence of TMP before treatment or not with PlaB (3 h). Graph shows relative Pol II IP normalized to occupancy over the GapmeR target site in each case. $n = 3$. Error bars are SEM. (*E*) Model. Normally, Pol II pauses beyond the PAS in a manner requiring PP1 activity and associated with Thr4p, which facilitates its pursuit and termination by XRN2. (-CPSF73) Pol II engages in runaway readthrough transcription; (-XRN2) paused Pol II is paused over 3′ flanking regions unable to terminate efficiently; (-XRN2 and PP1) this situation shows extended readthrough relative to loss of XRN2 alone, which is associated with reduced pausing after the PAS; (RNaseH/P/Z cleavage) bypasses of CPSF73/PAS requirement by promoting Pol II termination in association with 5′ → 3′ RNA degradation. Filled dots represent phosphates.

cleavage. While not mutually exclusive, we favor the latter because of our previous finding that inactive CPSF73 does not support termination (Eaton et al. 2018). Furthermore, PP1 depletion does not affect PAS cleavage (Shi et al. 2009), but still causes extended transcriptional readthrough.

Part of the mechanism for how PP1 activity causes Pol II to slow down is explained by recent data from fission yeast, in which dephosphorylation of SPT5 occurs specifically beyond the PAS (Parua et al. 2018). Consistently, and while we were revising this manuscript, the Bentley lab demonstrated dephosphorylation of SPT5 beyond the PAS in human cells associated with XRN2-dependent termination (Cortazar et al. 2019). Even so, other relevant factors may be subject to PP1 activity because phosphomimetic mutations in SPT5 do not lead to termination defects in fission yeast (Kecman et al. 2018). For instance, p54nrb is among the many PP1-interacting proteins and promotes the recruitment of XRN2 to genes (Kaneko et al. 2007; Heroes et al. 2013). Elucidating the complete network of phospho-regulation during transcriptional termination is an interesting and extensive area for future study.

The importance of Pol II slowing for XRN2-dependent termination likely relates to the respective rates of RNA transcription and degradation. Pol II elongation rate is on average ~2 kb/min with some variation on different genes (Singh and Padgett 2009; Fuchs et al. 2014). Although the rate of XRN2 degradation is unknown, the closely related 5′ → 3′ exonuclease, XRN1, has recently been measured at ~2 kb/min (Hoek et al. 2019). Assuming that XRN2 behaves similarly, rates of synthesis and degradation are closely matched. Without pausing, XRN2 may be unable to catch the polymerase where PAS cleavage is slow or when Pol II elongation is faster than the average rate. In support of closely matching rates, increasing Pol II elongation rate by only 220 nt/min induces a substantial downstream shift of termination positions (Fong et al. 2015).

Finally, the GapmeR and RNase P/Z experiments show that CPSF73-dependent events can be bypassed by rapid cleavage. This result also implies that the elemental role of CPSF73 (and the PAS) in termination is to provide an end for XRN2, otherwise alternative cleavage would not function in its absence. It also demonstrates that CPSF73-dependent changes occurring beyond the PAS facilitate termination (we propose, by slowing down Pol II) rather than directly contributing. This is consistent with the ability for XRN2 to promote termination in vitro in the absence of other factors (Park et al. 2015). This experiment also has additional implications for the study of noncoding RNA function, which often relies on the assumption that their targeting by antisense oligonucleotides does not affect transcription of their locus. We suggest that premature transcriptional termination might also occur in some of these experiments.

Materials and methods

Cell culture and transfections

HCT116 cells were cultured in DMEM (high glucose) containing penicillin/streptomycin and 10% FBS at 37°C. DNA transfections were done with Jetprime (polyplus) and RNAi with Lipofectamine RNAiMAX (Life Technologies) following the manufacturers' guidelines. Two siRNA transfections were used: the second 24 h after the first and RNA isolation 48 h after that. PP1 RNAi was performed using a single siRNA transfection with RNA isolated 72 h later. GapmeRs were transfected at 10 nM in 24-well or 100-mm dishes for RNA and ChIP experiments, respectively. This was done using 1.5 or 10 µL of Lipofectamine RNAiMAX in a final combined volume of 50 or 500 µL of Opti-MEM (Life Technologies). Dox was used at 1 µg/mL, auxin was used at 500 µM, tautomycetin was used at 500 nM, PlaB was used at 1 nM, and calyculin A was used at 5 nM. TMP was used at 20 µM and withdrawn for 12 h to deplete CPSF73-DHFR.

Cell lines and cloning

The *XRN2-AID* and *CPSF73-eDHFR* cell lines are previously described (Eaton et al. 2018). *NLS-RNASEH1* was synthesized by Integrated DNA Technologies and cloned into pMK243 (Addgene 72835) digested with MluI and BglII. This plasmid was transfected with AAVS1 T2 CRISPR plasmid (Addgene 72833), and colonies were selected in 1 µg/mL puromycin. The *CPSF73-AID* cell line was made by generating plasmids containing a 5′ homology arm – AID – P2A – HYG/NEO – SV40 PAS – 3′ homology arm. We have published the sequences of each of these elements (Eaton et al. 2018). These, together with px330 (Addgene 42230) containing a CPSF73 guide RNA sequence, were transfected into HCT116 cells containing dox-inducible TIR1 and colonies selected with 30 µg/mL hygromycin and 800 µg/mL neomycin. HCT116 cells containing inducible TIR1 were made by transfecting HCT116 cells with pMK243 and AAVS1 T2 CRISPR plasmid, followed by selection in 1 µg/mL Puromycin. The *RBM3* and *MORF4L2* gene insertions were made using coselection (Agudelo et al. 2017). gRNA sequences were cloned into Addgene plasmid 86611 and transfected with a plasmid containing insertion elements (see the Supplemental Material) flanked with homology arms. Positive clones were selected by growth in 0.5 µM ouabain. Insertions were confirmed by Sanger sequencing.

Antibodies

The following antibodies were used: CPSF73 (Bethyl Laboratories A301-090A), Tubulin (Abcam Ab7921), HA (Roche 3F10), Thr4p (Active Motif 6D7), EXOSC10 (Santa Cruz Biotechnology Sc-374595-X), PP1α (Bethyl Laboratories A300-904A), PP1β (Bethyl Laboratories A300-905A), and RNA Pol II (MBL Technologies CMA601 and Abcam 8WG16). Note that CMA601 was discontinued during our study and used in Figures 3A and 6D.

Total RNA isolation and qRT-PCR

RNA was isolated using tri-reagent and, following DNase treatment, 1 µg was reverse transcribed (Protoscript II, New England Biolabs) with random hexamers. cDNA was diluted to 50 µL in water and 1 µL was used per qPCR, performed with LUNA qPCR reagent (New England Biolabs) in a Qiagen Rotorgene instrument.

Chromatin-associated and nucleoplasmic RNA isolation

A six-well dish of cells was lysed in in HLB (10 mM Tris pH 7.5, 10 mM NaCl, 2.5 mM $MgCl_2$, 0.5% NP40); this was underlayered with HLB + 10% sucrose and spun at 500×*g* for 5 min. Nuclei were resuspended in 100 µL of NUN1 buffer (20 mM Tris.HCl pH 7.9, 75 mM NaCl, 0.5 mM EDTA, 50% glycerol, 0.85 mM DTT) and then topped up with 1 mL of NUN2 (20 mM HEPES pH 7.6, 1 mM DTT, 7.5 mM $MgCl_2$, 0.2 mM EDTA. 0.3 M NaCl, 1 M urea, 1% NP40). After incubation for 10 min on ice with mixing every 2–3 min, samples were spun at 13,000 rpm for 10 min. Chromatin pellets were resuspended in 500 µL of Tri-reagent and RNA was isolated with Trizol as above. Nucleoplsamic RNA was extracted from the supernatant by phenol chloroform and ethanol precipitation.

RNA sequencing

For chromatin-associated RNA-seq, cells were treated with dox/EtOH (18 h) and then with auxin/EtOH (3 h). Nuclear RNA-seq was performed following treatment with dox (18 h) and then auxin/EtOH (3 h). Five-hundred nanograms was rRNA depleted using the Ribozero kit, and libraries were prepared with the tru-seq stranded kit and sequenced on an illumina Hi-seq 2500. Data processing steps are described in the Supplemental Material.

Chromatin immunoprecipitation

A semi-confluent 100-mm dish of cells was cross-linked in 1% fomaldehyde for 10 min before quenching in 125 mM glycine.

Cells were pelleted and resuspended in 350 µL RIPA buffer (150 mM NaCl, 1% NP40, 0.5% sodium deoxycholate, 0.1% SDS, 50 mM Tris.HCl pH 8, 5 mM EDTA pH 8) before sonication in a Bioruptor for 30 sec on and 30 sec off 10 times on high. Cross-linked chromatin was divided into two: Half was incubated with 40 µL of sheep antimouse/sheep antirat Dynabeads (Life Technologies) preincubated with 3 µg of Pol II/Thr4p antibody for 2 h at 4°C, and the other half with beads incubated without antibody. Following rotation for 3 h at 4°C, beads were washed twice in RIPA buffer, three times in ChIP wash buffer (500 mM NaCl, 1% NP40, 1% sodium deoxycholate, 100 mM Tris.HCl pH 8.5), and twice in RIPA buffer (ChIP wash washes were not used for Thr4p). DNA was eluted in 0.1 M NaHCO3/1% SDS with rotation for 30 min at room temperature. Eluate was added to 30 µL of 5 M NaCl and cross-links were reversed overnight at 70°C. DNA was purified by phenol chloroform extraction and ethanol precipitation and resuspended in 50 µL of water. One microliter was used per qPCR.

Gene expression omnibus accession numbers

RNA-seq and mNET-seq of *XRN2-AID* cells (GSE109003); chromatin and nuclear RNA-seq of *CPSF73-AID* cells (GSE137727); CPSF73 RNAi chromatin RNA-seq (GSE60358).

DNA sequences

All DNA sequences are provided in the Supplemental Material.

