The functions of the RNA polymerase II CTD in transcription and RNA processing

Jing-Ping Hsin

Submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

in the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2013

Abstract

The functions of the RNA polymerase II CTD in transcription and RNA processing

Jing-Ping Hsin

RNA polymerase II (RNAP II), transcribing messenger RNAs (mRNAs), small nuclear RNAs (snRNAs), and non-coding RNAs (ncRNAs), is composed of 12 subunits. Rpb1, the largest subunit with catalytic polymerase activity, possesses a unique c-terminal domain (CTD) that consists of tandem heptad repeats with the consensus sequence of Tyr-Ser-Pro-Thr-Ser-Pro-Ser ($Y^1S^2P^3T^4S^5P^6S^7$). Somewhat reflecting the complexity of the organism, the number of repeats varies, from 26 in yeast to 52 in vertebrates. The CTD, intensively phosphorylated during transcription, serves a means to coordinate transcription and RNA processing- capping, splicing, and 3' end formation. For example, Ser 5, phosphorylated in the start of transcription, promotes the recruitment of capping enzyme, and Ser 2 phosphorylation facilitates RNA 3' end formation and transcription termination by acting as a landing pad for Pcf11. Detailed introduction is described in Chapter 1. Because of the importance of the CTD in these events, I created an Rpb1 conditional knock-out DT40 cell line (DT40-Rpb1) to further study the CTD with an initial focus on Thr 4, the function of which was unclear. Using DT40-Rpb1 system, we found that Thr 4 was phosphorylated in yeast, fly, chicken, and human cells, and cyclin-dependent kinase (CDK9) was likely the kinase to carry out this phosphorylation. We further provide evidence that Thr 4 functions in histone mRNA 3' end formation (presented mostly in chapter 2 of this thesis).

Chapter 3 mainly describes the studies regarding Ser 2, Ser 5, and Ser 7. Based on the DT40-Rpb1 cell line, I created stable cell lines expressing an Rpb1 carrying a CTD with Ser 2, Ser 5, or Ser 7 mutated to alanine, and investigated the phenotypes of these cells. We found that cells expressing an Rpb1 with S2A or S5A mutation were defective in transcription and RNA processing. Contrary to previous findings, we found Ser 7 was not involved in snRNA expression and 3' end processing. In fact, no phenotypes associated with Ser 7 mutation were detected by our measurements. Extending previous Thr 4 studies, we showed *in vitro* and *in vivo* that Fcp1 dephosphorylated Thr 4.

Finally, Chapter 4 describes what we have found the functions of CTD Tyr 1. Using the DT40-Rpb1 cells, I created stable cell lines expressing an Rpb1 with all Tyr residues mutated to phenylalanine (Phe). We found these cells were inviable, and the mutant Rpb1-Y1F was degraded to a CTD-less protein.

Interestingly, the instability of Rpb1-Y1F was restored by reintroduction of one Tyr residue at the last heptad repeat. Further analysis provided evidence showing the involvement of Tyr phosphorylation in preventing Rpb1 from degradation by the 20S proteasome. Next, using ChIP assay, we showed Tyr phosphorylation was detected mostly at promoters, indicating a function of Tyr phosphorylation in transcription initiation. Indeed, transcription initiation defects were uncovered by assessing the recruitment of general transcription factors in cells with Y1F mutation. Extending this, we found an accumulation of upstream antisense RNAs in about one hundred reference genes by RNA-Seq analysis.

TABLE OF CONTENTS

Acknowledgements

I would like first to thank my thesis research advisor, Jim Manley, for support over these years. From working in his lab, I have learned a great deal: how to write papers, defend ideas in the papers, and present data. I also like to thank my thesis committee members Larry Chasin and Carol Prives for support and many useful suggestions, and my other committee members, Liang Tong, Robert Fisher, and Stewart Shuman for taking the time to be on my thesis defense committee.

I enjoy very much working in our very diverse lab with members of different backgrounds and ways of thinking. In the beginning of my project, I worked closely with Tsuyoshi Kashima, who taught me how to build the Rpb1 conditional knock-out cells. To him, I owe many thanks. I am thankful that I often get useful ideas through discussions with many lab members, especially Patricia Richard, Kehui Xiang, Ying Feng, Dafne Campigli Di Giammartino, Chuck David, and Emanuel Rosonina. I am grateful that Patricia, Chong Han Ng, Dafne, Tristan Coady, and Emanuel rehearsed many talks with me. Other lab members have been helpful in many ways- the assistance from Amit Sheth, Alex Boyne often voluntarily watching the levels of my running buffer, and etc.

My family has been very supportive in this PhD journey. I appreciate vey much the encouragement they gave me when I was struggling with my research project.

Dedication

To my family

Preface

This thesis is divided into four chapters. Chapter 1 is an introduction to the functions of the CTD of RNA polymerase II, and has been published as a review article in *Genes and Development*. Chapter 2 is a research article entitled "RNAP II CTD phosphorylated on threonine 4 is required for histone mRNA 3' end processing," published in *Science*. Chapter 3 is a manuscript with a title "Function and control of RNA polymerase II CTD phosphorylation in vertebrate transcription and RNA processing" that has been submitted for publication. Chapter 4 is the draft of a manuscript in preparation for publication.

Chapter 1

The RNA polymerase II CTD coordinates transcription and RNA processing

Jing-Ping Hsin and James L. Manley

Department of Biological Sciences

Columbia University

New York, NY 10027, USA

Abstract

The carboxy terminal domain of the RNA polymerase II largest subunit (CTD) consists of multiple heptad repeats (consensus Tyr_1-Ser_2-Pro_3-Thr_4-Ser_5-Pro_6-Ser_7), varying in number from 26 in yeast to 52 in vertebrates. The CTD functions to help couple transcription and processing of the nascent RNA, and also plays roles in transcription elongation and termination. The CTD is subject to extensive posttranslational modification, most notably phosphorylation, during the transcription cycle, which modulates its activities in the above processes. Therefore, understanding the nature of CTD modifications, including how they function and how they are regulated, is essential to understanding the mechanisms that control gene expression. While the significance of phosphorylation of Ser2 and Ser5 residues has been studied and appreciated for some time, several additional modifications have more recently been added to the CTD repertoire, and insight into their function has begun to emerge. Here, we review findings regarding modification and function of the CTD, highlighting the important role this unique domain plays in coordinating gene activity.

Introduction

Eukaryotes have three nuclear DNA-dependent RNA polymerases, RNA polymerase (RNAP) I, II and III. RNAP II, responsible for the synthesis of all mRNA as well as many non-coding RNAs, consists of twelve polypeptides, of which Rpb1, which possesses the enzyme's catalytic activity, is the largest. Rpb1 also contains a carboxyl-terminal domain (CTD), composed of tandem heptad repeats, that constitutes a unique feature of RNAP II and distinguishes it from all other polymerases. The CTD is conserved from fungi to human although the number of repeats and their deviation from the consensus varies, reflecting to a large degree the complexity of the organism. The CTD plays important roles at all steps of the transcription process, including enhancing or modulating the efficiency of all the RNA processing

reactions required for completion of synthesis of the mature RNA. The phosphorylation state of the CTD is critical in determining its activity.

Transcription by RNAP II is an immensely complicated process, at each step from initiation to termination. Initiation requires the assembly of the preinitiation complex (PIC), composed of the general transcription factors (GTFs), TFIIA, B, D, E, F and H, the Mediator complex, and RNAP II with an unphosphorylated CTD (designated RNAP IIA) (reviewed by Cramer 2004; Hirose and Ohkuma 2007; Sikorski and Buratowski 2009). But concomitant with initiation and throughout the transcription cycle, the CTD becomes highly phosphorylated (RNAP IIO), with the Ser2 and Ser5 positions of the heptad being especially important sites of modification. A generalized model of CTD phosphorylation during transcription depicts that at the beginning of genes, the CTD is phosphorylated on Ser5 by the TFIIH-associated kinase CDK7, and as RNAP II elongates, Ser2 is increasingly phosphorylated, by CDK9, or pTEFb, while Ser5 phosphorylation is gradually removed by phosphatases. Therefore, a phosphorylation pattern emerges with Ser5 phosphorylation (Ser5-P) peaks around transcription start site and Ser2 phosphorylation (Ser2-P) accumulates towards the ends of transcribed genes (reviewed by Chapman et al. 2008; Egloff and Murphy 2008). But as we will discuss, the situation is now known to be considerably more complex.

The CTD orchestrates multiple events during the transcription process. Genes transcribed by RNAP II are bound dynamically by nucleosomes. The CTD, modulated by phosphorylation, provides a means to recruit histone modifiers and chromatin-remodeling complexes to influence transcription initiation, elongation and termination (reviewed by Spain and Govind 2011). RNA transcribed by RNAP II is processed to mature RNA through the steps of 5’ capping, intron removal, and 3’ end formation. These processes are frequently coupled with transcription, and all involve the CTD (reviewed by Hirose and Manley 2000; Maniatis and Reed 2002; Proudfoot et al. 2002). Through different patterns of

phosphorylation and other modifications, the CTD functions to coordinate these events during transcription (reviewed by Phatnani and Greenleaf 2006; Buratowski 2009; Perales and Bentley 2009). It is striking that the simplicity of CTD structure, i.e., YSPTSPS consensus repeats, is coupled with complex and dynamic patterns of modification to organize the numerous and even more complex nuclear events necessary to form mature transcripts. Below we review studies documenting the structure and function of the RNAP II CTD.

CTD architecture and modification

The structure of the RNAP II CTD displays a number of intriguing and idiosyncratic features. For example, vertebrate CTDs contain 52 tandem consensus $Y^1S^2P^3T^4S^5P^6S^7$ heptapeptide repeats, whereas the budding yeast CTD has only 26, and other eukaryotes have intermediate numbers. In vertebrate CTDs, 21 out of 52 tandem repeats match the consensus perfectly, while the remaining 31 heptads have one or more substitution, at position 2, 4, 5, and/or 7 (Corden et al. 1985; Chapman et al. 2008; Liu et al. 2010). Strictly conserved consensus heptads occur mostly in the N-terminal half of the CTD, whereas less conserved, divergent, heptads are found largely in the C-terminal half (Figure 1). The most conserved residues are tyrosine 1 (Tyr1) and proline 6 (Pro6), present in all 52 repeats, while serine 7 (Ser7) is the least conserved, found in only 26 repeats. A ten residue sequence is present at the very C terminus, which functions to help stabilize the CTD (Chapman et al. 2004). The CTD extends from the core of the enzyme to form a tail-like structure that provides binding sites for various factors involving in RNA processing and transcription-coupled modifications. The structural length of the tail-like CTD is determined by the phosphorylation status, with unphosphorylated yeast CTD likely forming a more compacted spiral (~100 Å), and the phosphorylated CTD forming a more extended tail (~650 Å), about four times the diameter of the surface of RNAP II (reviewed by Meinhart et al. 2005).

The CTD is subject to extensive modification. While the most widely studied of these is phosphorylation, glycosylation, ubiquitinylation, and methylation have also been reported (Kelly et al. 1993; Li et al. 2007; Sims et al. 2011). All five of the hydroxylated amino acids can be phosphorylated. Coupled with the fact that a number of the heptads are degenerate, the existence of different combinations of phosphorylated residues suggests that the CTD may form a remarkable number of distinct conformations or functional states. This potential complexity can be further increased by the activity of cis/trans prolyl isomerases. For example, Pin1 (Ess1 in yeast) specifically recognizes phosphoserine-proline bonds (Yaffe et al. 1997), and its activity can alter the structure (Verdecia et al. 2000) and function (Xu et al. 2003; Xu and Manley 2007; Xiang et al. 2010) of the CTD.

Genetic analyses in a number of organisms have provided insight into the number of heptads required for cell viability, as well as into the importance of specific residues in the repeat and the spacing between heptads. In general, cells expressing Rpb1 with CTDs consisting of only ~50% the natural numbers of heptads are viable. For example, the CTD of budding yeast contains 26 repeats, but only 8 heptads are required for cell viability and 13 for wild type-like growth (West and Gorden 1995). Similarly, chicken DT40 cells expressing an Rpb1 with only 26 all-consensus repeats are fully viable (Hsin et al. 2011). Mouse cells require for viability a CTD with more than 26 repeats (Bartolomei et al. 1988), and mice homozygous for a deletion of 13 repeats of the CTD are viable, but with high neonatal mortality (Litingtung et al. 1999). Requirements for specific residues can vary from species to species. Thus, in *S. cerevisiae* substitution of all Tyr1 residues to phenylalanine (phe), or Ser2 or Ser5 to alanine (ala) is lethal (West and Gorden 1995). However, surprisingly, in *S. pombe*, only Ser5 is absolutely essential, and cells with Tyr1 mutated to phe or Ser2 replaced with ala, although cold sensitive, are viable (Schwer and Shuman 2011). It will be interesting to learn the requirement of these residues for viability in vertebrates. Both budding and fission yeast with Thr4 or Ser7 substituted with ala are viable

(Stiller et al. 2000; Schwer and Shuman 2011). However, DT40 cells expressing an Rpb1 with all Thr4 residues changed to valine (val) are inviable (Hsin et al. 2011), as are human cells with Thr4 to ala substitutions (Hintermair et al. 2012). With respect to spacing between heptads, experiments in yeast defined a minimal functional unit that requires two consecutive heptads. Insertion of single ala residues between such diheptad units was without effect, while insertion of five alas resulted in slow-growth but cells remained viable, indicating a tolerance for considerable flexibility in the CTD (Stiller and Cook 2004; Liu et al. 2008).

CTD kinases

A number of kinases have been implicated in phosphorylation of the CTD. For example, the CTD is phosphorylated by several cyclin-dependent kinases, most notably CDK7 and CDK9. Ser5 is phosphorylated by CDK7, a component of TFIIH, and Ser2 is phosphorylated by CDK9, or P-TEFb in mammals (reviewed by Meinhart et al. 2005; Hirose and Ohkuma 2007; Egloff and Murphy 2008; Buratowski 2009). CDK7 also appears to phosphorylate Ser7 (Akhtar et al. 2009; Glover-Cutter et al. 2009), and CDK9 is required for phosphorylation of Thr4 (Hsin et al. 2011). Surprisingly, a Polo-like kinase (Plk3) has also recently been implicated in Thr4 phosphorylation (Hintermair et al. 2012). Tyrosine residues are phosphorylated by abl tyrosine kinases, Abl1 and Abl2 (Baskaran et al. 1993; 1997).

CDK7 is a component of the GTF TFIIH. Both yeast (Kin28) and human CDK7 were initially discovered as TFIIH-associated kinase activities (Feaver et al. 1991; Lu et al. 1992), and the activity of CDK7/cyclin H was shown to be necessary for reconstituted *in vitro* transcription (Akoulitchev et al. 1995). Consistent with its presence in TFIIH, CDK7 functions during transcription initiation, phosphorylating Ser5 (Hengartner et al. 1998). In yeast, mutations that reduce Kin28 kinase activity abolish Ser5 phosphorylation at promoters (Komarnitsky et al. 2000; Rodriguez et al. 2000). More

recently, evidence was presented that shows CDK7 also phosphorylates Ser7 *in vitro*, and inhibition of CDK7 activity decreases the levels of Ser7-P on a collection of genes *in vivo*. RNAP II in early transcription is thus phosphorylated on both Ser5 and Ser7 residues (Akhtar et al. 2009; Glover-Cutter et al. 2009; Kim et al. 2009).

The CTD can also be phosphorylated by a second PIC component, CDK8, a subunit of the Mediator. CDK8/Cyclin C was shown to phosphorylate the CTD, on both Ser2 and Ser5, *in vitro* (Liao et al. 1995; Sun et al. 1998). Furthermore, the ability of TFIIH to stimulate transcription was shown to be repressed by CDK8, indicating CDK8 functions as a negative transcription regulator, perhaps by its ability to phosphorylate the CTD (Hengartner et al. 1998; Akoulitchev et al. 2000). However, deleting *SRB10* (yeast CDK8) does not affect CTD phosphorylation level in cells (Rodriguez et al. 2000), and how much CDK8 in fact contributes to phosphorylating the CTD *in vivo* is unclear (reviewed by Galbraith et al. 2010). CDK8/Srb10 indeed targets a number of other transcription factors. For example, phosphorylation of the yeast transcription factor Gcn4 by Srb10 leads to Gcn4 promoter clearance (Rosonina et al. 2012) and degradation (Chi et al. 2001), and a number of studies have provided evidence that CDK8 can play a coactivator role (reviewed by Galbraith et al. 2010).

The CTD of elongating RNAP II is phosphorylated by P-TEFb. P-TEFb was initially found to overcome pausing of RNAP II near promoters and to stimulate transcription elongation *in vitro* (Marshall and Price 1995; Marshall et al. 1996). P-TEFb is composed of CDK9 and cyclin T, and CDK9 kinase activity is inhibited by 5,6-dichloro-1-β-D-ribofuranosylbenzimidazole (DRB) (Peng et al. 1998a, b), which was known to block transcription elongation *in vitro* (Marciniak and Sharp 1991). CDK9's ability to trigger the transition from transcription initiation to elongation was further substantiated by studies demonstrating that CDK9 stimulates transcript elongation from the HIV-1 promoter (Herrmann and Rice 1995; Fujinaga

et al. 1998). The HIV transcriptional transactivator Tat binds to the transactivation response element in the nascent RNA to enhance elongation by recruiting P-TEFb to phosphorylate the CTD (Zhou et al. 2000).

P-TEFb is also involved in regulating the activities of two elongation factors, DRB sensitivity-inducing factor (DSIF) and negative elongation factor (NELF). After transcription initiation, the two factors associate with RNAP II, and pause the elongating RNAP II downstream from the transcription start site (TSS) (Yamaguchi et al. 1999; Wu et al. 2003). To overcome this pausing, P-TEFb phosphorylates the two factors as well as the CTD, and elongation proceeds after dissociation of phosphorylated NELF from the RNAP II complex (Renner et al. 2001; Fujinaga et al. 2004; Cheng and Price 2007). Significantly, about one third of genes, both in fly and human cells, appear to contain a paused RNAP II downstream of the TSS (Core et al. 2008; Nechaev et al. 2010). Pausing is thought to constitute a mechanism to obtain rapid and coordinated transcription during development and in response to external stimuli (Muse et al. 2007; Zeitlinger et al. 2007). The detailed molecular mechanisms responsible for pausing remain under investigation (reviewed by Chiba et al. 2010; Levine 2011).

Budding yeast has two cyclin-dependent kinases, Bur1 and Ctk1, that function similarly to CDK9 of higher eukaryotes. CTDK-1, composed of Ctk1, the cyclin Ctk2, and a regulatory subunit, Ctk3, like pTEFb, is capable of stimulating transcription elongation by phosphorylating the CTD as shown by *in vitro* transcription assays (Sterner et al. 1995; Lee and Greenleaf 1997), and *CTK1* deletion abolishes Ser2 phosphorylation *in vivo* (Cho et al. 2001). Bur1, together with the Bur2 cyclin, can phosphorylate the CTD *in vitro* (Yao et al. 2000; Murray et al. 2001). Both Ctk1 and Bur1 function in transcription elongation, but the main function of Bur1 *in vivo* may not to be as a CTD kinase (Keogh et al. 2003). Instead, Bur1/Bur2 is involved in establishing histone H2B monoubiquitinylation and H3K4 trimethylation at promoters by recruiting a ubiquitin ligase and histone methyl-transferase, and the bulk

of Ser2 phosphorylation reflects the action of Ctk1 (Wood and Shilatifard 2006; although see Bartkowiak et al. 2010). Indeed, Bur kinase, recruited by the CTD phosphorylated on Ser5 by Kin28, phosphorylates Ser2 at promoter regions, and subsequent Ctk1 recruitment, stimulated by Bur1, phosphorylates Ser2 further downstream in the coding region (Qiu et al. 2009). This model is supported by a recent genome-wide analysis showing that Bur1 is recruited to promoter regions prior to Ctk1 (Zhang et al. 2012). The promoter-proximal preference of Bur1 explains why only a mild reduction of Ser2 phosphorylation was observed in cells with impaired Bur kinase, since the majority of Ser2 phosphorylation occurs more downstream (Keogh et al. 2003; Qiu et al. 2009).

The existence of two CDK9-like kinases in yeast begs the question, is there a second Ser2 kinase in higher eukaryotes? Recently, Greenleaf and colleagues (Bartkowiak et al. 2010) reported that *Drosophila* CDK12, which is conserved in humans, can phosphorylate the CTD on Ser2, and proposed, based on phylogenetic relationships (Guo and Stiller 2004), that the ortholog of Bur1 is CDK9, whereas CDK12 is the counterpart of Ctk1. Furthermore, depletion of CDK12 or the associated cyclin, CycK, in human cells reduced both total Ser2 levels and expression of select genes including a number of DNA damage response genes (Blazek et al. 2011). Global RNAP II transcription, however, was not affected detectably in the CDK12/CycK-depleted cells. Thus, the contribution of CDK12 to Ser2 phosphorylation, and overall RNAP II transcription, while intriguing, requires further study.

Ser2 and Ser5 are both followed by pro residues, which is characteristic of substrates of cyclin-dependent kinases, such as CDK7 and 9. But this structure has also raised the question whether the CTD might be phosphorylated by classical CDKs that function in cell-cycle control. Indeed, the first CTD kinase identified was the mitotic regulator Cdc2 (Cisek and Corden 1989), although the significance of this finding was unclear for some time. However, Xu et al. (2003) subsequently showed that the CTD can be phosphorylated in M phase by Cdc2/cyclin B, to generate a hyperphosphorylated RNAP II isoform

dubbed RNAP IIOO. This hyperphosphorylation, which requires Pin1 (also known to be a mitotic regulator; Lu et al. 1996), inhibits RNAP II, contributing to mitotic gene silencing. On the other hand, given the preference of CDKs for ser/thr-pro substrates, it is perhaps surprising that CDK7 and CDK9 appear to phosphorylate Ser7 and Thr4, respectively, although in each case precedents exist for targeting non-S/T-P substrates (Larochelle et al. 2006; Baumli et al. 2008).

CTD phosphatases

The pattern of CTD phosphorylation during the transcription cycle is highly dynamic, and requires the activity of dedicated phosphatases as well as kinases. For example, as discussed above, the CTD is hypophosphorylated in the PIC, and becomes highly phosphorylated on multiple residues during transcription. Thus, both to ready RNAP II for new rounds of transcription and to regulate the dynamic CTD phosphorylation status during transcription, specific CTD residues must be dephosphorylated throughout the transcription cycle. This involves the activities of two major phosphatases, Fcp1 and Ssu72, conserved from yeast to human, and others such as the small CTD phosphatases (SCPs).

Fcp1 (TFIIF-associating CTD phosphatase 1) was initially described as an activity in HeLa cells (Chesnut et al. 1992) and in yeast (Archambault et al. 1997) that can dephosphorylate the CTD. Human Fcp1 was found to function on elongating RNAP II, and to recycle RNAP II for PIC assembly (Cho et al. 1999), and inactivation of *FCP1* results in impaired transcription and lethality in yeast (Kobor et al. 1999). Fcp1 is capable of dephosphorylating both Ser2-P and Ser5-P (Lin et al. 2002), although it appears to prefer Ser2-P (Cho et al. 2001; Hausmann and Shuman 2002; Ghosh et al. 2008). Consistent with this, while Fcp1 is present on active genes at both 5' and 3' ends (Cho et al. 2001; Calvo and Manley 2005), in global analyses, higher levels were found in promoter-distal regions (Zhang et al. 2012) and mutations in *FCP1* lead to increased levels of Ser2-phosphorylation (Bataille et al. 2012). Thus Fcp1 is likely

responsible for dephosphorylating the CTD at the end of the transcription cycle, but may also function earlier in the transcription process.

Ssu72 was first found in a genetic screen as a suppressor of a defect in the GTF TFIIB (Sun and Hampsey 1996). Later it was identified as a component of yeast cleavage and polyadenylation factor (CPF), and shown to function in 3' end formation of both polyadenylated and non-polyadenylated RNA (Dichtl et al. 2002a; He et al. 2003; Nedea et al. 2003). Essentially simultaneously, Ssu72 was shown to be a CTD phosphatase with preference for Ser5-P (Ganem et al. 2003; Krishnamurthy et al. 2004). Consistent with both these observations, the genome-wide distribution of Ssu72 indicates that Ssu72 peaks at both the promoter and 3' end of genes, with more present in 3' end regions (Zhang et al. 2012), and depletion of Ssu72 leads to increased levels of Ser5-P towards the 3' end of genes (Bataille et al. 2012). The interaction with TFIIB now appears to reflect the role of these factors in the phenomenon of gene looping (Singh and Hampsey 2007). The phosphatase activity of yeast Ssu72 is stimulated *in vivo* by Pta1, a component of CPF (Ghazy et al. 2009). This interaction is conserved in higher eukaryotes, as human Ssu72 activity is stimulated *in vitro* by a direct interaction with the human homologue of Pta1, Symplekin (Xiang et al. 2010). Although the phosphatase activity of Ssu72 is not required for its function in 3' processing in yeast cell extracts (Ghazy et al. 2009), it is in HeLa extracts but only when the 3' processing reaction is coupled to transcription (Xiang et al. 2010). This suggests that Ser5 dephosphorylation is important for the role of the CTD in coupling transcription and polyadenylation (see below). An unexpected finding was that Ssu72 favors the CTD substrate with the Ser5-Pro6 peptide bond in the *cis* configuration, which contrasts with all other known CTD phosphatases (Xiang et al. 2010; Werner-Allen et al. 2011). Consistent with the fact that the *cis* configuration is energetically unfavored, dephosphorylation of the CTD by Ssu72 is enhanced by Pin1 (Xiang et al. 2010). In line with this, *ESS1* and *SSU72* interact genetically, and mutations in Ess1 result in accumulation of Ser5-P (Krishnamurthy et

al. 2009; Bataille et al. 2012). Recently, Ssu72 was suggested to dephosphorylate Ser7-P, as depletion of Ssu72 and mutation in *SSU72* also result in elevated levels of Ser7-P at the 3' end of genes, and Ssu72 dephosphorylates both Ser5-P and Ser7-P on CTD substrates *in vitro* (Bataille et al. 2012; Zhang et al. 2012).

In searching for human Fcp1 homologs, three related proteins containing a region with homology to the Fcp1 phosphatase domain were identified (SCP1-3) and found to dephosphorylate the CTD (Yeo et al. 2003; Kamenski et al. 2004). SCPs preferentially dephosphorylate Ser5-P (Yeo et al. 2003; Zhang et al. 2006), and were shown to be recruited by the transcription factor REST/NRSF to suppress transcription of neural genes in non-neuronal cells by removing Ser5 phosphorylation on promoter-proximal RNAP II (Yeo et al. 2005). The SCPs, which are not present in lower eukaryotes, thus appear to play a role in tissue-specific transcriptional regulation.

Finally, yeast Rtr1 (Regulator of transcription 1) has also been reported to be a CTD phosphatase, functioning to dephosphorylate Ser5-P *in vitro* and *in vivo* (Mosley et al. 2009). Rtr1 associates with RNAP II, localizes to the 5' ends of genes, and *RTR1* deletion causes increased levels of Ser5-P and reduced RNAP II occupancy on a collection of genes (Mosley et al. 2009). Unlike Ssu72, Rtr1 is required only for Ser5-P, not Ser7-P, dephosphorylation (Kim et al. 2009). RPAP2, the human homologue of Rtr1, has also been reported to possess Ser5-P phosphatase activity (Egloff et al. 2012). However, the structure of Rtr1 was recently solved and did not reveal an apparent active site, and enzymatic assays with purified Rtr1 or RPAP2 and CTD substrates failed to detect activity (Xiang et al. 2012). Thus the role of Rtr1 in CTD dephosphorylation is currently unclear.

Transcription elongation, chromatin and the CTD

As alluded to above, the CTD undergoes dynamic changes in phosphorylation during transcription elongation. Very simply, the CTD is phosphorylated on Ser2 by CDK9 while Ser5-P is removed by Ser5 phosphatase(s), and towards the end of genes, the CTD is dephosphorylated by Ser5 and Ser2 phosphatases. RNAP II with a hypophosphorylated CTD, RNAP IIA, can then be recycled for another round of transcription. This view has been expanded by inclusion of two additional phosphorylation sites, Ser7 and Thr4, requiring CDK7 and CDK9, respectively. Chromatin immunoprecipitation (ChIP) experiments in yeast revealed a resemblance of Ser7-P distribution to Ser5-P on select genes (Glover-Cutter et al. 2009; Kim et al. 2009), suggesting that, in general, Ser5 and Ser7 are both phosphorylated at the beginning of transcription. Thr4-P functions in facilitating histone mRNA 3' end processing, and Thr4 phosphorylation requires CDK9, independent of its role in Ser2 phosphorylation (Hsin et al. 2011), suggesting that the Thr4-P pattern may be similar to Ser2-P, increasing towards the 3' end of genes. Indeed, a recent study in human cells found the genome-wide profiles of Thr4-P overlapped with, but shifted slightly 3' to, Ser2-P (Hintermair et al. 2012). Therefore, a simple model of CTD phosphorylation is that Ser5 and Ser7 are phosphorylated at the beginning of transcription, and Ser2 and Thr4 phosphorylation occurs during elongation (Figure 2).

The above general model for the dynamics of CTD phosphorylation was, for the most part, based on analysis of a limited number of genes. This has more recently been reanalyzed by several genome-wide ChIP studies in yeast (Kim et al. 2010; Mayer et al. 2010; Tietjen et al. 2010; Bataille et al. 2012; Zhang et al. 2012). Although somewhat different conclusions were reached, probably reflecting use of different analytical methods and/or antibodies, so far the general model holds for the majority of genes (Bataille et al. 2012). In brief, these studies showed, with exceptions, that Ser5-P is enriched at the 5' ends of genes, and peaks around transcription start sites (TSS). As RNAP II elongates toward the 3' end, Ser5-P is gradually removed, Ser2-P increases, beginning to saturate at ~600 nucleotides

downstream from TSS, regardless of gene length, and sharply decreases at 100 nucleotides downstream of the poly(A) addition site (Mayer et al. 2010). A crossover of Ser5-P to Ser2-P was observed at ~450 nucleotides downstream of the TSS. Therefore, longer genes have phosphorylation patterns more consistent with the model, and shorter genes have higher levels of Ser5-P and lower levels of Ser2-P. Overall, Ser7-P and Ser5-P patterns overlap to some extent, although profiles of Ser7-P vary, with some genes having discrete peaks at the 5' and/or 3' end (Kim et al. 2010). In one study, uniform levels of Ser7-P were observed along genes, although this may have reflected a technical limitation of the ChIP assay (Bataille et al. 2012). In any event, it is intriguing that long and short genes have different patterns of CTD phosphorylation at their 3' ends. Given the role of the CTD in processes such as mRNA 3' end formation and transcription termination, the existence of this pattern points to possible differences in the mechanisms underlying these processes on different genes dependent on length (see below).

Just as CTD phosphorylation patterns change along the length of transcribed genes, so do histone modifications in the chromatin of active genes. This raises the possibility that changes in the CTD phosphorylation array may contribute to differential patterns of histone modification as transcription proceeds, and indeed considerable evidence suggests that this is the case (reviewed by Hirose and Ohkuma 2007; Egloff and Murphy 2008; Spain and Govind 2011). For example, in yeast the histone methyltransferases Set1 and Set2 are recruited to actively transcribed genes in a CTD-dependent manner. Set1 is recruited to the 5' ends of genes by Ser5-P, together with the Paf1 complex, and methylates histone H3 on lys 4 (H3K4) (Krogan et al. 2003; Ng et al. 2003). Set1 establishes two distinct chromatin zones on genes, with tri-methylated H3K4 at promoter regions and di-methylated H3K4 downstream (Kim and Buratowski 2009). H3K4 di-methylation and CTD Ser5-P are then involved in recruiting the Set3 complex, a histone deacetylase that deactylates histones H3 and H4, leading to

reduced histone acetylation levels at the 5' ends of genes, which promotes the association of RNAP II (Kim and Buratowski 2009; Govind et al. 2010).

Set2 functions towards the 3' ends of genes. It binds to phosphorylated CTD with a preference for Ser2/5-P (Kizer et al. 2005), and methylates H3 on lys 36 (H3K36) (Li et al. 2003; Xiao et al. 2003). After H3K36 methylation, the histone deacetylase Rpd3S is recruited to deacetylate H3 and H4, which results in reduced transcription elongation efficiency and prevents cryptic transcription in coding region. The recruitment of Rpd3S is mediated by two subunits, Eaf3 and Rco (Carrozza et al. 2005; Joshi and Struhl 2005; Keogh et al. 2005). However, initial Rpd3S recruitment is independent of H3K36 methylation, and occurs through direct binding to Ser2, Ser5-phosphorylated CTD (Govind et al. 2010). Rpd3S then transfers to H3 by interaction of the Eaf3 and Rco subunits with methylated H3K36, which is required for activation of acetylation activity (Drouin et al. 2010; Govind et al. 2010). Although recruitment of Rpd3S to coding regions, and subsequent histone deacetylation to prevent cryptic transcription in the wake of the elongating RNAP II, is well established, Rpd3S is not found on all transcribed genes, suggesting the existence of other mechanisms to prevent aberrant transcription (Drouin et al. 2010).

H3K4 and H3K36 methylation patterns similar to those in yeast have been observed in higher eukaryotes (Bannister et al. 2005; Barski et al. 2007; Edmunds et al. 2008). Mammalian cells contain six Set1-like factors, of which Set1A and Set1B are responsible for the bulk H3K4 methylation (reviewed by Shilatifard 2012). A component of Set1A and Set1B, Wrd82, binds to the CTD phosphorylated on Ser5 (Lee and Skalnik 2008), which facilitates recruitment of these factors. Additionally, association of the mixed lineage leukemia (MLL) Set1-like factors, MLL1 and MLL2, with Ser5-P has been reported (Hughes et al. 2004; Milne et al. 2005). MLL1/MLL2 are required for proper *Hox* gene expression, and translocation mutations of MLL are associated with the pathogenesis of leukemia (reviewed in

Shilatifard 2012). A human homolog of yeast Set2, Huntington interacting protein (HYPB, also known as Setd2), interacts with Ser2/5 phosphorylated CTD through its C-terminal region (Li et al. 2005) and methylates H3K36 (Sun et al. 2005). The association of HYPB with phosphorylated CTD is enhanced by an elongation factor, Iws1, that is cotranscriptionally recruited by the histone chaperon and transcription elongation factor, Spt6 (Yoh et al. 2007; Yoh et al. 2008). Spt6 directly binds to the phosphorylated CTD via its SH2 domain with a preference for Ser2-P (Yoh et al. 2007; Sun et al. 2010). Interestingly, Spt6 has also been shown to be involved in facilitating splicing and nuclear mRNA export (Yoh et al. 2007; Yoh et al. 2008).

RNA processing and the CTD

The CTD is now known to function in essentially all the RNA processing reactions involved in maturation of very likely all transcripts produced by RNAP II. This includes not only mRNAs, but also snRNAs, miRNAs and other ncRNAs. The CTD thus provides the basis for the coupling between transcription and RNA processing that is important for ensuring the efficiency and accuracy of the complex processing reactions required for production of functional RNAs. Here we will concentrate on the role of the CTD in mRNA processing events, both in the interests of space and because these are illustrative of the function of the CTD more generally. Reviews dealing with the function of the CTD in other aspects of RNA processing and metabolism have been published recently (Munoz et al. 2010; Pawlicki and Steitz 2010).

