The pol II CTD: new twists in the tail

Justyna Zaborowska[1], Sylvain Egloff[2] and Shona Murphy[1]*

[1]Sir William Dunn School of Pathology, University of Oxford, Oxford, UK
[2]Laboratoire de Biologie Moléculaire Eucaryote, CNRS, Université Paul Sabatier, 118 route de Narbonne, 31062 Toulouse cedex 9, France.

*corresponding author:shona.murphy@path.ox.ac.uk

Post-translational modifications of the pol II CTD regulate transcription and co-transcriptional RNA processing. Several new findings prompt a reassessment of the role of the CTD in networks controlling gene expression.

Introduction

RNA polymerase II (pol II) transcribes all protein-coding genes and a number of genes for long and small non-coding RNAs in eukaryotic genomes. The unique carboxyl-terminal domain (CTD) of the largest subunit of this polymerase consists of multiple repeats of the consensus motif Tyr1-Ser2-Pro3-Thr4-Ser5-Pro6-Ser7. Reversible modification of the residues within the repeats generates a code that changes during the transcription cycle and is read by CTD-binding factors. The CTD code and the function of the modifications have been extensively reviewed in recent years[1-3]. However, the CTD modification field continues to move rapidly and here we discuss the recent detailed mapping of CTD modifications in yeast and man by mass spectrometry and how newly described modifications update the CTD code in higher eucaryotes and expand the roles played by the CTD in regulating gene expression. Recent advances also emphasize that the pol II CTD is only one of several substrates in gene expression networks regulated by CTD modification enzymes.

Decorating the pol II CTD

The CTD is a complex and conserved structure only found in pol II. In mammals, the CTD consists of 21 consensus and 31 non-consensus heptad repeats[1,4]. The consensus repeats can be phosphorylated or glycosylated on Tyr1, Ser2, Thr4, Ser5, Ser7 and the prolines isomerized (Figure 1, left panel). Eight of the non-consensus repeats have substitutions of Ser7 to Lys7. Recent chromatin immunoprecipitation (ChIP) and mass spectrometry analyses indicate that these lysines can be modified by acetylation, mono-, di- and tri-methylation and ubiquitylation[5-7]. The arginine at position 1810 (Arg1810) can be asymetrically or symmetrically methylated[8,9] (Figure 1, right panel). Modification of the lysines and arginines in the non-consensus repeats provides the potential for specific codes localised to different regions of the CTD. Interestingly, almost all of the 26 CTD repeats in *S. cerevisiae* and 29 CTD repeats in *S. pombe* follow the consensus sequence[1]. This suggests that the non-consensus repeats have evolved to perform additional functions in higher eucaryotes, although a CTD with all consensus repeats supports viability in human cells[10].

ChIP, mammalian native elongating transcript sequencing (mNET-seq) and mCRAC analyses, using antibodies to specific CTD modifications, indicate that the pattern of CTD repeat modification changes through the transcription cycle of protein-coding genes in both the yeast *S. cerevisiae* and mammals[6,10-13] (Figure 2). The pattern of Ser2 and Ser5 phosphorylation has been most extensively studied and ChIP analysis indicates that Ser5 phosphorylation (Ser5P) is highest close to transcription start sites (TSS), where pol II levels are highest, whereas Ser2P is highest at the 3' end of genes[6,10,11,13]. mNET-seq also indicates that Ser2P is highest at the 3' end of human genes. However, mNET-seq does not show as much Ser5P enrichment at the beginning of genes as ChIP and rather detects high levels of Ser5P over exons[12]. In ChIP, the association of the modified pol II with DNA is being measured; mNET-seq instead measures the RNA in the active site of immunoprecipitated pol II and NET-seq was first developed in yeast to directly measure transcription at single nucleotide resolution[14]. It is therefore possible that much Ser5P at the beginning of genes is associated with pol II that has not yet made the 35nts of RNA necessary for capture by mNET-seq[12]. mNET-seq also detects a peak of splicing-dependent Ser5P at the end of exons, although it is not yet clear whether this is caused by transcriptionally-engaged pol II[12]. Importantly, mNET-seq provides a high-resolution pattern of CTD modification on transcribing pol II.

ChIP studies indicate that Tyr1P, Ser7P, K7ac and K7me1/2 are also highest at the beginning of mammalian protein-coding genes, whereas Thr4P, like Ser2P, is highest at 3' ends[5,6] (Figure 2). Tyr1P is also associated with antisense promoter transcription and active enhancers[11]. CTD ubiquitylation and glycosylation have not yet been mapped genomewide. Glycosylation is distributed throughout the whole human CTD and requires a minimum of 20 heptad repeats[15].

In general, the patterns of modification of the residues in the consensus repeats are similar in yeast and mammals[13,16-21]. The notable exceptions in yeast are Tyr1P, which is enriched in gene bodies and lower at the 5' and 3' ends of genes[16] and Thr4P, which is more uniformly distributed along genes[16], although this mark also peaks after poly(A) sites in yeast genes[17].

Recent genetic manipulation of the CTD combined with mass spectrometry has provided insights into the relative levels and spatial patterns of modifications of the CTD in both *S. cerevisiae*[22,23] and human cells[23]. The CTD is phosphorylated in a uniform manner across all repeats, including the Lys7-containing non-consensus repeats, whether original or introduced artificially to facilitate proteomic analysis. In addition, CTD phosphorylation is generally lower than previously anticipated. In human cells, Ser5P and Ser2P predominate in mono-phosphorylated repeats and make up about 75% of the total phospho-counts[23], consistent with their roles in major transcription and RNA processing events[1]. Ser7P, Tyr1P and Thr4P are present in lower levels (Thr4P-containing heptad repeats represent around 15% and Ser7P or Tyr1P around 5% of the total phospho-counts respectively)[23]. In one study of the yeast CTD, Ser5P and Ser2P together represent nearly all of the total phospho-counts and the level of Tyr1, Thr4 or Ser7 phosphorylation is almost two orders of magnitude lower[22]. However, Thr4P was found at a similar level to Ser2P in a different study on the yeast CTD[23]. Ser5P is the most prevalent modification in mono-phosphorylated repeats in yeast, whereas Ser5P and Ser2P are found in similar quantities on mono-phosphorylated repeats in the human CTD. Interestingly, in both man and yeast, two modifications on any one repeat is a relatively rare event; where this happens, it is generally Ser2P and Ser5P that are found together. In human cells, double-phospho marks across two neighbouring repeats are more common, consistent with the idea that the repeat unit of the CTD is two Tyr-Ser-Pro-Thr-Ser-Pro-Ser repeats[24].

