Taylor & Francis
Taylor & Francis Group

POINT-OF-VIEW

🔓 OPEN ACCESS

A single flexible RNAPII-CTD integrates many different transcriptional programs

Maria J. Aristizabal and Michael S. Kobor

Centre for Molecular Medicine and Therapeutics, Child and Family Research Institute, Department of Medical Genetics, University of British Columbia, Vancouver, British Columbia, Canada

ABSTRACT
The RNAPII-CTD functions as a binding platform for coordinating the recruitment of transcription associated factors. Altering CTD function results in gene expression defects, although mounting evidence suggests that these effects likely vary among species and loci. Here we highlight emerging evidence of species- and loci-specific functions for the RNAPII-CTD.

Over thirty years of research have illuminated salient aspects about the composition, function, and regulation of the C-terminal domain (CTD) of RNA polymerase II (RNAPII).[1,2] The CTD is composed of tandem heptapeptide repeats that typically follow a Y1-S2-P3-T4-S5-P6-S7 consensus motif.[3,4] Highlighting its importance, complete removal of the CTD results in lethality in all species tested, although shortened versions are tolerated across species.[5] The CTD is located on the catalytic subunit of RNAPII, although it is not required for RNA synthesis *in vitro*.[5] Instead, the CTD coordinates co-transcriptional activities such as mRNA capping, splicing, and polyadenylation, in part by acting as a dynamic platform for the recruitment of factors involved in these processes.[1,2] CTD activity is intimately linked to a growing repertoire of posttranslational modifications including phosphorylation, ubiquitination, methylation, glycosylation, acetylation, and *cis-trans* isomerization of proline residues. Some of these modifications occur in uniform patterns along the length of actively transcribed genes, forming a "CTD code" that marks progression through the transcription cycle and regulates ordered recruitment of transcription-associated factors.[2] CTD phosphorylation is the best understood of these

modifications, and it can occur on five of the seven residues making up the CTD repeat: Y1, S2, T4, S5, and S7. Of these, S5 and S2 phosphorylation (S5p and S2p respectively) are the most abundant forms of this mark,[6,7] and play key roles during the transcription cycle. Briefly, S5p levels are highest at the 5′ end of most genes and these function to recruit mRNA capping factors early in the transcription process.[2] In comparison, S2p levels increase toward the 3′ end of genes where they recruit elongation, splicing, and termination factors.

Understanding CTD function has been a community endeavor, capitalizing on the unique properties and experimental advantages of a variety of organisms. Collectively, findings from individual species have provided insight into conservation and divergence of RNAPII-CTD function across evolution. Following its discovery, numerous pieces of evidence suggested species-specific activities for the RNAPII-CTD, starting with the observation that CTD composition varies substantially between species.[5] For example, the CTDs of *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Drosophila melanogaster*, and mammals consist of 25–26, 29, 44, and 52 repeats, respectively. Of these, most repeats follow the

---

CONTACT Maria J. Aristizabal ✉ marist@cmmt.ubc.ca 📧 Centre for Molecular Medicine and Therapeutics, Child and Family Research Institute, Department of Medical Genetics, University of British Columbia, Vancouver, British Columbia V5Z 4H4, Canada.
Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/ktrn.