Most of the earliest studies linking the CTD to RNA processing were done in mammalian systems. Many of these used a “pseudo-genetic” analysis that takes advantage of the RNAP II inhibitor α-amanitin. In this system, first introduced by Corden and colleagues (Gerber et al. 1995), α-amanitin, which binds Rpb1, is added to the cell culture to inhibit endogenous RNAP II after a plasmid encoding an exogenous,

α-amanitin-resistant Rpb1 is introduced and expressed in the cells. Therefore the only functional RNAP II will contain an α-amanitin-resistant Rpb1 subunit, and the effects of mutations, specifically of the CTD, can be determined. Although not without difficulties (see below), this system has provided a wealth of information about the function of the CTD and of CTD modifications. For example, Bentley and colleagues used this method to provide the first direct evidence that the CTD indeed plays a role in mRNA processing. They showed, unexpectedly, that deletion of the CTD had no effects on transcription of a reporter plasmid, but was required for efficient 5' capping, splicing, 3' processing (McCracken et al. 1997a, b). While initial studies suggested that the C-terminal half of the CTD, rich in nonconsensus heptads, but not the more conserved N-terminal half, could support all three processing reactions (Fong and Bentley 2001), this was subsequently shown to reflect the presence of the 10 amino acid nonconsensus motif following heptad 52: When the N-terminal half was fused with this motif, it now fully supported processing (Fong et al. 2003). Indeed, another study indicated that only the number of heptads is important: Twenty two repeats, from either the conserved or divergent half of the CTD, were found to be sufficient to stimulate splicing and 3' end cleavage (Rosonina and Blencowe 2004). As mentioned above, the 10 amino acid motif functions in Rpb1 stability (Chapman et al. 2004; 2005), likely explaining its requirement in processing. Intriguingly, the c-abl tyrosine kinase interacts with the CTD through this motif (Baskaran et al. 1999), although there is no evidence indicating the involvement of abl kinase, or Tyr1 phosphorylation, in Rpb1 protein stability. More recently, the α-amanitin system has been used to determine the role of specific residues, and their modification, in CTD function. For example, it was used to provide evidence that Ser7 phosphorylation functions in the expression and 3' formation of snRNAs (Egloff et al. 2007). Below, we discuss the role of the CTD, and modifications, in capping splicing and 3' end formation, and review the factors (listed in Table 1) involved in RNA processing that bind the CTD (Figure 3 and 4).

Capping

5’ capping, the first step of mRNA processing, occurs very early in transcription, essentially as soon as the newly synthesized RNA is extruded from RNAP II (Coppola et al. 1983; Jove and Manley 1984). Only RNAP II products are capped and this specificity is due in large part to the CTD, which, when phosphorylated, recruits capping enzymes (Cho et al. 1997; McCracken et al. 1997a). In metazoans, a bifunctional capping enzyme with RNA triphosphatase and RNA guanylyltransferase activities binds to the phosphorylated CTD through the guanylyltransferase domain. This domain has two binding sites for phosphorylated CTD, one specific for the Ser2-phosphorylated CTD and the other, an allosteric activator site, for Ser5-phosphorylated CTD (Ho and Shuman 1999), which stimulates formation of enzyme-GMP intermediate (Ho and Shuman 1999; Ghosh et al. 2011). The interaction between guanylyltransferase and Ser5-P also occurs in yeast (*C. albicans*) (Fabrega et al. 2003). Interestingly, the budding yeast capping enzyme contains another binding site, for the Rpb1 “foot domain,” which works synergistically with the Ser5-P binding site to enhance interaction with RNAP II (Suh et al. 2010). Remarkably, it was recently shown that recruitment of capping enzymes is the sole essential function of Ser5, at least in fission yeast. As mentioned above, substitution of all Ser 5 residues with ala is lethal in *S. pombe*, but viability can be restored simply by tethering capping enzymes to the CTD (Schwer and Shuman 2011).

Splicing

Nearly all mammalian pre-mRNAs contain introns. In humans, the average size of an intron is about 3 kb, but many are considerably larger (Lander et al. 2001). Furthermore, nearly all human genes produce transcripts in which one or more intron is subject to alternative splicing (reviewed by Chen and Manley 2009). Given this complexity, coupled with the remarkable complexity of the spliceosome itself, it is perhaps not surprising that splicing and transcription are coupled, as this provides a mechanism to

enhance the fidelity and efficiency of splicing, as well as its regulation. And, while other factors function in this coupling (e.g., Luco et al. 2011), it is also not surprising that the CTD plays a central role.

The CTD interacts with a number of proteins that function in splicing. Early on, phosphorylated RNAP II (RNAP IIO) was found to associate with splicing complexes (Mortillaro et al. 1996). *In vitro* biochemical studies indicated that RNAP II0, but not RNAP IIA, can enhance splicing of an exogenously supplied substrate by influencing an early step of spliceosome assembly (Hirose et al. 1999). Splicing of pre-mRNAs transcribed *in vitro* by RNAP II is accelerated compared to those transcribed by T7 RNA polymerase (Ghosh and Garcia-Blanco 2000; Das et al. 2006), suggesting that the CTD can enhance the rate of splicing, likely by facilitating recruitment of splicing factors, and artificial tethering of the phosphorylated CTD to a pre-mRNA can also accelerate splicing *in vitro* (Millhouse and Manley 2005). Additionally, splicing factor SRSF2 (SC35) was shown to colocalize with nascent transcripts and associate with RNAP IIO in HeLa cells, but this colocalization and association was lost when the CTD was truncated (Misteli and Spector 1999). Additional splicing factors, including U1 snRNP-associated protein Prp40, involved in 5' splice site recognition (Morris and Greenleaf 2000; Phatnani et al. 2004), PSF, which binds sequences downstream from branch point (Emili et al. 2002; Rosonina et al. 2005), and U2AF, which helps define the 3'splice site (David et al. 2011), all appear to bind directly to the CTD (Figure 3B). Together these factors could help recruit the splicing machinery to both ends of the intron during transcription. Indeed, the U2AF interaction with the CTD recruits the Prp19 complex (PRP19C), an essential splicing factor required for activation of the spliceosome (David et al. 2011). An intriguing idea is that recruitment of splicing factors to nascent transcripts by the CTD is especially important for splicing of introns with weak splicing signals (reviewed by David and Manley 2011).

Recruitment of splicing factors and other RNA binding proteins to nascent transcripts can facilitate gene expression in additional ways. For example, in yeast the co-transcriptional recruitment of

PRP19C leads to corecruitment of the TREX complex, which functions in mRNA transport, thereby preparing a nascent transcript for transport to the cytoplasm as it is synthesized (Chanarat et al. 2011). These authors also provided evidence that PRP19C enhances RNAP II elongation, which provides support for the view that splicing also communicates back to modulate transcription (reviewed by Manley 2002). For example, the presence of splice sites on nascent transcripts can stimulate transcription, as reduced transcription can be seen when splice sites are impaired (Furger et al. 2002; Damgaard et al. 2008; Eberle et al. 2010); splicing factors, such as SKIP, Npl3 and SRSF2, can enhance transcriptional elongation (Bres et al. 2005; Dermody et al. 2008; Lin et al. 2008); and incorrect spliceosome recruitment and assembly can interfere with transcription termination (Martins et al. 2011). Finally, while histone modifications can help recruit splicing factors to active genes (Luco et al. 2010; reviewed by Hnilicova and Stanek 2011), splicing can also influence chromatin structure, for example by facilitating recruitment of HYPB and thereby increasing H3K36 methylation (de Almeida et al. 2011).

Cotranscriptional recruitment of RNA binding proteins has an additional important effect in cells, which is to prevent genomic instability and DNA rearrangements during transcription (reviewed by Li and Manley 2006; Aguilera and Garcia-Muse 2012). If nascent transcripts are not properly packaged or processed, they have the potential, especially in G-rich regions, to rehybridize with the template DNA strand, creating an R-loop structure in which the non-template strand is single stranded. Such structures can be the targets of DNA double strand breaks and rearrangements. The CTD plays an important role in this process, as prevention of R loops by SRSF1 in an *in vitro* reconstituted system was found to be dependent on the CTD (Li and Manley 2005).

3’ end processing

The CTD also plays an important role in 3' end processing of RNAP II-produced transcripts. This has been most extensively studied in the case of polyadenylated mRNAs. Polyadenylation, a relatively simple two-step reaction consisting of an endonucleolytic cleavage followed by poly(A) tail synthesis, involves a remarkably complex set of protein factors (Shi et al. 2009), and the CTD functions in recruitment and/or stabilization of this complex on the pre-mRNA. The studies mentioned above using the α-amanitin system to provide evidence that 3' end processing is impaired when Rpb1 is truncated also revealed that the CTD associates with two polyadenylation factors, CPSF and CstF (McCracken et al. 1997b). Hirose and Manley (1998) showed that the CTD is required for efficient 3' end processing *in vitro* in the absence of transcription, and that this function can be fulfilled by the CTD alone. Subsequent experiments revealed that 26 all-consensus repeats were required for reconstitution of CTD activity *in vitro* (Ryan et al. 2002), consistent with results *in vivo*. The CTD requirement is conserved in yeast, as CTD deletion was shown to cause defects in 3' processing (Licatalosi et al. 2002).

A number of 3' processing factors interact with the CTD, consistent with it playing a role as a scaffolding factor (Figure 4). For example, CstF-50, a component of CstF, and Yhh1, the yeast counterpart of the mammalian AAUAAA-binding CPSF-160, both interact with the CTD (Fong and Bentley 2001; Dichtl et al. 2002b), providing an explanation for the initial observations of McCracken et al. (1997b). Pcf11, which contains an N-terminal CTD interaction domain (CID) characteristic of several CTD binding proteins (reviewed by Corden and Patturajan 1997), binds the CTD in a manner enhanced by Ser2 phosphorylation (Barilla et al. 2001; Licatalosi et al. 2002; Meinhart and Cramer 2004). Indeed, Ser2-P plays an important role in facilitating polyadenylation. For example, impairment of Ser2 phosphorylation, by deletion of *CTK1* in yeast (Skaar and Greenleaf 2002; Ahn et al. 2004) or flavopiridol in metazoan cells (Ni et al. 2004), impairs recruitment of processing factors at the 3' ends of genes and subsequent polyadenylation. However, at least in yeast accumulation of these factors also requires an

intact polyadenylation signal, indicating not surprisingly that interactions with the pre-mRNA are also important (Kim et al. 2004a). Extending these findings, genome-wide ChIP experiments showed that peaks of 3' processing factors followed Ser2-P peaks, consistent with the view that both the polyA site in the RNA and CTD Ser2-P contribute to the recruitment/assembly of the polyadenylation complex (Kim et al. 2010; Mayer et al. 2010; Mayer et al. 2012b). Together, these experiments point to an important role for the CTD, and Ser2-P, in polyadenylation of mRNA precursors.

The involvement of Ser2-P and the observed accumulation of processing factors at the 3' ends of genes are consistent with the simple view that polyadenylation factors are recruited to genes near their site of action. However, considerable evidence also exists that polyadenylation factors are in fact recruited to genes at the promoter (reviewed by Calvo and Manley 2003; see also Glover-Cutter et al. 2008; Nagaike et al. 2011 and refs therein). Factors involved in this recruitment include the GTFs TFIID (Dantonel et al. 1997) and TFIIB (Sun and Hampsey 1996; Wang et al. 2010), the PAF complex (Rozenblatt-Rosen et al. 2009), as well as transcriptional activators (reviewed by Nagaike and Manley 2011). Indeed, as with splicing, these interactions may affect transcription, as polyA site mutations reduce transcription and reduce the levels of TFIIB and TFIID at promoters (Mapendano et al. 2010). While some of these interactions likely reflect gene looping (O'Sullivan et al. 2004; reviewed by Hampsey et al. 2011), they highlight the importance of coupling transcription and polyadenylation via multiple mechanisms that extend beyond the CTD.

The 3' ends of several types of RNAP II-transcribed RNAs, including small nuclear RNA (snRNA) and histone mRNA, are not polyadenylated, but also require the CTD. The CTD plays a role in snRNA 3' end formation, and inhibition of Ser2 phosphorylation was shown to impair 3' processing without affecting transcription (Jacobs et al. 2003; Medlin et al. 2003). Proper snRNA 3' end formation is dependent on both the promoter and a 3'-box, located just downstream of the snRNA-encoding region

(reviewed by Egloff et al. 2008). The 3'-box is recognized by a multisubunit RNA 3' end processing complex, the Integrator, and snRNAs are cleaved by the Int11 subunit (Baillat et al. 2005). The Integrator is recruited by interaction with the CTD phosphorylated on Ser2 and Ser7 (Egloff et al. 2010), consistent with earlier studies showing that mutation of all Ser7 residues to ala impairs 3' end processing of snRNAs, but not of polyadenylated mRNAs (Egloff et al. 2007). Recently, it was shown that RPAP2, which functions in Ser5-P dephosphorylation (see above), binds to the CTD with Ser7-P and facilitates Integrator recruitment (Egloff et al. 2012). These results support the view that Ser5-P dephosphorylation is important for both snRNA and mRNA (Xiang et al. 2010) 3' end formation. It is also noteworthy that methylation of a specific arg residue in a non-consensus heptad appears to dampen expression of certain snRNAs and sno (small nucleolar) RNAs, as mutation of this residue, to ala, resulted in enhanced accumulation of these RNAs in the α-amanitin assay (Sims et al. 2011). Although the mechanism remains to be determined, an intriguing possibility is that this modification in some way interferes with the positive effects of Ser2 and Ser7 phosphorylation.

Metazoan replication-dependent histone mRNAs are the only protein-coding transcripts that are not polyadenylated. The 3' end of histone mRNA contains two cis-elements, a stem-loop structure that is recognized by stem-loop binding protein (SLBP) and a downstream sequence bound by U7 snRNP. Together, SLBP and U7 snRNP recruit a cleavage complex, containing polyadenylation factors CPSF73 (the endonuclease), CPSF100 and Symplekin, to process histone mRNA 3' ends (reviewed by Marzluff et al. 2008). As with other RNAP II transcripts, the CTD also participates in histone 3' end processing. This was first suggested by the finding that CDK9 is required for proper histone mRNA 3' processing and SLBP recruitment (Pirngruber et al. 2009). Recently, using the DT40 system (see above), evidence was provided that Thr4-P is required for histone mRNA 3' end processing (Hsin et al. 2011). Specifically, cells expressing Rpb1 with a mutant CTD in which all Thr4 residues were replaced by val were found to be

defective in histone mRNA 3' end processing and recruitment of SLBP and CPSF100, while 3' end formation of polyadenylated RNAs, and overall transcription rates, were essentially unaffected. Inhibition of CDK9 prevented Thr4 phosphorylation, and inhibited histone 3' processing, consistent with the results of Pirngruber et al. (2009). Hintermair et al. (2012) analyzed a related Thr4 mutant Rpb1 using the α-amanitin system. These authors did not monitor 3' processing, but detected by RNAP II ChIP a global defect in elongation, characterized by promoter-proximal accumulation of initiated RNAP II. While Hsin et al. (2011) also observed some increase in promoter-proximal RNAP II on several genes, the effects were likely smaller than detected by Hintermair et al., as measurements of transcription rate and polyadenylated mRNA levels revealed at most modest effects.

The basis for the possible differences in the effects of Thr4 mutation is unclear, but one possibility may reflect features of the α-amanitin system. Although, as illustrated in the above discussion, a great deal has been learned about the functions of the CTD using this assay, there are several potential drawbacks. One of these is that transiently transfected reporter genes are often used to observe phenotypes of mutant CTDs because endogenous RNAs are stabilized by α-amanitin (Meininghaus et al. 2000; Chapman et al. 2004). Such reporter genes are not properly packaged into chromatin, which may affect results. Another drawback reflects the toxicity of α-amanitin. Thus, while stably transformed cells expressing α-amanitin resistant Rpb1 avoid the problems of transient transfection, such cells initially grow slowly with reduced viability in the presence of α-amanitin (Meininghaus et al. 2000; Chapman et al. 2004; 2005), indicating that they must undergo certain changes to survive. Indeed, 24 hour α-amanitin treatment was recently shown to bring about accelerated degradation of several proteins, including the transcription elongation factor DSIF, and this may complicate interpretation of experimental results (Tsao et al. 2012).

The CTD and transcription termination

Stopping RNAP II is much more difficult than any other RNA polymerase. This undoubtedly reflects not only the need to transcribe successfully immensely long genes, which can extend over 1 Mbp in vertebrates, but also the diversity of sequence landscapes presented by the thousands of genes that RNAP II must negotiate, which can pose a variety of challenges to the elongating polymerase. Consistent with this, the mechanisms leading to RNAP II termination are complex and remain incompletely understood. While considerable progress has been made in deciphering these mechanisms in recent years (reviewed by Richard and Manley 2009; Kuehner et al. 2011), here we will discuss RNAP II termination from the perspective of the CTD and its role in the process (Figure 5 and Table 1).

One of the earliest insights into the mechanism of RNAP II termination was the discovery in mammalian systems that a functional polyadenylation signal is required for subsequent termination (reviewed by Proudfoot 1989). This suggested that 3' processing of the mRNA precursor is required for termination, and subsequent experiments, principally in yeast and employing transcriptional run on and/or RNAP II ChIP to measure termination, revealed that a number of 3' cleavage factors are also necessary for termination (Birse et al. 1998; Dichtl et al. 2002b; Ganem et al. 2003; Nedea et al. 2003; Kim et al. 2010; Zhang et al. 2012). Given the requirement of the CTD for the 3' processing, it is not surprising that the CTD is also required for termination (McCracken et al. 1997b), with the N-terminal half being sufficient (Park et al. 2004), as determined using the α-amanitin system. One CTD binding cleavage factor, Pcf11, may play a particularly important role in termination. A mutant yeast Pcf11 that retains 3' cleavage activity but is defective in CTD binding was found to be defective in termination (Sadowski et al. 2003). *In vitro* studies suggested that both yeast and fly Pcf11 can disassociate RNAP II and the transcribed RNA from the DNA by bridging the CTD to RNA (Zhang et al. 2005; Zhang and Gilmour 2006). However, the dissociation activity is not specific to Pcf11 (Zhang et al. 2004), and Pcf11

cleavage activity, and not CTD binding ability, was subsequently found to be important for termination of protein coding genes (Kim et al. 2006), leaving the precise role of Pcf11 in termination unclear.

The discovery that 3' cleavage is necessary for termination led to a model in which a 5' to 3' exoribonuclease degrades the downstream nascent RNA, and in some way signals termination to the elongating RNAP II (Connelly and Manley 1988). Eventually, confirmation of this model came from the discovery that termination, indeed, requires such a nuclease: Rat1 in yeast (Kim et al. 2004b) and Xrn2 in humans (West et al. 2004). The CTD plays a role in Rat1/Xrn2 recruitment. Pcf11 is required in yeast for Rat1 association with active genes (Luo et al. 2006), and in HeLa cells for efficient degradation of the downstream RNA (West and Proudfoot 2008). In line with this, the genome-wide profile of yeast Pcf11 closely overlaps that of Rat1 (Kim et al. 2010). Rat1 recruitment to the CTD also requires the associated protein Rtt103, which specifically interacts with CTD Ser2-P (Kim et al. 2004b). Rtt103 and Pcf11 both achieve high-affinity for the CTD phosphorylated on Ser2 by cooperatively binding to neighboring Ser2-P residues (Lunde et al. 2010). In humans, Xrn2 associates with p54/PSF (Kaneko et al. 2007), a multifunctional protein dimer that binds to the CTD (Emili et al. 2002; Rosonina et al. 2005) and was identified as a component of the 3' processing complex (Shi et al. 2009). p54/PSF may facilitate termination by recruiting Xrn2, as accumulation of the downstream cleaved RNA and termination defects were observed *in vitro* and *in vivo* after p54 was depleted from HeLa cells (Kaneko et al. 2007).

The RNA-DNA helicase Sen1 -- well established to function in termination of snoRNAs in yeast (see below) -- also appears to function in the termination of some mRNA genes. Sen1 binds to the CTD phosphorylated on Ser2 (Ursic et al. 2004; Chinchilla et al. 2012), which likely facilitates its recruitment to multiple coding as well as non-coding genes, where it tends to accumulate towards the 3' end (Chinchilla et al. 2012). However, in cells expressing a catalytically inactive mutant Sen1, only a small set of protein-coding genes, usually short in length, are defective in termination, as shown by a genome-

wide Rpb1 ChIP analysis (Steinmetz et al. 2006). The human homologue of Sen1, Senataxin, has also been implicated in termination (Suraweera et al. 2009), perhaps by resolving RNA-DNA hybrids (R loops) formed at pause sites downstream of the polyadenylation signal, allowing degradation of the cleaved RNA by Xrn2 (Skourti-Stathaki et al. 2011).

Termination of transcription on genes encoding small noncoding RNAs shares some features in common with mRNA-encoding genes, but also displays differences. The requirement of Pcf11 in termination extends to yeast snoRNA genes. However, here the CTD binding activity of Pcf11 is necessary for termination, while the 3' cleavage function is dispensable (Kim et al. 2006). Termination on snoRNA genes requires a protein complex, the Nrd1 complex (Nrd1c), which consists of Nrd1, Nab3, and Sen1 (Steinmetz and Brow 1996; Conrad et al. 2000; Steinmetz et al. 2001; Kim et al. 2006). Nrd1 and Nab3 are RNA binding proteins, and their interaction with nascent RNAs helps recruit Nrd1c to target genes (Steinmetz and Brow 1998; Conrad et al. 2000; Steinmetz et al. 2001). However, Nrd1, like Sen1, binds the CTD, except displaying a preference for Ser5-P (Conrad et al. 2000; Steinmetz et al. 2001; Vasiljeva et al. 2008). This serves to help target Nrd1c to short genes, such as snoRNA genes, and genes encoding CUTs. CUTs (cryptic unstable transcripts) are short non-coding RNAP II-produced non-polyadenylated transcripts found in yeast that are rapidly degraded by the nuclear exosome (reviewed by Colin et al. 2011). Consistent with the requirement of Nrd1c, CUT termination is critically dependent on CTD phosphorylation status, requiring high levels of Ser5-P and low Ser2-P (Gudipati et al. 2008). However, the Ser5-P must be removed, as Ssu72 and Ess1 are required for snoRNA/CUT termination (Kim et al. 2006; Singh et al. 2009; Zhang et al. 2012). Precisely how Nrd1c brings about termination is not well understood, and may vary among different genes (reviewed by Richard and Manley 2009; Kuehner et al. 2011).

Nrd1c also associates with mRNA-encoding genes (Nedea et al. 2003; Kim et al. 2006). Genome-wide ChIP of Nrd1 revealed frequent colocalization with Pcf11 and, unexpectedly, Ser7-P at the 3' ends of genes (Kim et al. 2010). However, termination of a collection of protein-coding genes is not affected by mutations in Nrd1 (Kim et al. 2006). Perhaps the function of Nrd1c on such genes is to provide an alternative pathway to terminate aberrant transcription (Rondon et al. 2009) and/or a quality control point to degrade erroneous transcripts (Honorine et al. 2011).

An interesting question is whether a Nrd1c-like pathway for termination exists in higher eukaryotes. Although as mentioned above, a homologue of Sen1, Senataxin, exists in human cells, functional homologues of Nrd1 and Nab3 have not been identified, and there is no evidence for the existence of a Nrd1c-like complex. SCAF8, which shares sequence similarity with Nrd1, was in fact the first CID domain protein identified (Yuryev et al. 1996; Patturajan et al. 1998). While SCAF8 CTD binding has been extensively characterized (Becker et al. 2008), its function is unknown, but does not appear to involve termination. The absence of Nrd1c in metazoans is consistent with the facts that snoRNAs tend to be produced by a different mechanism (i.e., cleavage from introns; Richard and Kiss 2006) and, while RNAP II-transcribed non-coding RNAs are ubiquitous (reviewed by Jacquier 2009), it is not clear if CUT RNAs are indeed produced in higher eukaryotes. In the future, it will be interesting to determine if additional mechanisms of termination exist in metazoans, and if so how the CTD might participate.

Perspectives

A great deal has been learned over the course of the past two decades about the structure and function of the CTD, the nature and role of CTD modifications in gene expression, and how the CTD mediates the interplay between transcription and RNA processing. However, the goal of completely understanding the functions and importance of this deceptively simple domain remains a challenge. We

mention here just a few examples. What is the function of Tyr1 phosphorylation, and how do other modifications such as methylation, acetylation, and ubiquitination affect CTD function and gene expression? Intriguingly, evidence that Tyr1-P occurs in budding yeast and may play a general role in facilitating elongation, by blocking premature 3' end formation, was recently presented (Mayer et al. 2012a). These findings suggest an important, perhaps general, role for Tyr1-P in elongation. This is perhaps surprising given that tyr phosphorylation is rare in yeast (e.g., Gnad et al. 2009), and further characterization of Tyr1-P, in yeast and mammalian systems, will therefore be informative. Although a general model outlines changes in CTD phosphorylation along the length of transcribed genes, how common is regulated CTD modification in control of specific genes, and how is this achieved? A well-documented example is recruitment of P-TEFb by the oncogenic transcription factor c-Myc (Eberhardy and Farnham 2002; Kanazawa et al. 2003; Rahl et al. 2010), but is this a common mechanism and are other CTD modifiers targets for gene-specific recruitment? How many more CTD modifiers remain to be discovered, and might they work in gene- or cell-specific ways? For example, the Cdc14 cell-cycle phosphatase (Clemente-Blanco et al. 2011) and the bromodomain transcriptional regulator Brd4, apparently functioning as an atypical protein kinase (Devaiah et al. 2012), have both been reported to target the CTD. Future studies will continue to unravel the complex workings of the CTD, and, we suspect, consolidate its role as the conductor of the gene expression symphony.

Acknowledgements

We thank Manley lab members Emanuel Rosonina, Patricia Richard, Dafne Campigli Di Giammartino, and Kehui Xiang for discussion and critical reading of this manuscript. Work from our lab was supported by grants from the NIH.

References

Aguilera A, Garcia-Muse T. 2012. R loops: from transcription byproducts to threats to genome stability. *Mol Cell* **46**: 115-124.

Ahn SH, Kim M, Buratowski S. 2004. Phosphorylation of serine 2 within the RNA polymerase II C-terminal domain couples transcription and 3' end processing. *Mol Cell* **13**: 67-76.

Akhtar MS, Heidemann M, Tietjen JR, Zhang DW, Chapman RD, Eick D, Ansari AZ. 2009. TFIIH kinase places bivalent marks on the carboxy-terminal domain of RNA polymerase II. *Mol Cell* **34**: 387-393.

Akoulitchev S, CHuikov S, Reinberg D. 2000. TFIIH is negatively regulated by cdk8-containing mediator complexes. *Nature* **407**: 102-106.

Akoulitchev S, Makela TP, Weinberg RA, Reinberg D. 1995. Requirement for TFIIH Kinase Activity in Transcription by RNA Polymerase II. *Nature* **377**: 557-560.

Archambault J, Chambers RS, Kobor MS, Ho Y, Cartier M, Bolotin D, Andrews B, Kane CM, Greenblatt J. 1997. An essential component of a C-terminal domain phosphatase that interacts with transcription factor IIF in Saccharomyces cerevisiae. *Proc Natl Acad Sci U S A* **94**: 14300-14305.

Baillat D, Hakimi MA, Naar AM, 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.

Bannister AJ, Schneider R, Myers FA, Thorne AW, Crane-Robinson C, Kouzarides T. 2005. Spatial distribution of di- and tri-methyl lysine 36 of histone H3 at active genes. *J Biol Chem* **280**: 17732-17736.

Barilla D, Lee BA, Proudfoot NJ. 2001. Cleavage/polyadenylation factor IA associates with the carboxyl-terminal domain of RNA polymerase II in Saccharomyces cerevisiae. *Proc Natl Acad Sci U S A* **98**: 445-450.

Barski A, Cuddapah S, Cui K, Roh TY, Schones DE, Wang Z, Wei G, Chepelev I, Zhao K. 2007. High-resolution profiling of histone methylations in the human genome. *Cell* **129**: 823-837.

Bartkowiak B, Liu P, Phatnani HP, Fuda NJ, Cooper JJ, Price DH, Adelman K, Lis JT, Greenleaf AL. 2010. CDK12 is a transcription elongation-associated CTD kinase, the metazoan ortholog of yeast Ctk1. *Genes Dev* **24**: 2303-2316.

Bartolomei MS, Halden NF, Cullen CR, Corden JL. 1988. Genetic analysis of the repetitive carboxyl-terminal domain of the largest subunit of mouse RNA polymerase II. *Mol Cell Biol* **8**: 330-339.

Baskaran R, Chiang GG, Mysliwiec T, Kruch GD, Wang JYJ. 1997. Tyrosine phosphorylation of RNA polymerase II carboxyl-terminal domain by the Abl-related gene product. *J Biol Chem* **272**: 18905-18909.

Baskaran R, Dahmus ME, Wang JYJ. 1993. Tyrosine phosphorylation of mammalian RNA polymerase II carboxyl-terminal domain. *Proc Natl Acad Sci U S A* **90**: 11167-11171.

Baskaran R, Escobar SR, Wang JYJ. 1999. 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 & Differentiation* **10**: 387-396.

Bataille AR, Jeronimo C, Jacques PE, Laramee L, Fortin ME, Forest A, Bergeron M, Hanes SD, Robert F. 2012. A universal RNA polymerase II CTD cycle is orchestrated by complex interplays between kinase, phosphatase, and isomerase enzymes along genes. *Mol Cell* **45**: 158-170.

Baumli S, Lolli G, Lowe ED, Troiani S, Rusconi L, Bullock AN, Debreczeni JE, Knapp S, Johnson LN. 2008. The structure of P-TEFb (CDK9/cyclin T1), its complex with flavopiridol and regulation by phosphorylation. *The Embo Journal* **27**: 1907-1918.

Becker R, Loll B, Meinhart A. 2008. Snapshots of the RNA processing factor SCAF8 bound to different phosphorylated forms of the carboxyl-terminal domain of RNA polymerase II. *J Biol Chem* **283**: 22659-22669.

Birse CE, Minvielle-Sebastia L, Lee BA, Keller W, Proudfoot NJ. 1998. Coupling Termination of Transcription to Messenger RNA Maturation in Yeast. *Science* **280**: 298-301.

Blazek D, Kohoutek J, Bartholomeeusen K, Johansen E, Hulinkova P, Luo Z, Cimermancic P, Ule J, Peterlin BM. 2011. The Cyclin K/Cdk12 complex maintains genomic stability via regulation of expression of DNA damage response genes. *Genes Dev* **25**: 2158-2172.

Bres V, Gomes N, Pickle L, Jones KA. 2005. A human splicing factor, SKIP, associates with P-TEFb and enhances transcription elongation by HIV-1 Tat. *Genes Dev* **19**: 1211-1226.

Buratowski S. 2009. Progression through the RNA polymerase II CTD cycle. *Mol Cell* **36**: 541-546.

Calvo O, Manley JL. 2003. Strange bedfellows: polyadenylation factors at the promoter. *Genes Dev* **17**: 1321-1327.

Calvo O, Manley JL. 2005. The transcriptional coactivator PC4/Sub1 has multiple functions in RNA polymerase II transcription. *The Embo Journal* **24**: 1009-1020.

Carrozza MJ, Li B, Florens L, Suganuma T, Swanson SK, Lee KK, Shia WJ, Anderson S, Yates J, Washburn MP et al. 2005. Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. *Cell* **123**: 581-592.

Chanarat S, Seizl M, Strasser K. 2011. The Prp19 complex is a novel transcription elongation factor required for TREX occupancy at transcribed genes. *Genes Dev* **25**: 1147-1158.

Chapman RD, Conrad M, Eick D. 2005. Role of the mammalian RNA polymerase II C-terminal domain (CTD) nonconsensus repeats in CTD stability and cell proliferation. *Mol Cell Biol* **25**: 7665-7674.

Chapman RD, Heidemann M, Hintermair C, Eick D. 2008. Molecular evolution of the RNA polymerase II CTD. *Trends Genet* **24**: 289-296.

Chapman RD, Palancade B, Lang A, Bendaude O, Eick D. 2004. The last CTD repeat of the mammalian RNA polymerase II large subunit is important for its stability. *Nucleic Acids Res* **32**: 35-44.

Chen M, Manley JL. 2009. Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. *Nat Rev Mol Cell Biol* **10**: 741-754.

Cheng B, Price DH. 2007. Properties of RNA polymerase II elongation complexes before and after the P-TEFb-mediated transition into productive elongation. *J Biol Chem* **282**: 21901-21912.

Chesnut JD, Stephens JH, Dahmus ME. 1992. The interaction of RNA polymerase II with the adenovirus-2 major late promoter is precluded by phosphorylation of the C-terminal domain of subunit IIa. *J Biol Chem* **267**: 10500-10506.

Chi Y, Huddleston MJ, Zhang XL, Young RA, Annan RS, Carr SA, Deshaies RJ. 2001. Negative regulation of Gcn4 and Msn2 transcription factors by Srb10 cyclin-dependent kinase. *Genes Dev* **15**: 1078-1092.

Chiba K, Yamamoto J, Yamaguchi Y, Handa H. 2010. Promoter-proximal pausing and its release: molecular mechanisms and physiological functions. *Exp Cell Res* **316**: 2723-2730.

Chinchilla K, Rodriguez-Molina JB, Ursic D, Finkel JS, Ansari AZ, Culbertson MR. 2012. Interactions of Sen1, Nrd1, and Nab3 with multiple phosphorylated forms of the Rpb1 C-terminal domain in Saccharomyces cerevisiae. *Eukaryot Cell* **11**: 417-429.

Cho EJ, Kobor MS, Kim M, Greenblatt J, Buratowski S. 2001. Opposing effects of Ctk1 kinase and Fcp1 phosphatase at Ser 2 of the RNA polymerase II C-terminal domain. *Genes Dev* **15**: 3319-3329.

Cho EJ, Takagi T, Moore CR, Buratowski S. 1997. mRNA capping enzyme is recruited to the transcription complex by phosphorylation of the RNA polymerase II carboxy-terminal domain. *Genes Dev* **11**: 3319-3326.

Cho H, Kim TK, Mancebo H, Lane WS, Flores O, Reinberg D. 1999. A protein phosphatase functions to recycle RNA polymerase II. *Genes Dev* **13**: 1540-1552.

Cisek LJ, Corden JL. 1989. Phosphorylation of RNA polymerase by the murine homologue of the cell-cycle control protein cdc2. *Nature* **339**: 679-684.

Clemente-Blanco A, Sen N, Mayan-Santos M, Sacristan MP, Graham B, Jarmuz A, Giess A, Webb E, Game L, Eick D et al. 2011. Cdc14 phosphatase promotes segregation of telomeres through repression of RNA polymerase II transcription. *Nat Cell Biol* **13**: 1450-1456.

Colin J, Libri D, Porrua O. 2011. Cryptic transcription and early termination in the control of gene expression. *Genet Res Int* **2011**: 653494.

Connelly S, Manley JL. 1988. A functional mRNA polyadenylation signal is required for transcription termination by RNA polymerase II. *Genes Dev* **2**: 440-452.

Conrad NK, Wilson SM, Steinmetz EJ, Patturajan M, Brow DA, Swanson MS, Corden JL. 2000. A yeast heterogeneous nuclear ribonucleoprotein complex associated with RNA polymerase II. *Genetics* **154**: 557-571.

Coppola JA, Field AS, Luse DS. 1983. Promoter-proximal pausing by RNA polymerase II in vitro: transcripts shorter than 20 nucleotides are not capped. *Proc Natl Acad Sci U S A* **80**: 1251-1255.

Corden JL, Cadena DL, Ahearn JM, Jr., Dahmus ME. 1985. A unique structure at the carboxyl terminus of the largest subunit of eukaryotic RNA polymerase II. *Proc Natl Acad Sci U S A* **82**: 7934-7938.

Corden JL, Patturajan M. 1997. A CTD function linking transcription to splicing. *Trends in Biochemical Sciences* **22**: 413-416.

Core LJ, Waterfall JJ, Lis JT. 2008. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science* **322**: 1845-1848.

Cramer P. 2004. RNA polymerase II structure: from core to functional complexes. *Curr Opin Genet Dev* **14**: 218-226.

Damgaard CK, Kahns S, Lykke-Andersen S, Nielsen AL, Jensen TH, Kjems J. 2008. A 5' splice site enhances the recruitment of basal transcription initiation factors in vivo. *Mol Cell* **29**: 271-278.

Dantonel J-C, Murthy KGK, Manley JL, Tora L. 1997. Transcription factor TFIID recruits factor CPSF for formation of 3' end of mRNA. *Nature* **389**: 399-402.

Das R, Dufu K, Romney B, Feldt M, Elenko M, Reed R. 2006. Functional coupling of RNAP II transcription to spliceosome assembly. *Genes Dev* **20**: 1100-1109.

