PeerJ Preprints NOT PEER-REVIEWED

Transcriptional control by premature termination: a forgotten mechanism

Kinga Kamieniarz-Gdula[1,2,3,*] & Nick J. Proudfoot[1]

[1]Sir William Dunn School of Pathology, University of Oxford, South Parks Road, Oxford, OX1 3RE, UK

[2]Department of Molecular and Cellular Biology, Institute of Molecular Biology and Biotechnology, Faculty of Biology, Adam Mickiewicz University, Umultowska 89, 61-614 Poznań, Poland

[3]Center for Advanced Technology, Adam Mickiewicz University, Umultowska 89c, 61-614 Poznań, Poland

[*]Correspondence: kinga.kamieniarz-gdula@path.ox.ac.uk (K. Kamieniarz-Gdula)

Keywords: premature transcription termination, transcription attenuation, alternative polyadenylation (APA), intronic polyadenylation (IPA), alternative last exon (ALE)

Abstract

The concept of premature termination as an important means of transcriptional control is long established. Even so, its role in metazoan gene expression is underappreciated. Recent technological advances provide novel insights into premature transcription termination (PTT). PTT is very frequent and wide-spread, being either TSS-associated or intragenic. Stable prematurely terminated transcripts contribute to the transcriptome as instances of alternative polyadenylation. Independently of the transcript stability and function, PTT opposes the formation of full-length transcript thereby negatively regulating gene expression. In particular, the expression of many transcriptional regulators is controlled by PTT. PTT can be beneficial or harmful, depending on context. As a result multiple factors have evolved which control this process.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27590v1 | CC BY 4.0 Open Access | rec: 15 Mar 2019, publ: 15 Mar 2019

Transcriptional control by premature termination - revisited

The concept of premature termination as an important means of transcriptional control is long established. Premature termination of transcription (PTT), or 'attenuation', was demonstrated in bacteria as a key regulatory event for the synthesis of enzymes that make amino acids in the mid 1970s [1,2], and first reported to occur for RNA polymerase II (Pol II) transcription of a viral gene in mammalian cells in 1979 [3]. Many more cases of eukaryotic PTT have been identified, even though analysis was hampered by technical limitations and the highly unstable nature of prematurely terminated RNA. The recent development of next-generation sequencing technologies combined with novel methods to measure nascent transcription, single-molecule footprints, and advanced live-imaging, makes it possible to revisit this paradigm. In this review we present recent findings on metazoan PTT revealing its wide-spread nature and role in the regulation of protein coding genes. Our definition of PTT is Pol II release from the gene template between the genes transcription start site (TSS) and 3'UTR. While we focus on metazoans, a broader perspective of PTT in other kingdoms of life is summarized in Box 1. Please note the recent multiple reviews on RNA 3' processing and transcription termination [4–11] and alternative polyadenylation [12–16].

TSS-linked premature transcription termination

PTT of a protein-coding gene can occur close to the TSS or within the gene body (Figure 1). We predict that PTT at these two locations are likely to be functionally and mechanistically different, since they occur at different stages of the transcription cycle.

For many genes, their TSS is characterized by a high accumulation of Pol II, as measured by ChIP. This depends on the action of negative elongation factor (NELF) and DRB sensitivity-inducing factor (DSIF) typically occurring 30-50 bp downsteam of the TSS [17]. Such a Pol II "pileup" is usually interpreted as stable pausing of engaged Pol II. However, TSS-associated Pol II accumulation may also be due to PTT with concomitant Pol II turnover [18]. Several lines of evidence supported this latter possibility: Termination and RNA 3' processing factors have been observed to accumulate at the 5' ends of genes [19,20], short nuclear capped transcripts have been detected [21] and RNA cleavage sites near the TSS were identified [22]. In the past two

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27590v1 | CC BY 4.0 Open Access | rec: 15 Mar 2019, publ: 15 Mar 2019

years, more direct experiments have demonstrated that a high percentage of TSS-bound Pol II molecules terminate prematurely.

A dual-enzyme single-molecule footprinting method (dSMF) has been developed recently to probe Pol II binding to the genome at single molecule resolution, and applied to *Drosophila melanogaster* derived cell lines [23]. This revealed unexpectedly high levels of Pol II turnover at the promoters of paused genes. In particular, changes in Pol II occupancy upon inhibiting initiation revealed that Pol II half-life at model paused gene promoters, including *Hsp70*, was comparable to Pol II half-life at "non-paused", normally elongating genes. This suggests that Pol II accumulation at these gene promoters is a consequence of PTT, rather than stable pausing of transcription-competent polymerases. This interpretation is further supported by an independent study that generated GFP-RPB1 knock-in cells and analyzed the real time dynamics of Pol II in live human cells [24]. Fluorescence recovery after photobleaching (FRAP) experiments combined with the use of various transcriptional inhibitors revealed the existence of four kinetically distinct Pol II fractions: freely diffusing, or chromatin-bound as initiating, promoter-paused and elongating Pol II. Monte Carlo-based computational modelling of Pol II kinetics in unperturbed cells showed that initiating Pol II remains chromatin-bound for only 2.4 s and promoter-paused Pol II for merely 42 s. This is in contrast to elongating Pol II which remained chromatin-bound on average for 23 min. These big differences in Pol II residence times suggest that only a small fraction of initiating and pausing Pol II can proceed through a complete transcription cycle. It follows that most initiating and pausing Pol II is released from chromatin at the promoter. Indeed, the rate constants of the different Pol II fractions show that only 1 in 8 Pol II molecules that initiate transcription will go on to promoter pausing, and only 1 in 13 promoter-paused Pol II molecules continue into productive elongation. Taken together, this study indicates that 99% of transcription initiation events result in PTT at the promoter, with only 1% giving rise to mRNA [24]. This surprisingly inefficient transcription initiation process is consistent with previous Pol II measurements on a lacO array [25]. Furthermore, PTT is the most plausible explanation for the dramatic increase of promoter-associated Pol II within 2-3 mins after H_2O_2 addition to U2OS cells [26]. As a clever way to investigate this issue, the differential sensitivity of transcription initiation and elongation to high ionic strength was used [27]. This showed that blocking recruitment of Pol II to promoters (but not elongation) by high salt treatment affects its binding in ChIP-seq, and revealed an almost complete loss of Pol II from

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27590v1 | CC BY 4.0 Open Access | rec: 15 Mar 2019, publ: 15 Mar 2019

promoter-proximal pause sites within 2-5 minutes. This loss was rapidly reversible, and unaffected by transcriptional inhibitors such as DRB and α-amanitin. Therefore, Pol II removal from pause sites appears not to require elongation. Instead, a high rate of assembly and eviction of preinitiated Pol II complexes at TSS is predicted. These various studies used very different methodologies such as genome-wide single-molecule footprinting, FRAP, ChIP-seq, PRO-seq, or walk-on modification of the nuclear run-on procedure. Even so they all describe very high turnover rates of Pol II at *Drosophila* or human promoters in various cell types. It is therefore unlikely that observed turnover rates are an artefact from any one procedure, or the analysis of an unusual cell type. In conclusion, most initiating Pol II molecules terminate prematurely. It follows that the release of Pol II into productive elongation may be regulated by inhibition of this promoter proximal Pol II termination.