consensus sequence in *S. cerevisiae* and *S. pombe*, while in mammals only the proximal half and in *D. melanogaster* only a handful of repeats adhere to the consensus motif. Moreover, replacing the *S. cerevisiae* CTD with the longer mammalian version is supported *in vivo*, resulting in no obvious growth phenotypes, although substituting it with the more variable sequence from *D. melanogaster* results in lethality.[8] Consistent with these findings, sequence alterations are differentially tolerated among species. To illustrate, *S. cerevisiae* and *S. pombe* CTDs can be replaced with non-phosphorylatable residues at positions 2 and 4 of the heptapeptide repeat, while in mammals similar substitutions result in lethality.[9] Thus, early genetic evidence generally supported the notion of distinct functional requirements for the RNAPII-CTD across species. More recently, studies focused on the importance of posttranslational modifications in CTD function have provided additional evidence. These showed that distribution patterns of less well-characterized CTD phosphorylation marks differ between species. In mammals, Y1 phosphorylation is exclusive to gene promoters,[10] while in *S. cerevisiae*, it is found at the 3′ end of genes (Fig. 1).[11] Similarly, T4 phosphorylation increases toward the 3′ end of genes in mammals,[12] although is found uniformly along the length of genes in *S. cerevisiae*.[11] Most significantly, recent studies reported distinct functions for Y1 phosphorylation in *S. cerevisiae* and mammals. In *S. cerevisiae*, Y1 phosphorylation prevents premature recruitment of termination factors,[11] while in mammals it functions at gene promoters to regulate upstream antisense transcription.[13] Additional observations suggest that modifications other than phosphorylation may also contribute to CTD function in a species-specific manner. For example, CTD methylation and acetylation function exclusively in higher eukaryotes, given that the non-consensus residues targeted by these modifications are not conserved in yeast.[14-16] As a whole, these findings are an important stepping-stone in the larger goal of understanding how RNAPII accommodates differences in genome composition and transcriptional requirements across species.

