Transcriptional termination in mammals: stopping the RNA polymerase II juggernaut.

Nick J Proudfoot

Juggernaut was the name given for a huge wagon bearing an image of the god Krishna drawn annually in procession at Puri in Orissa, India from the 1600s. Some devotees were crushed under its wheels in sacrifice. Now means a massive, unstoppable vehicle.

Prokaryotes versus eukaryotes:

Genes are defined as regions of the genome that correspond to a single transcription unit (TU), starting from the promoter and ending at the terminator. Whilst promoters are often well characterised, less is known about the mechanism and regulation of transcriptional termination, even though defining the extent of transcription units across the genome is an essential aspect of genetics. For prokaryotic genes, protein expression units (cistrons) are usually clustered into tandem arrays transcribed as a single TU creating a polycistronic messenger (m) RNA. Failure to terminate transcription results in the inclusion of extra cistrons in the extended mRNA that may cause the production of unwanted proteins with adverse biological consequences (*1, 2*). The basic mechanism of termination in the model prokaryote *E.coli* is well defined. If the nascent transcript can form an RNA hairpin structure immediately followed by an oligo U sequence then this will trigger termination (*3*). Termination can also be facilitated by the ATP dependent translocase Rho which recognises a loosely defined C rich sequence (Rho utilisation transcript-RUT) on nascent RNA (*4*). Following initial polymerase binding, hexameric Rho translocates and unravels the nascent RNA in association with the elongating polymerase (*5*). Contacts between an RNA hairpin or Rho and the polymerase somehow trigger conformational changes that switch its enzymatic mode from elongation to termination. An important aspect of prokaryotic termination is that mRNA translation occurs on transcripts still being made by RNA polymerase (co-transcriptional). Ribosomes recognise purine rich sequence elements on the nascent transcript that are positioned just upstream of each initiation codon through base

pairing with the 3’ end of 16s ribosomal (r) RNA (*6*). Translation elongation along the mRNA template can remove RNA hairpin structures or block access of Rho to RUT sites. Either way translation can directly regulate termination and the consequent extent of transcription units (*7*).

Eukaryotic genes are fundamentally different to prokaryotes as their transcription occurs in the nucleus, separate from the cytoplasmic translation apparatus. Furthermore eukaryotes employ three different classes of RNA polymerase (Pol). Pol II transcribes all protein coding genes to generate mRNA as well as many noncoding (nc) RNA. ncRNA can either be abundant and stable such as small nuclear and small nucleolar (sn and sno) RNA, or low level and rapidly degraded such as longer (l) ncRNA that may run between or overlap with protein coding genes (*8*). In contrast Pol I transcribes the highly abundant rRNA precursor which is co-transcriptionally processed to mature 28s, 18s and 5.8s rRNA, while Pol III transcribes transfer (t) RNA and 5s rRNA. All eukaryotic mRNA are monocistronic with a short RNA tract before and a longer one following the coding region (5’ and 3’ untranslated regions or UTRs). The 5’UTR begins with a 5’ terminal Cap structure while the 3’UTR 3’ ends with a polyA tail. Both these terminal mRNA modifications are formed as part of pre-mRNA processing that occurs co-transcriptionally and is also coordinated with removal (splicing) of introns that separate the coding exons. These complex RNA processing reactions are all required to generate translatable mRNA which is then exported through the nuclear pore to sites of cytoplasmic translation. In particular translation initiation factors recognise the mRNA Cap structure which then recruit the ribosome to initiate translation at an adjacent initiation codon. In general translation can’t initiate at internal initiation codons unless a complex RNA structure exists upstream, referred to as an internal ribosome entry site (IRES) (*9*). IRES sequences have been identified in various eukaryotic viral genomes as well as on some regulated mRNA (*10*, *11*). An interesting new twist is that methylation of mRNA adenosines to 6meA by MTT3 methyl transferase is known to occur on mRNA (*12*). Furthermore 6meA marks on 5’UTR generate another signal that allows ribosome recruitment to adjacent initiation codons independently of the regular terminal Cap structure (*13*).

Failure to terminate transcription in eukaryotic genes may have severe consequences for gene expression. For protein coding genes arranged in tandem, read-through transcripts from a non-terminated upstream gene will run into the promoter of the downstream gene and restrict its activity by a process called transcriptional interference (*14*, *15*) . This will in turn prevent Cap addition to the downstream gene transcript as this can only occur on a triphosphorylated 5’ end. For genes arranged in convergent orientation, termination defects may result in the formation of overlapping transcripts that down-regulate gene expression by triggering RNA interference pathways (*16*). In severe cases, failure of convergent genes to terminate transcription will result in molecular collision between Pol II transcribing opposite DNA template strands (*17*, *18*). Another major deleterious effect of failed termination is that Pol II elongation complexes may run into regions of the genome undergoing DNA replication. Collision with DNA polymerase complexes may disrupt DNA synthesis and trigger DNA damage and genome instability (*19*). The extensive lncRNA transcriptome increases the likelihood of potential interference problems between TUs. Failure of lncRNA to terminate transcription may also cause interference with adjacent protein coding genes (*20*) even though our current understanding of lncRNA termination is rudimentary. The low level expression of many lncRNA may mean that

for a single DNA template only one of two potentially interfering TUs will be transcribed at one time thereby obviating interference (*21*).

Studies on transcription termination have undergone a major experimental shift in recent years. Previously gene transcription was measured by gene specific analysis using steady state RNA mapping techniques such as Northern blotting and RNase protection analysis. The extent of primary (unprocessed or nascent) transcription across the gene could only be determined by nuclear run on analysis involving the *in vitro* labelling of elongating Pol II in isolated nuclear preparations (*22*). These technologies are now largely superseded by the advent of high throughput sequencing where transcription profiles across all genes in the genome can be assessed in a single experiment (RNA-seq). While many RNA-seq analyses use steady state RNA fractions, it is possible to visualise nascent transcription using a global run on analysis procedure (GRO-seq) (*23*) or by sequencing RNA isolated from nuclear chromatin (*24*). Finally truly nascent RNA can be isolated and sequenced from within Pol II elongation complexes (NET-seq) (*24, 25*). Consequently it is now possible to directly map the extent of transcription past the polyA site (PAS) of a gene which remarkably reveals significant diversity as to where transcription terminates. While many Pol II genes display gradual termination profiles across multiple kilobases, others terminate abruptly at specific locations.

My review describes current understanding of how RNA polymerase II terminates transcription, mainly focusing on mammalian protein coding genes, but also with reference to other eukaryotic systems that exemplify specific features. I will start with a consideration of how the chromatin template signals Pol II to either slow down or completely stop. I will then consider how transcript processing and degradation can also trigger Pol II termination. Finally I will describe how termination can often be modulated to allow enhanced gene regulation or perturbed to cause genetic disease.

Intrinsic termination:

Pol II is uniquely endowed with an extra protein segment separate from the main globular enzyme that derives from the carboxy terminal domain (CTD) of Rpb1 (*26*). CTD plays a critical role in coordinating co-transcriptional RNA processing: capping, splicing and 3'end cleavage and polyadenylation. In mammals it comprises a relatively unstructured polypeptide of 52 heptad repeats (consensus $Y_1S_2P_3T_4S_5P_6S_7$) that are rich in differentially phosphorylated serine residues. In particular S_2P is associated with gene 3' ends and interacts with a large complex of cleavage and polyA factors (CPA) that generate mRNA 3' ends (*27-30*). A further important difference between Pol II and all other types of RNA polymerase transcription is that Pol II transcribed genes in mammals show an enormous range of genomic size, varying from a few hundred nucleotides for snRNA genes through to many protein coding genes exceeding 100 kb in length. Indeed the largest known human protein coding gene is Dystrophin with a primary transcript exceeding 2 mega bases (*31*). It is evident that Pol II must have the capacity to be highly processive to allow even a few Pol II elongation complexes to traverse the whole gene and thereby generate a translatable mRNA. Clearly when Pol II finally reaches the gene end, then termination mechanisms need to be sufficiently robust to stop this molecular juggernaut.

As a first line of defence it is apparent that Pol II is capable of directly sensing its passage across a functional PAS. This has been shown genome wide in recent mNET-seq analysis of human HeLa cells where significant levels of Pol II pausing are detected immediately following protein coding gene PAS. Furthermore depletion of key components of CPA such as CPSF-73 and CstF-64, reduces this Pol II pausing effect (*24*). Apparently CPA recruitment to the Pol II elongation complex as it traverses the PAS induces a significant slow-down effect on Pol II elongation. These studies are backed up by *in vitro* transcription experiments which show that the interaction of CPSF and CstF components with Pol II transcribing through a gene PAS can have significant pausing effects on transcription that results in the gradual release of Pol II from the DNA template (*32*). This termination process apparently occurs independently of PAS cleavage so arguing that Pol II conformational changes alone can induce significant levels of transcriptional termination.

