CellPress

Terminate and make a loop: regulation of transcriptional directionality

Pawel Grzechnik, Sue Mei Tan-Wong, and Nick J. Proudfoot

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

Bidirectional promoters are a common feature of many eukaryotic organisms from yeast to humans. RNA Polymerase II that is recruited to this type of promoter can start transcribing in either direction using alternative DNA strands as the template. Such promiscuous transcription can lead to the synthesis of unwanted transcripts that may have negative effects on gene expression. Recent studies have identified transcription termination and gene looping as critical players in the enforcement of promoter directionality. Interestingly, both mechanisms share key components. Here, we focus on recent findings relating to the transcriptional output of bidirectional promoters.

Transcription from bidirectional promoters

Many promoters for RNA polymerase II (Pol II) are bidirectional, as found in a wide range of organisms including humans, mice, *Arabidopsis thaliana*, and *Saccharomyces cerevisiae* [1–8]. Pol II recruited to such promoters is directionally unbiased and can transcribe DNA in both directions. Divergent transcription can produce either two mRNA (head-to-head genes) or a single mRNA and a corresponding upstream noncoding RNA (ncRNA) [4–8]. The presence of head-to-head genes increases with decreasing genome length. Thus, for the compressed *S. cerevisiae* genome, half of mRNA coding genes are divergent, whereas in humans, only 11% of genes are so organized [9]. The vast majority of bidirectional promoters produce only one mRNA together with a divergent, usually nonfunctional, ncRNA (Box 1).

The reason transcription initiates in both directions from a promoter region appears to be dictated by the chromatin structure. In general, Pol II promoters are nucleosome-free regions (NFRs) [10–14]. The lack of DNA packaging facilitates recruitment of transcriptional machinery, unwinding of DNA strands, and as a consequence, initiation of transcription. NFRs are partially determined by intrinsic promoter sequences, which disfavor nucleosome assembly and permit access to transcription factors (TFs) [15]. In yeast, this role is played by poly(A-T) tracts frequently occurring in promoter sequences [16–18]. In higher eukaryotes, CpG islands, which are present in more than half of human

and mouse promoters, are involved in regulation of nucleosome assembly [15]. Although CpG islands support nucleosome formation *in vitro* [19], they appear nucleosome depleted *in vivo* [20]. It has been suggested that CpG islands are resistant to higher-order chromatin compaction by linker histones such as H1, which is known to have a binding preference for A-T-rich DNA [15,21].

It appears that more complex mechanisms are responsible for keeping promoters in a nucleosome-free state. For instance, several groups have reported that NFRs over promoters result from competition for DNA binding between TFs and nucleosomes [22–25]. In particular, TFs bound to promoters introduce a barrier for chromatin organization. Promoters with paused Pol II have also been shown to restrict nucleosome formation [26]. A proposed model describes stretches of poly(A-T) tracts or CpG islands that act to maintain chromatin in a state that allows transient TF access. The consequent recruitment of TFs and then Pol II results in the expansion of accessible regions and thus maintenance of NFRs [15].

Although such chromatin organization of promoters may aid gene expression, its tendency to be bidirectional introduces the potential danger of deleterious transcript synthesis. Biased recruitment of the transcriptional machinery can be determined by the orientation specific binding of TATA-binding protein (TBP) to the TATA box [4,27]. Indeed, in *Drosophila melanogaster*, Pol II-transcribed promoters are normally unidirectional with the protein coding strand defined by prominent TATA box sequences [28]. In the absence of a TATA element, TBP is still delivered to promoters by the multi-subunit TFIID complex that lacks sequence specificity [27]. Consequently, Pol II occupation at the initiation sites of bidirectional promoters is evident in the coding and noncoding directions [7,8] (Figure 1). High-resolution chromatin immunoprecipitation preceded by exonucleolytic digestion (ChIP-exo analysis) of the preinitiation complex (PIC) in *S. cerevisiae* has revealed that bidirectional promoters expressing ncRNA–mRNA contain two independent PICs in inverted orientation, similar to bidirectional promoters controlling head-to-head protein coding genes [29]. However, despite the presence of separate PICs, transcription in one direction is dependent on the other. For example, the repression of mRNA synthesis by a TATA box mutation leads to an increase in transcription of antisense divergent RNA [4]. Furthermore, divergent promoters often share the same transcriptional activators, as indicated by the co-regulation of ncRNA–mRNA as well as head-to-head mRNA pairs [1,4,5,9].

CrossMark

Box 1. Promoter associated ncRNA species

In animals, promoter-associated ncRNA (upstream antisense RNA, uasRNA) are classified by their length. Short ncRNAs (<100 nt) are referred to as transcription start site-associated RNA (TSSa-RNA) [7]. TSSa-RNA are the products of 5'→3' degradation of the nascent transcript protected by paused Pol II [93]. ncRNA >100–200 nt are termed lncRNA and are produced divergently from the coding sequence by Pol II that has escaped from pausing [94]. A prominent group of this type is called PROMPTs. PROMPTs are 5' capped and 3' adenylated, but in contrast to mRNA, undergo quick degradation by the nuclear exosome [6,48]. lncRNA considerably longer than PROMPTs derived from enhancers are named enhancer RNA (eRNA) (Box 2) [32–34,94]. Other lncRNA that are transcribed from intergenic regions (devoid of protein-coding genes) and are called long intergenic ncRNA (lincRNA) [95,96].

In yeast, promoter-associated ncRNA are more extensively characterized. Similar to animals, the vast majority of ncRNA are detectable only in RNA-degradation-deficient mutants as they undergo rapid degradation in wild type cells. Cryptic unstable transcripts (CUTs) include ncRNA accumulating in 3'→5' exonucleolytic mutants of Rrp6 and exosome [4,5,97]. Their number varies from 925 when only *RRP6* is deleted, to ~1600 when both exosome and its cofactor Rrp6 are inactive. Stable unannotated transcripts (SUTs; 847 detected), are longer (median length is 761 nt compared to 440 nt for CUTs) and more resistant to exonucleolytic digestion [5]. CUTs and SUTs are transcribed mostly from divergent promoters, followed by terminators and least frequently from intragenic regions. Both classes are predominantly terminated by the NRD complex (68% of CUTs and 58% of SUTs) and have been proposed to be organized into a holo-class called Nrd1-dependent unterminated transcripts (NUTs) [40]. A total of 1526 NUTs were identified by deep sequencing after Nrd1 depletion. However, only 213 are novel ncRNA and these do not overlap with any specific genomic features. Another major class of ncRNA includes 605 Ssu72-restricted transcripts (SRTs) [73]. SRTs accumulate independently of the exonucleolytic machinery in an *ssu72-2* mutant which has lost gene looping, and are directly associated with promoter directionality. The most abundant group of ncRNA (1658 detected) accumulate in a Xrn1 5'→3' exonuclease mutant, and are called Xrn1-dependent unstable transcripts (XUTs) [98]. Most XUTs are expressed from terminators, as 66% are antisense to their ORFs. The lines of distinction between these different ncRNA classes are blurred, for example, elongated CUTs, (eCUTs), which have escaped their primary transcription termination pathway, are terminated at the nearest PAS and may become stable SUTs [51]. Moreover, 20% of CUTs and 75% of SUTs have been classified as XUTs [98].