RNAPII is responsible for the transcription of most protein-coding genes, some of which have unique regulatory and processing needs. For example, in contrast to the majority of mRNAs, histone transcripts have unique 3′ end processing requirement and are not polyadenylated.[17] Perhaps not surprisingly, emerging evidence suggests that the RNAPII-CTD orchestrates transcriptional programs in a gene-specific manner. Specifically, it has been observed that altering CTD function affects expression of relatively few RNAs, and that individual loci are uniquely dependent on specific letters of the CTD heptapeptide repeat for normal expression.[13,14,18-25] Systematic RNA-sequencing studies in *S. pombe*, wherein each letter of the heptapeptide repeat targeted by phosphorylation was replaced to a non-phosphorylatable residue, revealed that no more than 5% of protein-coding genes depended on any one letter of the CTD repeat for normal expression.[18] Similar findings were observed in *S. cerevisiae* strains containing truncated CTDs or CTDs with T4 residues replaced with valine,[20,24] and, in mammals, with R1810A substitutions, which abolish CTD methylation.[26] Sensitive loci in *S. pombe* include nucleotide, sugar, and amino acid metabolism genes whose mRNA levels decrease in Y1F and S2A substituted CTDs (Fig. 2); iron regulated genes whose mRNA levels increase upon Y1F and S2A substitutions; and meiotic genes, which are normally repressed but show increased levels when S7, Y1, and S2 residues are substituted with non-phosphorylatable residues.[18] The finding that altering CTD function leads to increases in gene expression suggests that the RNA-PII-CTD not only functions to facilitate transcription, but also to impede it, although it will be important to tease apart the degree to which this represents direct vs. indirect effects. Regardless, negative roles for the RNAPII-CTD have also been described in other species and will be discussed in more detail below. In mammals, candidate gene approaches showed that substituting T4 residues with valine decreased the mRNA levels of histone genes, while other mRNAs were unaffected (Fig. 2).[21] Importantly, this effect was not observed in genome-wide expression profiles of *S. cerevisiae* or *S. pombe* strains containing similar substitutions, highlighting an exclusive role for T4 residues in mammals. Instead, the yeast expression profiles showed transcriptional defects of genes involved in phosphate metabolism (Fig. 2).[18,20] Intriguingly, T4 substitutions in *S. cerevisiae* (T4A substitution) and *S. pombe* (T4V substitution) had opposite effects on the expression of phosphate metabolism genes, with these increasing in *S. cerevisiae* while decreasing in *S. pombe*.[18,20] Although suggestive of additional species-specific roles for T4 residues among different types of yeasts, one trivial
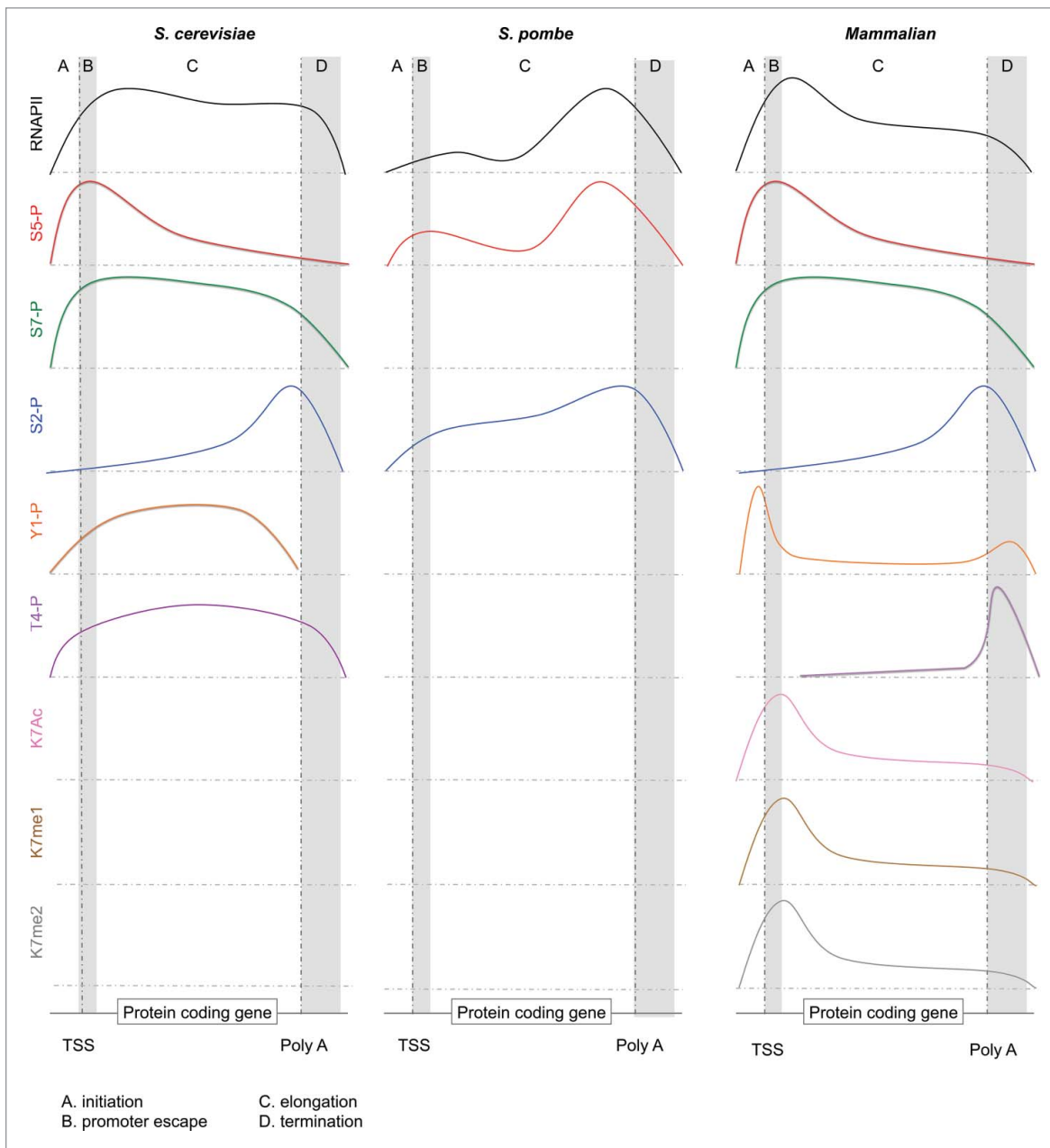
Figure 1. Differences in reported CTD modification patterns along the length of genes in *S. cerevisiae, S. pombe*, and mammals. Schematic of reported average gene profiles of CTD modifications in the indicated species. Adapted from Coudreuse et al. 2010, Eick and Geyer 2013, Descostes et al. 2014, Dias et al. 2015, and Voss et al. 2015.[2,10,16,19,30] TSS and Poly A refer to the transcription start site and polyadenylation signal respectively.