David CJ, Boyne AR, Millhouse SR, Manley JL. 2011. The RNA polymerase II C-terminal domain promotes splicing activation through recruitment of a U2AF65-Prp19 complex. *Genes Dev* **25**: 972-983.

David CJ, Manley JL. 2011. The RNA polymerase C-terminal domain: a new role in spliceosome assembly. *Transcription* **2**: 221-225.

de Almeida SF, Grosso AR, Koch F, Fenouil R, Carvalho S, Andrade J, Levezinho H, Gut M, Eick D, Gut I et al. 2011. Splicing enhances recruitment of methyltransferase HYPB/Setd2 and methylation of histone H3 Lys36. *Nat Struct Mol Biol* **18**: 977-983.

Dermody JL, Dreyfuss JM, Villen J, Ogundipe B, Gygi SP, Park PJ, Ponticelli AS, Moore CL, Buratowski S, Bucheli ME. 2008. Unphosphorylated SR-Like Protein Npl3 Stimulates RNA Polymerase II Elongation. *PLoS One* **3**: e3273.

Devaiah BN, Lewis BA, Cherman N, Hewitt MC, Albrecht BK, Robey PG, Ozato K, Sims RJ, 3rd, Singer DS. 2012. BRD4 is an atypical kinase that phosphorylates Serine2 of the RNA Polymerase II carboxy-terminal domain. *Proc Natl Acad Sci U S A* **109**: 6927-6932.

Dichtl B, Blank D, Ohnacker M, Friedlein A, Roeder D, Langen H, Keller W. 2002a. A role for SSU72 in balancing RNA polymerase II transcription elongation and termination. *Mol Cell* **10**: 1139-1150.

Dichtl B, Blank D, Sadowski M, Hubner W, Weiser S, Keller W. 2002b. Yhh1p/Cft1p directly links poly(A) site recognition and RNA polymerase II transcription termination. *The Embo Journal* **21**: 4125-4135.

Drouin S, Laramee L, Jacques PE, Forest A, Bergeron M, Robert F. 2010. DSIF and RNA polymerase II CTD phosphorylation coordinate the recruitment of Rpd3S to actively transcribed genes. *PLoS Genet* **6**: e1001173.

Eberhardy SR, Farnham PJ. 2002. Myc recruits P-TEFb to mediate the final step in the transcriptional activation of the cad promoter. *J Biol Chem* **277**: 40156-40162.

Eberle AB, Hessle V, Helbig R, Dantoft W, Gimber N, Visa N. 2010. Splice-site mutations cause Rrp6-mediated nuclear retention of the unspliced RNAs and transcriptional down-regulation of the splicing-defective genes. *PLoS One* **5**: e11540.

Edmunds JW, Mahadevan LC, Clayton AL. 2008. Dynamic histone H3 methylation during gene induction: HYPB/Setd2 mediates all H3K36 trimethylation. *The Embo Journal* **27**: 406-420.

Egloff S, Murphy S. 2008. Cracking the RNA polymerase II CTD code. *Trends Genet* **24**: 280-288.

Egloff S, O'Reilly D, Chapman RD, Taylor A, Tanzhaus K, Pitts L, Eick D, Murphy S. 2007. Serine-7 of the RNA polymerase II CTD is specifically required for snRNA gene expression. *Science* **318**: 1777-1779.

Egloff S, O'Reilly D, Murphy S. 2008. Expression of human snRNA genes from beginning to end. *Biochem Soc Trans* **36**: 590-594.

Egloff S, Szczepaniak SA, Dienstbier M, Taylor A, Knight S, Murphy S. 2010. The integrator complex recognizes a new double mark on the RNA polymerase II carboxyl-terminal domain. *J Biol Chem* **285**: 20564-20569.

Egloff S, Zaborowska J, Laitem C, Kiss T, Murphy S. 2012. Ser7 phosphorylation of the CTD recruits the RPAP2 Ser5 phosphatase to snRNA genes. *Mol Cell* **45**: 111-122.

Emili A, Shales M, McCracken S, Xie W, Tucker PW, Kobayashi R, Blencowe BJ, Ingles CJ. 2002. Splicing and transcription-associated proteins PSF and p54nrb/nonO bind to the RNA polymerase II CTD. *Rna* **8**: 1102-1111.

Fabrega C, Shen V, Shuman S, Lima CD. 2003. Structure of an mRNA Capping Enzyme Bound to the Phosphorylated Carboxy-Terminal Domain of RNA Polymerase II. *Mol Cell* **11**: 1549-1561.

Feaver WJ, Gileadi O, Li Y, Kornberg RD. 1991. Ctd Kinase Associated with Yeast Rna Polymerase-Ii Initiation Factor-B. *Cell* **67**: 1223-1230.

Fong N, Bentley DL. 2001. Capping, splicing, and 3' processing are independently stimulated by RNA polymerase II: different functions for different segments of the CTD. *Genes Dev* **15**: 1783-1795.

Fong N, Bird G, Vigneron M, Bentley DL. 2003. A 10 residue motif at the C-terminus of the RNA pol II CTD is required for transcription, splicing and 3' end processing. *Embo J* **22**: 4274-4282.

Fujinaga K, Cujec TP, Peng J, Garriga J, Price DH, Grana X, Peterlin BM. 1998. The ability of positive transcription elongation factor B to transactivate human immunodeficiency virus transcription depends on a functional kinase domain, cyclin T1, and Tat. *Journal of Virology* **72**: 7154-7159.

Fujinaga K, Irwin D, Huang Y, Taube R, Kurosu T, Peterlin BM. 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.

Furger A, O'Sullivan JM, Binnie A, Lee BA, Proudfoot NJ. 2002. Promoter proximal splice sites enhance transcription. *Genes Dev* **16**: 2792-2799.Galbraith MD, Donner AJ, Espinosa JM. 2010. CDK8: a positive regulator of transcription. *Transcription* **1**: 4-12.

Ganem C, Devaux F, Torchet C, Jacq C, Quevillon-Cheruel S, Labesse G, Facca C, Faye G. 2003. Ssu72 is a phosphatase essential for transcription termination of snoRNAs and specific mRNAs in yeast. *The Embo Journal* **22**: 1588-1598.

Gerber HP, Hagmann M, Seipel K, Georgiev O, West MAL, Litingtung Y, Schaffner W, Corden JL. 1995. RNA polymerase II C-terminal domain required for enhancer-driven transcription. *Nature* **374**: 660-662.

Ghazy MA, He X, Singh BN, Hampsey M, Moore C. 2009. The essential N terminus of the Pta1 scaffold protein is required for snoRNA transcription termination and Ssu72 function but is dispensable for pre-mRNA 3'-end processing. *Mol Cell Biol* **29**: 2296-2307.

Ghosh A, Shuman S, Lima CD. 2008. The structure of Fcp1, an essential RNA polymerase II CTD phosphatase. *Mol Cell* **32**: 478-490.

Ghosh A, Shuman S, Lima CD. 2011. Structural insights to how mammalian capping enzyme reads the CTD code. *Mol Cell* **43**: 299-310.

Ghosh S, Garcia-Blanco MA. 2000. Coupled in vitro synthesis and splicing of RNA polymerase II transcripts. *Rna* **6**: 1325-1334.

Glover-Cutter K, Kim S, Espinosa J, Bentley DL. 2008. RNA polymerase II pauses and associates with pre-mRNA processing factors at both ends of genes. *Nat Struct Mol Biol* **15**: 71-78.

Glover-Cutter K, Larochelle S, Erickson B, Zhang C, Shokat K, Fisher RP, Bentley DL. 2009. 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**: 5455-5464.

Gnad F, de Godoy LM, Cox J, Neuhauser N, Ren S, Olsen JV, Mann M. 2009. High-accuracy identification and bioinformatic analysis of in vivo protein phosphorylation sites in yeast. *Proteomics* **9**: 4642-4652.

Govind CK, Qiu H, Ginsburg DS, Ruan C, Hofmeyer K, Hu C, Swaminathan V, Workman JL, Li B, Hinnebusch AG. 2010. Phosphorylated Pol II CTD recruits multiple HDACs, including Rpd3C(S), for methylation-dependent deacetylation of ORF nucleosomes. *Mol Cell* **39**: 234-246.

Gudipati RK, Villa T, Boulay J, Libri D. 2008. Phosphorylation of the RNA polymerase II C-terminal domain dictates transcription termination choice. *Nat Struct Mol Biol* **15**: 786-794.

Guo Z, Stiller JW. 2004. Comparative genomics of cyclin-dependent kinases suggest co-evolution of the RNAP II C-terminal domain and CTD-directed CDKs. *BMC Genomics* **5**: 69.

Hampsey M, Singh BN, Ansari A, Laine JP, Krishnamurthy S. 2011. Control of eukaryotic gene expression: gene loops and transcriptional memory. *Adv Enzyme Regul* **51**: 118-125.

Hausmann S, Shuman S. 2002. Characterization of the CTD phosphatase Fcp1 from fission yeast. Preferential dephosphorylation of serine 2 versus serine 5. *J Biol Chem* **277**: 21213-21220.

He X, Khan AU, Cheng H, Pappas DL, Jr., Hampsey M, Moore CL. 2003. Functional interactions between the transcription and mRNA 3' end processing machineries mediated by Ssu72 and Sub1. *Genes Dev* **17**: 1030-1042.

Hengartner CJ, Myer VE, Liao SM, Wilson CJ, Koh SS, Young RA. 1998. Temporal regulation of RNA polymerase II by Srb10 and Kin28 cyclin-dependent kinases. *Mol Cell* **2**: 43-53.

Herrmann CH, Rice AP. 1995. Lentivirus Tat proteins specifically associate with a cellular protein kinase, TAK, that hyperphosphorylates the carboxyl-terminal domain of the large subunit of RNA polymerase II: candidate for a Tat cofactor. *Journal of Virology* **69**: 1612-1620.

Hintermair C, Heidemann M, Koch F, Descostes N, Gut M, Gut I, Fenouil R, Ferrier P, Flatley A, Kremmer E et al. 2012. Threonine-4 of mammalian RNA polymerase II CTD is targeted by Polo-like kinase 3 and required for transcriptional elongation. *The Embo Journal*.

Hirose Y, Manley JL. 1998. RNA polymerase II is an essential mRNA polyadenylation factor. *Nature* **395**: 93-96.

Hirose Y, Manley JL. 2000. RNA polymerase II and the integration of nuclear events. *Genes Dev* **14**: 1415-1429.

Hirose Y, Ohkuma Y. 2007. Phosphorylation of the C-terminal domain of RNA polymerase II plays central roles in the integrated events of eucaryotic gene expression. *J Biochem* **141**: 601-608.

Hirose Y, Tacke R, Manley JL. 1999. Phosphorylated RNA polymerase II stimulates pre-mRNA splicing. *Genes Dev* **13**: 1234-1239.

Hnilicova J, Stanek D. 2011. Where splicing joins chromatin. *Nucleus* **2**: 182-188.

Ho CK, Shuman S. 1999. Distinct roles for CTD Ser-2 and Ser-5 phosphorylation in the recruitment and allosteric activation of mammalian mRNA capping enzyme. *Mol Cell* **3**: 405-411.

Honorine R, Mosrin-Huaman C, Hervouet-Coste N, Libri D, Rahmouni AR. 2011. Nuclear mRNA quality control in yeast is mediated by Nrd1 co-transcriptional recruitment, as revealed by the targeting of Rho-induced aberrant transcripts. *Nucleic Acids Res* **39**: 2809-2820.

Hsin JP, Sheth A, Manley JL. 2011. RNAP II CTD phosphorylated on threonine-4 is required for histone mRNA 3' end processing. *Science* **334**: 683-686.

Hughes CM, Rozenblatt-Rosen O, Milne TA, Copeland TD, Levine SS, Lee JC, Hayes DN, Shanmugam KS, Bhattacharjee A, Biondi CA et al. 2004. Menin associates with a trithorax family histone methyltransferase complex and with the hoxc8 locus. *Mol Cell* **13**: 587-597.

Jacobs EY, Ogiwara I, Weiner AM. 2003. Role of the C-Terminal Domain of RNA Polymerase II in U2 snRNA Transcription and 3' Processing. *Mol Cell Biol* **24**: 846-855.

Jacquier A. 2009. The complex eukaryotic transcriptome: unexpected pervasive transcription and novel small RNAs. *Nat Rev Genet* **10**: 833-844.

Joshi AA, Struhl K. 2005. Eaf3 chromodomain interaction with methylated H3-K36 links histone deacetylation to Pol II elongation. *Mol Cell* **20**: 971-978.

Jove R, Manley JL. 1984. In vitro transcription from the adenovirus 2 major late promoter utilizing templates truncated at promoter-proximal sites. *J Biol Chem* **259**: 8513-8521.

Kamenski T, Heilmeier S, Meinhart A, Cramer P. 2004. Structure and mechanism of RNA polymerase II CTD phosphatases. *Mol Cell* **15**: 399-407.

Kanazawa S, Soucek L, Evan G, Okamoto T, Peterlin BM. 2003. c-Myc recruits P-TEFb for transcription, cellular proliferation and apoptosis. *Oncogene* **22**: 5707-5711.

Kaneko S, Rozenblatt-Rosen O, Meyerson M, Manley JL. 2007. The multifunctional protein p54nrb/PSF recruits the exonuclease XRN2 to facilitate pre-mRNA 3' processing and transcription termination. *Genes Dev* **21**: 1779-1789.

Kelly WG, Dahmus ME, Hart GW. 1993. RNA polymerase II is a glycoprotein. Modification of the COOH-terminal domain by O-GlcNAc. *J Biol Chem* **268**: 10416-10424.

Keogh MC, Kurdistani SK, Morris SA, Ahn SH, Podolny V, Collins SR, Schuldiner M, Chin K, Punna T, Thompson NJ et al. 2005. Cotranscriptional set2 methylation of histone H3 lysine 36 recruits a repressive Rpd3 complex. *Cell* **123**: 593-605.

Keogh MC, Podolny V, Buratowski S. 2003. Bur1 Kinase Is Required for Efficient Transcription Elongation by RNA Polymerase II. *Mol Cell Biol* **23**: 7005-7018.

Kim H, Erickson B, Luo W, Seward D, Graber JH, Pollock DD, Megee PC, Bentley DL. 2010. Gene-specific RNA polymerase II phosphorylation and the CTD code. *Nat Struct Mol Biol* **17**: 1279-1286.

Kim M, Ahn SH, Krogan NJ, Greenblatt JF, Buratowski S. 2004a. Transitions in RNA polymerase II elongation complexes at the 3' ends of genes. *The Embo Journal* **23**: 354-364.

Kim M, Krogen NJ, Vasiljeva L, Rando OJ, Nedea E, Greenblatt J, Buratowski S. 2004b. The yeast Rat1 exonuclease promotes transcription termination by RNA polymerase II. *Nature* **432**: 517-522.

Kim M, Suh H, Cho EJ, Buratowski S. 2009. Phosphorylation of the yeast Rpb1 C-terminal domain at serines 2, 5, and 7. *J Biol Chem* **284**: 26421-26426.

Kim M, Vasiljeva L, Rando OJ, Zhelkovsky A, Moore C, Buratowski S. 2006. Distinct pathways for snoRNA and mRNA termination. *Mol Cell* **24**: 723-734.

Kim T, Buratowski S. 2009. Dimethylation of H3K4 by Set1 recruits the Set3 histone deacetylase complex to 5' transcribed regions. *Cell* **137**: 259-272.

Kizer KO, Phatnani HP, Shibata Y, Hall H, Greenleaf AL, Strahl BD. 2005. A novel domain in Set2 mediates RNA polymerase II interaction and couples histone H3 K36 methylation with transcript elongation. *Mol Cell Biol* **25**: 3305-3316.

Kobor MS, Archambault J, Lester W, Holstege FCP, Gileadi O, Jansma DB, Jennings EG, Kouyoumdjian F, Davidson AR, Young RA et al. 1999. An unusual eukaryotic protein phosphatase required for transcription by RNA polymerase II and CTD dephosphorylation in S. cerevisiae. *Mol Cell* **4**: 55-62.

Komarnitsky P, Cho EJ, Buratowski S. 2000. Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. *Genes Dev* **14**: 2452-2460.

Krishnamurthy S, Ghazy MA, Moore C, Hampsey M. 2009. Functional interaction of the Ess1 prolyl isomerase with components of the RNA polymerase II initiation and termination machineries. *Mol Cell Biol* **29**: 2925-2934.

Krishnamurthy S, He X, Reyes-Reyes M, Moore C, Hampsey M. 2004. Ssu72 Is an RNA polymerase II CTD phosphatase. *Mol Cell* **14**: 387-394.

Krogan NJ, Dover J, Wood A, Schneider J, Heidt J, Boateng MA, Dean K, Ryan OW, Golshani A, Johnston M et al. 2003. The Paf1 Complex Is Required for Histone H3 Methylation by COMPASS and Dot1p: Linking Transcriptional Elongation to Histone Methylation. *Mol Cell* **11**: 721-729.

Kuehner JN, Pearson EL, Moore C. 2011. Unravelling the means to an end: RNA polymerase II transcription termination. *Nat Rev Mol Cell Biol* **12**: 283-294.

Lander ES Linton LM Birren B Nusbaum C Zody MC Baldwin J Devon K Dewar K Doyle M FitzHugh W et al. 2001. Initial sequencing and analysis of the human genome. *Nature* **409**: 860-921.

Larochelle S, Batliner J, Gamble MJ, Barboza NM, Kraybill BC, Blethrow JD, Shokat KM, Fisher RP. 2006. Dichotomous but stringent substrate selection by the dual-function Cdk7 complex revealed by chemical genetics. *Nat Struct Mol Biol* **13**: 55-62.

Lee JH, Skalnik DG. 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.

Lee JM, Greenleaf AL. 1997. Modulation of RNA polymerase II elongation efficiency by C-terminal heptapeptide repeat domain kinase I. *J Biol Chem* **272**: 10990-10993.

Levine M. 2011. Paused RNA polymerase II as a developmental checkpoint. *Cell* **145**: 502-511.

Li B, Howe L, Anderson S, Yates JR, 3rd, Workman JL. 2003. The Set2 histone methyltransferase functions through the phosphorylated carboxyl-terminal domain of RNA polymerase II. *J Biol Chem* **278**: 8897-8903.

Li H, Zhang Z, Wang B, Zhang J, Zhao Y, Jin Y. 2007. Wwp2-mediated ubiquitination of the RNA polymerase II large subunit in mouse embryonic pluripotent stem cells. *Mol Cell Biol* **27**: 5296-5305.

Li M, Phatnani HP, Guan Z, Sage H, Greenleaf AL, Zhou P. 2005. Solution structure of the Set2-Rpb1 interacting domain of human Set2 and its interaction with the hyperphosphorylated C-terminal domain of Rpb1. *Proc Natl Acad Sci U S A* **102**: 17636-17641.

Li X, Manley JL. 2005. Inactivation of the SR protein splicing factor ASF/SF2 results in genomic instability. *Cell* **122**: 365-378.

Li X, Manley JL. 2006. Cotranscriptional processes and their influence on genome stability. *Genes Dev* **20**: 1838-1847.

Liao SM, Zhang J, Jeffery DA, Koleske AJ, Thompson CM, Chao DM, Viljoen M, van Vuuren HJJ, Young RA. 1995. A kinase-cyclin pair in the RNA polymerase II holoenzyme. *Nature* **374**: 193-196.

Licatalosi DS, Geiger G, Minet M, Schroeder S, Cilli K, McNeil JB, Bentley DL. 2002. Functional interaction of yeast pre-mRNA 3' end processing factors with RNA polymerase II. *Mol Cell* **9**: 1101-1111.

Lin PS, Dubois MF, Dahmus ME. 2002. TFIIF-associating carboxyl-terminal domain phosphatase dephosphorylates phosphoserines 2 and 5 of RNA polymerase II. *J Biol Chem* **277**: 45949-45956.

Lin S, Coutinho-Mansfield G, Wang D, Pandit S, Fu XD. 2008. The splicing factor SC35 has an active role in transcriptional elongation. *Nat Struct Mol Biol* **15**: 819-826.

Litingtung Y, Lawler AM, Sebald SM, Lee E, Gearhart JD, Westphal H, Corden JL. 1999. Growth retardation and neonatal lethality in mice with a homozygous deletion in the C-terminal domain of RNA polymerase II. *Mol Gen Genet* **261**: 100-105.

Liu P, Greenleaf AL, Stiller JW. 2008. The essential sequence elements required for RNAP II carboxyl-terminal domain function in yeast and their evolutionary conservation. *Mol Biol Evol* **25**: 719-727.

Liu P, Kenney JM, Stiller JW, Greenleaf AL. 2010. Genetic organization, length conservation, and evolution of RNA polymerase II carboxyl-terminal domain. *Mol Biol Evol* **27**: 2628-2641.

Lu H, Zawel L, Fisher L, Egly JM, Reinberg D. 1992. Human General Transcription Factor IIH Phosphorylates the C-Terminal Domain of Rna Polymerase II. *Nature* **358**: 641-645.

Lu KP, Hanes SD, Hunter T. 1996. A human peptidyl-prolyl isomerase essential for regulation of mitosis. *Nature* **380**: 544-547.

Luco RF, Allo M, Schor IE, Kornblihtt AR, Misteli T. 2011. Epigenetics in alternative pre-mRNA splicing. *Cell* **144**: 16-26.

Luco RF, Pan Q, Tominaga K, Blencowe BJ, Pereira-Smith OM, Misteli T. 2010. Regulation of alternative splicing by histone modifications. *Science* **327**: 996-1000.

Lunde BM, Reichow SL, Kim M, Suh H, Leeper TC, Yang F, Mutschler H, Buratowski S, Meinhart A, Varani G. 2010. Cooperative interaction of transcription termination factors with the RNA polymerase II C-terminal domain. *Nat Struct Mol Biol* **17**: 1195-1201.

Luo W, Johnson AW, Bentley DL. 2006. The role of Rat1 in coupling mRNA 3'-end processing to transcription termination: implications for a unified allosteric-torpedo model. *Genes Dev* **20**: 954-965.

Maniatis T, Reed R. 2002. An extensive network of coupling among gene expression machines. *Nature* **416**: 499-506.

Manley JL. 2002. Nuclear coupling: RNA processing reaches back to transcription. *Nat Struct Biol* **9**: 790-791.

Mapendano CK, Lykke-Andersen S, Kjems J, Bertrand E, Jensen TH. 2010. Crosstalk between mRNA 3' end processing and transcription initiation. *Mol Cell* **40**: 410-422.

Marciniak RA, Sharp PA. 1991. HIV-1 Tat protein promotes formation of more-processive elongation complexes. *The Embo Journal* **10**: 4189-4196.

Marshall NF, Peng J, Xie Z, Price DH. 1996. Control of RNA polymerase II elongation potential by a novel carboxyl-terminal domain kinase. *J Biol Chem* **271**: 27176-27183.

Marshall NF, Price DH. 1995. Purification of P-TEFb, a transcription factor required for the transition into productive elongation. *J Biol Chem* **270**: 12335-12338.

Martins SB, Rino J, Carvalho T, Carvalho C, Yoshida M, Klose JM, de Almeida SF, Carmo-Fonseca M. 2011. Spliceosome assembly is coupled to RNA polymerase II dynamics at the 3' end of human genes. *Nat Struct Mol Biol* **18**: 1115-1123.

Marzluff WF, Wagner EJ, Duronio RJ. 2008. Metabolism and regulation of canonical histone mRNAs: life without a poly(A) tail. *Nat Rev Genet* **9**: 843-854.

Mayer A, Heidemann M, Lidschreiber M, Schreieck A, Sun M, Hintermair C, Kremmer E, Eick D, Cramer P. 2012a. CTD tyrosine phosphorylation impairs termination factor recruitment to RNA polymerase II. *Science* **336**: 1723-1725.

Mayer A, Lidschreiber M, Siebert M, Leike K, Soding J, Cramer P. 2010. Uniform transitions of the general RNA polymerase II transcription complex. *Nat Struct Mol Biol* **17**: 1272-1278.

Mayer A, Schreieck A, Lidschreiber M, Leike K, Martin DE, Cramer P. 2012b. The spt5 C-terminal region recruits yeast 3' RNA cleavage factor I. *Mol Cell Biol* **32**: 1321-1331.

McCracken S, Fong N, Rosonina E, Yankulov K, Brothers G, Siderovski D, Hessel A, Foster S, Amgen EST Program, Shuman S et al. 1997a. 5'-Capping enzymes are targeted to pre-mRNA by binding to the phosphorylated carboxy-terminal domain of RNA polymerase II. *Genes Dev* **11**: 3306-3318.

McCracken S, Fong N, Yankulov K, Ballantyne S, Pan G, Greenblatt J, Patterson SD, Wickens M, Bentley DL. 1997b. The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. *Nature* **385**: 357-361.

Medlin JE, Uguen P, Taylor A, Bentley DL, Murphy S. 2003. The C-terminal domain of pol II and a DRB-sensitive kinase are required for 3' processing of U2 snRNA. *The Embo Journal* **22**: 925-934.

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

Meinhart A, Kamenski T, Hoeppner S, Baumli S, Cramer P. 2005. A structural perspective of CTD function. *Genes Dev* **19**: 1401-1415.

Meininghaus M, Chapman RD, Horndasch M, Eick D. 2000. Conditional expression of RNA polymerase II in mammalian cells. Deletion of the carboxyl-terminal domain of the large subunit affects early steps in transcription. *J Biol Chem* **275**: 24375-24382.

Millhouse S, Manley JL. 2005. The C-terminal domain of RNA polymerase II functions as a phosphorylation-dependent splicing activator in a heterologous protein. *Mol Cell Biol* **25**: 533-544.

Milne TA, Dou Y, Martin ME, Brock HW, Roeder RG, Hess JL. 2005. MLL associates specifically with a subset of transcriptionally active target genes. *Proc Natl Acad Sci U S A* **102**: 14765-14770.

Misteli T, Spector DL. 1999. RNA polymerase II targets pre-mRNA splicing factors to transcription sites in vivo. *Mol Cell* **3**: 697-705.

Morris DP, Greenleaf AL. 2000. The splicing factor, Prp40, binds the phosphorylated carboxyl-terminal domain of RNA polymerase II. *J Biol Chem* **275**: 39935-39943.

Mortillaro MJ, Blencowe BJ, Wei X, Nakayasu H, Du L, Warren SL, Sharp PA, Berezney R. 1996. A hyperphosphorylated form of the large subunit of RNA polymerase II is associated with splicing complexes and the nuclear matrix. *Proc Natl Acad Sci U S A* **93**: 8253-8257.

Mosley AL, Pattenden SG, Carey M, Venkatesh S, Gilmore JM, Florens L, Workman JL, Washburn MP. 2009. Rtr1 is a CTD phosphatase that regulates RNA polymerase II during the transition from serine 5 to serine 2 phosphorylation. *Mol Cell* **34**: 168-178.

Munoz MJ, de la Mata M, Kornblihtt AR. 2010. The carboxy terminal domain of RNA polymerase II and alternative splicing. *Trends in Biochemical Sciences* **35**: 497-504.

Murray S, Udupa R, Yao S, Hartzog G, Prelich G. 2001. Phosphorylation of the RNA polymerase II carboxy-terminal domain by the Bur1 cyclin-dependent kinase. *Mol Cell Biol* **21**: 4089-4096.

Muse GW, Gilchrist DA, Nechaev S, Shah R, Parker JS, Grissom SF, Zeitlinger J, Adelman K. 2007. RNA polymerase is poised for activation across the genome. *Nat Genet* **39**: 1507-1511.

Nagaike T, Logan C, Hotta I, Rozenblatt-Rosen O, Meyerson M, Manley JL. 2011. Transcriptional activators enhance polyadenylation of mRNA precursors. *Mol Cell* **41**: 409-418.

Nagaike T, Manley JL. 2011. Transcriptional activators enhance polyadenylation of mRNA precursors. *RNA Biol* **8**: 964-967.

Nechaev S, Fargo DC, dos Santos G, Liu L, Gao Y, Adelman K. 2010. Global analysis of short RNAs reveals widespread promoter-proximal stalling and arrest of Pol II in Drosophila. *Science* **327**: 335-338.

Nedea E, He X, Kim M, Pootoolal J, Zhong G, Canadien V, Hughes T, Buratowski S, Moore CL, Greenblatt J. 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.

Ng HH, Robert F, Young RA, Struhl K. 2003. Targeted Recruitment of Set1 Histone Methylase by Elongating Pol II Provides a Localized Mark and Memory of Recent Transcriptional Activity. *Mol Cell* **11**: 709-719.

Ni Z, Schwartz BE, Werner J, Suarez JR, Lis JT. 2004. Coordination of transcription, RNA processing, and surveillance by P-TEFb kinase on heat shock genes. *Mol Cell* **13**: 55-65.

O'Sullivan JM, Tan-Wong SM, Morillon A, Lee B, Coles J, Mellor J, Proudfoot NJ. 2004. Gene loops juxtapose promoters and terminators in yeast. *Nat Genet* **36**: 1014-1018.Park NJ, Tsao DC, Martinson HG. 2004. The two steps of poly(A)-dependent termination, pausing and release, can be uncoupled by truncation of the RNA polymerase II carboxyl-terminal repeat domain. *Mol Cell Biol* **24**: 4092-4103.

Patturajan M, Wei X, Berezney R, Corden JL. 1998. A nuclear matrix protein interacts with the phosphorylated C-terminal domain of RNA polymerase II. *Mol Cell Biol* **18**: 2406-2415.

Pawlicki JM, Steitz JA. 2010. Nuclear networking fashions pre-messenger RNA and primary microRNA transcripts for function. *Trends Cell Biol* **20**: 52-61.

Peng J, Marshall NF, Price DH. 1998a. Identification of a cyclin subunit required for the function of Drosophila P-TEFb. *J Biol Chem* **273**: 13855-13860.

Peng J, Zhu Y, Milton JT, Price DH. 1998b. Identification of multiple cyclin subunits of human P-TEFb. *Genes Dev* **12**: 755-762.

Perales R, Bentley D. 2009. "Cotranscriptionality": the transcription elongation complex as a nexus for nuclear transactions. *Mol Cell* **36**: 178-191.

Phatnani HP, Greenleaf AL. 2006. Phosphorylation and functions of the RNA polymerase II CTD. *Genes Dev* **20**: 2922-2936.

Phatnani HP, Jones JC, Greenleaf AL. 2004. Expanding the functional repertoire of CTD kinase I and RNA polymerase II: novel phosphoCTD-associating proteins in the yeast proteome. *Biochemistry* **43**: 15702-15719.

Pirngruber J, Shchebet A, Schreiber L, Shema E, Minsky N, Chapman RD, Eick D, Aylon Y, Oren M, Johnsen SA. 2009. CDK9 directs H2B monoubiquitination and controls replication-dependent histone mRNA 3'-end processing. *EMBO Rep* **10**: 894-900.

Proudfoot NJ. 1989. How RNA polymerase II terminates transcription in higher eukaryotes. *Trends in Biochemical Sciences* **14**: 105-110.

Proudfoot NJ, Furger A, Dye MJ. 2002. Integrating mRNA processing with transcription. *Cell* **108**: 501-512.

Qiu H, Hu C, Hinnebusch AG. 2009. Phosphorylation of the Pol II CTD by KIN28 enhances BUR1/BUR2 recruitment and Ser2 CTD phosphorylation near promoters. *Mol Cell* **33**: 752-762.

Rahl PB, Lin CY, Seila AC, Flynn RA, McCuine S, Burge CB, Sharp PA, Young RA. 2010. c-Myc regulates transcriptional pause release. *Cell* **141**: 432-445.

Renner DB, Yamaguchi Y, Wada T, Handa H, Price DH. 2001. A highly purified RNA polymerase II elongation control system. *J Biol Chem* **276**: 42601-42609.

Richard P, Kiss T. 2006. Integrating snoRNP assembly with mRNA biogenesis. *EMBO Rep* **7**: 590-592.

Richard P, Manley JL. 2009. Transcription termination by nuclear RNA polymerases. *Genes Dev* **23**: 1247-1269.

Rodriguez CR, Cho EJ, Keogh MC, Moore CL, Greenleaf AL, Buratowski S. 2000. Kin28, the TFIIH-Associated Carboxy-Terminal Domain Kinase, Facilitates the Recruitment of mRNA Processing Machinery to RNA Polymerase II. *Mol Cell Biol*: 104-112.

Rondon AG, Mischo HE, Kawauchi J, Proudfoot NJ. 2009. Fail-safe transcriptional termination for protein-coding genes in S. cerevisiae. *Mol Cell* **36**: 88-98.

Rosonina E, Blencowe BJ. 2004. Analysis of the requirement for RNA polymerase II CTD heptapeptide repeats in pre-mRNA splicing and 3'-end cleavage. *Rna* **10**: 581-589.

Rosonina E, Duncan SM, Manley JL. 2012. Sumoylation of transcription factor Gcn4 facilitates its Srb10-mediated clearance from promoters in yeast. *Genes Dev* **26**: 350-355.

Rosonina E, Ip JY, Calarco JA, Bakowski MA, Emili A, McCracken S, Tucker P, Ingles CJ, Blencowe BJ. 2005. Role for PSF in mediating transcriptional activator-dependent stimulation of pre-mRNA processing in vivo. *Mol Cell Biol* **25**: 6734-6746.

Rozenblatt-Rosen O, Nagaike T, Francis JM, Kaneko S, Glatt KA, Hughes CM, LaFramboise T, Manley JL, Meyerson M. 2009. The tumor suppressor Cdc73 functionally associates with CPSF and CstF 3' mRNA processing factors. *Proc Natl Acad Sci U S A* **106**: 755-760.

Ryan K, Murthy KGK, Kaneko S, Manley JL. 2002. Requirements of the RNA Polymerase II C-Terminal Domain for Reconstituting Pre-mRNA 3' Cleavage. *Mol Cell Biol* **22**: 1684-1692.

Sadowski M, Dichtl B, Hubner W, Keller W. 2003. Independent functions of yeast Pcf11p in pre-mRNA 3' end processing and in transcription termination. *The Embo Journal* **22**: 2167-2177.

Schwer B, Shuman S. 2011. Deciphering the RNA polymerase II CTD code in fission yeast. *Mol Cell* **43**: 311-318.

Shi Y, Di Giammartino DC, Taylor D, Sarkeshik A, Rice WJ, Yates JR, 3rd, Frank J, Manley JL. 2009. Molecular architecture of the human pre-mRNA 3' processing complex. *Mol Cell* **33**: 365-376.

Shilatifard A. 2012. The COMPASS Family of Histone H3K4 Methylases: Mechanisms of Regulation in Development and Disease Pathogenesis. *Annu Rev Biochem* **81**: 65-95.

Sikorski TW, Buratowski S. 2009. The basal initiation machinery: beyond the general transcription factors. *Curr Opin Cell Biol* **21**: 344-351.

Sims RJ, 3rd, Rojas LA, Beck D, Bonasio R, Schuller R, Drury WJ, 3rd, Eick D, Reinberg D. 2011. The C-terminal domain of RNA polymerase II is modified by site-specific methylation. *Science* **332**: 99-103.

Singh BN, Hampsey M. 2007. A transcription-independent role for TFIIB in gene looping. *Mol Cell* **27**: 806-816.

Singh N, Ma Z, Gemmill T, Wu X, Defiglio H, Rossettini A, Rabeler C, Beane O, Morse RH, Palumbo MJ et al. 2009. The Ess1 prolyl isomerase is required for transcription termination of small noncoding RNAs via the Nrd1 pathway. *Mol Cell* **36**: 255-266.

Skaar DA, Greenleaf AL. 2002. The RNA polymerase II CTD kinase CTDK-I affects pre-mRNA 3' cleavage/polyadenylation through the processing component Pti1p. *Mol Cell* **10**: 1429-1439.

Skourti-Stathaki K, Proudfoot NJ, Gromak N. 2011. Human senataxin resolves RNA/DNA hybrids formed at transcriptional pause sites to promote Xrn2-dependent termination. *Mol Cell* **42**: 794-805.

Spain MM, Govind CK. 2011. A role for phosphorylated Pol II CTD in modulating transcription coupled histone dynamics. *Transcription* **2**: 78-81.