Transcriptional pausing; chromatin and R-loops (see Figure 1A):

As described above the recognition of PAS by CPA factors promotes transcriptional pausing that may in turn induce some degree of intrinsic termination. However other features of the chromatin template may also induce pausing that will in turn increase the dwell time over the PAS. This will increase the likelihood of CPA association with the PAS and the consequent completion of 3' end processing and coupled transcriptional termination (see below). Perhaps the most common type of transcriptional pausing for Pol II is caused by chromatin structure, especially the core nucleosome. While it is clear that Pol II transcribes nucleosomal templates, it is also the case that nucleosome free or depleted templates are more readily transcribed. Nucleosome profiles across genes indicate that promoter regions are generally nucleosome depleted regions (NDR) extending into the first exon of genes (*33*). At this point there is often a build-up or pausing of Pol II prior to elongation across the gene body. This pausing is associated with a switch to higher nucleosomal occupancy on the edges of the NDR (*34*, *35*). Promoter proximal Pol II pausing (see below) is a widespread feature of regulated or inducible genes and is likely to be an important aspect of mammalian gene regulation as extensively reviewed elsewhere (*34*). Similarly there is substantial experimental evidence that exons are associated with higher nucleosome density and higher Pol II pausing levels (*36*) which may facilitate co-transcriptional splicing. A slow-down in Pol II elongation rates may provide the necessary time for spliceosome assembly to occur on intronic splice sites.

Pol II pausing can also be induced by hybridisation of the nascent transcript with the antisense DNA strand outside the elongation complex. This results in the formation of RNA:DNA hybrids and displacement of the sense DNA strand, a structure referred to as an R-loop (*37, 38*). R-loop formation is favoured by the act of Pol II transcription, since the DNA template behind the elongation complex is depleted in nucleosomes that are transiently displaced during the transcription process. The DNA double helix is also torsionally under-wound (negatively supercoiled). Once formed R-loops can remain as a stable structure especially as the RNA:DNA hybrid is thermodynamically stronger than duplex DNA. Also they are often associated with G rich regions of transcribed genes because the displaced sense DNA can form stabilising G quadruplex structures (*38*). R-loops were originally observed in budding yeast when pre-mRNA packaging by the THO complex was inactivated by gene deletion (*39*). Similarly defective splicing can lead to accumulation of R-loops (*40*). In both cases

accumulation of nascent RNA in close proximity to the under-wound, just transcribed DNA template leads to RNA:DNA hybrid formation.

Interestingly a clear effect of R-loop accumulation induced by mRNA packaging or splicing defects is increased DNA damage. This is due to the mutagenic nature of the single strand template DNA in the R-loop which leads to single and then double strand brakes with elevated levels of DNA recombination (*37*). The realisation that the even wild type levels of R-loop formation still requires R-loop resolution to prevent DNA damage led to the identification of helicases such as Sen1 in yeast that act to remove these potentially harmful structures. Indeed loss of Sen1 gives an equally severe DNA damage phenotype to loss of THO (*41*). In mammalian cells the homologue of Sen1 called Senataxin (a causative gene for various neurological diseases; AOA and ALS) is similarly required to resolve R-loops but also plays a direct role in promoting more efficient termination (*42*). The recruitment of Senataxin to terminator regions is likely mediated by the creation of a specific Pol II CTD mark on an arginine residue present at the 7^{th} position of the variant 31^{st} heptad repeat (R1810). Symmetric dimethylation of this residue by the methyl transferase PRMT5 recruits first SMN (survival of motor neuron disease associated protein) which then recruits Senataxin. Loss of any of these factors causes an accumulation of R-loops and a defect in Pol II termination (*43*). Remarkably an alternative Senataxin recruitment pathway involves the DNA repair factor BRCA1. This is recruited to R-loops especially at Pol II terminators and in turn directly recruits Senataxin. This will afford quick resolution of R-loops, both to promote Pol II termination and also to prevent DNA damage. Loss of SMN, BRCA1 or co-recruited Senataxin each cause an accumulation of the DNA damage histone mark γH2AX. Furthermore cells derived from cancer patients with loss of both BRCA1 alleles show increased recombination over R-loop associated termination regions (*44*).

While R-loop structures display an intrinsic slow-down effect on Pol II elongation they have also been shown to induce low level antisense transcription. This may result in the formation of transient double stranded RNA which will in turn trigger an RNA interference effect mediated by nuclear Dicer and Ago proteins. This leads to dimethylation of histone H3K9 by the histone methyl transferase enzyme G9a/GLP and consequent recruitment of HP1γ, effectively creating localised patches of repressed chromatin (*45*). These will act to perpetuate and enhance Pol II pausing over R-loop associated termination regions and are a feature of relatively short and ubiquitously expressed genes. Interestingly a natural example of R-loop formation in prokaryotes is the CRISPR cas9 gene targeting system where cas9 promotes base pairing of guide RNA to complementary DNA sequence, resulting in the formation of a transient R-loop structure (*46*). While Cas9 goes on to cleave the associated duplex DNA, mutant versions of Cas9 have been engineered that lack cleavage activity but still pause Pol II at the targeted position (*47*).

Transcriptional arrest (see Figure 1B):

Transcriptional arrest is taken to mean the irreversible association of Pol II with the DNA template such that it cannot be displaced by PAS mediated termination mechanisms to allow efficient recycling. Instead it is targeted for proteolytic degradation. PAS dependent termination may be viewed as a productive mechanism allowing reuse of Pol II while Pol II arrest can be viewed as non-productive since Pol II is degraded on the DNA template.

The chemical modification/damage of DNA by oxidation or UV treatment can directly lead to arrested Pol II transcription complexes. In such cases it is known that ubiquitin ligases are recruited to the complex resulting in Pol II degradation, effectively clearing the DNA template to allow new rounds of transcription (*48*). R-loops may also restrict the passage of the replication fork formed in the wake of DNA replication (*49*). Such collisions between replication and stalled transcription complexes have been shown to be the cause of fragile sites in the genome which are in effect positions where high levels of DNA recombination occur. These are often a feature of very long mammalian genes: the length of time it takes to complete their transcription exceeds the length of the cell cycle (*50*) necessitating the co-occurrence of transcription and replication. If transcription is paused at R-loops then replication will be slowed or stopped, resulting in collapsed replication forks and consequent DNA damage. Fascinating examples of transcriptional road blocks come from termination studies on non-mammalian genes. Thus in some trypanosomes a DNA modification process occurs where enzymes convert T to glucosyl hydroxymethyl uracil, called base J (*51*). While base J is mainly found in repetitive DNA it is also strategically positioned at the end of many TUs, especially those arranged in convergent orientation. Remarkably transcript reads precisely terminate at J sites. Deletion of genes encoding the enzymes that perform this T-J conversion result in read-through transcription and ultimately cell death (*51*). As with road blocks caused by DNA mutation the only way to release Pol II from base J road blocks must be to destroy it by ubiquitin triggered proteolysis. Several unexplained issues remain with this trypanosomal termination mechanism. First it is unclear how trypanosomes can tolerate the likely high turnover of Pol II necessitated by its release from DNA through proteolytic destruction. Also the selection mechanism that places base J at the end of transcription units has yet to be determined (*52*). Back to *S. cerevisiae* another clear case of a protein-induced DNA road block has come to light with the DNA binding protein Reb1. This protein was well known to act as a transcription factor for ribosomal protein coding genes, through its capacity to form nucleosome depleted promoter regions. This is in part achieved by the recruitment of chromatin remodelling complexes such as RSC (*53*). However as Reb1 binds very tightly to the DNA template it can also block the passage of Pol II. Remarkably Reb1 binding may be a general feature of yeast intergenic sequence where it restricts read-through transcription acting as a failsafe mechanism to prevent global transcriptional interference between genes. Like DNA damage and T->J mediated termination, Reb1 termination causes Pol II arrest so that the only way to remove the Pol II complex appears to be through ubiquitin mediated Pol II destruction (*54*).

Torpedo termination:

Ever since it was realised that the 3' end of eukaryotic mRNA is formed by an RNA processing mechanism rather than direct transcription termination, how termination connects with transcript cleavage has been intensively studied. It is now well appreciated that the CPA complex assembles onto the pre-mRNA PAS as it is extruded from the RNA exit channel of Pol II at the gene 3' end. This is facilitated by prior recruitment of CPA to the nearby Pol II CTD (*55*). Two models are widely cited for how PAS recognition triggers termination. One model already referred to above as intrinsic termination, but also dubbed the allosteric model (indicative of Pol II conformational change) is that Pol II elongation somehow senses its passage through a functional PAS (*32*). Likely this is caused by the association of the very

large CPA complex with Pol II CTD. This may induce a conformational change within the Pol II active site resulting in first pausing and then somehow Pol II release. However an alternative model relates to the nascent transcript and what happens to the RNA still being synthesised by Pol II following cleavage at the PAS and subsequent release of the polyadenylated mRNA from its chromatin cradle.