Acknowledgments

We thank members of the laboratory for critical discussions. We are also grateful to Exeter Sequencing Service and Computational core facilities at the University of Exeter. This work was supported by a Wellcome Trust Investigator Award (WT107791/Z/15/Z) and a Lister Institute Research Fellowship held by S.W., as well as a Medical Research Council Clinical Infrastructure Award (MR/M008924/1), the Wellcome Trust Institutional Strategic Support Fund (WT097835MF), a Wellcome Trust Multi User Equipment Award (WT101650MA), and a Biotechnology and Biological Sciences Research Council Longer and Larger (LoLa) Award (BB/K003240/1).

Author contributions: J.D.E. and S.W. conceived the study. J.D.E. and L.D. analyzed the data and curated the study. J.D.E., L.F., L.D., and S.W. performed the investigation and validated the results. J.D.E. and S.W. wrote the manuscript. S.W. supervised the study and acquired the funding.

References

Agudelo D, Duringer A, Bozoyan L, Huard CC, Carter S, Loehr J, Synodinou D, Drouin M, Salsman J, Dellaire G, et al. 2017. Marker-free coselection for CRISPR-driven genome editing in human cells. *Nat Methods* **14:** 615–620. doi:10.1038/nmeth.4265

Ahn SH, Kim M, Buratowski S. 2004. Phosphorylation of serine 2 within the RNA polymerase II C-terminal domain couples transcription and 3′ end processing. *Mol Cell* **13:** 67–76. doi:10.1016/S1097-2765(03)00492-1

Baejen C, Andreani J, Torkler P, Battaglia S, Schwalb B, Lidschreiber M, Maier KC, Boltendahl A, Rus P, Esslinger S, et al. 2017. Genome-wide analysis of RNA polymerase II termination at protein-coding genes. *Mol Cell* **66:** 38–49.e6. doi:10.1016/j.molcel.2017.02.009

Calvo O, Manley JL. 2001. Evolutionarily conserved interaction between CstF-64 and PC4 links transcription, polyadenylation, and termination. *Mol Cell* **7:** 1013–1023. doi:10.1016/S1097-2765(01)00236-2

Chapman EG, Costantino DA, Rabe JL, Moon SL, Wilusz J, Nix JC, Kieft JS. 2014. The structural basis of pathogenic subgenomic flavivirus RNA (sfRNA) production. *Science* **344:** 307–310. doi:10.1126/science.1250897

Choy MS, Swingle M, D'Arcy B, Abney K, Rusin SF, Kettenbach AN, Page R, Honkanen RE, Peti W. 2017. PP1:Tautomycetin complex reveals a path toward the development of PP1-specific inhibitors. *J Am Chem Soc* **139:** 17703–17706. doi:10.1021/jacs.7b09368

Connelly S, Manley JL. 1988. A functional mRNA polyadenylation signal is required for transcription termination by RNA polymerase II. *Genes Dev* **2:** 440–452. doi:10.1101/gad.2.4.440

Cortazar MA, Sheridan RM, Erickson B, Fong N, Glover-Cutter K, Brannan K, Bentley DL. 2019. Control of RNA Pol II speed by PNUTS-PP1 and Spt5 dephosphorylation facilitates termination by a 'sitting duck torpedo' mechanism. *Mol Cell* doi:10.1016/j.molcel.2019.09.031

Davidson L, Francis L, Cordiner RA, Eaton JD, Estell C, Macias S, Cáceres JF, West S. 2019. Rapid Depletion of DIS3, EXOSC10, or XRN2 reveals the immediate impact of exoribonucleolysis on nuclear RNA metabolism and transcriptional control. *Cell Rep* **26:** 2779–2791.e5. doi:10.1016/j.celrep.2019.02.012

Eaton JD, West S. 2018. An end in sight? Xrn2 and transcriptional termination by RNA polymerase II. *Transcription* **9:** 321–326.

Eaton JD, Davidson L, Bauer DLV, Natsume T, Kanemaki MT, West S. 2018. Xrn2 accelerates termination by RNA polymerase II, which is underpinned by CPSF73 activity. *Genes Dev* **32:** 127–139. doi:10.1101/gad.308528.117

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

Fong N, Öhman M, Bentley DL. 2009. Fast ribozyme cleavage releases transcripts from RNA polymerase II and aborts co-transcriptional pre-mRNA processing. *Nat Struct Mol Biol* **16:** 916–922. doi:10.1038/nsmb.1652

Fong N, Brannan K, Erickson B, Kim H, Cortazar MA, Sheridan RM, Nguyen T, Karp S, Bentley DL. 2015. Effects of transcription elongation rate and Xrn2 exonuclease activity on RNA polymerase II termination suggest widespread kinetic competition. *Mol Cell* **60:** 256–267. doi:10.1016/j.molcel.2015.09.026

Fuchs G, Voichek Y, Benjamin S, Gilad S, Amit I, Oren M. 2014. 4sUDRB-seq: measuring genomewide transcriptional elongation rates and initiation frequencies within cells. *Genome Biol* **15:** R69. doi:10.1186/gb-2014-15-5-r69

Gregersen LH, Mitter R, Ugalde AP, Nojima T, Proudfoot NJ, Agami R, Stewart A, Svejstrup JQ. 2019. SCAF4 and SCAF8, mRNA anti-terminator proteins. *Cell* **177:** 1797–1813.e18. doi:10.1016/j.cell.2019.04.038

Gromak N, West S, Proudfoot NJ. 2006. Pause sites promote transcriptional termination of mammalian RNA polymerase II. *Mol Cell Biol* **26:** 3986–3996. doi:10.1128/MCB.26.10.3986-3996.2006

Heinz S, Texari L, Hayes MGB, Urbanowski M, Chang MW, Givarkes N, Rialdi A, White KM, Albrecht RA, Pache L, et al. 2018. Transcription elongation can affect genome 3D structure. *Cell* **174:** 1522–1536.e22. doi:10.1016/j.cell.2018.07.047

Heroes E, Lesage B, Görnemann J, Beullens M, Van Meervelt L, Bollen M. 2013. The PP1 binding code: a molecular-lego strategy that governs specificity. *FEBS J* **280:** 584–595. doi:10.1111/j.1742-4658.2012.08547.x

Hoek TA, Khuperkar D, Lindeboom RGH, Sonneveld S, Verhagen BMP, Boersma S, Vermeulen M, Tanenbaum ME. 2019. Single-molecule imaging uncovers rules governing nonsense-mediated mRNA decay. *Mol Cell* **75:** 324–339.e11. doi:10.1016/j.molcel.2019.05.008

Hori S, Yamamoto T, Obika S. 2015. XRN2 is required for the degradation of target RNAs by RNase H1-dependent antisense oligonucleotides. *Biochem Biophys Res Commun* **464:** 506–511. doi:10.1016/j.bbrc.2015.06.171

Horvathova I, Voigt F, Kotrys AV, Zhan Y, Artus-Revel CG, Eglinger J, Stadler MB, Giorgetti L, Chao JA. 2017. The dynamics of mRNA turnover revealed by single-molecule imaging in single cells. *Mol Cell* **68:** 615–625.e9. doi:10.1016/j.molcel.2017.09.030

Kaida D, Berg MG, Younis I, Kasim M, Singh LN, Wan L, Dreyfuss G. 2010. U1 snRNP protects pre-mRNAs from premature cleavage and polyadenylation. *Nature* **468:** 664–668. doi:10.1038/nature09479

Kaneko S, Rozenblatt-Rosen O, Meyerson M, Manley JL. 2007. The multifunctional protein p54nrb/PSF recruits the exonuclease XRN2 to facilitate pre-mRNA 3′ processing and transcription termination. *Genes Dev* **21:** 1779–1789. doi:10.1101/gad.1565207

Kecman T, Kuś K, Heo DH, Duckett K, Birot A, Liberatori S, Mohammed S, Geis-Asteggiante L, Robinson CV, Vasiljeva L. 2018. Elongation/termination factor exchange mediated by PP1 phosphatase orchestrates transcription termination. *Cell Rep* **25:** 259–269.e5. doi:10.1016/j.celrep.2018.09.007

Kim M, Ahn SH, Krogan NJ, Greenblatt JF, Buratowski S. 2004a. Transitions in RNA polymerase II elongation complexes at the 3′ ends of genes. *EMBO J* **23:** 354–364. doi:10.1038/sj.emboj.7600053

Kim M, Krogan NJ, Vasiljeva L, Rando OJ, Nedea E, Greenblatt JF, Buratowski S. 2004b. The yeast Rat1 exonuclease promotes transcription termination by RNA polymerase II. *Nature* **432:** 517–522. doi:10.1038/nature03041

Logan J, Falck-Pedersen E, Darnell JE Jr, Shenk T. 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 β maj-globin gene. *Proc Natl Acad Sci* **84:** 8306–8310. doi:10.1073/pnas.84.23.8306

Luo W, Johnson AW, Bentley DL. 2006. The role of Rat1 in coupling mRNA 3′-end processing to transcription termination: implications for a unified allosteric-torpedo model. *Genes Dev* **20:** 954–965. doi:10.1101/gad.1409106

Mandel CR, Kaneko S, Zhang H, Gebauer D, Vethantham V, Manley JL, Tong L. 2006. Polyadenylation factor CPSF-73 is the pre-mRNA 3′-end-processing endonuclease. *Nature* **444:** 953–956. doi:10.1038/nature05363

Miki TS, Carl SH, Großhans H. 2017. Two distinct transcription termination modes dictated by promoters. *Genes Dev* **31:** 1870–1879. doi:10.1101/gad.301093.117

Muniz L, Davidson L, West S. 2015. Poly(A) polymerase and the nuclear poly(A) binding protein, PABPN1, coordinate the splicing and degradation of a subset of human pre-mRNAs. *Mol Cell Biol* **35:** 2218–2230. doi:10.1128/MCB.00123-15

