Review Article

ZC3H4/Restrictor Exerts a Stranglehold on Pervasive Transcription

Chris Estell * **and Steven West** *

The Living Systems Institute, *University of Exeter, Stocker Road, Exeter EX4 4QD, UK*

Correspondence to Chris Estell and Steven West: *c.estell@exeter.ac.uk (C. Estell), s.west@exeter.ac.uk (S. West)*
https://doi.org/10.1016/j.jmb.2024.168707
Edited by Stephen Buratowski

Abstract

The regulation of transcription by RNA polymerase II (RNAPII) underpins all cellular processes and is perturbed in thousands of diseases. In humans, RNAPII transcribes ~20000 protein-coding genes and engages in apparently futile non-coding transcription at thousands of other sites. Despite being so ubiquitous, this transcription is usually attenuated soon after initiation and the resulting products are immediately degraded by the nuclear exosome. We and others have recently described a new complex, "Restrictor", which appears to control such unproductive transcription. Underpinned by the RNA binding protein, ZC3H4, Restrictor curtails unproductive/pervasive transcription genome-wide. Here, we discuss these recent discoveries and speculate on some of the many unknowns regarding Restrictor function and mechanism.

Introduction

Approximately 80% of the human genome can be transcribed by RNA polymerase II (RNAPII).[1–3] While some RNA synthesis derives from protein-coding genes, most occurs in non-coding regions. This abundance of apparently unproductive/non-coding (nc)RNA synthesis is often termed pervasive transcription (PT). PT can be defined as any transcription that occurs at a genomic region that does not generate classical "functional" RNAs, such as mRNA, snRNA, rRNA, miRNA.[4,5] Regions of the genome that are susceptible to PT are enriched for CpG islands and contain a low number of nucleosomes – so-called nucleosome depleted regions (NDR).[6–8] While most PT RNA products have no described function, PT may play important roles (e.g. by maintaining open chromatin). Some of these transcripts have a poly(A) tail,[9] providing functional potential; however, most PT products are rapidly degraded and these are a focus of this review.

Although PT can occur in intergenic regions, a classic example is the generation of PROMoter uPstream Transcripts (PROMPTs).[10,11] Here, as a polymerase transcribes a protein-coding gene creating an mRNA, RNAPII initiates transcription upstream and in the opposite direction to synthesise a PROMPT (Figure 1).[12] This upstream transcription might initiate opportunistically due to the open chromatin generated by initiation in the sense direction or from an independent promoter structure. Regardless, antisense transcription usually terminates within a few kilobases (kb), with the resulting short RNA being degraded by the exosome *via* its exoribonuclease, DIS3.[10,13,14] This contrasts with genic transcription, where polymerases can transcribe long distances and generate stable and functional RNAs. This poses two fundamental questions: what selectively terminates PT, and what promotes protein-coding transcription?

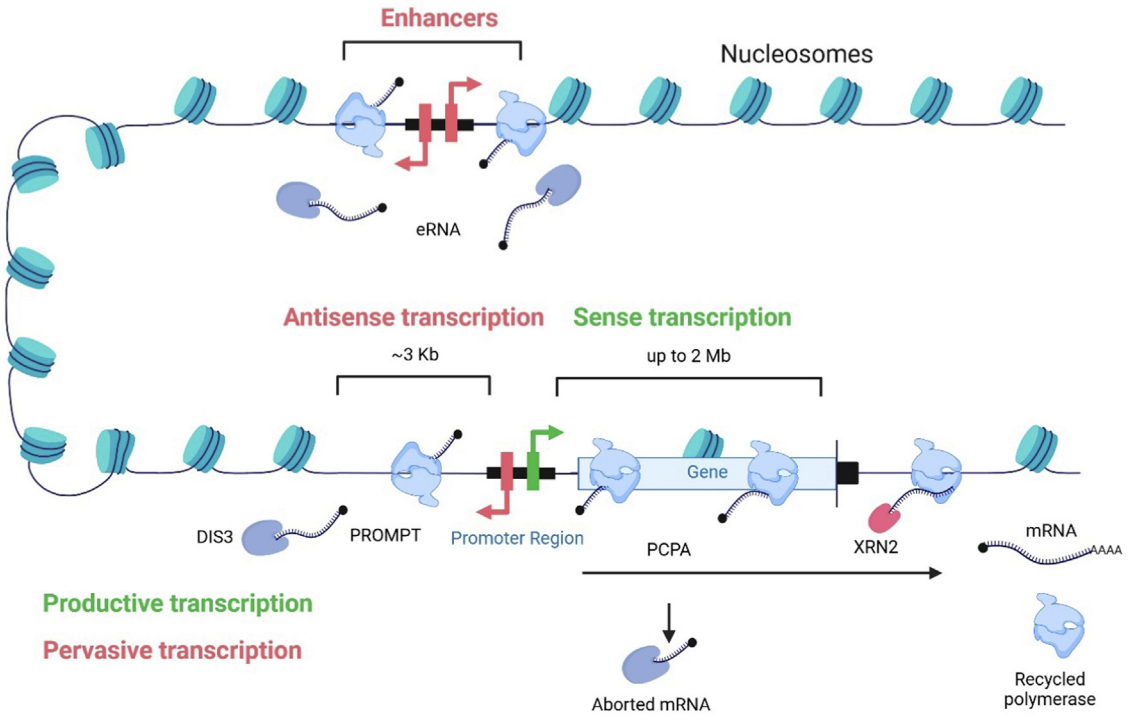

Figure 1. Pervasive transcription. Nucleosome-depleted regions are found around the promoters of protein-coding genes and enhancers, where RNAPII can access the DNA and start transcribing. Productive transcription (green) can occur over great distances before being terminated beyond the polyadenylation signal (PAS) *via* CPSF and the XRN2 torpedo. PCPA denotes the premature cleavage and polyadenylation that can occur if U1 snRNA levels are deficient. Pervasive transcription (red) is often targeted for attenuation by Integrator or Restrictor, with the resulting RNAs degraded by the nuclear exosome *via* its catalytic DIS3 subunit.

Metazoans have numerous complexes that promote transcriptional termination. The archetypal example operates at the 3′ end of protein-coding genes and depends on 3′ end processing of nascent transcripts by cleavage and polyadenylation (CPA), which requires a polyadenylation signal (PAS) that is recognised and cleaved by the CPA complex.[15,16] As a result, an RNAPII-associated 3′ cleavage product is exposed to 5′ → 3′ degradation by XRN2, leading to transcriptional termination.[17–20] Although required to terminate most protein-coding transcription, CPA and XRN2 are less frequently employed to terminate PT.[13,18,21] Instead, additional complexes are emerging, which selectively control such unproductive/pervasive transcription.

One of these complexes is Integrator, which contains 15 subunits.[22] Its INTS11 subunit has endonuclease activity, while its INTS6 component links Integrator to the PP2A phosphatase.[23,24] Integrator was originally discovered through its role in the 3′ end processing of snRNAs.[25] However, another major role of Integrator is to control promoter-proximal transcription, which it does genome-wide.[26–29] Through this activity, Integrator attenuates a large fraction of PT, *via* its interactions with promoter-proximal RNAPII and negative elongation factor (NELF).[30,31] Integrator is not the major subject of this review, but its molecular mechanism was recently described and its roles in transcription are reviewed in detail elsewhere.[22,30,32]

We and others recently discovered another factor, ZC3H4, which also restricts PT.[33–38] ZC3H4 interacts with several factors including WDR82, ARS2, the Nuclear EXosome Targeting (NEXT) complex and Casein Kinase 2 (CK2).[33,34,36,38,39] Based on its impact on transcription, this complex has been named Restrictor.[37] In this review, we will outline what is known about Restrictor, and speculate on the answers to many questions regarding its function(s).

Su(s)sing out the origin of ZC3H4

Although ZC3H4/Restrictor has recently come to the fore, these discoveries were foreshadowed by work from the lab of Lillie Searles who was investigating the mechanism behind suppressor genes in *Drosophila melanogaster*.[40] In genetics, a secondary mutation can be described as a suppressor if it rescues (suppresses) the phenotype of a primary mutation elsewhere within the genome. A popular model to study suppressors was the *vermilion* gene, which encodes tryptophan oxygenase (TO) – the enzyme responsible for brown eye colour in *Drosophila melanogaster*.[41] Normally, fly

eye colour is a mosaic of red and brown, but spontaneous mutations of the *vermilion* gene give rise to a red eye phenotype due to the absence of TO. Genetically, these spontaneous mutations can be caused by inserting a *412*-type transposon into the 5′ untranslated region (5′ UTR) of the *vermilion* gene.[40,42] Interestingly, some wild-type (WT) *vermilion* mRNA was still detected *via* northern blot, even when *412* was present in both alleles.[43] This is because the *vermilion* locus was still transcribed following the transposon insertion and cryptic splice sites, found within *412,* were being recognised by the spliceosome. Utilisation of these splice sites resulted in most of the transposon being spliced out, giving rise to some near-wildtype (WT) *vermilion* mRNA – although at levels too low to maintain normal eye colour.[44] This primary mutation-induced red eye phenotype can be reversed (suppressed) by secondary mutations in the fly orthologue of ZC3H4 – suppressor of sable (Su(s)). The low level of WT-TO mRNA in primary mutant *vermilion* strains is increased when Su(s) is also mutated, suggesting that Su(s) could regulate the *vermilion* locus by influencing splicing or transcription.

Su(s) is approximately 150 kDa (kd) in size and localises to the nucleoplasm and chromatin. Immunofluorescence of polytene spreads shows that while Su(s) binds throughout the genome, it is strongly enriched at a subset of loci.[45,46] It contains two arginine-rich RNA binding motifs: (ARM1/2) and an acidic-rich region (ACR) upstream of two CCCH zinc fingers (ZF).[47,48] Identifying its specific RNA substrates was challenging as recombinant Su(s) bound a broad range of RNAs.[46] Subsequently, the systematic evolution of ligands by exponential enrichment (SELEX) and RNA footprinting was employed to identify substrates. These techniques revealed that Su(s) can bind to a UCAGUAGUCU motif with downstream G/U-rich regions as well as regions of GU dinucleotides.[46,48] Biochemical studies revealed that Su(s) binding to the UCAGUAGUCU consensus came from its ARM1 alone rather than involving the ZFs.[48] Interestingly, ARM2 was found to bind consensus site mutants stronger than ARM1, while only ARM1 was required for Su(s) to bind polytene chromosomes. The GU motifs within this consensus may have some capacity to bind spliceosomal U1 small nuclear ribonucleoprotein complexes (herein abbreviated to U1) since GU is the most important dinucleotide motif within the 5′ splice site. Indeed, it was proposed that Su(s) served to destabilise RNA transcribed from the *412* transposon of *vermilion* mutants by competing with spliceosome assembly at suboptimal splice sites.[49] Consistently, the *412* silencing effects, alleviated by Su(s) depletion, are also reversed by converting these suboptimal splice sites to consensus sequences.[49] However, SELEX utilises a library of random sequences flanked by priming/T7 RNA polymerase promoter sites, which are then used for amplification and *in vitro* transcription. It was noted that the identified UCAGUAGUCU motif was predicted to form a hairpin with the 5′ priming region, raising the possibility that Su(s) could also/instead recognises secondary structure.

The last three studies of Su(s) by Searles and colleagues moved away from *vermilion,* broadening its substrates, and providing further mechanistic understanding. Using a reporter, they demonstrated that artificially tethered Su(s) suppressed transcription – providing initial evidence that Su(s) impacts RNA at the level of transcription.[50] Next, they reported that Su(s) regulates aberrant RNAs generated after heat shock, which are subsequently targeted by the exosome.[51] Interestingly, the ZF region of Su(s), and not the ARMS, were required for the recognition of these RNAs. This contrasts with how Su(s) binds polytene chromosomes/RNA in a WT context, demonstrating a multi-modal interaction of Su(s) with its substrates. They identified the first interacting partner of Su(s), WDR82, which is essential for its function and demonstrated that Su(s) acts co-transcriptionally.[52] Importantly, this study provided the first evidence that Su(s) and WDR82 induce transcriptional termination. Finally, they hypothesised that ZC3H4 was the human orthologue of Su(s). Although there is limited amino acid conservation between the two proteins, their regions of predicted order and disorder show striking overlap.