A key feature of this model, originally proposed in the 1980s (*56, 57*) is that the nuclear 5'-3'exonuclease Xrn2 is recruited to PAS and progressively degrades the downstream product, starting from the 5'phosphorylated end generated by 3' end processing. In effect degradation of this RNA acts in kinetic competition with ongoing Pol II elongation. If Xrn2 catches Pol II then this somehow acts as a molecular trigger to release Pol II from the DNA template. Clearly pausing of Pol II, caused for instance by R-loop mediated heterochromatic marks (Figure 1A), will enhance Xrn2 mediated termination (*45*). This proposed mechanism is evocatively named the torpedo model where in navel vernacular, Pol II is the battleship and Xrn2 the torpedo. Direct evidence for the torpedo model came 15 years later when it was shown that depletion of Xrn2 in mammalian cells or Rat1 in *S. cerevisiae* provokes a significant loss in termination, as judged by increased transcriptional read-through (*58, 59*). In particular analysis of Rat1 revealed that it was part of a complex including Rai1 which tightly interacts with Rat1 and possesses both pyrophophatase and some 5'-3' exonuclease activity (*60*). Both these activities likely cooperate with Rat1 to promote more efficient RNA degradation. A third member of the torpedo complex Rtt103, possesses a CTD interaction domain (CID) possibly accounting for Rat1 recruitment to Pol II (*58*). Indeed other CPA factors may aid Rat1 recruitment including Pcf11 which also possesses a CID (*61*). Since this time, opinion on the veracity of this model has fluctuated. In general the depletion of Xrn2 by RNAi technology in mammalian cells produced varying degrees of termination defect leading to the view that other factors might cooperate with Xrn2/Rat1 to achieve more efficient termination (*24*). One such factor is the RNA:DNA helicase Sen1 in yeast or Senataxin in mammals which may act to expose the cleaved downstream RNA product to Xrn2 degradation. This may be particularly important if RNA is entwined with the DNA template in an R-loop structure (RNA:DNA hybrid) (*41, 42*). Interestingly Sen1 displays termination activity independently of RNA degradation (*62*) indicating that the simple act of unravelling RNA structure is enough to destabilise Pol II and promote termination. Such a mechanism is very similar to the termination activity of *E. coli* Rho. Recent experiments appear to have finally laid to rest remaining doubts about the role of Xrn2 in Pol II termination. While depletion of Xrn2 by RNAi often causes only a marginal termination defect, combining RNAi treatment with expression of dominant negative Xrn2 (mutated in its active site) gives a satisfyingly large termination defect for most protein coding genes (*63*). It appears that depletion of Xrn2 cellular levels (by RNAi) may not adequately reduce levels of Xrn2 actively engaged in termination.

An underlying feature of productive termination is that 3' end cleavage of the nascent transcript at the PAS facilitates termination by allowing Xrn2 "torpedo" action. The RNAse III endonuclease activity of *S. cerevisiae* Rnt1 is a good case in point. Rnt1 recognises hairpin structures crowned with an AGNN tetra loop and carries out a double endonuclease cut across the hairpin. Several yeast genes employ Rnt1 cleavage as alternative 3' end processing events that allow PAS independent Xrn2 mediated termination (*64, 65*). Remarkably a very similar mechanism operates on lncRNA derived primary microRNA (miRNA) which are independently transcribed rather than more usually residing in the introns of protein coding

genes. Lnc-miRNA are co-transcriptionally cleaved by Drosha, a distant relative of Rnt1. RNA cleavage releases pre-miRNA which are then exported from the nucleus and converted into miRNA using another Rnt 1 relative called Dicer. Significantly Drosha cleavage not only generates miRNA but also promotes lncRNA gene termination further downstream, again likely involving the action of Xrn2 degradation (*20*, *66*). To underline the effectiveness of Rnt1 cleavage as a termination mechanism, Pol I transcription is also terminated by Rnt1 cleavage of a tetra loop hairpin at the 3' end of rRNA genes. This again elicits termination by a combination of Xrn2 and Sen1 action (*67*). Going back to PAS dependent termination, a further category of Pol II termination has been uncovered. In this case following passage of the PAS the polymerase continues to generate an extended transcript multiple kilobases into the gene 3' flanking region. Termination eventually occurs, coincident with a terminal co-transcriptionally cleaved transcript (CoTC sequence) which generates cleavage products closely associated with Pol II. These are degraded by Xrn2 and so promote close by Pol II termination (*59*, *68*). Remarkably CoTC termination still requires the presence of an upstream PAS. This suggests that the conformational change induced by CPA recognition of the PAS is required for downstream CoTC mediated termination (*69*). Cleavage at the PAS to finally release polyadenylated mRNA may occur following CoTC mediated termination, effectively after the Pol II complex, with its still associated pre-mRNA is released into the nucleoplasm (*70*).

Transcriptional backtracking to promote termination:

All RNA polymerases have the capacity to oscillate transcription in both forward and reverse directions on the DNA template. Forward movement results in template dependent RNA synthesis with the nascent transcript emerging from the RNA exit channel. Backward movement (backtracking) results in extrusion of the already synthesised nascent transcript out of the secondary channel (also referred to as the nucleotide entry channel) (*71*). During transcriptional elongation such backtracking is widely employed by Pol II as a proof reading mechanism. The general transcription factor TFIIS enhances an intrinsic endonuclease activity of Pol II that promotes cleavage of the mismatched extruded RNA. This allows transcription to resume, reinstating the correct nucleotide into the nascent RNA (*72*, *73*). For both *E. coli* RNA polymerase and eukaryotic Pol III, backtracking has been directly implicated in termination. For intrinsic termination in *E. coli*, a model is envisaged where oligo U sequences promote polymerase pausing which then favours backtracking. If an RNA hairpin forms on the upstream transcript, it can be forced into the RNA exit channel which in turn triggers a conformational change in the polymerase that promotes its release from the DNA template (*3*). Remarkably Pol III appears to adopt a similar strategy as transcript 3' ends are normally oligo U sequences which pause the polymerase and so encourage backtracking. If backward polymerase movement encounters a hairpin structure then termination ensues (*74*) (Figure 2A).

It is tempting to speculate that Pol II may employ a related backtracking mechanism to promote termination. Suggestive evidence comes from the yeast *S. pombe* where inactivation of the RNA exosome displays a clear general termination defect which is counteracted by simultaneous loss of TFIIS (*75*). Since the multi subunit exosome possesses two separate 3'-5'exonucleases, one of these may act on the extruded RNA formed by backtracking. Possibly this pushes the polymerase further backwards and by so doing induces conformational changes in the Pol II active site that promote termination. Further evidence for such a mechanism comes

from *in vitro* termination experiments using purified yeast Pol II, Rat1 and Rai1 together with immobilised DNA templates, where transcription is artificially blocked by omitting specific nucleotides (*76*). While earlier *in vitro* experiments failed to observe Pol II termination with these minimal components (*77*), significant termination was observed if the wrong nucleotide is forced onto the transcript 3' end so arresting Pol II at this mismatch position. This has the effect of inducing Pol II backtracking and also remarkably promotes termination when coupled with degradation of the upstream transcript by Xrn2 up to the arrested Pol II (*76*). In effect these experiments argue that the act of removing RNA up to or into the Pol II active site by either degradation of backtracked extruded transcript (reverse torpedo) or degradation of upstream RNA up to backtracked Pol II (forward torpedo) induces conformational changes to Pol II that promote termination (Figure 2B). It remains for structural biology to establish the true nature of such conformational changes and how they ultimately promote Pol II release from the DNA template.

Premature termination versus transcriptional elongation:

S. cerevisiae possesses a secondary Pol II termination mechanism that operates on short Pol II transcripts. This involves the NRD complex (Nrd1, Nab3 and Sen1) (*78*) which promotes termination of ncRNA particularly derived from antisense promoter activity associated with promoters of protein coding genes. It also functions on genes encoding small stable RNA, snRNA and snoRNA and plays a major role in regulating a subset of protein coding genes by promoting their premature termination. NRD promotes termination in a sequence specific manner (through RNA recognition domains on Nrd1 and Nab3) and recruits the exosome to rapidly degrade these transcripts. This will occur unless they are protected by RNA binding proteins that package snRNA and snoRNA into functional splicing or RNA modification complexes respectively.

While mammalian genes have no clear counterpart to NRD, most protein coding genes display substantial promotor proximal pausing. This is manifest by an accumulation of actively transcribing Pol II localised to the first few hundred nucleotides of the gene (*34*). In contrast *S. cerevisiae* genes show little Pol II pausing at the promoter. It would appear that the mammalian transcriptome, possibly due to its greater complexity, has evolved mechanisms of transcriptional regulation more focused on post initiation events. However some of these early transcripts are abortive, needing to be quickly terminated or they may be intermediates waiting to be converted into full length gene transcripts. Interestingly *bona fide* termination clearly operates on these TSS transcripts. Firstly decapping of the transcript can occur by Dcp1 action followed by Xrn2 degradation (*79*). Furthermore mis-spliced transcripts are somehow detected by nuclear surveillance and are similarly degraded by decapping and Xrn2 degradation (*80*). Promoter proximal PAS are also thought to be actively recognised by CPA and Xrn2 to promote early termination. Thus depletion of either CPSF components or Xrn2 both increase levels of TSS associated transcripts (*24*). Surprisingly a recently identified feature of these early terminated TSS transcripts shows that some form hairpin structures that are directly converted to functional pre-microRNA without the involvement of the microprocessor. Decapping on the 5' side and exonuclease degradation on their 3' side can generate functional pre-microRNA (*81*). It is apparent that this early termination is regulated by 5' splice sites which block TSS associated PAS and thereby favour continued elongation into the gene body.

This phenomenon was first shown in viruses such as HIV-1 (*82*) but has also been revealed as a general mechanism that acts to block cryptic PAS recognition and consequent premature termination. In particular depletion of U1 snRNA activates TSS proximal PAS as well as numerous PAS present across genes, often within their extensive intronic regions (*83*).

As well as controlled RNA processing, transcription elongation is tightly controlled across genes (*84*). Specific check points operate to enforce premature termination that acts to prevent inappropriate transcription (Figure 3). Early transcription elongation is restricted by two negative elongation factors DSIF and NELF which are regulated by the major Pol II CTD S_2 kinase Cdk9. As well as acting on CTD this kinase further phosphorylates DSIF and NELF. Thus when Cdk9 is experimentally inhibited by drugs such as DRB or KM, a substantial increase in TSS proximal transcripts is observed with greatly reduced transcription downstream into the gene body (*85*). Once Pol II escapes from TSS proximal check points and elongation is fully underway then numerous elongation factors come into play. These promote efficient transcription across the TU, be it a modest 1 kb or much longer 1 Mb gene. Many of these elongation factors act to remodel nucleosomes encountered by elongating Pol II as well as to coordinate efficient and often regulated (alternative) intron splicing (*84*). Recently a further example of regulated premature termination has been observed near the end of gene transcription units (3' end check point) that may act as a last ditch control mechanism to prevent the production of a translatable but potentially flawed mRNA. Like the TSS check point, the 3' end check point is controlled by Cdk9 activity. In this case the substrate may be Xrn2 which is significantly activated by Cdk9 mediated phosphorylation (*86*). Inhibition of Cdk9 causes a non-productive termination mechanism that still appears to utilise component parts of CPA, but apparently does not allow the formation of functional polyadenylated mRNA (*85*).