Natsume T, Kiyomitsu T, Saga Y, Kanemaki MT. 2016. Rapid protein depletion in human cells by auxin-inducible degron tagging with short homology donors. *Cell Rep* **15:** 210–218. doi:10.1016/j.celrep.2016.03.001

Park J, Kang M, Kim M. 2015. Unraveling the mechanistic features of RNA polymerase II termination by the 5′–3′ exoribonuclease Rat1. *Nucleic Acids Res* **43:** 2625–2637. doi:10.1093/nar/gkv133

Parua PK, Booth GT, Sansó M, Benjamin B, Tanny JC, Lis JT, Fisher RP. 2018. A Cdk9–PP1 switch regulates the elongation-termination transition of RNA polymerase II. *Nature* **558:** 460–464. doi:10.1038/s41586-018-0214-z

Preker P, Nielsen J, Kammler S, Lykke-Andersen S, Christensen MS, Mapendano CK, Schierup MH, Jensen TH. 2008. RNA exosome depletion reveals transcription upstream of active human promoters. *Science* **322:** 1851–1854. doi:10.1126/science.1164096

Proudfoot NJ. 1989. How RNA polymerase II terminates transcription in higher eukaryotes. *Trends Biochem Sci* **14:** 105–110. doi:10.1016/0968-0004(89)90132-1

Proudfoot NJ. 2011. Ending the message: poly(A) signals then and now. *Genes Dev* **25:** 1770–1782. doi:10.1101/gad.17268411

Schlackow M, Nojima T, Gomes T, Dhir A, Carmo-Fonseca M, Proudfoot NJ. 2017. Distinctive patterns of transcription and RNA processing for human lincRNAs. *Mol Cell* **65:** 25–38. doi:10.1016/j.molcel.2016.11.029

Schreieck A, Easter AD, Etzold S, Wiederhold K, Lidschreiber M, Cramer P, Passmore LA. 2014. RNA polymerase II termination involves C-terminal-domain tyrosine dephosphorylation by CPF subunit Glc7. *Nat Struct Mol Biol* **21:** 175–179. doi:10.1038/nsmb.2753

Shi Y, Di Giammartino DC, Taylor D, Sarkeshik A, Rice WJ, Yates JR 3rd, Frank J, Manley JL. 2009. Molecular architecture of the human pre-mRNA 3′ processing complex. *Mol Cell* **33:** 365–376. doi:10.1016/j.molcel.2008.12.028

Singh J, Padgett RA. 2009. Rates of in situ transcription and splicing in large human genes. *Nat Struct Mol Biol* **16:** 1128–1133. doi:10.1038/nsmb.1666

Stevens A, Maupin MK. 1987. A 5′–3′ exoribonuclease of human placental nuclei: purification and substrate specificity. *Nucleic Acids Res* **15:** 695–708. doi:10.1093/nar/15.2.695

Vilborg A, Passarelli MC, Yario TA, Tycowski KT, Steitz JA. 2015. Widespread inducible transcription downstream of human genes. *Mol Cell* **59:** 449–461. doi:10.1016/j.molcel.2015.06.016

West S, Gromak N, Proudfoot NJ. 2004. Human 5′ → 3′ exonuclease Xrn2 promotes transcription termination at co-transcriptional cleavage sites. *Nature* **432:** 522–525. doi:10.1038/nature03035

Wilusz JE, Freier SM, Spector DL. 2008. 3′ end processing of a long nuclear-retained noncoding RNA yields a tRNA-like cytoplasmic RNA. *Cell* **135:** 919–932. doi:10.1016/j.cell.2008.10.012

Zhang Z, Gilmour DS. 2006. Pcf11 is a termination factor in *Drosophila* that dismantles the elongation complex by bridging the CTD of RNA polymerase II to the nascent transcript. *Mol Cell* **21:** 65–74. doi:10.1016/j.molcel.2005.11.002

Zhang H, Rigo F, Martinson HG. 2015. 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:** 437–448. doi:10.1016/j.molcel.2015.06.008

A unified allosteric/torpedo mechanism for transcriptional termination on human protein-coding genes

Joshua D. Eaton, Laura Francis, Lee Davidson, et al.

Genes Dev. 2020, **34:** originally published online December 5, 2019
Access the most recent version at doi:10.1101/gad.332833.119

iv. Termination of Transcription by RNA Polymerase II: BOOM!

Eaton, J.D., and West, S. (2020). Termination of Transcription by RNA Polymerase II: BOOM! Trends in Genetics 36, 664–675.

Trends in Genetics

CellPress REVIEWS

Review

Termination of Transcription by RNA Polymerase II: BOOM!

Joshua D. Eaton[1] and Steven West[1,*]

RNA polymerase II (Pol II) transcribes hundreds of thousands of transcription units – a reaction always brought to a close by its termination. Because Pol II transcribes multiple gene types, its termination occurs in a variety of ways, with the polymerase being responsive to different inputs. Moreover, it is not just a default process occurring at the end of genes. Promoter-proximal and premature termination is common and might in turn regulate gene expression levels. Although some transcription termination mechanisms have been debated for decades, research is only just underway on emergent processes. We provide an updated view of transcription termination in human cells, highlighting common themes and some interesting differences between the contexts in which it occurs.

Highlights

PP1-induced Pol II pausing and XRN2 exoribonuclease activity constitute the major mechanism of poly(A) signal-dependent termination.

Antisense oligonucleotides and other cotranscriptional RNA cleavage events promote XRN2-dependent termination of Pol II.

Diverse termination mechanisms operate across different scenarios.

Promoter-proximal Pol II is often turned over rapidly by termination.

Stopping Pol II at the Beginning, Middle, and End of Genes

RNA polymerase II (Pol II) is a multisubunit complex of proteins that transcribes many gene classes (see Glossary). It is distinguished from other RNA polymerases by the carboxyl-terminal domain (CTD) of its largest subunit. This consists of repeated heptad amino acid motifs (most commonly YSPTSPS) that are modified, mostly by phosphorylation, to link transcription with RNA maturation [1]. **Transcription termination** brings an end to every Pol II transcription cycle, which can be after tens of nucleotides or over a million. The most studied termination process is at the end of protein-coding genes and depends on a **polyadenylation signal** (PAS), which directs **cleavage and polyadenylation** (CPA) of pre-mRNA [2,3]. However, recent evidence reveals that Pol II is vulnerable to termination as soon as transcription begins [4,5]. Indeed, **premature termination** can occur almost anywhere on a gene (Figure 1) [6–10]. Although PAS-dependent termination is the best-characterized mechanism, a diverse range of processes exist in different contexts or on different gene classes [11–13]. More broadly, inefficient termination might account for a significant portion of intergenic transcription and aid our understanding of these enigmatic RNAs [14]. Moreover, because termination is affected in certain genetic diseases, its correction could be therapeutically advantageous [15–17]. By understanding the mechanisms of transcriptional termination we can establish its regulatory functions and take advantage of opportunities to control it.

PAS-Dependent Termination

Most protein-coding genes possess a PAS, which defines the mRNA 3′ end and directs Pol II termination. A PAS includes an AAUAAA motif often with upstream U-rich and downstream U/GU-rich sequences [3]. These elements are recognized by a CPA complex that is assembled from cleavage and polyadenylation specificity factor (CPSF), cleavage stimulatory factor and cleavage factors I and IIm subcomplexes. The association of these factors with the PAS leads to cleavage of the pre-mRNA (between the AAUAAA and downstream element) followed by polyadenylation of the upstream mRNA. During CPA, AAUAAA is recognized by the CPSF30 and WDR33 components of CPSF and PAS cleavage is accomplished by the CPSF73 endonuclease [18–21,107]. PAS mutations also abolish transcriptional termination – an observation that is a foundation for understanding its mechanism [17]. Two different models were envisaged to

[1]The Living Systems Institute, University of Exeter, Exeter EX4 4QD, UK

*Correspondence: s.west@exeter.ac.uk (S. West).

Trends in Genetics

CellPress
REVIEWS

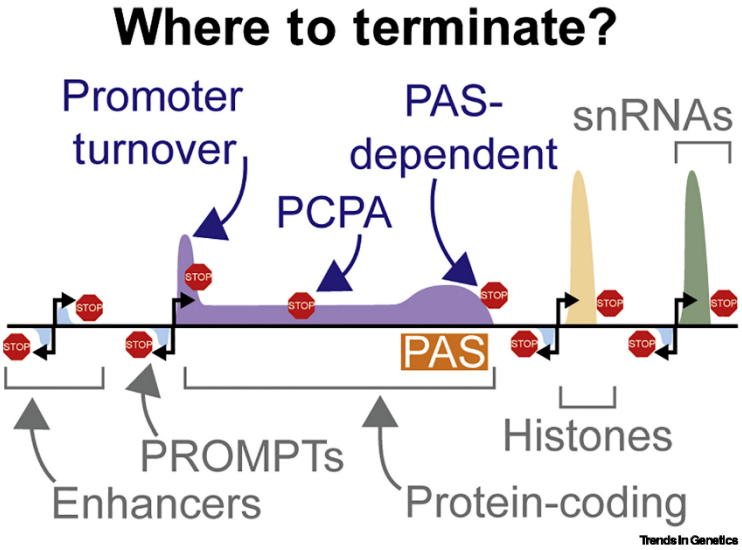


Figure 1. Termination Occurs at All Points in the Transcription Cycle. Summary of where RNA polymerase II termination can occur (signified by stop signs). This can be at the end of known transcription units which are labelled in gray or at upstream positions (such as promoter turnover and PCPA) labelled in blue over the protein-coding gene. Polymerase occupancies at the various gene classes are depicted as filled areas above (sense) or below (anti-sense) the central black line. These approximate to what has been observed experimentally. Abbreviations: PAS, polyadenylation signal; PCPA, premature cleavage and polyadenylation; PROMPT, promoter upstream transcript; snRNA, small nuclear RNA.

explain PAS-dependent termination. These are referred to as the **allosteric/antiterminator** and **torpedo** models and their relevance and contribution to the mechanism will be introduced in subsequent paragraphs [22–24].