From fly to human

The earliest mention of ZC3H4 in humans derived from its presence in an interactome analysis of WDR82 – a WD repeat-containing protein.[53] Although the ZC3H4-WDR82 interaction was not characterised further, this study showed that WDR82 forms distinct complexes with PNUTS-PP1 phosphatase and the SET1A/B histone methyl transferase complex.[54] As such, these data uncovered WDR82 as a lynchpin in several assemblies – the relevance of which will be discussed below. We serendipitously uncovered the link between ZC3H4 and transcription by screening for proteins with reduced proximity to RNAPII following the elimination of the PAS recognition factor CPSF30.[34] Apart from other CPA factors, ZC3H4 and its paralogue, ZC3H6, were the most prominent hits from this screen. We hoped that ZC3H4 and/or ZC3H6 would turn out to be part of the PAS processing and transcriptional termination pathway. However, to our surprise (and initial dismay), this was not the case. Instead, ZC3H4 depletion (but not ZC3H6 loss) induced widespread upregulation and extension of ncRNAs highlighting it as a factor that restricts PT.

The clearest effects of ZC3H4 loss were on PROMPTs and enhancer RNAs; with clusters of the latter (so-called super-enhancers (SEs)) being most obviously impacted.[55,34] SEs contain multiple closely spaced promoters and their short RNAs

are degraded by DIS3, giving the impression of transcriptional "islands" when viewed *via* a genome browser after DIS3 depletion. When ZC3H4 is depleted, these normally short RNAs are extended. The clear difference in these two transcriptional profiles likely reflects the order in which these two factors operate. As DIS3 recognises free RNA 3′ ends, it presumably degrades the products of transcriptional termination. However, the longer transcripts observed following ZC3H4 depletion represent a failure of transcriptional termination and suggest that Restrictor acts upstream of DIS3 – as was proposed for Su(s).[33,34,52] Consistent with ZC3H4 somehow promoting transcriptional termination, these extended PROMPT and SE transcripts remain associated with RNAPII following its depletion.[33]

ZC3H4 and its interactors

Figure 2 displays a linear view of ZC3H4, its known domains, and its major interactors described to date: WDR82, CK2, ARS2 and ZCCHC8 (a component of the NEXT complex). The three CCCH ZFs within ZC3H4 encompass amino acids 396–467. Upstream of these ZFs is an RGG/RG repeat region, which is required for ZC3H4 to bind RNA.[33] Deletion of an N-terminal portion of ZC3H4 that incorporated these repeats corresponds to the region of Su(s) that mediates its RNA binding *via* ARM regions.[33,48] ARM regions have widely been reported to bind secondary structures such as hairpins, while RGG/RG regions bind poly(G, U, C) tracks.[56–59] Although we demonstrated that the RGG/RG regions are likely important for ZC3H4 to bind RNA, the relevance of the ZF regions is unclear.[33] This is because a ZC3H4 derivative bearing three CCCH > AAAA mutations restored a large fraction of PT after the loss of endogenous ZC3H4.[33] In contrast, ZC3H4 lacking the RNA binding N-terminal region cannot support its transcriptional restriction function. Perhaps the ZF's of ZC3H4 function in the same manner as those in Su(s) and help to recognise induced substrates i.e. those arising from heat shock. Another possibility is that ZC3H4 interacts with RNA in a

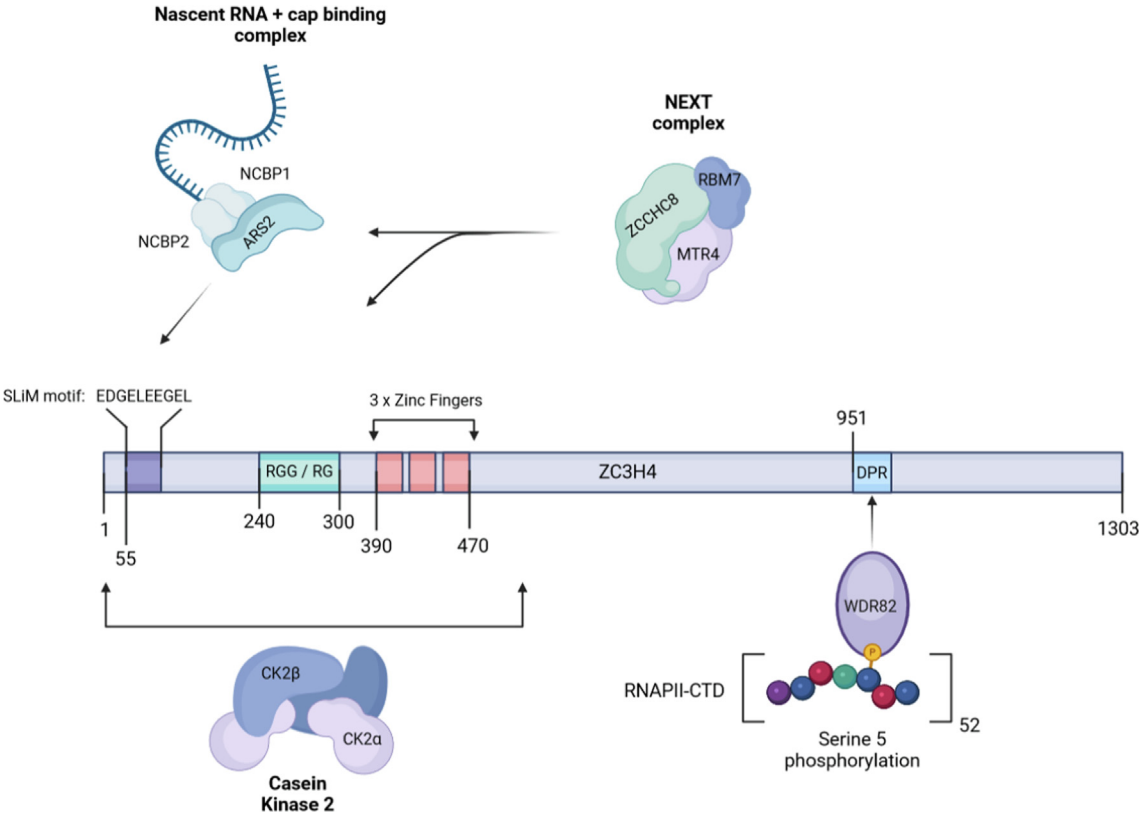

Figure 2. ZC3H4 and its validated interactors. ZC3H4 is largely disordered, with only the three CCCH zinc fingers and WDR82 binding regions showing some order. The N-terminal region interacts with RNA, presumably mediated *via* the RGG / RG-rich regions. The DPR amino acid triplet within ZC3H4, important for WDR82 binding, spans amino acids 951–953. WDR82 binds Ser5p-modified RNAPII and may thereby recruit ZC3H4 to such polymerases. ARS2 bridges Restrictor to its substrates by binding both Nuclear Cap Binding Proteins (NCBP1/2) and the SLiM motif (EDGELEEGEL) within ZC3H4. ARS2 and ZC3H4 recruit the NEXT complex to target Restrictor-attenuated RNAs to the exosome. Casein Kinase 2α binds the CK2 tetramer to the N-terminal region between amino acids 1 – 530.

bipartite manner like the RNA binding protein FUS – where the RGG/RG regions enhance RNA binding and its ZFs bind a specific sequence.[60] Interestingly, FUS uses this binding mode to interact with, and open, structured substrates and it will be interesting to determine whether ZC3H4 can also use this mechanism.

WDR82 is a member of the large family of WD repeat-containing proteins, the most prominent ZC3H4 interactor, and associates with the Carboxyl-terminal domain (CTD) of RNAPII's large subunit when it is phosphorylated on Serine 5 (Ser5p).[38,61] Ser5p is most abundant in proximity to the TSS,[62] and an attractive model is that the interaction between WDR82 and RNAPII supports the promoter-proximal location and function of ZC3H4. The importance of WDR82 for controlling extragenic PT was first revealed by Austenaa et al., who described the widespread extension of such transcription following its depletion.[36] The WDR82-ZC3H4 interaction was initially characterised as between amino acids 804–1057.[35] Further dissection revealed an aspartate-proline-arginine (DPR) amino acid triplet (starting at D951) within this region, which mediates this interaction.[33] This triplet has structural overlap with a DPR motif in SET1A, which is also required for its association with WDR82.[33,63] This may explain the mutually exclusive association of ZC3H4 and SET1 complexes with WDR82.[35,53] The functional relevance of the WDR82-ZC3H4 interaction is clear because, when over-expressed, the 804–1057 fragment of ZC3H4 acts in a dominant-negative fashion to dysregulate PT.[35] A likely explanation is that this fragment acts as a "sponge" for free WDR82 thereby preventing its association with endogenous ZC3H4. Consistently, mutating this fragment to prevent its interaction with WDR82 abolishes its ability to inhibit Restrictor.[33] Although expression of the WDR82-interacting fragment of ZC3H4 disrupts Restrictor function, WDR82 appears to play a broader role in PT regulation than ZC3H4.[33] This is presumably because WDR82 loss could compromise multiple transcription regulatory assemblies that it is part of.

ZC3H4 also interacts with Casein Kinase 2 (CK2), which is a tetrameric complex comprising two catalytic (CK2α) and two regulatory subunits (CK2β).[64] CK2 interacts with the N-terminal region of ZC3H4 and was found to phosphorylate the critical RNAPII elongation factor, SPT5.[38,65,66] Phosphorylation of SPT5, by cyclin-dependent kinase 9 (CDK9), is required for RNAPII to escape the promoter-proximal pause and engage in efficient elongation.[67–69] However, CK2 phosphorylates multiple serine and tyrosine residues in the N-terminal region of SPT5, distinct from those known to be targeted by CDK9.[38] In the authors' model, phosphorylation of the N-terminal region of SPT5 aids RNAPII in transitioning from early to productive elongation in the sense direction. Depletion of ZC3H4 disturbs this balance, resulting in a down-regulation of genic transcription and an increase in upstream antisense transcription. However, ZC3H4 depletion does not generally reduce sense transcription suggesting that CK2 activity may act *via* additional mechanisms to restrict PT. For instance, the phosphorylation of SPT5 by CK2 may play an inhibitory role by restricting the elongation potential of RNAPII during PT.

The N-terminal region of ZC3H4 also interacts with ARS2, a multi-domain protein that can act in various transcriptional termination and maturation pathways.[39,70–72] ARS2 was originally shown to preferentially promote transcriptional termination of short RNAs by an unknown mechanism.[70,72] ARS2 binds the cap-binding complex (CBC) where it coordinates transcriptional termination and degradation by the exosome *via* the NEXT complex, which contains the helicase MTR4, the ZCCHC8 scaffolding factor, and the RNA binding protein RBM7.[73–75] NEXT recruitment to ARS2-CBC conventionally involves ZC3H18 as an adaptor.[72] However, ZC3H4 can recruit the NEXT complex independently of ZC3H18 supporting a model for the direct handover of Restrictor-terminated RNA to the degradation machinery.[39] ZC3H4 was identified as a prominent ARS2 interactor *via* a Short Linear Motif (SLiM) of amino acids (EDGELEEGEL) within its N-terminal region.

The same SLiM is present in numerous ARS2 interactors, and its likely function is to ensure the mutually exclusive participation of ARS2 in different complexes to coordinate distinct aspects of ncRNA biogenesis.[76,77] In this respect, the SLiM motif has some analogy to the DPR motif that ZC3H4 and SET1A use for mutually exclusive interaction with WDR82. Importantly, the recruitment of ZC3H4 to selected loci was strongly diminished by the depletion of ARS2.[39] Moreover, ZC3H4 lacking the ability to bind ARS2 loses its restriction capacity.[33] Therefore, the ZC3H4-ARS2 interaction may explain the long-known function of ARS2 in terminating ncRNA transcription and provide one way of locating Restrictor to relevant substrates via an interaction with the 5′ end of the nascent RNA. However, at some Restrictor targets, depleting ZC3H4 induces a stronger transcriptional defect than the loss of ARS2.[33] As was suggested for Su(s), this may reflect multiple modes of recruiting Restrictor to targets (e.g. *via* the interaction of WDR82 with Ser5p RNAPII).

What defines a Restrictor target?

The function of ZC3H4 in restricting unproductive transcription was independently described by Gioacchino Natoli's lab, which similarly showed its dominant impact on extragenic PT.[35] They also found that not all PT is sensitive to ZC3H4, again suggesting that it recognises an RNA or DNA feature. In an elegant experiment, they ruled out the participation of promoter elements in this process.

At a bidirectional promoter, where transcription in the upstream antisense direction was Restrictor sensitive and sense transcription was not, flipping the promoter around did not affect the Restrictor-sensitivity of the antisense transcript. Therefore, an element in the transcribed RNA is important.