Steinmetz EJ, Brow DA. 1996. Repression of gene expression by an exogenous sequence element acting in concert with a heterogeneous nuclear ribonucleoprotein-like protein, Nrd1, and the putative helicase Sen1. *Mol Cell Biol* **16**: 6993-7003.

Steinmetz EJ, Brow DA. 1998. Control of pre-mRNA accumulation by the essential yeast protein Nrd1 requires high-affinity transcript binding and a domain implicated in RNA polymerase II association. *Proc Natl Acad Sci U S A* **95**: 6699-6704.

Steinmetz EJ, Conrad NK, Brow DA, Corden JL. 2001. RNA-binding protein Nrd1 directs poly(A)-independent 3'-end formation of RNA polymerase II transcripts. *Nature* **413**: 327-331.

Steinmetz EJ, Warren CL, Kuehner JN, Panbehi B, Ansari AZ, Brow DA. 2006. Genome-wide distribution of yeast RNA polymerase II and its control by Sen1 helicase. *Mol Cell* **24**: 735-746.

Sterner DE, Lee JM, Hardin SE, Greenleaf AL. 1995. The yeast carboxyl-terminal repeat domain kinase CTDK-I is a divergent cyclin-cyclin-dependent kinase complex. *Mol Cell Biol* **15**: 5176-5724.

Stiller JW, Cook MS. 2004. Functional unit of the RNA polymerase II C-terminal domain lies within heptapeptide pairs. *Eukaryot Cell* **3**: 735-740.

Stiller JW, McConaughy BL, Hall BD. 2000. Evolutionary complementation for polymerase II CTD function. *Yeast* **16**: 57-64.

Suh MH, Meyer PA, Gu M, Ye P, Zhang M, Kaplan CD, Lima CD, Fu J. 2010. A dual interface determines the recognition of RNA polymerase II by RNA capping enzyme. *J Biol Chem* **285**: 34027-34038.

Sun M, Lariviere L, Dengl S, Mayer A, Cramer P. 2010. A tandem SH2 domain in transcription elongation factor Spt6 binds the phosphorylated RNA polymerase II C-terminal repeat domain (CTD). *J Biol Chem* **285**: 41597-41603.

Sun X, Zhang Y, Cho H, Rickert P, Lees E, Lane W, Reinberg D. 1998. NAT, a human complex containing Srb polypeptides that functions as a negative regulator of activated transcription. *Mol Cell* **2**: 213-222.

Sun XJ, Wei J, Wu XY, Hu M, Wang L, Wang HH, Zhang QH, Chen SJ, Huang QH, Chen Z. 2005. Identification and characterization of a novel human histone H3 lysine 36-specific methyltransferase. *J Biol Chem* **280**: 35261-35271.

Sun ZW, Hampsey M. 1996. Synthetic enhancement of a TFIIB defect by a mutation in SSU72, an essential yeast gene encoding a novel protein that affects transcription start site selection in vivo. *Mol Cell Biol*: 1557-1566.

Suraweera A, Lim Y, Woods R, Birrell GW, Nasim T, Becherel OJ, Lavin MF. 2009. Functional role for senataxin, defective in ataxia oculomotor apraxia type 2, in transcriptional regulation. *Hum Mol Genet* **18**: 3384-3396.

Tietjen JR, Zhang DW, Rodriguez-Molina JB, White BE, Akhtar MS, Heidemann M, Li X, Chapman RD, Shokat K, Keles S et al. 2010. Chemical-genomic dissection of the CTD code. *Nat Struct Mol Biol* **17**: 1154-1161.

Tsao DC, Park NJ, Nag A, Martinson HG. 2012. Prolonged alpha-amanitin treatment of cells for studying mutated polymerases causes degradation of DSIF160 and other proteins. *Rna* **18**: 222-229.

Ursic D, Chinchilla K, Finkel JS, Culbertson MR. 2004. Multiple protein/protein and protein/RNA interactions suggest roles for yeast DNA/RNA helicase Sen1p in transcription, transcription-coupled DNA repair and RNA processing. *Nucleic Acids Res* **32**: 2441-2452.

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. *Nat Struct Mol Biol* **15**: 795-804.

Verdecia MA, Bowman ME, Lu KP, Hunter T, Noel JP. 2000. Structural basis for phosphoserine-proline recognition by group IV WW domains. *Nature* **7**: 639-643.

Wang Y, Fairley JA, Roberts SG. 2010. Phosphorylation of TFIIB links transcription initiation and termination. *Curr Biol* **20**: 548-553.

Werner-Allen JW, Lee CJ, Liu P, Nicely NI, Wang S, Greenleaf AL, Zhou P. 2011. cis-Proline-mediated Ser(P)5 dephosphorylation by the RNA polymerase II C-terminal domain phosphatase Ssu72. *J Biol Chem* **286**: 5717-5726.

West ML, Gorden JL. 1995. Construction and analysis of yeast RNA polymerase II CTD deletion and substitution mutations. *Genetics* **140**: 1223-1233.

West S, Gromak N, Proudfoot NJ. 2004. Human 5' --> 3' exonuclease Xrn2 promotes transcription termination at co-transcriptional cleavage sites. *Nature* **432**: 522-525.

West S, Proudfoot NJ. 2008. Human Pcf11 enhances degradation of RNA polymerase II-associated nascent RNA and transcriptional termination. *Nucleic Acids Res* **36**: 905-914.

Wood A, Shilatifard A. 2006. Bur1/Bur2 and the Ctk complex in yeast: the split personality of mammalian P-TEFb. *Cell Cycle* **5**: 1066-1068.

Wu CH, Yamaguchi Y, Benjamin LR, Horvat-Gordon M, Washinsky J, Enerly E, Larsson J, Lambertsson A, Handa H, Gilmour D. 2003. NELF and DSIF cause promoter proximal pausing on the hsp70 promoter in Drosophila. *Genes Dev* **17**: 1402-1414.

Xiang K, Nagaike T, Xiang S, Kilic T, Beh MM, Manley JL, Tong L. 2010. Crystal structure of the human symplekin-Ssu72-CTD phosphopeptide complex. *Nature* **467**: 729-733.

Xiang K, Manley JL, Tong L. 2012. The yeast regulator of transcription protein Rtr1 lacks an active site and phohphatase activity. *Nat Commun* **3**: 946. doi:10.1038/ncomms1947

Xiao T, Hall H, Kizer KO, Shibata Y, Hall MC, Borchers CH, Strahl BD. 2003. Phosphorylation of RNA polymerase II CTD regulates H3 methylation in yeast. *Genes Dev* **17**: 654-663.

Xu YX, Hirose Y, Zhou XZ, Lu KP, Manley JL. 2003. Pin1 modulates the structure and function of human RNA polymerase II. *Genes Dev* **17**: 2765-2776.

Xu YX, Manley JL. 2007. Pin1 modulates RNA polymerase II activity during the transcription cycle. *Genes Dev* **21**: 2950-2962.

Yaffe MB, Schutkowski M, Shen M, Zhou XZ, Stukenberg PT, Rahfeld JU, Xu J, Kuang J, Kirchner MW, Fischer G et al. 1997. Sequence-Specific and Phosphorylation-Dependent Proline Isomerization: A Potential Mitotic Regulatory Mechanism. *Science* **278**: 1957-1960.

Yamaguchi Y, Takagi T, Wada T, Yano K, Furuya A, Sugimoto S, Hasegawa J, Handa H. 1999. NELF, a multisubunit complex containing RD, cooperates with DSIF to repress RNA polymerase II elongation. *Cell* **97**: 41-51.

Yao S, Neiman A, Prelich G. 2000. BUR1 and BUR2 encode a divergent cyclin-dependent kinase-cyclin complex important for transcription in vivo. *Mol Cell Biol* **20**: 7080-7087.

Yeo M, Lee SK, Lee B, Ruiz EC, Pfaff SL, Gill GN. 2005. Small CTD phosphatases function in silencing neuronal gene expression. *Science* **307**: 596-600.

Yeo M, Lin PS, Dahmus ME, Gill GN. 2003. A novel RNA polymerase II C-terminal domain phosphatase that preferentially dephosphorylates serine 5. *J Biol Chem* **278**: 26078-26085.

Yoh SM, Cho H, Pickle L, Evans RM, Jones KA. 2007. The Spt6 SH2 domain binds Ser2-P RNAPII to direct Iws1-dependent mRNA splicing and export. *Genes Dev* **21**: 160-174.

Yoh SM, Lucas JS, Jones KA. 2008. The Iws1:Spt6:CTD complex controls cotranscriptional mRNA biosynthesis and HYPB/Setd2-mediated histone H3K36 methylation. *Genes Dev* **22**: 3422-3434.

Yuryev A, Patturajan M, Litingtung Y, Joshi RV, Gentile C, Gebara M, Corden JL. 1996. The C-terminal domain of the largest subunit of RNA polymerase II interacts with a novel set of serine/arginine-rich proteins. *Proc Natl Acad Sci U S A* **93**: 6975-6980.

Zeitlinger J, Stark A, Kellis M, Hong JW, Nechaev S, Adelman K, Levine M, Young RA. 2007. RNA polymerase stalling at developmental control genes in the Drosophila melanogaster embryo. *Nat Genet* **39**: 1512-1516.

Zhang DW, Mosley AL, Ramisetty SR, Rodriguez-Molina JB, Washburn MP, Ansari AZ. 2012. 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**: 8541-8551.

Zhang Y, Kim Y, Genoud N, Gao J, Kelly JW, Pfaff SL, Gill GN, Dixon JE, Noel JP. 2006. Determinants for dephosphorylation of the RNA polymerase II C-terminal domain by Scp1. *Mol Cell* **24**: 759-770.

Zhang Z, Fu J, Gilmour DS. 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.

Zhang Z, Gilmour DS. 2006. Pcf11 is a termination factor in Drosophila that dismantles the elongation complex by bridging the CTD of RNA polymerase II to the nascent transcript. *Mol Cell* **21**: 65-74.

Zhang Z, Wu CH, Gilmour DS. 2004. Analysis of polymerase II elongation complexes by native gel electrophoresis. Evidence for a novel carboxyl-terminal domain-mediated termination mechanism. *J Biol Chem* **279**: 23223-23228.

Zhou M, Halanski MA, Radonovich MF, Kashanchi F, Peng J, Price DH, Brady JN. 2000. Tat modifies the activity of CDK9 to phosphorylate serine 5 of the RNA polymerase II carboxyl-terminal domain during human immunodeficiency virus type 1 transcription. *Mol Cell Biol* **20**: 5077-5086.

Figure Legends

Figure 1. Comparison of select CTD sequences. The CTD sequences of fission and budding yeast, zebrafish and human are shown and aligned to display the context of the heptad repeats. All-consensus YSPTSPS repeats heptads are in red, and the number next to the parenthesis indicates the repeat number. The CTD of fission yeast contains 29 heptads, 24 of which are all-consensus, whereas the CTD of budding yeast consists of 26 heptads, 19 of which are perfect consensus. The fish and human CTDs

each consist of 52 repeats, with a 10-11 residue motif (in bold case) at the very C terminus. Highlighting the conservation of the CTD among vertebrates, fish and human CTDs are 97% identical. Residues in the fish CTD that differ from human are in yellow, and human residues that deviate from the consensus are in blue.

Figure 2. Dynamic modification of the CTD during the transcription cycle. At transcription initiation, CDK7 phosphorylates Ser5 and Ser7 residues. Later, during elongation CDK9 phosphorylates Ser2 and perhaps Thr4, while the phosphate groups on Ser5 and Ser7 are gradually removed by phosphatases. For example Rtr1, likely indirectly with another phosphatase, and Ssu72, with the aid of the prolyl isomerase Pin1, dephosphorylate Ser5-P early and late during elongation, respectively. Ssu72 also dephosphorylates Ser7-P. CDK12 likely also contributes to Ser2 phosphorylation during elongation, at least of some genes. As RNAP II nears termination, Fcp1 dephosphorylates Ser2-P, regenerating unphosphorylated RNAP II that can be recycled for another round of transcription.

Figure 3. The CTD facilitates capping and splicing by recruitment of RNA processing factors. (A) Capping enzyme (CE) is recruited to the vicinity of nascent mRNA by the CTD phosphorylated on Ser5. (B) During transcription, the CTD is phosphorylated on Ser2 while the Ser5-P is dephosphorylated, and is involved in recruiting the indicated splicing factors, which defines splice sites and facilitates assembly of the spliceosome. In this and subsequent figures, green spheres above the CTD represent relevant CTD-binding proteins, while assembled functional complexes are indicated below.

Figure 4. The CTD functions in 3' processing of both polyadenylated and non-polyadenylated RNAs. (A) At 3' ends of polyadenylated mRNA, Ser2-P serves to recruit Pcf11, a component of CFII (human nomenclature is used for all factors). Other 3' end factors, such as CstF50 and AAUAAA-binding factor CPSF-160 (Yhh1 in yeast), also bind the CTD, whereas Ssu72, with the aid of Pin1 (Ess1) must

dephosphorylate Ser5-P. The AAUAAA element and G/U elements are bound by CPSF and CstF, respectively. Loading of some factors, including CPSF and CstF, may occur upstream, perhaps at the promoter (see text). (B) The 3' end of histone pre-mRNA contains a stem-loop motif bound by SLBP and a downstream element recognized by U7 snRNP. A complex containing CPSF73, CPSF100, and Symplekin, is recruited for 3' cleavage. Thr4-P facilitates this process, likely through a yet-to-be-identified factor. (C) The 3' end of snRNA genes contains a 3' box that interacts with the Integrator complex. RPAP2 binds to Ser7-P on the CTD and to recruit the Integrator, and Int 11, an Integrator subunit, cleaves the RNA.

Figure 5. The CTD facilitates different termination mechanisms for protein-coding and non-coding genes. (A) Poly(A)-dependent termination pathway. RNA is cleaved by 3' end processing factors at the polyadenylation site. The CTD with Ser2-P is involved in recruiting factors including Pcf11, Rtt103, p54/PSF and Sen1 to facilitate termination of long polyadenylated transcripts. Pcf11 and Rtt103 are required for the recruitment of exoribonuclease Rat1 in yeast, while Xrn2 is recruited by p54/PSF in humans. Sen1 (Senataxin in humans) may function on some of these genes by resolving RNA-DNA hybrids. (B) Nrd1c-dependent termination pathway. The Nrd1 complex (Nrd1-Nab3-Sen1) interacts via Nrd1 with the CTD phosphorylated on Ser5, which is present at the 3' ends of short genes, such as snoRNAs and CUTs. Ssu72 and Ess1 are also required to dephosphorylate Ser5-P, although the exact mechanism of Nrd1c-dependent termination awaits further studies.

Factor	P-CTD Binding Preference	Reference
Set1 (histone methylase)	Ser5-P	Krogan et al. 2003; Ng et al. 2003
Set1A/1B (histone methylase)	Ser5-P	Lee and Skalnik 2008
MLL1/2 (histone methylase)	Ser5-P	Hughes et al. 2004; Milne et al. 2005
Set2 (histone methylase)	Ser2/5-P	Li et al. 2003; Xiao et al. 2003; Kizer et al. 2005
HYPB (histone methylase)	Ser2/5-P	Li et al. 2005; Sun et al. 2005
Rpd3S (histone deacetylase)	P-CTD	Drouin et al. 2010; Govind et al. 2010
Spt6	Ser2-P	Yoh et al. 2007; Sun et al. 2010
Guanylyltransferase (capping)	Ser5-P	Ho and Shuman 1999; Fabrega et al. 2003; Ghosh et al. 2011; Schwer and Shuman 2011
Prp40 (U1 snRNP)	P-CTD	Morris and Greenleaf 2000
PSF/p54 (multifunctional)	CTD, P-CTD	Emili et al. 2002; Rosonina et al. 2005
U2AF65 (U2 snRNP)	P-CTD	David et al. 2011
CstF50 (CstF)	CTD, P-CTD	Fong and Bentley 2001
Yhh1 (CPSF)	P-CTD	Dichtl et al. 2002b
Ssu72	Ser5-P	Ganem et al. 2003; Krishnamurthy et al. 2004; Xiang et al. 2010; Werner-Allen et al. 2011
Ess1/Pin1	Ser5-P	Yaffe et al. 1997; Verdecia et al. 2000
Pcf11 (CF II)	Ser2-P	Barilla et al. 2001; Licatalosi et al. 2002; Meinhart and Cramer 2004
Rtt103 (termination factor)	Ser2-P	Kim et al. 2004b; Lunde et al. 2010
Sen1 (termination factor)	Ser2-P	Ursic et al. 2004; Chinchilla et al. 2012
Nrd1 (termination factor)	Ser5-P	Conrad et al. 2000; Steinmetz et al. 2001; Vasiljeva et al. 2008

Table 1. CTD binding proteins. The CTD binding preference of factors involved in transcription elongation, RNA processing and termination. The multi-functional protein complex PSF/p54 and CstF50 can bind to either unphosphorylated CTD or phosphorylated CTD (P-CTD). Pcf11, Rtt103, and Nrd1 bind to the CTD phosphorylated on Ser2 or Ser5 through a conserved CID.

Figure 1.

S. Cerevisiae
(budding yeast)

FSPTSPT
YSPTSPA
(YSPTSPS) 3~16
YSPTSPA
(YSPTSPS) 18~21
YSPTSPN
(YSPTSPS) 23
YSPTSPG
YSPGSPA
(YSPKQDE) 26
QKHNENENSR

S. Pombe
(fission yeast)

YGLTSPS
YSPSSPG
YSTSPA
YMPSSPS
(YSPTSPS) 5~8
YSATSPS
(YSPTSPS) 10~29

D. rerio
(zebrafish)

YSPTSPA
YEPRSPGGG
YTPQSPG
(YSPTSPS) 4~5
YSPTSPN
(YSPTSPS) 7~21
(YSPTSPS) 22
YSPTSPN
YTPTSPS
(YSPTSPS) 25
(YSPTSPS) 26
YSPTSPN
YTPTSPN
(YSPTSPS) 29~30
YSPSSPR
YTPQSPT
YTPSSPS
YSPSSPS
YSPTSPK
YTPTSPS
YSPSSPE
YTPTSPK
YSPTSPK
YSPTSPK
YSPTSPT
YSPTTPK
YSPTSPT
YSPTSPT
YTPTSPK
YSPTSPT
YSPTSPK
YSPTSPT
YSPTSPKGST
YSPTSPG
YSPTSPT
(YSPA- - -) 52
ISPDDSDEENN

H. Sapiens
(human)

YSPTSPA
YEPRSPGG
YTPQSPS
(YSPTSPS) 4~5
YSPTSPN
(YSPTSPS) 7~21
YSPTSPN
YSPTSPN
YTPTSPS
(YSPTSPS) 25
YSPTSPN
YTPTSPN
(YSPTSPS) 28
(YSPTSPS) 29~30
YSPSSPR
YTPQSPT
YTPSSPS
YSPSSPS
YSPTSPK
YTPTSPS
YSPSSPE
YTPTSPK
YSPTSPK
YSPTSPK
YSPTSPT
YSPTTPK
YSPTSPT
YSPTSPV
YTPTSPK
YSPTSPT
YSPTSPK
YSPTSPT
YSPTSPKGST
YSPTSPG
YSPTSPT
(YSLTSPA) 52
ISPDDSDEEN

Figure 2.

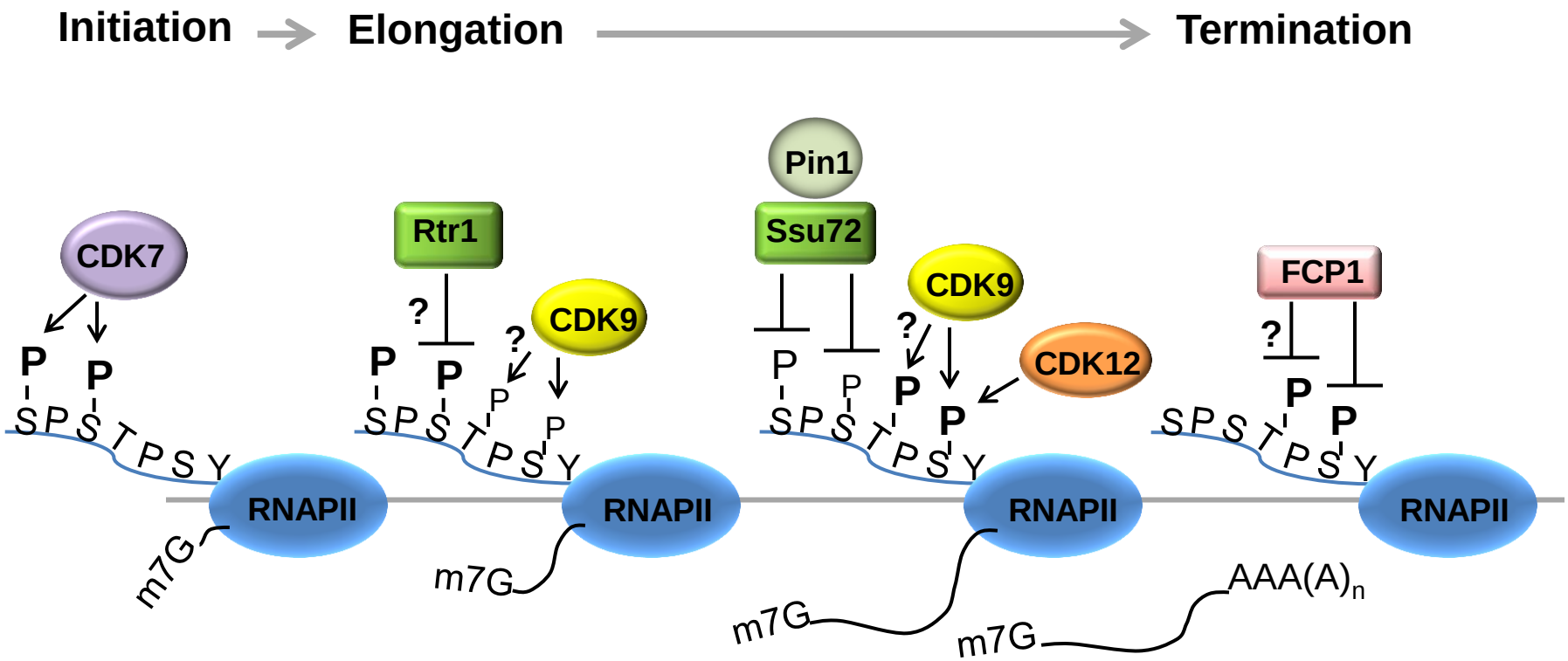
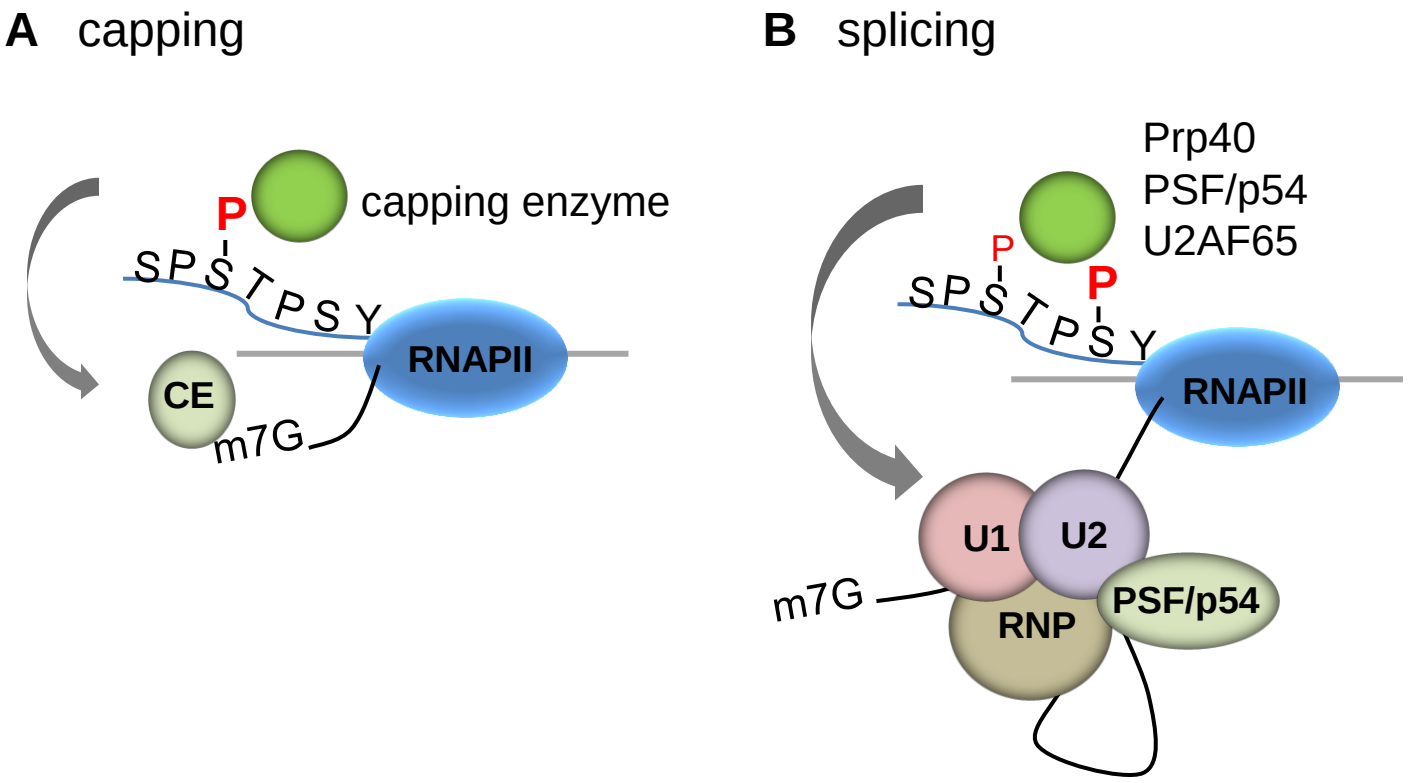

Initiation
Elongation
Termination
CDK7
Rtr1
CDK9
Pin1
Ssu72
CDK9
CDK12
FCP1
RNAPII
RNAPII
RNAPII
RNAPII
m7G
m7G
m7G
AAA(A)n

Figure 3.


A capping
P
capping enzyme
S P S T P S Y
RNAPII
CE
m7G
B splicing
Prp40
PSF/p54
U2AF65
P
P
S P S T P S Y
RNAPII
U1
U2
m7G
RNP
PSF/p54

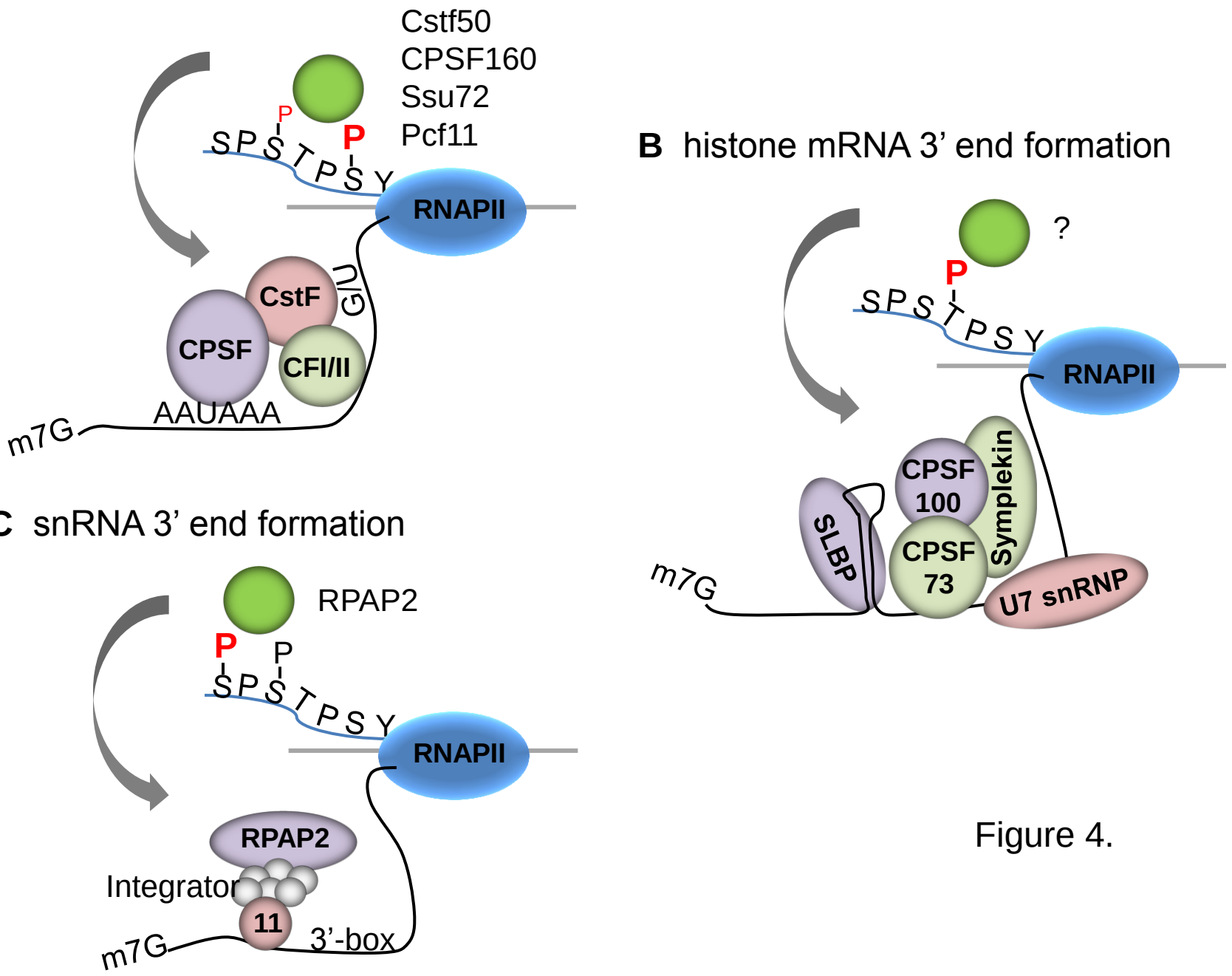
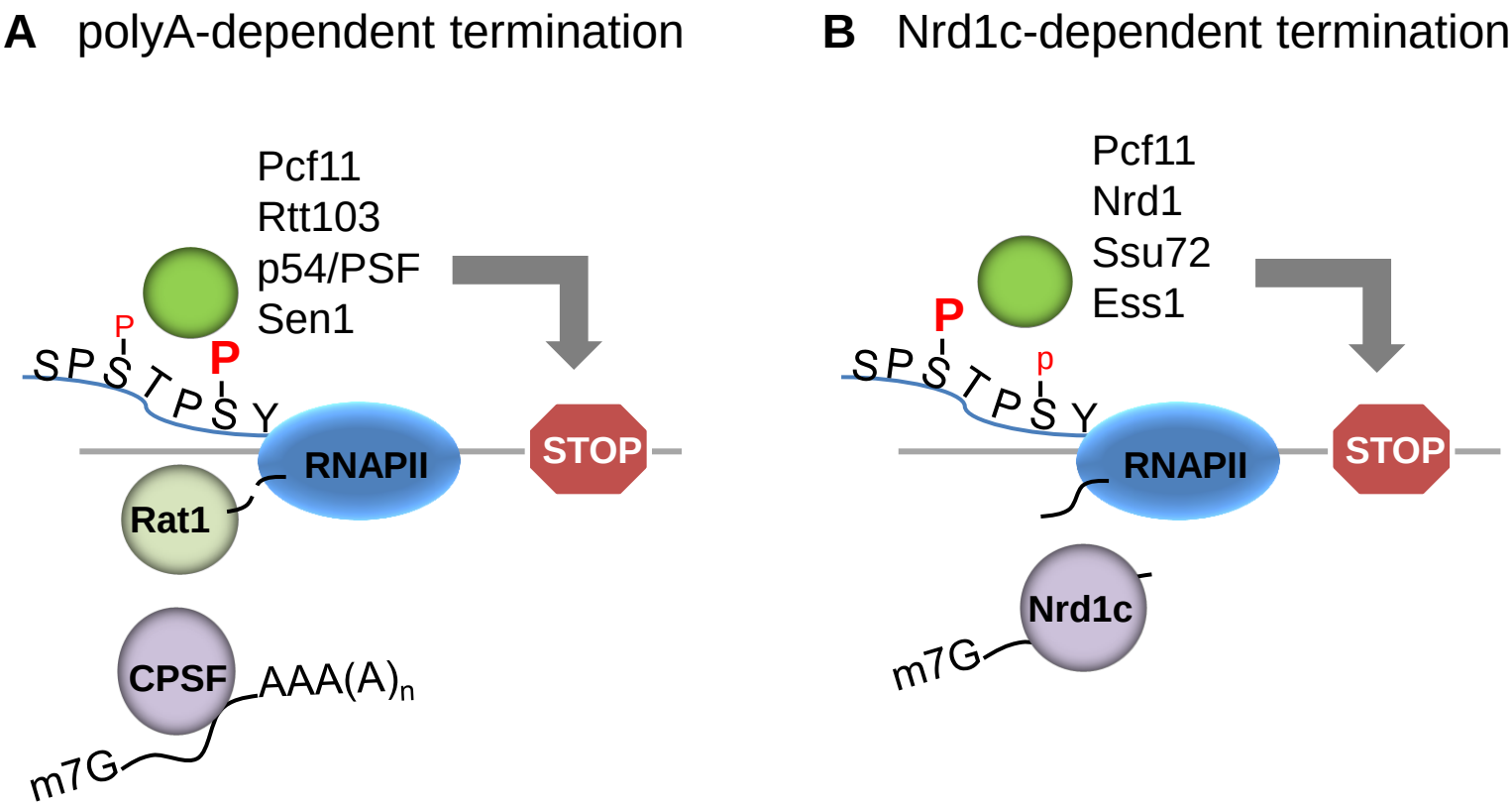

A polyA mRNA 3' end formation
Cstf50
CPSF160
Ssu72
Pcf11
P
P
S P S T P S Y
RNAPII
CstF
CPSF
CFI/II
GU
AAUAAA
m7G
B histone mRNA 3' end formation
?
P
S P S T P S Y
RNAPII
CPSF 100
Symplekin
SLBP
CPSF 73
U7 snRNP
m7G
C snRNA 3' end formation
RPAP2
P
P
S P S T P S Y
RNAPII
RPAP2
Integrator
11
3'-box
m7G

Figure 4.

Figure 5.


A polyA-dependent termination
Pcf11
Rtt103
p54/PSF
Sen1
P
P
S P S T P S Y
RNAPII
STOP
Rat1
CPSF
AAA(A)n
m7G
B Nrd1c-dependent termination
Pcf11
Nrd1
Ssu72
Ess1
P
p
S P S T P S Y
RNAPII
STOP
Nrd1c
m7G

Chapter 2

RNAP II CTD Phosphorylated on Threonine 4 Is Required for Histone mRNA 3' end Processing

Jing-Ping Hsin, Amit Sheth, James L. Manley

Department of Biological Sciences

Columbia University

New York, NY 10027, USA

ABSTRACT

The RNA polymerase II (RNAP II) largest subunit contains a C-terminal domain (CTD) with up to 52 Tyr^{1}-Ser^{2}-Pro^{3}-Thr^{4}-Ser^{5}-Pro^{6}-Ser^{7} consensus repeats. Serines 2, 5 and 7 are known to be phosphorylated, and these modifications help to orchestrate the interplay between transcription and processing of messenger RNA (mRNA) precursors. Here, we provide evidence that phosphorylation of CTD Thr^{4} residues is required specifically for histone mRNA 3' end processing, functioning to facilitate recruitment of 3' processing factors to histone genes. Like Ser^{2}, Thr^{4} phosphorylation requires the CTD kinase CDK9 and is evolutionarily conserved from yeast to human. Our data thus illustrate how a CTD modification can play a highly specific role in facilitating efficient gene expression.