These studies are a necessary complement to the ChIP, mNET-seq, mCRAC and western blot analyses by anti-CTD modification antibodies that cannot discriminate between the regions of the CTD modified or provide comparative quantitation. However, it is important to note that a low level of modification does not indicate a lack of function (see below).

Who's doing the decorating?

Many of the enzymes involved in CTD modification have been identified (Table 1). These enzymes can also modify a range of other substrates and, at least in some cases, are likely to modify the CTD and other substrates at the same time (see below).

Several kinases are implicated in CTD phosphorylation, although further investigation is needed to determine the specificity of mammalian CTD kinases, as the *in vivo* roles of some have been difficult to pin down. Inhibition of analog-sensitive versions of the CDK7 subunit of TFIIH indicates that this is the major Ser5 and Ser7 kinase in both yeast and mammals[25,26]. CDK8, a subunit of Mediator complex, has Ser2 and Ser5 phosphorylation activity *in vitro*. However, its role in CTD phosphorylation *in vivo* remains unclear[27]. CDK19, a paralog of CDK8, is required for induction of hypoxia inducible factor 1A (HIF1A) target genes in response to hypoxia[28]. A recent study using a drug that inhibits both CDK8 and CDK19 coupled with quantitative phosphoproteomics, suggests that both kinases are regulators of chromatin and pol II activity[29].

CDK9 and CDK12/13 are the metazoan homologs of the *S. cerevisiae* CTD Ser2 kinases Bur1 and Ctk1[30] and the *S. pombe* CTD Ser2 kinases CDK9 and Lsk1[31,32], respectively.

However, the identity of the major mammalian Ser2 kinase is still unclear. The CDK9 subunit of P-TEFb is a good candidate as treatment of cells with the CDK9 inhibitors DRB, KM05382 and Flavopiridol reduces Ser2P on western blots (e.g.[33,34]) and Flavopiridol specifically affects Ser2P in mass spectrometry studies[23]. However, recent studies demonstrate that CDK9 colocalizes with Ser5P in nuclei[35] and preferentially phosphorylates Ser5 of the CTD *in vitro*[36]. CDK12 is also implicated in Ser2 CTD phosphorylation[30,37-39] and is recruited to the transcription units of pol II transcribed genes[40-42]. However, knockdown of CDK12 or inhibition of an analog-sensitive (as) version of this kinase does not always greatly affect Ser2 CTD phosphorylation measured by western blotting[43,44]. *In vitro*, CDK12 can phosphorylate Ser5 and Ser7 in addition to Ser2[45]. Knockdown of CDK12 causes a specific defect in expression of DNA damage response genes[37,43], casting doubt on its role as a general Ser2 kinase. Both CDK13 and the dual specificity tyrosine-regulated kinase 1A (DYRK1A) can phosphorylate the CTD at both Ser2 and Ser5 *in vitro*[34,46]. However, DYRK1A appears to be a gene-type specific kinase, which regulates growth-related genes. CDK11, recruited by TREX/THOC to elongating pol II, promotes Ser2 phosphorylation at the 3′ end of HIV templates, which leads to an increase in the levels of cleavage and polyadenylation factors[47]. The bromodomain-containing protein 4 (Brd4) can also function as a Ser2 kinase in the absence of P-TEFb recruitment to genes and in stem cell lines deficient in P-TEFb[48]. There is therefore a wide choice of potential mammalian Ser2 kinases.

Knockdown experiments suggest that Polo-like kinase 3 (PLK3) is a Thr4 specific kinase in human cells[10], whereas the CDK9 inhibitor, DRB affects Thr4P in chicken cells[49]. The kinase phosphorylating Tyr1 has also not been unequivocally identified. C-Abl was thought to be the Tyr1 kinase[50]. However, c-Abl inhibitors have little effect on Tyr1 phosphorylation on transcribing pol II in human cells (*Laitem and Murphy, unpublished observations*[33]).

Symmetric and asymmetric dimethylation of Arg1810 of the human CTD are placed by protein arginine methyltransferase (PRMT5)[9] and coactivator-associated arginine methyltransferase 1 (CARM1)[8], respectively. Schroeder et al. demonstrated that the human CTD is acetylated by p300/KAT3B acetyltransferase[51]. Finally, the enzymes responsible for Lys7 methylation are still unknown.