explanation for this variability is that the constructs employed differed in the amino acid used to replace the T4 residue and CTD length (Fig. 2; * indicates amino acid substitutions generated in shortened versions of the CTD). Collectively, these studies performed under basal growth conditions demonstrate that protein-coding genes naturally vary in their CTD requirements for normal expression. In addition, candidate gene approaches have shown that this is also the case for genes transcribed upon changes in

environmental conditions. For example, in *S. cerevisiae* expression of the phosphate-induced gene *PHO5* and galactose-induced genes *GAL1/7* is reduced when T4 residues are replaced with valine (Fig. 2).[20] Similarly, in *S. pombe, PHO1* and the starvation response gene *STE11* show reduced induction when S5 or S2 residues are changed to alanine respectively (Fig. 2).[19,27] Importantly, beyond protein-coding genes, RNAPII is also involved in the transcription of many types of non-coding elements, and loci-specific
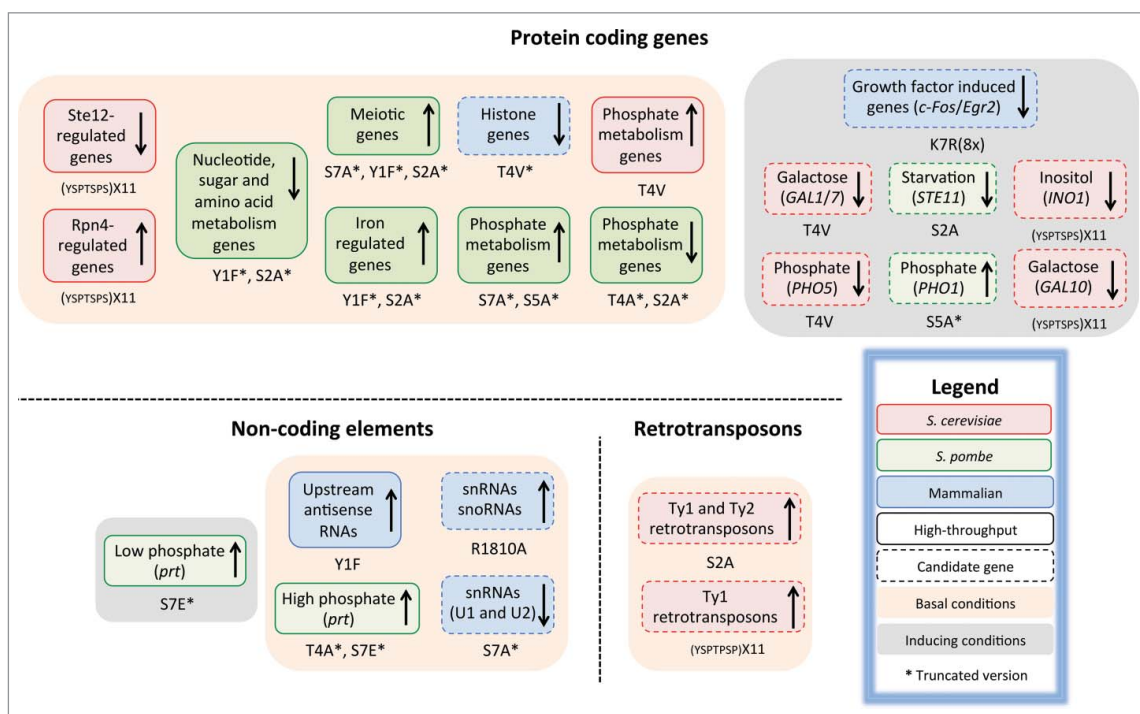
Figure 2. Summary of the effects of altering RNAPII-CTD composition on basal and induced gene expression. Each box represents a gene group that is uniquely affected by the CTD alteration indicated below. Arrows indicate the direction of the expression change. For inducing conditions, the arrow represents greater or lower levels of induction compared to wild type. "High-throughput" and "candidate gene" refer to the type of gene expression evidence giving rise to that observation.