NFRs present elsewhere across the genome can also be used for transcription initiation. For example, NFRs over terminators of protein coding genes appear to be formed similarly to those in promoter regions. Their sequences may impede nucleosome formation [17,18,30], but transcription-dependent mechanisms may again play a role in NFR maintenance [31]. ncRNA classes initiated from terminators are identical to those expressed from the promoters [118]. However, terminator-derived transcripts may be selectively fired in the direction of the mRNA promoter, so that the ncRNA is complementary to the corresponding mRNA and may have regulatory function. Inter- or intragenically located NFRs may be also used for transcription initiation. However, a likely future focus of research will be NFRs located over enhancers which are the source of enhancer RNA (eRNA) [32–34] (Box 2). Although terminator or enhancer-associated NFRs are Pol II promoters, they appear to produce only ncRNA. In many cases it is not clear whether it is RNA synthesis or just the process of transcription itself that is relevant. Pol II initiation in these regions may also

be effectively random and merely represent byproducts of NFRs. In this review we focus only on promoters engaged in protein-coding gene expression and the mechanisms that direct Pol II into the productive synthesis of mRNA.

Divergent transcription from bidirectional promoters can be deleterious to cells. Transcription of noncoding regions may result in the downregulation of coexpressed mRNA and can also lead to accumulation of ncRNA. This may be harmful for the cell because such ncRNA could be translated into a toxic product or compete with mRNA for RNA-binding proteins and their regulatory factors [35]. In spite of this and for the reasons that are discussed in this review, divergent transcription is a common feature in many eukaryotes. Therefore, mechanisms have evolved to direct Pol II into the coding region and limit transcription of divergent ncRNA as described later.

Transcriptional directionality is controlled by termination

The simplest approach to block unwanted transcription is the immediate termination and degradation of newly synthesized ncRNA [36–40]. In this situation, Pol II transcribing ncRNA encounters intrinsic sequences that trigger transcription termination and subsequent degradation. Therefore, despite efficient transcription initiation in both directions, only Pol II transcribing the protein coding sequence successfully enters into a full elongation phase.

Multiple mechanisms can promote transcription termination of Pol II in eukaryotic organisms [41,42]. For protein-coding genes, the major transcription termination mechanism uses a poly(A) site (PAS) comprising a central AAUAAA sequence in humans and less conserved, degenerate $Py(A)_n$ sequence in *S. cerevisiae*. PAS also include 5' positioned U-rich and 3' positioned GU/U-rich sequences. PAS-dependent termination (Box 3) is mediated by macromolecular complexes generally referred to as the cleavage and polyadenylation complex (CPAC), which is remarkably conserved during evolution [43]. CPAC recognizes the PAS and cleaves the nascent RNA. Following release of the mRNA, Pol II transcribes further downstream, up to 200 nucleotides (nt) in yeast and 1500 nt in mammals, before being dismantled from the DNA template. A subset of human mRNA coding genes uses a variation of PAS-dependent termination for which a distal AU-rich region acts to mediate cotranscriptional cleavage (CoTC) with rapid 5'→3' degradation of nascent RNA leading to Pol II release [44,45].

Recently, a balance between PAS recognition and PAS blockage by U1 small nuclear RNA (snRNA) was predicted to regulate transcription termination and thus define the directionality of Pol II transcription in mammals [38,39]. Deep sequencing of polyadenylated RNA 3' ends from mouse embryonic cells identified cleavage sites associated with the PAS hexamer AAUAAA or close variants, proximal to Pol II transcription start sites (TSS) and at least 5 kb away from known transcription termination sites. Upstream antisense regions contain twofold more cleavage sites, peaking at 700 bp from a TSS, compared to the protein-coding sense sequences. Consistent with this,
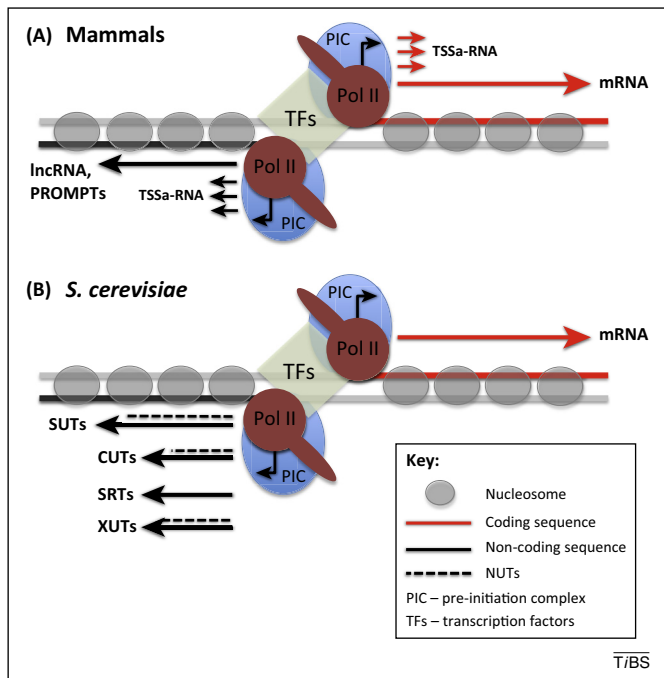
Figure 1. Promoter-associated noncoding RNA. **(A)** Major classes of promoter-associated ncRNA in animals. **(B)** ncRNA transcribed from divergent promoters in *Saccharomyces cerevisiae*. NFR acts as a Pol II promoter for both the protein-coding sequence (marked as a red line) and antisense noncoding sequence (black line). The open and unbiased nature of NFR allows for formation of two independent PICs, which share TFs. Such promoter structure allows Pol II to transcribe in both directions. In mammals the upstream regions are transcribed into PROMPTs and longer ncRNA generally referred to as lncRNA. TSSa-RNA, related to Pol II pausing, are synthesized in both directions. *S. cerevisiae* bidirectional promoters for protein-coding genes are similarly used to initiate transcription of upstream noncoding regions. Transcribed ncRNA are classified as SUTs, CUTs, or XUTs by their susceptibility to different degrading enzymes (Box 1). SRTs are synthesized when interactions between the promoter and the sense open reading frame terminator is disrupted. All ncRNA initiated from Pol II promoters, except for SRTs, may undergo NRD dependent termination and so are classified as NUTs (marked by dotted line). The most common classes of promoter-associated ncRNA are SUTs and CUTs, whereas occurrence of SRTs and XUTs is similar. Due to the lack of Pol II promoter pausing TSSa-RNA are not present in *S. cerevisiae*. Abbreviations: CUT, cryptic unstable transcript; lncRNA, long noncoding RNA; ncRNA, noncoding RNA; NFR, nucleosome-free region; NRD, NRD complex; NUT, Nrd1-dependent unterminated transcript; PIC, preinitiation complex; Pol II, RNA polymerase II; PROMPT, promoter upstream transcript; SRT, Ssu72-restricted transcript; SUT, stable unannotated transcript; TF, transcription factor; TSSa-RNA, transcription start site-associated RNA; XUT, Xrn1-depednent transcript.