Austenaa et al., also noted that at least some ncRNAs that were stabilised by ZC3H4 or WDR82 loss underwent chaotic splicing suggesting that Restrictor depletion unmasked otherwise inefficient splice sites.[35] This synergised with the work on Su(s) where it was hypothesised that Su(s) targeted transcripts when spliceosome assembly was inefficient.[49] Importantly, the deletion of promoter-proximal cryptic (or chaotic) splice sites stabilised Restrictor substrates that were consequently no longer affected by depleting WDR82 or ZC3H4.[35] Therefore, for both Su(s) and Restrictor, there is a clear connection between the sensitivity of transcription to this pathway and the efficiency of splicing/spliceosome assembly. Nevertheless, while Austenaa et al., showed that RNA is stabilised by deleting the splice sites, Fridell et al., noted that strengthening weak splice sites provided resistance to Su(s). This suggests that the relationship between the splicing potential of a transcript and its sensitivity to Restrictor is not straightforward and exactly what underpins this interplay remains to be fully elucidated. Notably, the splice sites that become active upon the depletion of ZC3H4/WDR82 show little or no sequence deviation from those in transcripts unaffected by Restrictor.[35] Therefore, additional sequence elements or characteristics of RNAPII may also be important (more on the latter below).

In HeLa and HCT116 cells, ZC3H4 depletion mainly affects PT. However, a subset of mRNAs is upregulated in its absence, suggesting a competition between their attenuation by Restrictor and their full-length expression.[34,35] In mouse embryonic stem cells (mESCs), a large number (over 2500) of protein-coding genes are affected by ZC3H4.[63] That study demonstrated that the histone methyl transferase, SET1A, promotes the expression of low to moderately-expressed genes transcribed from CpG island promoters. While the transcription of such genes was downregulated by depleting SET1A, their expression was restored by co-depleting ZC3H4. This suggests that SET1A supports the transcription of these genes by opposing Restrictor. We recently demonstrated that ZC3H4 and WDR82 target the transcription of hundreds of protein-coding genes when U1 cannot be recruited to the nascent RNA.[33] In analogy to SET1A, U1 is well-known to enhance RNAPII elongation and suppress premature transcriptional termination.[78,79] These data suggest that many transcripts have the potential to be targeted by Restrictor unless opposed by elongation-promoting activities. Indeed, U1 sites are known to be rarer in PROMPT transcripts which are often constitutively targeted by Restrictor.[80,81] Because U1 can enhance both splicing and elongation, effects on the latter might be an important determinant of Restrictor-sensitivity additional to the above-mentioned impact of splicing and spliceosome assembly.

These observations suggest that Restrictor affects protein-coding transcription in certain contexts, whereas many ncRNAs are constitutively sensitive to Restrictor loss. Thus, in addition to recognising nucleic acid features, Restrictor function may be enabled by/require certain forms of RNAPII. This could explain why ZC3H4 is recruited to most RNAPII promoters while its loss has selective effects on transcription.[33] An extension of this idea is that factors such as SET1A and U1 directly or indirectly promote the eviction of Restrictor as RNAPII complexes mature for full elongation.

Proximal labelling of ZC3H4 reveals neighbouring pathways

Despite the above-discussed transcriptomics and interactome studies, how ZC3H4 promotes transcriptional termination remains a mystery. Of the prominent ZC3H4 interacting factors, none have the known ability to evict RNAPII from chromatin. This might indicate that ZC3H4 indirectly promotes termination by other pathways, or that other relevant factors are only transiently associated with it. To address this further, we performed a proximity interactome study of ZC3H4 using TurboID.[33,82] This revealed Restrictor to be proximal to multiple complexes, the most relevant to transcriptional control and termination being PNUTS-PP1, PAF1, and CPSF.

PP1 Nuclear-Targeting Subunit (PNUTS) is a highly conserved protein that interacts with PP1 phosphatases to regulate their nuclear functions.[83] PNUTS-PP1 is part of the CPA complex and dephosphorylates SPT5 to slow transcription and enable Torpedo-dependent termination beyond the PAS.[18,20,84,85] This negative regulation of transcription by PNUTS may sensitise RNAPII to Restrictor. Alternatively, PNUTS may cooperate with Restrictor to keep RNAPII in a poorly processive state. The promotion of restriction by a negative transcriptional regulator is reciprocal to the function of a positive factor (U1) in counteracting Restrictor.

Interestingly, PNUTS-PP1 orthologues also play a critical role in terminating kinetoplastid RNAPII.[86–88] In that mechanism, polymerase arrest at a hypermodified “base J” helps promote transcriptional termination.[89] By restricting the elongation capacity of RNAPII, Restrictor may similarly enable PNUTS-mediated transcriptional termination at some loci. Recent *in vitro* studies suggest that the yeast orthologue of PNUTS (Ref2) may terminate transcription directly by promoting an RNAPII dimer that is incompatible with transcription.[90]

A three-dimensional structure of this arrested dimer has been solved and it will be exciting to see if this dimerization is observed in whole cells. Furthermore, a plant orthologue of PNUTS (LUMINIDEPENDENS) associates with a histone demethylase complex (FLD-SDG26) that removes H3K4me1 from 3′ end nucleosomes to create a less favourable environment for future rounds of transcription.[91]

We performed a limited analysis of ZC3H4-affected ncRNAs, which confirmed their sensitivity to both Restrictor and PNUTS depletion.[33] Because the depletion of both factors did not cause synergistic effects, we suggested that Restrictor and PNUTS act in the same pathway in those cases.[33] A separate analysis of the proximal ZC3H4 interactome similarly identified PNUTS but reached some different conclusions on its role in Restrictor function.[92] This was based on transcriptome-wide analyses following the depletion of each factor individually or both together. These analyses revealed some overlap between PNUTS and Restrictor effects, again indicating that they comprise a common pathway. However, the transcriptome-wide approach uncovered many transcripts that did not fit this model, suggesting the synergistic action of PNUTS and Restrictor in other cases. Interestingly, some of the exemplar transcripts used in both studies were affected differently by PNUTS loss. For example, we showed that the PROMPT upstream of *ITPRID2* is substantially extended after PNUTS depletion whereas Russo et al. described no effect whatsoever. This discrepancy is not explained by cell line differences since both studies used HCT116 cells. However, Russo et al., employed dTAG-13 to deplete PNUTS-dTAG, which works *via* a cereblon (CRBN) E3 ligase. The CRBN-based dTAG degraders are known to function poorly in HCT116 cells,[93,94] which may explain the partial depletion of PNUTS observed by Russo et al. and why some transcripts we observed as strongly affected by PNUTS loss were unaffected in their study. Despite this discrepancy, these studies highlight a connection between Restrictor and PNUTS-PP1 that will be a fertile area for future research.

Many components of the protein-coding 3′ end formation machinery are also proximal to Restrictor. The most prominent is SYMPLEKIN, which is part of the CPSF complex.[95] Co-depletion of SYMPLEKIN with ZC3H4 revealed a synergistic increase of PROMPT levels, suggesting they function in separate or redundant pathways, something we also suggested when evaluating the interaction between ZC3H4 and the 3′ processing endonuclease CPSF73.[19,33] Interestingly, the transcriptional readthrough at protein-coding genes caused by SYMPLEKIN depletion was sometimes reversed by simultaneous ZC3H4 depletion. This is thought to be because limiting levels of SYMPLEKIN are sequestered away from protein-coding 3′ ends in a ZC3H4-dependent manner. As such, co-depletion of ZC3H4 makes these trace levels of SYMPLEKIN available once again for CPA-dependent termination. The proximity of Restrictor and the CPA machinery may be why we initially discovered the reduced proximity of ZC3H4 and RNAPII after depleting CPSF30.[34] These data highlight the interesting possibility of a network of regulation between different transcriptional termination pathways exposed by tuning the levels of their key components.

Finally, the PAF complex is also proximal to ZC3H4. This comprises PAF1, CTR9, RTF1, CDC73, LEO1, and WDR61 (in humans only).[96–98] It was first identified by two groups who found it associated with RNAPII, and it has now been linked with regulating pause release in addition to stimulating RNAPII elongation.[99–102] Interestingly, RNAi depletion of the PAF complex was shown to impair the degradation of ncRNAs, including Restrictor substrates (i.e. PROMPTs).[103] As the PAF complex is proximal to Restrictor, it may connect RNAPII elongation with Restrictor function and ncRNA turnover by the exosome. However, the rapid elimination of PAF components does not clearly impact PT, suggesting that its role in ncRNA turnover might be a secondary effect.[104,105] As for when ZC3H4 targets protein-coding genes, any functional interaction with PAF may be context-dependent. The association of ZC3H4 with PAF distinguishes its transcriptional attenuation function from that of Integrator which associates with NELF (more below).[31] Interestingly, ZC3H4 shows a reduced association with RNAPII following the acute depletion of PAF[106] – an observation consistent with Restrictor recruitment after PAF complex loading. Table 1 summarises the factors that have been linked to the Restrictor pathway so far:

Different RNAPII termination pathways related or relevant to Restrictor

ZC3H4 and its orthologues appear to be largely restricted to Metazoans.[34] Nevertheless, PT is common to all cells and organisms and a variety of dedicated transcriptional termination mechanisms are used to control it. In the below section, we discuss some other examples with similarity to Restrictor, focussing on eukaryotic RNAPII.

The yeast Nrd1-Nab3-Sen1 complex

Nrd1-Nab3-Sen1 (NNS) is a complex that regulates the transcriptional termination of budding yeast transcripts including snRNA, snoRNAs, cryptic unstable transcripts (CUTs, which are the likely equivalent of mammalian PROMPTs), and some protein-coding RNAs.[107–110] At CUTs, it has been widely reported that Nrd1-Nab3 are recruited to Ser5p RNAPII, where this heterodimer then recognises UGUAG and GNUCUUGU motifs (respectively) within nascent

Table 1 Factors related to Restrictor function.

Interactor	Direct	Proximal	Other	Experimental evidence	Role	Notes
Factors that support Restrictor function						
WDR82	✓			IP-MS, IP-WB, *in-vitro* reconstitution	Bridges ZC3H4 to RNAPII *via* Ser5p residues in RNAPIIs CTD.	Binds ZC3H4 at DPR$^{951–953}$ and also part of SET1A, PNUTS-PP1, and PCF11 complexes.
ARS2	✓			IP-MS, IP-WB, *in-vitro* reconstitution	Recruits ZC3H4 to chromatin.	Binds capped RNA via CBC where it and multiple factors interacts with ZC3H4s N-terminal region.
CK2	✓			IP-MS, IP-WB, *in-vitro* reconstitution	Both subunits form the tetrameric casein kinase 2 which phosphorylates SPT5.	Ser/Thr kinase that phosphorylates multiple targets in the Restrictor pathway. May phosphorylate other factors in the Restrictor pathway.
RNAPII (CTD)	✓			IP-WB. *in-vitro* reconstitution	Transcription.	Restrictor acts co-transcriptionally on Ser5p RNAPII, to which it could be recruited by WDR82.
ZCCHC8	✓			IP-MS, IP-WB,	Part of the NEXT complex, interacts with Restrictor and hands aborted transcripts to the exosome for turnover.	Forms NEXT with MTR4 and RBM7.
SYMPLEKIN		✓		ZC3H4-Turbo	Not directly involved in the Restrictor pathway but acts synergistically in the same regions.	Part of the CPA machinery.
PAF1	‘	✓		ZC3H4-Turbo, RNAPII IP-MS	Unknown. ZC3H4 no longer associates with RNAPII following PAF1 depletion.	Interacts with RNAPII as part of the PAF1 complex where it stimulates elongation.
PNUTS		✓		ZC3H4-Turbo	Dephosphorylates unknown factors.	Dephosphorylates SPT5 and RNAPII facilitating transcription termination. It is unknown if this antagonises ZC3H4-CK2 at SPT5.
SPT5			✓	*in-vitro* kinase assay	Residues in its N-terminal domain selectively impact antisense transcription and are phosphorylated by ZC3H4-CK2.	Forms part of DSIF along with SPT4; phosphorylated by CDK9 to promote productive transcription.
Factors that oppose Restrictor function						
U1 snRNP			✓	Co-depletion, INSERT-seq	Promotes RNAPII elongation, shielding protein-coding transcription from Restrictor	Binds the 5′ splice site of mRNA during spliceosome assembly; acts as an elongation factor.
SET1A			✓	Co-depletion	Promotes the expression of moderately expressed genes by opposing Restrictor.	Binds WDR82 *via* a DPR motif in its N-terminal region.