Several previous studies described much longer median half-lives of paused Pol II. One possible explanation for this discrepancy is that typically triptolide was used in these studies to block initiation. This blocks open complex formation by inhibition of TFIIH-associated XPB ATPase [28,29] and it was assumed this prevents recruitment of stable Pol II complexes at promoters. However, triptolide disturbs transcriptional regulation and Pol II stability. Additionally there is a lag in the onset of XBP inhibition which may prevent accurate half-life determination using this drug [24,27]. Further studies using different drugs and methods are required to resolve this discrepancy and so determine the relative contribution of promoter-proximal Pol II pausing versus PTT, to Pol II occupancy at promoters and to control of productive elongation. Furthermore, the mechanism of TSS-associated PTT remains unclear and in particular whether it requires RNA cleavage.

Intragenic premature termination of transcription

Once Pol II overcomes the TSS-associated elongation checkpoint, subsequent PTT likely requires a co-transcriptional cleavage reaction at a cryptic polyadenylation site (PAS), mainly located in introns. PAS mediated cleavage within internal exons appears a small fraction compared to introns, and leads to transcripts without a stop codon, which typically get rapidly degraded through the non-stop decay pathway [30], although in rare cases can result in truncated proteins [31]. PTT frequently occurs in the beginning of the first intron [32–34], partially

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27590v1 | CC BY 4.0 Open Access | rec: 15 Mar 2019, publ: 15 Mar 2019

coinciding with a second Pol II stalling event at stable nucleosomes downstream of CpG island promoters [35]. This typically leads to unstable transcripts that undergo rapid degradation (see below). When premature cleavage occurs downstream of the second exon, it is more likely to be accompanied by productive 3' processing, including polyadenylation and stabilization of the transcript. Stable premature polyadenylated transcripts can be readily detected by various RNA sequencing methodologies and are classified as instances of alternative polyadenylation (APA), reviewed in [14]. APA within the gene body, upstream of the 3' UTR has been referred to as upstream region APA (UR-APA), alternative last exon (ALE) or intronic polyadenylation (IPA or IpA). We will use the IPA term. IPA had been already described for specific genes in 1980 [36–38], but only recently has its widespread nature and significance been appreciated. Thus about 40% of murine genes [39] and 16% of genes in human immune cells [40] have been reported to express IPA isoforms. In particular, IPA is shown to be a frequent genome-wide event, which plays diverse roles in immune cells [41], inactivation of tumour suppressor genes [42] and regulating DNA repair genes [43]. Notably, a systematic analysis of IPA in normal human tissues, primary immune cells and multiple myeloma samples has been used to create an atlas of 4927 high-confidence IPA events in these cell types [41]. IPA isoforms were shown to yield stable transcripts, which in the case of 5'- proximal IPA tend to produce ncRNAs. In contrast, more 3'-proximal IPA events tend to produce truncated proteins [41].

Fate and function of prematurely terminated transcripts

PTT may lead to three outcomes: rapid transcript degradation, formation of a more stable ncRNA, or production of a protein-coding, polyadenylated mRNA isoform (Figure 2). The more 5'-proximal a transcript is terminated, the more likely it is to be degraded. TSS-linked PTT might not require cleavage and could lead to unprocessed, unprotected poly(A)- RNA. Further downstream, transcripts may be cleaved and at least partially polyadenylated. Short transcripts appear to generally undergo quick degradation, similar to PROMPTs/upstream antisense RNAs, which also often yield exosome-sensitive RNAs [44,45,22,46]. In the case of short sense transcripts, while PAS usage allows for transcript cleavage, this might not lead to efficient polyadenylation. Another explanation for the instability of short transcripts is that the 3' entry site for the nuclear exosome in such transcripts is physically close to the 5' capped end of the

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27590v1 | CC BY 4.0 Open Access | rec: 15 Mar 2019, publ: 15 Mar 2019

transcript which is bound by the cap-binding complex (CBC), known to contribute to exosomal decay, especially of some classes of ncRNAs [34,47,48].

Stable prematurely terminated RNAs are presumably mainly polyadenylated, and contribute to the IPA transcriptome. They can be further subdivided into ncRNAs and protein-coding mRNA isoforms (Figure 2). Some ncRNAs contain an ORF, therefore might produce a micropeptide (<100AA). Other ncRNAs might serve specific functions. For example, PTT of *ASCC3* leads to the formation of a ncRNA which is critical for recovery of transcription following UV damage [49]. IPA-generated ncRNAs also bind to RNA-binding proteins, which are normally enriched in the 3'UTRs of coding transcripts, such as FUS, ELAVL1, PUM2, TAF15 and TIAL1 [41]. Such ncRNA could potentially act as scaffolds/sponges for RBPs and so regulate other RNAs in *trans*.

In the third scenario, protein-coding IPA isoforms (Figure 2, bottom) contribute to the diversity of the proteome, for example by generating proteins with physiologically distinct functions, or truncated proteins with dominant negative functions. In the classic case of B cell expressed immunoglobulin M (IgM) heavy chain mRNA, cellular activation causes a switch from full-length to IPA mRNA isoforms, resulting in a change from membrane-bound to secreted forms of that antibody [36–38]. IPA appears to regulate membrane anchoring properties of many other proteins. Computational analysis revealed that 376 mouse genes are likely to use the IPA/ALE mechanism to generate proteins with transmembrane domain (TMD) changes [50]. In human cells, while TMD-containing proteins are significantly depleted among IPA genes, 499 genes encoding transmembrane proteins undergo IPA, and in 152 cases these lead to loss of the TMD [41]. For example, IPA generates various mRNA isoforms of the transmembrane T-cell co-stimulator CD46, which are predicted to form soluble CD46 [51]. Similarly, prematurely polyadenylated transcript variants encoding proteins with dominant negative functions, such as retinoblastoma-binding protein 6 (RBBP6) [52,53], MAGI3 [54,55] or platelet-derived growth factor receptor α (PDGFR α) have been also described [56]. Interestingly, the site of premature polyadenylation is frequently located within a domain mediating either protein-protein interactions, DNA- or RNA binding, such as zinc finger arrays [41]. The partial loss of such interaction surfaces may lead to altered binding affinity to protein interaction partners and altered nucleic acid-binding specificity, respectively. Therefore, IPA may be physiologically relevant by allowing a diversification of protein function. Alternatively the widespread pathological use of

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27590v1 | CC BY 4.0 Open Access | rec: 15 Mar 2019, publ: 15 Mar 2019

premature polyadenylation has been uncovered in chronic lymphocytic leukaemia (CLL) [42]. Here, mRNA truncations by IPA are recurrent and predominantly affect genes with tumour-suppressive functions. This leads either to their inactivation, or to transformation into oncogenes. In conclusion, proteome alteration occurring as a result of IPA can be either beneficial or harmful.

Gene regulation by premature transcription termination

Regardless of whether the prematurely terminated transcript is stable or unstable, and independently of its ability to produce a functional ncRNA or protein, a potential outcome of early termination is the repression of the corresponding full-length mRNA (Figure 3). Such negative regulation had been demonstrated for the *CSTF3* gene (known also as *CSTF77*), originally in *Drosophila* and then in humans [57–59]. CSTF3 is a CPA complex subunit, stimulating cleavage. Premature CPA of *CSTF3* is induced by high levels of CSTF3 protein, thus forming a negative feedback loop controlling its own activity, which is important for cell cycle control [59]. The CSTF3 example raised the possibility that PTT could be a more general mechanism to repress gene expression in metazoans.