48% of divergent promoters produce ncRNA susceptible to PAS-dependent antisense cleavage events in proximity to the TSS. Moreover, antisense divergent PAS are bound by cleavage and polyadenylation factors in a similar manner to mRNA PAS. This suggests that mammalian promoter-associated ncRNA (in particular promoter upstream transcripts, PROMPTs; Box 1) are generally prematurely terminated by early PAS selection. However, the frequency of candidate PAS sequences in TSS proximal sense regions (up to 6 kb) is only reduced by 33% as compared to antisense regions. Such a sequence bias does not adequately account for the discrepancy in cleavage distribution. Consequently, additional elements were searched for to explain better the asymmetric pattern of cleavage events for sense and antisense transcripts. This led to finding that 5′ splice site related sequences (5′SS; U1 sites), which are recognized by U1 snRNA are more enriched in the coding sequence direction as compared to antisense divergent sequence. This provides a possible explanation for the

biased PAS-mediated cleavage pattern, because the U1 small nuclear ribonucleoprotein (snRNP) complex is known to suppress cleavage and polyadenylation within a 1-kb vicinity of 5′SS, by inhibiting cleavage and poly-adenylation specificity factor (CPSF) [46,47]. Indeed, downregulation of U1 increases cleavage events over coding regions, but has little impact on antisense regions [38].

Parallel studies characterizing PROMPT 3′ end formation in HeLa cells [39] have validated some of the above mechanisms. This genome-wide analysis shows that these ncRNA contain functional PAS with AWUAAA, GU/U and U-rich motifs in positions resembling mRNA 3′ ends. PROMPT PAS cloned into reporter genes generate unstable polyadenylated RNA. Mutation of the GU/U PAS motif causes significant read-through transcription of the PROMPT PAS, whereas mutation of AWUAAA hexamer completely abolishes 3′ end formation [39]. Overall, the involvement of a PAS-dependent pathway in directing expression from bidirectional promoters in mammalian cells appears unequivocal. However, a few questions remain unanswered. In contrast to protein-coding genes, PAS in a PROMPT context yield unstable RNA that is subjected to degradation [6,48]. It is therefore speculated that PAS-dependent termination close to promoters occurs in suboptimal conditions, which somehow facilitates transcript degradation. This is achieved by the recruitment of the multimeric degradation complex called exosome via interaction of the cap binding complex (CBP) with the nuclear exosome targeting complex (NEXT) [39,49]. Nonetheless, the mechanism whereby the PAS-associated degradation pathway differentiates ncRNA from mRNA remains unclear.

In *S. cerevisiae*, apart from PAS-dependent termination, Pol II also uses a separate NRD-dependent pathway (Box 3) for termination of short transcripts [42]. The NRD complex consists of three polypeptides (Nrd1, Nab3, and Sen1) and is recruited at early stages of transcription. Furthermore, it responds to specific short sequences in the nascent RNA (NRD binding sites, NBS) and acts to promote transcription termination in the downstream region. NRD complex interacts with CBP and recruits the exosome together with the associated exonuclease Rrp6 [50]. In contrast to mRNA, NRD-terminated transcripts are subjected to rapid 3′→5′ exonucleolytic digestion. This connection between a termination complex and RNA-degrading enzymes allows the NRD pathway to maintain proper transcriptional directionality from bidirectional promoters. Unwanted transcription is effectively restricted close to the transcription start site and its product, ncRNA, is immediately degraded.

Recent extensive genome-wide studies [40] on the function of NRD have confirmed and broadened previous findings [36,37,51,52] by demonstrating that ~50% of known ncRNA in *S. cerevisiae* are controlled by NRD. Following nuclear depletion of Nrd1, RNA-seq analysis of newly synthesized RNA showed termination defects for many ncRNA, with 55% of them arising from bidirectional promoters. Lack of NRD caused a twofold increase in divergent transcription. Furthermore, photoactivable ribonucleoside enhanced crosslinking (PAR-CLIP), a method used for analysis of protein–RNA binding, has revealed

Box 2. eRNA

Enhancers are described as DNA elements that activate promoters from a variable distance. The ~55 000 enhancers identified in the human genome play regulatory roles in development, differentiation, and tissue-specific gene expression [99,100]. Like promoters, enhancers comprise accessible chromatin structure bound by TFs and Pol II [32,33,101]. However, in contrast to promoters, only 5% of enhancer sequences consist of CpG islands, and enhancers have significantly lower levels of H3K4me3 marks, which are responsible for PIC assembly stimulation [84,99,102]. Instead, enhancers are generally more defined by H3K4me1 and H3K327 acetylation marks. Analysis of ncRNA in human embryonic stem cells, murine embryonic stem cells, and human endodermal cells has revealed that RNA initiated from enhancers (eRNA) comprises, respectively, 19%, 27%, and 23% of all lncRNA [34]. However, if only lincRNA are taken into account, eRNA comprise the vast majority.

Enhancer regions promoting transcription (referred as e-promoters) can be either uni- or bidirectional [32,33,103]. Genome-wide studies in mouse cells have revealed that eRNA transcribed from bidirectional e-promoters are relatively short (0.5–2 kb) and non-polyadenylated [103]. By contrast, unidirectional e-promoters generate longer transcripts (often >4 kb) and are normally polyadenylated. Although features of Pol II transcribing long eRNA appear similar to protein-coding genes, its C-terminal domain (CTD) phosphorylation pattern differs: CTD Ser2 levels are low along the enhancer-derived transcription unit. Also, the transcribed region lacks H3K36me3 elongation marks [103]. A poly(A) tail indicates the presence of a PAS at the 3′ end of long eRNA and implies termination by CPAC. This may control directionality of transcription initiated in enhancers by the mechanisms as described in this review. Studies on the murine α-globin locus in mouse erythrocytes have shed significant light on long eRNA [104]. Transcription from intragenic α-globin enhancers results in short nonpolyadenylated eRNA in both directions. However, because these enhancers are located in the intron of a separate host gene *Nprl3*, they can also act as alternative promoters for *Nprl3*. This results in directional transcription of lncRNA referred to as multiexonic RNA (meRNA), which are spliced and polyadenylated as mRNA. Such meRNA are prevalent in the human genome [104]. This observation raises the intriguing question: how can bidirectional intergenic e-promoters convert to unidirectionality? Does such a mechanism act to enforce transcriptional directionality in a similar manner to that of protein-coding genes? Taking into account the exceptionally important role of enhancers in the regulation of basic biological processes, the better understanding of enhancer-associated transcription may significantly broaden our knowledge of how protein-coding genes are activated or silenced.

relatively higher NRD binding affinity for divergent ncRNA than for other Pol II transcripts, and has identified several RNA-binding motifs including those already known (Box 3). As with human PAS distribution [38,39], NBS are rarer in protein-coding sequences than in antisense ncRNA. However, the mechanism that antagonizes NRD termination in mRNA sequences, mimicking the mammalian PAS–U1 axis, has not yet been established. Another PAR-CLIP study has revealed that a quarter of protein-coding genes in yeast are bound by NRD [53], but only 302 genes show significant concentration of crosslinked NRD at the 5′ end, suggesting the location of functional terminators. Nevertheless, even a single NBS is sufficient to trigger transcription termination [54], therefore, many genes theoretically may be susceptible to the NRD termination pathway. However, given that only 44 mRNA coding genes have been shown to be prematurely terminated by NRD [40,52,55,56], a mechanism that negatively affects NRD-dependent termination over the protein-coding sequences may exist. Possibly, suppression of NRD

termination relies on phosphorylation of Pol II C-terminal domain (CTD) Tyr-1, which may impair Nrd1–CTD interaction and so encourage elongation [57]. Other mechanisms that control promoter directionality at the transcription termination level in yeast have yet to be characterized.

In higher eukaryotes no homologs of the NRD complex have been identified. However, surprisingly, yeast NRD termination also requires CPAC factors. Although the functional links between CPAC and NRD are still under investigation, it appears that many CPAC components (e.g., Pcf11, Ssu72, Glc7) are crucial for NRD termination [58–64]. We speculate that in the course of evolution, NRD has been lost and entirely replaced with CPAC to terminate ncRNA in higher organisms. Alternatively, yeast might be considered to have evolved NRD to facilitate the recruitment of CPAC close to the transcription start site.