Alternative polyadenylation and termination:

The previous section emphasises the major regulatory role played by termination at early stages of gene transcription. Needless to say termination at gene 3' ends is also subject to intense regulation. A principal feature of mRNA 3' ends relates to the startling fact that many if not most mRNA possess variable lengths of 3' untranslated sequence defined by the selective usage of different PAS (*87, 88*). Since mRNA 3' UTRs define mRNA cytoplasmic functions including RNA stability, translatability and localisation, the use of alternative polyA sites (APA) can constitute a key regulatory process in gene expression. However the differential stability of different mRNA 3'UTR isoforms in the cytoplasm must be distinguished from actual PAS selection during pre-mRNA synthesis. Analysing total cellular mRNA isoform levels will mainly show up differential stability, while analysis of nuclear mRNA isoform levels more closely reflects PAS selection. In fact nuclear APA shows rather less variation and consequently gene regulation then initially envisaged (*89*). An issue often overlooked is whether APA constitutes alternative RNA processing alone or whether selective termination defines which PAS is utilised. Furthermore APA may also occur at more internal positions within a gene where it will result in truncated mRNA which in some cases can be translated into shorter protein isoforms with important biological functions. A classic example of this phenomenon is the alternative use of an internal PAS in IgM that generates secreted antibody versus a distal PAS that produces membrane bound antibody. This APA switch is regulated by levels of CstF64 which is depleted in early B cells and favours distal PAS selection (*90*). In

most examples of APA (as in IgM) the first (proximal) PAS will have weaker sequence features such as the lack of an exact AAUAAA sequence or no clear upstream or downstream sequence element (*91*). Consequently CPA factors will associate with such PAS less efficiently, in effect giving more time for later transcribed downstream (distal) PAS to be recognised. Notably distal PAS tend to possess canonical PAS elements (*87, 91*). As mentioned above Pol II pausing will also have significant effects here, as pausing between proximal and distal PAS favours the proximal PAS. Once a PAS is selected this will trigger Pol II termination downstream and so preclude distal PAS usage.

Several specific conditions are known to affect APA. Thus the kinetics of transcription can have a major effect. In general rapidly proliferating cells favour proximal PAS usage (*87*). Exactly how a gene promoter is set up to recruit elongation factors such as the PAF complex can also impact on downstream Pol II elongation and consequent PAS selection (*92*). Finally directly slowing down Pol II elongation by specific Pol II mutation can favour proximal PAS usage (*93*). Interestingly either slowing down or speeding up Pol II elongation by specific Pol II mutation can be seen to globally shorten or extend Pol II TUs, implying effects on both APA and coupled termination (*63*). Added to the above Pol II kinetic considerations, the availability of CPA factors can also impact on APA. Several studies show that the systematic depletion of individual CPA components by RNAi treatment can influence APA (*94, 95*). For example depletion of CFI components and PABPN1 reduce while Pcf11 and Fip1 depletion enhance distal PAS usage. However it is generally unclear whether such artificial CPA factor depletions can naturally occur. Even so in human pathology CF1m25 is shown to be over-expressed in glioblastoma cells resulting in global proximal PAS selection (*96*). Also PABN1 can be lost in a rare form of muscular dystrophy called OPMD where a triplet expansion in the gene generates an inactive protein containing N-terminal polyalanine. This depletion of functional PABN1 causes a clear APA shift to distal PAS usage (*97, 98*). Yet another fascinating example of factor levels affecting APA is U1snRNA, already mentioned as a means to block premature PAS usage at upstream gene positions. While substantial depletion of U1snRNA causes premature termination (*83*), more subtle titration of U1snRNA levels can impact on 3'UTR APA. In this case lowering U1snRNA levels favours proximal PAS usage. It is possible that in activated neurons, U1snRNA may be sufficiently depleted to favour proximal PAS usage (*99*). An encouraging avenue that leads to better understanding of how APA is regulated comes from non CPA RNA binding factors that are known to be tightly regulated in amount and/or cellular location. These include CPEB, FUS, ELAV and MBNL (*100-103*). All of these factors display some degree of RNA binding specificity and where their binding sites are near PAS may directly influence CPA binding/recruitment (Figure 3).

The actual mechanism behind how APA may be regulated in these diverse situations remains largely unknown. However the consequences of a mRNA having a more or less extended 3'UTR clearly relates to key functions of these mRNA regulatory regions, including microRNA binding sites and RNA stability elements. Remarkably the actual selection of different PAS can also result in the recruitment of protein factors to the 3'UTR which when the mRNA is translated can directly associate with the nascent protein to modulate its function (*104*).

Mis-regulated termination:

As emphasised above transcriptional termination is a highly robust process capable of stopping the Pol II juggernaut at gene 3' ends. It is therefore paradoxical that this basic mechanism appears to be readily subverted for particular cellular or viral needs. Firstly a notable feature of repressed termination can be seen in the primary piRNA clusters of *Drosophila*. These clusters harbour a wide range of transposon related sequences and in *Drosophila* are often transcribed on both DNA strands. Primary piRNA are processed into small, 20 nucleotide piRNA by dedicated RNA processing enzymes, including specific argonaute proteins. PiRNA act to block the spread of transposons and retroposons in a wide range of eukaryotes by targeting transposon derived mRNA, using a mechanism analogous to miRNA (*105*). Notably piRNA clusters lack independent promoters, but are positioned between convergent genes so that they are transcribed by read-through transcription (*106-108*). Due to dsRNA formation in these clusters, the primary piRNA chromatin is marked by H3K9me3, normally a feature of repressed HPI associated heterochromatin. However these TUs are bound by a different HPI protein called Rhino which is associated with other proteins including a Rai1 homologue called Cutoff. Remarkably Rhino enhances rather than represses Pol II transcription across the piRNA TUs and Cutoff further aids this read-through transcription process by blocking Pol II termination. Possibly Cutoff acts as a dominant negative regulator of the Xrn2 torpedo (*63*). Not only are the normal termination sites of the flanking convergent genes blocked so promoting efficient Pol II read-through transcription across both strands of the primary piRNA TUs, but also terminators prevalent in transposon termini are similarly restricted (Figure 4A).

Further examples of blocked termination comes from cells undergoing stress. Thus the artificial induction of osmotic stress in cultured cells results in a large number of genes failing to efficiently terminate at the normal gene 3' end. Instead read-through transcripts are detectible that may extend through intergenic regions and invade downstream positioned genes. These downstream of gene transcripts (DOGs) while potentially deleterious due to interference effects, have been postulated to possess protective features for the overall integrity of the nucleus under cellular stress (*109*). Similarly cancer cells have been widely reported to display complex read-through transcription profiles that may be related to DOGs (*110-112*). Indeed for renal cancer, mutations in the methyl transferase gene SETD2 correlate with read-through transcription profiles (*113*). Interestingly Setd2 adds the histone H3K36me3 mark to genic nucleosomes which is required for Pol II elongation and termination. Setd2 also cooperates with Pol II elongation factors and facilitates Pol II CTD S_2P formation which will ultimately lead to CPA recruitment and termination (*114*).

Viral infection may be considered an extreme form of cellular stress. Remarkably at least two viruses are known to drastically perturb the termination efficiency of their host genomes. Thus influenza virus essentially blocks host transcription termination, genome-wide. In detail the viral protein NS1 has high affinity for CPSF30 and through this interaction destroys CPA complex integrity. This leads to a general loss in 3' end processing with commensurate read-through transcription (*115*) (E. Fodor, personal communication). Herpes simplex virus upon infecting host cells similarly causes a massive mis-regulation of host gene transcription. Extensive read-through transcription occurs with all the associated interference and mis-splicing of these extended transcripts (*116*) (Figure 4B). Arguably the destruction of normal regulated Pol II termination in host cells infected with these common human viruses at least in part explains their pathogenicity.

Conclusions:

This review charts our ever increasing understanding of how transcriptional termination impacts on many aspects of eukaryotic gene expression. Far from acting as a constitutive mechanism to separate TUs across the genome, termination can be seen as an intricate process that displays remarkable flexibility. At the beginning of the gene, termination regulates transcript release into productive elongation. It also acts as a checkpoint to prevent the synthesis of defective mRNA, which could be translated into multiple copies of a toxic (dominant negative) protein. At the end of the gene, termination dictates which mRNA isoform is formed by APA, thereby conferring selective expression properties on the mRNA. Finally termination can be overridden to adjust cells to stress conditions or to adapt cells into a more pliable host for viral replication. It is likely that future analysis of the termination process has yet more surprises in store.