RNA.[111–113] Transcriptional termination is thought to occur after the recognition of these motifs, when the RNA:DNA helicase Sen1 is loaded onto the NNS substrate and dislodges RNAPII from the DNA template/nascent RNA using its RNA:DNA helicase activity.[114] Next, the CTD interacting domain of Nrd1 interacts with Trf4 of the TRf4-Air2-Mtr4p Polyadenylation (TRAMP) complex, which polyadenylates the CUTs and hands them over to the exosome for degradation.[115,116,12] More recently, Collin et al. only observed read-through of CUTs when Tyr1 of the CTD was mutated, not Ser5 (the transcriptional profile of Ser5 loss was a lack of promoter escape).[117] They also observed pause sites at NNS substrates that were abolished in the Tyr1 mutant and upon depletion of the phosphatase Glc7. Their favoured model is that Glc7 dephosphorylates Tyr1 of the CTD to slow or pause transcription. This pause is required because *in vitro* studies suggest that Sen1 is only capable of terminating paused transcription complexes due to being a poorly processive helicase.[118]

It had been thought that NNS has no mammalian equivalent because no obvious sequence similarity exists between Nrd1/Nab3 and human proteins; it should also be noted that, despite being orthologues, the sequence conservation between ZC3H4 and Su(s) is limited.[52] There is a striking similarity between several aspects of NNS function and those of Restrictor: both pathways regulate the termination of ncRNA, hand terminated transcripts over to the exosome for degradation, and regulate a subset of protein-coding genes.[34,35] Recognition of these substrates occurs *via* RNA binding with GU-rich motifs reported as preferred sites for both complexes, which appear to regulate transcriptional termination independent of RNA cleavage.[33] Interestingly, both ZC3H4 and Nrd1 autoregulate the expression of their own transcripts.[119,34,35] As for Nrd1, ZC3H4 interacts with Ser5p (although possibly indirectly *via* WDR82).[38,111] However, an unbiased proteomic study indicated a potential interaction between ZC3H4 and the RPB6 subunit of RNAPII.[120] This could mean that ZC3H4 interacts with RNAPII in a bimodal way, as recently suggested for NNS.[117] In yeast, Ser5p of the CTD is widely reported to recruit NNS to enable the termination of snoRNA and CUTs.[121,122] However, Collin et al., observed that while the CTD was required for termination, it was not required to recruit Nrd1 or Sen1 to snoRNA loci. Using this analogy, perhaps ZC3H4 mainly associates with RPB6 in an inactive state before moving to the CTD to enact Restriction. This could explain the strong similarities between ZC3H4 and RNAPII chromatin occupancy profiles, although not the limited number of affected PROMPTs. As a general note, whether ZC3H4 associates with all/most promoter-proximal RNAPII or is recruited in response to a problem with transcription is unknown.

A key component of the NNS pathway proposed by Collin et al. is the dephosphorylation of Tyr1 by Glc7. This is interesting as, in humans, Tyr1 phosphorylation is enriched at PROMPT and eRNA regions where ZC3H4 acts.[123] Furthermore, Glc7 interacts with Swd2 and Ref2, the budding yeast orthologues of WDR82 and PNUTS, respectively.[124] Finally, the human orthologue of Sen1 (SETX) has been reported to play a role in both protein-coding gene transcriptional termination and the resolution of r-loops.[125] Whether SETX is involved in Restrictor function is untested although it was found (albeit low down the list) as proximal to ZC3H4 *via* TurboID.[33] The genome-wide role of SETX in mammalian RNAPII transcriptional termination remains to be fully determined.

Regulation of HIV transcription

Although not a eukaryote, attenuation of human immunodeficiency virus (HIV) transcription shares features with the control of PT in mammals. HIV provirus is transcribed by RNAPII, which pauses shortly after initiating and transcribing a so-called *trans*-activation response (TAR) element. Release of paused RNAPII occurs by the recruitment of CDK9 to TAR by the *trans*-activator of transcription (TAT) protein.[126] Thus, promoter-proximal regulation of HIV transcription is analogous to most mammalian protein-coding genes. This promoter-proximally paused RNAPII can be attenuated by PCF11, a CPA factor at protein-coding genes.[127] Interestingly, PCF11 is one of the few factors shown to dissociate RNAPII from DNA in a purified system.[128] More recently, PCF11 was found at the HIV promoter in complex with WDR82 providing some analogy with the control of PT in mammals.[129] Like ZC3H4, PCF11 terminates polymerases using a combination of RNA and RNAPII binding. While it is unknown if ZC3H4 can dissociate RNAPII from DNA in the same way, these findings indicate commonality between these two WDR82-dependent pathways. As for Restrictor function on endogenous genes, attenuation of HIV transcription is opposed by U1.[130,131] In this original example of the suppression of early transcriptional termination by a 5′ splice site, U1 prevents the use of a PAS within the 5′ long terminal repeat (LTR) to ensure full-length HIV transcription. Finally, transcription from the HIV promoter can be attenuated by Integrator in the absence of TAT.[132,133] Therefore, and like most mammalian promoters, positive functions of U1 and CDK9 promote transcription at least in part by opposing multiple early termination mechanisms.

Intrinsic/DNA-directed transcriptional termination

In bacteria, transcribed poly(U) stretches cause RNA polymerases to terminate due to unstable

rU-dA hybrids within the polymerase active site.[134] Upstream of these stretches are inverted GC-rich repeats. These form hairpins and cause a conformational change in paused polymerases.[135,136] This forces the dissolution of the ternary polymerase-RNA-DNA complex, constituting a so-called intrinsic termination mechanism.[137] We demonstrated that RNAPII terminates over T-rich regions (in the coding DNA strand) downstream of snRNA genes suggesting that this mechanism is conserved throughout evolution.[138] More recently, DNA-directed RNAPII termination was shown to occur in budding yeast and we have shown that, in mammals, this mechanism is frequently employed in promoter-proximal regions where PT is common.[139,140] Recently, Vlaming et al., developed INtegrated Sequences on Expression of RNA and Translation, using high-throughput sequencing (INSERT-seq) to establish the effect of nucleic acid sequences on transcription.[141] They report that the most negative predictor of transcription was the AT content of coding DNA strands, as opposed to specific motifs. Considering these data, Restrictor may enable an intrinsic-type termination activity. Such a pathway may be effective because of poor elongation enforced by Restrictor and PNUTS-PP1 and explain why no cleavage site has so far been observed in the Restrictor pathway. Among the thousands of sequences assayed by Vlaming et al, a 5′ splice site had the greatest positive impact on transcription. This synergises with observations that U1 inhibition sensitises some protein-coding transcription to Restrictor,[33] that chaotic splicing of some extragenic PT is seen in the absence of Restrictor,[35] and supports the original hypothesis that Su(s) competes with the assembly of splicing complexes to suppress RNA levels.[49] Table 2 summarises the RNAPII termination pathways discussed above:

Unanswered questions and future directions

What is the relationship between Integrator and Restrictor?

Why does a cell utilise both Integrator and Restrictor to regulate PT? Integrator functions at the promoter-proximal pause, when RNAPII is associated with NELF.[142–144] The structure of this complex has been solved, and importantly it precludes the simultaneous binding of PAF components.[30] When RNAPII escapes the promoter-proximal pause, PAF replaces NELF in an SPT6-dependent process.[106] Interestingly, NELF is absent in ZC3H4 complexes and proximity interactomes but PAF components are highly enriched.[33] A plausible order of events is the surveillance of initial promoter-proximal RNAPII by Integrator (associated with NELF-bound RNAPII) followed by subsequent monitoring of transcription by Restrictor (associated with PAF-bound RNAPII). If true, then early events in elongation may be more complex than currently envisaged, with early termination frequently occurring even after the displacement of NELF by productive elongation factors. Once elongation is established, it appears to be resistant to both termination complexes.

Interestingly, the only PAF component not proximal to Restrictor is RTF1, which enhances RNAPII velocity.[104,145] Therefore, RTF1 could potentially replace Restrictor at the commencement of productive transcription. Interestingly, the PAF complex (CTR9, LEO1, PAF1) and SPT6 are predicted to be CK2 targets.[146] This could mean that ZC3H4 interacts with these components transiently *via* CK2 in a similar way that has been shown for SPT5, although this has not been explored.[38] A further distinction between Restrictor and Integrator is

Table 2 Comparison of different RNAPII termination mechanisms in eukaryotes.

Pathway	Regions a pathway is active at	Method of termination
CPA (all)	3′ end of protein-coding genes and some PROMPTs	PAS cleavage by CPSF73 followed by degradation of the RNAPII-associated product by XRN2. Termination is enabled by PNUTS-PP1 dephosphorylation of SPT5, which slows transcription beyond the PAS.
Integrator (Metazoan)	Promoter proximal regions	INTS6-PP2A counteracts CDK9 to negatively affect elongation, INTS11 endonuclease cleaves nascent transcripts to promote termination, and the INTS13/14 “sting” domain prevents the association of positive elongation factors to favour termination.
Restrictor (Metazoan)	PROMPTs, enhancers, 5′ end of some mRNA	ZC3H4 interacts with the RNA 5′ end via ARS2, Ser5p via WDR82, and RNA via its ARMS domain. Modulation of SPT5 phosphorylation by CK2 and, potentially, PNUTS-PP1 favour termination. How RNAPII is evicted from the template is unknown.
NNS (Yeast)	CUTs, snoRNA, some mRNA (aberrant)	UGUAG / GNUCUUGU motifs in nascent RNA are recognised by NNS which interacts with Ser5p of RNAPIIs CTD – RNAPII is removed from template DNA by the RNA-DNA helicase Sen1.
DNA- directed (all)	snRNA, promoter-proximal regions	RNAPII terminates over T-rich elements in the DNA coding strand. The mechanism is possibly analogous to intrinsic termination in bacteria, which is favoured by the weak rU: dA hybrid within the RNAPII active centre.

suggested by experiments showing that ARS2 (which binds Restrictor) and NELFE (which associates with Integrator) interact with CBC in a mutually exclusive manner.[147]

What provides the termination activity in the Restrictor pathway?

In the review that coined the name "Restrictor", the authors asked whether ZC3H4 is a nuclease.[37] In flies and plants, orthologues of CPSF30 have endonucleolytic potential which, importantly, derives from their CCCH ZFs – the same type of ZF present in ZC3H4.[148–150] With the 3′ end processing machinery in mind, it could mean that PNUTS, which is proximal to ZC3H4, acts to slow RNAPII before ZC3H4 itself cuts the nascent RNA. This would reveal a 3′ end that Restrictor and ARS2 could hand over to DIS3 for turnover. On the face of it, this is an attractive model; however, the available evidence suggests this is not the case. For instance, while mutating the third CCCH ZF finger in plant CPSF30 rendered it inert, mutating all ZFs in ZC3H4 does not fully ablate ZC3H4 restriction activity.[33,150] As there are no other annotated domains within ZC3H4, nuclease activity is either absent, weak/irrelevant, or provided *via* another region of ZC3H4.

If ZC3H4 is not an endonuclease, another (CPSF73 or INTS11) might be employed as part of the Restrictor pathway. Rapid CPSF73 depletion causes transcriptional termination defects at a subset of ncRNAs.[14,33] At some of these loci, CPSF73 tended to synergise with ZC3H4, implying that they act in different pathways as Russo et al. also suggested for the CPA factor SYMPLEKIN.[33,92] When artificially tethering ZC3H4 to a reporter construct *via* MS2 hairpins its ability to silence reporter RNA is unaffected by Integrator loss.[33] As discussed above, the available structural and interactome data indicate that Integrator and Restrictor associate with distinct RNAPII complexes. Importantly, endonucleolytic cleavage of RNA creates a 5′ end which would be recognised by the major nuclear 5′ → 3′ exonuclease XRN2. However, while rapid depletion of XRN2 resulted in the expected transcriptional read-through beyond protein-coding genes, no termination defects were observed at PROMPTs.[13,33] Because XRN2 requires an RNAPII-associated cleavage product to induce termination, this observation reduces the likelihood of such an event being part of the Restrictor pathway. A caveat of this interpretation is that RNAPII elongation capacity may be low during PT, which might preclude the observation of long read-through transcription that is conventionally expected after depleting factors such as XRN2.

Ultimately, Restrictor-mediated termination is likely to happen in one of three ways; *via* ZC3H4 directly, *via* an additional recruited factor, or ZC3H4 independently. Establishing whether ZC3H4 promotes termination itself likely requires *in vitro* systems to recapitulate Restrictor function and/or a detailed structure that may allow more definitive mutations to be tested *in vivo*. As discussed above, ZC3H4 was not found in association with known termination factors. On the one hand, this appears to rule out their function in its mechanism. However, another possibility is that ZC3H4 enables transcriptional termination by an established mechanism that responds to its restriction of RNAPII elongation capacity. Another indirect consequence of ZC3H4-restricted elongation might be an intrinsic-like mechanism of RNAPII termination.

What underpins the sensitivity of transcription to ZC3H4?

The highest genomic occupancy observed for ZC3H4 is at histones and snRNA loci, yet defects in their transcription are not observed following Restrictor loss.[34] Histone gene transcriptional termination has been studied for decades and Restrictor has never been implicated.[151,152] As a result, the presence of ZC3H4 over these regions may constitute an artifact, or there may be a context in which Restrictor works here that we are unaware of as was illuminated at other protein-coding genes by SET1A depletion and U1 inhibition.[33,63] For example, Integrator has been proposed as a backup to the 3′ end processing machinery at certain genes under stress conditions – Downstream Of Genes (DOGs).[153,154] Although nobody has evaluated ZC3H4 as a stress responder, Su(s) does recognise additional targets following heat shock.[51]

Does ZC3H4 function beyond PT?