Regardless of the organism, ncRNA are enriched in termination signals. These act to trigger recruitment of termination factors resulting in premature transcription termination and subsequent transcript degradation, thus maintaining transcriptional directionality. Remarkably, termination factors are also essential for another mechanism that acts to enforce promoter directionality: gene loops.

Gene loops enhance promoter directionality

As described above, early transcription termination and subsequent RNA degradation act to remove unwanted divergent transcription. However, it is now apparent that a specific gene looping mechanism may further select productive gene transcription by encouraging Pol II to transcribe into the sense coding sequence. The fact that functional and physical interaction between initiation and termination factors leads to the juxtaposition of promoter and terminator regions has been known for a decade [65,66]. Thus ChIP analysis in *S. cerevisiae* of the general transcription factor TFIIB, which functions in promoter regions, also detected it over the 3′ ends of mRNA coding genes [67]. Conversely, some CPAC components crosslinked to the 5′ ends of genes [66,68,69]. Among them, were the Pol II CTD phosphatase Ssu72, which interacts genetically with TFIIB [70] and the whole CPAC subcomplex called cleavage factor IA (CF IA) [66,68]. Formation of gene loops was also independently confirmed by chromosome conformation capture analysis (3C) and shown to be dependent on active transcription [65,66].

Although gene loop formation is transcription dependent, mRNA synthesis is not strongly affected when the loop is disrupted. However, gene loops were shown to contribute to transcriptional memory in yeast. Thus, although transcriptional reactivation of inducible *GAL10* and *HXK1* genes was faster than the first round of activation, this effect was lost following gene loop disruption [71,72]. Physical proximity of the promoter and termination regions may facilitate reinitiation of RNA polymerase and therefore improve gene expression [66]. However, it has not been directly proven that the same Pol II molecule released at the 3′ end of the gene is immediately transferred to the promoter. Such evidence can only be obtained from single molecule experiments.

Box 3. PAS- and NRD-dependent transcription termination

PAS-dependent transcription termination is tightly coupled with 3' end processing. Despite the lack of PAS sequence conservation between mammals and yeast, most of mammalian termination and processing factors have homologs in *S. cerevisiae*. Throughout this review we generally refer to these factors as components of the CPAC. In higher eukaryotes, CPAC consists of four multi-subunit complexes: CPSF, CstF, and two cleavage factors (CF Im and CF IIm). In yeast, homologous proteins are organized into subcomplexes; CPF and CF IA and CF IB [42,105]. The eukaryotic termination and processing machinery also contains poly(A) polymerase (PAP) and poly(A) binding proteins (PABP). CPAC recognizes and binds the PAS. The CPAC component CPSF-73 (Brr5 in yeast) cleaves the nascent RNA releasing pre-mRNA [106,107], which is subsequently polyadenylated. The downstream RNA generated by the cleavage is digested by 5'→3' exoribonuclease Xrn2 (Rat1 in yeast). This is accompanied by an RNA:DNA helicase called Senataxin which unwinds potential R-loops formed with the DNA template [108]. Xrn2 degrades RNA and together with the other CPAC components, displaces Pol II from DNA [109,110].

Recruitment of CPAC also depends on the Pol II CTD. Pcf11, a component of CF IIm in humans and CF IA in yeast, is crucial for transcription termination [111]. This protein contains a CTD interacting domain (CID) and is recruited to Ser-2 phosphorylated CTD [112], characteristic of late stages of transcription. The presence of termination factors on Pol II approaching the end of the gene enhances the termination process.

In *S. cerevisiae* Pol II can be also terminated by the NRD complex consisting of Nrd1, Nab3, and RNA:DNA helicase Sen1 [50,113]. Similarly to CPAC, NRD recognizes specific sequences in the nascent RNA [the strongest motifs are UGUA/GUAG and UCUU/CUUG for the RNA recognition motifs (RRMs) of Nrd1 and Nab3, respectively] and interacts with CTD via the Nrd1 CID [52,114,115]. However, in contrast to Pcf11, Nrd1 has a stronger affinity to Ser-5 phosphorylated CTD and is therefore recruited at an earlier stage of transcription [114]. Remarkably, Ser2-P marks are also present over NRD terminators [59,116] and most CPAC mutants, including Pcf11 but excluding Brr5, display a termination defect for NRD-dependent genes [58–64]. However, NRD–CPAC interplay remains poorly understood. The current model, backed by *in vitro* data, assumes that Sen1, delivered to the nascent RNA by NRD, translocates in a 5'→3' direction and somehow displaces Pol II [117].

Recently, intragenic gene loops have also been shown to enhance transcriptional directionality from bidirectional promoters in *S. cerevisiae* [73]. Mutations of Ssu72 affects loop formation, which in turn results in Pol II relocation on the promoter towards the production of divergent ncRNA and their consequent accumulation. *ssu72-2* mutant cells show the appearance of 605 novel ncRNA called Ssu72 restricted transcripts (SRTs); the majority of which are associated with bidirectional promoters. Mutation of the TFIIB component Sua7 and other CPAC factors involved in gene looping, Pta1, Rna14 and RNA15, also causes an increase in promoter-associated ncRNA. Furthermore, replacement of the PAS with an Rnt1 cleavage signal (RCS) results in a similar phenotype. RCS can act as an alternative terminator for Pol-II-transcribed genes [74,75]. However, the lack of PAS disrupts the association of CPAC and consequently prevents gene loop formation. This results in threefold higher transcription of the noncoding divergent region compared to the gene possessing a normal PAS. Overall, gene loop formation favors transcription in the direction that forms a looped structure. It must be mentioned that Ssu72 is also essential for the NRD pathway. By controlling both processes of early transcription termination and gene looping, this protein appears to be the determinant factor for promoter directionality. The CPAC subcomplex CF IA has consistently been shown to be involved in gene loop formation, where it promotes Pol II reinitiation and transcription of the coding region [69]. Intrinsic DNA sequences can also modulate transcriptional directionality through physical properties that facilitate looping. Genes where TFIIB was detected at the 3' end are likely to form loops [76,77]. Interestingly, their gene sequences display higher flexibility in the middle of the open reading frame (ORF) than in nonlooped genes [77].

In higher organisms the process of gene looping has not been investigated in depth. However there are strong indications that the same mechanism exists in animals and plants. Human TFIIB recruits both CPSF and CstF to the promoter at an early transcriptional stage, before Pol II enters into the elongation phase [78]. TFIIB is also necessary for CstF recruitment to the terminator region and interacts directly with human Ssu72 and CstF-64 [68,78]. Similarly, in *D. melanogaster* TFIIB is required for termination of the *polo* gene. In this case, 3C analysis shows the existence of a TFIIB-dependent gene loop between promoter and terminator region [79]. Looping of the mammalian genes CD68 and BRCA1 as well as the plant gene FLC has also been reported [80–82]. Finally, a gene loop was detected across integrated HIV-1 provirus following transcriptional activation [83].

Gene loops provide an additional mechanism to enforce promoter directionality. In contrast to transcription termination, which aborts transcription of noncoding upstream regions, gene loops direct Pol II into the protein-coding sequences. However, their genome-wide prevalence is yet to be estimated.