Allosteric/Antiterminator Model

The allosteric/antiterminator model proposes that transcription of a PAS causes the dissociation of factors from or a conformational change within Pol II that promotes its termination (Figure 2A, Key Figure). Antitermination factors are predicted to suppress termination until the desired PAS has been transcribed. Two examples are SCAF4 and SCAF8, both of which interact with Pol II. When these two factors are knocked out, premature CPA (PCPA) happens at PAS sequences within the gene body [25]. PCPA is observed under a variety of other conditions including the depletion of U1 **small nuclear RNA** (snRNA), cyclin-dependent kinase (CDK)12 or nuclear poly(A) binding protein (PABPN1) [26–28]. Premature termination of many transcripts is also affected by the CPA factor, PCF11, the levels of which influence early PAS usage [29]. The antiterminator model envisages the dissociation of proteins that prevent termination. However, some CPA factors are recruited to Pol II after a PAS and many are important for termination – although not always directly (more in following text) [30].

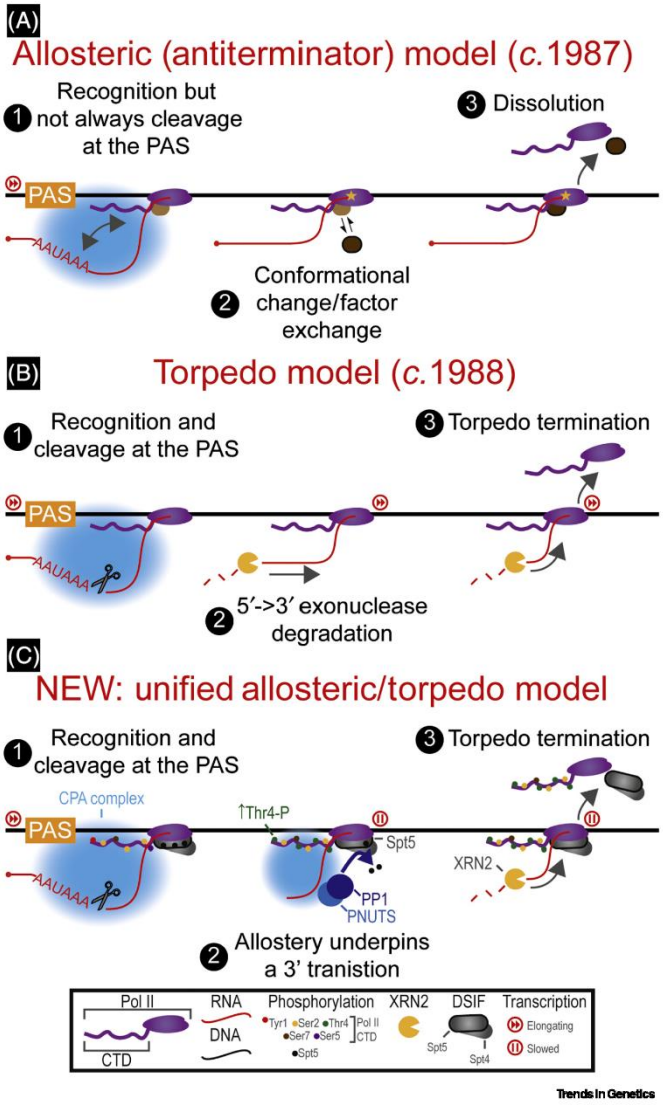
The allosteric model also allows for Pol II conformational changes that drive termination. PAS transcription induces Pol II conformational changes *in vitro* that promote some termination even without transcript cleavage [31]. Cotranscriptional PAS cleavage is also rarely seen when transcription

Glossary

Allosteric/antiterminator model: this proposes that transcription termination is caused by conformational changes within or proteins recruited to/dissociating from, Pol II.
Antisense oligonucleotides: chemically modified nucleic acid sequences that are extremely stable. When bound to an RNA target, the resulting hybrid is recognized and cut by RNAseH1.
Auxin-inducible degron: a protein tag that can be appended to a protein of interest and induce its rapid degradation in the presence of the plant hormone auxin.
Cleavage and polyadenylation (CPA): the process of mRNA 3′-end formation comprising PAS cleavage by CPSF73 and polyadenylation of the upstream cleavage product.
Exosome: a complex of proteins that catalyzes RNA degradation in a 3′→5′ direction.
Polyadenylation signal: a hexamer sequence, AAUAAA, followed by a U/GU rich element that directs cleavage and polyadenylation (CPA) at the 3′ end of most protein-coding RNAs.
Premature termination: transcriptional termination occurring before the annotated end of a gene.
Read through: transcription beyond the normal site of termination.
RNA polymerase II: a multisubunit enzyme that transcribes various RNAs in a DNA-dependent reaction.
Small nuclear RNA: short noncoding RNA crucial for removing noncoding introns from pre-mRNA by a process called splicing.
Torpedo model: this proposes that termination is caused by 5′→3′ degradation of the Pol-II-associated RNA product of PAS cleavage.
Transcription termination: cessation of RNA polymerase transcription and dissolution of the protein, RNA and DNA ternary complex.
XRN2: a nuclear enzyme that degrades RNA in a 5′→3′ direction and requires a 5′ phosphate on its substrates.

Key Figure

A Unified Model for PAS-Dependent Pol II Termination



(See figure legend at the bottom of the next page.)

is imaged with electron microscopy, which has been used as support for an allosteric model [32]. However, rapid 5′→3′ degradation of the downstream cleavage product could account for this result. Termination before PAS cleavage is probably uncommon in cells and might conflict with RNA quality control. This is suggested by findings that PAS mutation leads to the retention of transcripts at the gene locus [33]. There is nevertheless an evolutionary precedent for cleavage-independent termination of some RNA polymerases, especially for those where termination defines the RNA 3′ end. For example, eukaryotic Pol III terminates over T runs in the non-template DNA strand, as does prokaryotic RNA polymerase (when these are preceded by transcribed RNA hairpins) [34,35].

Torpedo Model

In the torpedo model, a 5′→3′ exonuclease degrades the Pol-II-associated product of PAS cleavage and promotes termination on reaching the polymerase (Figure 2B). This explanation requires PAS-cleavage (or another RNA cleavage) to precede Pol II release and is supported by observations that CPA factors are generally needed for termination [36–38]. Human **XRN2** and budding yeast Rat1 are nuclear 5′→3′ exonucleases and their depletion leads to a termination defect, identifying them as possible molecular torpedoes [39,40]. Although Rat1 was shown to promote termination at most yeast protein-coding genes, XRN2 depletion initially had little impact other than on reporter plasmids [38,41,42]. With hindsight, this was due to its incomplete loss because overexpressing inactive XRN2, following RNAi depletion of the endogenous protein, induces a general termination defect [13]. Rapid elimination of XRN2 using an **auxin-inducible degron** (AID) also generally inhibits PAS-dependent termination without having to express the inactive protein and underlines its specific role in the process (Box 1) [12]. Mutating the XRN2 active site or blocking its progress along the RNA delays termination and argues that the pursuit of Pol II by XRN2 is important [12,37]. Consistently, when the speed of Pol II is altered by mutations, faster Pol II evades termination more effectively than a slower derivative [13].

Cotranscriptional degradation of PAS cleavage products in budding yeast can continue in the absence of Rat1 (via the related Xrn1 5′→3′ exonuclease) without leading to termination [43]. However, XRN2 depletion is sufficient to stabilize nascent RNAs downstream of the PAS in human cells [12]. Moreover, indiscriminately blocking other 5′→3′ exonucleases, using Xrn-resistant RNAs (xrRNAs) [44], produces a near-equivalent effect to XRN2 depletion, arguing against relevant alternative activities [37].

Allosteric Switches in a Unified Model

Recent experiments provide strong evidence for a termination mechanism that unifies both original models. This is initiated by an allosteric switch that decelerates Pol II beyond the PAS – a step that was first uncovered in fission yeast where the protein phosphatase 1 (PP1) enzyme, Dis2, dephosphorylates the Spt5 elongation factor [45,46]. Spt5 is initially dephosphorylated at promoters where its phosphorylation by CDK9 stimulates efficient Pol II elongation. Dis2-catalyzed dephosphorylation of Spt5 occurs after the PAS and is predicted

Figure 2. (A) Allosteric/anti-terminator model proposes that a conformational change or factors recruited to or dissociated from Pol II render it termination competent. (B) In the torpedo model, PAS cleavage generates a Pol-II-associated RNA that is degraded 5′→3′ leading to termination. (C) A combined allosteric/torpedo model, PAS cleavage (or late-stage CPA complex assembly) promotes Pol II slowing which is caused by dephosphorylation of SPT5 by PNUTS/PP1 and constitutes an allosteric switch. PAS-cleavage is also associated with increased threonine 4 phosphorylation on Pol II CTD. This switch renders Pol II stranded on the template and easily terminated by XRN2, which degrades the polymerase-associated product of PAS cleavage. Abbreviations: CPA, cleavage and polyadenylation; CTD, C-terminal domain; DSIF, DRB sensitivity-inducing factor; PAS, polyadenylation signal; PNUTS, PP1 nuclear targeting factor; PP1, protein phosphatase 1; Pol II, RNA polymerase II; XRN2, nuclear enzyme that degrades RNA in a 5′→3′ direction.

Box 1. Termination Defects Using Different Experimental Systems

RNAi-mediated protein depletion is frequently used to study the function of termination factors but genetically implemented protein degron approaches are now increasingly used (e.g. AID, [100]). These offer faster (1–3 h vs 2–3 days) and typically more complete depletion, enriching direct effects and limiting the influence of redundant mechanisms. This matters because the elimination of XRN2–AID identified XRN2 as a general termination factor whereas RNAi alone did not [12,38]. Similarly, RNAi of CPSF73 induces a partial termination defect at protein-coding genes that is not always apparent at noncoding loci [101]. A partial termination defect can be explained by the existence of alternative mechanisms that are independent of the protein that was depleted. However, a similar result will be obtained if the canonical pathway is delayed because of reduced levels of an important component. The latter should be considered because the acute loss of CPSF73–AID inhibits termination more profoundly than its depletion by RNAi, inducing a massive defect at all protein-coding and many noncoding loci [37]. In summary, mild termination defects seen with RNAi give important clues about protein function but rapid depletion techniques can provide a more complete picture (see Figure I and [12,37,38,101]).