Although Restrictor's role in PT regulation has been well described above, there remains a possibility that ZC3H4 has a role beyond this. For example, Restrictor is essential during early embryogenesis, where ZC3H4$^{-/-}$ embryos are non-recoverable post-gastrulation, and blastocysts exhibit extensive DNA damage.[155] The cause of this damage was not elucidated, but a potential explanation could be that dysregulated polymerases might collide and interfere with the DNA replication machinery.[156,157] Interestingly, PNUTS has been reported to play a role in the resolution of transcription-replication conflicts, leading to the possibility that ZC3H4 might also be involved.[158] Genome-wide sequencing studies have not been performed in ZC3H4$^{-/-}$ blastocysts, so it is unclear to what degree transcription is dysregulated.

It was also observed that there were ~50% fewer cells containing the essential transcription factor OCT4 in ZC3H4$^{-/-}$ embryos, with the suggestion that ZC3H4 worked upstream of it.[155] OCT4 is a "master regulator" responsible for maintaining pluripotency, with thousands of binding sites documented within the mouse genome.[159,160] Interestingly, OCT4 is regulated by both distal and

proximal enhancers, with the deletion of the distal enhancer causing a loss of expression.[161] As documented,[34,35] ZC3H4 is active at enhancer regions and it would be interesting to see if the cause of OCT4 loss in ZC3H4$^{-/-}$ embryos was due to dysregulation of enhancer transcription.[34,155] Finally, because ZC3H4 (and Su(s)) suppress sub-optimal splice sites, their depletion could also disrupt pre-mRNA splicing.

The multiple faces of WDR82

WDR82 is reported to form complexes with ZC3H4 and PNUTS mutually exclusively.[35] As such, the proximity of ZC3H4 and PNUTS may be explained by multiple WDR82-containing complexes on the same RNAPII (i.e. one containing ZC3H4 and another bound to PNUTS). Interestingly, PNUTS and ZC3H4 are found in the same fraction following size exclusion chromatography supporting the possibility that they form part of a larger common assembly affecting PT.[63] In contrast, ZC3H4 and SET1 do not co-sediment, consistent with their identical mode of interaction with WDR82 and their opposite effects on transcription (SET1 promotes transcription whereas ZC3H4 restricts it). How and whether proximal ZC3H4 and PNUTS combine to elicit their effects on transcription is an important question. One possibility is that two separate WDR82-containing complexes act independently to ensure the robust restriction of PT. This is consistent with recent data from the Natoli lab which identifies some extragenic transcription that is separately affected by ZC3H4 and PNUTS.[92] Alternatively, WDR82 may form a single transient complex with ZC3H4 and PNUTS to elicit transcriptional termination. Such a complex was recently predicted by AlphaFold modelling and further suggested to occur *via* immunoprecipitation.[166] This complex might be unstable and dissociate once RNAPII is evicted from the chromatin template. The possibility that functionally relevant Restrictor complexes might be transient is supported by our observation that the termination inactive WDR82-interacting fragment of ZC3H4 co-precipitates RNAPII more strongly than the termination active full-length protein.[33]

The recently described interaction between WDR82 and PCF11 that attenuates HIV transcription adds to the repertoire of WDR82-containing complexes. How and why multiple WDR82-complexes form is an important future question. Moreover, it will be interesting to determine whether WDR82 functions merely to tether complexes to Ser5p RNAPII or if it additionally plays a role in mediating critical structural reconfigurations of its interacting partners.

The budding yeast homologue of WDR82 (Swd2) also binds Ser5p, forms a complex with SET1, and is part of the CPA complex.[162–164] As mentioned above, budding yeast does not have an obvious ZC3H4 orthologue, yet Restrictor shares some features with the NNS complex. In analogy to WDR82-PNUTS-PP1, Swd2 associates with the putative PNUTS homologue, Ref2, and the Glc7 phosphatase – a complex known to function in transcriptional termination.[90,124] Interestingly, a functional link between Glc7 and NNS was established through observations that termination defects caused by Swd2 loss could be suppressed by over-expressing the Sen1 helicase.[165] One consequence of this was an increased association of Glc7 with NNS. Given the plethora of WDR82 interactions, depletion of factors such as ZC3H4 or PNUTS may alter transcriptional profiles by changing the abundance of the various WDR82 complexes. For example, the loss of ZC3H4 may make more WDR82 available for interactions with PNUTS, SET1, or PCF11. How these various WDR82-containing complexes are regulated to control transcription is an important question for the future.

Is Restrictor a single complex or multiple assemblies?

WDR82 is the strongest ZC3H4 interactor and was also identified as a partner for Su(s).[52] Interestingly, polyacrylamide gel-based analysis of the Su(s) interactome suggested WDR82 to be its sole interactor or, at least, its most prominent. However, ZC3H4 interacts with multiple factors based on immunoprecipitation from mammalian cells.[33,38,39] An obvious possibility is that a core dimer of ZC3H4 and WDR82 participates in multiple Restrictor subcomplexes. Indeed, ARS2 and CK2 bind to the N-terminal region of ZC3H4 and it would be interesting to test whether they are mutually exclusive. Moreover, the effects of depleting ARS2 from cells are sometimes distinct from those resulting from ZC3H4 loss and the latter causes the longer extension of some Restrictor targets.[33,39]

What is the function of the ZC3H4 paralogue, ZC3H6?

Like ZC3H4, ZC3H6 is proximal to RNAPII and is Metazoan-specific.[34] At present, it has no known function, and our previous analyses suggest it is not redundant with ZC3H4.[34] Some relationship between the two proteins is indicated by the fact that the transcription of *ZC3H6* is increased following ZC3H4 depletion. However, its identification in an RNAPII proximity screen following the perturbation of the CPA machinery makes it likely that it does have a function related to transcription. At present, the function of ZC3H6 in transcription or otherwise is unknown.

Summary

The Restrictor complex is an emerging regulator of PT. A possible model for Restriction is depicted in Figure 3. Here, ZC3H4 targets a

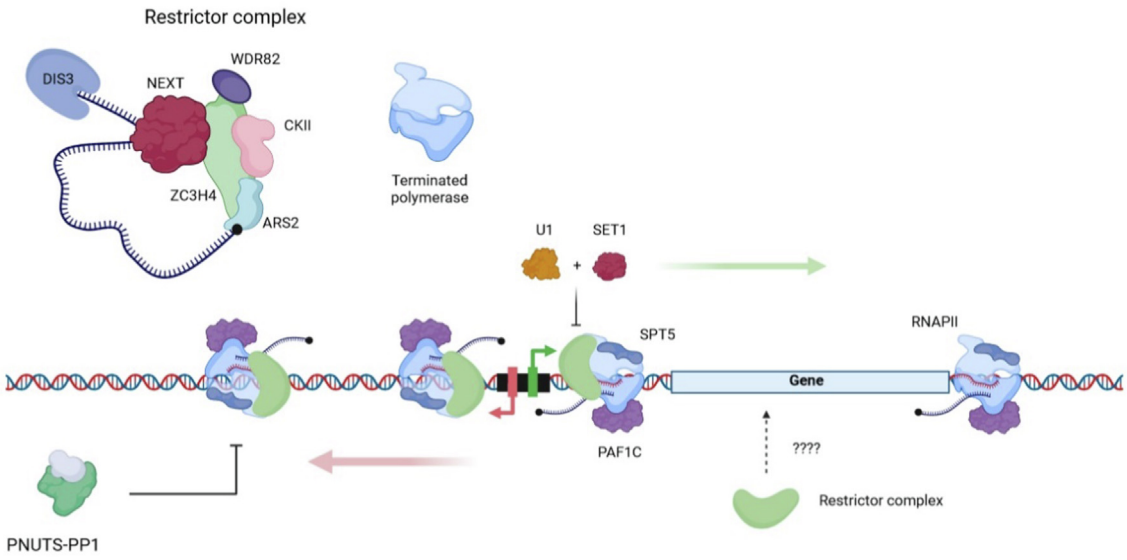

Figure 3. The Restrictor pathway. Restrictor interacts with RNAPII at most promoter-proximal regions. This could occur via the ARS2-mediated interaction with RNA 5′ ends and/or the WDR82-mediated interaction with Ser5p RNAPII. At protein-coding genes, Restrictor is opposed in the sense direction by U1 (yellow) and the SET1A complex (red) which enable RNAPII to transition to productive elongation. In the antisense direction, ZC3H4-CK2 phosphorylates the N-terminus of SPT5, while PNUTS-PP1 dephosphorylates currently unidentified factors that may include SPT5. These activities maintain a pervasively transcribing state and favour early termination by an unknown mechanism. At the same time, the liberated RNA-ZC3H4-ARS2 complex interacts with NEXT, handing over terminated transcripts to DIS3 for degradation.

subpopulation of PAF and PNUTS-bound polymerases engaged in PT. The basis for Restrictor targeting is yet to be fully established but accumulating evidence suggests some selectivity toward poorly spliced RNAs or those that fail to assemble splicing complexes. Transcript targeting might be achieved by a combination of the direct recognition of RNA/DNA sequences by ZC3H4, by Restrictor recruitment to the 5′ end of RNA via ARS2, or by targeting ZC3H4 to Ser5p *via* its interaction with WDR82. It remains to be seen whether these different modes of Restrictor recruitment can be employed separately or if there are distinct Restrictor subcomplexes deployed under certain circumstances. When bound to CK2, Restrictor can control RNAPII elongation via phosphorylation of SPT5. Similarly, Restrictor function may be enabled by PP1-PNUTS, which distinctly dephosphorylates SPT5 to maintain poor RNAPII elongation. Perhaps the most important question is how transcriptional termination ultimately occurs for Restrictor-sensitive RNA. This could be by Restrictor itself destabilising/evicting elongation complexes, by an unknown factor, or *via* an established mechanism that acts in a Restrictor-dependent manner. Following termination, the liberated RNA is rapidly targeted to the exosome via ARS2-bound ZC3H4 and NEXT.[39]

CRediT authorship contribution statement

Chris Estell: Writing – review & editing, Writing – original draft. **Steven West:** Writing – review & editing, Writing – original draft, Supervision, Funding acquisition.

DECLARATION OF COMPETING INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We thank The Wellcome Trust for funding this work through two Investigator Awards to SW (WT107791/Z/15/Z and WT223106/Z/21/Z). We also thank Jane Roscow (Chris Estell's mum) for proofreading the manuscript. Figures were prepared using Biorender.

Received 14 May 2024;
Accepted 9 July 2024;
Available online 22 July 2024

Keywords:
ZC3H4;
restrictor;
RNA polymerase II;
transcription termination;
non-coding RNA;
PNUTS