Supporting this possibility, the highly abundant U1 snRNP has been shown to function not only in splicing, but also by blocking widespread PTT in thousands of vertebrate genes [32,33]. Vertebrate genes often contain large introns, with numerous cryptic polyadenylation signals. They are therefore inherently susceptible to PTT, although this is suppressed by U1. Interestingly, PTT is strongly activated upon UV treatment of cells, and correlates with decreased U1 snRNA levels [60], and a slowdown in transcription elongation [49]. The above studies demonstrated the widespread predisposition of vertebrate genes to IPA. However, it wasn't clear whether PTT serves a regulatory role genome-wide, or is a harmful genomic accident, induced by DNA damage, and suppressed by U1. In view of the high abundance of U1 snRNP at about 1 million copies per human cell, it is difficult to reconcile how it could serve as a natural regulator of PTT.

Recently, we have established in human cells, as well as during zebrafish embryogenesis, that a subset of protein-coding genes are downregulated by PTT under physiological conditions. This is

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27590v1 | CC BY 4.0 Open Access | rec: 15 Mar 2019, publ: 15 Mar 2019

triggered by a 3' processing and termination factor called PCF11 [61]. Like CSTF3, PCF11 employs PTT as an autoregulation mechanism [61,62]. However, we have additionally identified 218 other human genes, which are attenuated by PCF11-mediated PTT. We predict that this is likely an underestimate, because many genes undergoing PCF11-mediated PTT might also be dependent on PCF11 for efficient full-length transcript expression. Interestingly, half of PCF11-attenuated genes show PTT without detectable IPA products [61]. This suggests that many transcripts attenuated by PTT under physiological conditions are unstable, and that IPA corresponds to only a subset of intragenic PTT. Therefore, it will be important that future PTT studies also assay the nascent transcriptome. Notably, the genes undergoing PCF11-mediated PTT are enriched for transcriptional regulators, both in humans and in zebrafish [61]. Similarly, in primary human tissues, IPA has also been found to occur preferentially on genes encoding transcriptional regulators [41]. We have also found that PCF11 levels are an order of magnitude lower compared to other 3' processing factors, and vary between tissues, making PCF11 a likely regulatory factor [61]. Overall, our data indicate that PTT is a naturally occurring, wide-spread and controlled phenomenon in vertebrates.

The role of CPA and termination factors in premature termination

Since PCF11 acts at both gene ends and throughout the gene body, it appears likely that other canonical CPA and termination factors are involved in PTT.

Notably, XRN2, the nuclear 5′–3′ exonuclease "torpedo" that facilitates transcription termination at the 3′ ends of genes, has been shown by ChIP-seq to also localize near TSS, and interact with decapping factors [19]. Coupled decapping of nascent transcripts and PTT has been suggested to limit bidirectional Pol II elongation. We found PCF11 as well as CPSF73 enrichment at TSS in ChIP-seq [61]. Additionally, in vivo cross-linking and immunoprecipitation (CLIP) revealed that all analyzed CPA factors (CPSF73, CstF64, CstF64t, CPSF160, CPSF30, and CFIm25) were significantly detected on both strands within 500 nt of the TSS [63,64]. Finally, depletion of CPA and termination factors (CPSF73, CSTF64, XRN2 and PCF11) resulted in increased mNET-seq signal specifically at the TSS, in both sense and antisense direction [64,61]. This suggests that promoter-associated non-productive RNA synthesis is terminated and thereby

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27590v1 | CC BY 4.0 Open Access | rec: 15 Mar 2019, publ: 15 Mar 2019

controlled by the same factors that are responsible for 3' processing and transcription termination at gene ends.

The involvement of the termination machinery in premature cleavage within the gene body is less well defined. However, IPA globally correlates with the preferential use of proximal PASs in 3'UTRs [65,39,66,41], which are both more prevalent in proliferating cells but less so during cell differentiation. Consequently, IPA and 3'UTR APA are likely to be co-regulated. Beside PCF11, FIP1 is a likely candidate for PTT stimulation, as it is the only other CPA factor known to promote early PAS usage [67].

Non-canonical factors involved in premature termination

Besides the factors that are known to mediate 3' processing and transcription termination at gene ends, others have emerged that stimulate PTT. In budding yeast, premature termination is tightly linked with RNA degradation by the nuclear exosome [48]. Similarly, mutation of the catalytic subunit of the human nuclear exosome complex, DIS3, resulted in accumulation of truncated RNAs, presumably PTT products [68]. Furthermore, the mammalian nuclear exosome interacts with the cap-binding complex (CBC). Two CBC-associated proteins, ZC3H18 and SRRT/ARS2, have also been connected with PTT events in the first introns of protein-coding genes [35].

Another candidate factor for PTT is the Integrator complex (INT). This multiprotein complex interacts with Pol II and was initially identified as a RNA 3' processing complex for non-coding small nuclear RNAs [69,70]. Two INT subunits, IntS9 and IntS11, are homologous to CPA factors CPSF100 and CPSF73, respectively [71]. More recently, experimental evidence emerged which suggests that INT plays also a role in the activation of protein-coding genes, particularly in the Pol II pause-release and elongation [72–75]. The subunit containing endonuclease activity, Ints11, was found to bind around TSS and to be required for TSS-proximal Pol II pause-release [73]. This function was dependent on its catalytic activity. It is likely that the Integrator complex is involved in TSS-associated PTT.

In terms of the genetic elements that promote PTT, cryptic poly(A) signals within introns are likely candidates. Interestingly, intragenic enhancers also appear to stimulate PTT, as

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27590v1 | CC BY 4.0 Open Access | rec: 15 Mar 2019, publ: 15 Mar 2019

transcription at these enhancers interferes with and attenuates host gene transcription during productive elongation [76]. Possibly further genetic elements exist that can trigger PTT.

Factors opposing premature termination

Since PTT is disruptive to full-length transcription, many mechanisms and factors must exist to suppress this process. A major factor in preventing PTT is U1 snRNP, as described above. Furthermore, DNA sequences likely evolved to prevent harmful yet reinforce beneficial termination events. Around the TSS, poly(A) signals are depleted in the sense direction relative to the upstream antisense direction, while U1 snRNP recognition sites show the opposite pattern. This sequence asymmetry has been proposed to control promoter directionality [22,46]. It has been also reported that codon usage biases co-evolve with transcription termination machinery to suppress PTT and allow optimal gene expression [77]. Termination inhibition can also occur in a chromatin-dependent, and sequence-independent manner, as shown in *Drosophila* for piRNA transcription regulator Cutoff. Cutoff prevents cleavage of nascent RNA at PAS by interfering with recruitment of the CPA complex, and also protects processed transcripts from degradation [78]. Also RNA-binding proteins can protect against PTT. For example in neural progenitor cells, Sam68 binding to an intronic poly(A) site in *Aldh1a3* prevents its recognition and consequent PTT. This promotes cell self-renewal [79]. Furthermore, a negative correlation between N6-methyladenosine (m^6A) modification of mRNA and PTT has been reported [55].