Removal of the modifications at the right time is likely to be just as important as putting the modifications on in the first place. For example, dephosphorylation of the CTD residues is essential to produce the classic pattern of CTD phosphorylation across genes and replacement of serine 2, 5, or 7 by a phosphomimic is not tolerated[1,4]. The transition to productive elongation is accompanied by loss of several modifications (Figure 2). Human RPAP2 and its yeast orthologue, Rtr1 dephosphorylate Ser5P[52,53], possibly soon after transcription initiation. RPAP2 is part of a cascade of CTD phosphorylation/dephosphorylation regulating expression of snRNA genes and is recruited early in the transcription cycle by CTD Ser7P. RPAP2 is recruited to protein-coding genes instead by RPRD1A/B[54]. Ssu72 dephosphorylates Ser5 and Ser7[18,55,56] and is associated with both ends of protein-coding genes[56]. Ssu72 depletion causes an increase in the levels of Ser5 and Ser7 at the 3'end of genes and a defect in termination of transcription, suggesting a role for CTD dephosphorylation in termination of transcription[18,57]. The evolutionary-conserved TFIIF-associated CTD phosphatase 1 (Fcp1), which preferentially dephosphorylates Ser2P, is also associated with both ends of protein-coding genes[58] but also appears to be most active at the end of the transcription cycle[18]. How Glc7, the proposed Tyr1P phosphatase in yeast[59], is recruited/activated to remove Tyr1P close to the polyA site is still unclear. It was recently demonstrated that Rtr1, in addition to Ser5, can also dephosphorylate Tyr1[60]. The enzymes removing the phosphate from Thr4P, the acetyl group from Lys7ac, the methyl groups from Lys7me1/2 or the asymmetric or symmetric methylation from Arg1810me are currently unknown.

Importantly, there are indications that CTD modification patterns are also shaped by RNA processing factors that interact with the CTD. For example, knockdown of the polyadenylation factor CPSF73, that interacts with the unphosphorylated or phosphorylated CTD[61], affects Ser2P levels at the 3' end of human genes[40] and mutation of the CTD-interaction domain of Pcf11 that specifically recognises Ser2P affects Ser2P levels at the 3' end of yeast genes[62].

The Role of CTD modification

The changes in CTD modification through the transcription cycle are thought to help co-ordinate the sequential recruitment of transcription and RNA processing factors and histone modification enzymes, often through direct interaction with the CTD (Figure 1). Several factors have been shown to specifically interact with different CTD modifications or combinations of modifications[2] and there is a clear logic to the interactions; factors needed early in transcription eg capping enzymes recognise modifications placed early in transcription eg Ser5P and factors that function in RNA 3' end formation recognise marks placed late in the transcription cycle, eg the polyadenylation factor Pcf11 binds to Ser2P (see Figure 1). Ser5P at the end of exons is splicing-dependent, pointing to a role for this modification in coupling splicing to transcription[12]. Indeed, a recent study demonstrated that Ser5P helps to recruits the spliceosome and regulate splicing in yeast[17]. An exciting new discovery is that symmetric dimethylation of Arg1810 is recognised by the Tudor domain of survival of motor neuron (SMN), which is mutated in spinal muscular atrophy. SMN in turn recruits the RNA/DNA hybrid helicase senataxin, which is required to resolve R-loops for efficient termination of transcription. Accordingly, mutation of Arg1810 to alanine causes a termination defect on protein-coding genes[9] (Figure 1, right panel). Mutation of Arg1810 to Ala also causes an increase in expression of snRNAs and snoRNAs[8]. Asymmetric dimethylation of Arg1810 is recognised by the tudor domain of TDRD3 rather than SMN. However, knockdown of TDRD3 has no effect on sn/snoRNA levels, suggesting that recruitment of TDRD3 to the CTD has a role unrelated to sn/snoRNA gene expression. These findings underscore the power of modification of a single amino acid to influence the function of the highly repetitive CTD (Figure 1, Box 1).

It is not yet clear what factors, if any, are recruited by Tyr1P, Ser7P, Kac and Kme during expression of mammalian protein-coding genes. In yeast, Tyr1P is found within coding regions and is associated with elongation of transcription (Figure 2). Phosphorylation of both Ser2P and Tyr1P stabilises binding of the elongation factor Spt6 to CTD repeats *in vitro*[16,63]. As Tyr1P inhibits binding of the termination factors, including Pcf11, *in vitro*[16], loss of this modification at the end of yeast genes could control termination of transcription. This is an interesting example of a modification working to repel rather than recruit a factor. The finding that Ser2P and Tyr1P in the same repeat is relatively rare in both yeast and man readily explains how Pcf11 can still be recruited to the termination regions of human genes when Tyr1P is present[11,33] but raises the question of how phosphorylation of Tyr1 can block the recruitment of termination factors to the coding region of yeast genes, *in vivo*[16].

In yeast, Thr4P helps transcription termination by recruiting the termination factor Rtt103 and regulating pol II pausing after polyadenylation sites. This mark also helps ensure efficient splicing[17]. In chicken DT40 cells, this modification has been linked to 3'end processing of histone transcripts[49]. Thr4P is associated with efficient transcription elongation and M phase progression in human cells[10,64].

Ser7P helps to recruit the Integrator complex to human snRNA genes[53] but whether this modification specifically recruits factors to protein-coding genes is unclear. Tietjen and colleagues have suggested that Ser7P could facilitate elongation and suppress cryptic

transcription[20]. It has also been shown that this mark stimulates CTD Ser2 phosphorylation by CDK9 and CDK12 *in vitro*[36,45].
Acetylation and mono- and di-methylation of Lys7 residues is enriched close to transcription start sites of mammalian genes[5,6] suggesting a role in initiation of transcription or the transition to elongation. Alternatively, pol II paused close to the promoter might be, at least in part, an accidental substrate for acetyltransferases and methylases. Importantly, Lys7 acetylation is required for the transcription activation of c-Fos and Egr[51], indicating that this modification has an important biological role. In addition, functional and evolutionary analyses demonstrate that the emergence of CTD-Lys7 correlates with the origin of Metazoan lineages[65]. Lys7 is also subject to ubiquitylation[7], which is implicated in turnover of the large subunit. However, the importance of CTD ubiquitylation to regulation of gene expression is unclear. Glycosylation occurs before pol II recruitment and its removal may regulate initiation[66].