functions for the RNAPII-CTD have also been described for these. In mammals, S7A and Y1F substitutions uniquely affect expression of small nuclear RNAs (snRNAs)[22] and upstream antisense RNAs respectively.[13] Similarly, in *S. pombe* basal expression of the phosphate regulated non-coding RNA, *prt*, is affected when the RNAPII-CTD contains T4A or S7E substitutions, while its induction is elevated in strains containing S7E substituted CTDs.[27] As such, the evidence of loci-specific functions for the RNAPII-CTD is widespread across species and different types of loci.

Currently, the molecular underpinnings of loci-specific functions for the RNAPII-CTD remain unclear. Offering a clue, most CTD substitutions that elicit loci-specific effects on gene expression predominantly abolish specific phosphorylation marks, suggesting that these may underlie loci-specific activities for the RNAPII-CTD. However, CTD phosphorylation profiles have not been examined in detail at genes uniquely dependent on individual letters of the CTD repeat for normal expression. Thus, it remains to be determined if and how CTD phosphorylation events contribute to loci-specific transcription regulation. It is also likely that other CTD modifications may underlie gene-specific

functions for the RNAPII-CTD, given that similar to CTD phosphorylation, substitutions that affect CTD methylation and acetylation also affect gene expression in loci-specific manner.[14,15,26] Briefly, in mammals, loss of CTD methylation through a R1810A substitution increases the levels of snRNAs.[14,26] Similarly, changing all lysine residues at positions 7 of the heptapeptide repeat to arginine [Fig. 2 denoted K7R(8x)] both prevents CTD acetylation and reduces the induction of *c-FOS* and *ERG2* upon addition of growth factors.[15] Focusing on the latter, wider peaks of CTD acetylation were observed at actively transcribed genes compared to those with paused RNAPIIs. However, whether CTD acetylation patterns change in a gene-specific manner upon addition of growth factors was not explored.

In addition to CTD modifications, CTD length also contributes to CTD function. In fact, early evidence revealed that organisms have minimal CTD length requirements to support life, while mutants carrying viable but shortened versions display growth, gene expression, and RNA processing defects.[5] For example, *S. cerevisiae* strains with shortened CTDs exhibit reduced growth fitness,

mammalian cells show reduced size, and mice display increased neonatal lethality.[5] Focusing on gene expression, shortening the CTD affects mRNA levels of a subset of genes, suggesting that like CTD sequence, CTD length contributes to gene expression in a loci-specific manner.[23-25] For instance, early work in *S. cerevisiae* unveiled a role for the CTD in activated transcription in particular conditions by showing reduced induction of the *GAL10* and *INO1* genes when the CTD was truncated, although induction of *HIS4* was unaffected [Fig. 2 denoted (YSPTSPS)x11].[25] This effect extends to genes expressed under normal growth conditions, given that in *S. cerevisiae* progressively truncating the RNAPII-CTD resulted in increasing number and severity of gene expression changes.[24] Focusing on the affected genes, these were relatively few and mainly regulated by Ste12 and Rpn4, gene-specific transcription factors that regulate mating and proteasome genes, respectively. Interestingly, reducing CTD length increased the levels of genes regulated by Rpn4, providing further evidence of a negative role for the RNAPII-CTD in transcription.[24] Most recently, the RNAPII-CTD was also shown to limit the expression of Ty1 retrotransposons, thus directly functioning in the maintenance of genome stability.[23] Although the molecular underpinnings of negative functions for the RNAPII-CTD remain to be fully determined, the increased expression levels of Rpn4-regulated genes and Ty1 retrotransposons were mediated by changes in promoter activity and suppressed by loss of the *SRB10/CDK8* gene, which encodes the kinase subunit of the Mediator complex identified in the original SRB screen.[24] Thus, gene-specific signaling pathways that contribute to negative functions for the RNAPII-CTD are beginning to emerge. CTD length has also been shown to be important for splicing, 3′ processing, and transcription termination at a subset of representative genes, although how it generally affects mRNA processing remains unknown.[5] Furthermore, CTD length also influences mRNA capping-CTD interactions, although its role in the recruitment of mRNA capping factors may be more nuanced and gene-specific than previously thought.[24] More specifically, genome-wide maps of capping factor occupancy in *S. cerevisiae* strains containing CTDs with only 11 heptapeptide repeats showed significant loss of capping factor