Modifying Pol II activity may be another way to oppose PTT. When Pol II mis-incorporates a base or runs into a “roadblock” like a nucleosome, it arrests and backtracks. Pol II elongation is rescued by TFIIS stimulating the RNA endonuclease activity of Pol II, which produces a new 3’ end in the active site. It has recently been shown that TFIIS stimulation of Pol II cleavage activity antagonizes both premature TSS-proximal and gene end termination [80]. Pol II is further regulated by phosphorylation of the CTD of its largest subunit. Depletion of the CTD kinase CDK12 results in genome-wide increased IPA, indicating that CDK12 suppresses PTT [43]. Interestingly, many DNA repair genes harbour more IPA sites than other expressed genes, and are particularly sensitive to loss of CDK12. IPA is also opposed by CPA factors which favour distal APA events, especially the CFIm complex subunits, PABPN1 and PABPC1 [67,81].

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27590v1 | CC BY 4.0 Open Access | rec: 15 Mar 2019, publ: 15 Mar 2019

Since premature cleavage and polyadenylation are mutually exclusive with splicing of the intron in which the PAS is located, a competition between splicing and PTT can be anticipated. Indeed, knock-down of various splicing factors as well as inhibition of 5' splice site recognition by antisense nucleotide consistently results in IPA [67,42]. Therefore, antagonism between splicing and PTT goes beyond U1 snRNP-mediated PAS blocking as described above. Further to splicing, Pol II elongation rates have been shown to influence PTT, as IPA events increase under conditions where Pol II elongation is slowed down [82–84]. Slower Pol II provides a longer window of opportunity for CPA before the 3' splice site is reached. Hence changes in the balance between splicing and PTT, together with Pol II pausing predisposes it to termination.

Concluding remarks and future perspectives

We have outlined recent evidence that demonstrates the widespread occurrence of PTT in protein-coding genes, both TSS-proximal and intragenic. PTT limits and opposes full-length transcription. Consequently, it may contribute to pathological processes, such as host gene downregulation during viral infection [85], and to carcinogenesis by inactivation of tumour suppressor genes [42]. PTT is globally suppressed by genic DNA sequence biases, and various cellular factors, such as U1 snRNP. On the other hand, PTT can be also beneficial: it diversifies the transcriptome and proteome, and contributes to gene regulation. Gene regulation by PTT occurs in at least four kingdoms of life: Eubacteria, Fungi, Plantae, and Animalia (Box 1). Yet, the regulatory aspect of PTT has been largely overlooked in metazoan research. This is partially due to the fact that the machinery triggering PTT varies in different organisms. The well researched NNS complex regulating PTT in *S. cerevisiae* is not only absent in plants and animals, but even in less evolutionarily distant *S. pombe*. Even so, the mechanisms modulating PTT appear conserved. In budding yeast NNS cooperates with PCF11 [86], and both NNS and PCF11 levels are controlled by PTT [87,88]. In vertebrates, PCF11 employs PTT to regulate own levels, as well as the levels of other 3' processing factors and transcriptional regulators [61]. Therefore, vertebrate PCF11 may have at least partially taken over the function of the yeast NNS complex. We anticipate that PTT in metazoans is a critical feature of gene regulation. It is also possible that this process can be manipulated to achieve clinical and biotechnological benefit (Box 2).

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27590v1 | CC BY 4.0 Open Access | rec: 15 Mar 2019, publ: 15 Mar 2019

Acknowledgements

We are grateful to Sebastian Marquardt and N.J.P. group members for helpful discussions. This work was supported by Wellcome Trust Investigator Award (107928/Z/15/Z), ERC Advanced Grant (339170) to N.J.P., and Marie Curie fellowship from EU FP7 (327985) to K.K.-G.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27590v1 | CC BY 4.0 Open Access | rec: 15 Mar 2019, publ: 15 Mar 2019

References

1 Bertrand, K. *et al.* (1975) New features of the regulation of the tryptophan operon. *Science* 189, 22–26

2 Artz, S.W. and Broach, J.R. (1975) Histidine regulation in Salmonella typhimurium: an activator attenuator model of gene regulation. *Proc. Natl. Acad. Sci. U.S.A.* 72, 3453–3457

3 Evans, R. *et al.* (1979) Premature termination during adenovirus transcription. *Nature* 278, 367–370

4 Di Giammartino, D.C. and Manley, J.L. (2014) New links between mRNA polyadenylation and diverse nuclear pathways. *Mol. Cells* 37, 644–649

5 Lemay, J.-F. and Bachand, F. (2015) Fail-safe transcription termination: Because one is never enough. *RNA Biol* 12, 927–932

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

7 Loya, T.J. and Reines, D. (2016) Recent advances in understanding transcription termination by RNA polymerase II. *F1000Res* 5,

8 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* advance online publication,

9 Porrua, O. *et al.* (2016) Transcription Termination: Variations on Common Themes. *Trends Genet.* 32, 508–522

10 Proudfoot, N.J. (2016) Transcriptional termination in mammals: Stopping the RNA polymerase II juggernaut. *Science* 352, aad9926

11 Shi, Y. and Manley, J.L. (2015) The end of the message: multiple protein-RNA interactions define the mRNA polyadenylation site. *Genes Dev.* 29, 889–897

12 Mayr, C. (2016) Evolution and Biological Roles of Alternative 3′UTRs. *Trends in Cell Biology* 26, 227–237

13 Neve, J. and Furger, A. (2014) Alternative polyadenylation: less than meets the eye? *Biochem. Soc. Trans.* 42, 1190–1195

14 Tian, B. and Manley, J.L. (2017) Alternative polyadenylation of mRNA precursors. *Nat. Rev. Mol. Cell Biol.* 18, 18–30

15 Turner, R.E. *et al.* (2018) Alternative polyadenylation in the regulation and dysregulation of gene expression. *Seminars in Cell & Developmental Biology* 75, 61–69

16 Zheng, D. and Tian, B. (2014) RNA-binding proteins in regulation of alternative cleavage and polyadenylation. *Adv. Exp. Med. Biol.* 825, 97–127

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

18 Ehrensberger, A.H. *et al.* (2013) Mechanistic interpretation of promoter-proximal peaks and RNAPII density maps. *Cell* 154, 713–715

19 Brannan, K. *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

20 Wagschal, A. *et al.* (2012) Microprocessor, Setx, Xrn2, and Rrp6 co-operate to induce premature termination of transcription by RNAPII. *Cell* 150, 1147–1157

21 Nechaev, S. *et al.* (2010) Global analysis of short RNAs reveals widespread promoter-proximal stalling and arrest of Pol II in Drosophila. *Science* 327, 335–338

22 Almada, A.E. *et al.* (2013) Promoter directionality is controlled by U1 snRNP and polyadenylation signals. *Nature* 499, 360–363

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27590v1 | CC BY 4.0 Open Access | rec: 15 Mar 2019, publ: 15 Mar 2019

23 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

24 Steurer, B. *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

25 Darzacq, X. *et al.* (2007) In vivo dynamics of RNA polymerase II transcription. *Nat. Struct. Mol. Biol.* 14, 796–806

26 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

27 Erickson, B. *et al.* (2018) Dynamic turnover of paused Pol II complexes at human promoters. *Genes Dev.* 32, 1215–1225