CTD modification at the centre of networks controlling gene expression

The changing profile of pol II activity across mammalian genes reflects the existence of transition points during the transcription cycle and changes in modification of the CTD are intimately connected to these transitions (Figure 2). It is becoming increasingly apparent that the CTD is usually only one of several substrates modified by "CTD modification enzymes". This is emphasized by the recent findings that "histone modification enzymes" can also modify the CTD[6,8,9,51]. The simultaneous modification of the CTD and several other factors involved in transcription or RNA processing can therefore be at the heart of sophisticated networks capable of exquisite control of gene expression.
The best-characterised example to date of a network including CTD modification involves the activity of the CDK9 subunit of P-TEFb in metazoan cells[67,68] (Figure 3). DRB-sensitivity-inducing factor (DSIF) and negative elongation factor (NELF) halt pol II soon after initiation, which provides a window of opportunity for capping to occur[69]. Phosphorylation of the Spt5 subunit of DSIF and the E subunit of NELF by CDK9 is required to release pol II from this early elongation checkpoint and concomitant phosphorylation of the CTD by CDK9 ensures recruitment of elongation and RNA processing factors that will be needed downstream. CDK9 therefore regulates the transition to productive elongation. CDK9 also phosphorylates histone H1, promoter-binding transcription factors and co-repressors, which regulates initiation of transcription[70,71]. P-TEFb also targets the AFF1 subunit of the super elongation complex (SEC) and the ubiquitin-conjugating enzyme E2A (UBE2A) required for H2B ubiquitylation, which are associated with productive elongation[72,73]. P-TEFb has also been reported to regulate recruitment of the pol II-associated factor 1 complex (PAF1C), which is a critical regulator of paused pol II release[42,74]. It has been recently shown that PAF1C is in turn necessary for recruitment of CDK12 early in the transcription cycle[42]. In addition, CDK12 associates with the exon junction complex (EJC) and SR splicing factors (SRSFs), which also help to recruit it to the RNAPII elongation complex[41]. The recent mNET-seq data suggests that there is a transition in pol II phosphorylation associated with splicing[12] (Figure 2). This echoes the findings that in yeast, pol II is subject to an elongation checkpoint regulated by the formation of competent splicing complexes[75]. Interestingly, splicing factors can help recruit P-TEFb[76] and inhibition of splicing decreases phosphorylation of Ser2 of the CTD[77], indicating a two-way conversation between splicing and CTD modification.
At the other end of genes, CDK9 activity appears to regulate continued elongation and the transition to termination downstream of functional poly(A) sites [33]. Inhibition of CDK9 by DRB, KM05382 and Flavopiridol causes premature termination to occur close to the poly(A) site, at what we have termed the poly(A)-associated checkpoint, rather than a few kilobases

downstream[33]. High resolution mCRAC analysis reveals that changes in pol II and CTD modification levels occurs close to the poly(A) site also during transcription of yeast protein-coding genes[13]. Davidson et al. have suggested that poly(A) signals and CPSF mediate pol II pausing[40]. In turn, pol II pausing may promote phosphorylation of Ser2 of pol II CTD by CDK12, which is important for recruitment of CstF77 and optimal 3' end processing, at least for some genes[40]. There are therefore clear parallels between events at the early elongation checkpoint and the poly(A) site. CDK9 also phosphorylates the termination exoribonuclease Xrn2 at target residue Thr439[73]. Inhibition of CDK9 or mutation of Xrn2-Thr439 to a nonphosphorylatable Ala residue causes termination defects[73], indicating that CDK9 directly regulates transcriptional termination. Thus, CDK9 is a master regulator that drives transcription forward through phosphorylation of targets implicated in successive steps of the pol II transcription cycle, including the CTD.

Future prospects

Although we know a lot more about the placement, removal and role of CTD modifications than we did just a couple of years ago, there are still big gaps in our knowledge.