The carboxyl-terminal domain (CTD) of the RNA polymerase II (RNAP II) largest subunit (Rpb1) consists of Tyr^1-Ser^2-Pro^3-Thr^4-Ser^5-Pro^6-Ser^7 (YSPTSPS) consensus tandem repeats, conserved from yeast to human. The CTD, through phosphorylation on serine (ser) residues, links transcription to mRNA processing events (*1-3*). Ser^5 is phosphorylated by cyclin-dependent kinase 7, CDK7, a subunit of the general transcription factor TFIIH, and this modification functions to facilitate capping (*4-5*). During transcriptional elongation, Ser^2 is phosphorylated by CDK9/P-TEFb, which helps to coordinate RNA 3' end processing and transcription termination (*6-7*). Ser^7 phosphorylation has been implicated in transcription and 3' end processing of genes encoding certain small non-coding RNAs (*8, 9*). Tyr^1 can also be phosphorylated, by the c-Abl kinase (*10*). While Thr^4 has been reported to be phosphorylated in fission yeast (*11*), there is no evidence that this residue is modified in other species, or what might the function of Thr^4 be.

To investigate CTD function in a genetically tractable vertebrate cell system, we created an Rpb1 conditional knock-out chicken cell line, DT40-Rpb1, using methods developed previously to study other conserved proteins (*12, 13*; Fig. S1). These cells express, as the only source of Rpb1, a tetracycline (tet)-repressible cDNA encoding HA-tagged human Rpb1 (human and chicken Rpb1 are 97% identical, and the CTD is very highly conserved among vertebrates; Fig. S2). After addition of tet, Rpb1 became undetectable between 12 and 18 hours (Figs. 1A and S3), and DT40-Rpb1 cells stopped growing and began to die by 24 hours (Fig. 1B). An important use of DT40 derivatives such as DT40-Rpb1 is for genetic complementation (e.g., *14*). Here, a plasmid expressing FLAG-tagged wild-type (wt) Rpb1 was found to restore viability to DT40-Rpb1 cells in the presence of tet. Expression of an Rpb1 derivative (26r) containing 26 all-consensus repeats, plus the ten C-terminal most residues that confer stability (*15*), also restored viability. However, a comparable derivative containing 30 repeats with all Thr^4 residues mutated to Val (T4V) was inviable. All three Rpb1 derivatives were expressed equivalently (Fig. 1C), and

wt and 26r-expressing cells grew similarly following addition of tet, while T4V cells stopped growing and began dying by 24 hours (Fig. 1D).

We next investigated the possible basis for T4V cell inviability. Comparison of overall transcription (by ^{3}H pulse labeling) in wt, 26r and TV4 cells treated with tet for 40 hours (Fig. 2A) revealed only a modest decrease (20%) in polyadenylated [poly $(A)^{+}$] and poly $(A)^{-}$ RNA fractions in T4V compared to 26r cells, indicating that total transcription and synthesis of poly $(A)^{+}$ mRNA was not significantly affected by the T4V substitutions. A 20-min induction of the inducible *Egr1* gene (*16*) in T4V and 26r cells, grown in tet for 24 hours, resulted in an equivalent ~300-fold mRNA increase in both (Fig. 2B). Additionally, no differences were observed in unspliced/spliced and uncleaved/total (a measure of 3' processing) Egr1 mRNA (Fig. S4) and 3' processing of several additional poly $(A)^{+}$ mRNAs was similar in T4V and 26r cells (Fig. S5). Thus, transcription and processing of poly $(A)^{+}$ mRNAs was not significantly affected in T4V cells.

We next measured accumulation of several types of RNAP II products in T4V and 26r cells treated with tet for 40 hours (Fig. 2C). Levels of U1 snRNA were unaffected, as were levels of two poly $(A)^{+}$ mRNAs, RplPl and GAPDH. However, levels of all four non-poly $(A)^{+}$ replication-dependent histone mRNAs were significantly reduced in two independent T4V cell lines, to ~15% the levels observed in 26r cells. Consistent with this, histone H3 protein levels were reduced in T4V cells ~50% after 30 or 40 hours tet treatment (Fig. 2D). Levels of other proteins tested were not significantly affected (Figs. 2D and S6).

Decreased histone gene transcription could explain the defect in histone mRNA accumulation in T4V cells. However, nuclear run-on analysis showed no differences, or even a slight increase, with three histone genes in 26r and T4V cells after 24 or 40 hours tet treatment (Fig. S7A and B). We also performed chromatin immunoprecipitation (ChIP) with anti-Rpb1 antibodies (N20) in T4V and 26r cells treated with tet for 30 hours. RNAP II levels at the transcription start site (TSS) and 3' end of *H2A*, *H2B* and *H3* genes were not reduced, and indeed slightly elevated at the TSS, in T4V cells (Fig. 3A, left).

Similar results were observed along the length of the *RPLP1* gene (Fig. 3A, right). We conclude that histone gene transcription was unaffected in T4V mutant cells.

The role of the CTD in 3' end formation suggested that Thr^4 might be required specifically in histone mRNA 3' formation, which uses a distinct mechanism from poly $(A)^+$ mRNAs (*17*). When histone 3' processing is defective, cryptic downstream polyadenylation sites can be used, resulting in formation of poly $(A)^+$ histone mRNAs (*18*). 26r and T4V cells were grown in the presence of tet for 40 hours, and for all four histone genes, ~two fold more poly $(A)^+$ mRNA was recovered from T4V cells than from 26r cells (Fig. 3B, top), and the fraction of histone mRNA in these cells that was poly $(A)^+$ increased 10-50 fold (Fig. 3B, bottom).

Histone pre-mRNA 3' processing requires a complex set of evolutionarily conserved factors, only some of which are shared with the poly $(A)^+$ mRNA 3' processing machinery (*17*). Using ChIP, we examined recruitment of one common (CPSF-100) and one histone-specific (SLBP) factor to several histone genes and to the *RPLP1* gene. CPSF-100 levels on the *RPLP1* gene peaked at the promoter, consistent with previous observations (e.g., *19*), and were equivalent in 26r and T4V cells treated with tet for 30 hours (Fig. 3C, right). On the three histone genes tested, CPSF-100 levels were slightly reduced at the TSS in T4V cells but significantly lower at the 3' end (Fig. 3C, left), consistent with the defect in 3' processing. As expected, in 26r cells SLBP was detected at background levels on the *RPLP1* gene (Fig. 3D), while significant levels were found associated with the two histone genes tested. However, SLBP levels were significantly reduced both at the TSS and at the 3' processing site in T4V cells (Fig. 3D). Total levels of SLBP, CPSF-100 and other 3' processing factors tested were equivalent in the two cell lines (Figs. 2D and S6).

The Thr to Val mutation is a conservative change, and our results thus suggest that modification (e.g., phosphorylation) of Thr^4 is important for its function. To investigate whether Thr^4 is phosphorylated, we used an antibody raised against and specific for a Thr^4-phosphorylated CTD heptad

repeat (P-Thr4; fig. S8). IP-Westerns with extracts from 26r and T4V cells, using anti-FLAG for IP and P-Thr4 for Western (Fig. 4A) detected a band corresponding to a protein of the expected size in both input and IP from the 26r cells, but this species was detected only at most at very low levels in the T4V samples, indicating that Thr^4 is phosphorylated in the 26r CTD. Thr^4 is highly conserved throughout eukaryotes, and Western analysis with P-Thr4 of lysates from yeast, fly (KC), 26r and human (HeLa) cells revealed a band the size expected for Thr^4 phosphorylated Rpb1 in all samples (Fig. 4B).

To identity the possible kinase responsible for Thr^4 phosphorylation, we examined whether Thr^4 phosphorylation in DT40 cells was sensitive to the CDK9/P-TEFb inhibitors DRB and flavopiridol (*20*), and found that it was strongly inhibited by both (Figs. 4C and S9). This did not reflect a requirement for prior Ser^2 phosphorylation, as Thr^4 phosphorylation was detected in DT40-Rpb1 cells expressing an Rpb1 derivative with all Ser^2 residues mutated to Ala (S2A) (Fig. 4D). Linking this to the defect in histone mRNA processing in T4V cells, histone mRNA levels in flavopiridol-treated 26r cells were reduced to levels found in T4V cells (Fig. S10). Consistent with this, knockdown of CDK9 in HEK293 cells was previously shown to impair SLBP recruitment to histone genes and lead to accumulation of poly $(A)^+$ histone mRNA (*21*).

Our experiments have provided new insight into the intricate mechanisms used by cells to couple transcription by RNAP II to subsequent RNA processing. While our data provide evidence that histone mRNA 3' end formation specifically requires Thr^4 phosphorylation, it is likely that this modification is important for other CTD functions. Although our experiments have not uncovered evidence for this, the fact that Thr^4 is phosphorylated in yeast, which produces histone mRNA 3' ends by the same mechanism as other mRNAs, suggests the existence of additional functions in mRNA synthesis and/or processing.

References and Notes

1. Y. Hirose, J. L. Manley, RNA polymerase II and the integration of nuclear events. *Genes & Development* **14**, 1415 (2000).

2. T. Maniatis, R. Reed, An extensive network of coupling among gene expression machines. *Nature* **416**, 499 (2002).

3. M. J. Moore, N. J. Proudfoot, Pre-mRNA processing reaches back to transcription and ahead to translation. *Cell* **136**, 688 (2009).

4. P. Komarnitsky, E. J. Cho, S. Buratowski, Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. *Genes & Development* **14**, 2452 (2000).

5. S. C. Schroeder, B. Schwer, S. Shuman, D. L. Bentley, Dynamic association of capping enzymes with transcribing RNA polymerase II. *Genes & Development* **14**, 2435 (2000).

6. S. H. Ahn, M. Kim, S. Buratowski, Phosphorylation of serine 2 within the RNA polymerase II C-terminal domain couples transcription and 3' end processing. *Molecular Cell* **13**, 67 (2004).

7. Z. Ni, B. E. Schwartz, J. Werner, J.-R. Suarez, J. T. Lis, Coordination of transcription, RNA processing, and surveillance by P-TEFb kinase on heat shock genes. *Molecular Cell* **13**, 55 (2004).

8. R. D. Chapman *et al.*, Transcribing RNA polymerase II is phosphorylated at CTD residue serine-7. *Science* **318**, 1780 (2007).

9. S. Egloff *et al.*, Serine-7 of the RNA polymerase II CTD is specifically required for snRNA gene expression. *Science* **318**, 1777 (2007).

10. R. Baskaran, M. E. Dahmus, J. Y. J. Wang, Tyrosine phosphorylation of mammalian RNA polymerase II carboxyl-terminal domain. *Proc. Natl. Acad. Sci. U.S.A.* **90**, 11167 (1993).

11. H. Sakurai, A. Ishihama, Level of the RNA polymerase II in the fission yeast stays constant but phosphorylation of its carboxyl terminal domain varies depending on the phase and rate of cell growth. *Genes to Cells* **7**, 273 (2002).

12. J. Wang, Y. Takagaki, J. L. Manley, Targeted disruption of an essential vertebrate gene: ASF/SF2 is required for cell viability. *Genes & Development* **10**, 2588 (1996).

13. Y. Takagaki, J. L. Manley, Levels of polyadenylation factor CstF-64 control IgM heavy chain mRNA accumulation and other events associated with B cell differentiation. *Molecular Cell* **2**, 761 (1998).

14. J. Wang, S. H. Xiao, J. L. Manley, Genetic analysis of the SR protein ASF/SF2: interchangeability of RS domains and negative control of splicing. *Genes & Development* **12**, 2222 (1998).

15. R. D. Chapman, M. Conrad, D. Eick, Role of the mammalian RNA polymerase II C-terminal domain (CTD) nonconsensus repeats in CTD stability and cell proliferation. *Molecular and Cellular Biology* **25**, 7665 (2005).

16. C. Y. Chen, L. W. Forman, D. V. Faller, Calcium-dependent immediate-early gene induction in lymphocytes is negatively regulated by $p21^{Ha\text{-}ras}$. *Molecular and Cellular Biology* **16**, 6582 (1996).

17. W. F. Marzluff, E. J. Wagner, R. J. Duronio, Metabolism and regulation of canonical histone mRNAs: life without a poly(A) tail. *Nat. Rev. Genet.* **9**, 843 (2008).

18. K. D. Sullivan, T. E. Mullen, W. F. Marzluff, E. J. Wagner, Knockdown of SLBP results in nuclear retention of histone mRNA. *Rna* **15**, 459 (2009).

19. O. Rozenblatt-Rosen *et al.*, The tumor suppressor Cdc73 functionally associates with CPSF and CstF 3' mRNA processing factors. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 755 (2009).

20. S. H. Chao, D. H. Price, Flavopiridol inactivates P-TEFb and blocks most RNA polymerase II transcription in vivo. *The Journal of Biological Chemistry* **276**, 31793 (2001).

21. J. Pirngruber *et al.*, CDK9 directs H2B monoubiquitination and controls replication-dependent histone mRNA 3'-end processing. *EMBO reports* **10**, 894 (2009).

Acknowledgement: We thank T. Kashima for plasmids and help with construction of the DT40-Rpb1 cell line, W. Marzluff for the SLBP antibody, Z. Dominski for the Lsm11 antibody, and Novus Biologicals for the P-Thr4 antibody. This work was supported by NIH grant R01 GM28983.

FIGURE LEGENDS

Figure 1. Growth properties of DT40-Rpb1 cells and derivatives.

(A) Western blot of lysates from DT40, DT40 *rpb1*(+/-) and DT40-Rpb1 cells treated with tet as indicated. Blots were probed with anti-Rpb1 (N20) and anti-actin antibodies. (B) Growth curves of DT40, DT40-*rpb1*(+/-) and DT40-Rpb1 cells, in the presence or absence of tet. N=2. (C) Western blot of lysates of DT40-Rpb1 cells expressing the indicated Flag-tagged Rpb1 derivative and treated with tet for 24 hours. (D) Growth curves of wt, 26r and T4V cells grown in the presence of tet for 4 days. Cells were split after two days. N=3. Error bars indicate standard deviation.

Figure 2. T4V cells are defective in histone expression.

(A) Wt, 26r and T4V cells were treated with tet for 40 hours, RNA was labeled for 30 min, purified and poly $(A)^+$ and poly $(A)^-$ fractions quantitated by scintillation counting. Counts per minute (CPM) relative to 26r are shown. N=3. (B) T4V and 26r cells were treated with tet for 24 hours before induction with ionomycin and PMA for 20 min. EGR1 mRNA levels were determined by RT-qPCR following normalization to 18S RNA levels. N=3. (C) T4V (two independent lines) and 26r cells were treated with tet for 40 hours. RNA was isolated and indicated RNAs quantitated by RT-qPCR, normalized to 5S RNA levels and displayed relative to 26r. N=3. (D) T4V and 26r cells were treated with tet for 30 or 40 hours. Cell lysates were analyzed by Western blotting (top). Ratios of H3/actin and H3/GAPDH were determined and are displayed relative to values obtained with 26r. N=3. Error bars display standard deviation.

Figure 3. T4V cells are defective in histone mRNA 3' end processing.

(A) Rpb1 CHIP. Diagrams depict genes analyzed. Thick lines represent genes and dashed lines, transcripts. Triangles denote 3' cleavage sites. Positions of amplicons are indicated. T4V and 26r cells were grown in the presence of tet for 30 hours, then used for ChIP. DNA samples were analyzed by qPCR. Signals were normalized as described in Methods. N=3. (B) T4V and 26r cells were treated with tet for 40 hours, RNA isolated and RT performed using oligo-dT or random hexamer oligos. qPCR was performed, values were normalized to Rplp1 mRNA levels and are shown relative to 26r (top). The ratio of poly $(A)^+$/total histone transcripts was calculated and plotted (bottom). N=3. (C) CPSF100 CHIP. N=5. (D) SLBP CHIP. N=3. Error bars indicate standard deviation.

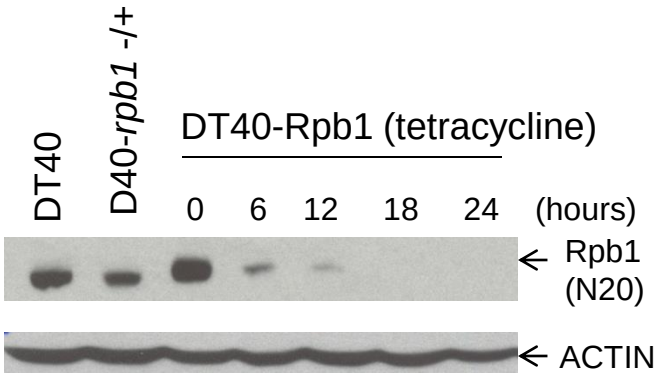
Figure 4. Thr^4 is phosphorylated throughout eukaryotes and requires CDK 9.

(A) T4V and 26r cells were treated with tet for 30 hours. Western blotting of input and Flag IP samples was done with P-Thr4 (right) and N20 (left) antibodies. (B) Total cell lysates from yeast, KC, 26r and Hela cells were prepared and used for Western blotting. Total Rpb1 was detected by 8WG16 (left) and Thr^4-

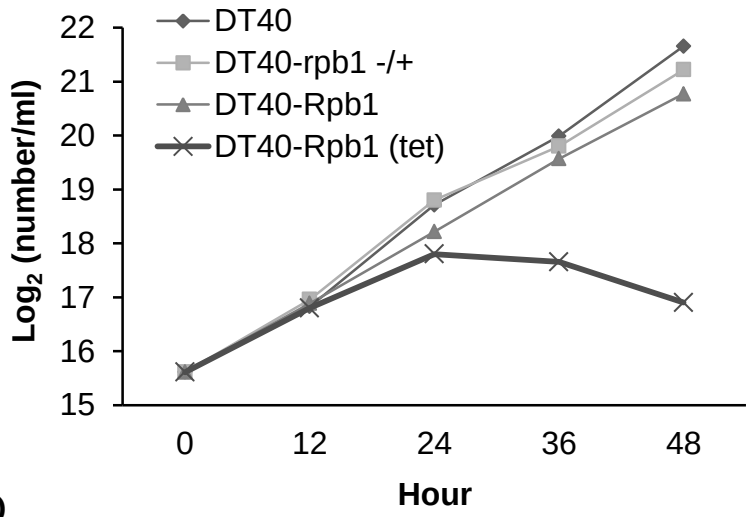
phosphorylated Rpb1 with P-Thr4 (right). Band indicated with asterisk is a cross-reacting protein. (C) DT40 cells were treated with tet for 24 hours, the indicated concentrations of flavopiridol were added and cells grown for an additional 6 hours. Cell lysates were analyzed by Western blotting. (D) Lysates from 26r and S2A cells treated with tet for 30 hours were analyzed by Western blotting.

Figure 1.

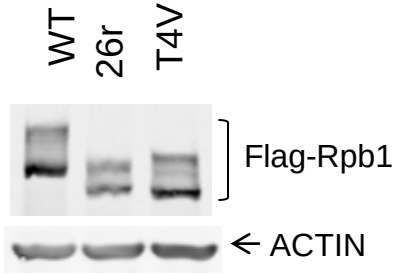
A



B



C



D

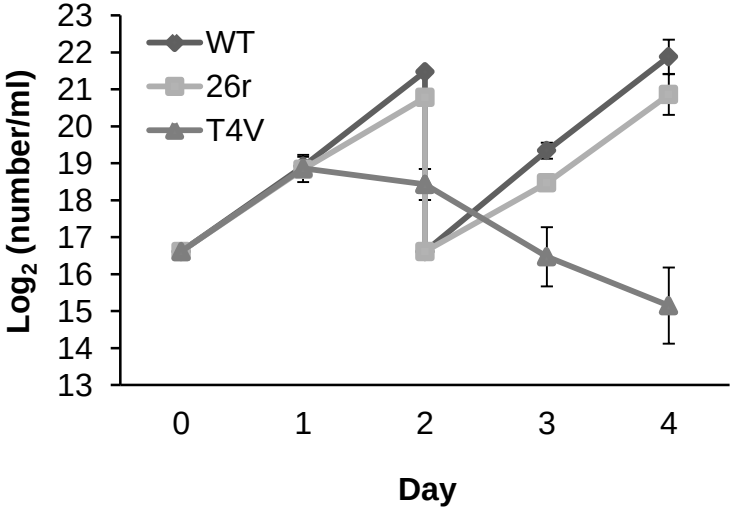
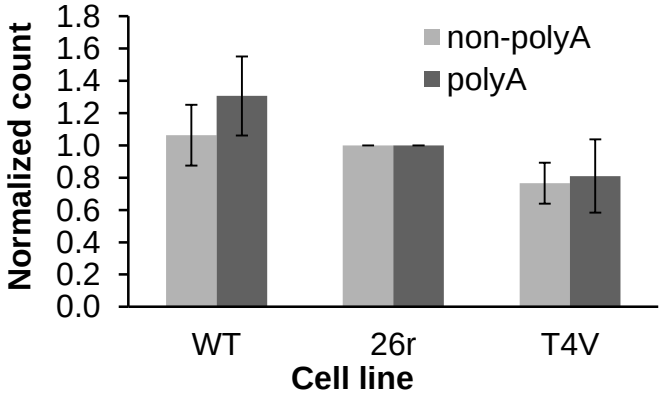
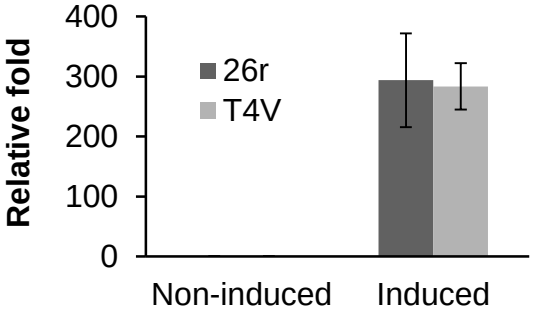


Figure 2.

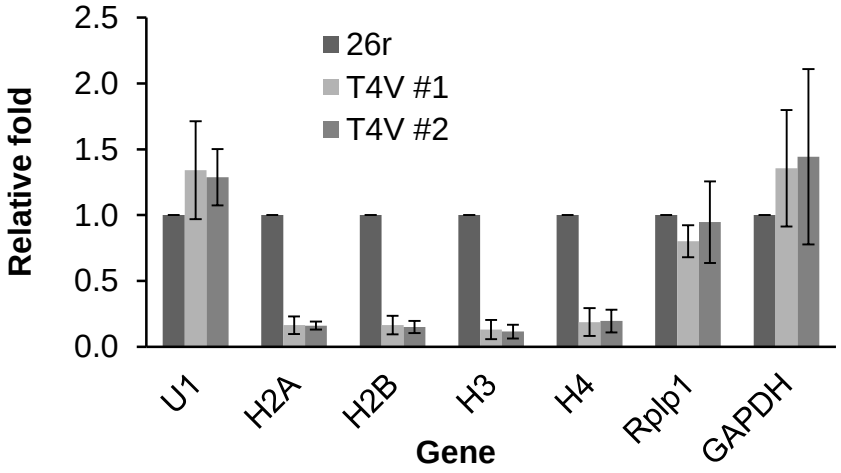
A



B



C



D

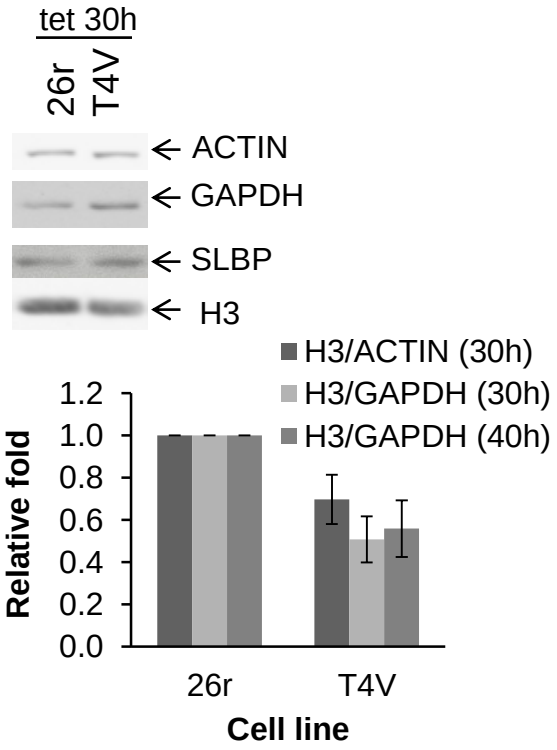


Figure 3.

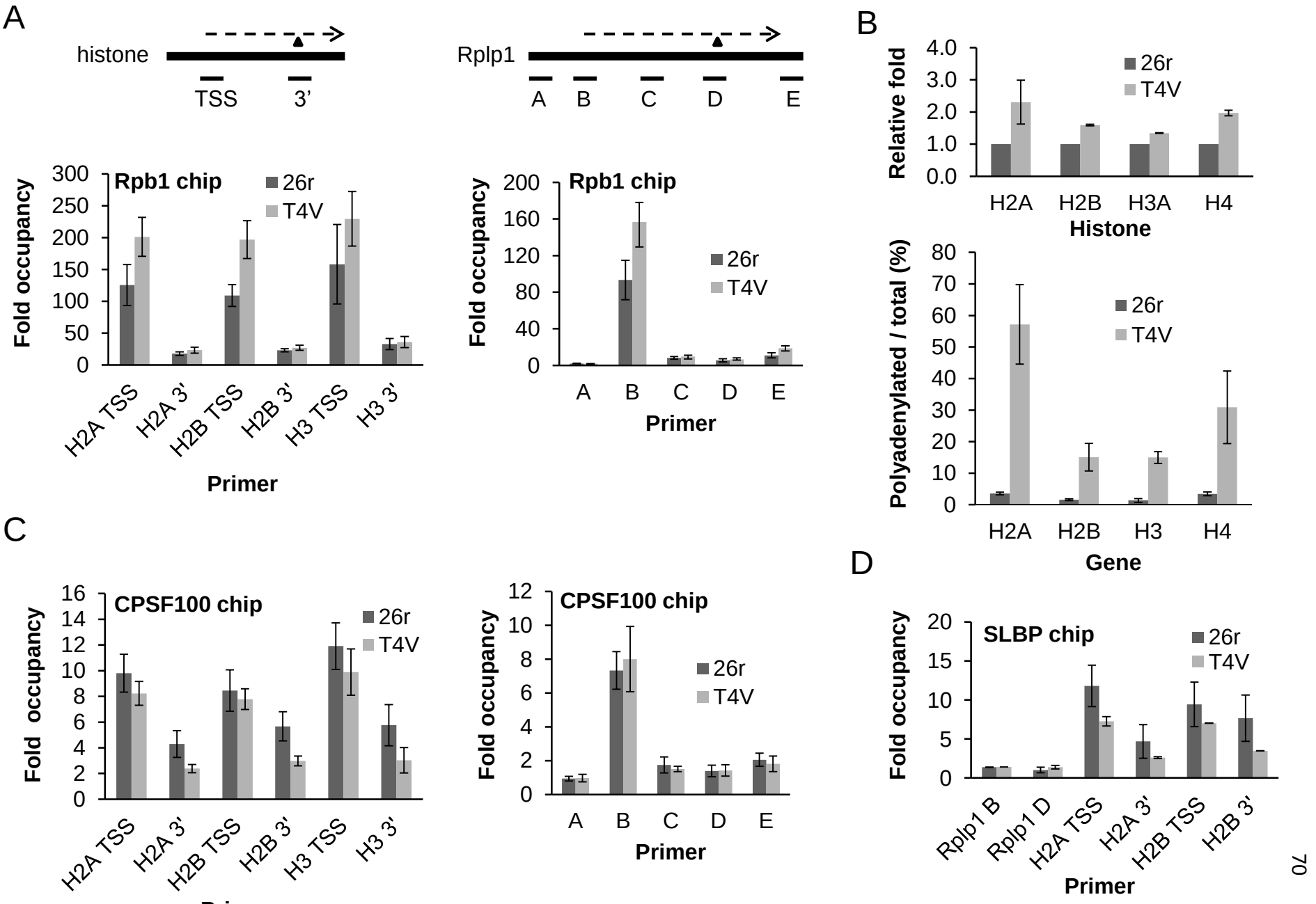
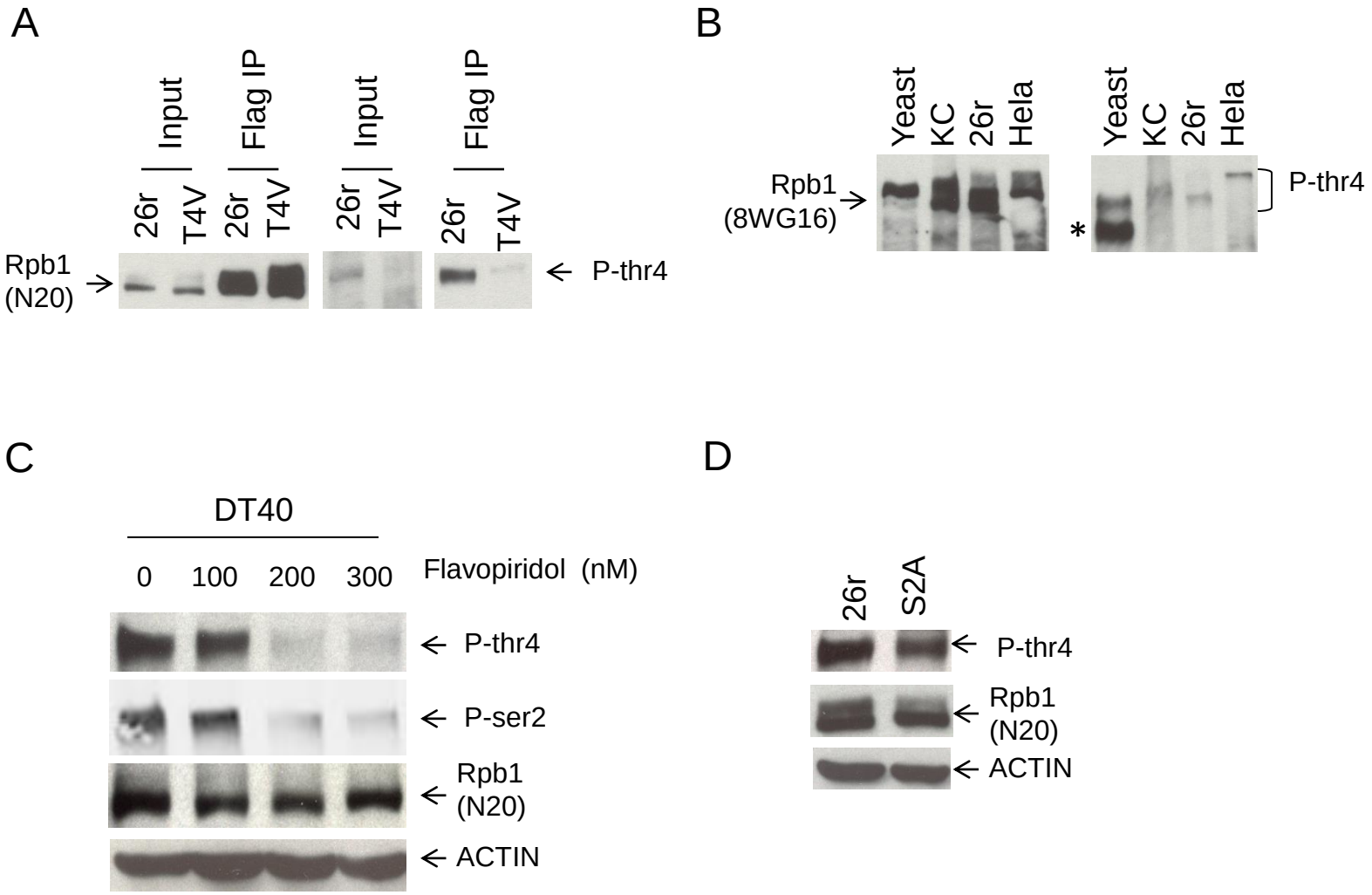

A
histone
TSS
3'
Rplp1
A
B
C
D
E
Rpb1 chip
26r
T4V
Fold occupancy
300
250
200
150
100
50
0
H2A TSS
H2A 3'
H2B TSS
H2B 3'
H3 TSS
H3 3'
Primer
Rpb1 chip
26r
T4V
Fold occupancy
200
160
120
80
40
0
A
B
C
D
E
Primer
B
Relative fold
4.0
3.0
2.0
1.0
0.0
26r
T4V
H2A
H2B
H3A
H4
Histone
Polyadenylated / total (%)
80
70
60
50
40
30
20
10
0
26r
T4V
H2A
H2B
H3
H4
Gene
C
Fold occupancy
16
14
12
10
8
6
4
2
0
CPSF100 chip
26r
T4V
H2A TSS
H2A 3'
H2B TSS
H2B 3'
H3 TSS
H3 3'
Primer
CPSF100 chip
Fold occupancy
12
10
8
6
4
2
0
26r
T4V
A
B
C
D
E
Primer
D
Fold occupancy
20
15
10
5
0
SLBP chip
26r
T4V
Rplp1 B
Rplp1 D
H2A TSS
H2A 3'
H2B TSS
H2B 3'
Primer

Figure 4.


A
Input
Flag IP
Input
Flag IP
26r
T4V
26r
T4V
26r
T4V
26r
T4V
Rpb1
(N20)
P-thr4
B
Yeast
KC
26r
Hela
Yeast
KC
26r
Hela
Rpb1
(8WG16)
*
P-thr4
C
DT40
0
100
200
300
Flavopiridol (nM)
P-thr4
P-ser2
Rpb1
(N20)
ACTIN
D
26r
S2A
P-thr4
Rpb1
(N20)
ACTIN

Supplemental material for

RNA Pol II CTD Phosphorylated on Threonine 4 Is Required for Histone mRNA 3' end Processing

This document includes:

Materials and Methods

Cell culture and cloning

DT40 cells and derivatives were cultured in RPMI1640 medium containing 10% FBS and 1% chicken serum at 37^{o}C with 5% CO_2.

Targeting vectors: The genomic PCR product containing chicken Rpb1 exon 11, intron 11, and exon 12 (11F2/12R) was inserted into pBlueScript KS by EcoRV to build pKS-11F2/12R. A second genomic PCR product containing chicken Rpb1 intron 9, and exon 10 was inserted into pBlueScript KS by EcoRV to build pKS-9/10R4. Next, a hygromycin- or zeocin-resistance gene driven by the chicken β-actin promoter was inserted into pKS-11F2/12R to obtain pKS-hygromycin-11F2/12R and pKS-zeocin-11F2/12R. The pKS-9/10R4 was then cut by HindIII (filled in by Klenow) and EagI. The fragment containing intron 9 and exon 10 from pKS-9/10R4 was inserted upstream of the β-actin promoter of pKS-hygromycin-11F2/12R and pKS-zeocin-11F2/12R to create pKS-9/10R4-hygromycin-11F2/12R and pKS-9/10R4-zeocin-11F2/12R, which were later linearized by XhoI for electroporation.

HA-tagged Rpb1: The pTRE-HA (Clontech), modified by mutating the ccatg sequence after the HA-tag to cc, was used to construct conditionally expressed Rpb1. Human Rpb1 was cloned into pTRE-HA. The TRE-HA-Rpb1 (TRE promoter, HA tag, and Rpb1) fragment was excised and cloned into pPGK-HisD (a gift from Tsuyoshi Kashima), which contains a histidinol-resistance gene under the PGK promoter. pTRE-HA-Rpb1-HisD was linearized by RsrII for electroporation.

The pTRE-puro-CMV-tTA plasmid: The tTA fragment from ptTA (a gift from T. Kashima) was used to replace the tetR-SP1 fragment of the tTA-SP1 plasmid (a gift from A. Kornblihtt) to obtain CMV-tTA. The puromycin-resistance gene was cloned into pTRE2 (a gift from T. Kashima), which contains the TRE promoter and GH poly-A signal. A fragment containing TRE, puromycin-resistance gene, and GH poly-A

signal was then inserted into the CMV-tTA plasmid to obtain pTRE-puro-CMV-tTA, which was later linearized by HindIII for electroporation.

Rpb1 derivatives with mutant CTDs: The chicken beta-actin promoter from pAZeo plasmid (a gift from T. Kashima) was ligated with a FLAG-tag sequence from p3XFLAG-CMV7 (SIGMA). The ligated fragment of beta-actin and FLAG was inserted into pBlueScript. The Neomycin resistance gene from pNeo (*1*) was then inserted into the plasmid to obtain Actin-Flag-Neo. The human *RPB1* body without the CTD was inserted between the FLAG tag and Neomycin resistance gene. Various CTD fragments obtained by PCR or by ligating oligos to form concatemers (*2*) were inserted directly 3' to the *RPB1* body. For electroporation, plasmids were linearized with NotI. All constructs were verified by DNA sequencing. Primer and oligo sequences are available upon request.