28 Titov, D.V. *et al.* (2011) XPB, a subunit of TFIIH, is a target of the natural product triptolide. *Nat. Chem. Biol.* 7, 182–188

29 Alekseev, S. *et al.* (2017) Transcription without XPB Establishes a Unified Helicase-Independent Mechanism of Promoter Opening in Eukaryotic Gene Expression. *Mol. Cell* 65, 504-514.e4

30 Vasudevan, S. *et al.* (2002) Non-stop decay—a new mRNA surveillance pathway. *BioEssays* 24, 785–788

31 Yao, P. *et al.* (2012) Coding region polyadenylation generates a truncated tRNA synthetase that counters translation repression. *Cell* 149, 88–100

32 Kaida, D. *et al.* (2010) U1 snRNP protects pre-mRNAs from premature cleavage and polyadenylation. *Nature* 468, 664–668

33 Venters, C.C. *et al.* (2019) U1 snRNP Telescripting: Suppression of Premature Transcription Termination in Introns as a New Layer of Gene Regulation. *Cold Spring Harb Perspect Biol* 11,

34 Iasillo, C. *et al.* (2017) ARS2 is a general suppressor of pervasive transcription. *Nucleic Acids Res.* 45, 10229–10241

35 Chiu, A.C. *et al.* (2018) Transcriptional Pause Sites Delineate Stable Nucleosome-Associated Premature Polyadenylation Suppressed by U1 snRNP. *Mol. Cell* 69, 648-663.e7

36 Alt, F.W. *et al.* (1980) Synthesis of secreted and membrane-bound immunoglobulin mu heavy chains is directed by mRNAs that differ at their 3' ends. *Cell* 20, 293–301

37 Early, P. *et al.* (1980) Two mRNAs can be produced from a single immunoglobulin mu gene by alternative RNA processing pathways. *Cell* 20, 313–319

38 Rogers, J. *et al.* (1980) Two mRNAs with different 3' ends encode membrane-bound and secreted forms of immunoglobulin mu chain. *Cell* 20, 303–312

39 Hoque, M. *et al.* (2013) Analysis of alternative cleavage and polyadenylation by 3′ region extraction and deep sequencing. *Nat Meth* 10, 133–139

40 Lianoglou, S. *et al.* (2013) Ubiquitously transcribed genes use alternative polyadenylation to achieve tissue-specific expression. *Genes Dev.* 27, 2380–2396

41 Singh, I. *et al.* (2018) Widespread intronic polyadenylation diversifies immune cell transcriptomes. *Nat Commun* 9, 1716

42 Lee, S.-H. *et al.* (2018) Widespread intronic polyadenylation inactivates tumour suppressor genes in leukaemia. *Nature* 561, 127

43 Dubbury, S.J. *et al.* (2018) CDK12 regulates DNA repair genes by suppressing intronic polyadenylation. *Nature* 564, 141–145

44 Preker, P. *et al.* (2008) RNA Exosome Depletion Reveals Transcription Upstream of Active Human Promoters. *Science* 322, 1851–1854

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27590v1 | CC BY 4.0 Open Access | rec: 15 Mar 2019, publ: 15 Mar 2019

45 Seila, A.C. *et al.* (2008) Divergent transcription from active promoters. *Science* 322, 1849–1851

46 Ntini, E. *et al.* (2013) Polyadenylation site–induced decay of upstream transcripts enforces promoter directionality. *Nat Struct Mol Biol* 20, 923–928

47 Andersen, P.R. *et al.* (2013) The human cap-binding complex is functionally connected to the nuclear RNA exosome. *Nat Struct Mol Biol* 20, 1367–1376

48 Schmid, M. and Jensen, T.H. (2018) Controlling nuclear RNA levels. *Nature Reviews Genetics* 19, 518

49 Williamson, L. *et al.* (2017) UV Irradiation Induces a Non-coding RNA that Functionally Opposes the Protein Encoded by the Same Gene. *Cell* 168, 843-855.e13

50 Davis, M.J. *et al.* (2006) Differential use of signal peptides and membrane domains is a common occurrence in the protein output of transcriptional units. *PLoS Genet.* 2, e46

51 Ly, P.T. *et al.* (2017) Alternative polyadenylation expands the mRNA isoform repertoire of human CD46. *Gene* 625, 21–30

52 Mbita, Z. *et al.* (2012) De-regulation of the RBBP6 isoform 3/DWNN in human cancers. *Mol Cell Biochem* 362, 249–262

53 Giammartino, D.C.D. *et al.* (2014) RBBP6 isoforms regulate the human polyadenylation machinery and modulate expression of mRNAs with AU-rich 3′ UTRs. *Genes Dev.* 28, 2248–2260

54 Ni, T.K. and Kuperwasser, C. (2016) Premature polyadenylation of MAGI3 produces a dominantly-acting oncogene in human breast cancer. *Elife* 5,

55 Ni, T.K. *et al.* (2018) Premature polyadenylation of MAGI3 is associated with diminished N6-methyladenosine in its large internal exon. *Sci Rep* 8, 1415

56 Mueller, A.A. *et al.* (2016) Intronic polyadenylation of PDGFRα in resident stem cells attenuates muscle fibrosis. *Nature* 540, 276–279

57 Audibert, A. and Simonelig, M. (1998) Autoregulation at the level of mRNA 3' end formation of the suppressor of forked gene of Drosophila melanogaster is conserved in Drosophila virilis. *Proc. Natl. Acad. Sci. U.S.A.* 95, 14302–14307

58 Pan, Z. *et al.* (2006) An intronic polyadenylation site in human and mouse CstF-77 genes suggests an evolutionarily conserved regulatory mechanism. *Gene* 366, 325–334

59 Luo, W. *et al.* (2013) The Conserved Intronic Cleavage and Polyadenylation Site of CstF-77 Gene Imparts Control of 3′ End Processing Activity through Feedback Autoregulation and by U1 snRNP. *PLOS Genetics* 9, e1003613

60 Devany, E. *et al.* (2016) Intronic cleavage and polyadenylation regulates gene expression during DNA damage response through U1 snRNA. *Cell Discov* 2, 16013

61 Kamieniarz-Gdula, K. *et al.* (2019) Selective Roles of Vertebrate PCF11 in Premature and Full-Length Transcript Termination. *Molecular Cell* 0,

62 Wang, R. *et al.* (2019) Regulation of Intronic Polyadenylation by PCF11 Impacts mRNA Expression of Long Genes. *Cell Reports* 26, 2766-2778.e6

63 Martin, G. *et al.* (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

64 Nojima, T. *et al.* (2015) Mammalian NET-Seq Reveals Genome-wide Nascent Transcription Coupled to RNA Processing. *Cell* 161, 526–540

65 Elkon, R. *et al.* (2012) E2F mediates enhanced alternative polyadenylation in proliferation. *Genome Biology* 13, R59

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27590v1 | CC BY 4.0 Open Access | rec: 15 Mar 2019, publ: 15 Mar 2019