References

1. Consortium, E.P., Birney, E., Stamatoyannopoulos, J.A., Dutta, A., Guigo, R., Gingeras, T.R., et al., (2007). Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447**, 799–816.
2. Djebali, S., Davis, C.A., Merkel, A., Dobin, A., Lassmann, T., Mortazavi, A., et al., (2012). Landscape of transcription in human cells. *Nature* **489**, 101–108.
3. Piovesan, A., Pelleri, M.C., Antonaros, F., Strippoli, P., Caracausi, M., Vitale, L., (2019). On the length, weight and GC content of the human genome. *BMC. Res. Notes* **12**, 106.
4. Jacquier, A., (2009). The complex eukaryotic transcriptome: unexpected pervasive transcription and novel small RNAs. *Nature Rev. Genet.* **10**, 833–844.
5. Jensen, T.H., Jacquier, A., Libri, D., (2013). Dealing with pervasive transcription. *Mol. Cell* **52**, 473–484.
6. Seila, A.C., Calabrese, J.M., Levine, S.S., Yeo, G.W., Rahl, P.B., Flynn, R.A., et al., (2008). Divergent transcription from active promoters. *Science* **322**, 1849–1851.
7. Core, L.J., Waterfall, J.J., Lis, J.T., (2008). Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science* **322**, 1845–1848.
8. Carninci, P., Sandelin, A., Lenhard, B., Katayama, S., Shimokawa, K., Ponjavic, J., et al., (2006). Genome-wide analysis of mammalian promoter architecture and evolution. *Nature Genet.* **38**, 626–635.
9. Gockert, M., Schmid, M., Jakobsen, L., Jens, M., Andersen, J.S., Jensen, T.H., (2022). Rapid factor depletion highlights intricacies of nucleoplasmic RNA degradation. *Nucleic Acids Res.* **50**, 1583–1600.
10. Preker, P., Nielsen, J., Kammler, S., Lykke-Andersen, S., Christensen, M.S., Mapendano, C.K., et al., (2008). RNA exosome depletion reveals transcription upstream of active human promoters. *Science* **322**, 1851–1854.
11. Neil, H., Malabat, C., d'Aubenton-Carafa, Y., Xu, Z., Steinmetz, L.M., Jacquier, A., (2009). Widespread bidirectional promoters are the major source of cryptic transcripts in yeast. *Nature* **457**, 1038–1042.
12. Wyers, F., Rougemaille, M., Badis, G., Rousselle, J.C., Dufour, M.E., Boulay, J., et al., (2005). Cryptic pol II transcripts are degraded by a nuclear quality control pathway involving a new poly(A) polymerase. *Cell* **121**, 725–737.
13. Davidson, L., Francis, L., Cordiner, R.A., Eaton, J.D., Estell, C., Macias, S., et al., (2019). Rapid depletion of DIS3, EXOSC10, or XRN2 reveals the immediate impact of exoribonucleolysis on nuclear RNA metabolism and transcriptional control. *Cell Rep.* **26**, 2779–2791.e5.
14. Lykke-Andersen, S., Zumer, K., Molska, E.S., Rouviere, J. O., Wu, G., Demel, C., et al., (2021). Integrator is a genome-wide attenuator of non-productive transcription. *Mol. Cell* **81**, 514–529.e6.
15. Rodriguez-Molina, J.B., West, S., Passmore, L.A., (2023). Knowing when to stop: Transcription termination on protein-coding genes by eukaryotic RNAPII. *Mol. Cell* **83**, 404–415.
16. Proudfoot, N.J., (2011). Ending the message: poly(A) signals then and now. *Genes Dev.* **25**, 1770–1782.
17. Eaton, J.D., West, S., (2020). Termination of transcription by RNA polymerase II: BOOM! *Trends Genet.*.
18. Eaton, J.D., Francis, L., Davidson, L., West, S., (2020). A unified allosteric/torpedo mechanism for transcriptional termination on human protein-coding genes. *Genes Dev.* **34**, 132–145.
19. Eaton, J.D., Davidson, L., Bauer, D.L.V., Natsume, T., Kanemaki, M.T., West, S., (2018). Xrn2 accelerates termination by RNA polymerase II, which is underpinned by CPSF73 activity. *Genes Dev.* **32**, 127–139.
20. Cortazar, M.A., Sheridan, R.M., Erickson, B., Fong, N., Glover-Cutter, K., Brannan, K., et al., (2019). Control of RNA Pol II speed by PNUTS-PP1 and Spt5 Dephosphorylation facilitates termination by a "Sitting Duck Torpedo" mechanism. *Mol. Cell* **76**, 896–908.e4.
21. Joshua, D., Eaton, J.B., Davidson, L., Estell, C., West, S., (2024). Human promoter directionality is determined by transcriptional initiation and the opposing activities of INTS11 and CDK9. *eLife*.
22. Wagner, E.J., Tong, L., Adelman, K., (2023). Integrator is a global promoter-proximal termination complex. *Mol. Cell* **83**, 416–427.
23. Huang, K.L., Jee, D., Stein, C.B., Elrod, N.D., Henriques, T., Mascibroda, L.G., et al., (2020). Integrator recruits protein phosphatase 2A to prevent pause release and facilitate transcription termination. *Mol. Cell* **80**, 345–358.e9.
24. Albrecht, T.R., Wagner, E.J., (2012). snRNA 3′ end formation requires heterodimeric association of integrator subunits. *Mol. Cell. Biol.* **32**, 1112–1123.
25. Baillat, D., Hakimi, M.A., Naar, A.M., Shilatifard, A., Cooch, N., Shiekhattar, R., (2005). Integrator, a multiprotein mediator of small nuclear RNA processing, associates with the C-terminal repeat of RNA polymerase II. *Cell* **123**, 265–276.
26. Stein, C.B., Field, A.R., Mimoso, C.A., Zhao, C., Huang, K.L., Wagner, E.J., et al., (2022). Integrator endonuclease drives promoter-proximal termination at all RNA polymerase II-transcribed loci. *Mol. Cell* **82**, 4232–4245.e11.
27. Elrod, N.D., Henriques, T., Huang, K.L., Tatomer, D.C., Wilusz, J.E., Wagner, E.J., et al., (2019). The integrator complex attenuates promoter-proximal transcription at protein-coding genes. *Mol. Cell* **76**, 738–752.e7.
28. Lykke-Andersen, S., Zumer, K., Molska, E.S., Rouviere, J. O., Wu, G., Demel, C., et al., (2020). Integrator is a genome-wide attenuator of non-productive transcription. *Mol. Cell*.
29. Tatomer, D.C., Elrod, N.D., Liang, D., Xiao, M.S., Jiang, J. Z., Jonathan, M., et al., (2019). The Integrator complex cleaves nascent mRNAs to attenuate transcription. *Genes Dev.* **33**, 1525–1538.

30. Fianu, I., Chen, Y., Dienemann, C., Dybkov, O., Linden, A., Urlaub, H., et al., (2021). Structural basis of Integrator-mediated transcription regulation. *Science* **374**, 883–887.
31. Yamamoto, J., Hagiwara, Y., Chiba, K., Isobe, T., Narita, T., Handa, H., et al., (2014). DSIF and NELF interact with Integrator to specify the correct post-transcriptional fate of snRNA genes. *Nature Commun.* **5**, 4263.
32. Fianu, I., Ochmann, M., Walshe, J.L., Dybkov, O., Cruz, J. N., Urlaub, H., et al., (2024). Structural basis of Integrator-dependent RNA polymerase II termination. *Nature* **629**, 219–227.
33. Estell, C., Davidson, L., Eaton, J.D., Kimura, H., Gold, V. A.M., West, S., (2023). A restrictor complex of ZC3H4, WDR82, and ARS2 integrates with PNUTS to control unproductive transcription. *Mol. Cell* **83**, 2222–2239.e5.
34. Estell, C., Davidson, L., Steketee, P.C., Monier, A., West, S., (2021). ZC3H4 restricts non-coding transcription in human cells. *Elife*, 10.
35. Austenaa, L.M.I., Piccolo, V., Russo, M., Prosperini, E., Polletti, S., Polizzese, D., et al., (2021). A first exon termination checkpoint preferentially suppresses extragenic transcription. *Nature Struct. Mol. Biol.* **28**, 337–346.
36. Austenaa, L.M., Barozzi, I., Simonatto, M., Masella, S., Della Chiara, G., Ghisletti, S., et al., (2015). Transcription of mammalian cis-regulatory elements is restrained by actively enforced early termination. *Mol. Cell* **60**, 460–474.
37. Nojima, T., Proudfoot, N.J., (2022). Mechanisms of lncRNA biogenesis as revealed by nascent transcriptomics. *Nature Rev. Mol. Cell Biol.* **23**, 389–406.
38. Park, K., Zhong, J., Jang, J.S., Kim, J., Kim, H.J., Lee, J. H., et al., (2022). ZWC complex-mediated SPT5 phosphorylation suppresses divergent antisense RNA transcription at active gene promoters. *Nucleic Acids Res.* **50**, 3835–3851.
39. Rouviere, J.O., Salerno-Kochan, A., Lykke-Andersen, S., Garland, W., Dou, Y., Rathore, O., et al., (2023). ARS2 instructs early transcription termination-coupled RNA decay by recruiting ZC3H4 to nascent transcripts. *Mol. Cell* **83**, 2240–2257.e6.
40. Searles, L.L., Voelker, R.A., (1986). Molecular characterization of the Drosophila vermilion locus and its suppressible alleles. *PNAS* **83**, 404–408.
41. Baglioni, C., (1960). Genetic control of tryptophan pyrrolase in Drosophila melanogaster and Drosophila virilis. *Heredity* **15**
42. Walker, A.R., Howells, A.J., Tearle, R.G., (1986). Cloning and characterization of the vermilion gene of Drosophila melanogaster. *Mol. Gen. Genet.* **202**, 102–107.
43. Searles, L.L., Ruth, R.S., Pret, A.M., Fridell, R.A., Ali, A.J., (1990). Structure and transcription of the Drosophila melanogaster vermilion gene and several mutant alleles. *Mol. Cell Biol.* **10**, 1423–1431.
44. Fridell, R.A., Pret, A.M., Searles, L.L., (1990). A retrotransposon 412 insertion within an exon of the Drosophila melanogaster vermilion gene is spliced from the precursor RNA. *Genes Dev.* **4**, 559–566.
45. Voelker, R.A., Gibson, W., Graves, J.P., Sterling, J.F., Eisenberg, M.T., (1991). The Drosophila suppressor of sable gene encodes a polypeptide with regions similar to those of RNA-binding proteins. *Mol. Cell Biol.* **11**, 894–905.
46. Murray, M.V., Turnage, M.A., Williamson, K.J., Steinhauer, W.R., Searles, L.L., (1997). The Drosophila suppressor of sable protein binds to RNA and associates with a subset of polytene chromosome bands. *Mol. Cell. Biol.* **17**, 2291–2300.
47. Mello, C.C., Schubert, C., Draper, B., Zhang, W., Lobel, R., Priess, J.R., (1996). The PIE-1 protein and germline specification in C. elegans embryos. *Nature* **382**, 710–712.
48. Turnage, M.A., Brewer-Jensen, P., Bai, W.L., Searles, L. L., (2000). Arginine-rich regions mediate the RNA binding and regulatory activities of the protein encoded by the Drosophila melanogaster suppressor of sable gene. *Mol. Cell. Biol.* **20**, 8198–8208.
49. Fridell, R.A., Searles, L.L., (1994). Evidence for a role of the Drosophila melanogaster suppressor of sable gene in the pre-mRNA splicing pathway. *Mol. Cell. Biol.* **14**, 859–867.
50. Kuan, Y.S., Brewer-Jensen, P., Searles, L.L., (2004). Suppressor of sable, a putative RNA-processing protein, functions at the level of transcription. *Mol. Cell. Biol.* **24**, 3734–3746.
51. Kuan, Y.S., Brewer-Jensen, P., Bai, W.L., Hunter, C., Wilson, C.B., Bass, S., et al., (2009). Drosophila suppressor of sable protein [Su(s)] promotes degradation of aberrant and transposon-derived RNAs. *Mol. Cell. Biol.* **29**, 5590–5603.
52. Brewer-Jensen, P., Wilson, C.B., Abernethy, J., Mollison, L., Card, S., Searles, L.L., (2016). Suppressor of sable [Su(s)] and Wdr82 down-regulate RNA from heat-shock-inducible repetitive elements by a mechanism that involves transcription termination. *RNA* **22**, 139–154.
53. Lee, J.H., You, J., Dobrota, E., Skalnik, D.G., (2010). Identification and characterization of a novel human PP1 phosphatase complex. *J. Biol. Chem.* **285**, 24466–24476.
54. Lee, J.H., Skalnik, D.G., (2005). CpG-binding protein (CXXC finger protein 1) is a component of the mammalian Set1 histone H3-Lys4 methyltransferase complex, the analogue of the yeast Set1/COMPASS complex. *J. Biol. Chem.* **280**, 41725–41731.
55. Loven, J., Hoke, H.A., Lin, C.Y., Lau, A., Orlando, D.A., Vakoc, C.R., et al., (2013). Selective inhibition of tumor oncogenes by disruption of super-enhancers. *Cell* **153**, 320–334.
56. Su, L., Radek, J.T., Hallenga, K., Hermanto, P., Chan, G., Labeots, L.A., et al., (1997). RNA recognition by a bent alpha-helix regulates transcriptional antitermination in phage lambda. *Biochemistry* **36**, 12722–12732.
57. Tan, R., Frankel, A.D., (1995). Structural variety of arginine-rich RNA-binding peptides. *PNAS* **92**, 5282–5286.
58. Ohno, T., Ouchida, M., Lee, L., Gatalica, Z., Rao, V.N., Reddy, E.S., (1994). The EWS gene, involved in Ewing family of tumors, malignant melanoma of soft parts and desmoplastic small round cell tumors, codes for an RNA binding protein with novel regulatory domains. *Oncogene* **9**, 3087–3097.
59. Kiledjian, M., Dreyfuss, G., (1992). Primary structure and binding activity of the hnRNP U protein: binding RNA through RGG box. *EMBO J.* **11**, 2655–2664.
60. Loughlin, F.E., Lukavsky, P.J., Kazeeva, T., Reber, S., Hock, E.M., Colombo, M., et al., (2019). The solution