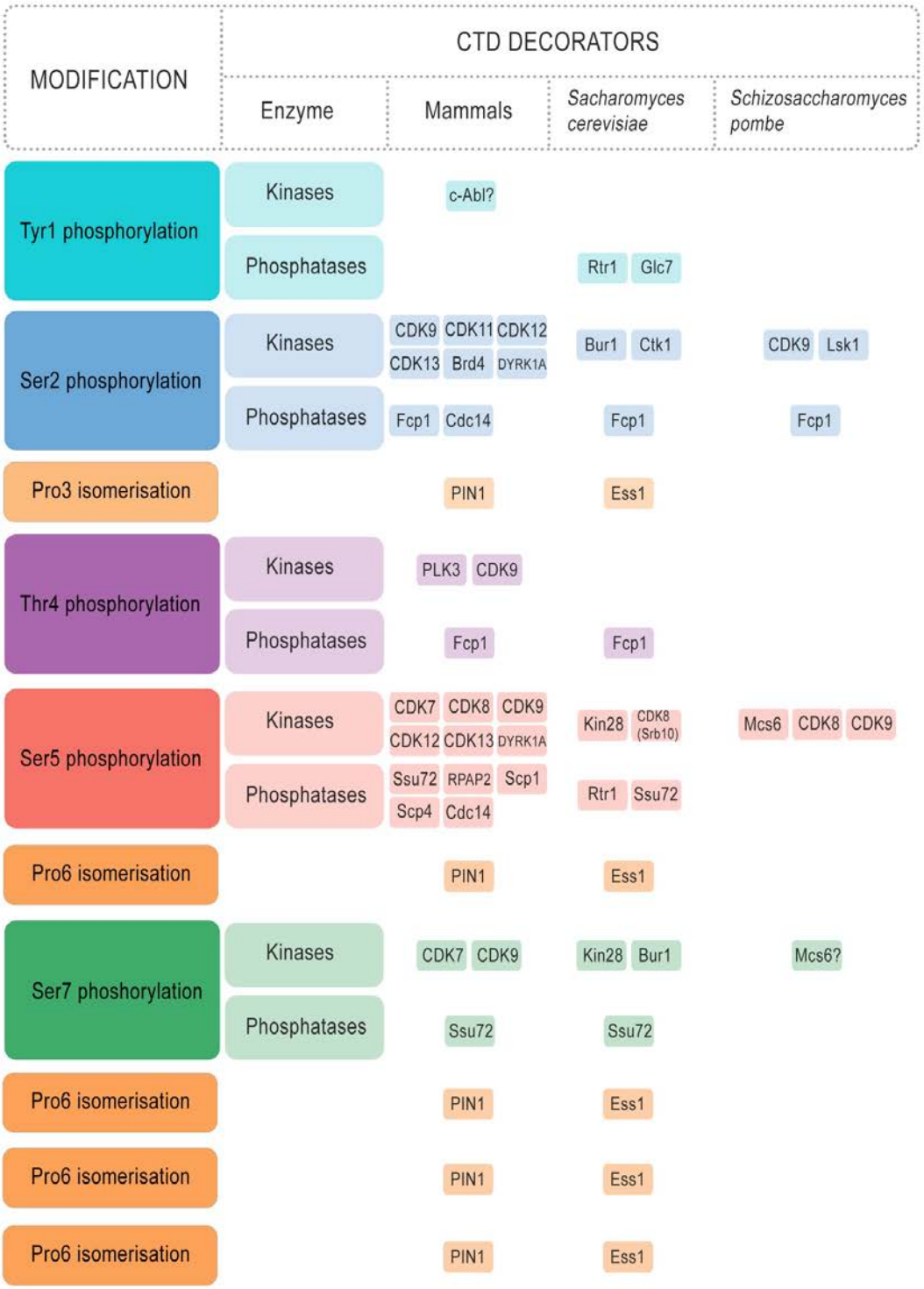
Important goals will be to complete the identification of CTD modification enzymes and characterise the full range of their substrates. The advent of CRISPR/Cas9 genome editing has already facilitated the production of human cell lines bearing analog-sensitive (as)CDK12[44], suggesting that we will soon have more metazoan cell lines carrying as versions of CTD kinases to add to the repertoire already available. Combining the use of asCDKs with phosphoproteomics will help to determine the precise contribution of each CDK to CTD phosphorylation. Understanding the biological roles of the CTD-modifying CDKs is particularly important given their relevance to human diseases including cancer and infection by viruses including HIV, HSV and Influenza[68].

It may also be possible to carry out mass spectrometry analysis on specific subpopulations of Pol II, eg initiating polymerase, pausing polymerase, elongating polymerase or polymerase transcribing a specific class of genes, by purifying pol II associated with transcription cycle stage or gene-type-specific factors. This in turn will help to fully elucidate the role of different CTD modifications in recruitment of CTD binding factors.

Another particularly intriguing question is whether the CTD modification patterns associated with transitions in the transcription cycle are cause or consequence.

We look forward to the answers to the many questions that remain.

Tables



MODIFICATION	CTD DECORATORS			
	Enzyme	Mammals	*Sacharomyces cerevisiae*	*Schizosaccharomyces pombe*
Tyr1 phosphorylation	Kinases	c-Abl?		
	Phosphatases		Rtr1, Glc7	
Ser2 phosphorylation	Kinases	CDK9, CDK11, CDK12, CDK13, Brd4, DYRK1A	Bur1, Ctk1	CDK9, Lsk1
	Phosphatases	Fcp1, Cdc14	Fcp1	Fcp1
Pro3 isomerisation		PIN1	Ess1	
Thr4 phosphorylation	Kinases	PLK3, CDK9		
	Phosphatases	Fcp1	Fcp1	
Ser5 phosphorylation	Kinases	CDK7, CDK8, CDK9, CDK12, CDK13, DYRK1A	Kin28, CDK8 (Srb10)	Mcs6, CDK8, CDK9
	Phosphatases	Ssu72, RPAP2, Scp1, Scp4, Cdc14	Rtr1, Ssu72	
Pro6 isomerisation		PIN1	Ess1	
Ser7 phoshorylation	Kinases	CDK7, CDK9	Kin28, Bur1	Mcs6?
	Phosphatases	Ssu72	Ssu72	
Pro6 isomerisation		PIN1	Ess1	
Pro6 isomerisation		PIN1	Ess1	
Pro6 isomerisation		PIN1	Ess1	

Table 1. Summary of known CTD-modifying enzymes in mammals and yeast [1,3,5,6,8-10,15,26,27,43,46,47,52,53,55,56,59,60,87-90].

Acknowledgements
We would like to thank Lidia Vasiljeva, Nick Proudfoot, Michael Tellier and Ivan Ferrer-Vicens for discussions and critical comments and the reviewers for insightful comments. The authors also thank Flor Vazquez-Arango and Michael Tellier for help with the figures.
This work was supported by grants from the Wellcome Trust (WT106134A) to SM and by Fondation ARC to SE.