recruitment to most genes, although a subset of highly transcribed genes were unaffected.[24] Nevertheless, whether these observations translate to general mRNA capping defects has not been examined.

Recent mass spectrometry studies in *S. cerevisiae* and mammals revealed that the entire length of the CTD can undergo phosphorylation at Y, S, and T residues, suggesting that shortening this domain likely reduces its modification potential.[6,7] In support of this relationship, shortening the mammalian CTD to 24 repeats largely abolishes S7 phosphorylation[28] and, in *S. cerevisiae*, reducing CTD length or preventing S2 phosphorylation results in similar gene expression defects at Ty1 retrotransposons and galactose-induced genes (Fig. 2).[23] However, thus far, not all gene expression alterations associated with CTD truncation mutants are recapitulated in strains with CTD sequence alterations,[23,27] indicating that CTD length may also contribute to CTD function independently of its effect on CTD phosphorylation. On its own, CTD sequence is also important for function, as illustrated in mammals, where CTDs composed entirely of non-consensus repeats are unable to support viability.[29] Additionally, although truncation of 20 repeats are tolerated in mammals with minimal effects to cell growth and viability, loss of a single non-consensus repeat at position 52 in otherwise full-length CTDs results in Rpb1 protein instability and lethality. Teasing apart the relationship between CTD sequence and length is important, given that many CTD mutants employed so far have been generated in strains carrying viable but shortened versions of the CTD (Fig. 2). Overall, beyond the snapshot of current knowledge described here, much remains to be learned about the molecular underpinnings of species- and loci-specific activities for the RNAPII-CTD.

Disclosure of potential conflicts of interest

Acknowledgments

Funding

References

[1] Srivastava R, Ahn SH. Modifications of RNA polymerase II CTD: Connections to the histone code and cellular function. Biotechnol Adv 2015; 33:856-72; PMID:26241863; http://dx. doi.org/10.1016/j.biotechadv.2015.07.008

[2] Eick D, Geyer M. The RNA polymerase II carboxy-terminal domain (CTD) code. Chem Rev 2013; 113:8456-90; PMID:23952966; http://dx.doi.org/10.1021/cr400071f

[3] Allison LA, Moyle M, Shales M, Ingles CJ. Extensive homology among the largest subunits of eukaryotic and prokaryotic RNA polymerases. Cell 1985; 42:599-610; PMID: 3896517; http://dx.doi.org/10.1016/0092-8674(85)90117-5

[4] Corden JL, Cadena DL, Ahearn JM, Dahmus ME. A unique structure at the carboxyl terminus of the largest subunit of eukaryotic RNA polymerase II. Proc Natl Acad Sci USA 1985; 82:7934-8; PMID:2999785; http://dx.doi.org/10.1073/pnas.82.23.7934

[5] Corden JL. RNA polymerase II C-terminal domain: Tethering transcription to transcript and template. Chem Rev 2013; 113:8423-55; PMID:24040939; http://dx.doi.org/10.1021/cr400158h

[6] Schüller R, Forné I, Straub T, Schreieck A, Texier Y, Shah N, Decker TM, Cramer P, Imhof A, Eick D. Heptad-Specific Phosphorylation of RNA Polymerase II CTD. Mol Cell 2016; 61:305-14; PMID:26799765; http://dx.doi.org/10.1016/j.molcel.2015.12.003