structure of FUS bound to RNA reveals a bipartite mode of RNA recognition with both sequence and shape specificity. *Mol. Cell* **73**, 490–504.e6.
61. Lee, J.H., Skalnik, D.G., (2008). Wdr82 is a C-terminal domain-binding protein that recruits the Setd1A Histone H3-Lys4 methyltransferase complex to transcription start sites of transcribed human genes. *Mol. Cell. Biol.* **28**, 609–618.
62. Buratowski, S., (2003). The CTD code. *Nature Struct. Biol.* **10**, 679–680.
63. Hughes, A.L., Szczurek, A.T., Kelley, J.R., Lastuvkova, A., Turberfield, A.H., Dimitrova, E., et al., (2023). A CpG island-encoded mechanism protects genes from premature transcription termination. *Nature Commun.* **14**, 726.
64. Litchfield, D.W., (2003). Protein kinase CK2: structure, regulation and role in cellular decisions of life and death. *Biochem. J* **369**, 1–15.
65. Trembley, J.H., Kren, B.T., Afzal, M., Scaria, G.A., Klein, M.A., Ahmed, K., (2023). Protein kinase CK2 - diverse roles in cancer cell biology and therapeutic promise. *Mol. Cell. Biochem.* **478**, 899–926.
66. Burnett, G., Kennedy, E.P., (1954). The enzymatic phosphorylation of proteins. *J. Biol. Chem.* **211**, 969–980.
67. Yamada, T., Yamaguchi, Y., Inukai, N., Okamoto, S., Mura, T., Handa, H., (2006). P-TEFb-mediated phosphorylation of hSpt5 C-terminal repeats is critical for processive transcription elongation. *Mol. Cell* **21**, 227–237.
68. Marshall, N.F., Peng, J., Xie, Z., Price, D.H., (1996). Control of RNA polymerase II elongation potential by a novel carboxyl-terminal domain kinase. *J. Biol. Chem.* **271**, 27176–27183.
69. Fujinaga, K., Irwin, D., Huang, Y., Taube, R., Kurosu, T., Peterlin, B.M., (2004). Dynamics of human immunodeficiency virus transcription: P-TEFb phosphorylates RD and dissociates negative effectors from the transactivation response element. *Mol. Cell Biol.* **24**, 787–795.
70. Hallais, M., Pontvianne, F., Andersen, P.R., Clerici, M., Lener, D., Benbahouche Nel, H., et al., (2013). CBC-ARS2 stimulates 3′-end maturation of multiple RNA families and favors cap-proximal processing. *Nature Struct. Mol. Biol.* **20**, 1358–1366.
71. Lykke-Andersen, S., Rouviere, J.O., Jensen, T.H., (2021). ARS2/SRRT: at the nexus of RNA polymerase II transcription, transcript maturation and quality control. *Biochem. Soc. Trans.* **49**, 1325–1336.
72. Andersen, P.R., Domanski, M., Kristiansen, M.S., Storvall, H., Ntini, E., Verheggen, C., et al., (2013). The human cap-binding complex is functionally connected to the nuclear RNA exosome. *Nature Struct. Mol. Biol.* **20**, 1367–1376.
73. Gruber, J.J., Zatechka, D.S., Sabin, L.R., Yong, J., Lum, J.J., Kong, M., et al., (2009). Ars2 links the nuclear cap-binding complex to RNA interference and cell proliferation. *Cell* **138**, 328–339.
74. Winczura, K., Schmid, M., Iasillo, C., Molloy, K.R., Harder, L.M., Andersen, J.S., et al., (2018). Characterizing ZC3H18, a multi-domain protein at the interface of RNA production and destruction decisions. *Cell Rep.* **22**, 44–58.
75. Lubas, M., Andersen, P.R., Schein, A., Dziembowski, A., Kudla, G., Jensen, T.H., (2015). The human nuclear exosome targeting complex is loaded onto newly synthesized RNA to direct early ribonucleolysis. *Cell Rep.* **10**, 178–192.
76. Foucher, A.E., Touat-Todeschini, L., Juarez-Martinez, A. B., Rakitch, A., Laroussi, H., Karczewski, C., et al., (2022). Structural analysis of Red1 as a conserved scaffold of the RNA-targeting MTREC/PAXT complex. *Nature Commun.* **13**, 4969.
77. Dobrev, N., Ahmed, Y.L., Sivadas, A., Soni, K., Fischer, T., Sinning, I., (2021). The zinc-finger protein Red1 orchestrates MTREC submodules and binds the Mtl1 helicase arch domain. *Nature Commun.* **12**, 3456.
78. Mimoso, C.A., Adelman, K., (2023). U1 snRNP increases RNA Pol II elongation rate to enable synthesis of long genes. *Mol. Cell*.
79. Kaida, D., Berg, M.G., Younis, I., Kasim, M., Singh, L.N., Wan, L., et al., (2010). U1 snRNP protects pre-mRNAs from premature cleavage and polyadenylation. *Nature* **468**, 664–668.
80. Almada, A.E., Wu, X., Kriz, A.J., Burge, C.B., Sharp, P.A., (2013). Promoter directionality is controlled by U1 snRNP and polyadenylation signals. *Nature* **499**, 360–363.
81. Ntini, E., Jarvelin, A.I., Bornholdt, J., Chen, Y., Boyd, M., Jorgensen, M., et al., (2013). Polyadenylation site-induced decay of upstream transcripts enforces promoter directionality. *Nature Struct. Mol. Biol.* **20**, 923–928.
82. Branon, T.C., Bosch, J.A., Sanchez, A.D., Udeshi, N.D., Svinkina, T., Carr, S.A., et al., (2018). Efficient proximity labeling in living cells and organisms with TurboID. *Nature Biotechnol.* **36**, 880–887.
83. Ferreira, M., Beullens, M., Bollen, M., Van Eynde, A., (2019). Functions and therapeutic potential of protein phosphatase 1: Insights from mouse genetics. *Biochim. Biophys. Acta* **1866**, 16–30.
84. West, S., Gromak, N., Proudfoot, N.J., (2004). Human 5' –> 3′ exonuclease Xrn2 promotes transcription termination at co-transcriptional cleavage sites. *Nature* **432**, 522–525.
85. Zeng, Y., Zhang, H.W., Wu, X.X., Zhang, Y., (2024). Structural basis of exoribonuclease-mediated mRNA transcription termination. *Nature*.
86. Kieft, R., Zhang, Y., Marand, A.P., Moran, J.D., Bridger, R., Wells, L., et al., (2020). Identification of a novel base J binding protein complex involved in RNA polymerase II transcription termination in trypanosomes. *PLoS Genet.* **16**, e1008390
87. Kieft, R., Zhang, Y., Yan, H., Schmitz, R.J., Sabatini, R., (2023). Knockout of protein phosphatase 1 in Leishmania major reveals its role during RNA polymerase II transcription termination. *Nucleic Acids Res.* **51**, 6208–6226.
88. Jensen, B.C., Phan, I.Q., McDonald, J.R., Sur, A., Gillespie, M.A., Ranish, J.A., et al., (2021). Chromatin-associated protein complexes link DNA base J and transcription termination in Leishmania. *mSphere* **6**
89. van Luenen, H.G., Farris, C., Jan, S., Genest, P.A., Tripathi, P., Velds, A., et al., (2012). Glucosylated hydroxymethyluracil, DNA base J, prevents transcriptional readthrough in Leishmania. *Cell* **150**, 909–921.
90. Carminati, M., Rodriguez-Molina, J.B., Manav, M.C., Bellini, D., Passmore, L.A., (2023). A direct interaction between CPF and RNA Pol II links RNA 3′ end processing to transcription. *Mol. Cell* **83**, 4461–4478.e13.

91. Mateo-Bonmati, E., Montez, M., Maple, R., Fiedler, M., Fang, X., Saalbach, G., et al., (2024). A CPF-like phosphatase module links transcription termination to chromatin silencing. *Mol. Cell* **84**, 2272–2286.e7.
92. Russo, M., Piccolo, V., Polizzese, D., Prosperini, E., Borriero, C., Polletti, S., et al., (2023). Restrictor synergizes with Symplekin and PNUTS to terminate extragenic transcription. *Genes Dev.* **37**, 1017–1040.
93. Nabet, B., Roberts, J.M., Buckley, D.L., Paulk, J., Dastjerdi, S., Yang, A., et al., (2018). The dTAG system for immediate and target-specific protein degradation. *Nature Chem. Biol.* **14**, 431–441.
94. Jaeger, M.G., Schwalb, B., Mackowiak, S.D., Velychko, T., Hanzl, A., Imrichova, H., et al., (2020). Selective Mediator dependence of cell-type-specifying transcription. *Nature Genet.* **52**, 719–727.
95. Takagaki, Y., Manley, J.L., (2000). Complex protein interactions within the human polyadenylation machinery identify a novel component. *Mol. Cell. Biol.* **20**, 1515–1525.
96. Squazzo, S.L., Costa, P.J., Lindstrom, D.L., Kumer, K.E., Simic, R., Jennings, J.L., et al., (2002). The Paf1 complex physically and functionally associates with transcription elongation factors in vivo. *EMBO J.* **21**, 1764–1774.
97. Mueller, C.L., Jaehning, J.A., (2002). Ctr9, Rtf1, and Leo1 are components of the Paf1/RNA polymerase II complex. *Mol. Cell Biol.* **22**, 1971–1980.
98. Zhu, B., Mandal, S.S., Pham, A.D., Zheng, Y., Erdjument-Bromage, H., Batra, S.K., et al., (2005). The human PAF complex coordinates transcription with events downstream of RNA synthesis. *Genes Dev.* **19**, 1668–1673.
99. Wade, P.A., Werel, W., Fentzke, R.C., Thompson, N.E., Leykam, J.F., Burgess, R.R., et al., (1996). A novel collection of accessory factors associated with yeast RNA polymerase II. *Protein Expr. Purif.* **8**, 85–90.
100. Shi, X., Finkelstein, A., Wolf, A.J., Wade, P.A., Burton, Z. F., Jaehning, J.A., (1996). Paf1p, an RNA polymerase II-associated factor in Saccharomyces cerevisiae, may have both positive and negative roles in transcription. *Mol. Cell Biol.* **16**, 669–676.
101. Yu, M., Yang, W., Ni, T., Tang, Z., Nakadai, T., Zhu, J., et al., (2015). RNA polymerase II-associated factor 1 regulates the release and phosphorylation of paused RNA polymerase II. *Science* **350**, 1383–1386.
102. Chen, F.X., Woodfin, A.R., Gardini, A., Rickels, R.A., Marshall, S.A., Smith, E.R., et al., (2015). PAF1, a molecular regulator of promoter-proximal pausing by RNA polymerase II. *Cell* **162**, 1003–1015.
103. Liu, X., Guo, Z., Han, J., Peng, B., Zhang, B., Li, H., et al., (2022). The PAF1 complex promotes 3′ processing of pervasive transcripts. *Cell Rep.* **38**, 110519
104. Zumer, K., Maier, K.C., Farnung, L., Jaeger, M.G., Rus, P., Winter, G., et al., (2021). Two distinct mechanisms of RNA polymerase II elongation stimulation in vivo. *Mol. Cell* **81**, 3096–3109.e8.
105. Wang, Z., Song, A., Xu, H., Hu, S., Tao, B., Peng, L., et al., (2022). Coordinated regulation of RNA polymerase II pausing and elongation progression by PAF1. *Sci. Adv.* **8**, eabm5504
106. Aoi, Y., Shah, A.P., Ganesan, S., Soliman, S.H.A., Cho, B.K., Goo, Y.A., et al., (2022). SPT6 functions in transcriptional pause/release via PAF1C recruitment. *Mol. Cell* **82**, 3412–3423.e5.
107. Arigo, J.T., Eyler, D.E., Carroll, K.L., Corden, J.L., (2006). Termination of cryptic unstable transcripts is directed by yeast RNA-binding proteins Nrd1 and Nab3. *Mol. Cell* **23**, 841–851.
108. Thiebaut, M., Kisseleva-Romanova, E., Rougemaille, M., Boulay, J., Libri, D., (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.
109. Steinmetz, E.J., Conrad, N.K., Brow, D.A., Corden, J.L., (2001). RNA-binding protein Nrd1 directs poly(A)-independent 3′-end formation of RNA polymerase II transcripts. *Nature* **413**, 327–331.
110. Kim, M., Vasiljeva, L., Rando, O.J., Zhelkovsky, A., Moore, C., Buratowski, S., (2006). Distinct pathways for snoRNA and mRNA termination. *Mol. Cell* **24**, 723–734.
111. Vasiljeva, L., Kim, M., Mutschler, H., Buratowski, S., Meinhart, A., (2008). The Nrd1-Nab3-Sen1 termination complex interacts with the Ser5-phosphorylated RNA polymerase II C-terminal domain. *Nature Struct. Mol. Biol.* **15**, 795–804.
112. Creamer, T.J., Darby, M.M., Jamonnak, N., Schaughency, P., Hao, H., Wheelan, S.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
113. Conrad, N.K., Wilson, S.M., Steinmetz, E.J., Patturajan, M., Brow, D.A., Swanson, M.S., et al., (2000). A yeast heterogeneous nuclear ribonucleoprotein complex associated with RNA polymerase II. *Genetics* **154**, 557–571.
114. Porrua, O., Libri, D., (2013). A bacterial-like mechanism for transcription termination by the Sen1p helicase in budding yeast. *Nature Struct. Mol. Biol.* **20**, 884–891.
115. Vasiljeva, L., Buratowski, S., (2006). Nrd1 interacts with the nuclear exosome for 3′ processing of RNA polymerase II transcripts. *Mol. Cell* **21**, 239–248.
116. LaCava, J., Houseley, J., Saveanu, C., Petfalski, E., Thompson, E., Jacquier, A., et al., (2005). RNA degradation by the exosome is promoted by a nuclear polyadenylation complex. *Cell* **121**, 713–724.
117. Collin, P., Jeronimo, C., Poitras, C., Robert, F., (2019). RNA polymerase II CTD tyrosine 1 is required for efficient termination by the Nrd1-Nab3-Sen1 pathway. *Mol. Cell* **73**, 655–669.e7.
118. Han, Z., Libri, D., Porrua, O., (2017). Biochemical characterization of the helicase Sen1 provides new insights into the mechanisms of non-coding transcription termination. *Nucleic Acids Res.* **45**, 1355–1370.
119. Arigo, J.T., Carroll, K.L., Ames, J.M., Corden, J.L., (2006). Regulation of yeast NRD1 expression by premature transcription termination. *Mol. Cell* **21**, 641–651.
120. Li, Y., Huang, J., Zhu, J., Bao, L., Wang, H., Jiang, Y., et al., (2022). Targeted protein degradation reveals RNA Pol II heterogeneity and functional diversity. *Mol. Cell* **82**, 3943–3959.e11.
121. Kubicek, K., Cerna, H., Holub, P., Pasulka, J., Hrossova, D., Loehr, F., et al., (2012). Serine phosphorylation and proline isomerization in RNAP II CTD control recruitment of Nrd1. *Genes Dev.* **26**, 1891–1896.
122. Tudek, A., Porrua, O., Kabzinski, T., Lidschreiber, M., Kubicek, K., Fortova, A., et al., (2014). Molecular basis for coordinating transcription termination with noncoding RNA degradation. *Mol. Cell* **55**, 467–481.