Chromatin marks reinforce promoter directionality

On leaving the promoter and transcribing into the gene body, Pol II encounters nucleosomes; chromatin structural elements that consist of histones. Modified histones act as modulators of transcription, including regulation of transcription from bidirectional promoters. Divergent PICs on the promoters are in general compositionally equivalent [29]. Consequently, in human cells, histone H3 lysine 4 is trimethylated (H3K4me3) on both sides of the promoter and acts to stimulate PIC formation [7,84]. By contrast, the marker of active transcription, H3 lysine 79 dimethylation (H3K79me2) is found only over the coding region [7].

In yeast, the *ssu72-2* mutation results in H4 acetylation in the upstream noncoding region of the bidirectional promoters [73]. This suggests that H4 deacetylation normally acts to reinforce transcriptional directionality established by Ssu72-dependent gene loops. Deacetylated histones are thought to form more compact chromatin than when acetylated. This restricts access to transcription factors and so affects transcription [85].

Why is divergent transcription needed?

Promoter bidirectionality as a general feature of eukaryotic genes could facilitate genome evolution by providing a pool of noncoding transcripts that may potentially cause a gain of function. It has been suggested that some long ncRNA (lncRNA) might be evolutionary precursors of mRNA-coding genes [38]. Indeed, the distribution of

PAS and U1 sites over mRNA–PROMPT, mRNA–lncRNA, and head to head mRNA–mRNA genes, reveals that mRNA–lncRNA have an intermediate pattern. Consistent with this, evolutionary analysis of mouse genes has uncovered that the gain of U1 sites and loss of PAS at gene 5′ ends is correlated with the age of the gene [38]. This reinforces promoter directionality towards the coding sequence and promotes the stability and inheritance of the gene. The role of divergent transcription in genome evolution has been recently reviewed [86].

Bidirectional promoters offer an additional possibility to control gene expression. The negative impact on expression of protein-coding genes by competition for TFs between sense and antisense PICs [4] was mentioned above. However, divergent transcription can also positively affect gene expression. PICs formed in a noncoding direction may help to maintain NFRs and locally provide a pool of available TFs, which may then be used to enhance transcription of the coding region, if so needed. Negative supercoils generated behind actively transcribing Pol II may create tension that helps unwind the DNA duplex in promoters and so increases the efficiency of transcription initiation in either direction [87]. Finally, head-to-head protein-coding genes often encode proteins involved in the same biological process. This arrangement appears to enhance their co-regulation and gives an advantage in rapid response to different stimuli [88]. The best examples are genes encoding DNA repair enzymes, where bidirectional promoters, both expressing proteins respond to DNA damage, are five times more represented than unidirectional [1].

Concluding remarks

Transcriptional initiation *de novo* appears to be directionally unbiased. Pol II starts transcribing in both directions making either mRNA or ncRNA. Uncontrolled transcription of noncoding regions may lead to downregulation of divergent mRNA and potentially toxic accumulation of ncRNA. However conditions that facilitate transcription initiation of ncRNA also promote mRNA synthesis. Finally, divergent transcription provides the opportunity for the evolution of new genes. Therefore eukaryotes have retained mechanisms that allow for the existence of bidirectional ncRNA–mRNA promoters, by restricting the synthesis of unwanted ncRNA (Figure 2).
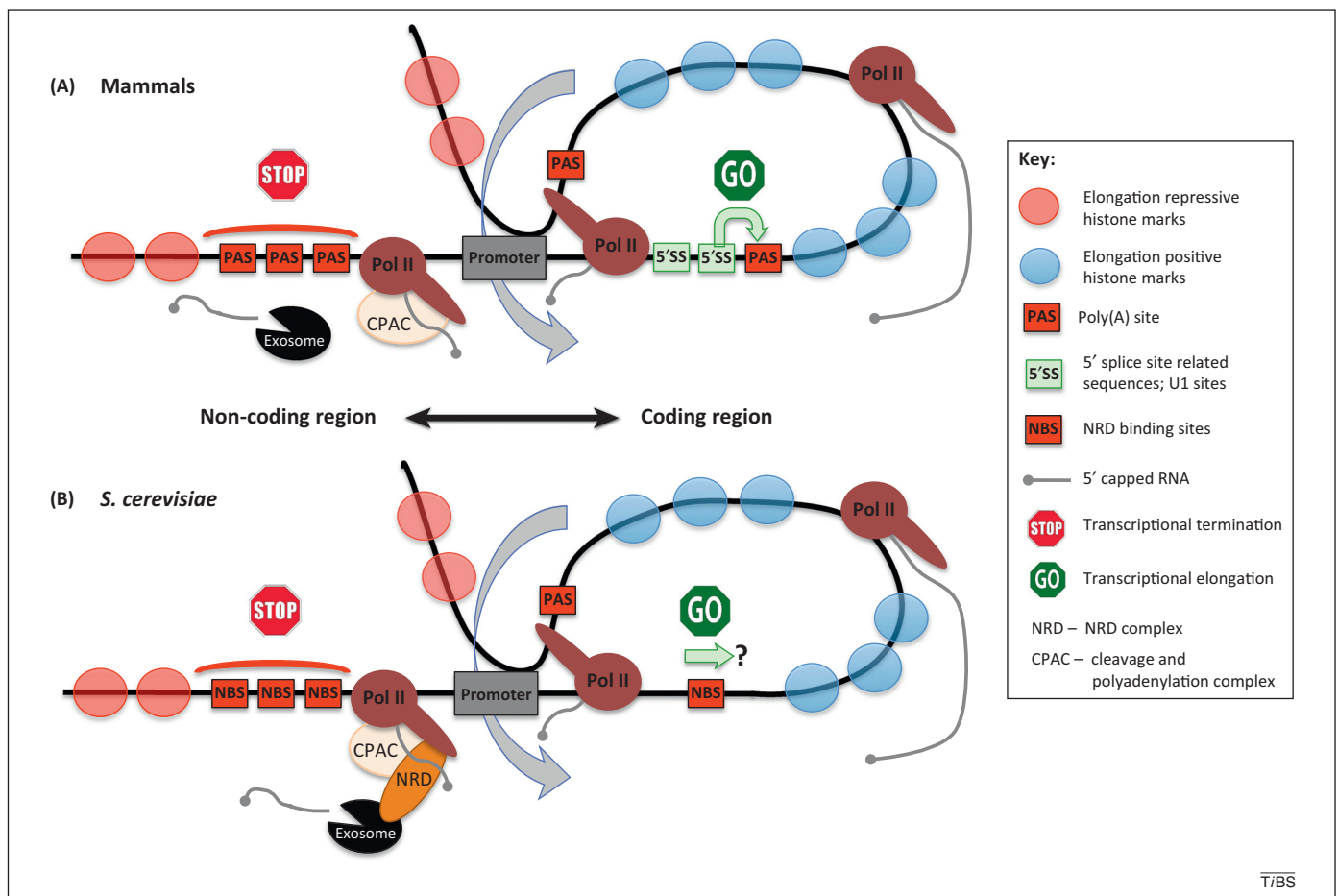

Figure 2. Mechanisms enforcing promoter directionality. **(A)** In mammalian cells; **(B)** in *Saccharomyces cerevisiae*. Transcription of the upstream noncoding region is restricted by over-representation of termination signals (STOP sign, red rectangles), PAS in mammals or NBS in *S. cerevisiae*. NBS trigger the NRD termination pathway, which depends on both the NRD complex and CPAC. In contrast PAS-dependent termination is mediated by CPAC only. In both cases released ncRNA is degraded by the exosome. PAS and NBS are depleted over the coding regions. Moreover, in mammalian cells, the coding sequence is enriched in 5′ splice-site-related sequences recognized by U1 small nuclear RNA (snRNA), which inhibits PAS-dependent transcription termination (GO sign). A similar mechanism in yeast is yet to be discovered. Transcriptional directionality is also reinforced by chromatin modification. Nucleosomes with transcription-positive histone marks are presented as blue semitransparent circles, whereas nucleosomes with negative marks are shown in red. A pioneer round of transcription of the coding region establishes a gene loop. Termination factors recruited to Pol II transcribing in the vicinity of the terminator region interacts with initiation factors on the promoter and these juxtapose both regions. The formed loop enhances transcriptional reinitiation into the coding sequence (denoted by bent blue arrows). Abbreviations: CPAC, cleavage and polyadenylation complex; NBS, NRD-binding sites; ncRNA, noncoding RNA; NRD, NRD complex; PAS, poly(A) site; Pol II, RNA polymerase II.