[7] Suh H, Ficarro SB, Kang UB, Chun Y, Marto JA, Buratowski S. Direct Analysis of Phosphorylation Sites on the Rpb1 C-Terminal Domain of RNA Polymerase II. Mol Cell 2016; 61:297-304; PMID:26799764; http://dx.doi.org/10.1016/j.molcel.2015.12.021

[8] Allison LA, Wong JK, Fitzpatrick VD, Moyle M, Ingles CJ. The C-terminal domain of the largest subunit of RNA polymerase II of Saccharomyces cerevisiae, Drosophila melanogaster, and mammals: a conserved structure with an essential function. Mol Cell Biol 1988; 8:321-9; PMID:3122024; http://dx.doi.org/10.1128/MCB.8.1.321

[9] Hsin JP, Manley JL. The RNA polymerase II CTD coordinates transcription and RNA processing. Genes Dev 2012; 26:2119-37; PMID:23028141; http://dx.doi.org/10.1101/gad.200303.112

[10] Descostes N, Heidemann M, Spinelli L, Schüller R, Maqbool MA, Fenouil R, Koch F, Innocenti C, Gut M, Gut I, et al. Tyrosine phosphorylation of RNA polymerase II CTD is associated with antisense promoter transcription and active enhancers in mammalian cells. Elife 2014; 3:e02105; PMID:24842994; http://dx.doi.org/10.7554/eLife.02105

[11] Mayer A, Heidemann M, Lidschreiber M, Schreieck A, Sun M, Hintermair C, Kremmer E, Eick D, Cramer P. CTD tyrosine phosphorylation impairs termination factor recruitment to RNA polymerase II. Science 2012; 336:1723-5; PMID:22745433; http://dx.doi.org/10.1126/science.1219651

[12] Hintermair C, Heidemann M, Koch F, Descostes N, Gut M, Gut I, Fenouil R, Ferrier P, Flatley A, Kremmer E, et al. Threonine-4 of mammalian RNA polymerase II CTD is targeted by Polo-like kinase 3 and required for transcriptional elongation. EMBO J 2012; 31:2784-97; PMID:22549466; http://dx. doi.org/10.1038/emboj.2012.123

[13] Hsin J-P, Li W, Hoque M, Tian B, Manley JL. RNAP II CTD tyrosine 1 performs diverse functions in vertebrate cells. Elife 2014; 3:e02112; PMID:24842995; http://dx.doi.org/10.7554/eLife.02112

[14] Sims RJ, Rojas LA, Beck D, Bonasio R, Schüller R, Drury WJ, Eick D, Reinberg D. The C-terminal domain of RNA polymerase II is modified by site-specific methylation. Science 2011; 332:99-103; PMID:21454787; http://dx.doi.org/10.1126/science.1202663

[15] Schröder S, Herker E, Itzen F, He D, Thomas S, Gilchrist DA, Kaehlcke K, Cho S, Pollard KS, Capra JA, et al. Acetylation of RNA polymerase II regulates growth-factor-induced gene transcription in mammalian cells. Mol Cell 2013; 52:314-24; PMID:24207025; http://dx.doi.org/10.1016/j.molcel.2013.10.009

[16] Dias JD, Rito T, Torlai Triglia E, Kukalev A, Ferrai C, Chotalia M, Brookes E, Kimura H, Pombo A. Methylation of RNA polymerase II non-consensus Lysine residues marks early transcription in mammalian cells. Elife 2015; 4:387; http://dx.doi.org/10.7554/eLife.11215

[17] Hsin J-P, Xiang K, Manley JL. Function and control of RNA polymerase II C-terminal domain phosphorylation in vertebrate transcription and RNA processing. Mol Cell Biol 2014; 34:2488-98; PMID:24752900; http://dx.doi.org/10.1128/MCB.00181-14

[18] Schwer B, Bitton DA, Sanchez AM, Bähler J, Shuman S. Individual letters of the RNA polymerase II CTD code govern distinct gene expression programs in fission yeast. Proc Natl Acad Sci USA 2014; 111:4185-90; PMID:24591591; http://dx.doi.org/10.1073/pnas.1321842111