1. S. Adhya, M. Gottesman, Control of transcription termination. *Annu Rev Biochem* **47**, 967 (1978).
2. T. Platt, Transcription termination and the regulation of gene expression. *Annu Rev Biochem* **55**, 339 (1986).
3. V. Epshtein, C. J. Cardinale, A. E. Ruckenstein, S. Borukhov, E. Nudler, An allosteric path to transcription termination. *Mol Cell* **28**, 991 (Dec 28, 2007).
4. J. P. Richardson, Preventing the synthesis of unused transcripts by Rho factor. *Cell* **64**, 1047 (Mar 22, 1991).
5. V. Epshtein, D. Dutta, J. Wade, E. Nudler, An allosteric mechanism of Rho-dependent transcription termination. *Nature* **463**, 245 (Jan 14, 2010).
6. J. A. Steitz, K. Jakes, How ribosomes select initiator regions in mRNA: base pair formation between the 3' terminus of 16S rRNA and the mRNA during initiation of protein synthesis in Escherichia coli. *Proc Natl Acad Sci U S A* **72**, 4734 (Dec, 1975).
7. T. M. Henkin, C. Yanofsky, Regulation by transcription attenuation in bacteria: how RNA provides instructions for transcription termination/antitermination decisions. *Bioessays* **24**, 700 (Aug, 2002).
8. B. S. Gloss, M. E. Dinger, The specificity of long noncoding RNA expression. *Biochim Biophys Acta*, (Aug 19, 2015).
9. N. Sonenberg, A. G. Hinnebusch, Regulation of translation initiation in eukaryotes: mechanisms and biological targets. *Cell* **136**, 731 (Feb 20, 2009).
10. G. Lozano, E. Martinez-Salas, Structural insights into viral IRES-dependent translation mechanisms. *Current opinion in virology* **12**, 113 (Jun, 2015).
11. K. A. Spriggs, M. Bushell, A. E. Willis, Translational regulation of gene expression during conditions of cell stress. *Mol Cell* **40**, 228 (Oct 22, 2010).
12. D. Dominissini *et al.*, Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. *Nature* **485**, 201 (May 10, 2012).
13. K. D. Meyer *et al.*, 5' UTR m(6)A Promotes Cap-Independent Translation. *Cell* **163**, 999 (Nov 5, 2015).
14. I. H. Greger, N. J. Proudfoot, Poly(A) signals control both transcriptional termination and initiation between the tandem GAL10 and GAL7 genes of Saccharomyces cerevisiae. *EMBO J* **17**, 4771 (Aug 17, 1998).
15. K. E. Shearwin, B. P. Callen, J. B. Egan, Transcriptional interference--a crash course. *Trends Genet* **21**, 339 (Jun, 2005).

16. M. Gullerova, N. J. Proudfoot, Cohesin complex promotes transcriptional termination between convergent genes in S. pombe. *Cell* **132**, 983 (Mar 21, 2008).
17. E. M. Prescott, N. J. Proudfoot, Transcriptional collision between convergent genes in budding yeast. *Proc Natl Acad Sci U S A* **99**, 8796 (Jun 25, 2002).
18. D. J. Hobson, W. Wei, L. M. Steinmetz, J. Q. Svejstrup, RNA polymerase II collision interrupts convergent transcription. *Mol Cell* **48**, 365 (Nov 9, 2012).
19. H. Gaillard, T. Garcia-Muse, A. Aguilera, Replication stress and cancer. *Nature reviews. Cancer* **15**, 276 (May, 2015).
20. A. Dhir, S. Dhir, N. J. Proudfoot, C. L. Jopling, Microprocessor mediates transcriptional termination of long noncoding RNA transcripts hosting microRNAs. *Nat Struct Mol Biol* **22**, 319 (Apr, 2015).
21. M. Castelnuovo *et al.*, Bimodal expression of PHO84 is modulated by early termination of antisense transcription. *Nat Struct Mol Biol* **20**, 851 (Jul, 2013).
22. M. J. Dye, N. J. Proudfoot, Terminal exon definition occurs cotranscriptionally and promotes termination of RNA polymerase II. *Mol Cell* **3**, 371 (Mar, 1999).
23. L. J. Core, J. J. Waterfall, J. T. Lis, Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science* **322**, 1845 (Dec 19, 2008).
24. T. Nojima *et al.*, Mammalian NET-Seq Reveals Genome-wide Nascent Transcription Coupled to RNA Processing. *Cell* **161**, 526 (Apr 23, 2015).
25. L. S. Churchman, J. S. Weissman, Native elongating transcript sequencing (NET-seq). *Current protocols in molecular biology / edited by Frederick M. Ausubel ... [et al.]* **Chapter 4**, Unit 4 14 1 (Apr, 2012).
26. A. Vannini, P. Cramer, Conservation between the RNA polymerase I, II, and III transcription initiation machineries. *Mol Cell* **45**, 439 (Feb 24, 2012).
27. S. McCracken *et al.*, The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. *Nature* **385**, 357 (Jan 23, 1997).
28. M. Heidemann, C. Hintermair, K. Voss, D. Eick, Dynamic phosphorylation patterns of RNA polymerase II CTD during transcription. *Biochim Biophys Acta* **1829**, 55 (Jan, 2013).
29. J. P. Hsin, J. L. Manley, The RNA polymerase II CTD coordinates transcription and RNA processing. *Genes Dev* **26**, 2119 (Oct 1, 2012).
30. S. Egloff, M. Dienstbier, S. Murphy, Updating the RNA polymerase CTD code: adding gene-specific layers. *Trends Genet* **28**, 333 (Jul, 2012).
31. C. N. Tennyson, H. J. Klamut, R. G. Worton, The human dystrophin gene requires 16 hours to be transcribed and is cotranscriptionally spliced. *Nat Genet* **9**, 184 (Feb, 1995).
32. H. Zhang, F. Rigo, H. G. Martinson, 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 (Aug 6, 2015).
33. J. L. Workman, Nucleosome displacement in transcription. *Genes Dev* **20**, 2009 (Aug 1, 2006).
34. K. Adelman, J. T. Lis, Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. *Nat Rev Genet* **13**, 720 (Oct, 2012).
35. D. A. Gilchrist *et al.*, Pausing of RNA polymerase II disrupts DNA-specified nucleosome organization to enable precise gene regulation. *Cell* **143**, 540 (Nov 12, 2010).
36. S. Naftelberg, I. E. Schor, G. Ast, A. R. Kornblihtt, Regulation of alternative splicing through coupling with transcription and chromatin structure. *Annu Rev Biochem* **84**, 165 (2015).
37. A. Aguilera, T. Garcia-Muse, R loops: from transcription byproducts to threats to genome stability. *Mol Cell* **46**, 115 (Apr 27, 2012).
38. K. Skourti-Stathaki, N. J. Proudfoot, A double-edged sword: R loops as threats to genome integrity and powerful regulators of gene expression. *Genes Dev* **28**, 1384 (Jul 1, 2014).
39. S. Jimeno, A. G. Rondon, R. Luna, A. Aguilera, The yeast THO complex and mRNA export factors link RNA metabolism with transcription and genome instability. *EMBO J* **21**, 3526 (Jul 1, 2002).

40. X. Li, J. L. Manley, Inactivation of the SR protein splicing factor ASF/SF2 results in genomic instability. *Cell* **122**, 365 (Aug 12, 2005).
41. H. E. Mischo *et al.*, Yeast Sen1 helicase protects the genome from transcription-associated instability. *Mol Cell* **41**, 21 (Jan 7, 2011).
42. K. Skourti-Stathaki, N. J. Proudfoot, N. Gromak, Human senataxin resolves RNA/DNA hybrids formed at transcriptional pause sites to promote Xrn2-dependent termination. *Mol Cell* **42**, 794 (Jun 24, 2011).
43. D. Yanling Zhao *et al.*, SMN and symmetric arginine dimethylation of RNA polymerase II C-terminal domain control termination. *Nature*, (Dec 23, 2015).
44. E. Hatchi *et al.*, BRCA1 recruitment to transcriptional pause sites is required for R-loop-driven DNA damage repair. *Mol Cell* **57**, 636 (Feb 19, 2015).
45. K. Skourti-Stathaki, K. Kamieniarz-Gdula, N. J. Proudfoot, R-loops induce repressive chromatin marks over mammalian gene terminators. *Nature*, (Oct 5, 2014).
46. H. Nishimasu *et al.*, Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell* **156**, 935 (Feb 27, 2014).
47. L. S. Qi *et al.*, Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* **152**, 1173 (Feb 28, 2013).
48. J. Q. Svejstrup, The interface between transcription and mechanisms maintaining genome integrity. *Trends Biochem Sci* **35**, 333 (Jun, 2010).
49. J. M. Santos-Pereira, A. Aguilera, R loops: new modulators of genome dynamics and function. *Nat Rev Genet* **16**, 583 (Oct, 2015).
50. A. Helmrich, M. Ballarino, L. Tora, Collisions between replication and transcription complexes cause common fragile site instability at the longest human genes. *Mol Cell* **44**, 966 (Dec 23, 2011).
51. H. G. van Luenen *et al.*, Glucosylated hydroxymethyluracil, DNA base J, prevents transcriptional readthrough in Leishmania. *Cell* **150**, 909 (Aug 31, 2012).
52. P. A. Genest *et al.*, Defining the sequence requirements for the positioning of base J in DNA using SMRT sequencing. *Nucleic Acids Res* **43**, 2102 (Feb 27, 2015).
53. P. D. Hartley, H. D. Madhani, Mechanisms that specify promoter nucleosome location and identity. *Cell* **137**, 445 (May 1, 2009).
54. J. Colin *et al.*, Roadblock termination by reb1p restricts cryptic and readthrough transcription. *Mol Cell* **56**, 667 (Dec 4, 2014).
55. N. Proudfoot, New perspectives on connecting messenger RNA 3' end formation to transcription. *Curr Opin Cell Biol* **16**, 272 (Jun, 2004).
56. N. J. Proudfoot, How RNA polymerase II terminates transcription in higher eukaryotes. *Trends Biochem Sci* **14**, 105 (Mar, 1989).
57. S. Connelly, J. L. Manley, A functional mRNA polyadenylation signal is required for transcription termination by RNA polymerase II. *Genes Dev* **2**, 440 (Apr, 1988).
58. M. Kim *et al.*, The yeast Rat1 exonuclease promotes transcription termination by RNA polymerase II. *Nature* **432**, 517 (Nov 25, 2004).
59. S. West, N. Gromak, N. J. Proudfoot, Human 5' --> 3' exonuclease Xrn2 promotes transcription termination at co-transcriptional cleavage sites. *Nature* **432**, 522 (Nov 25, 2004).
60. S. Xiang *et al.*, Structure and function of the 5'-->3' exoribonuclease Rat1 and its activating partner Rai1. *Nature* **458**, 784 (Apr 9, 2009).
61. D. D. Licatalosi *et al.*, Functional interaction of yeast pre-mRNA 3' end processing factors with RNA polymerase II. *Mol Cell* **9**, 1101 (May, 2002).
62. O. Porrua, D. Libri, A bacterial-like mechanism for transcription termination by the Sen1p helicase in budding yeast. *Nat Struct Mol Biol* **20**, 884 (Jul, 2013).
63. N. Fong *et al.*, Effects of Transcription Elongation Rate and Xrn2 Exonuclease Activity on RNA Polymerase II Termination Suggest Widespread Kinetic Competition. *Mol Cell* **60**, 256 (Oct 15, 2015).