References:
1. Eick, D. & Geyer, M. The RNA polymerase II carboxy-terminal domain (CTD) code. *Chem Rev* **113**, 8456-90 (2013).
2. Jeronimo, C., Collin, P. & Robert, F. The RNA Polymerase II CTD: The increasing complexity of a low-complexity protein domain. *J Mol Biol* (2016).
3. Srivastava, R. & Ahn, S.H. Modifications of RNA polymerase II CTD: Connections to the histone code and cellular function. *Biotechnol Adv* **33**, 856-72 (2015).
4. Egloff, S., Dienstbier, M. & Murphy, S. Updating the RNA polymerase CTD code: adding gene-specific layers. *Trends Genet* **28**, 333-41 (2012).
5. Voss, K. et al. Site-specific methylation and acetylation of lysine residues in the C-terminal domain (CTD) of RNA polymerase II. *Transcription*, 0 (2015).
6. Dias, J.D. et al. Methylation of RNA polymerase II non-consensus Lysine residues marks early transcription in mammalian cells. *Elife* **4**(2015).
7. Li, H. et al. Wwp2-mediated ubiquitination of the RNA polymerase II large subunit in mouse embryonic pluripotent stem cells. *Mol Cell Biol* **27**, 5296-305 (2007).
8. Sims, R.J., 3rd et al. The C-terminal domain of RNA polymerase II is modified by site-specific methylation. *Science* **332**, 99-103 (2011).
9. Zhao, D.Y. et al. SMN and symmetric arginine dimethylation of RNA polymerase II C-terminal domain control termination. *Nature* **529**, 48-53 (2016).
10. Hintermair, C. et al. Threonine-4 of mammalian RNA polymerase II CTD is targeted by Polo-like kinase 3 and required for transcriptional elongation. *EMBO J* **31**, 2784-97 (2012).
11. Descostes, N. et al. Tyrosine phosphorylation of RNA polymerase II CTD is associated with antisense promoter transcription and active enhancers in mammalian cells. *Elife* **3**, e02105 (2014).
12. Nojima, T. et al. Mammalian NET-Seq Reveals Genome-wide Nascent Transcription Coupled to RNA Processing. *Cell* **161**, 526-40 (2015).
13. Milligan, L. et al. Strand-specific, high-resolution mapping of modified RNA polymerase II. *Mol Syst Biol* **12**, 874 (2016).
14. Churchman, L.S. & Weissman, J.S. Nascent transcript sequencing visualizes transcription at nucleotide resolution. *Nature* **469**, 368-73 (2011).
15. Lu, L. et al. Distributive O-GlcNAcylation on the Highly Repetitive C-Terminal Domain of RNA Polymerase II. *Biochemistry* **55**, 1149-58 (2016).
16. Mayer, A. et al. CTD tyrosine phosphorylation impairs termination factor recruitment to RNA polymerase II. *Science* **336**, 1723-5 (2012).
17. Harlen, K.M. et al. Comprehensive RNA Polymerase II Interactomes Reveal Distinct and Varied Roles for Each Phospho-CTD Residue. *Cell Rep* **15**, 2147-2158 (2016).
18. Bataille, A.R. et al. A universal RNA polymerase II CTD cycle is orchestrated by complex interplays between kinase, phosphatase, and isomerase enzymes along genes. *Mol Cell* **45**, 158-70 (2012).

19. Mayer, A. et al. Uniform transitions of the general RNA polymerase II transcription complex. *Nat Struct Mol Biol* **17**, 1272-8 (2010).
20. Tietjen, J.R. et al. Chemical-genomic dissection of the CTD code. *Nat Struct Mol Biol* **17**, 1154-61 (2010).
21. Kim, H. et al. Gene-specific RNA polymerase II phosphorylation and the CTD code. *Nat Struct Mol Biol* **17**, 1279-86 (2010).
22. Suh, H. et al. Direct Analysis of Phosphorylation Sites on the Rpb1 C-Terminal Domain of RNA Polymerase II. *Mol Cell* **61**, 297-304 (2016).
23. Schuller, R. et al. Heptad-Specific Phosphorylation of RNA Polymerase II CTD. *Mol Cell* **61**, 305-14 (2016).
24. Stiller, J.W. & Cook, M.S. Functional unit of the RNA polymerase II C-terminal domain lies within heptapeptide pairs. *Eukaryot Cell* **3**, 735-40 (2004).
25. Akhtar, M.S. et al. TFIIH kinase places bivalent marks on the carboxy-terminal domain of RNA polymerase II. *Mol Cell* **34**, 387-93 (2009).
26. Glover-Cutter, K. et al. TFIIH-associated Cdk7 kinase functions in phosphorylation of C-terminal domain Ser7 residues, promoter-proximal pausing, and termination by RNA polymerase II. *Mol Cell Biol* **29**, 5455-64 (2009).
27. Galbraith, M.D., Donner, A.J. & Espinosa, J.M. CDK8: a positive regulator of transcription. *Transcription* **1**, 4-12 (2010).
28. Galbraith, M.D. et al. HIF1A employs CDK8-mediator to stimulate RNAPII elongation in response to hypoxia. *Cell* **153**, 1327-39 (2013).
29. Poss, Z.C. et al. Identification of Mediator Kinase Substrates in Human Cells using Cortistatin A and Quantitative Phosphoproteomics. *Cell Rep* **15**, 436-50 (2016).
30. Bartkowiak, B. et al. CDK12 is a transcription elongation-associated CTD kinase, the metazoan ortholog of yeast Ctk1. *Genes Dev* **24**, 2303-16 (2010).
31. Pei, Y., Schwer, B. & Shuman, S. Interactions between fission yeast Cdk9, its cyclin partner Pch1, and mRNA capping enzyme Pct1 suggest an elongation checkpoint for mRNA quality control. *J Biol Chem* **278**, 7180-8 (2003).
32. Karagiannis, J. & Balasubramanian, M.K. A cyclin-dependent kinase that promotes cytokinesis through modulating phosphorylation of the carboxy terminal domain of the RNA Pol II Rpb1p sub-unit. *PLoS One* **2**, e433 (2007).
33. Laitem, C. et al. CDK9 inhibitors define elongation checkpoints at both ends of RNA polymerase II-transcribed genes. *Nat Struct Mol Biol* **22**, 396-403 (2015).
34. Greifenberg, A.K. et al. Structural and Functional Analysis of the Cdk13/Cyclin K Complex. *Cell Rep* **14**, 320-31 (2016).
35. Ghamari, A. et al. In vivo live imaging of RNA polymerase II transcription factories in primary cells. *Genes Dev* **27**, 767-77 (2013).
36. Czudnochowski, N., Bosken, C.A. & Geyer, M. Serine-7 but not serine-5 phosphorylation primes RNA polymerase II CTD for P-TEFb recognition. *Nat Commun* **3**, 842 (2012).
37. Blazek, D. et al. The Cyclin K/Cdk12 complex maintains genomic stability via regulation of expression of DNA damage response genes. *Genes Dev* **25**, 2158-72 (2011).
38. Bowman, E.A., Bowman, C.R., Ahn, J.H. & Kelly, W.G. Phosphorylation of RNA polymerase II is independent of P-TEFb in the C. elegans germline. *Development* **140**, 3703-13 (2013).