[19] Coudreuse D, van Bakel H, Dewez M, Soutourina J, Parnell T, Vandenhaute J, Cairns B, Werner M, Hermand D. A gene-specific requirement of RNA polymerase II CTD phosphorylation for sexual differentiation in S. pombe. Curr Biol 2010; 20:1053-64; PMID:20605454; http://dx.doi.org/10.1016/j.cub.2010.04.054

[20] Rosonina E, Yurko N, Li W, Hoque M, Tian B, Manley JL. Threonine-4 of the budding yeast RNAP II CTD couples transcription with Htz1-mediated chromatin remodeling. Proc Natl Acad Sci USA 2014; 111:11924-31; PMID:25071213; http://dx.doi.org/10.1073/pnas.1412802111

[21] Hsin J-P, Sheth A, Manley JL. RNAP II CTD phosphorylated on threonine-4 is required for histone mRNA 3' end processing. Science 2011; 334:683-6; PMID:22053051; http://dx.doi.org/10.1126/science.1206034

[22] Egloff S, O'Reilly D, Chapman RD, Taylor A, Tanzhaus K, Pitts L, Eick D, Murphy S. Serine-7 of the RNA polymerase II CTD is specifically required for snRNA gene expression. Science 2007; 318:1777-9; PMID:18079403; http://dx.doi.org/10.1126/science.1145989

[23] Aristizabal MJ, Negri GL, Kobor MS. The RNAPII-CTD Maintains Genome Integrity through Inhibition of Retrotransposon Gene Expression and Transposition. PLoS

Genet 2015; 11:e1005608; PMID:26496706; http://dx.doi. org/10.1371/journal.pgen.1005608

[24] Aristizabal MJ, Negri GL, Benschop JJ, Holstege FCP, Krogan NJ, Kobor MS. High-throughput genetic and gene expression analysis of the RNAPII-CTD reveals unexpected connections to SRB10/CDK8. PLoS Genet 2013; 9:e1003758; PMID:24009531; http://dx.doi.org/ 10.1371/journal.pgen.1003758

[25] Scafe C, Chao D, Lopes J, Hirsch JP, Henry S, Young RA. RNA polymerase II C-terminal repeat influences response to transcriptional enhancer signals. Nature 1990; 347:491-4; PMID:2215664; http://dx.doi.org/ 10.1038/347491a0

[26] Zhao DY, Gish G, Braunschweig U, Li Y, Ni Z, Schmitges FW, Zhong G, Liu K, Li W, Moffat J, et al. SMN and symmetric arginine dimethylation of RNA polymerase II C-terminal domain control termination. Nature 2016; 529:48-53; PMID:26700805; http://dx.doi.org/10.1038/nature16469

[27] Schwer B, Sanchez AM, Shuman S. RNA polymerase II CTD phospho-sites Ser5 and Ser7 govern phosphate homeostasis

in fission yeast. RNA 2015; 21:1770-80; PMID:26264592; http://dx.doi.org/10.1261/rna.052555.115

[28] Chapman RD, Heidemann M, Albert TK, Mailhammer R, Flatley A, Meisterernst M, Kremmer E, Eick D. Transcribing RNA polymerase II is phosphorylated at CTD residue serine-7. Science 2007; 318:1780-2; PMID:18079404; http://dx. doi.org/10.1126/science.1145977

[29] Chapman RD, Conrad M, Eick D. Role of the mammalian RNA polymerase II C-terminal domain (CTD) nonconsensus repeats in CTD stability and cell proliferation. Mol Cell Biol 2005; 25:7665-74; PMID: 16107713; http://dx.doi.org/10.1128/MCB.25.17.7665-7674.2005

[30] Voß K, Forné I, Descostes N, Hintermair C, Schüller R, Maqbool MA, Heidemann M, Flatley A, Imhof A, Gut M, et al. Site-specific methylation and acetylation of lysine residues in the C-terminal domain (CTD) of RNA polymerase II. Transcription 2015; 6:91-101; http://dx.doi.org/10.1080/21541264. 2015.1114983