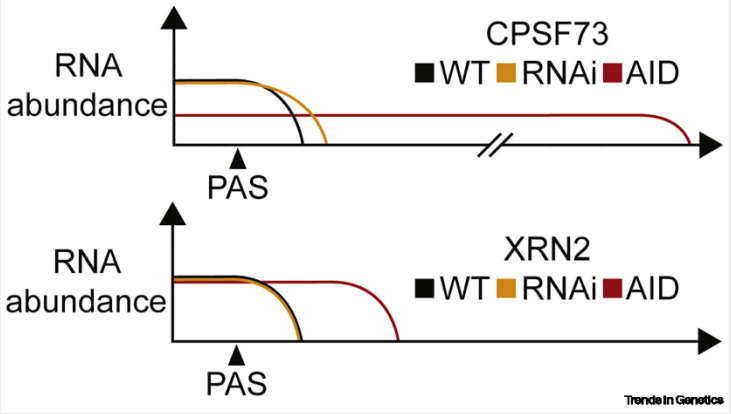


Figure I. Schematic of RNAi versus AID Depletion Effects for CPSF73 (A) or XRN2 (B). RNAi of CPSF73 induces a mild read-through effect (yellow) compared to CPSF73-AID depletion (red). Note that acute CPSF73-AID loss also reduces gene body RNA signals. RNAi of XRN2 by itself does not cause general read-through but rapid XRN2-AID depletion does. Loss of CPSF73-AID causes longer read-through than XRN2-AID elimination (see main review for explanation).

to reduce Pol II speed [45]. SPT5 phosphorylation by CDK9 is also required for efficient elongation in human cells and its dephosphorylation by PP1 occurs downstream of the PAS [47,48]. Human PP1 enzyme and its nuclear targeting factor, PNUTS, are present in CPA complexes providing a connection between 3′-end processing and this termination mechanism [49]. SPT5 is not the only substrate of PP1 relevant to termination as PP1 targets phosphorylated tyrosine 1 within the CTD of budding yeast Pol II, which coordinates the recruitment of CPA factors [50]. Mutations of most tyrosine residues in the human Pol II CTD also cause termination defects [51]. Finally, a list of human PP1 substrates includes p54/nrb, which is known to aid the recruitment of XRN2 to genes [52,53].

A number of other mechanisms contribute to Pol II elongation capacity beyond the PAS, which might be aided by PP1-catalyzed events or be independent processes. Pol II termination sites are associated with polymerase pausing and nucleotide sequences conducive to backtracking [54,55]. RNA can also hybridize with single-stranded DNA in the wake of Pol II and these so-called R loops slow transcription by generating torsional stress [56]. R loops are most prevalent at promoters and terminators, where SPT5 is dephosphorylated and termination commonly

Trends in Genetics

CellPress
REVIEWS

occurs [57,58]. Moreover, they are resolved by RNA:DNA helicases, among which SETX has been proposed to facilitate XRN2-dependent termination [59].

Threonine 4 phosphorylation of the Pol II CTD might signify imminent termination or allosteric changes. It is prominent beyond the PAS and depends on CPSF73 [37,38,46]. PAS cleavage possibly transmits the allosteric termination signal because catalytically inactive CPSF73 is associated with read-through comparable with CPSF73 loss [12]. However, some contacts between CPSF73 and other CPA factors might be important as these are also weakened by inactivating CPSF73 [60]. Reduced CDK9 activity could serve to enhance PP1 impact after the PAS as well [45]. Whatever their constellation, the effect of PAS-dependent allosteric switches described previously is dramatic: Pol II elongation rate drops from ~2 kb/min upstream of the PAS to as little as 0.1 kb/min afterwards [47].

Torpedo Termination in a Unified Model

In the torpedo model, XRN2 needs to outpace Pol II. Although its degradation speed is unknown, the related cytoplasmic exonuclease, XRN1, achieves ~2.3 kb/min [61]. The considerable slowing of Pol II after the PAS is well beneath this estimate and renders it a 'sitting duck' for the pursuing XRN2 torpedo [47]. This deceleration of Pol II after PAS cleavage also explains the different effects of XRN2 versus CPA factor loss on termination [12,37]. CPA factor depletion generally induces longer **read through** than XRN2 loss, which was interpreted to mean an underpinning role for the former with XRN2 part of a supporting act. We now know that loss of the PAS endonuclease, CPSF73, induces read through of hundreds of kilobases because the allosteric switch has not been activated. Because XRN2 acts after this PP1-catalyzed event, its depletion causes Pol II to pile up at terminator regions instead of continuing to rapidly transcribe. It is this, rather than alternative pathways, that might explain the lack of longer read through following XRN2 loss.

The mechanism derived from the data described above is consistent with kinetic competition between XRN2 and the transcribing polymerase [13], with a potential modification. As PAS activity virtually stops Pol II in its tracks, the effect of transcription speed on the position of termination might reflect where Pol II reaches before allosteric modifications come into effect (i.e., the point at which it is rendered a sitting duck) rather than the distance required for XRN2 to chase it down. Our understanding of PAS-dependent termination is therefore as follows: PAS cleavage promotes dephosphorylation of SPT5, and possibly other factors, by PP1/PNUTS. This slows Pol II down and facilitates its termination by XRN2 (Figure 2C).

Termination Activity of XRN2

There is good evidence that XRN2 possesses a general Pol II termination activity that is not just confined to the PAS-dependent mechanism. First, it promotes termination following the activities of other endonucleases, including RNase P/Z and microprocessor [12,13]. Second, XRN2-dependent termination is induced by **antisense oligonucleotides** (ASOs) that direct transcript cleavage by RNaseH1 [15,16,37]. This activity is not restricted to gene 3′ ends and can even be used to promote termination in the absence of CPSF73 [37]. These findings imply that PAS-dependent allosteric modifications are not strictly required for XRN2 to terminate Pol II. However, slowing down Pol II is clearly advantageous, especially where it has proceeded far beyond any upstream XRN2 entry site. Additionally, effects of RNA cleavage or XRN2 activity on Pol II behavior cannot be excluded considering that shortening nascent RNA reduces its elongation capacity *in vitro* [62]. An interesting implication of ASO-promoted termination outside of a PAS context is that natural RNaseH1 substrates (R loops) might terminate transcription in a variety of contexts. The ability of ASOs to promote termination and the possibility of controlling this provides some important experimental considerations and therapeutic opportunities (Box 2).

Box 2. Termination on Demand

The discovery that ASOs can direct XRN2-dependent termination provides several experimental possibilities and opportunities [15,16,37]. Some genetic diseases are characterized by mutations encompassing the PAS that result in transcriptional read through, which could interfere with and reduce the expression of downstream genes [17]. By promoting termination at the affected upstream gene ASOs might have the potential to ameliorate these interference effects and perhaps reduce pathology. ASO-driven premature termination could also be used to selectively prevent pathological RNA isoforms from being produced. The finding that ASOs promote transcriptional termination is additionally important for studying nuclear noncoding RNAs. ASOs have long been used to separate the function of noncoding transcription from that of the noncoding transcript on the assumption that they only target the latter. The realization that they promote termination means that judicious ASO design is required to achieve separation of function. Specifically, ASOs directed toward the 3′ ends of noncoding RNAs have the greatest potential to selectively target the transcript without affecting transcription of the locus [15,16].

It is worth speculating about how XRN2 terminates Pol II, which could entail a specific interaction between the two. XRN2 and Rat1 terminate Pol II *in vitro* but not *Escherichia coli* RNA polymerase [63]. Similarly, prokaryotic RNA polymerase is released more efficiently by the 5′→3′ exonuclease activity of bacterial RNaseJ than it is by Xrn1 (which can terminate Pol II *in vitro*) [64]. An interaction between Pol II and XRN2 requires degradation to continue until the polymerase active center is reached, which is supported by experiments discussed previously. Polymerases that are terminated by XRN2 beyond a PAS have diminished elongation capacity and enhanced propensity to stall at motifs predicted to destabilize the RNA:DNA hybrid in its active center [54]. Analogously, dA:rU hybrids are associated with the termination of prokaryotic polymerases that become trapped over these regions by a stem loop formed behind the transcription complex [65]. Such hybrids have particularly low thermodynamic stability (compared with other combinations), which can assist in pulling nascent RNA from the RNA:DNA duplex upon upstream stem-loop formation [66–68]. It is conceivable that XRN2-captured Pol II stalled over similar motifs might share some features with this bacterial mechanism but this remains to be tested.

Are There Other PAS-Dependent Mechanisms?

Given the importance of stopping transcription, failsafe pathways would be sensible though might only apply following PAS cleavage because CPSF73 depletion causes such substantial readthrough. Failsafe termination after PAS cleavage exists in budding yeast where the Nrd1 RNA-binding protein terminates Pol II following Rat1 inactivation [69]. XRN2-independent mechanisms were also hypothesized based on the relatively minor termination defect associated with its absence; however, PP1-induced Pol II pausing under these circumstances probably underestimates the impact of XRN2 loss [37,47]. 3′ ends of RNAs released by XRN2-independent termination should be degraded 3′→5′ by the **exosome** but this is not often observed and 3′-flanking RNAs from PAS-containing genes are not prominent exosome substrates [8,37]. Nevertheless, the products of XRN2-independent termination might be degraded another way or acquire a stabilizing feature like a poly(A) tail. When transcription is inhibited in XRN2-depleted cells, polymerases do eventually terminate but at a slower rate than control cells [47]. If this is not due to trace levels of XRN2, then this approach might identify any XRN2-independent mechanisms. Finally, Pol II synthesizes unstable promoter upstream transcripts (PROMPTs [70], also see later section). These can use a form of PAS-dependent processing, but they are not all terminated by XRN2 suggesting alternative mechanisms [8,71].