The first line of defense against pervasive transcription is early transcription termination of ncRNA. This is mediated by PAS- and NRD-dependent pathways in mammals and yeast, respectively. In both cases termination signals are unequally positioned in each direction. Protein-coding sequences are depleted of PAS in human or NBS in yeast, whereas divergent noncoding regions are enriched in these sequences. Conversely, in mammals there is an opposite distribution of U1 sites that inhibit premature PAS recognition in protein-coding sequences. In addition, the first round of transcription of the coding sequence assists in establishing chromatin marks that promote efficient elongation. Finally, when transcribing Pol II approaches the 3′ end of the coding sequence, CPAC is recruited. This establishes interactions with TFIIB on the promoter and so creates a gene loop, which in turn strengthens transcriptional directionality. Taken together, bidirectional promoters are controlled at two stages. First, Pol II elongation directionality is regulated by transcriptional termination, and second, gene loops selectively enhance Pol II transcription on coding sequences.

Many questions concerning transcriptional directionality remain unanswered. How is it regulated in the case of head-to-head protein-coding genes when both divergent genes form loops? Can the promoter and PAS of ncRNA juxtapose? How widely distributed are gene loops in yeast and higher eukaryotes? Although the mammalian CoTC-type termination mechanism is as yet poorly understood, it will be interesting to investigate whether this type of PAS-dependent transcription termination is also involved in promoter-dependent antisense termination.

Transcriptional directionality can also potentially be regulated by RNA editing. Adenosine deaminase acting on RNA (ADAR) proteins can convert adenosine to inosine (reviewed in [89]). Inosine is then recognized as guanosine by the splicing and translation machineries. ADAR has been shown to introduce alternative 5′ (U1 site) and 3′ splicing sites [90,91]. Similarly, recently described R-loop-related editing of nascent transcripts frequently introduces U/A-to-G changes [92], which could also preferentially alter RNA sequence. It is therefore tempting to speculate that such RNA editing contributes to transcriptional directionality by creating U1 sites or by destroying PAS sequences, thus selectively enhancing mRNA synthesis.

Directing Pol II towards the coding region and restricting transcription of noncoding regions ensures proper gene expression. Moreover, bidirectional promoters offer an opportunity to create additional layers of either positive or negative regulation on protein-coding genes. Immediate transcription termination in the noncoding regions and subsequent degradation of unwanted RNA reduces the potential selective pressure to convert bidirectional into unidirectional promoters. This may in turn facilitate the origin of new genes and thus evolutionary progress.

References

1 Trinklein, N.D. *et al.* (2004) An abundance of bidirectional promoters in the human genome. *Genome Res.* 14, 62–66
2 Koyanagi, K.O. *et al.* (2005) Comparative genomics of bidirectional gene pairs and its implications for the evolution of a transcriptional regulation system. *Gene* 353, 169–176
3 Wang, Q. *et al.* (2009) Searching for bidirectional promoters in *Arabidopsis thaliana*. *BMC Bioinform.* 10 (Suppl. 1), S29
4 Neil, H. *et al.* (2009) Widespread bidirectional promoters are the major source of cryptic transcripts in yeast. *Nature* 457, 1038–1042
5 Xu, Z. *et al.* (2009) Bidirectional promoters generate pervasive transcription in yeast. *Nature* 457, 1033–1037
6 Preker, P. *et al.* (2008) RNA exosome depletion reveals transcription upstream of active human promoters. *Science* 322, 1851–1854
7 Seila, A.C. *et al.* (2008) Divergent transcription from active promoters. *Science* 322, 1849–1851
8 Core, L.J. *et al.* (2008) Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science* 322, 1845–1848
9 Xu, C. *et al.* (2012) The preservation of bidirectional promoter architecture in eukaryotes: what is the driving force? *BMC Syst. Biol.* 6 (Suppl. 1), S21
10 Yuan, G.C. *et al.* (2005) Genome-scale identification of nucleosome positions in *S. cerevisiae*. *Science* 309, 626–630
11 Albert, I. *et al.* (2007) Translational and rotational settings of H2A.Z. nucleosomes across the *Saccharomyces cerevisiae* genome. *Nature* 446, 572–576
12 Mito, Y. *et al.* (2005) Genome-scale profiling of histone H3.3 replacement patterns. *Nat. Genet.* 37, 1090–1097
13 Ozsolak, F. *et al.* (2007) High-throughput mapping of the chromatin structure of human promoters. *Nat. Biotechnol.* 25, 244–248
14 Mavrich, T.N. *et al.* (2008) Nucleosome organization in the *Drosophila* genome. *Nature* 453, 358–362
15 Guertin, M.J. and Lis, J.T. (2013) Mechanisms by which transcription factors gain access to target sequence elements in chromatin. *Curr. Opin. Genet. Dev.* 23, 116–123
16 Sekinger, E.A. *et al.* (2005) Intrinsic histone-DNA interactions and low nucleosome density are important for preferential accessibility of promoter regions in yeast. *Mol. Cell* 18, 735–748
17 Segal, E. *et al.* (2006) A genomic code for nucleosome positioning. *Nature* 442, 772–778
18 Kaplan, N. *et al.* (2009) The DNA-encoded nucleosome organization of a eukaryotic genome. *Nature* 458, 362–366
19 Valouev, A. *et al.* (2011) Determinants of nucleosome organization in primary human cells. *Nature* 474, 516–520
20 Wang, J. *et al.* (2012) Sequence features and chromatin structure around the genomic regions bound by 119 human transcription factors. *Genome Res.* 22, 1798–1812
21 Cui, F. and Zhurkin, V.B. (2009) Distinctive sequence patterns in metazoan and yeast nucleosomes: implications for linker histone binding to AT-rich and methylated DNA. *Nucleic Acids Res.* 37, 2818–2829
22 Floer, M. *et al.* (2010) A RSC/nucleosome complex determines chromatin architecture and facilitates activator binding. *Cell* 141, 407–418
23 Bai, L. *et al.* (2011) Multiple sequence-specific factors generate the nucleosome-depleted region on CLN2 promoter. *Mol. Cell* 42, 465–476
24 Ozonov, E.A. and van Nimwegen, E. (2013) Nucleosome free regions in yeast promoters result from competitive binding of transcription factors that interact with chromatin modifiers. *PLoS Comput. Biol.* 9, e1003181
25 Zhang, Y. *et al.* (2009) Intrinsic histone-DNA interactions are not the major determinant of nucleosome positions in vivo. *Nat. Struct. Mol. Biol.* 16, 847–852
26 Gilchrist, D.A. *et al.* (2010) Pausing of RNA polymerase II disrupts DNA-specified nucleosome organization to enable precise gene regulation. *Cell* 143, 540–551
27 Papai, G. *et al.* (2011) New insights into the function of transcription factor TFIID from recent structural studies. *Curr. Opin. Genet. Dev.* 21, 219–224
28 Core, L.J. *et al.* (2012) Defining the status of RNA polymerase at promoters. *Cell Rep.* 2, 1025–1035