123. Descostes, N., Heidemann, M., Spinelli, L., Schuller, R., Maqbool, M.A., Fenouil, R., et al., (2014). Tyrosine phosphorylation of RNA polymerase II CTD is associated with antisense promoter transcription and active enhancers in mammalian cells. *Elife* **3**, e02105
124. Nedea, E., He, X., Kim, M., Pootoolal, J., Zhong, G., Canadien, V., et al., (2003). Organization and function of APT, a subcomplex of the yeast cleavage and polyadenylation factor involved in the formation of mRNA and small nucleolar RNA 3′-ends. *J. Biol. Chem.* **278**, 33000–33010.
125. Skourti-Stathaki, K., Proudfoot, N.J., Gromak, N., (2011). Human senataxin resolves RNA/DNA hybrids formed at transcriptional pause sites to promote Xrn2-dependent termination. *Mol. Cell* **42**, 794–805.
126. Wei, P., Garber, M.E., Fang, S.M., Fischer, W.H., Jones, K.A., (1998). A novel CDK9-associated C-type cyclin interacts directly with HIV-1 Tat and mediates its high-affinity, loop-specific binding to TAR RNA. *Cell* **92**, 451–462.
127. Zhang, Z., Klatt, A., Henderson, A.J., Gilmour, D.S., (2007). Transcription termination factor Pcf11 limits the processivity of Pol II on an HIV provirus to repress gene expression. *Genes Dev.* **21**, 1609–1614.
128. Zhang, Z., Fu, J., Gilmour, D.S., (2005). CTD-dependent dismantling of the RNA polymerase II elongation complex by the pre-mRNA 3′-end processing factor, Pcf11. *Genes Dev.* **19**, 1572–1580.
129. Ait Said, M., Bejjani, F., Abdouni, A., Segeral, E., Emiliani, S., (2023). Premature transcription termination complex proteins PCF11 and WDR82 silence HIV-1 expression in latently infected cells. *Proc. Natl. Acad. Sci. U.S.A.* **120**, e2313356120
130. Ashe, M.P., Pearson, L.H., Proudfoot, N.J., (1997). The HIV-1 5' LTR poly(A) site is inactivated by U1 snRNP interaction with the downstream major splice donor site. *EMBO J.* **16**, 5752–5763.
131. Ashe, M.P., Griffin, P., James, W., Proudfoot, N.J., (1995). Poly(A) site selection in the HIV-1 provirus: inhibition of promoter-proximal polyadenylation by the downstream major splice donor site. *Genes Dev.* **9**, 3008–3025.
132. Eaton, J.D., Board, J., Davidson, L., Estell, C., S, W., (2023). Human promoter directionality is determined by transcriptional initiation and the opposing activities of INTS11 and CDK9. *bioRxiv*.
133. Stadelmayer, B., Micas, G., Gamot, A., Martin, P., Malirat, N., Koval, S., et al., (2014). Integrator complex regulates NELF-mediated RNA polymerase II pause/release and processivity at coding genes. *Nature Commun.* **5**, 5531.
134. Gusarov, I., Nudler, E., (1999). The mechanism of intrinsic transcription termination. *Mol. Cell* **3**, 495–504.
135. Peters, J.M., Vangeloff, A.D., Landick, R., (2011). Bacterial transcription terminators: the RNA 3′-end chronicles. *J. Mol. Biol.* **412**, 793–813.
136. Ray-Soni, A., Bellecourt, M.J., Landick, R., (2016). Mechanisms of bacterial transcription termination: all good things must end. *Annu. Rev. Biochem* **85**, 319–347.
137. You, L., Omollo, E.O., Yu, C., Mooney, R.A., Shi, J., Shen, L., et al., (2023). Structural basis for intrinsic transcription termination. *Nature* **613**, 783–789.
138. Davidson, L., Francis, L., Eaton, J.D., West, S., (2020). Integrator-dependent and allosteric/intrinsic mechanisms ensure efficient termination of snRNA transcription. *Cell Rep.* **33**, 108319
139. Han, Z., Moore, G.A., Mitter, R., Lopez Martinez, D., Wan, L., Dirac Svejstrup, A.B., et al., (2023). DNA-directed termination of RNA polymerase II transcription. *Mol. Cell* **83**, 3253–3267.e7.
140. Davidson, L.R., Sousa-Luís, J.O., Nojima, R., Proudfoot, T., Heick Jensen, N., West, T.S., (2024). DNA-directed termination of mammalian RNA polymerase II. *bioRxiv*.
141. Vlaming, H., Mimoso, C.A., Field, A.R., Martin, B.J.E., Adelman, K., (2022). Screening thousands of transcribed coding and non-coding regions reveals sequence determinants of RNA polymerase II elongation potential. *Nature Struct. Mol. Biol.*.
142. Lai, F., Gardini, A., Zhang, A., Shiekhattar, R., (2015). Integrator mediates the biogenesis of enhancer RNAs. *Nature* **525**, 399–403.
143. Fianu, I., Ochmann, M., Walshe, J.L., Dybkov, O., Cruz, J. N., Urlaub, H., et al., (2024). Structural basis of Integrator-dependent RNA polymerase II termination. *Nature*.
144. Zheng, H., Qi, Y., Hu, S., Cao, X., Xu, C., Yin, Z., et al., (2020). Identification of Integrator-PP2A complex (INTAC), an RNA polymerase II phosphatase. *Science* **370**
145. Vos, S.M., Farnung, L., Linden, A., Urlaub, H., Cramer, P., (2020). Structure of complete Pol II-DSIF-PAF-SPT6 transcription complex reveals RTF1 allosteric activation. *Nature Struct. Mol. Biol.* **27**, 668–677.
146. Bian, Y., Ye, M., Wang, C., Cheng, K., Song, C., Dong, M., et al., (2013). Global screening of CK2 kinase substrates by an integrated phosphoproteomics workflow. *Sci. Rep.* **3**, 3460.
147. Schulze, W.M., Cusack, S., (2017). Structural basis for mutually exclusive co-transcriptional nuclear cap-binding complexes with either NELF-E or ARS2. *Nature Commun.* **8**, 1302.
148. Bai, C., Tolias, P.P., (1998). Drosophila clipper/CPSF 30K is a post-transcriptionally regulated nuclear protein that binds RNA containing GC clusters. *Nucleic Acids Res.* **26**, 1597–1604.
149. Bai, C., Tolias, P.P., (1996). Cleavage of RNA hairpins mediated by a developmentally regulated CCCH zinc finger protein. *Mol. Cell Biol.* **16**, 6661–6667.
150. Addepalli, B., Hunt, A.G., (2007). A novel endonuclease activity associated with the Arabidopsis ortholog of the 30-kDa subunit of cleavage and polyadenylation specificity factor. *Nucleic Acids Res.* **35**, 4453–4463.
151. Marzluff, W.F., Koreski, K.P., (2017). Birth and Death of Histone mRNAs. *Trends Genet.* **33**, 745–759.
152. Sun, Y., Zhang, Y., Aik, W.S., Yang, X.C., Marzluff, W.F., Walz, T., et al., (2020). Structure of an active human histone pre-mRNA 3′-end processing machinery. *Science* **367**, 700–703.
153. Rosa-Mercado, N.A., Zimmer, J.T., Apostolidi, M., Rinehart, J., Simon, M.D., Steitz, J.A., (2021). Hyperosmotic stress alters the RNA polymerase II interactome and induces readthrough transcription despite widespread transcriptional repression. *Mol. Cell* **81**, 502–513.e4.

154. Vilborg, A., Passarelli, M.C., Yario, T.A., Tycowski, K.T., Steitz, J.A., (2015). Widespread inducible transcription downstream of human genes. *Mol. Cell* **59**, 449–461.
155. Su, J., Miao, X., Archambault, D., Mager, J., Cui, W., (2021). ZC3H4-a novel Cys-Cys-Cys-His-type zinc finger protein-is essential for early embryogenesis in micedagger. *Biol. Reprod.* **104**, 325–335.
156. Gomez-Gonzalez, B., Aguilera, A., (2019). Transcription-mediated replication hindrance: a major driver of genome instability. *Genes Dev.* **33**, 1008–1026.
157. Hamperl, S., Cimprich, K.A., (2016). Conflict resolution in the genome: how transcription and replication make it work. *Cell* **167**, 1455–1467.
158. Landsverk, H.B., Sandquist, L.E., Bay, L.T.E., Steurer, B., Campsteijn, C., Landsverk, O.J.B., et al., (2020). WDR82/PNUTS-PP1 prevents transcription-replication conflicts by promoting RNA polymerase II degradation on chromatin. *Cell Rep.* **33**, 108469
159. Loh, Y.H., Wu, Q., Chew, J.L., Vega, V.B., Zhang, W., Chen, X., et al., (2006). The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. *Nature Genet.* **38**, 431–440.
160. Mehravar, M., Ghaemimanesh, F., Poursani, E.M., (2021). An overview on the complexity of OCT4: at the Level of DNA, RNA and protein. *Stem Cell Rev. Rep.* **17**, 1121–1136.
161. Moon, B.S., Huang, D., Gao, F., Cai, M., Lyu, G., Zhang, L., et al., (2023). Long range inter-chromosomal interaction of Oct4 distal enhancer loci regulates ESCs pluripotency. *Cell Death Discov.* **9**, 61.
162. Roguev, A., Schaft, D., Shevchenko, A., Pijnappel, W.W., Wilm, M., Aasland, R., et al., (2001). The Saccharomyces cerevisiae Set1 complex includes an Ash2 homologue and methylates histone 3 lysine 4. *EMBO J.* **20**, 7137–7148.
163. Bae, H.J., Dubarry, M., Jeon, J., Soares, L.M., Dargemont, C., Kim, J., et al., (2020). The Set1 N-terminal domain and Swd2 interact with RNA polymerase II CTD to recruit COMPASS. *Nature Commun.* **11**, 2181.
164. Soares, L.M., Buratowski, S., (2012). Yeast Swd2 is essential because of antagonism between Set1 histone methyltransferase complex and APT (associated with Pta1) termination factor. *J. Biol. Chem.* **287**, 15219–15231.
165. Nedea, E., Nalbant, D., Xia, D., Theoharis, N.T., Suter, B., Richardson, C.J., 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.
166. Erickson, B., Fedoryshchak, R., Fong, N., Sheridan, R., Larson, K.Y., Saviola, A.J., Mouilleron, S., Hansen, K.C., Treisman, R., Bentley D.L., (2024). PP1 PNUTS binds the “restrictor” and dephosphorylates RNA pol II CTD Ser5 to stimulate transcription termination, bioRxiv.