39. Cheng, S.W. et al. Interaction of cyclin-dependent kinase 12/CrkRS with cyclin K1 is required for the phosphorylation of the C-terminal domain of RNA polymerase II. *Mol Cell Biol* **32**, 4691-704 (2012).
40. Davidson, L., Muniz, L. & West, S. 3' end formation of pre-mRNA and phosphorylation of Ser2 on the RNA polymerase II CTD are reciprocally coupled in human cells. *Genes Dev* **28**, 342-56 (2014).
41. Eifler, T.T. et al. Cyclin-dependent kinase 12 increases 3' end processing of growth factor-induced c-FOS transcripts. *Mol Cell Biol* **35**, 468-78 (2015).
42. Yu, M. et al. RNA polymerase II-associated factor 1 regulates the release and phosphorylation of paused RNA polymerase II. *Science* **350**, 1383-6 (2015).
43. Liang, K. et al. Characterization of human cyclin-dependent kinase 12 (CDK12) and CDK13 complexes in C-terminal domain phosphorylation, gene transcription, and RNA processing. *Mol Cell Biol* **35**, 928-38 (2015).
44. Bartkowiak, B., Yan, C. & Greenleaf, A.L. Engineering an analog-sensitive CDK12 cell line using CRISPR/Cas. *Biochim Biophys Acta* **1849**, 1179-87 (2015).
45. Bosken, C.A. et al. The structure and substrate specificity of human Cdk12/Cyclin K. *Nat Commun* **5**, 3505 (2014).
46. Di Vona, C. et al. Chromatin-wide profiling of DYRK1A reveals a role as a gene-specific RNA polymerase II CTD kinase. *Mol Cell* **57**, 506-20 (2015).
47. Pak, V. et al. CDK11 in TREX/THOC Regulates HIV mRNA 3' End Processing. *Cell Host Microbe* **18**, 560-70 (2015).
48. Devaiah, B.N. et al. BRD4 is an atypical kinase that phosphorylates serine2 of the RNA polymerase II carboxy-terminal domain. *Proc Natl Acad Sci U S A* **109**, 6927-32 (2012).
49. Hsin, J.P., Sheth, A. & Manley, J.L. RNAP II CTD phosphorylated on threonine-4 is required for histone mRNA 3' end processing. *Science* **334**, 683-6 (2011).
50. Baskaran, R., Escobar, S.R. & Wang, J.Y. Nuclear c-Abl is a COOH-terminal repeated domain (CTD)-tyrosine (CTD)-tyrosine kinase-specific for the mammalian RNA polymerase II: possible role in transcription elongation. *Cell Growth Differ* **10**, 387-96 (1999).
51. Schroder, S. et al. Acetylation of RNA polymerase II regulates growth-factor-induced gene transcription in mammalian cells. *Mol Cell* **52**, 314-24 (2013).
52. Mosley, A.L. et al. Rtr1 is a CTD phosphatase that regulates RNA polymerase II during the transition from serine 5 to serine 2 phosphorylation. *Mol Cell* **34**, 168-78 (2009).
53. Egloff, S., Zaborowska, J., Laitem, C., Kiss, T. & Murphy, S. Ser7 phosphorylation of the CTD recruits the RPAP2 Ser5 phosphatase to snRNA genes. *Mol Cell* **45**, 111-22 (2012).
54. Ni, Z. et al. RPRD1A and RPRD1B are human RNA polymerase II C-terminal domain scaffolds for Ser5 dephosphorylation. *Nat Struct Mol Biol* **21**, 686-95 (2014).
55. Krishnamurthy, S., He, X., Reyes-Reyes, M., Moore, C. & Hampsey, M. Ssu72 Is an RNA polymerase II CTD phosphatase. *Mol Cell* **14**, 387-94 (2004).
56. Zhang, D.W. et al. Ssu72 phosphatase-dependent erasure of phospho-Ser7 marks on the RNA polymerase II C-terminal domain is essential for viability and transcription termination. *J Biol Chem* **287**, 8541-51 (2012).
57. O'Reilly, D. et al. Human snRNA genes use polyadenylation factors to promote efficient transcription termination. *Nucleic Acids Res* **42**, 264-75 (2014).