29 Rhee, H.S. and Pugh, B.F. (2012) Genome-wide structure and organization of eukaryotic pre-initiation complexes. *Nature* 483, 295–301

30 Kaplan, N. *et al.* (2010) Nucleosome sequence preferences influence in vivo nucleosome organization. *Nat. Struct. Mol. Biol.* 17, 918–920, author reply 920–912

31 Fan, X. *et al.* (2010) Nucleosome depletion at yeast terminators is not intrinsic and can occur by a transcriptional mechanism linked to 3′-end formation. *Proc. Natl. Acad. Sci. U.S.A.* 107, 17945–17950

32 Kim, T.K. *et al.* (2010) Widespread transcription at neuronal activity-regulated enhancers. *Nature* 465, 182–187

33 De Santa, F. *et al.* (2010) A large fraction of extragenic RNA pol II transcription sites overlap enhancers. *PLoS Biol.* 8, e1000384

34 Sigova, A.A. *et al.* (2013) Divergent transcription of long noncoding RNA/mRNA gene pairs in embryonic stem cells. *Proc. Natl. Acad. Sci. U.S.A.* 110, 2876–2881

35 Jensen, T.H. *et al.* (2013) Dealing with pervasive transcription. *Mol. Cell* 52, 473–484

36 Arigo, J.T. *et al.* (2006) Termination of cryptic unstable transcripts is directed by yeast RNA-binding proteins Nrd1 and Nab3. *Mol. Cell* 23, 841–851

37 Thiebaut, M. *et al.* (2006) Transcription termination and nuclear degradation of cryptic unstable transcripts: a role for the nrd1-nab3 pathway in genome surveillance. *Mol. Cell* 23, 853–864

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

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

40 Schulz, D. *et al.* (2013) Transcriptome surveillance by selective termination of noncoding RNA synthesis. *Cell* 155, 1075–1087

41 Richard, P. and Manley, J.L. (2009) Transcription termination by nuclear RNA polymerases. *Genes Dev.* 23, 1247–1269

42 Mischo, H.E. and Proudfoot, N.J. (2013) Disengaging polymerase: terminating RNA polymerase II transcription in budding yeast. *Biochim. Biophys. Acta* 1829, 174–185

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

44 White, E. *et al.* (2012) AT-rich sequence elements promote nascent transcript cleavage leading to RNA polymerase II termination. *Nucleic Acids Res.* 41, 1797–1806

45 Nojima, T. *et al.* (2013) Definition of RNA polymerase II CoTC terminator elements in the human genome. *Cell Rep.* 3, 1080–1092

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

47 Berg, M.G. *et al.* (2012) U1 snRNP determines mRNA length and regulates isoform expression. *Cell* 150, 53–64

48 Preker, P. *et al.* (2011) PROMoter uPstream Transcripts share characteristics with mRNAs and are produced upstream of all three major types of mammalian promoters. *Nucleic Acids Res.* 39, 7179–7193

49 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

50 Vasiljeva, L. and Buratowski, S. (2006) Nrd1 interacts with the nuclear exosome for 3′ processing of RNA polymerase II transcripts. *Mol. Cell* 21, 239–248

51 Marquardt, S. *et al.* (2011) Distinct RNA degradation pathways and 3′ extensions of yeast non-coding RNA species. *Transcription* 2, 145–154

52 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

53 Webb, S. *et al.* (2014) PAR-CLIP data indicate that Nrd1-Nab3-dependent transcription termination regulates expression of hundreds of protein coding genes in yeast. *Genome Biol.* 15, R8

54 Kim, K.Y. and Levin, D.E. (2011) Mpk1 MAPK association with the Paf1 complex blocks Sen1-mediated premature transcription termination. *Cell* 144, 745–756

55 Kuehner, J.N. and Brow, D.A. (2008) Regulation of a eukaryotic gene by GTP-dependent start site selection and transcription attenuation. *Mol. Cell* 31, 201–211

56 Thiebaut, M. *et al.* (2008) Futile cycle of transcription initiation and termination modulates the response to nucleotide shortage in *S. cerevisiae*. *Mol. Cell* 31, 671–682

57 Mayer, A. *et al.* (2012) CTD tyrosine phosphorylation impairs termination factor recruitment to RNA polymerase II. *Science* 336, 1723–1725

58 Al Husini, N. *et al.* (2013) A role for CF1A 3′ end processing complex in promoter-associated transcription. *PLoS Genet.* 9, e1003722

59 Kim, H. *et al.* (2010) Gene-specific RNA polymerase II phosphorylation and the CTD code. *Nat. Struct. Mol. Biol.* 17, 1279–1286

60 Kim, M. *et al.* (2006) Distinct pathways for snoRNA and mRNA termination. *Mol. Cell* 24, 723–734

61 Singh, N. *et al.* (2009) The Ess1 prolyl isomerase is required for transcription termination of small noncoding RNAs via the Nrd1 pathway. *Mol. Cell* 36, 255–266

62 Steinmetz, E.J. and Brow, D.A. (2003) Ssu72 protein mediates both poly(A)-coupled and poly(A)-independent termination of RNA polymerase II transcription. *Mol. Cell. Biol.* 23, 6339–6349

63 Nedea, E. *et al.* (2008) The Glc7 phosphatase subunit of the cleavage and polyadenylation factor is essential for transcription termination on snoRNA genes. *Mol. Cell* 29, 577–587

64 Lenstra, T.L. *et al.* (2014) The role of Ctk1 kinase in termination of small non-coding RNAs. *PLoS ONE* 8, e80495

65 O'Sullivan, J.M. *et al.* (2004) Gene loops juxtapose promoters and terminators in yeast. *Nat. Genet.* 36, 1014–1018

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

67 Singh, B.N. and Hampsey, M. (2007) A transcription-independent role for TFIIB in gene looping. *Mol. Cell* 27, 806–816

68 Calvo, O. and Manley, J.L. (2001) Evolutionarily conserved interaction between CstF-64 and PC4 links transcription, polyadenylation, and termination. *Mol. Cell* 7, 1013–1023

69 Medler, S. *et al.* (2011) Evidence for a complex of transcription factor IIB with poly(A) polymerase and cleavage factor 1 subunits required for gene looping. *J. Biol. Chem.* 286, 33709–33718

70 Sun, Z.W. and Hampsey, M. (1996) Synthetic enhancement of a TFIIB defect by a mutation in SSU72, an essential yeast gene encoding a novel protein that affects transcription start site selection in vivo. *Mol. Cell. Biol.* 16, 1557–1566

71 Laine, J.P. *et al.* (2009) A physiological role for gene loops in yeast. *Genes Dev.* 23, 2604–2609

72 Tan-Wong, S.M. *et al.* (2009) Gene loops function to maintain transcriptional memory through interaction with the nuclear pore complex. *Genes Dev.* 23, 2610–2624