Construction of Rpb1 conditional knock-out DT40 cells

Procedures for establishing *Rpb1* knock-out cell lines were essentially as described (*1, 3*). The first *RPB1* allele was disrupted using the pKS-9/10R4-hygromycin-11F2/12R targeting vector. A plasmid encoding HA-tagged human Rpb1 under the control of the tet-repressible CMV promoter was used to rescue the Rpb1 knock-out cells. After pTRE-puro-CMV-tTA and pTRE-HA-Rpb1-HisD were introduced into the heterozygous *Rpb1* knock-out cells, the second *RPB1* allele was disrupted with pKS-9/10R4-zeocin-11F2/12R. Southern blotting and PCR were performed to confirm disruption of the *Rpb1* alleles.

Southern blotting

Ten ug of genomic DNA was digested with BamH1, separated on an 0.7% agarose gel, and processed for southern blotting as described (*1*). Probes specific to exon 9 of chicken Rpb1 were made using Rediprime kit (Amersham). Hybridization was carried out at 65^{o}C overnight. Membranes were washed under high stringency.

Complementation assays and construction of stable cell lines

About 10^7 cells were harvested, and washed once in cold PBS, then resuspended in 0.5 ml cold PBS and transferred to a cuvette (0.4 cm, Bio-Rad). About 15 ug of linearized DNA (in 50-100 ul PBS) were added, and the cuvette was incubated on ice for 10 min. After electroporation (550 V, 25u F with a Bio-Rad Gene Pulser II), cells were chilled on ice for 10 min, and allowed to recover in 20 ml of medium (without antibiotics). After 24 hours, cells were resuspended in 70 ml of medium containing appropriate antibiotics, plated into four 96-well plates (about 150 ul per well), and selected for 6-9 days. Surviving clones were picked and analyzed by western blotting. Concentration of antibiotics for selection: G418 (neomycin) 1.5 mg/ml, puromycin 0.25 ug/ml, hygromycin B 1.5 mg/ml, histidinol 0.5 mg/ml, zeocin 0.5 mg/ml, and tetracycline 1 ug/ml.

In vivo labeling of nascent RNA

$4x10^6$ cells were incubated with 200 uCi of ^{3}H-uridine (1 mCi/ml) for 30 min. Total RNA was extracted using Trizol (Invitrogen) and dissolved in 10 mM Tris-HCl (pH 8). For poly(A) selection, each RNA sample was divided into 3 equal parts to control for possible variations during selection. Selections were performed using oligo(dT) magnetic beads (Novagen), following procedures provided by the supplier. Unbound and eluted RNAs were collected for scintillation counting. Average counts from the three aliquots from each RNA sample were calculated and considered to be one independent value.

Nuclear run-on

$2x10^7$ cells were harvested, washed in cold PBS, and processed for nuclear run on as described (*4*) with the following modifications. Nuclear pellets were washed once in ice-cold NP-40-free lysis buffer and resuspended in 50 ul of ice-cold storage buffer. Nuclei were used immediately or were quick frozen on dry ice and stored at -80^{o}C. For one reaction, 25 ul nuclei were mixed with 25 ul 2X reaction buffer (10 mM Tris-HCl pH 8, 5 mM $MgCl_2$, 300 mM KCl, 5 mM DTT, 1 mM ATP, CTP, and GTP, 0.01 mM UTP, 100 uCi of α-^{32}P UTP (800 Ci/mmol) and incubated at 30^{o}C for 30 min. RNA was extracted and purified using

Trizol (Invitrogen). To facilitate RNA precipitation, one tenth volume of 3 M NaOAc and glycogen (~1 ug) were added prior to addition of isopropanol. RNA was dissolved in 20 ul formamide for slot blotting.

Slot blotting

5 ug samples of synthesized single-strand DNA oligo in 200 ul of 10X SSC were loaded onto Nytran N (0.2 um, moderately charged) filter. Membranes were washed with 400 ul 10X SSC, and DNA was cross-linked to the membrane by UV irradiation. Prior to hybridization, membranes were boiled in 0.1X SSC, 0.5% SDS for 5 min. RNAs from nuclear run-ons were denatured at 85^{o}C for 5 min. About 1-5x10^{6} cpm of labeled RNA was used in 500 ul hybridization buffer (1 mM EDTA, 0.5 M NaPO4 pH 7.2, 7% SDS). Hybridization was carried out at 65^{o}C overnight. Membranes were washed first with 1X SSC, 0.1% SDS at room temperature for 5 min, then with 1X SSC, 0.1% SDS at 45^{o}C for 15 min twice. DNA oligo sequences are available upon request.

RT-qPCR

RNA was extracted using Trizol (Invitrogen) and further treated with DNase I. Reverse transcription was performed using Maxima Reverse Transcriptase (Fermentas) and random hexamer or oligo-dT primers. cDNA was further diluted and analyzed by qPCR (ABI StepOnePlus). Cyber green PCR reagents for qPCR were from Fermentas. To analyze the relative mRNA level, the 2-ΔΔCt method was used (*5*). The target amplification efficiency of primers was determined, and all primers had amplification efficiencies >0.95. Primer sequences will be provided upon request.

Immunoprecipitation

About 3x10^{7} cells were collected, and washed with PBS. Then, 1 ml cold RIPA (150 mM NaCl, 1 mM EDTA, 50 mM Tris-HCl pH 7.4, 0.5% NP-40, 0.25% sodium deoxycholate, 50 mM NaF, 10 mM β-glycerolphosphate) buffer and 10 ul of protease inhibitors (140 ug/mL Pepstatin A, 35 ug/mL Leupeptin, and 170 ug/mL Aprotinin in 90% ethanol) was added. After brief sonication, samples were rotated at 4^{o}C

for 15 min. Debris was centrifuged at 12,000 g, 4^{o}C, for 15 min, and the supernatant was removed to a new tube. 50 ul of the lysate were kept for input control, and the rest of the extract was incubated with 20 ul of pre-washed protein G Sepharose and 1-4 ug of antibody. Samples were rotated at 4^{o}C for 2 hr, and beads were washed with cold RIPA buffer for 5 min three times, and then were resuspended in 100 ul of 1X SDS sample buffer for western blotting.

Chromatin immunoprecipitation assay (ChIP)

Cells were grown in the presence of tet (1ug/ml) for 30 hr to 70% confluence (2-3x10^{6}/ml), cross-linked with 1% formaldehyde for 10 min, and processed for ChIP as described (*6*) with the following modifications. Samples were sonicated until the length of chromatin fragments was in a range of 100-400 bp. Immunoprecipitation was carried out at 4^{o}C for 4 hr or overnight. Lysates of equal number of cells (1-2x10^{7}), 1-4 ug antibody, and 20 ul Sepharose G (GE Healthcare) were used per IP. After IP, beads were washed sequentially with WB1 (150 mM NaCl, 2 mM EDTA, 20 mM Tris-HCl (pH 8), 0.1% SDS, 1% Triton X-100), WB2 (500 mM NaCl, 2 mM EDTA, 20 mM Tris-HCl (pH 8), 0.1% SDS, 1% Triton X-100), WB3 (250 mM LiCl, 1 mM EDTA, 50 mM Hepes-KOH (pH 7.5), 0.7% sodium deoxycholate, 1% NP-40), and with TE-plus (1X TE, 50 mM NaCl), 5 min for each wash. DNA was eluted with 100 ul of elution buffer (50 mM Tris-HCl pH 8, 2 mM EDTA, 1%SDS, 200 ug/ml RNase A) at 65^{o}C for 4 hr or overnight. A PCR purification kit (QIAGEN) was used to purify DNA for analysis by qPCR (ABI StepOnePlus). Fold occupancy was calculated as described (*7*). The background signal was determined from an internal region of the 18S rRNA gene. Primer sequences will be provided upon request. ChIP was performed using antibodies against Rpb1 (N20; Santa Cruz), CPSF100 and SLBP (a gift from W. Marzluff).

Western blotting

Cell lysates were resolved in composite SDS-PAGE with two resolving layers, 6% and 12%. Western blotting was performed using standard protocols with antibodies against CTD (8WG16; Covance), actin (Sigma), CPSF73 (Bethyl Laboratories), CPSF100, Flag (M2) (Sigma), GAPDH (Sigma), histone H3 (Abcam),

HA (Covance), Lsm11 (a gift from Z. Dominski), mouse IgG (Sigma), mouse IgG (IRDye 800; Li-Cor), rabbit IgG (Sigma), rabbit IgG (IRDye 680; Li-Cor), rat IgG (IRDye 680; Li-Cor), Rpb1 (N20; Santa Cruz), phospho-ser2 CTD heptad (3E10; Millipore), phospho-thr4 CTD heptad (Novus #NBP1-49546), SLBP (a gift from W. Marzluff), and symplekin (Bethyl Laboratories). Western blots were quantified using ImageJ.

Supplemental Figure Legends

Figure S1. Construction of DT40-Rpb1 cells. (A) Diagrams depict four exons, 9~12, of chicken Rpb1, and the targeting vectors designed to delete the two *Rpb1* alleles. A region encompassing exon 10, intron 10, and exon 11 of Rpb1 was disrupted by targeting vectors carrying hygromycin-resistance gene (Hygro^r) or zeocin-resistance gene (Zeo^r). The vertical bars represent BamH1 digestion sites, and the thick line underneath exon 9, the probe used in Southern blotting. The two alleles of *Rpb1* were disrupted sequentially. Before the second allele was disrupted, a tet repressible HA-tagged human Rpb1 was introduced. (B) Genomic DNA samples were digested with BamH1, and analyzed by Southern blotting using exon 9 as a probe.

Figure S2. The CTD is highly conserved in vertebrates. Human and zebra fish CTDs both consist of 52 repeats and are 97% identical. Variant residues occur mostly at ser 7 positions. Likewise, the *X. laevis* and human CTDs are 97% identical, and the *A. carolinensis* (lizard) CTD is 100% identical to the human CTD. Although the sequence of the chicken genome was reported in 2004 (*8*), the *RPB1* sequence remains incomplete. The available partial chicken *RPB1* sequence, which lacks regions that encode 60 N terminal amino acids and 220 amino acids at the C terminus, is 97% identical to human, and the first 22 CTD repeats are 100% identical between chicken and human.

Figure S3. Expression of HA-tagged Rpb1. Western blot of lysates prepared from DT40, DT40 *rpb1(+/-)* and DT40-Rpb1 cells, treated with tet as indicated and for the times shown. Blots were probed with anti-HA and anti-actin antibodies.

Figure S4. Egr1 mRNA splicing and 3'-end processing. 26r and T4V cells were treated with tet for 40 hr, and egr1 expression was induced by ionomycin and PMA for 20 min. The unspliced, spliced, uncleaved, and total egr1 mRNA levels were determined by qPCR. The ratios of unspliced/spliced and uncleaved/total were calculated and are plotted relative to values obtained with 26r. N=3. Error bars indicate standard deviation.

Figure S5. 3' end cleavage efficiency of several polyadenylated mRNAs. RNA from 26r and T4V cells, treated with tet for 40 hr, was isolated and analyzed by RT-qPCR using primers to detect uncleaved and total mRNA. The ratios of uncleaved/total were calculated and are plotted relative to values obtained with 26r. Actb: beta-actin, Rps11: ribosomal protein S11, Rpl32: ribosomal protein L32. Rplp1: ribosomal protein, large, P1. N=3. Error bars indicate standard deviation.

Figure S6. Western blot analysis of select 3' processing factors. T4V and 26r cells were treated with tet for 30 hr. Cell lysates were prepared and analyzed by Western blotting using antibodies against the indicated proteins. Actin and GAPDH serve as loading controls.

Figure S7. Nuclear run-on analysis. (A) Cells were treated with tet for 24 hr and nuclei isolated. Following incubation with ^{32}P-UTP for 30 min, RNA was purified and analyzed by slot blot containing DNA oligos corresponding to the indicated genes (Top). NC1, NC2: negative controls (anti-sense U1 and U2). Signals from each gene were normalized to 18S, and plotted (Bottom). (B) Cells were treated with tet for 40 hours and analyzed as in (A). N=3. Error bars display standard deviation.

Figure S8. Dot blot analysis of P-thr4 antibody specificity. Indicated amounts of heptad peptides with or without phosphorylation on thr, and of two oligo-peptides, with phosphorylation on ser 5 (SYSPTS(p)PSYS), or on both ser 2 and ser 5 (SYS(p)PTS(p)PSYS), were spotted on membrane. Blot was probed with P-thr4 antibody.

Figure S9. DRB sensitivity of thr 4 phosphorylation. Wild-type DT40 cells were treated with tet for 24 hr and the indicated concentrations of DRB were added and cells grown for an additional 6 hr. Cell lysates were analyzed by Western blotting.

Figure S10. Flavopiridol inhibits expression of histone mRNAs. 26r cells were treated with tet for 24 hr and grown for an additional 16 hr with or without flavopiridol (300 nM). RNA was isolated and mRNA levels determined by RT-qPCR. Values relative to the untreated samples were plotted. N=3. Error bars indicate standard deviation.

References

1. J. Wang, Y. Takagaki, J. L. Manley, Targeted disruption of an essential vertebrate gene: ASF/SF2 is required for cell viability. *Genes & Development* **10**, 2588 (1996).

2. M. L. West, J. L. Croden, Construction and analysis of yeast RNA polymerase II CTD deletion and substitution mutations. *Genetics* **140**, 1223 (1995).

3. Y. Takagaki, J. L. Manley, Levels of polyadenylation factor CstF-64 control IgM heavy chain mRNA accumulation and other events associated with B cell differentiation. *Molecular Cell* **2**, 761 (1998).

4. R. Lemaire *et al.*, Stability of a PKCI-1-related mRNA is controlled by the splicing factor ASF/SF2: a novel function for SR proteins. *Genes & Development* **16**, 594 (Mar 1, 2002).

5. K. Livak, Analysis of relative gene expression data using real-time quantitative PCR and the 2−ΔΔCT method. *Methods* **25**, 402 (2001).

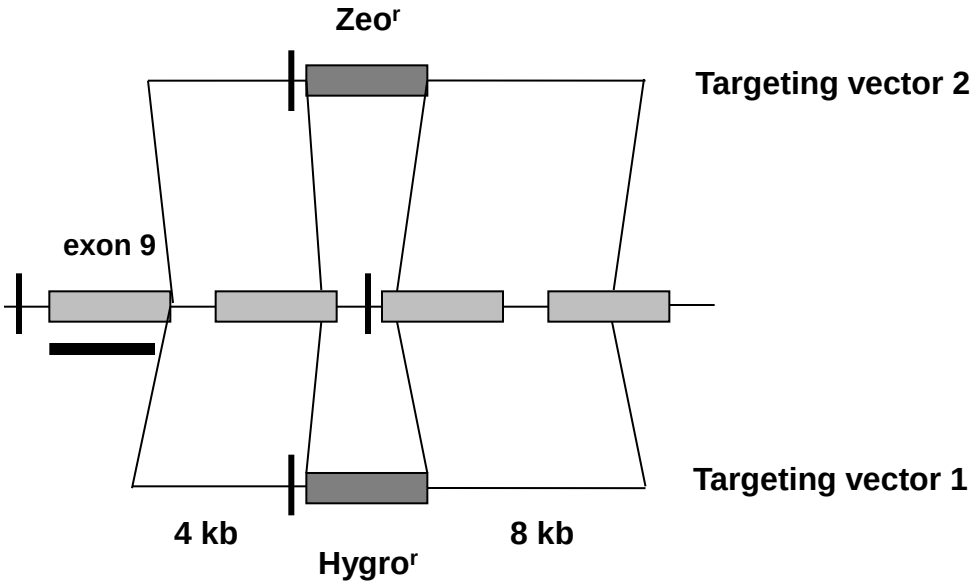
6. M. M. Chong *et al.*, Epigenetic propagation of CD4 expression is established by the Cd4 proximal enhancer in helper T cells. *Genes & Development* **24**, 659 (Apr 1, 2010).

7. E. Rosonina, S. M. Duncan, J. L. Manley, SUMO functions in constitutive transcription and during activation of inducible genes in yeast. *Genes & Development* **24**, 1242 (2010).

8. I. C. G. S. Consortium, Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* **432**, 695 (9 2004).

Figure S1.

A


Zeo^r
Targeting vector 2
exon 9
Targeting vector 1
4 kb
8 kb
Hygro^r

B

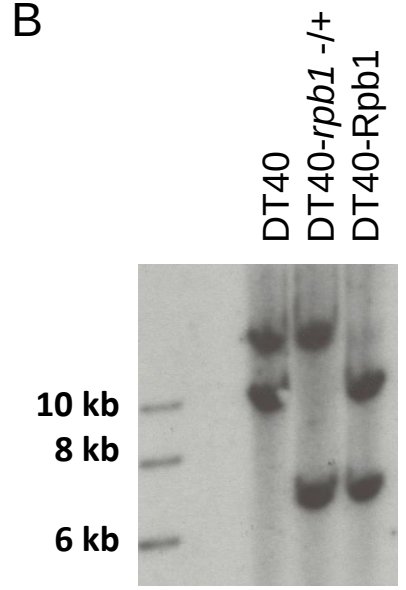

DT40
DT40-rpb1 -/+
DT40-Rpb1
10 kb
8 kb
6 kb

Figure S2.

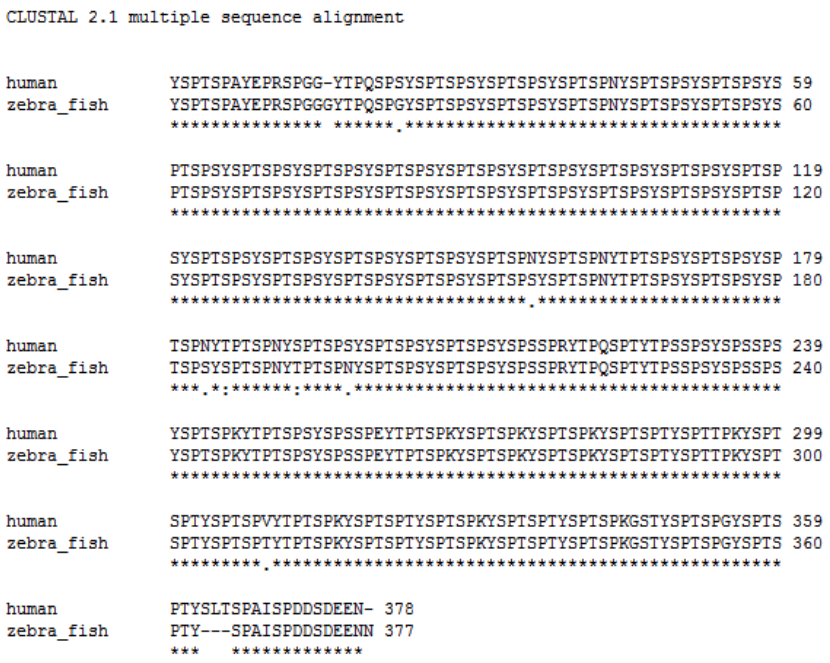

CLUSTAL 2.1 multiple sequence alignment
human YSPTSPAYEPRSPGG-YTPQSPSYSPTSPSYSPTSPSYSPTSPNYSPTSPSYSPTSPSYS 59
zebra_fish YSPTSPAYEPRSPGGGYTPQSPGYSPTSPSYSPTSPSYSPTSPNYSPTSPSYSPTSPSYS 60
human PTSPSYSPTSPSYSPTSPSYSPTSPSYSPTSPSYSPTSPSYSPTSPSYSPTSPSYSPTSP 119
zebra_fish PTSPSYSPTSPSYSPTSPSYSPTSPSYSPTSPSYSPTSPSYSPTSPSYSPTSPSYSPTSP 120
human SYSPTSPSYSPTSPSYSPTSPSYSPTSPSYSPTSPNYSPTSPNYTPTSPSYSPTSPSYSP 179
zebra_fish SYSPTSPSYSPTSPSYSPTSPSYSPTSPSYSPTSPSYSPTSPNYTPTSPSYSPTSPSYSP 180
human TSPNYTPTSPNYSPTSPSYSPTSPSYSPTSPSYSPSSPRYTPQSPTYTPSSPSYSPSSPS 239
zebra_fish TSPSYSPTSPNYTPTSPNYSPTSPSYSPTSPSYSPSSPRYTPQSPTYTPSSPSYSPSSPS 240
human YSPTSPKYTPTSPSYSPSSPEYTPTSPKYSPTSPKYSPTSPKYSPTSPTYSPTTPKYSPT 299
zebra_fish YSPTSPKYTPTSPSYSPSSPEYTPTSPKYSPTSPKYSPTSPKYSPTSPTYSPTTPKYSPT 300
human SPTYSPTSPVYTPTSPKYSPTSPTYSPTSPKYSPTSPTYSPTSPKGSTYSPTSPGYSPTS 359
zebra_fish SPTYSPTSPTYTPTSPKYSPTSPTYSPTSPKYSPTSPTYSPTSPKGSTYSPTSPGYSPTS 360
human PTYSLTSPAISPDDSDEEN- 378
zebra_fish PTY---SPAISPDDSDEENN 377

Figure S3.

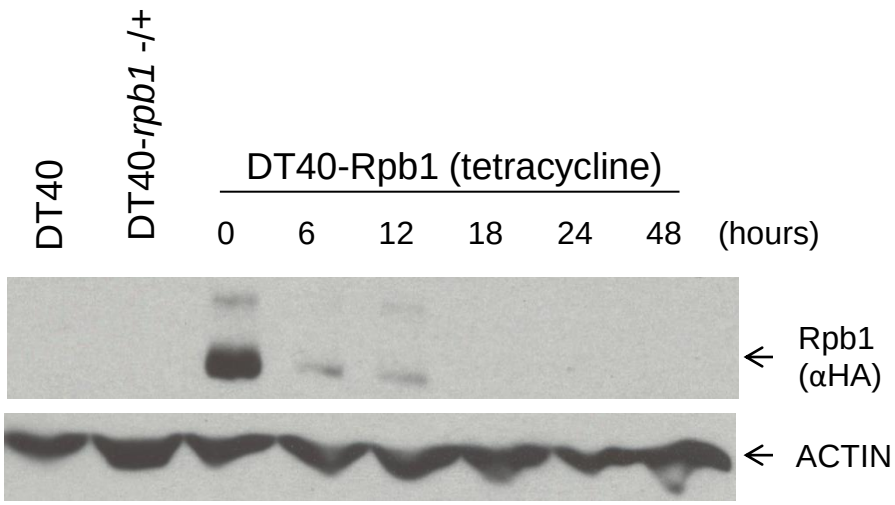

DT40
DT40-rpb1 -/+
DT40-Rpb1 (tetracycline)
0
6
12
18
24
48
(hours)
Rpb1
(αHA)
ACTIN

Figure S4.

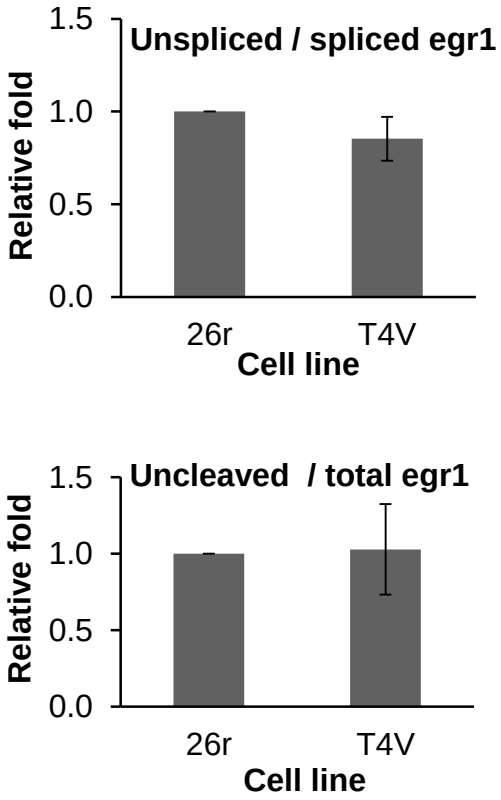

Unspliced / spliced egr1
Relative fold
1.5
1.0
0.5
0.0
26r
T4V
Cell line
Uncleaved / total egr1
Relative fold
1.5
1.0
0.5
0.0
26r
T4V
Cell line

Figure S5.

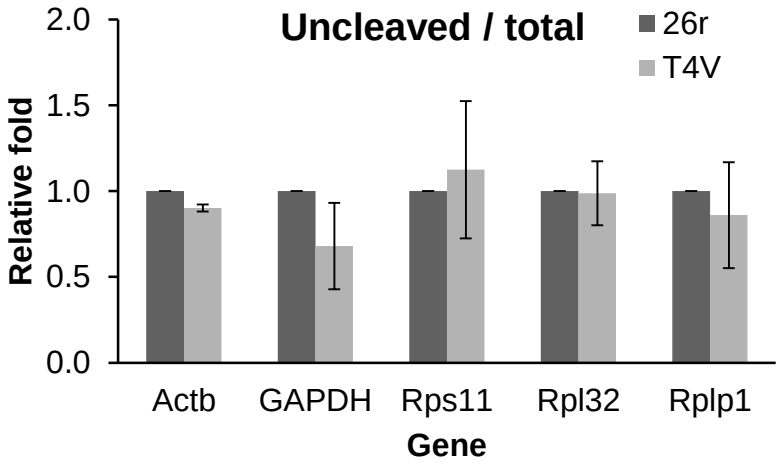

Uncleaved / total
26r
T4V
Relative fold
2.0
1.5
1.0
0.5
0.0
Actb
GAPDH
Rps11
Rpl32
Rplp1
Gene

Figure S6.

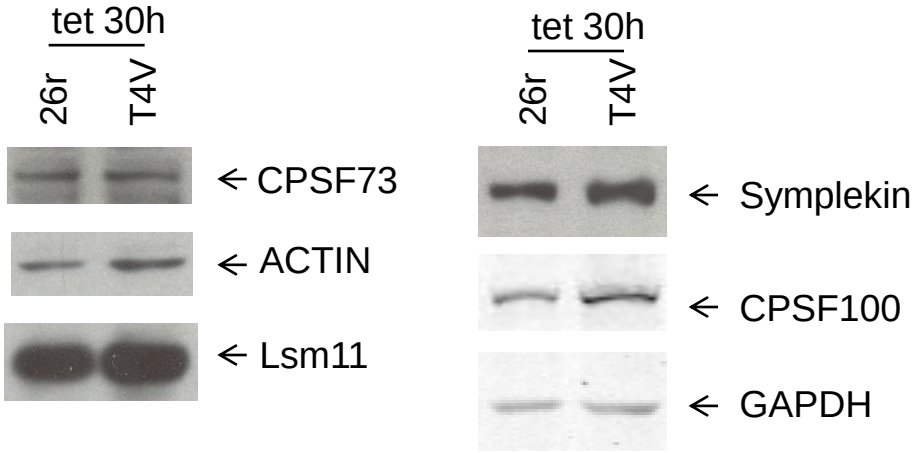

tet 30h
26r
T4V
CPSF73
ACTIN
Lsm11
tet 30h
26r
T4V
Symplekin
CPSF100
GAPDH

Figure S7.

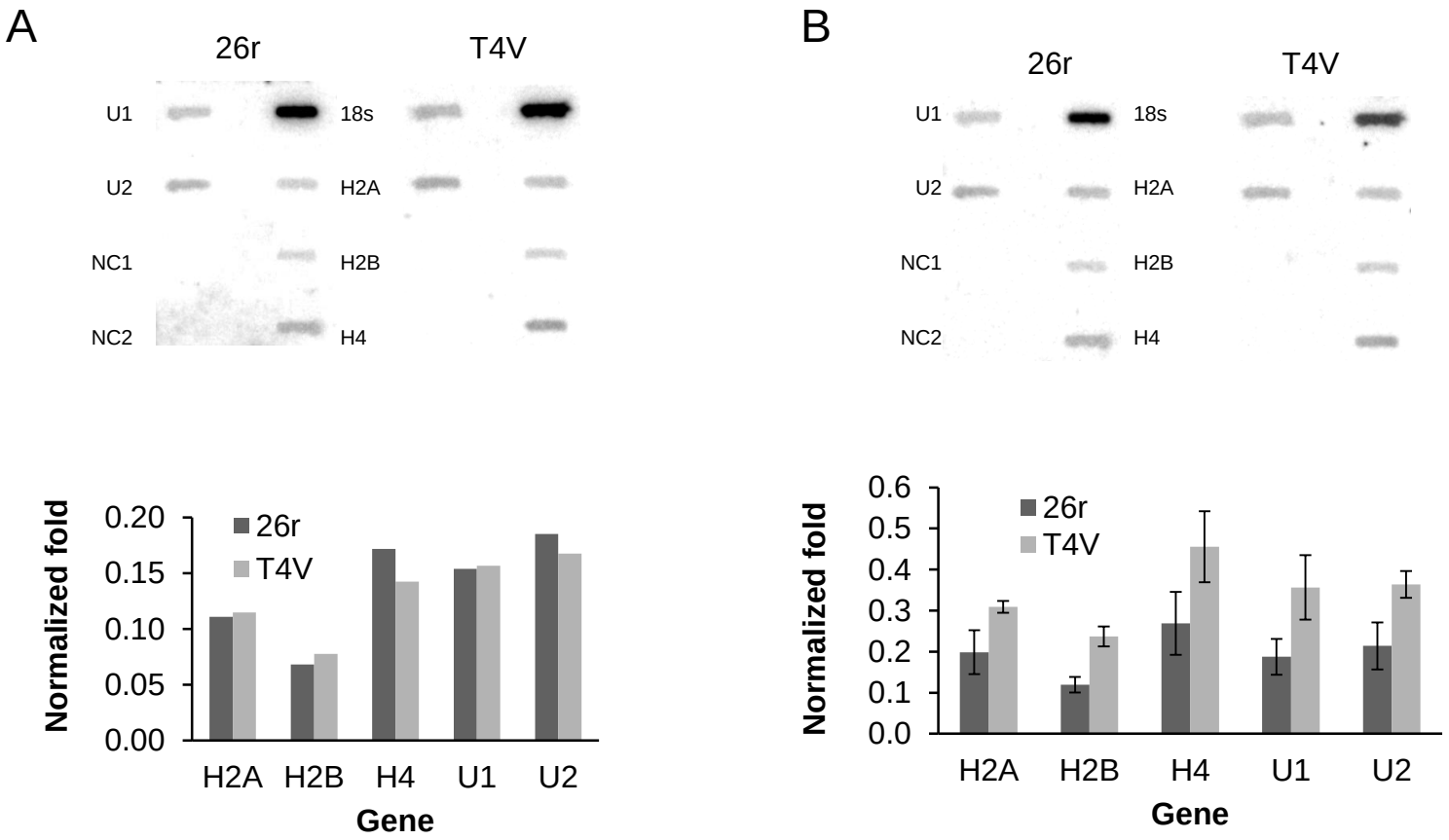
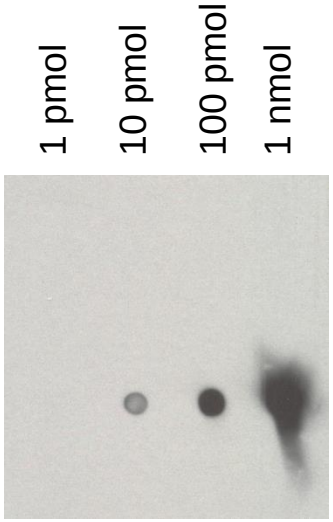

A
26r
T4V
U1
U2
NC1
NC2
18s
H2A
H2B
H4
Normalized fold
0.20
0.15
0.10
0.05
0.00
26r
T4V
H2A
H2B
H4
U1
U2
Gene
B
26r
T4V
U1
U2
NC1
NC2
18s
H2A
H2B
H4
Normalized fold
0.6
0.5
0.4
0.3
0.2
0.1
0.0
26r
T4V
H2A
H2B
H4
U1
U2
Gene

Figure S8.



SYS(p)PTS(p)PSYS

Heptad peptide

P-thr 4

SYSPTS(p)PSYS

Figure S9.

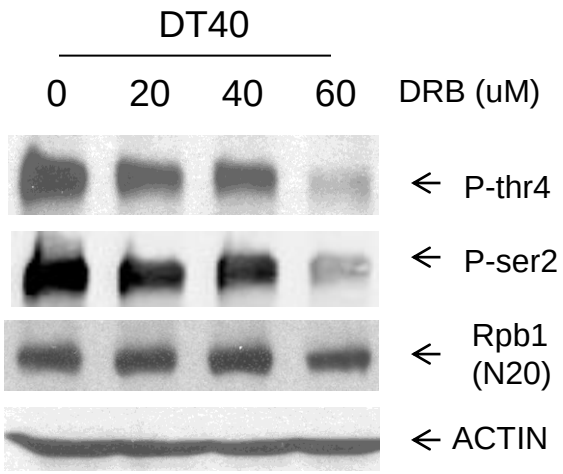
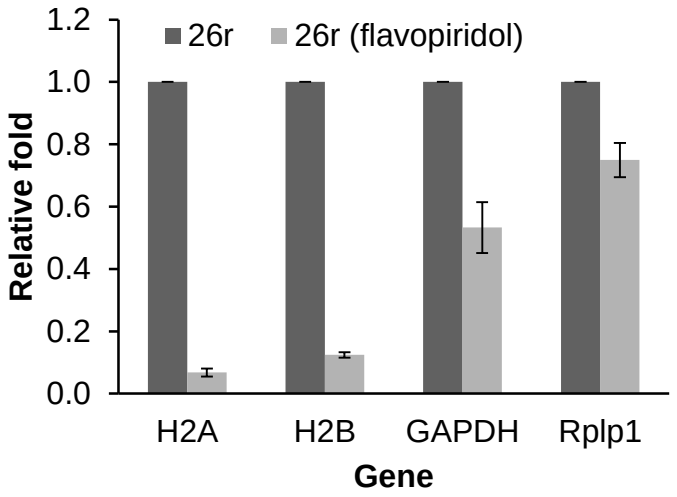


Figure S10.


1.2
1.0
0.8
0.6
0.4
0.2
0.0
Relative fold
26r
26r (flavopiridol)
H2A
H2B
GAPDH
Rplp1
Gene

Chapter 3

Function and control of RNA polymerase II CTD phosphorylation in vertebrate transcription and RNA processing

Jing-Ping Hsin, Amit Sheth, Kehui Xiang, James L. Manley

Department of Biological Sciences

Columbia University

New York, NY 10027, USA

Abstract

The C-terminal domain of the RNA polymerase II largest subunit (the Rpb1 CTD) is composed of tandem heptad repeats of the consensus $Y^1S^2P^3T^4S^5P^6S^7$, and is known to function, in a phosphorylation-dependent manner, in transcription and RNA processing. We reported previously that Thr 4 is phosphorylated throughout evolution, and provided evidence that this modification functions in histone mRNA 3' end formation in a chicken DT40 cell line. Here, we have extended our studies to other CTD mutations using these cells. We found that Rpb1 containing only the N-terminal half of the CTD, which consists of predominantly consensus repeats, conferred full viability, while the C-terminal half, with more divergent repeats, did not. Strikingly, this reflected a strong, specific defect in snRNA 3' end formation. Mutation of all Ser 2 (S2A) or Ser 5 (S5A) residues resulted in lethality, while Ser 7 (S7A) mutants displayed full viability. Further analysis of S2A and S5A cells revealed that they were defective in transcription and, to different extents, RNA processing. However, contrary to expectations from previous findings that Ser 7 plays a critical role in snRNA expression, no defects in this process were detected in S7A cells. Finally, we showed previously that phosphorylation of Thr 4 *in vivo* is inhibited by DRB and flavopiridol, specific inhibitors for the Ser 2 kinase CDK9. Here we show that Thr 4 is phosphorylated *in vitro* by purified CDK9. We also provide evidence that the Ser 2 phosphatase Fcp1 dephosphorylates Thr 4 *in vitro* and *in vivo*. Together, our data provide new insight into the function and conservation of CTD phosphorylation.