Termination at Replication-Dependent Histone and snRNA Genes

Replication-dependent histone (RDH) transcripts do not utilize a PAS. However, their 3′ ends are still cleaved by CPSF73, which is instead recruited by U7 snRNA as part of a histone cleavage complex (HCC) [72,73]. Manipulation of RDH 3′-end processing elements causes a termination defect [74], which predicts a similar mechanism to PAS-containing genes. When inactive XRN2

is expressed following depletion of the endogenous protein, RDH termination is mildly affected; however, there is little effect when XRN2–AID is rapidly eliminated [12,13]. These contrasting results could be explained by the different approaches: coexpression of the inactive enzyme might prevent other 5′→3′ exonucleases from accessing exposed 5′ ends whereas the role of XRN2 is more specifically tested by its AID depletion. Therefore, RDH termination might use a different or redundant 5′→3′ exonuclease and CPSF73 has been proposed as such [75,76]. Although CPSF73 exonuclease activity is not demonstrated directly, it is a similar protein to bacterial RNaseJ, which has dual endo- and exonuclease activities [77]. Degradation of the RDH 3′ cleavage product *in vitro* can proceed from a 5′OH (as well as the natural 5′ phosphate) and in the presence of high chelator concentrations, while XRN2 can only degrade 5′ phosphate and has a divalent cation in its active center [78,79]. Furthermore, archaeal RNA polymerase is terminated by a homolog of CPSF73, called FttA, which has dual endo- and exonuclease activities [80]. In any case, it is still not clear whether 5′→3′ degradation of RDH 3′ cleavage products is absolutely necessary for termination. As such, allosteric termination (or efficient failsafe mechanisms) of RDH transcription should be considered as a possibility.

snRNAs are 3′-end processed by the multisubunit integrator complex, which contains a paralog of the CPSF73 endonuclease called INTS11 [81]. Integrator binds Pol II, processes snRNAs via a 3′ box RNA element and its depletion causes snRNA termination defects [11,82]. The catalytic activity of INTS11 is implicated in its termination functions [6,7,83], but XRN2 effects are very modest or absent at snRNA genes [12,13]. It could be that transcript cleavage promotes termination without needing subsequent 5′→3′ activities or that Pol-II-associated products are too short to be detected. Still, a lack of XRN2 effect is also compatible with allosteric termination of snRNA transcription. Products released by allosteric (or cleavage independent) termination are predicted to be degraded by the exosome as alluded to previously. While 3′ flanking transcripts from protein-coding genes are not generally exosome substrates, at least some snRNA read-through products are bound by its DIS3 catalytic component [84]. If, as seems reasonable, these substrates are liberated by transcriptional termination then snRNA genes may sometimes use a mechanism distinct from protein-coding genes that does not require 5′→3′ degradation and better resembles an allosteric process. This could operate alongside or together with the established integrator-dependent mechanism. Notably, several other proteins are implicated in snRNA termination including negative elongation factor (NELF), the CPA complex members PCF11 and SSU72, and the cap binding complex-associating protein, ARS2 [11,85,86]. Among these, PCF11 can terminate Pol II without the need for any 5′→3′ exonuclease activity and could conceivably function this way at snRNA genes [87]. There is also evidence that RNA sequences, additional to the 3′ box, play a role in termination of at least some snRNAs [88].

Promoter-Proximal Termination

Most human promoters are loaded with Pol II, presumably as a result of elongation complexes pausing soon after transcription is initiated [89]. These polymerases might remain at the promoter until they receive a signal to resume transcription. However, by using new techniques that include live-cell imaging, single-molecule foot-printing, and chemical treatments, recent studies suggest that many promoter-proximal polymerases are rapidly turned over [4,5,90,91]. Mechanisms of promoter-proximal termination await full elucidation, but integrator depletion inhibits this process on many genes suggesting a prominent role in Pol II turnover early in transcription [6,7,83,92]. The nuclease activity of INTS11 is required for promoter-proximal termination and cleavage sites are found close to transcript 5′ ends. However, XRN2 is not yet implicated in this pathway showing some parallels with snRNA transcription [7]. Integrator is also involved in the termination of some PROMPTs and enhancer RNAs, which are both short exosome sensitive transcript classes [93,94]. However, its depletion is not associated with runaway read through in either case and

XRN2 effects are again not obvious [8,93,94]. Like snRNA maturation, the sensitivity of transcriptional termination to integrator depends on promoter identity and its specificity to short transcripts is consistent with observations that snRNA processing signals are most efficient when positioned close to the transcription start site [6,95,96].

Some promoter-proximal termination might use similar machinery to the PAS-dependent process. XRN2 degrades decapped transcripts and those that are inefficiently processed [42,97]. PAS activity is observed on promoter-proximal transcripts and CPA factors directly bind some of these [38,71,98]. Interestingly the promoter effects of CPSF and XRN2 are more prominent for Pol II possessing serine 2 phosphorylated (S2p) CTD [38]. As S2p is associated with efficient elongation this pathway could apply to a subpopulation of promoter-proximal Pol II that has acquired the capacity to use a PAS-dependent process. Overall, the mechanism of promoter-proximal and PROMPT termination is not well defined, perhaps because Pol II at these regions is unstable and vulnerable to many destabilizing events (Box 3). Consequently, only a fraction of polymerases reach the 3′ end of a gene as a result of common premature termination. We have described our new understanding of how they are terminated, especially at protein-coding genes, but anticipate additional mechanisms for the more recently uncovered premature termination processes.

Concluding Remarks

Most interest in transcription has historically focused on initiation and elongation, which have a more intuitive influence on gene regulation than termination. However, there is increasing appreciation that the vast majority of Pol II might terminate before it even reaches the 3′ end of a gene [99]. This implies that most transcriptional termination has scarcely been studied and that there is significant attrition of Pol II throughout a gene, some of which could be subject to regulation. We now have a mechanistic framework to explain termination at most protein-coding genes as reviewed here: transcription of a PAS results in deceleration of Pol II which is then terminated by XRN2. Even so, an important challenge will be to address promoter-proximal and premature termination events in more detail (see Outstanding Questions). Whether the PAS-dependent mechanism provides a paradigm for understanding termination elsewhere and how broadly endo- and exonuclease are used will be important and interesting to establish (Figure 3). Future experiments will enrich our understanding and, like many stories, there will be more intriguing twists before the end.

Acknowledgments

We thank members of the laboratory and Hannah Mischo for comments. The laboratory is supported by grants from the Lister Institute and The Wellcome Trust (WT107791/Z/15/Z).

Box 3. Termination at Short versus Long Transcription Units

Although integrator and CPSF have an impact on termination at shorter transcription units, their depletion is not associated with the runaway read through caused by CPSF73 loss at longer PAS-containing genes [37,38,94]. Moreover, many short transcripts are exosome substrates whereas RNA downstream of a PAS is not [8]. This indicates the potential for termination of some early transcription to directly release 3′ ends. Promoter-proximal polymerases are also significantly impeded by nucleosomes [10,102], whereas Pol II readily transcribes hundreds of kilobases of chromatin when CPA is inhibited [37,103]. These differences could be caused by the instability of early elongation, which is perhaps susceptible to a variety of termination activities (a state that might be re-established at the 3′ end of genes). Factors that influence the vulnerability of early Pol II to termination might be those promoting productive elongation (and consequent resistance to termination) such as pTEFb [104], SPT5 phosphorylation [48], U1 snRNA [27], promoter identity [105], or other elongation factors. Pol II CTD is also influential because its tyrosine residues impact on read through (perhaps indirectly by controlling integrator and mediator recruitment) and its phosphorylation status dictates termination decisions in budding yeast [51,106]. Finally, because XRN2 often terminates Pol II that is incapable of efficient elongation, its broader impact on termination could have been overlooked due to the expectation of longer read through in its absence. More generally, as promoter-proximal Pol II might not be capable of rapid elongation, it will be harder to identify termination defects over this region simply by looking for read through even if XRN2 is not involved.

Outstanding Questions

What proteins comprise the PP1-controlled network contributing to the allosteric switch for termination?

How does XRN2 promote termination when it catches Pol II?

What are the mechanisms of promoter-proximal termination and that at short gene classes?

How is premature termination activated or repressed to control gene output?

Trends in Genetics

CellPress
REVIEWS

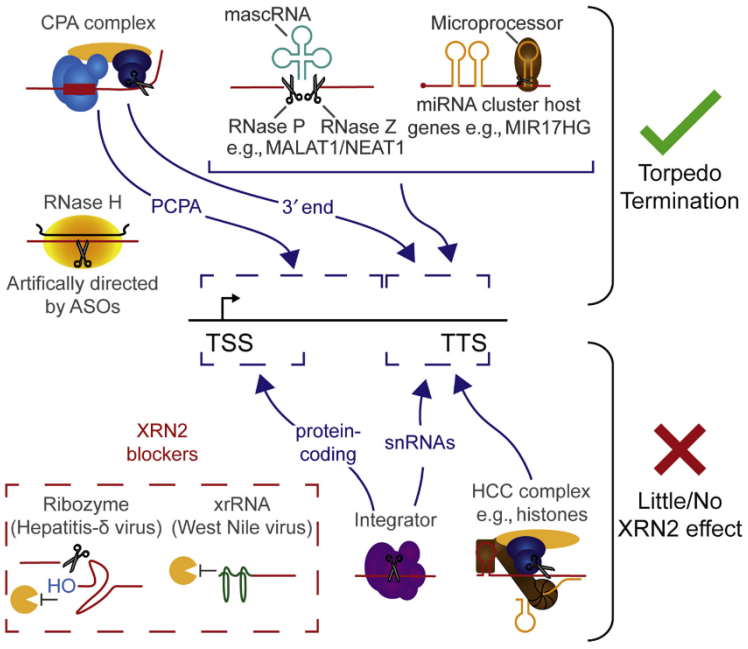


Figure 3. Cleavage and XRN2 Effects on Termination across the Gene. In the center is a schematic of a gene with the conventional transcription start and termination sites (TSS/TTS) indicated. Different cleavage activities can act at different steps in the transcription cycle or at different gene classes and these are depicted and labelled. Some of these promote XRN2 (torpedo)-dependent termination (green tick) whereas an XRN2 effect is less obvious in other cases or for cleavage in certain locations (red cross). Mechanisms to definitively block XRN2, via a hammerhead ribozyme (yielding an incompatible 5′OH) or xrRNA, are shown in the broken red box. Abbreviations: ASOs, **antisense oligonucleotides**; CPA, cleavage and polyadenylation; HCC, histone cleavage complex; mascRNA, MALAT1-associated small cytoplasmic RNA; PCPA, premature cleavage and polyadenylation; snRNA, small nuclear RNA; TSS, transcription start site; TTS, transcription termination site XRN2, nuclear enzyme that degrades RNA in a 5′→3′ direction; xrRNA, Xrn-resistant RNA.