73 Tan-Wong, S.M. *et al.* (2012) Gene loops enhance transcriptional directionality. *Science* 338, 671–675

74 Rondon, A.G. *et al.* (2009) Fail-safe transcriptional termination for protein-coding genes in *S. cerevisiae*. *Mol. Cell* 36, 88–98

75 Ghazal, G. *et al.* (2009) Yeast RNase III triggers polyadenylation-independent transcription termination. *Mol. Cell* 36, 99–109

76 Mavrich, T.N. *et al.* (2008) A barrier nucleosome model for statistical positioning of nucleosomes throughout the yeast genome. *Genome Res.* 18, 1073–1083

77 Dai, Z. *et al.* (2014) The pattern and evolution of looped gene bendability. *Mol. Biol. Evol.* 31, 319–329

78 Wang, Y. *et al.* (2010) Phosphorylation of TFIIB links transcription initiation and termination. *Curr. Biol.* 20, 548–553

79 Henriques, T. *et al.* (2012) Transcription termination between polo and snap, two closely spaced tandem genes of *D. melanogaster*. *Transcription* 3, 198–212

80 Tan-Wong, S.M. *et al.* (2008) Dynamic interactions between the promoter and terminator regions of the mammalian BRCA1 gene. *Proc. Natl. Acad. Sci. U.S.A.* 105, 5160–5165

81 O'Reilly, D. and Greaves, D.R. (2007) Cell-type-specific expression of the human CD68 gene is associated with changes in Pol II phosphorylation and short-range intrachromosomal gene looping. *Genomics* 90, 407–415

82 Crevillen, P. *et al.* (2013) A gene loop containing the floral repressor FLC is disrupted in the early phase of vernalization. *EMBO J.* 32, 140–148

83 Perkins, K.J. *et al.* (2008) Transcription-dependent gene looping of the HIV-1 provirus is dictated by recognition of pre-mRNA processing signals. *Mol. Cell* 29, 56–68

84 Lauberth, S.M. *et al.* (2013) H3K4me3 interactions with TAF3 regulate preinitiation complex assembly and selective gene activation. *Cell* 152, 1021–1036

85 Shahbazian, M.D. and Grunstein, M. (2007) Functions of site-specific histone acetylation and deacetylation. *Annu. Rev. Biochem.* 76, 75–100

86 Wu, X. and Sharp, P.A. (2013) Divergent transcription: a driving force for new gene origination? *Cell* 155, 990–996

87 Parvin, J.D. and Sharp, P.A. (1993) DNA topology and a minimal set of basal factors for transcription by RNA polymerase II. *Cell* 73, 533–540

88 Li, Y.Y. *et al.* (2006) Systematic analysis of head-to-head gene organization: evolutionary conservation and potential biological relevance. *PLoS Comput. Biol.* 2, e74

89 Mallela, A. and Nishikura, K. (2012) A-to-I editing of protein coding and noncoding RNAs. *Crit. Rev. Biochem. Mol. Biol.* 47, 493–501

90 Rueter, S.M. *et al.* (1999) Regulation of alternative splicing by RNA editing. *Nature* 399, 75–80

91 Solomon, O. *et al.* (2013) Global regulation of alternative splicing by adenosine deaminase acting on RNA (ADAR). *RNA* 19, 591–604

92 Wang, I.X. *et al.* (2014) RNA-DNA differences are generated in human cells within seconds after RNA exits polymerase II. *Cell Rep.* 6, 906–915

93 Valen, E. *et al.* (2011) Biogenic mechanisms and utilization of small RNAs derived from human protein-coding genes. *Nat. Struct. Mol. Biol.* 18, 1075–1082

94 Kapranov, P. *et al.* (2007) RNA maps reveal new RNA classes and a possible function for pervasive transcription. *Science* 316, 1484–1488

95 Guttman, M. *et al.* (2009) Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* 458, 223–227

96 Cabili, M.N. *et al.* (2011) Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev.* 25, 1915–1927

97 Gudipati, R.K. *et al.* (2012) Extensive degradation of RNA precursors by the exosome in wild-type cells. *Mol. Cell* 48, 409–421

98 van Dijk, E.L. *et al.* (2011) XUTs are a class of Xrn1-sensitive antisense regulatory non-coding RNA in yeast. *Nature* 475, 114–117

99 Heintzman, N.D. *et al.* (2009) Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* 459, 108–112

100 Levine, M. (2010) Transcriptional enhancers in animal development and evolution. *Curr. Biol.* 20, R754–R763

101 Orom, U.A. and Shiekhattar, R. (2011) Long non-coding RNAs and enhancers. *Curr. Opin. Genet. Dev.* 21, 194–198

102 Natoli, G. and Andrau, J.C. (2012) Noncoding transcription at enhancers: general principles and functional models. *Annu. Rev. Genet.* 46, 1–19

103 Koch, F. *et al.* (2011) Transcription initiation platforms and GTF recruitment at tissue-specific enhancers and promoters. *Nat. Struct. Mol. Biol.* 18, 956–963

104 Kowalczyk, M.S. *et al.* (2012) Intragenic enhancers act as alternative promoters. *Mol. Cell* 45, 447–458

105 Chan, S. *et al.* (2011) Pre-mRNA 3'-end processing complex assembly and function. *Wiley Interdiscip Rev. RNA* 2, 321–335

106 Chanfreau, G. *et al.* (1996) Essential yeast protein with unexpected similarity to subunits of mammalian cleavage and polyadenylation specificity factor (CPSF). *Science* 274, 1511–1514

107 Mandel, C.R. *et al.* (2006) Polyadenylation factor CPSF-73 is the pre-mRNA 3'-end-processing endonuclease. *Nature* 444, 953–956

108 Skourti-Stathaki, K. *et al.* (2011) Human senataxin resolves RNA/DNA hybrids formed at transcriptional pause sites to promote Xrn2-dependent termination. *Mol. Cell* 42, 794–805

109 West, S. *et al.* (2004) Human 5'→3' exonuclease Xrn2 promotes transcription termination at co-transcriptional cleavage sites. *Nature* 432, 522–525

110 Kim, M. *et al.* (2004) The yeast Rat1 exonuclease promotes transcription termination by RNA polymerase II. *Nature* 432, 517–522

111 Sadowski, M. *et al.* (2003) Independent functions of yeast Pcf11p in pre-mRNA 3' end processing and in transcription termination. *EMBO J.* 22, 2167–2177

112 Meinhart, A. and Cramer, P. (2004) Recognition of RNA polymerase II carboxy-terminal domain by 3'-RNA-processing factors. *Nature* 430, 223–226

113 Steinmetz, E.J. *et al.* (2001) RNA-binding protein Nrd1 directs poly(A)-independent 3'-end formation of RNA polymerase II transcripts. *Nature* 413, 327–331

114 Vasiljeva, L. *et al.* (2008) The Nrd1-Nab3-Sen1 termination complex interacts with the Ser5-phosphorylated RNA polymerase II C-terminal domain. *Nat. Struct. Mol. Biol.* 15, 795–804

115 Carroll, K.L. *et al.* (2004) Identification of cis elements directing termination of yeast nonpolyadenylated snoRNA transcripts. *Mol. Cell. Biol.* 24, 6241–6252

116 Mayer, A. *et al.* (2010) Uniform transitions of the general RNA polymerase II transcription complex. *Nat. Struct. Mol. Biol.* 17, 1272–1278

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

[118] Murray, S.C. *et al.* (2012) A pre-initiation complex at the 3'-end of genes drives antisense transcription independent of divergent sense transcription. *Nucleic Acids Res.* 40, 2432–2444