64. A. G. Rondon, H. E. Mischo, J. Kawauchi, N. J. Proudfoot, Fail-safe transcriptional termination for protein-coding genes in S. cerevisiae. *Mol Cell* **36**, 88 (Oct 9, 2009).
65. G. Ghazal *et al.*, Yeast RNase III triggers polyadenylation-independent transcription termination. *Mol Cell* **36**, 99 (Oct 9, 2009).
66. M. Ballarino *et al.*, Coupled RNA processing and transcription of intergenic primary microRNAs. *Mol Cell Biol* **29**, 5632 (Oct, 2009).
67. J. Kawauchi, H. Mischo, P. Braglia, A. Rondon, N. J. Proudfoot, Budding yeast RNA polymerases I and II employ parallel mechanisms of transcriptional termination. *Genes Dev* **22**, 1082 (Apr 15, 2008).
68. M. J. Dye, N. J. Proudfoot, Multiple transcript cleavage precedes polymerase release in termination by RNA polymerase II. *Cell* **105**, 669 (Jun 1, 2001).
69. S. West, N. J. Proudfoot, M. J. Dye, Molecular dissection of mammalian RNA polymerase II transcriptional termination. *Mol Cell* **29**, 600 (Mar 14, 2008).
70. T. Nojima, M. Dienstbier, S. Murphy, N. J. Proudfoot, M. J. Dye, Definition of RNA polymerase II CoTC terminator elements in the human genome. *Cell reports* **3**, 1080 (Apr 25, 2013).
71. E. Nudler, RNA polymerase backtracking in gene regulation and genome instability. *Cell* **149**, 1438 (Jun 22, 2012).
72. S. Sigurdsson, A. B. Dirac-Svejstrup, J. Q. Svejstrup, Evidence that transcript cleavage is essential for RNA polymerase II transcription and cell viability. *Mol Cell* **38**, 202 (Apr 23, 2010).
73. M. G. Izban, D. S. Luse, The RNA polymerase II ternary complex cleaves the nascent transcript in a 3'----5' direction in the presence of elongation factor SII. *Genes Dev* **6**, 1342 (Jul, 1992).
74. S. Nielsen, Y. Yuzenkova, N. Zenkin, Mechanism of eukaryotic RNA polymerase III transcription termination. *Science* **340**, 1577 (Jun 28, 2013).
75. J. F. Lemay *et al.*, The RNA exosome promotes transcription termination of backtracked RNA polymerase II. *Nat Struct Mol Biol* **21**, 919 (Oct, 2014).
76. J. Park, M. Kang, M. Kim, Unraveling the mechanistic features of RNA polymerase II termination by the 5'-3' exoribonuclease Rat1. *Nucleic Acids Res* **43**, 2625 (Mar 11, 2015).
77. S. Dengl, P. Cramer, Torpedo nuclease Rat1 is insufficient to terminate RNA polymerase II in vitro. *J Biol Chem* **284**, 21270 (Aug 7, 2009).
78. H. E. Mischo, N. J. Proudfoot, Disengaging polymerase: terminating RNA polymerase II transcription in budding yeast. *Biochim Biophys Acta* **1829**, 174 (Jan, 2013).
79. K. Brannan *et al.*, mRNA decapping factors and the exonuclease Xrn2 function in widespread premature termination of RNA polymerase II transcription. *Mol Cell* **46**, 311 (May 11, 2012).
80. L. Davidson, A. Kerr, S. West, Co-transcriptional degradation of aberrant pre-mRNA by Xrn2. *EMBO J* **31**, 2566 (May 30, 2012).
81. M. Xie *et al.*, Mammalian 5'-capped microRNA precursors that generate a single microRNA. *Cell* **155**, 1568 (Dec 19, 2013).
82. M. P. Ashe, P. Griffin, W. James, N. J. Proudfoot, Poly(A) site selection in the HIV-1 provirus: inhibition of promoter-proximal polyadenylation by the downstream major splice donor site. *Genes Dev* **9**, 3008 (Dec 1, 1995).
83. D. Kaida *et al.*, U1 snRNP protects pre-mRNAs from premature cleavage and polyadenylation. *Nature* **468**, 664 (Dec 2, 2010).
84. H. Kwak, J. T. Lis, Control of transcriptional elongation. *Annu Rev Genet* **47**, 483 (2013).
85. C. Laitem *et al.*, CDK9 inhibitors define elongation checkpoints at both ends of RNA polymerase II-transcribed genes. *Nat Struct Mol Biol* **22**, 396 (May, 2015).
86. M. Sanso *et al.*, P-TEFb regulation of transcription termination factor Xrn2 revealed by a chemical genetic screen for Cdk9 substrates. *Genes Dev* **30**, 117 (Jan 1, 2016).
87. B. Tian, J. L. Manley, Alternative cleavage and polyadenylation: the long and short of it. *Trends Biochem Sci* **38**, 312 (Jun, 2013).
88. R. Elkon, A. P. Ugalde, R. Agami, Alternative cleavage and polyadenylation: extent, regulation and function. *Nat Rev Genet* **14**, 496 (Jul, 2013).

89. J. Neve *et al.*, Subcellular RNA profiling links splicing and nuclear DICER1 to alternative cleavage and polyadenylation. *Genome Res*, (Nov 6, 2015).
90. Y. Takagaki, R. L. Seipelt, M. L. Peterson, J. L. Manley, The polyadenylation factor CstF-64 regulates alternative processing of IgM heavy chain pre-mRNA during B cell differentiation. *Cell* **87**, 941 (Nov 29, 1996).
91. N. J. Proudfoot, Ending the message: poly(A) signals then and now. *Genes Dev* **25**, 1770 (Sep 1, 2011).
92. T. Nagaike *et al.*, Transcriptional activators enhance polyadenylation of mRNA precursors. *Mol Cell* **41**, 409 (Feb 18, 2011).
93. P. A. Pinto *et al.*, RNA polymerase II kinetics in polo polyadenylation signal selection. *EMBO J* **30**, 2431 (Jun 15, 2011).
94. W. Li *et al.*, Systematic profiling of poly(A)+ transcripts modulated by core 3' end processing and splicing factors reveals regulatory rules of alternative cleavage and polyadenylation. *PLoS Genet* **11**, e1005166 (Apr, 2015).
95. G. Martin, A. R. Gruber, W. Keller, M. Zavolan, 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 (Jun 28, 2012).
96. C. P. Masamha *et al.*, CFIm25 links alternative polyadenylation to glioblastoma tumour suppression. *Nature* **510**, 412 (Jun 19, 2014).
97. M. Jenal *et al.*, The poly(A)-binding protein nuclear 1 suppresses alternative cleavage and polyadenylation sites. *Cell* **149**, 538 (Apr 27, 2012).
98. E. de Klerk *et al.*, Poly(A) binding protein nuclear 1 levels affect alternative polyadenylation. *Nucleic Acids Res* **40**, 9089 (Oct, 2012).
99. M. G. Berg *et al.*, U1 snRNP determines mRNA length and regulates isoform expression. *Cell* **150**, 53 (Jul 6, 2012).
100. F. A. Bava *et al.*, CPEB1 coordinates alternative 3'-UTR formation with translational regulation. *Nature* **495**, 121 (Mar 7, 2013).
101. A. Masuda *et al.*, Position-specific binding of FUS to nascent RNA regulates mRNA length. *Genes Dev* **29**, 1045 (May 15, 2015).
102. V. Hilgers, S. B. Lemke, M. Levine, ELAV mediates 3' UTR extension in the Drosophila nervous system. *Genes Dev* **26**, 2259 (Oct 15, 2012).
103. R. Batra *et al.*, Loss of MBNL leads to disruption of developmentally regulated alternative polyadenylation in RNA-mediated disease. *Mol Cell* **56**, 311 (Oct 23, 2014).
104. B. D. Berkovits, C. Mayr, Alternative 3' UTRs act as scaffolds to regulate membrane protein localization. *Nature* **522**, 363 (Jun 18, 2015).
105. E. M. Weick, E. A. Miska, piRNAs: from biogenesis to function. *Development* **141**, 3458 (Sep, 2014).
106. F. Mohn, G. Sienski, D. Handler, J. Brennecke, The rhino-deadlock-cutoff complex licenses noncanonical transcription of dual-strand piRNA clusters in Drosophila. *Cell* **157**, 1364 (Jun 5, 2014).
107. Z. Zhang *et al.*, The HP1 homolog rhino anchors a nuclear complex that suppresses piRNA precursor splicing. *Cell* **157**, 1353 (Jun 5, 2014).
108. A. Le Thomas *et al.*, Transgenerationally inherited piRNAs trigger piRNA biogenesis by changing the chromatin of piRNA clusters and inducing precursor processing. *Genes Dev* **28**, 1667 (Aug 1, 2014).
109. A. Vilborg, M. C. Passarelli, T. A. Yario, K. T. Tycowski, J. A. Steitz, Widespread Inducible Transcription Downstream of Human Genes. *Mol Cell* **59**, 449 (Aug 6, 2015).
110. C. A. Maher *et al.*, Transcriptome sequencing to detect gene fusions in cancer. *Nature* **458**, 97 (Mar 5, 2009).
111. K. Kannan *et al.*, Recurrent chimeric RNAs enriched in human prostate cancer identified by deep sequencing. *Proc Natl Acad Sci U S A* **108**, 9172 (May 31, 2011).