References

1. Eick, D. and Geyer, M. (2013) The RNA polymerase II carboxy-terminal domain code. *Chem. Rev.* 113, 8456–8490
2. Proudfoot, N.J. (2016) Transcriptional termination in mammals: stopping the RNA polymerase II juggernaut. *Science* 352, aad9926
3. Proudfoot, N.J. (2012) Ending the message: poly(A) signals then and now. *Genes Dev.* 25, 1770–1782
4. Krebs, A.R. *et al.* (2017) Genome-wide single-molecule footprinting reveals high RNA polymerase II turnover at paused promoters. *Mol. Cell* 67, 411–422.e4
5. Steurer, B. *et al.* (2018) Live-cell analysis of endogenous GFP-RPB1 uncovers rapid turnover of initiating and promoter-paused RNA Polymerase II. *Proc. Natl. Acad. Sci. U. S. A.* 115, E4368–E4376
6. Tatomer, D.C. *et al.* (2019) The Integrator complex cleaves nascent mRNAs to attenuate transcription. *Genes Dev.* 33, 1525–1538
7. Elrod, N.D. *et al.* (2019) The integrator complex attenuates promoter-proximal transcription at protein-coding genes. *Mol. Cell* 76, 738–752.e7
8. Davidson, L. *et al.* (2019) Rapid depletion of DIS3, EXOSC10, or XRN2 reveals the immediate impact of exoribonucleolysis on nuclear RNA metabolism and transcriptional control. *Cell Rep.* 26, 2779–2791.e5
9. Oh, J.M. *et al.* (2017) U1 snRNP telescripting regulates a size-function-stratified human genome. *Nat. Struct. Mol. Biol.* 24, 993–999
10. Chu, A.C. *et al.* (2018) Transcriptional pause sites delineate stable nucleosome-associated premature polyadenylation suppressed by U1 snRNP. *Mol. Cell* 69, 648–663.e7
11. O'Reilly, D. *et al.* (2014) Human snRNA genes use polyadenylation factors to promote efficient transcription termination. *Nucleic Acids Res.* 42, 264–275
12. Eaton, J.D. *et al.* (2018) Xrn2 accelerates termination by RNA polymerase II, which is underpinned by CPSF73 activity. *Genes Dev.* 32, 127–139
13. Fong, N. *et al.* (2015) Effects of transcription elongation rate and Xrn2 exonuclease activity on RNA polymerase II termination suggest widespread kinetic competition. *Mol. Cell* 60, 256–267

14. Agostini, F. *et al.* (2020) Intergenic RNA mainly derives from nascent transcripts of known genes. *bioRxiv*. Published online January 9, 2020. https://doi.org/10.1101/2020.01.08.898478
15. Lee, J.S. and Mendell, J.T. (2020) Antisense-mediated transcript knockdown triggers premature transcription termination. *Mol. Cell* 77, 1044–1054.e3
16. Lai, F. *et al.* (2020) Directed RNase H cleavage of nascent transcripts causes transcription termination. *Mol. Cell* 77, 1032–1043.e4
17. Whitelaw, E. and Proudfoot, N. (1986) Alpha-thalassaemia caused by a poly(A) site mutation reveals that transcriptional termination is linked to 3' end processing in the human alpha 2 globin gene. *EMBO J.* 5, 2915–2922
18. Clerici, M. *et al.* (2017) Structural insights into the assembly and polyA signal recognition mechanism of the human CPSF complex. *eLife* 6, e33111
19. Schonemann, L. *et al.* (2014) Reconstitution of CPSF active in polyadenylation: recognition of the polyadenylation signal by WDR33. *Genes Dev.* 28, 2381–2393
20. Chan, S.L. *et al.* (2014) CPSF30 and Wdr33 directly bind to AAUAAA in mammalian mRNA 3' processing. *Genes Dev.* 28, 2370–2380
21. Sun, Y. *et al.* (2018) Molecular basis for the recognition of the human AAUAAA polyadenylation signal. *Proc. Natl. Acad. Sci. U. S. A.* 115, E1419–E1428
22. Logan, J. *et al.* (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, 8306–8310
23. Connelly, S. and Manley, J.L. (1988) A functional mRNA polyadenylation signal is required for transcription termination by RNA polymerase II. *Genes Dev.* 2, 440–452
24. Proudfoot, N.J. (1989) How RNA polymerase II terminates transcription in higher eukaryotes. *Trends Biochem. Sci.* 14, 105–110
25. Gregersen, L.H. *et al.* (2019) SCAF4 and SCAF8, mRNA anti-terminator proteins. *Cell* 177, 1797–1813.e18
26. Dubbury, S.J. *et al.* (2018) CDK12 regulates DNA repair genes by suppressing intronic polyadenylation. *Nature* 564, 141–145
27. Kaida, D. *et al.* (2010) U1 snRNP protects pre-mRNAs from premature cleavage and polyadenylation. *Nature* 468, 664–668
28. Jenal, M. *et al.* (2012) The poly(A)-binding protein nuclear 1 suppresses alternative cleavage and polyadenylation sites. *Cell* 149, 538–553
29. Kamieniarz-Gdula, K. *et al.* (2019) Selective roles of vertebrate PCF11 in premature and full-length transcript termination. *Mol. Cell* 74, 158–172.e9
30. Ahn, S.H. *et al.* (2004) Phosphorylation of serine 2 within the RNA polymerase II C-terminal domain couples transcription and 3' end processing. *Mol. Cell* 13, 67–76
31. Zhang, H. *et al.* (2015) 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, 437–448
32. Osheim, Y.N. *et al.* (2002) EM visualization of Pol II genes in *Drosophila*: most genes terminate without prior 3' end cleavage of nascent transcripts. *Chromosoma* 111, 1–12
33. Custodio, N. *et al.* (1999) Inefficient processing impairs release of RNA from the site of transcription. *EMBO J.* 18, 2855–2866
34. Zenkin, N. (2014) Ancient RNA stems that terminate transcription. *RNA Biol.* 11, 295–297
35. Nielsen, S. *et al.* (2013) Mechanism of eukaryotic RNA polymerase III transcription termination. *Science* 340, 1577–1580
36. Birse, C.E. *et al.* (1998) Coupling termination of transcription to messenger RNA maturation in yeast. *Science* 280, 298–301
37. Eaton, J.D. *et al.* (2020) A unified allosteric/torpedo mechanism for transcriptional termination on human protein-coding genes. *Genes Dev.* 34, 132–145
38. Nojima, T. *et al.* (2015) Mammalian NET-Seq reveals genome-wide nascent transcription coupled to RNA processing. *Cell* 161, 526–540
39. West, S. *et al.* (2004) Human 5′→3′ exonuclease Xrn2 promotes transcription termination at co-transcriptional cleavage sites. *Nature* 432, 522–525
40. Kim, M. *et al.* (2004) The yeast Rat1 exonuclease promotes transcription termination by RNA polymerase II. *Nature* 432, 517–522
41. Baejen, C. *et al.* (2017) Genome-wide analysis of RNA polymerase II termination at protein-coding genes. *Mol. Cell* 66, 38–49.e6
42. Brannan, K. *et al.* (2012) mRNA decapping factors and the exonuclease Xrn2 function in widespread premature termination of RNA polymerase II transcription. *Mol. Cell* 46, 311–324
43. Luo, W. *et al.* (2006) The role of Rat1 in coupling mRNA 3′-end processing to transcription termination: implications for a unified allosteric-torpedo model. *Genes Dev.* 20, 954–965
44. Chapman, E.G. *et al.* (2014) RNA structures that resist degradation by Xrn1 produce a pathogenic Dengue virus RNA. *eLife* 3, e01892
45. Parua, P.K. *et al.* (2018) A Cdk9-PP1 switch regulates the elongation-termination transition of RNA polymerase II. *Nature* 558, 460–464
46. Kecman, T. *et al.* (2018) Elongation/termination factor exchange mediated by PP1 phosphatase orchestrates transcription termination. *Cell Rep.* 25, 259–269.e5
47. Cortazar, M.A. *et al.* (2019) Control of RNA Pol II speed by PNUTS-PP1 and Spt5 dephosphorylation facilitates termination by a "sitting duck torpedo" mechanism. *Mol. Cell* 76, 896–908.e4
48. Yamada, T. *et al.* (2006) P-TEFb-mediated phosphorylation of hSpt5 C-terminal repeats is critical for processive transcription elongation. *Mol. Cell* 21, 227–237
49. Shi, Y. *et al.* (2009) Molecular architecture of the human pre-mRNA 3′ processing complex. *Mol. Cell* 33, 365–376
50. Schreieck, A. *et al.* (2014) RNA polymerase II termination involves C-terminal-domain tyrosine dephosphorylation by CPF subunit Glc7. *Nat. Struct. Mol. Biol.* 21, 175–179
51. Shah, N. *et al.* (2018) Tyrosine-1 of RNA polymerase II CTD controls global termination of gene transcription in mammals. *Mol. Cell* 69, 48–61.e6
52. Heroes, E. *et al.* (2013) The PP1 binding code: a molecular-lego strategy that governs specificity. *FEBS J.* 280, 584–595
53. Kaneko, S. *et al.* (2007) The multifunctional protein p54nrb/PSF recruits the exonuclease XRN2 to facilitate pre-mRNA 3′ processing and transcription termination. *Genes Dev.* 21, 1779–1789
54. Schwalb, B. *et al.* (2016) TT-seq maps the human transient transcriptome. *Science* 352, 1225–1228
55. Sheridan, R.M. *et al.* (2019) Widespread backtracking by RNA Pol II is a major effector of gene activation, 5′ pause release, termination, and transcription elongation rate. *Mol. Cell* 73, 107–118.e4
56. Crossley, M.P. *et al.* (2019) R-loops as cellular regulators and genomic threats. *Mol. Cell* 73, 398–411
57. Tan-Wong, S.M. *et al.* (2019) R-loops promote antisense transcription across the mammalian genome. *Mol. Cell* 76, 600–616.e6
58. Chen, L. *et al.* (2017) R-ChIP using inactive RNase H reveals dynamic coupling of R-loops with transcriptional pausing at gene promoters. *Mol. Cell* 68, 745–757.e5
59. Skourti-Stathaki, K. *et al.* (2011) Human senataxin resolves RNA/DNA hybrids formed at transcriptional pause sites to promote Xrn2-dependent termination. *Mol. Cell* 42, 794–805
60. Kolev, N.G. *et al.* (2008) Conserved motifs in both CPSF73 and CPSF100 are required to assemble the active endonuclease for histone mRNA 3′-end maturation. *EMBO Rep.* 9, 1013–1018
61. Hoek, T.A. *et al.* (2019) Single-molecule imaging uncovers rules governing nonsense-mediated mRNA decay. *Mol. Cell* 75, 324–339.e11
62. Ujvari, A. *et al.* (2002) RNA polymerase II transcription complexes may become arrested if the nascent RNA is shortened to less than 50 nucleotides. *J. Biol. Chem.* 277, 32527–32537
63. Park, J. *et al.* (2015) Unraveling the mechanistic features of RNA polymerase II termination by the 5′-3′ exoribonuclease Rat1. *Nucleic Acids Res.* 43, 2625–2637
64. Sikova, M. *et al.* (2020) The torpedo effect in *Bacillus subtilis*: RNase J1 resolves stalled transcription complexes. *EMBO J.* 39, e102500