66 Taliaferro, J.M. *et al.* (2016) Distal Alternative Last Exons Localize mRNAs to Neural Projections. *Molecular Cell* 61, 821–833
67 Li, W. *et al.* (2015) 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
68 Szczepińska, 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
69 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
70 Rienzo, M. and Casamassimi, A. (2016) Integrator complex and transcription regulation: Recent findings and pathophysiology. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* 1859, 1269–1280
71 Dominski, Z. *et al.* (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
72 Gardini, A. *et al.* (2014) Integrator Regulates Transcriptional Initiation and Pause Release following Activation. *Molecular Cell* 56, 128–139
73 Stadelmayer, B. *et al.* (2014) Integrator complex regulates NELF-mediated RNA polymerase II pause/release and processivity at coding genes. *Nat Commun* 5, 5531
74 Skaar, J.R. *et al.* (2015) The Integrator complex controls the termination of transcription at diverse classes of gene targets. *Cell Research* 25, 288–305
75 Baillat, D. and Wagner, E.J. (2015) Integrator: surprisingly diverse functions in gene expression. *Trends in Biochemical Sciences* 40, 257–264
76 Cinghu, S. *et al.* (2017) Intragenic Enhancers Attenuate Host Gene Expression. *Mol. Cell* 68, 104-117.e6
77 Zhou, Z. *et al.* (2018) Codon usage biases co-evolve with transcription termination machinery to suppress premature cleavage and polyadenylation. *Elife* 7,
78 Chen, Y.-C.A. *et al.* (2016) Cutoff Suppresses RNA Polymerase II Termination to Ensure Expression of piRNA Precursors. *Mol. Cell* 63, 97–109
79 La Rosa, P. *et al.* (2016) Sam68 promotes self-renewal and glycolytic metabolism in mouse neural progenitor cells by modulating Aldh1a3 pre-mRNA 3'-end processing. *Elife* 5,
80 Sheridan, R.M. *et al.* (2018) Widespread Backtracking by RNA Pol II Is a Major Effector of Gene Activation, 5' Pause Release, Termination, and Transcription Elongation Rate. *Mol. Cell* DOI: 10.1016/j.molcel.2018.10.031
81 Zhu, Y. *et al.* (2018) Molecular Mechanisms for CFIm-Mediated Regulation of mRNA Alternative Polyadenylation. *Mol. Cell* 69, 62-74.e4
82 Cui, Y. and Denis, C.L. (2003) In vivo evidence that defects in the transcriptional elongation factors RPB2, TFIIS, and SPT5 enhance upstream poly(A) site utilization. *Mol. Cell. Biol.* 23, 7887–7901
83 Yang, Y. *et al.* (2016) PAF Complex Plays Novel Subunit-Specific Roles in Alternative Cleavage and Polyadenylation. *PLoS Genet.* 12, e1005794
84 Liu, X. *et al.* (2017) Transcription elongation rate has a tissue-specific impact on alternative cleavage and polyadenylation in Drosophila melanogaster. *RNA* 23, 1807–1816

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27590v1 | CC BY 4.0 Open Access | rec: 15 Mar 2019, publ: 15 Mar 2019

85 Tang, S. *et al.* (2016) Herpes simplex virus ICP27 regulates alternative pre-mRNA polyadenylation and splicing in a sequence-dependent manner. *Proc. Natl. Acad. Sci. U.S.A.* 113, 12256–12261

86 Grzechnik, P. *et al.* (2015) Pcf11 orchestrates transcription termination pathways in yeast. *Genes Dev.* 29, 849–861

87 Arigo, J.T. *et al.* (2006) Regulation of Yeast NRD1 Expression by Premature Transcription Termination. *Molecular Cell* 21, 641–651

88 Creamer, T.J. *et al.* (2011) Transcriptome-wide binding sites for components of the Saccharomyces cerevisiae non-poly(A) termination pathway: Nrd1, Nab3, and Sen1. *PLoS Genet.* 7, e1002329

89 Bastet, L. *et al.* (2018) Maestro of regulation: Riboswitches orchestrate gene expression at the levels of translation, transcription and mRNA decay. *RNA Biol* 15, 679–682

90 Sherwood, A.V. and Henkin, T.M. (2016) Riboswitch-Mediated Gene Regulation: Novel RNA Architectures Dictate Gene Expression Responses. *Annual Review of Microbiology* 70, 361–374

91 Colin, J. *et al.* (2011) Cryptic transcription and early termination in the control of gene expression. *Genet Res Int* 2011, 653494

92 Arndt, K.M. and Reines, D. (2015) Termination of Transcription of Short Noncoding RNAs by RNA Polymerase II. *Annu. Rev. Biochem.* 84, 381–404

93 Merran, J. and Corden, J.L. (2017) Yeast RNA-Binding Protein Nab3 Regulates Genes Involved in Nitrogen Metabolism. *Mol. Cell. Biol.* 37,

94 Bresson, S. *et al.* (2017) Nuclear RNA Decay Pathways Aid Rapid Remodeling of Gene Expression in Yeast. *Mol. Cell* 65, 787-800.e5

95 Whalen, C. *et al.* (2018) RNA Polymerase II Transcription Attenuation at the Yeast DNA Repair Gene, DEF1, Involves Sen1-Dependent and Polyadenylation Site-Dependent Termination. *G3 (Bethesda)* 8, 2043–2058

96 Quesada, V. *et al.* (2003) Autoregulation of FCA pre-mRNA processing controls Arabidopsis flowering time. *EMBO J.* 22, 3142–3152

97 Hornyik, C. *et al.* (2010) The spen family protein FPA controls alternative cleavage and polyadenylation of RNA. *Dev. Cell* 18, 203–213

98 Liu, F. *et al.* (2010) Targeted 3' processing of antisense transcripts triggers Arabidopsis FLC chromatin silencing. *Science* 327, 94–97

99 Swiezewski, S. *et al.* (2009) Cold-induced silencing by long antisense transcripts of an Arabidopsis Polycomb target. *Nature* 462, 799–802

100 Gong, Q. *et al.* (2014) Upregulation of functional Kv11.1 isoform expression by inhibition of intronic polyadenylation with antisense morpholino oligonucleotides. *J. Mol. Cell. Cardiol.* 76, 26–32

101 Melamed, Z. *et al.* (2019) Premature polyadenylation-mediated loss of stathmin-2 is a hallmark of TDP-43-dependent neurodegeneration. *Nature Neuroscience* DOI: 10.1038/s41593-018-0293-z

102 Gong, Q. and Zhou, Z. (2017) Regulation of Isoform Expression by Blocking Polyadenylation Signal Sequences with Morpholinos. *Methods Mol. Biol.* 1565, 141–150

103 Gao, Y. *et al.* (2016) Genome Therapy of Myotonic Dystrophy Type 1 iPS Cells for Development of Autologous Stem Cell Therapy. *Mol. Ther.* 24, 1378–1387

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27590v1 | CC BY 4.0 Open Access | rec: 15 Mar 2019, publ: 15 Mar 2019

PeerJ Preprints NOT PEER-REVIEWED

104 Araki, S. *et al.* (2018) Decoding Transcriptome Dynamics of Genome-Encoded Polyadenylation and Autoregulation with Small-Molecule Modulators of Alternative Polyadenylation. *Cell Chemical Biology* DOI: 10.1016/j.chembiol.2018.09.006

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27590v1 | CC BY 4.0 Open Access | rec: 15 Mar 2019, publ: 15 Mar 2019

Box 1

PTT in bacteria, yeast and plants

PTT is long known to be a key regulatory event in bacteria, referred to as attenuation. Classically attenuation was shown to control the expression of enzymes involved in amino acid biosynthesis, such as the tryptophan and histidine operons [1,2]. Bacterial terminators can be intrinsic, associated with a hairpin RNA structure, or factor dependent, usually involving the RNA helicase Rho. Attenuation occurs when an antiterminator hairpin RNA forms ahead of an intrinsic terminator positioned near the 5' end of an operon. Formation of the antiterminator hairpin blocks formation the intrinsic terminator hairpin and so allows transcription to read into the operon and express its protein coding regions. Switching between the antiterminator and terminator hairpins is controlled by diverse regulators [89,90]. As translation occurs co-transcriptionally, PTT is closely coupled to translation regulation. This differentiates it from eukaryotic regulation.