Introduction

RNA polymerase II (RNAP II) consists of 12 subunits and transcribes all mRNA and many non-coding RNA genes. Rpb1, the largest subunit, possesses a unique C-terminal domain (CTD) that consists of tandem heptad repeats, with a consensus sequence of tyr-ser-pro-thr-ser-pro-ser ($Y^1S^2P^3T^4S^5P^6S^7$). The number of repeats, generally reflecting the complexity of the organism, ranges from 26 in yeast to 52 in vertebrates. The CTD plays important roles in connecting transcription with all the steps of RNA production, including capping, splicing and 3' end processing, and these activities are modulated by posttranslational modification, mainly phosphorylation. Thus, the CTD can be envisioned functioning as a platform to recruit factors needed for proper RNA synthesis and maturation (for review, see refs 1-3).

The CTD is subject to extensive phosphorylation. All five hydroxylated amino acids are potential phosphorylation sites, and several CTD kinases and phosphatases have been discovered (for review, see refs 1, 4). Briefly, serine 5 (Ser 5) and serine 7 (Ser 7) are phosphorylated by the cyclin-dependent kinase CDK7, a component of TFIIH[5,6], whereas serine 2 (Ser 2) is phosphorylated by CDK9 (P-TEFb)[7,8]. Tyrosine 1 (Tyr1) is phosphorylated in mammals by the abl tyrosine kinase[9]. Threonine 4 (Thr 4) phosphorylation has been shown to be inhibited by specific CDK9 inhibitors[10,11], indicating the involvement of CDK9 in phosphorylating Thr 4. However, *in vitro* kinase assays showed that Thr 4 was phosphorylated by Polo-like kinase 3, questioning the role of CDK9 in Thr 4 phosphorylation[11]. The CTD is dephosphorylated by two major phosphatases, Fcp1 and Ssu72. Fcp1 has been reported capable of dephosphorylating Ser 2 and Ser 5, with a preference towards Ser 2[7,12], and Ssu72 can dephosphorylate both Ser 5 and Ser 7[13,14].

The CTD phosphorylation pattern corresponds in general to the position of RNAP II along a transcribed gene. Several genome-wide ChIP analyses of CTD phosphorylation provided evidence that early during transcription, the CTD is phosphorylated on Ser 5 and Ser 7, with Ser 5 gradually removed during elongation, while Ser 2 and Thr 4 phosphorylation increases as RNAP II progresses along the

gene[11,13,15-17]. Tyr 1 phosphorylation, analyzed in yeast, was also reported to increase as transcription progresses, but then to decrease as RNAP II approaches the transcript 3' end[18]. All phosphorylation is then "cleared" at or around transcription termination, which prepares RNAP II for reinitiation. While the majority of these studies have been performed in yeast, this general pattern of CTD phosphorylation is thought to be universal throughout eukaryotes (for reviews see refs 1, 4, 19).

This temporal pattern of CTD phosphorylation helps to link transcription with RNA processing events. For example, Ser 5 phosphorylation facilitates capping enzyme recruitment and indeed enhances the capping reaction, the first step of RNA processing[20-22]. Splicing factors, such as Prp40 and U2AF65, bind to the phosphorylated CTD, which facilitates recruitment and/or activation of the splicing machinery[23,24]. Recruitment of several mRNA 3' end processing factors to the vicinity of the nascent RNA[25] and 3' processing site[26] is enhanced by Ser 2 phosphorylation. Ser 7 phosphorylation has been implicated in recruitment of the Integrator complex, which functions in snRNA 3' end formation[27], while Thr 4 is important for efficient recruitment of 3' processing factors to histone genes[10].

The requirement of the phosphorylatable CTD residues for viability varies among species. For example, in *S. pombe*, only Ser 5 is essential[22], whereas in *S. cerevisiae*, Tyr 1 and Ser 2, as well as Ser 5, appear to be essential[28,29]. Although Thr 4 and Ser 7 residues are dispensable for yeast cell viability[29], this may not to be the case in higher eukaryotes. For example, mutation of Ser 7 residues to Ala (S7A) has been shown to compromise cell viability[30], while mutation of Thr 4 to Val (T4V) in chicken cells[10] or to Ala in human cells[11] was lethal.

Most functional studies of the CTD in mammalian cells have been performed using the Rpb1 inhibitor α-amanitin. In this assay, α-amanitin is added to the cell culture, and the effects of exogenously expressed α-amanitin-resistant Rpb1 derivatives carrying CTD mutations examined. Using this "pseudo-

genetic" system, considerable insight into the functions of the CTD has been obtained. For example, Bentley and colleagues first provided evidence that the CTD plays critical roles in mRNA processing[31,32]. This assay was also used to provide evidence that both the consensus-rich N-terminal half of the CTD and the more divergent C-terminal half were capable of effectively supporting mRNA processing[33,34], and that Ser 7 functions in snRNA gene expression[27]. More recently, methylation of a divergent Arg residue in the CTD C-terminal half was shown to function in repressing expression of certain snRNA and small nucleolar RNA genes[35], impaired splicing and 3' end cleavage were documented in cells expressing an α-amanitin-resistant Rpb1 with Ser 2 residues mutated to Ala[36], and Rpb1 with Thr 4 residues replaced with Ala displayed defects in transcription elongation[11]. Significantly, the T4V chicken cell line showed only a very modest effect on elongation[10]. Although there are several explanations for this discrepancy, it might result at least in part from the toxicity of α-amanitin (discussed in ref. 1). For example, cells expressing α-amanitin-resistant Rpb1 have been shown to grow slowly in the presence of α-amanitin[37,38], and accelerated degradation of several proteins, such as the transcription elongation factor DSIF, was reported[39]. These studies suggest that it may be valuable to extend genetic analyses of the vertebrate CTD with approaches that do not depend on α-amanitin.

In this report, we examined functions of the CTD in transcription and RNA processing using the genetically tractable DT40 cell system we described previously, and also extended our analysis of T4V modification. Surprisingly, while Rpb1 containing only the consensus-rich N-terminal half of the CTD conferred full viability, the more divergent C-terminal half did not, and we show that this reflects a strong and specific defect in snRNA 3' end formation. We found that Ser 2 and Ser 5 residues are essential for cell growth, as complete substitution of these residues with Ala impaired cell viability. Extending the importance of Ser 2 and Ser 5, we found that cells expressing Rpb1-S2A and Rpb1-S5A were defective in transcription, mRNA splicing and 3' end processing, albeit to different extents. In

contrast to previous findings indicating that Ser 7 is important for snRNA expression, Rpb1-S7A cells displayed no defect in snRNA expression and were fully viable. Finally, extending our analysis of Thr 4, and consistent with our previous results implicating CDK9 in Thr 4 phosphorylation *in vivo*[10], we found that purified CDK9 phosphorylates Thr 4 *in vitro*, and furthermore, that the Ser 2 phosphatase Fcp1 dephosphorylates Thr 4 *in vitro* and *in vivo*.

Results

Genetic complementation analysis reveals distinct requirements of CTD residues

To investigate genetically the functions of the CTD in vertebrate cells, we constructed a number of Flag-tagged Rpb1 derivatives with various CTD mutations (Table 1), and used an Rpb1 conditional knock-out cell line (tetracycline (tet) sensitive) described previously, DT40-Rpb1[10], to study the phenotypes of the Rpb1 mutants. The vertebrate CTD, which is almost invariant among species, contains 52 tandem heptads. The N-terminal heptads deviate little from the consensus, while the C-terminal repeats display more variation (Figure 1A). We showed previously that an Rpb1 derivative (26r) containing 26 all-consensus repeats plus the very C-terminal ten residues, which are important for Rpb1 stability[37,38], was able to completely rescue viability of DT40-Rpb1 cells in the presence of tet[10]. To determine whether the natural N-terminal or C-terminal 26 repeats are also sufficient for viability, two CTD derivatives, Rpb1 (1~26) and Rpb1 (27~52), were constructed (Figure 1A). Rpb1 (1~26) contains the first 26 heptads plus the C-terminal 10 residues, while Rpb1 (27~52) consists of the second half of the CTD, including the C-terminal 10 residues. Analogous to our studies with Thr 4[10], we also mutated all Ser residues to Ala, constructing Rpb1 derivatives with 26 repeats of YAPTSPS (S2A), 28 repeats of YSPTAPS (S5A), or 30 repeats of YSPTSPA (S7A). While as mentioned above 26 all-consensus repeats are sufficient to confer full viability, we also constructed an Rpb1 derivative with 19 YSPTSPS repeats (19r) to determine whether this lower number of repeats also confers viability.

We first asked if cells expressing these mutant Rpb1 derivatives are viable in the absence of wild-type Rpb1. Plasmids encoding the Rpb1 mutants were introduced into DT40-Rpb1 cells, and the transfected cells were grown in the presence of tet. The identity of surviving cell colonies was confirmed by western blotting with anti-FLAG antibodies (results not shown). As shown in Table 1, Rpb1 (1~26) allowed growth in tet, but Rpb1 (27~52) failed to give rise to tet-resistant colonies. As shown previously[10], 26r cells were fully viable, but the 19r cells were not, in line with previous observations in yeast[28], and mouse[40] cells. Significantly, cells expressing the Rpb1 derivatives S2A and S5A were inviable, as we observed previously with T4V cells[10]. However, somewhat unexpectedly, the S7A cells were fully viable. Together, these results revealed that the length and the conservation of the CTD repeats are crucial determinants for cell viability, and that Ser 2, 5 and Thr 4 are essential for viability, while Ser 7 is not.

To investigate further the phenotypes of the above mutant CTDs, we established stable DT40-Rpb1 cell lines expressing each of the Rpb1 derivatives described above, and analyzed how these mutants affect cell growth, transcription and RNA processing. In agreement with the results of the complementation analysis above, stable cell lines expressing wild-type Rpb1, Rpb1 (1~52), Rpb1 (1~26), 26r, and S7A were viable and displayed similar growth rates in medium containing tet (Figure 1B, left). Cells expressing Rpb1 (27~52) ceased growth ~48 hrs after tet addition, and S2A, T4V, and S5A cells all stopped growing after ~24-hrs tet treatment (Figure 1B). All Rpb1 derivatives were expressed at similar levels in these cell lines (Figure 1C).

Impact of CTD mutations on transcription and RNA processing

We next examined the effects of the CTD mutations on transcription, splicing and 3' processing, initially by using an inducible endogenous gene, *Egr1*. Cells were cultured in the presence of tet for 24

hrs, and *Egr1* was induced by addition of ionomycin and PMA. Following a 20-min induction, *Egr1* mRNA levels in 26r cells were strikingly increased, by about 300 fold as measured by RT-qPCR (Figure 2A; ref. 10). In contrast, *Egr1* induction levels were 4-5 fold lower in S2A and S5A cells. Small decreases in *Egr1* mRNA levels were observed in Rpb1 (1~26), Rpb1 (27~52), and S7A cells, but compared to Rpb1 (1~52) cells, the decreases were not significant (Figure 2B). We next used *Egr1* as a model to study how the Rpb1 CTD derivatives affect splicing and 3'-end processing. RT-qPCR assays designed to detect spliced and unspliced RNAs revealed that ~2 fold more unspliced *Egr1* mRNA was detected in S2A and S5A cells than in 26r cells (Figure 2C). Similarly, probes to distinguish total and 3' uncleaved RNAs detected 4-6 fold more uncleaved *Egr1* RNA in these cells (Figure 2D). More uncleaved *Egr1* RNA was detected in S2A cells than S5A cells, consistent with the known importance of Ser 2 phosphorylation in 3' cleavage (see above). Splicing and 3' processing were not affected detectably in any of the other mutants, although it is notable that 3'processing but not splicing appeared somewhat more efficient in Rpb1 (1~52) than in 26r or any of the other derivatives (Figure 2D).

Given that Rpb1 (27~52) cells are inviable, it was somewhat unexpected that we did not observe significant changes in *Egr1* transcription and/or RNA processing. To investigate if overall transcription and/or polyadenylation were affected in Rpb1 (27~52) cells, we performed *in vivo* labeling with ^{3}H uridine, and found that newly synthesized non-polyadenylated and polyadenylated RNA levels in Rpb1 (27~52) cells were similar to 26r and Rpb1 (1~52) cells (Figure 3A), indicating that these processes were not impaired by deletion of the N-terminal, consensus-rich half of the CTD. In agreement with the full viability of S7A cells (Table 1 and Figure 1B), overall transcription and polyadenylation in these cells was also comparable to Rpb1 (1~52) and 26r cells (Figure 3A).

The above data provided evidence that S2A and S5A cells are defective in *Egr1* transcription and RNA processing, which is largely consistent with previous studies. However, to investigate this further,

the effects of S2A and S5A mutations on transcription of several highly expressed genes were examined using Rpb1 ChIP assays. Consistent with the *Egr1* results, both S2A and S5A cells displayed impaired recruitment of RNAP II on several genes, a ribosomal protein gene, *Rplp1* (Figure 3B), *β-actin* (Figure 3C), a histone H2A gene (Figure 3D). Compared to 26r cells, decreased levels of Rpb1 were observed at the transcription start sites, coding regions, and 3' cleavage sites of all three genes in both S2A and S5A cells, more so in S2A cells than in S5A cells. Given the sequential phosphorylation of Ser 5 and Ser 2 along genes, it is possible that elongation was impaired in S5A cells. To address this, we reanalyzed Rpb1-S5A ChIP data by normalizing signals to the value at the TSS, and found that Rpb1-S5A levels along the *Rplp1* and *β-actin* genes were unaffected (Figure S1), providing evidence that Ser 5 is not necessary for efficient elongation. Consistent with this, Ser 2 phosphorylation was not strongly reduced by the S5A mutation (Figure S2A; the small decrease observed may reflect reduced recognition of S5A by the antibody).

We also examined RNAP II levels on U1 and U2 snRNA genes. Strikingly, Rpb1 levels on these genes were strongly reduced in S2A cells, but not significantly in S5A cells (Figure 3D), suggesting that Ser 5 is not necessary for U1 and U2 transcription. However, given that Ser 5 functions in facilitating capping enzyme[20] and Nrd1 termination factor[41] recruitment, and that transcription termination of snRNA genes is via a Nrd1-dependent pathway (for review, see refs 1, 42), it was possible that Ser 5 is required for other aspects of U snRNA gene expression, and indeed that appears to be the case (see below). Ser 7 phosphorylation was unaffected in S2A cells, indicating that the strong decrease in snRNA transcription was not due to a defect in this modification (Figure S2B).

Together, our results support the well-established involvement of Ser 2/Ser 5 phosphorylation in transcription and RNA processing, although there are some differences with previous studies. These are discussed below.

Distinct requirements for snRNA expression

We next examined snRNA expression in the mutant cells in more detail. Ser 7 phosphorylation has been shown to play an important role in snRNA gene expression and processing[27,43], and to be present on transcribing RNAP II on many genes[30]. It was unexpected therefore that S7A cells were fully viable and displayed no defects in overall transcription (see above). To investigate the role of Ser 7 in snRNA expression in DT40-Rpb1 cells, we first performed nuclear run-on (NRO) assays to determine the effect of the S7A mutation on snRNA transcription. Cells were treated with tet for 40 hrs, and nuclei were isolated and processed for NRO. Essentially identical levels of U1 and U2 transcription were detected in Rpb1 (1~52), 26r, and S7A cells (Figure 4A). We also measured steady-state levels of U1 and U2 snRNA in cells treated with tet for 6 days. RT-qPCR with total cellular RNA demonstrated that the levels of U1 and U2 snRNA, in two independent S7A cell lines, were equivalent to those in Rpb1 (1~52) and 26r cells (Figure 4B). We next examined U2 snRNA levels by RT-qPCR in several of the mutant cell lines after 40 hrs tet treatment. Total U2 snRNA levels were sharply reduced in both S2A and, importantly, S5A cells, and to a lesser degree in Rpb1 (27~52 cells). U2 snRNA levels were not reduced, and indeed very slightly increased, in S7A cells (Figure 4C). We also measured U2 snRNA 3' end formation, again by RT-qPCR, as depicted in Figure 4D. S7A cells were essentially identical to 26r cells, and a modest increase in the ratio of uncleaved to total U2 snRNA (~2 fold) was observed in S5A cells. However, a greater increase (~6 fold) was detected in S2A cells, and a striking ~14 fold increase was observed in Rpb1 (27~52) cells (Figure 4D). It is possible that this defect explains, at least in part, the inviability of Rpb1 (27~52) cells. Very much the same conclusions were drawn from analysis of U1 snRNA expression (Figure S3). It is intriguing that Rpb1 (27~52) cells, but not S7A cells, were defective in snRNA 3'end formation. We discuss these results below, and how they might be reconciled with previous findings.

Thr4 is phosphorylated by CDK9 in vitro, and dephosphorylated in vitro and in vivo by Fcp1

The kinases and phosphatases responsible for the reversible phosphorylation of Ser 2, 5 and 7 have been well studied, but less is known about the enzymes responsible for Thr 4 modification. We showed previously that inhibition of CDK9 prevented Thr 4 phosphorylation *in vivo*[10], while a Polo-like kinase (Plk3) has also been implicated in this modification[11]. Nothing is known about what enzyme dephosphorylates Thr 4. To investigate further the role of CDK9 in Thr 4 phosphorylation, we performed an *in vitro* kinase assay, using bacterially expressed and purified GST-CTD as a substrate and recombinant CDK9 isolated from baculovirus-infected insect cells (see Figure S4 for SDS gels of purified proteins). As shown in Figure 5A, CDK9 was capable of phosphorylating Thr 4 in a concentration-dependent manner, as judged by western blot with a previously described Thr 4-specific antibody[10]. Phosphorylation was inhibited by the CDK9 inhibitors, DRB and Flavopiridol (Figure 5A), ruling out the possibility that phosphorylation was mediated by a contaminating activity in the CK9 preparation.

We next wished to identify the phosphatase responsible for Thr 4 dephosphorylation. As mentioned in the Introduction, the two major CTD kinases are Fcp1 and Ssu72. Given that Fcp1 is known to dephosphorylate Ser 2 (e.g., refs 7, 12), that genome-wide ChIP studies indicated that the Thr 4 phosphorylation pattern is similar to Ser2, peaking toward the 3' ends of genes[11], and that preliminary experiments provided no evidence that Ssu72 could not dephosphorylate Thr 4 (results not shown), we reasoned that Fcp1 could be the Thr 4 phosphatase. To investigate this possibility, we first performed *in vitro* phosphatase assays, using recombinant Fcp1 (residues 168-606; ref. 44) and a catalytically inactive derivative (M271E; ref. 45) purified from bacteria cells and GST-CTD phosphorylated in HeLa nuclear extract and repurified (SDS gels shown in Figure S4). The results, again using western blot with the Thr 4-specific antibody, demonstrate that Thr4 was indeed dephosphorylated by wild-type Fcp1, but not by the M271E derivative (Figure 5B). We next asked whether Fcp1 can dephosphorylate Thr 4 *in vivo.* To this end, we designed and utilized three distinct shRNAs to target Fcp1 mRNA in HEK293 cells, and then

measured their effect, relative to a control shRNA, on Rpb1 Thr 4 phosphorylation by western blot (Figures 5C and D). Although with all three shRNAs, Fcp1 mRNA (see Figure S5) and protein levels were only modestly reduced (protein levels were decreased ~3 fold in all cases), Thr 4 phosphorylation levels were increased, ~2 fold, by all three shRNAs. Our data thus support the idea that Fcp1 dephosphorylates Thr 4, extending its previously described role in Ser 2, and possibly Ser 5, dephosphorylation.

Discussion

In this paper, we described experiments analyzing properties of the RNAP II CTD, primarily utilizing genetically tractable chicken DT40 cells. While some of our results serve to confirm findings obtained in other contexts, for example in yeast or using the α-amanitin system in mammalian cells, we have also provided a number of new and in some cases unexpected insights into properties of different CTD residues. For example, we found the first half of the CTD, consisting of mostly all-consensus repeats, as well as Ser 2, Thr 4 (ref. 10), and Ser 5 residues were essential for cell survival, while Ser 7 was not. Further analysis of these cells showed that, consistent with previous studies, Ser 2 and Ser 5 were required for optimal transcription and RNA splicing/polyadenylation. Although RNAP II levels on several genes in cells expressing Rpb1-S5A were reduced, transcription elongation was unaffected by the S5A mutation, consistent with the fact that Ser 2 phosphorylation was unaffected. In contrast to expectations from previous studies, we did not observe any defects in snRNA gene expression in Rpb1-S7A cells. Instead, significant defects in snRNA expression, and specifically in 3' processing, were detected in cells expressing Rpb1 (27~52), which consists predominantly of non-consensus repeats. Finally, we reported previously that CDK9 was required for Thr 4 phosphorylation *in vivo*[10], and we showed here that recombinant CDK9 phosphorylates Thr 4 *in vitro*, strengthening the case that Thr 4 is indeed a direct CDK 9 target. We also provided evidence that Fcp1 dephosphorylates Thr 4 *in vitro* and *in*

vivo. Below we discuss the significance of these findings, especially how they compare and contrast with previous related studies.

Our experiments have shown striking differences in the behavior of the N- and C-terminal halves of the CTD. Despite the fact the CTD is highly conserved from yeast to humans, and nearly invariant in vertebrates, it is well known from studies in yeast[28], mouse[40] and chicken[10] cells that only ~50% of the CTD is sufficient to confer cell viability. Furthermore, in human cells using the α-amanitin system, both the N-terminal consensus-rich and the more divergent C-terminal half tend to behave similarly in functional assays (e.g., ref. 34), although in some case displaying modest differences[33,46]. In agreement with this, we showed here that the N-terminal half of the CTD behaved essentially identical to the full-length CTD with respect to cell viability and in all functional assays tested, with the exception that mRNA 3' processing efficiency with transcripts from the inducible *Egr1* gene was slightly reduced. Cells expressing Rpb1 (27~52), containing only the C-terminal half of the CTD, likewise behaved similarly to wild-type cells in assays measuring transcription, splicing and polyadenylation. But surprisingly, these cells were completely inviable, revealing that these heptads cannot perform an essential function of the CTD. As discussed below, this is likely to be the observed defect in snRNA 3'processing.

Our studies have provided direct evidence that Ser 2 and Ser 5 are important for efficient mRNA splicing and 3' processing. Previously, 3' end processing defects were observed in human cells upon CDK7 inhibition[5,47], and in cells expressing an α-amanitin-resistant Rpb1-S5A[27]. 3' end cleavage defects were documented upon CDK9 deletion in yeast[48] or inhibition in Drosophila[26], and both splicing and 3' processing defects were observed in Xenpus oocytes treated with CDK9 inhibitors[49], and in human cells expressing an α-amanitin-resistant Rpb1-S2A[36]. Extending these results, our analysis in S2A and S5A cells provided evidence that Ser 2 phosphorylation is important for optimal splicing and 3' end formation of transcripts produced from the inducible *Egr1* gene. Significantly, we also found that Ser 5

phosphorylation, in addition to its involvement in 3'-end processing, is also required for proper splicing. This was not due to a defect in Ser 2 phosphorylation, as this was unaffected in S5A cells. Although it is perhaps not surprising that Ser 5 phosphorylation functions to enhance splicing, to our knowledge ours is the first demonstration that this is the case

Our data have also provided new insights into the role of Ser 5 phosphorylation in transcription elongation. Transcription in cells expressing the Rpb1-S5A derivative, as measured by Pol II ChIP, was not as efficient as in the corresponding Rpb1-26r cells. This is consistent with previous reports that Cdk7 inhibition results in a decrease of Rpb1 levels on examined genes at the TSS, coding region and 3' end[5,47], and our Rpb1-S5A ChIP assays extend these results by demonstrating similar reductions on several genes. Recent studies in yeast suggested that Ser 5 phosphorylation by the CDK7 homolog Kin28 primes the CTD, and enhances recruitment of the CDK9 homolog, Bur1/Bur2[50,51]. Furthermore, phosphorylation on Ser 7, not Ser 5, has been suggested to prime the CTD for CDK9 recruitment[52]. We further investigated the physiological significance of these findings in our DT40 cells. For this, we normalized our Rpb1-S5A ChIP data to signals at the promoter region, and found that Rpb1-S5A levels downstream to the promoter were comparable to Rpb1-26r levels, indicating that Ser 5 phosphorylation was not required for elongation. Furthermore, Ser 2 phosphorylation was not significantly affected in the S5A cells. Likewise, S7A cells were not defective in transcription elongation, or indeed any aspect of transcription. These lines of evidence suggest that CDK9 recruitment is not impaired by S5A or S7A mutations, and thus that Ser 2 phosphorylation is independent of Ser 5/7 phosphorylation. These findings are consistent with and extend previous reports that recruitment of CDK9 was not reduced by CDK7 inhibition in *Drosophila*[53] and human[47] cells. Importantly, our data provide evidence that Ser 5/7 phosphorylation is not involved in CDK9 recruitment, or in transcriptional elongation, in vertebrate cells.

Our results also extend studies examining the role of Ser 2 phosphorylation on transcription. Our ChIP analyses showed that Rpb1-S2A levels were decreased, more severely than Rpb1-S5A, all along the length of the genes examined. This contrasts with previous observations that inhibition of CDK9 resulted in only slight changes of RNAP II density on examined genes[26], and an accumulation of RNAP II around the TSS[54]. This apparent discrepancy could be explained by the existence of additional Ser 2 kinases, e.g., CDK12 and 13[55,56], which may not be inhibited by the CDK9 inhibitors[57], and thus could perhaps partially compensate for CDK9. In any event, our results provide strong confirmatory evidence for the importance of Ser 2 phosphorylation in transcription, fully consistent with the observed inviability of S2A cells. It is notable that an S2A derivative similar to the one we analyzed here was found to confer full viability, under normal growth conditions, in *S. pombe*[22]. This finding was somewhat surprising, given the important functions attributed to Ser 2 phosphorylation in our studies and others, and additional work will be required to understand the basis for this.

Perhaps the most unexpected of our findings is the complete dispensability of Ser 7. Previous studies provided evidence that Ser7 phosphorylation plays an important role in snRNA expression and 3' processing[27,43]. However, DT40 cells expressing Rpb1-S7A were fully viable, and displayed no defects in any aspect of gene expression, including expression of U1 and U2 snRNA genes. What might be the basis for this discrepancy? One possibility is a difference between chicken and human cells in the mechanism of U snRNA gene expression. However, we consider this unlikely, as the snRNA genes and factors involved in their expression are all highly conserved, and we are unaware of any evidence suggesting evolutionarily-based differences in the basic steps of gene expression among vertebrate organisms. A second possibility stems from the usage of α-amanitin in the previous study. As mentioned above, a recent study[39] showed that α-amanitin treatment accelerates the degradation of several proteins, including the transcription elongation factor DSIF, known to be important for U gene expression (e.g.,

ref. 46). Thus we suggest that the S7A mutation coupled with reduced accumulation of a required factor such as DSIF in the presence of α-amanitin results in a "synthetic" phenotype reflected in defective snRNA expression. This model nonetheless envisions a role for Ser 7 phosphorylation, and indeed our results provide considerable albeit indirect support for this. Specifically, we suggest that the strong and specific defects in snRNA 3' end processing we observed in Rpb1 (27~52) cells reflects changes in Ser 7. As mentioned above, the heptads in the C-terminal half of the CTD display considerable divergence from the consensus, and this divergence is by far the greatest at Ser 7. Only 6 of the 26 C-terminal heptads contain Ser at this position, and 9 contain a basic residue (8 Lys residues and one Arg; indeed, the arg residue, via methylation, exerts a negative effect on snRNA gene expression[35]). Other deviations from the consensus are mostly conservative. Thus, we suggest the defective snRNA 3' end formation detected in Rpb1 (27~52) cells resulted as a consequence of these non-conservative changes in Ser 7 residues. As we detected no other defects in the Rpb1 (27~52) cells, it may be that defective snRNA gene expression underlies the inviability of these cells.

We have also provided new insight into how phosphorylation of another CTD residue, Thr 4, is controlled. It was previously shown that Thr 4 phosphorylation was inhibited *in vivo* by the specific CDK9 inhibitors DRB and flavopiridol[10,11]. However, Hintermair et al. (2012) found that CDK9 was unable to phosphorylate Thr 4 *in vitro,* and instead provided compelling evidence that Polo-like kinase 3 (Plk3) could do so. In addition, these authors were unable to detect Thr 4 phosphorylation *in vivo* in the context of an S2A Rpb1 derivative, perhaps suggesting that the apparent CDK 9 requirement may instead indicate that Ser 2 phosphorylation is a prerequisite for Thr 4 phosphorylation[58]. However, our own experiments also examined Thr 4 phosphorylation on an S2A derivative (the one analyzed here), and found that Thr 4 phosphorylation levels were in fact unaffected[10]. One explanation for this discrepancy may be the antibodies employed. A negative result analyzing mutant CTD derivatives may

be difficult to interpret depending on the sensitivity of the antibody to alterations in its epitope. In any event, the *in vitro* kinase assays presented here, showing that Thr 4 in GST-CTD can be phosphorylated by purified recombinant CDK9/cyclin T, strengthen the case that CDK9 is indeed the main Thr 4 kinase. In keeping with this, it would be somewhat surprising if Plk3 was the principal Thr 4 kinase, as it has well-established roles in other cellular processes, e.g., the cell cycle and stress response, and has been reported to be localized primarily in the nucleolus[59]. In addition, a systematic proteomic study designed to identify RNAP II-associated proteins detected CDK7, CDK9, and CDK12, but not Plk3[60]. None of this however rules out the possibility that Plk3 may indeed phosphorylate Thr 4 under some conditions. This may be analogous to the case with Ser 5, which we showed previously can be modified during M phase by the cell-cycle kinase Cdc2/cyclinB[61]. Finally, our data providing evidence that Fcp1 dephosphorylates Thr4 *in vitro* and *in vivo* both explains how Thr 4 is dephosphorylated and also extends the roles of this well-studied CTD phosphatase. While Fcp1 can dephosphorylate both Ser 2 and Ser 5, its primary function is thought to dephosphorylate Ser 2 at the 3' ends of genes (reviewed in ref. 1). Thus, an attractive model is that CDK9 phosphorylates both Ser 2 and Thr 4 during transcription elongation, and that the same two residues are dephosphorylated by Fcp1 at the ends of genes.

Acknowledgements

We like to thank lab members, especially Emanuel Rosonina, for discussion and critical reading of this manuscript. We thank Dr. Stephane Larochelle and Dr. Robert Fisher for CDK9 baculoviruses, and Dr. Patrick Cramer for the Fcp1 expression vector. This work was supported by grants from the NIH.

Materials and Methods

Cell culture and cloning

DT40 cells and HEK293 cells were cultured at 37^{o}C with 5% CO_2 in RPMI1640 medium containing 10% FBS and 1% chicken serum, and in DMEM medium containing 10% FBS, respectively.

Rpb1 CTD derivatives were cloned as previously described[10]. Briefly, a fragment of beta-actin promoter and FLAG tag was inserted into pBlueScript containing Neomycin resistance gene. The human *Rpb1* body without the CTD was inserted behind the FLAG tag, and various CTD fragments were inserted directly 3' to the *Rpb1* body.

Fcp1 knockdown shRNA constructs were generated with DNA oligonucleotides targeting the following sequences, AAGAGGAAGCTGAATGAAGAGGA, AAGTATGACCGCTACCTCAACAA and AATCATTCTCGAGGCACTGAGGT in Fcp1. Synthesized oligonucleotides were cloned into HuSH pRS vector (Origene) and verified by sequencing.

Complementation and construction of stable cell lines

Procedures for complementation assays and for constructing stable cell clines were as previously described[10]. Briefly, 10^7 cells were transfected with ~15 ug of linearized DNA, and selected in the presence of appropriate antibiotics. Surviving cell clones were isolated, and further analyzed using western blotting.

Western blotting

Cells lysates were resolved in SDS-PAGE with indicated percentage of acrylamide. Western blotting was performed using standard protocols. Briefly, protein samples were transferred to nitrocellulose membrane, blocked in 5% milk, and incubated with primary antibodies diluted in PBST (0.1% Tween20). Membranes were then washed in PBST, and incubated with appropriate secondary antibodies. For quantification, western blots were analyzed by ImageJ. Antibodies used in this paper as follows: Flag tag

(M2; Sigma), actin (Sigma), phospho-ser7 CTD heptad (4E12; Millipore), phospho-thr4 CTD heptad (Novus), phospho-ser5 (H14, Covance), phospho-ser2 (H5, Covance), phospho-ser2 (3E10, Millipore), GST tag (Invitrogen), Rpb1 (N20; Santa Cruz), Fcp1 (Bethyl Laboratories).

RT-qPCR

RNA was extracted using Trizol (Invitrogen), and further treated with DNase I. Reverse transcription and qPCR analysis were performed as previously described[10]. Primer sequences will be provided upon request.

In vivo labeling of nascent RNA , nuclear run-on and slot blotting

In vivo ^{3}H uridine labeling and NRO assays was performed as previously described[10].

Chromatin immunoprecipitation (ChIP)

Cells were grown to 70% confluence (~$2x10^{6}$/ml), cross-linked with 1% formaldehyde for 10 minutes, and processed for ChIP as previously described[10]. ChIP was performed using antibody against Flag tag (M2; Sigma). Primers sequences will be available upon request.

***In vitro* kinase and phosphatase assays**

CDK9 protein complexes were expressed in insect cells as described[62], and purified using Ni-NTA agarose (QIAGEN). GST-CTD was expressed in *E. coli* as previously described[63], and purified using glutathione Sepharose 4B (GE Healthcare). 0.75 ug of GST-CTD was phosphorylated by CDK9 complexes at 37^{o}C for 1 hour in a kinase buffer (25 mM Hepes PH 7.5, 10 mM $MgCl_2$, 150 mM NaCl, 1 mM ATP). *In vitro* phosphorylation of GST-CTD by Hela nuclear extract was performed as previously described[63]. Baculoviruses expressing CDK9/cyclin T were gifts from Dr. Robert Fisher (Mt. Sinai).

For phosphatase assays, *S. cerevisiae* Fcp1 and a catalytic mutant Fcp1 (M271E) were expressed in *E. coli* and purified as described[44]. 200 ng of GST-CTD, phosphorylated by Hela nuclear extract, was incubated with Fcp1 proteins in a buffer containing 50 mM Tris-Acetate pH 5.5, 10 mM $MgCl_2$ at 30^{o}C for 90 minutes. A vector expressing functional Fcp1 (residues 168-606) was a gift from Dr. Patrick Cramer (Munich). Based on previous mutational studies[45], a point mutation was introduced into the Fcp1 (168-606) expression vector to create the catalytic mutant Fcp1 (M271E).

References

1 Hsin, J. P. & Manley, J. L. The RNA polymerase II CTD coordinates transcription and RNA processing. *Genes Dev* **26**, 2119-2137 (2012).

2 Buratowski, S. Progression through the RNA polymerase II CTD cycle. *Mol Cell* **36**, 541-546 (2009).

3 Moore, M. J. & Proudfoot, N. J. Pre-mRNA processing reaches back to transcription and ahead to translation. *Cell* **136**, 688-700 (2009).

4 Egloff, S., Dienstbier, M. & Murphy, S. Updating the RNA polymerase CTD code: adding gene-specific layers. *Trends Genet* **28**, 333-341 (2012).

5 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**, 5455-5464 (2009).

6 Akhtar, M. S. *et al.* TFIIH kinase places bivalent marks on the carboxy-terminal domain of RNA polymerase II. *Mol Cell* **34**, 387-393 (2009).

7 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**, 3319-3329 (2001).

8 Marshall, N. F., Peng, J., Xie, Z. & Price, D. H. Control of RNA polymerase II elongation potential by a novel carboxyl-terminal domain kinase. *J Biol Chem* **271**, 27176-27183 (1996).

9 Baskaran, R., Dahmus, M. E. & Wang, J. Y. J. Tyrosine phosphorylation of mammalian RNA polymerase II carboxyl-terminal domain. *Proc Natl Acad Sci U S A* **90**, 11167-11171 (1993).

10 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**, 683-686 (2011).