65. Ray-Soni, A. *et al.* (2016) Mechanisms of bacterial transcription termination: all good things must end. *Annu. Rev. Biochem.* 85, 319–347
66. Nudler, E. and Gottesman, M.E. (2002) Transcription termination and anti-termination in *E. coli*. *Genes Cells* 7, 755–768
67. Gusarov, I. and Nudler, E. (1999) The mechanism of intrinsic transcription termination. *Mol. Cell* 3, 495–504
68. Martin, F.H. and Tinoco Jr., I. (1980) DNA-RNA hybrid duplexes containing oligo(dA:rU) sequences are exceptionally unstable and may facilitate termination of transcription. *Nucleic Acids Res.* 8, 2295–2299
69. Rondon, A.G. *et al.* (2009) Fail-safe transcriptional termination for protein-coding genes in *S. cerevisiae*. *Mol. Cell* 36, 88–98
70. Preker, P. *et al.* (2008) RNA exosome depletion reveals transcription upstream of active human promoters. *Science* 322, 1851–1854
71. Ntini, E. *et al.* (2013) Polyadenylation site-induced decay of upstream transcripts enforces promoter directionality. *Nat. Struct. Mol. Biol.* 20, 923–928
72. Sun, Y. *et al.* (2020) Structure of an active human histone pre-mRNA 3′-end processing machinery. *Science* 367, 700–703
73. Marzluff, W.F. and Koreski, K.P. (2017) Birth and death of histone mRNAs. *Trends Genet.* 33, 745–759
74. Chodchoy, N. *et al.* (1991) An intact histone 3′-processing site is required for transcription termination in a mouse histone H2a gene. *Mol. Cell. Biol.* 11, 497–509
75. Yang, X.C. *et al.* (2009) Studies of the 5′ exonuclease and endonuclease activities of CPSF-73 in histone pre-mRNA processing. *Mol. Cell. Biol.* 29, 31–42
76. Bucholc, K. *et al.* (2020) Composition and processing activity of a semi-recombinant holo U7 snRNP. *Nucleic Acids Res.* 48, 1508–1530
77. Richards, J. and Belasco, J.G. (2011) Ribonuclease J: how to lead a double life. *Structure* 19, 1201–1203
78. Stevens, A. and Maupin, M.K. (1987) A 5′–3′ exoribonuclease of human placental nuclei: purification and substrate specificity. *Nucleic Acids Res.* 15, 695–708
79. Xiang, S. *et al.* (2009) Structure and function of the 5′→3′ exoribonuclease Rat1 and its activating partner Rai1. *Nature* 458, 784–788
80. Sanders, T.J. *et al.* (2020) FttA is a CPSF73 homologue that terminates transcription in Archaea. *Nat. Microbiol.* 5, 545–553
81. Baillat, D. *et al.* (2005) Integrator, a multiprotein mediator of small nuclear RNA processing, associates with the C-terminal repeat of RNA polymerase II. *Cell* 123, 265–276
82. Egloff, S. *et al.* (2010) The integrator complex recognizes a new double mark on the RNA polymerase II carboxyl-terminal domain. *J. Biol. Chem.* 285, 20564–20569
83. Stadelmayer, B. *et al.* (2014) Integrator complex regulates NELF-mediated RNA polymerase II pause/release and processivity at coding genes. *Nat. Commun.* 5, 5531
84. Szczepinska, T. *et al.* (2015) DIS3 shapes the RNA polymerase II transcriptome in humans by degrading a variety of unwanted transcripts. *Genome Res.* 25, 1622–1633
85. Yamamoto, J. *et al.* (2014) DSIF and NELF interact with Integrator to specify the correct post-transcriptional fate of snRNA genes. *Nat. Commun.* 5, 4263
86. Hallais, M. *et al.* (2013) CBC-ARS2 stimulates 3′-end maturation of multiple RNA families and favors cap-proximal processing. *Nat. Struct. Mol. Biol.* 20, 1358–1366
87. Zhang, Z. and Gilmour, D.S. (2006) Pcf11 is a termination factor in *Drosophila* that dismantles the elongation complex by bridging the CTD of RNA polymerase II to the nascent transcript. *Mol. Cell* 21, 65–74
88. Cuello, P. *et al.* (1999) Transcription of the human U2 snRNA genes continues beyond the 3′ box *in vivo*. *EMBO J.* 18, 2867–2877
89. Adelman, K. and Lis, J.T. (2012) Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. *Nat. Rev. Genet.* 13, 720–731
90. Erickson, B. *et al.* (2018) Dynamic turnover of paused Pol II complexes at human promoters. *Genes Dev.* 32, 1215–1225
91. Nilson, K.A. *et al.* (2017) Oxidative stress rapidly stabilizes promoter-proximal paused Pol II across the human genome. *Nucleic Acids Res.* 45, 11088–11105
92. Gardini, A. *et al.* (2014) Integrator regulates transcriptional initiation and pause release following activation. *Mol. Cell* 56, 128–139
93. Lai, F. *et al.* (2015) Integrator mediates the biogenesis of enhancer RNAs. *Nature* 525, 399–403
94. Nojima, T. *et al.* (2018) Deregulated expression of mammalian lncRNA through loss of SPT6 induces R-loop formation, replication stress, and cellular senescence. *Mol. Cell* 72, 970–984.e7
95. Hernandez, N. and Weiner, A.M. (1986) Formation of the 3′ end of U1 snRNA requires compatible snRNA promoter elements. *Cell* 47, 249–258
96. Ramamurthy, L. *et al.* (1996) Increasing the distance between the snRNA promoter and the 3′ box decreases the efficiency of snRNA 3′-end formation. *Nucleic Acids Res.* 24, 4525–4534
97. Davidson, L. *et al.* (2012) Co-transcriptional degradation of aberrant pre-mRNA by Xrn2. *EMBO J.* 31, 2566–2578
98. Almada, A.E. *et al.* (2013) Promoter directionality is controlled by U1 snRNP and polyadenylation signals. *Nature* 499, 360–363
99. Kamieniarz-Gdula, K. and Proudfoot, N.J. (2019) Transcriptional control by premature termination: a forgotten mechanism. *Trends Genet.* 35, 553–564
100. Natsume, T. *et al.* (2016) Rapid protein depletion in human cells by auxin-inducible degron tagging with short homology donors. *Cell Rep.* 15, 210–218
101. Schlackow, M. *et al.* (2017) Distinctive patterns of transcription and RNA processing for human lincRNAs. *Mol. Cell* 65, 25–38
102. Skene, P.J. *et al.* (2014) The nucleosomal barrier to promoter escape by RNA polymerase II is overcome by the chromatin remodeler Chd1. *eLife* 3, e02042
103. Heinz, S. *et al.* (2018) Transcription elongation can affect genome 3D structure. *Cell* 174, 1522–1536.e22
104. Peterlin, B.M. and Price, D.H. (2006) Controlling the elongation phase of transcription with P-TEFb. *Mol. Cell* 23, 297–305
105. Miki, T.S. *et al.* (2017) Two distinct transcription termination modes dictated by promoters. *Genes Dev.* 31, 1870–1879
106. Gudipati, R.K. *et al.* (2008) Phosphorylation of the RNA polymerase II C-terminal domain dictates transcription termination choice. *Nat. Struct. Mol. Biol.* 15, 786–794
107. Mandel, C.R. *et al.* (2006) Polyadenylation factor CPSF-73 is the pre-mRNA 3′-end-processing endonuclease. *Nature* 444, 953–956