PTT is a well-recognized regulatory mechanism in *S. cerevisiae*, mediated by the NNS (Nrd1–Nab3–Sen1) complex. The first example of attenuation by NNS was demonstrated for the *NRD1* gene, which is autoregulated by PTT in response to Nrd1 activity [87]. NNS-mediated PTT further regulates genes involved in nucleotide and amino acid biosynthesis, as well as nitrogen metabolism, and is physiologically relevant upon nutritional shift [91–94]. The prematurely terminated transcripts sometimes initiate at a TSS upstream of the protein-coding gene [91]. Interestingly, it was recently shown that the DNA repair gene *DEF1* is attenuated by Sen1 and CPA factors, without Nrd1 and Nab3 involvement [95], therefore PTT in *S. cerevisiae* might not be limited to the NNS pathway.

There are no Nrd1/Nab3 homologs known in plants. However, PTT plays an elaborate role in the control of flowering time in *Arabidopsis thaliana*. FLC is a transcription factor which acts as a master regulator of flowering. It is carefully titrated – with small changes in *FLC* transcript levels significantly affecting flowering. The accumulation of *FLC* mRNA is prevented by FCA and FPA: two RNA-binding proteins associated with RNA 3' processing factors. FCA and FPA autoregulate their own levels by premature polyadenylation and termination, independently of each other [96,97]. Interestingly, they also promote early termination of the lncRNA *COOLAIR* [98,97]. *COOLAIR* is an antisense transcript to *FLC*, and functions in early cold induced

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27590v1 | CC BY 4.0 Open Access | rec: 15 Mar 2019, publ: 15 Mar 2019

silencing of *FLC* transcription [99]. As a result, several layers of premature termination of coding and non-coding transcripts act to control the timing of plant flowering. Likely other examples of PTT will be established in plants.

Box 2

Implications of PTT in medicine and biotechnology

Aberrant PTT is associated with several pathological processes: inactivation of tumour suppressor genes in chronic lymphocytic leukaemia [42], downregulation of the expression of Kv11.1 potassium channel leading to a heart rhytm condition [100], and loss of neuronal growth-associated factor stathmin-2 in neurodegeneration [101]. Blocking pathological PTT by antisense morpholinos has been suggested as a potential therapeutic strategy [102]. Conversely, induction of PTT to prevent formation of disease-inducing transcripts, has also been proposed as a therapeutic approach. In particular, inserting intronic poly(A) signals upstream of toxic expanded CTG repeats in the *DMPK* gene as associated with myotonic dystrophy type 1 was demonstrated to revert the pathological phenotype of patient-derived iPS cells [103]. Similar approaches could also be used in biotechnology. Of note, two small-molecule modulators of APA have been discovered, which promote distal-to-proximal APA usage [104]. These could prove a useful tool for APA and PTT manipulations.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27590v1 | CC BY 4.0 Open Access | rec: 15 Mar 2019, publ: 15 Mar 2019

Figure 1

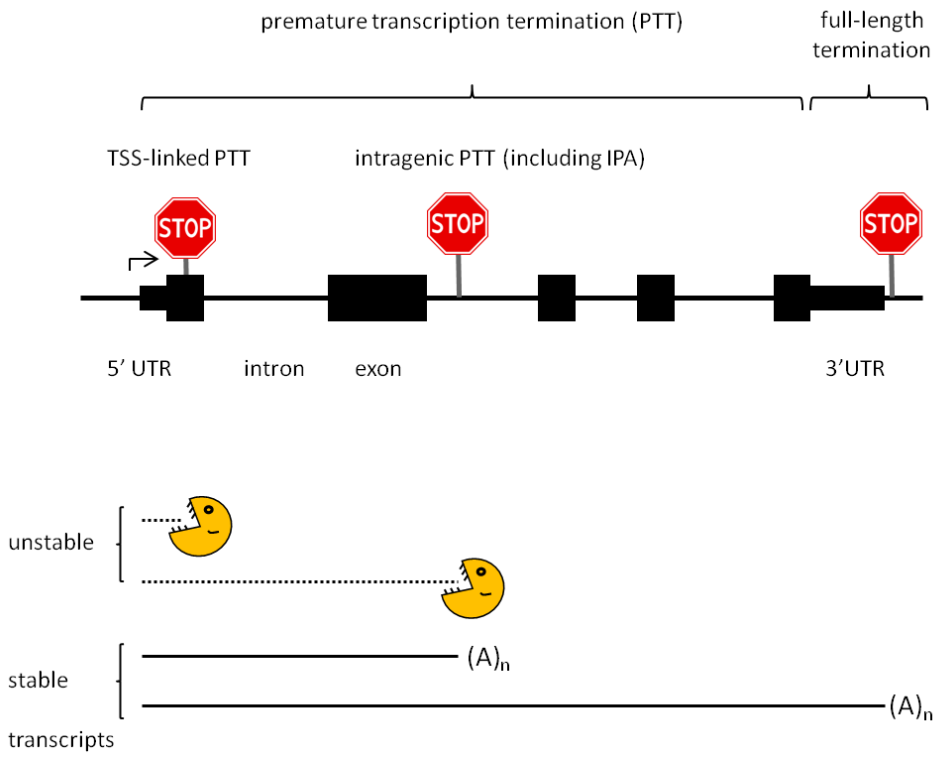

Figure 1. Localization of premature and full-length transcription termination.