112. K. E. Varley *et al.*, Recurrent read-through fusion transcripts in breast cancer. *Breast cancer research and treatment* **146**, 287 (Jul, 2014).
113. A. R. Grosso *et al.*, Pervasive transcription read-through promotes aberrant expression of oncogenes and RNA chimeras in renal carcinoma. *eLife* **4**, (Nov 17, 2015).
114. S. F. de Almeida, M. Carmo-Fonseca, Reciprocal regulatory links between cotranscriptional splicing and chromatin. *Semin Cell Dev Biol* **32**, 2 (Aug, 2014).
115. M. E. Nemeroff, S. M. Barabino, Y. Li, W. Keller, R. M. Krug, 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 (Jun, 1998).
116. A. J. Rutkowski *et al.*, Widespread disruption of host transcription termination in HSV-1 infection. *Nature communications* **6**, 7126 (2015).

Acknowledgements

I am grateful to my lab colleagues, especially Mick Dye and Hannah Mischo for advise on the text. I am indebted to Hannah Mischo for the illustratations. My laboratory is supported by a Wellcome Trust Investigator award (107928/Z/15/Z) and European Research Council Advanced Grant (339270).

Figure Legends

Figure 1

(A) Three different types of Pol II pausing induced by recognition of PAS by CPA, by R-loop formation and by heterochromatin patch. Elongating Pol II (red) shown transcribing DNA template with extruded, capped RNA transcript (blue). Nucleosomes depicted by yellow barrels with histone N-terminal tails indicated. Pol II CTD shown as extended tail. Red dots on CTD and histone tails denote methylation. Hand denotes pausing.

(B) Pol II arrested by base J or Reb1 DNA binding protein. Pol II is then ubiquinated and degraded by the proteasome.

Figure 2

(A) Bacterial RNA polymerase or Pol III terminate at oligo U transcript which pauses polymerase and promotes backtracking. Upstream RNA hairpin is forced into polymerase active site inducing a conformational change resulting in termination.

(B) Pol II moves forward to synthesise or backwards to extrude transcript (oscillation). Forward transcript, once cleaved at PAS to release mRNA is then degraded by Xrn2. Backtracked transcript is degraded by exosome. Removal of RNA up to Pol II (forward or reverse torpedo) induces termination.

Figure 3

Diagram depicting early release of Pol II prior to PAS or alternative PAS selection at gene 3' ends. This later process is mediated by competitive association of CPA versus other RNA binding factors to alternative PAS.

Figure 4

(A) Primary piRNA clusters in *Drosophila* placed between convergent genes. Double strand RNA induces heterochromatic histone tail modification (H3K9me3). This in turn recruits HPI-like factor Rhino together with Cutoff, an anti-terminator that promotes read-through transcription.

(B) Inactivation of termination by cancer mutation (SETD2 mutation), osmotic stress or viral infection all induce Pol II read-through and interference with downstream gene expression.

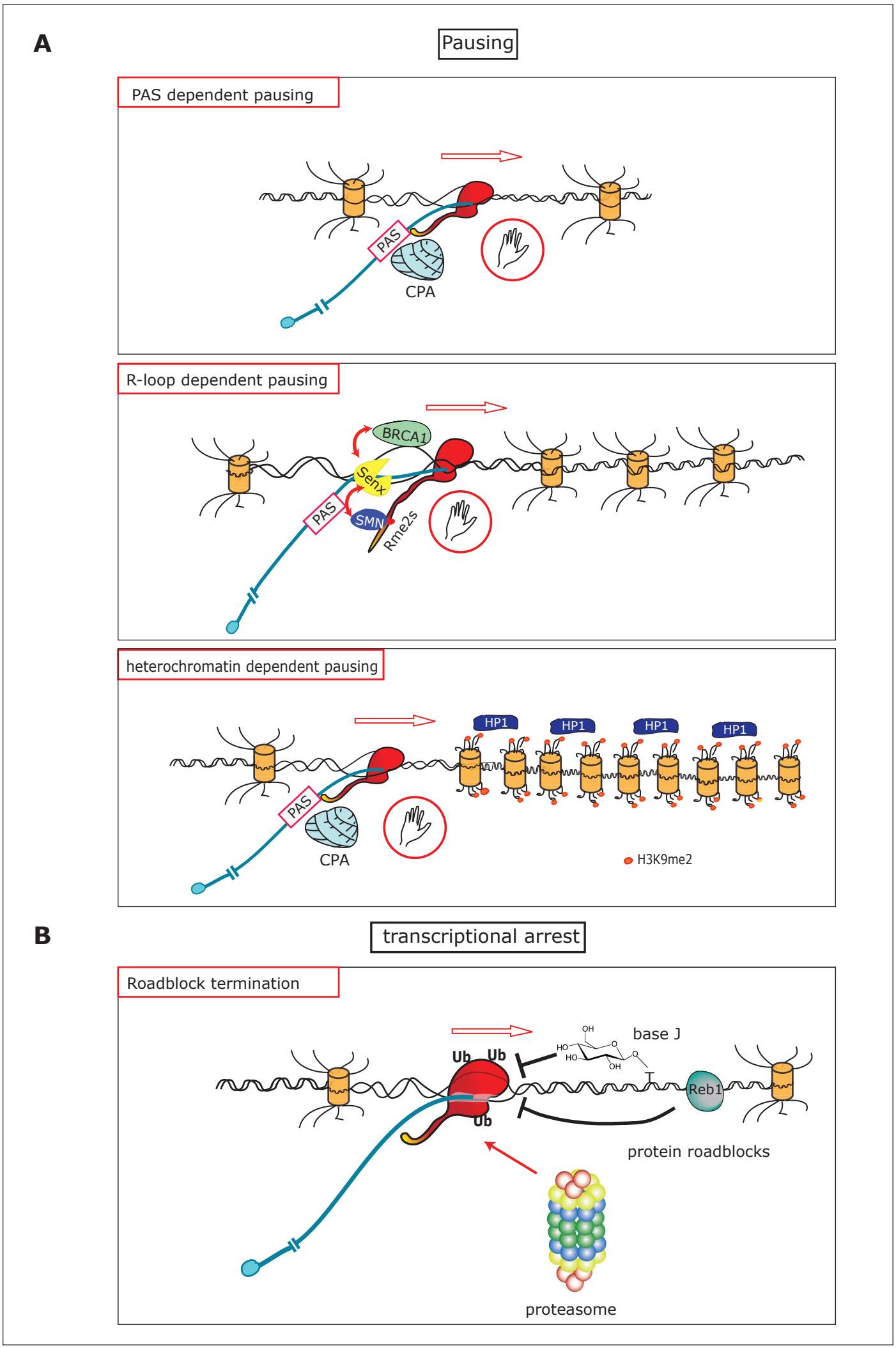

A
Pausing
PAS dependent pausing
PAS
CPA
R-loop dependent pausing
BRCA1
Senx
PAS
SMN
Rme2s
heterochromatin dependent pausing
HP1
HP1
HP1
HP1
PAS
CPA
H3K9me2
B
transcriptional arrest
Roadblock termination
base J
Ub
Ub
Ub
Reb1
protein roadblocks
proteasome