58. Cho, E.J., Kobor, M.S., Kim, M., Greenblatt, J. & Buratowski, S. Opposing effects of Ctk1 kinase and Fcp1 phosphatase at Ser 2 of the RNA polymerase II C-terminal domain. *Genes Dev* **15**, 3319-29 (2001).
59. Schreieck, A. et al. RNA polymerase II termination involves C-terminal-domain tyrosine dephosphorylation by CPF subunit Glc7. *Nat Struct Mol Biol* **21**, 175-9 (2014).
60. Hsu, P.L. et al. Rtr1 is a dual specificity phosphatase that dephosphorylates Tyr1 and Ser5 on the RNA polymerase II CTD. *J Mol Biol* **426**, 2970-81 (2014).
61. McCracken, S. et al. The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. *Nature* **385**, 357-61 (1997).
62. Grzechnik, P., Gdula, M.R. & Proudfoot, N.J. Pcf11 orchestrates transcription termination pathways in yeast. *Genes Dev* **29**, 849-61 (2015).
63. Burugula, B.B. et al. Histone deacetylases and phosphorylated polymerase II C-terminal domain recruit Spt6 for cotranscriptional histone reassembly. *Mol Cell Biol* **34**, 4115-29 (2014).
64. Hintermair, C. et al. Specific threonine-4 phosphorylation and function of RNA polymerase II CTD during M phase progression. *Sci Rep* **6**, 27401 (2016).
65. Simonti, C.N. et al. Evolution of lysine acetylation in the RNA polymerase II C-terminal domain. *BMC Evol Biol* **15**, 35 (2015).
66. Ranuncolo, S.M., Ghosh, S., Hanover, J.A., Hart, G.W. & Lewis, B.A. Evidence of the involvement of O-GlcNAc-modified human RNA polymerase II CTD in transcription in vitro and in vivo. *J Biol Chem* **287**, 23549-61 (2012).
67. Bres, V., Yoh, S.M. & Jones, K.A. The multi-tasking P-TEFb complex. *Curr Opin Cell Biol* **20**, 334-40 (2008).
68. Zaborowska J, I.N., Murphy S. P-TEFb goes viral. *Inside Cell* **1:1-11;**(2016).
69. Kwak, H. & Lis, J.T. Control of transcriptional elongation. *Annu Rev Genet* **47**, 483-508 (2013).
70. O'Brien, S.K., Cao, H., Nathans, R., Ali, A. & Rana, T.M. P-TEFb kinase complex phosphorylates histone H1 to regulate expression of cellular and HIV-1 genes. *J Biol Chem* **285**, 29713-20 (2010).
71. Kim, J.B. & Sharp, P.A. Positive transcription elongation factor B phosphorylates hSPT5 and RNA polymerase II carboxyl-terminal domain independently of cyclin-dependent kinase-activating kinase. *J Biol Chem* **276**, 12317-23 (2001).
72. Shchebet, A., Karpiuk, O., Kremmer, E., Eick, D. & Johnsen, S.A. Phosphorylation by cyclin-dependent kinase-9 controls ubiquitin-conjugating enzyme-2A function. *Cell Cycle* **11**, 2122-7 (2012).
73. Sanso, M. et al. P-TEFb regulation of transcription termination factor Xrn2 revealed by a chemical genetic screen for Cdk9 substrates. *Genes Dev* **30**, 117-31 (2016).
74. Chen, F.X. et al. PAF1, a Molecular Regulator of Promoter-Proximal Pausing by RNA Polymerase II. *Cell* **162**, 1003-15 (2015).
75. Chathoth, K.T., Barrass, J.D., Webb, S. & Beggs, J.D. A splicing-dependent transcriptional checkpoint associated with prespliceosome formation. *Mol Cell* **53**, 779-90 (2014).
76. Fong, Y.W. & Zhou, Q. Stimulatory effect of splicing factors on transcriptional elongation. *Nature* **414**, 929-33 (2001).
77. Koga, M., Hayashi, M. & Kaida, D. Splicing inhibition decreases phosphorylation level of Ser2 in Pol II CTD. *Nucleic Acids Res* **43**, 8258-67 (2015).

78. Yoh, S.M., Cho, H., Pickle, L., Evans, R.M. & Jones, K.A. The Spt6 SH2 domain binds Ser2-P RNAPII to direct Iws1-dependent mRNA splicing and export. *Genes Dev* **21**, 160-74 (2007).
79. Vasiljeva, L., Kim, M., Mutschler, H., Buratowski, S. & Meinhart, A. The Nrd1-Nab3-Sen1 termination complex interacts with the Ser5-phosphorylated RNA polymerase II C-terminal domain. *Nat Struct Mol Biol* **15**, 795-804 (2008).
80. Noble, C.G. et al. Key features of the interaction between Pcf11 CID and RNA polymerase II CTD. *Nat Struct Mol Biol* **12**, 144-51 (2005).
81. Werner-Allen, J.W. et al. cis-Proline-mediated Ser(P)5 dephosphorylation by the RNA polymerase II C-terminal domain phosphatase Ssu72. *J Biol Chem* **286**, 5717-26 (2011).
82. Kubicek, K. et al. Serine phosphorylation and proline isomerization in RNAP II CTD control recruitment of Nrd1. *Genes Dev* **26**, 1891-6 (2012).
83. Rosonina, E. et al. Threonine-4 of the budding yeast RNAP II CTD couples transcription with Htz1-mediated chromatin remodeling. *Proc Natl Acad Sci U S A* **111**, 11924-31 (2014).
84. Lewis, B.A. O-GlcNAcylation at promoters, nutrient sensors, and transcriptional regulation. *Biochim Biophys Acta* **1829**, 1202-6 (2013).
85. Chatterjee, D., Sanchez, A.M., Goldgur, Y., Shuman, S. & Schwer, B. Transcription of lncRNA prt, clustered prt RNA sites for Mmi1 binding, and RNA polymerase II CTD phospho-sites govern the repression of pho1 gene expression under phosphate-replete conditions in fission yeast. *RNA* (2016).
86. Coudreuse, D. et al. A gene-specific requirement of RNA polymerase II CTD phosphorylation for sexual differentiation in S. pombe. *Curr Biol* **20**, 1053-64 (2010).
87. Clemente-Blanco, A. et al. Cdc14 phosphatase promotes segregation of telomeres through repression of RNA polymerase II transcription. *Nat Cell Biol* **13**, 1450-6 (2011).
88. Kim, M., Suh, H., Cho, E.J. & Buratowski, S. Phosphorylation of the yeast Rpb1 C-terminal domain at serines 2, 5, and 7. *J Biol Chem* **284**, 26421-6 (2009).
89. Zhang, M. et al. Structural and kinetic analysis of prolyl-isomerization/phosphorylation cross-talk in the CTD code. *ACS Chem Biol* **7**, 1462-70 (2012).
90. Wani, S., Sugita, A., Ohkuma, Y. & Hirose, Y. Human SCP4 is a chromatin-associated CTD phosphatase and exhibits the dynamic translocation during erythroid differentiation. *J Biochem* (2016).