Premature transcription termination (PTT) can occur in the vicinity of the transcription start site (TSS) in the case of TSS-linked PTT, or further downstream in the gene body as for intragenic PTT. We classify any transcription termination events occurring within the 3'UTR of the full-length transcript, or downstream of the annotated 3'UTR, as full-length, whereas those occurring between the TSS and before the 3'UTR of the full-length mRNA as premature. PTT can generate unstable (dotted line) or stable (solid line), polyadenylated transcripts. If PTT generates a stable, polyadenylated transcript, it can be classified as alternative polyadenylation (APA), or intronic polyadenylation (IPA). Abbreviations: 5' UTR - 5' untranslated region; 3' UTR - 3' untranslated region; (A)n – polyadenylation.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27590v1 | CC BY 4.0 Open Access | rec: 15 Mar 2019, publ: 15 Mar 2019

Figure 2

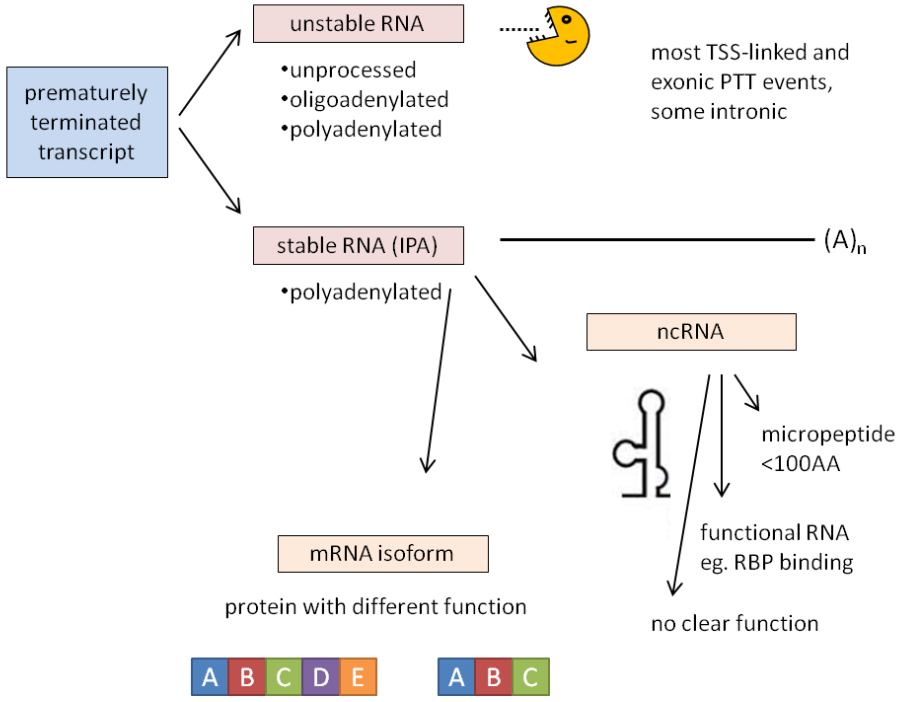

Figure 2. Fate and function of prematurely terminated transcripts.

Most TSS-linked, exonic and some intronic prematurely terminated transcripts are unstable, and presumably lacking cellular function. Stable prematurely terminated transcript generated as a result of intronic polyadenylation (IPA) form either a non-coding (nc) RNA or protein-coding mRNA. Contrary to their name, ncRNAs sometimes contain small open reading frames which may be translated into micropeptides. Other ncRNAs can serve cellular functions, for example as scaffolds for RNA binding-proteins (RBP) [41]. Many ncRNAs have no clear function determined to date. Protein-coding mRNA isoforms generated by PTT diversify the proteome. They lack the C-terminal domain(s) present in the full-length protein, and may have different properties. This includes membrane binding versus soluble [36–38,41,50], altered specificity or affinity for binding to nucleic acid or protein partners [41], and in some cases dominant negative functions [52–56].

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27590v1 | CC BY 4.0 Open Access | rec: 15 Mar 2019, publ: 15 Mar 2019

Figure 3

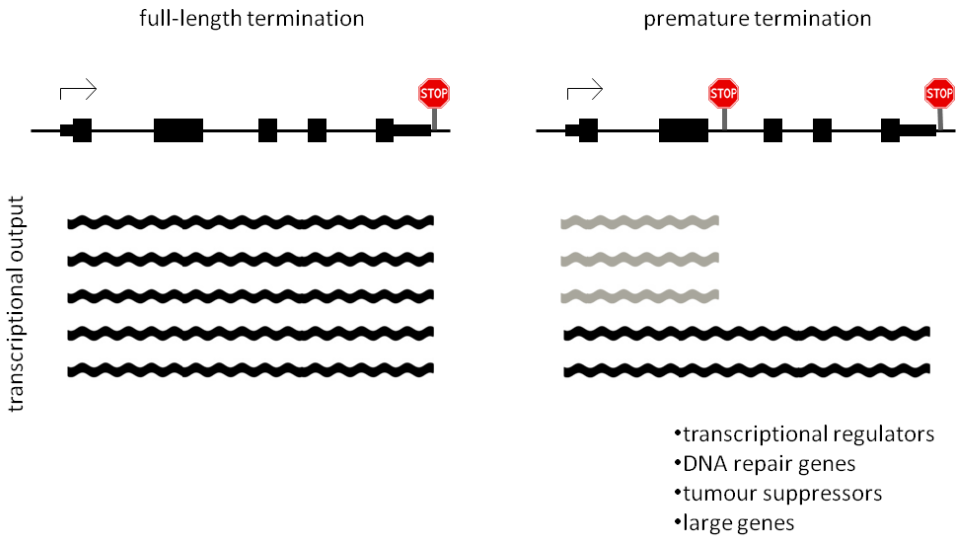

Figure 3. Transcriptional control by premature termination.

PTT is mutually exclusive with full-length transcription. Independent of the fate or function of prematurely terminated transcripts, PTT negatively regulates the expression of the full-length transcripts. Genes controlled by PTT include transcriptional regulators [41,61], DNA repair genes [43], tumour suppressor genes [42], and tend to be larger than average genes [33,62].

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27590v1 | CC BY 4.0 Open Access | rec: 15 Mar 2019, publ: 15 Mar 2019

Figure 4

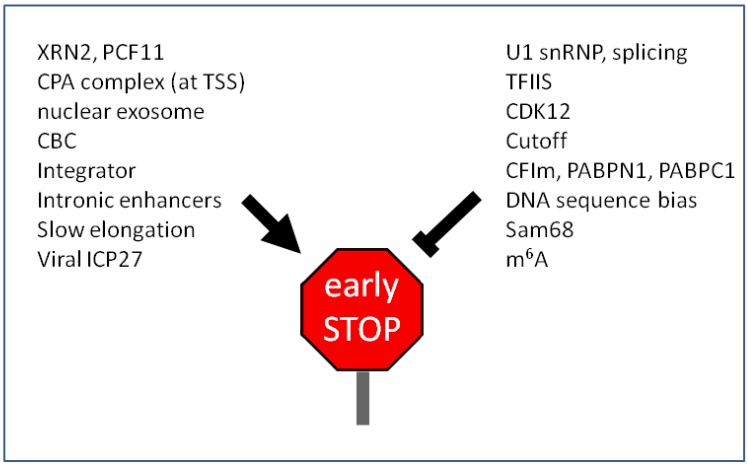

Figure 4. Factors triggering and suppressing premature transcription termination.
The molecular mechanism of PTT is unclear. However, multiple factors may trigger or prevent PTT. PTT triggering factors include canonical proteins involved in RNA 3' processing and transcription termination at gene ends (XRN2, PCF11, CPA complex), but also other factors, mainly involved in RNA processing (nuclear exosome, cap binding complex, integrator complex). Intragenic PTT is mutually exclusive with splicing, therefore the process of splicing, and splicing-promoting factors are major factors suppressing PTT. Further players preventing spurious PTT are also listed, and discussed in the text. Abbreviations: CPA – cleavage and polyadenylation; CPC – cap binding complex, CFIm – mammalian cleavage factor I; m^6A - modification of RNA by N6-methyladenosine.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27590v1 | CC BY 4.0 Open Access | rec: 15 Mar 2019, publ: 15 Mar 2019