Figure 2

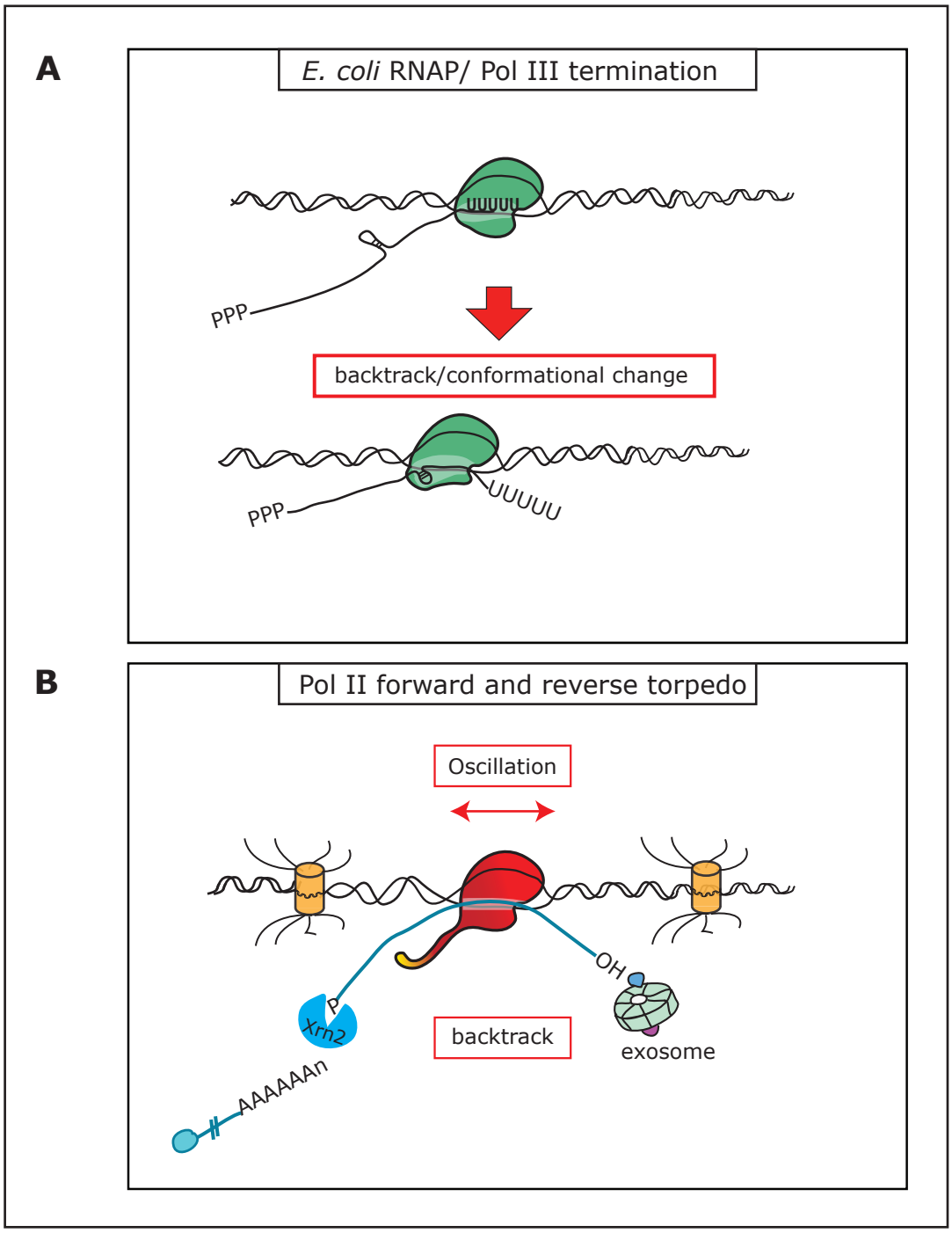

A
E. coli RNAP/ Pol III termination
UUUUU
PPP
backtrack/conformational change
PPP
UUUUU
B
Pol II forward and reverse torpedo
Oscillation
OH
P
Xrn2
backtrack
exosome
AAAAAAn

Figure 3

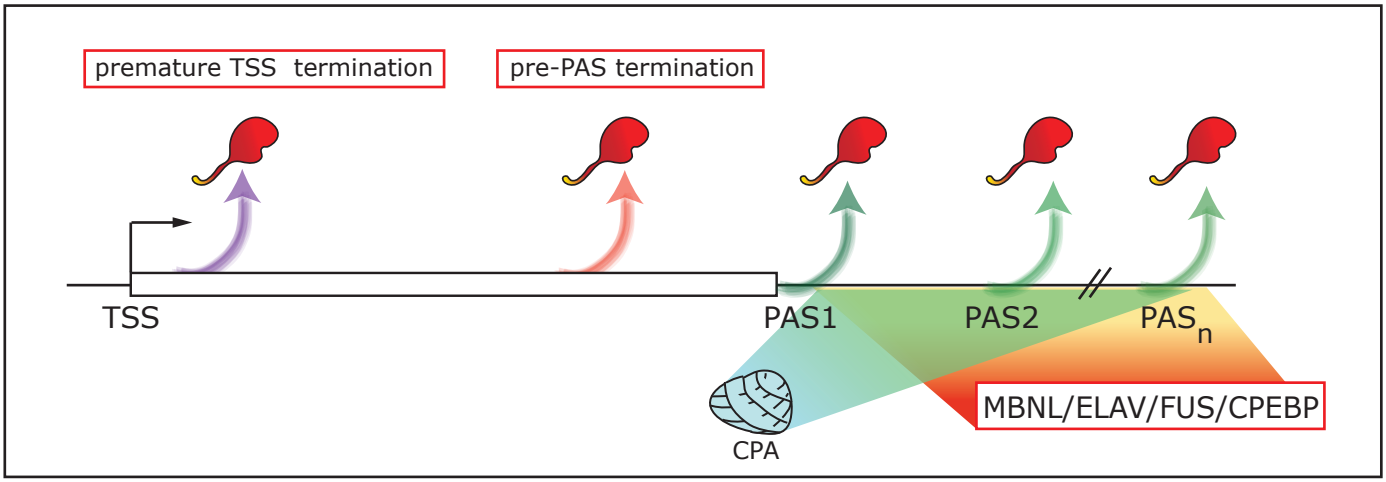
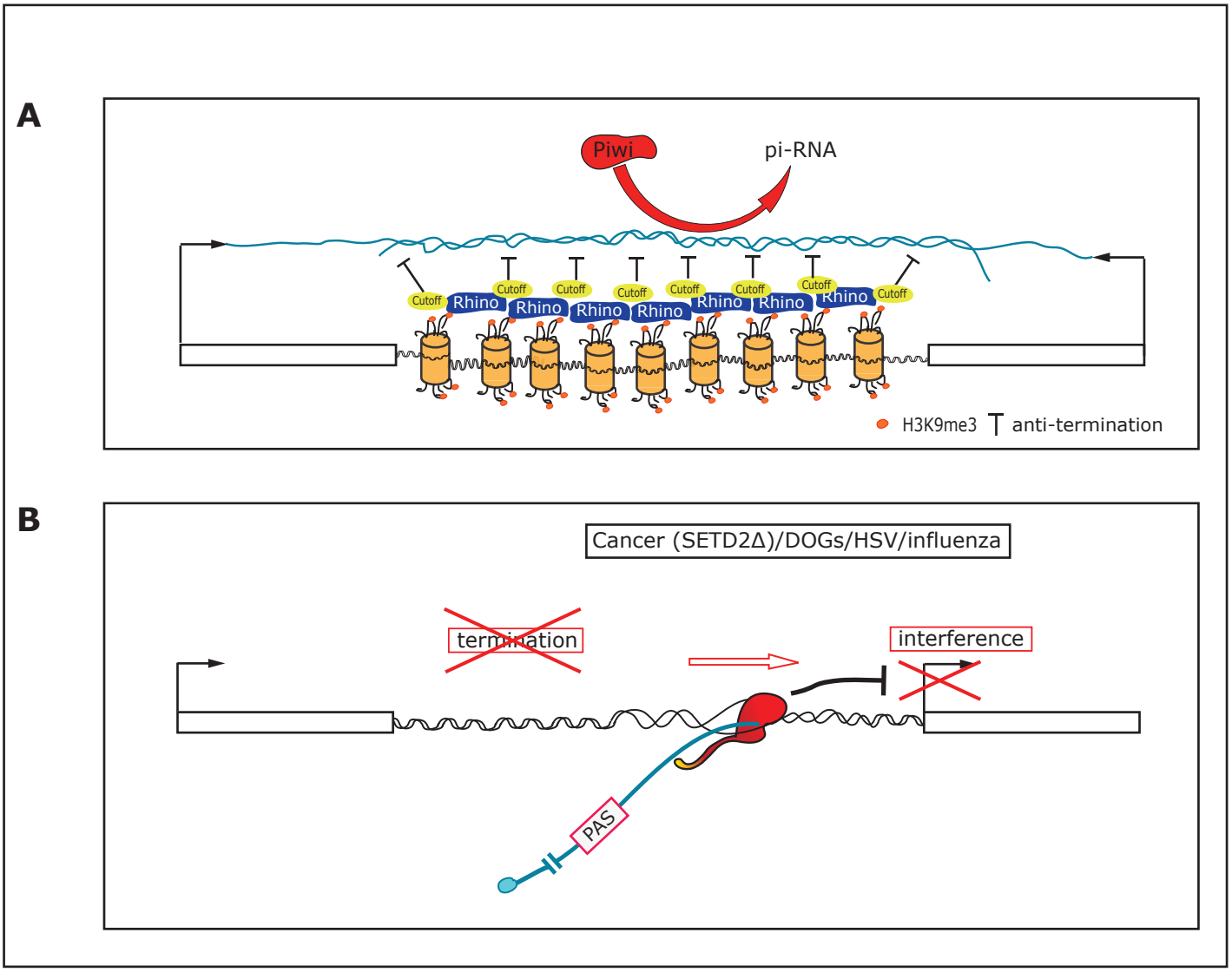

premature TSS termination
pre-PAS termination
TSS
PAS1
PAS2
PAS$_n$
MBNL/ELAV/FUS/CPEBP
CPA

Figure 4


A
Piwi
pi-RNA
Cutoff
Rhino
H3K9me3
anti-termination
B
Cancer (SETD2Δ)/DOGs/HSV/influenza
termination
interference
PAS