11 Hintermair, C. *et al.* Threonine-4 of mammalian RNA polymerase II CTD is targeted by Polo-like kinase 3 and required for transcriptional elongation. *The Embo Journal* **31,** 2784-2797 (2012).

12 Hausmann, S. & Shuman, S. Characterization of the CTD phosphatase Fcp1 from fission yeast. Preferential dephosphorylation of serine 2 versus serine 5. *J Biol Chem* **277**, 21213-21220 (2002).

13 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**, 158-170 (2012).

14 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**, 8541-8551 (2012).

15 Tietjen, J. R. *et al.* Chemical-genomic dissection of the CTD code. *Nat Struct Mol Biol* **17**, 1154-1161 (2010).

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

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

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

19 Heidemann, M., Hintermair, C., Voss, K. & Eick, D. Dynamic phosphorylation patterns of RNA polymerase II CTD during transcription. *Biochim Biophys Acta* **1829**, 55-62 (2013).

20 Ho, C. K. & Shuman, S. Distinct roles for CTD Ser-2 and Ser-5 phosphorylation in the recruitment and allosteric activation of mammalian mRNA capping enzyme. *Mol Cell* **3**, 405-411 (1999).

21 Fabrega, C., Shen, V., Shuman, S. & Lima, C. D. Structure of an mRNA Capping Enzyme Bound to the Phosphorylated Carboxy-Terminal Domain of RNA Polymerase II. *Mol Cell* **11**, 1549-1561 (2003).

22 Schwer, B. & Shuman, S. Deciphering the RNA polymerase II CTD code in fission yeast. *Mol Cell* **43**, 311-318 (2011).

23 Morris, D. P. & Greenleaf, A. L. The splicing factor, Prp40, binds the phosphorylated carboxyl-terminal domain of RNA polymerase II. *J Biol Chem* **275**, 39935-39943 (2000).

24 David, C. J., Boyne, A. R., Millhouse, S. R. & Manley, J. L. The RNA polymerase II C-terminal domain promotes splicing activation through recruitment of a U2AF65-Prp19 complex. *Genes Dev* **25**, 972-983 (2011).

25 Fong, N. & Bentley, D. L. Capping, splicing, and 3' processing are independently stimulated by RNA polymerase II: different functions for different segments of the CTD. *Genes Dev* **15**, 1783-1795 (2001).

26 Ni, Z., Schwartz, B. E., Werner, J., Suarez, J. R. & Lis, J. T. Coordination of transcription, RNA processing, and surveillance by P-TEFb kinase on heat shock genes. *Mol Cell* **13**, 55-65 (2004).

27 Egloff, S. *et al.* Serine-7 of the RNA polymerase II CTD is specifically required for snRNA gene expression. *Science* **318**, 1777-1779 (2007).

28 West, M. L. & Gorden, J. L. Construction and analysis of yeast RNA polymerase II CTD deletion and substitution mutations. *Genetics* **140**, 1223-1233 (1995).

29 Stiller, J. W., McConaughy, B. L. & Hall, B. D. Evolutionary complementation for polymerase II CTD function. *Yeast* **16**, 57-64 (2000).

30 Chapman, R. D. *et al.* Transcribing RNA polymerase II is phosphorylated at CTD residue serine-7. *Science* **318**, 1780-1782 (2007).

31 McCracken, S. *et al.* The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. *Nature* **385**, 357-361 (1997).

32 McCracken, S. *et al.* 5'-Capping enzymes are targeted to pre-mRNA by binding to the phosphorylated carboxy-terminal domain of RNA polymerase II. *Genes Dev* **11**, 3306-3318 (1997).

33 Fong, N., Bird, G., Vigneron, M. & Bentley, D. L. A 10 residue motif at the C-terminus of the RNA pol II CTD is required for transcription, splicing and 3' end processing. *Embo J* **22**, 4274-4282 (2003).

34 Rosonina, E. & Blencowe, B. J. Analysis of the requirement for RNA polymerase II CTD heptapeptide repeats in pre-mRNA splicing and 3'-end cleavage. *Rna* **10**, 581-589 (2004).

35 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).

36 Gu, B., Eick, D. & Bensaude, O. CTD serine-2 plays a critical role in splicing and termination factor recruitment to RNA polymerase II in vivo. *Nucleic Acids Res* **41**, 1591-1603 (2013).

37 Chapman, R. D., Palancade, B., Lang, A., Bensaude, O. & Eick, D. The last CTD repeat of the mammalian RNA polymerase II large subunit is important for its stability. *Nucleic Acids Res* **32**, 35-44 (2004).

38 Chapman, R. D., 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* **25**, 7665-7674 (2005).

39 Tsao, D. C., Park, N. J., Nag, A. & Martinson, H. G. Prolonged alpha-amanitin treatment of cells for studying mutated polymerases causes degradation of DSIF160 and other proteins. *Rna* **18**, 222-229 (2012).

40 Bartolomei, M. S., Halden, N. F., Cullen, C. R. & Corden, J. L. Genetic analysis of the repetitive carboxyl-terminal domain of the largest subunit of mouse RNA polymerase II. *Mol Cell Biol* **8**, 330-339 (1988).

41 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).

42 Richard, P. & Manley, J. L. Transcription termination by nuclear RNA polymerases. *Genes Dev* **23**, 1247-1269 (2009).

43 Egloff, S. *et al.* The integrator complex recognizes a new double mark on the RNA polymerase II carboxyl-terminal domain. *J Biol Chem* **285**, 20564-20569 (2010).

44 Kamenski, T., Heilmeier, S., Meinhart, A. & Cramer, P. Structure and mechanism of RNA polymerase II CTD phosphatases. *Mol Cell* **15**, 399-407 (2004).

45 Ghosh, A., Shuman, S. & Lima, C. D. The structure of Fcp1, an essential RNA polymerase II CTD phosphatase. *Mol Cell* **32**, 478-490 (2008).

46 Medlin, J. E., Uguen, P., Taylor, A., Bentley, D. L. & Murphy, S. The C-terminal domain of pol II and a DRB-sensitive kinase are required for 3' processing of U2 snRNA. *The Embo Journal* **22**, 925-934 (2003).

47 Larochelle, S. *et al.* Cyclin-dependent kinase control of the initiation-to-elongation switch of RNA polymerase II. *Nat Struct Mol Biol* **19**, 1108-1115 (2012).

48 Ahn, S. H., Kim, M. & Buratowski, S. Phosphorylation of serine 2 within the RNA polymerase II C-terminal domain couples transcription and 3' end processing. *Mol Cell* **13**, 67-76 (2004).

49 Bird, G., Zorio, D. A. & Bentley, D. L. RNA polymerase II carboxy-terminal domain phosphorylation is required for cotranscriptional pre-mRNA splicing and 3'-end formation. *Mol Cell Biol* **24**, 8963-8969 (2004).

50 Viladevall, L. *et al.* TFIIH and P-TEFb coordinate transcription with capping enzyme recruitment at specific genes in fission yeast. *Mol Cell* **33**, 738-751 (2009).

51 Qiu, H., Hu, C. & Hinnebusch, A. G. Phosphorylation of the Pol II CTD by KIN28 enhances BUR1/BUR2 recruitment and Ser2 CTD phosphorylation near promoters. *Mol Cell* **33**, 752-762 (2009).

52 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).

53 Schwartz, B. E., Larochelle, S., Suter, B. & Lis, J. T. Cdk7 is required for full activation of Drosophila heat shock genes and RNA polymerase II phosphorylation in vivo. *Mol Cell Biol* **23**, 6876-6886 (2003).

54 Glover-Cutter, K., Kim, S., Espinosa, J. & Bentley, D. L. RNA polymerase II pauses and associates with pre-mRNA processing factors at both ends of genes. *Nat Struct Mol Biol* **15**, 71-78 (2008).

55 Bartkowiak, B. *et al.* CDK12 is a transcription elongation-associated CTD kinase, the metazoan ortholog of yeast Ctk1. *Genes Dev* **24**, 2303-2316 (2010).

56 Blazek, D. *et al.* The Cyclin K/Cdk12 complex maintains genomic stability via regulation of expression of DNA damage response genes. *Genes Dev* **25**, 2158-2172 (2011).

57 Bartkowiak, B. & Greenleaf, A. L. Phosphorylation of RNAPII: To P-TEFb or not to P-TEFb? *Transcription* **2**, 115-119 (2011).

58 Svejstrup, J. Q. Transcription: another mark in the tail. *Embo J* **31**, 2753-2754 (2012).

59 Zimmerman, W. C. & Erikson, R. L. Finding Plk3. *Cell Cycle* **6**, 1314-1318 (2007).

60 Jeronimo, C. *et al.* Systematic analysis of the protein interaction network for the human transcription machinery reveals the identity of the 7SK capping enzyme. *Mol Cell* **27**, 262-274 (2007).

61 Xu, Y. X., Hirose, Y., Zhou, X. Z., Lu, K. P. & Manley, J. L. Pin1 modulates the structure and function of human RNA polymerase II. *Genes Dev* **17**, 2765-2776 (2003).

62 Larochelle, S. *et al.* Dichotomous but stringent substrate selection by the dual-function Cdk7 complex revealed by chemical genetics. *Nat Struct Mol Biol* **13**, 55-62 (2006).

63 Hirose, Y. & Manley, J. L. RNA polymerase II is an essential mRNA polyadenylation factor. *Nature* **395**, 93-96 (1998).

Figure legends

Figure 1. CTD mutants and growth properties. (A) The CTD composition of human Rpb1. The CTD of human Rpb1 consists of 52 tandem heptapeptide repeats. 21 out of 52 repeats are composed of conserved consensus $Y^1S^2P^3T^4S^5P^6S^7$ repeats, and the remaining 31 repeats are less conserved. The Rpb1 (1~26) expression vector contains the first half of the CTD, whereas Rpb1 (27~52) contains the second

half. White box: consensus repeat; dark grey box: less conserved repeat; grey box: Rpb1 body without CTD; black box: C terminal 10 amino-acid motif. (B) Growth curves of various mutant cell lines cultured in medium containing tetracycline (tet). Average from two independent experiments was plotted. Cells were split every two days. (C) Western blot shows the protein expression profile of various CTD mutants. Samples were treated with tet (1 ug/ml) for 24 hr.

Figure 2. **Impact of CTD mutations on *Egr1* transcription, splicing and 3' end processing.** (A) 26r cells were cultured in medium containing 1ug/ml tet for 24 hours, and then treated with DMSO or ionomycin and phorbol 12-myristate 13-acetate (PMA) for 20 min to induce *Egr1* expression. *Egr1* induction levels were measured by RT-qPCR, and plotted relative to non-induced control (DMSO treatment). *N*=3. (B) Analysis of *Egr1* induction levels in various cell lines. Cells were treated with tet for 24 hr, and induced by ionomycin and PMA for 20 minutes. *Egr1* mRNA levels were measured using RT-qPCR, and plotted relative to 26r cells. *N*=3. (C) Cells were treated as in (B). The ratios of unspliced to spliced *Egr1* mRNA were determined using RT-qPCR, and plotted relative to the ratios of 26r cells. Diagram depicts the two-exon *Egr1* gene. The two arrows depict primers used to detect spliced products, and the primer set (top) detects unspliced *Egr1*. *N*=3. (D) Cells were treated as in (B). Ratios of uncleaved to total *Egr1* mRNA were measured using RT-qPCR, and plotted relative to the ratio detected in 26r cells. The left primer set measures total *Egr1* mRNA, and the right set detects uncleaved *RNA*. *N*=3. Error bars display standard deviation.

Figure 3. Analysis of transcription/polyadenylation in CTD mutant cells. (A) Cells were treated with tet for 40 hr, and nascent RNA was labeled with ^{3}H uridine for 30 min. Extracted RNA was separated into non-polyadenylated (non-polyA) and polyadenylated (polyA) fractions. Non-polyA and polyA RNAs were quantified by scintillation counting, and counts per minute (CPM) relative to 26r cells were plotted. *N*=3 for Rpb1 (1~52) and S7A samples. *N*=2 for Rpb1 (27~52) sample. (B) Cells were treated with tet for 24 hr.

Chromatin immunoprecipitation (ChIP) was performed as described. The distribution of Flag-tagged Rpb1 on the *Rplp1* gene was determined using qPCR. Diagrams depict genes analyzed. Thick lines represent genes, and dashed lines display transcripts. Triangle denotes 3' cleavage site. For *Rplp1*, amplicon B is at the transcription start site (TSS) and amplicon D covers the 3' cleavage site. *N*=3. (C) ChIP was performed in the β-actin (*Actb*) gene as in (B). Amplicon A is at the TSS and amplicon C covers the 3' cleavage site. *N*=3. (D) ChIP analysis of Rpb1 recruitment on histone H2A, U1, and U2 genes. On H2A, Rpb1 levels at the TSS and 3' cleavage site were measured, whereas on the U1 and U2 genes, Rpb1 levels on the gene body were determined. ChIP was performed as in (B). *N*=3. Error bar indicates standard deviation.

Figure 4. snRNA expression is not affected by S7A mutation. (A) Effects of S7A mutation on snRNA transcription were analyzed using nuclear run-on experiments. Nuclei were harvested from cells treated with tet for 40 hr. Nascent RNA was labeled with ^{32}P-UTP for 30 min and purified and analyzed by slot blot containing DNA oligos as indicated on the top panel. NC: negative control (anti-sense U1). Signals from each gene were normalized to 18S and plotted (bottom panel). *N*=3. (B) Steady-state levels of U1 and U2 snRNAs in two independent cell lines expressing Rpb1-S7A compared with Rpb1 (1~52) and 26r cells. Cells were treated with tet for 6 days. Extracted RNA was analyzed using RT-qPCR. Levels of U1 and U2 RNA were normalized to 18S, and values relative to 26r cells were plotted. *N*=3. Error bars indicate standard deviation. (C) Total U2 snRNA levels in Rpb1 (27~52), S2A and S5A cells. Cells were treated with tet for 40 hours. Levels of total U2 snRNA were measured using RT-qPCR, and plotted relative to 26r cells. *N*=3. (D) U2 snRNA 3' cleavage in Rpb1 (27~52) and S2A cells. Cells were treated as in (C). Ratios of uncleaved / total U2 snRNA were measured using RT-qPCR, and plotted relative to 26r cells. *N*=3. Error bars display standard deviation.

Figure 5. Phosphorylation and dephosphorylation of Thr 4. (A) *In vitro* phosphorylation of Thr 4. GST-CTD was incubated with HeLa nuclear extract (left lane) or the indicated amount of CDK9/cyclinT plus or minus indicated inhibitors. Samples were analyzed by western blotting with antibodies that detect P-Thr 4 (top panel) or P-Ser 2 (bottom panel). (B) *In vitro* dephosphorylation of Thr 4. GST-CTD was phosphorylated by Hela nuclear extract as in (A) and incubated with Fcp1 or Fcp1 (M271E). Samples were analyzed by western blotting with antibodies that detect P-Thr 4 (top panel) or GST (bottom panel). (C) Knock down of Fcp1 causes increased levels of P-Thr4. HEK293 cells were transfected with vectors expressing one of three shRNAs targeting Fcp1 or an shRNA targeting GFP (con). 30 hr after transfection, cell lysates were analyzed by western blotting with the indicated antibodies. (D) Quantification of Fcp1 and P-Thr4 levels after knock-down. Fcp1 protein levels were normalized to actin (left panel) and P-Thr4 was normalized to total Rpb1 (right panel). Ratios relative to control were plotted. *N*=3. Error bars display standard deviation.

Rpb1-CTD mutant	Viable?
Rpb1(1~52)	Yes
Rpb1(1~26)	Yes
Rpb1(27~52)	No
26r $(YSPTSPS)_{26}$	Yes
19r $(YSPTSPS)_{19}$	No
S2A $(YAPTSPS)_{26}$	No
T4V $(YSPVSPS)_{30}$	No
S5A $(YSPTAPS)_{28}$	No
S7A $(YSPTSPA)_{30}$	Yes

Table 1. Viability of various Rpb1-CTD mutants. Cells were selected in the presence of tetracycline (1 ug/ml). Surviving cell clones were isolated, and the identity of these cells was confirmed using western blotting with an anti-Flag Ab. Multiple verified clones were isolated for the viable derivatives, and no clones were detected for any of the non-viable constructs. The number next to the parentheses indicates the number of heptads. Each construct contains an N-terminal Flag-tag and the natural 10 residue C-terminal sequence (ISPDDSDEEN) at its C-terminus.

Supplemental figure legends

Figure S1. Normalization and quantitation of Rpb1 ChIP data. (A) Data from Figure 3B were reanalyzed by normalizing Rpb1 signals to that of TSS. (B) Data from Figure 3C were analyzed as in (A). N=3. Error bars indicate standard deviation.

Figure S2. Ser 2, Ser 5 and Ser 7 phosphorylation patterns in mutant cells. Cells were treated with tet for 24 hours. Cell lysates were analyzed by western blotting. H5, H14, and 4E12 antibodies as indicated were used to detect Ser 2-P, Ser 5-P, and Ser7-P, respectively. (A) Ser 2 and Ser 5 levels. (B) Ser 7 levels.

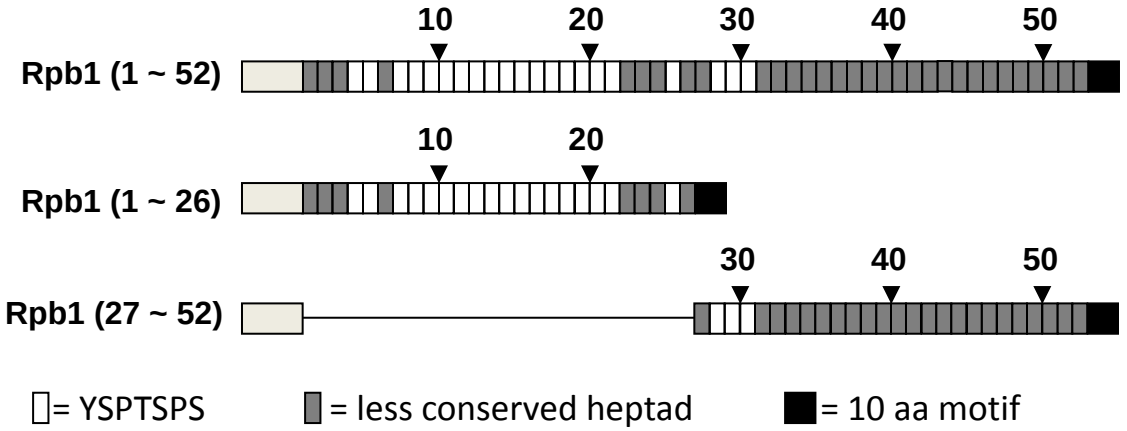
Figure S3. (A) Total U1 snRNA levels in S2A and S5A cells. Cells were treated as in Figure 4C. Levels of total U1 snRNA were measured using RT-qPCR, and plotted relative to 26r cells. *N*=3. (B) U1 snRNA 3' end cleavage in Rpb1 (27~52) and S2A cells. Ratios of uncleaved / total U1 snRNA were measured using RT-qPCR, and plotted relative to 26r cells. *N*=3. Error bar represents standard deviation.

Figure S4. Coomassie blue staining of purified GST-CTD, Fcp1, Fcp1 (M271E), and CDK9/cyclin T complex. Arrow indicates the purified protein resolved by SDS-PAGE.

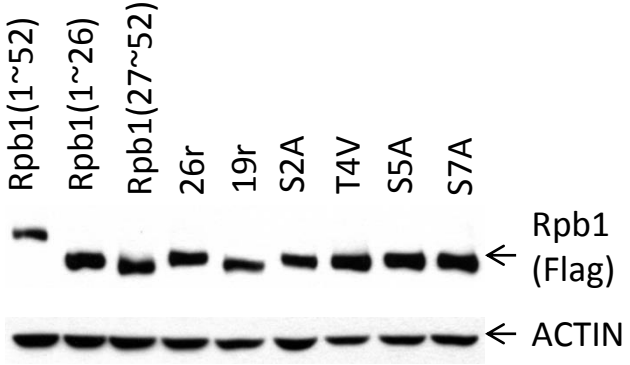
Figure S5. Fcp1 mRNA levels in shRNA-treated HEK293 cells. Cells were transfected with plasmids expressing shRNAs targeting Fcp1 or GFP (control). RNA was extracted 22 h or 41 h after transfection, and RT-qPCR was performed to determine Fcp1 mRNA levels. Levels of Fcp1 mRNA in Fcp1 shRNA expressing cells relative to control were plotted.

Figure 1.

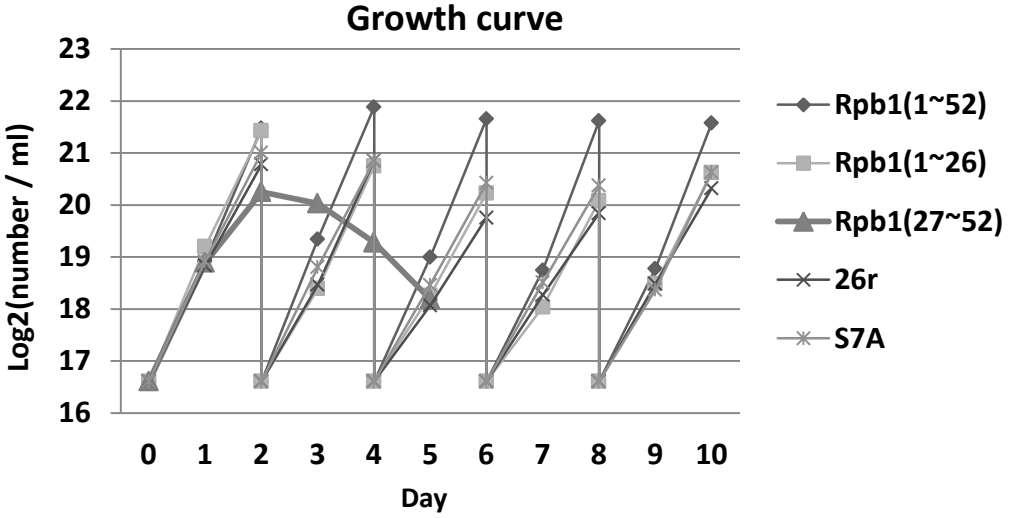
A


10
20
30
40
50
Rpb1 (1 ~ 52)
10
20
Rpb1 (1 ~ 26)
30
40
50
Rpb1 (27 ~ 52)
= YSPTSPS
= less conserved heptad
= 10 aa motif

C


Rpb1(1~52)
Rpb1(1~26)
Rpb1(27~52)
26r
19r
S2A
T4V
S5A
S7A
Rpb1 (Flag)
ACTIN

B


Growth curve
Log2(number / ml)
Day
Rpb1(1~52)
Rpb1(1~26)
Rpb1(27~52)
26r
S7A

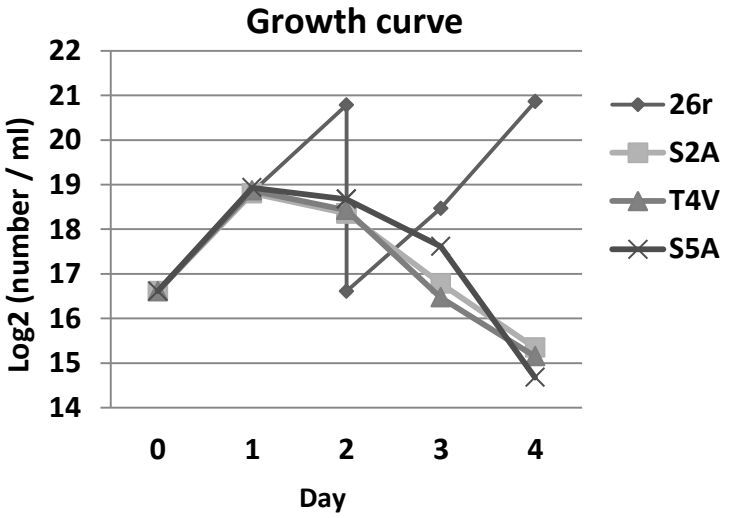
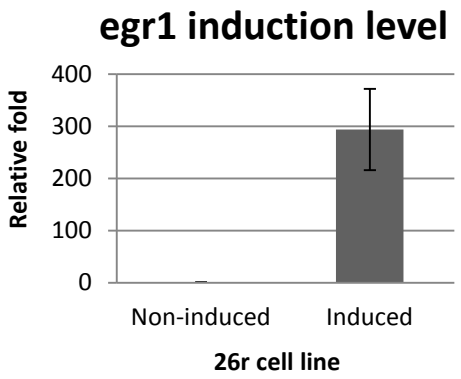
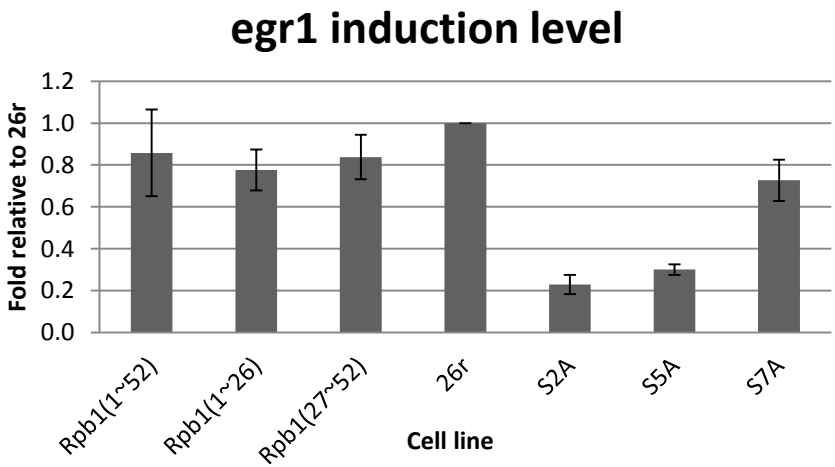

Growth curve
Log2 (number / ml)
Day
26r
S2A
T4V
S5A

Figure 2.

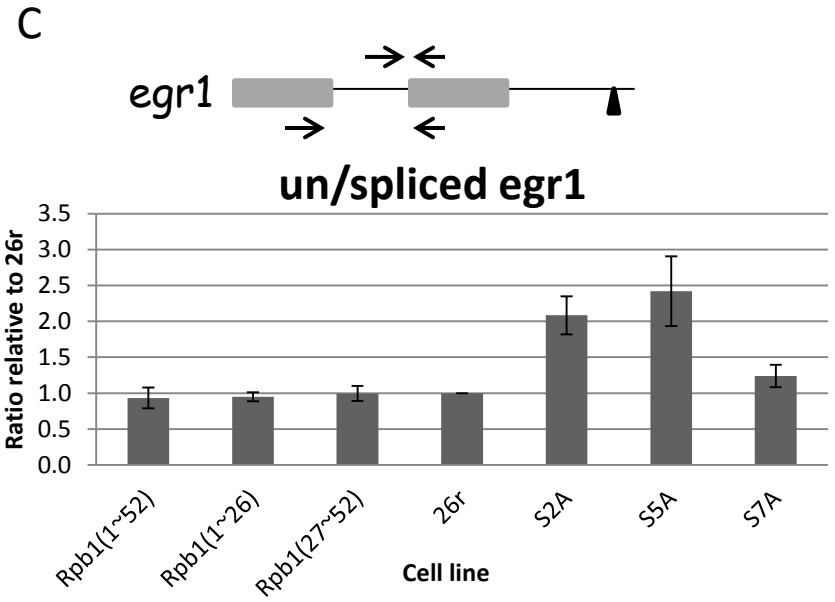
A



B



C



D

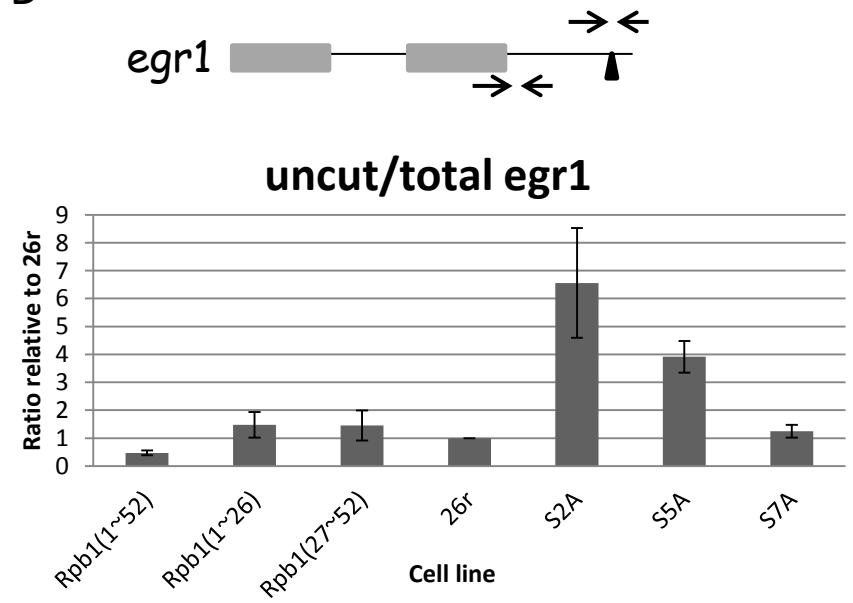
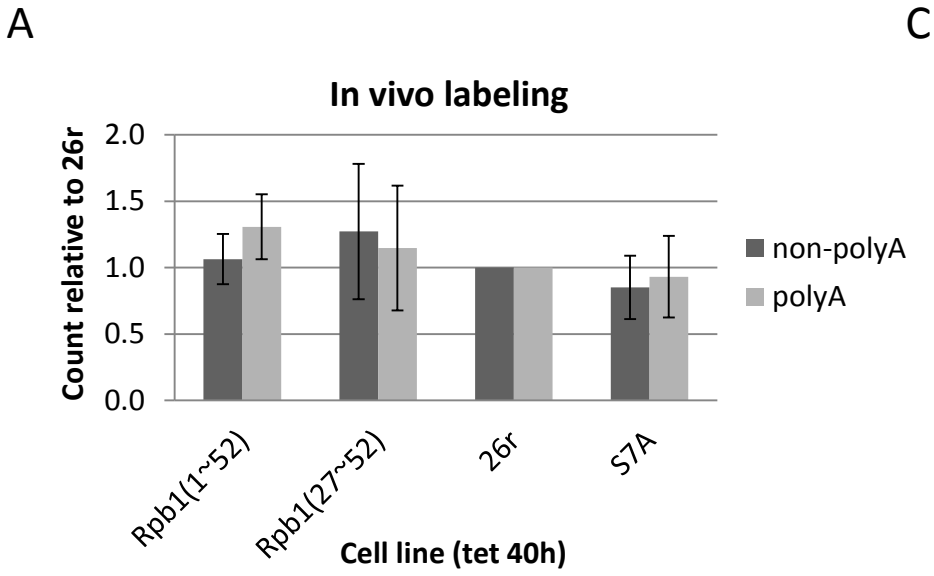
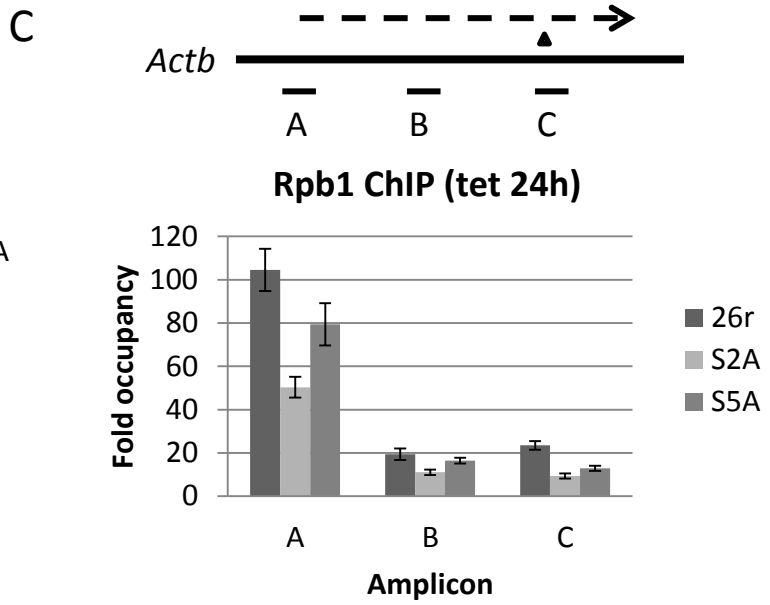
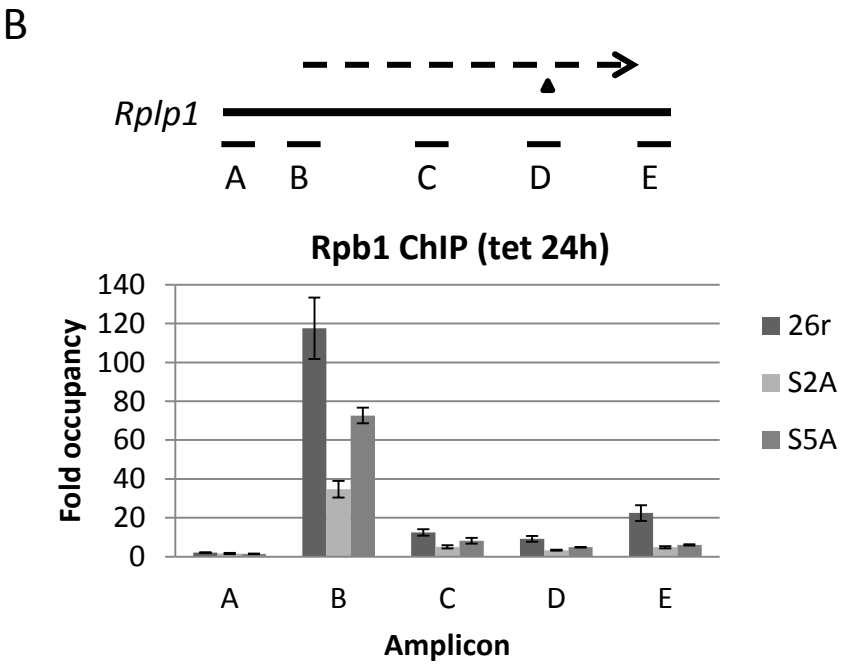


Figure 3.


A
In vivo labeling
Count relative to 26r
2.0
1.5
1.0
0.5
0.0
non-polyA
polyA
Rpb1(1~52)
Rpb1(27~52)
26r
S7A
Cell line (tet 40h)


C
Actb
A
B
C
Rpb1 ChIP (tet 24h)
Fold occupancy
120
100
80
60
40
20
0
26r
S2A
S5A
A
B
C
Amplicon


B
Rplp1
A
B
C
D
E
Rpb1 ChIP (tet 24h)
Fold occupancy
140
120
100
80
60
40
20
0
26r
S2A
S5A
A
B
C
D
E
Amplicon

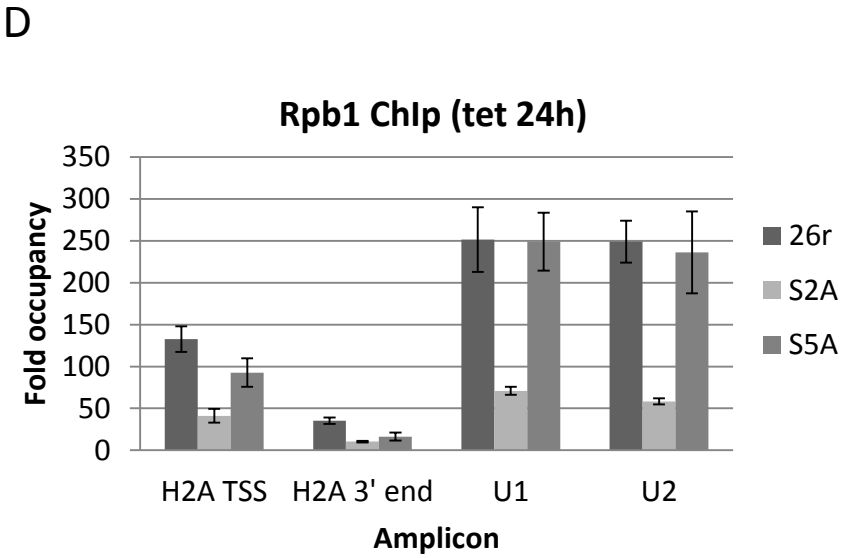
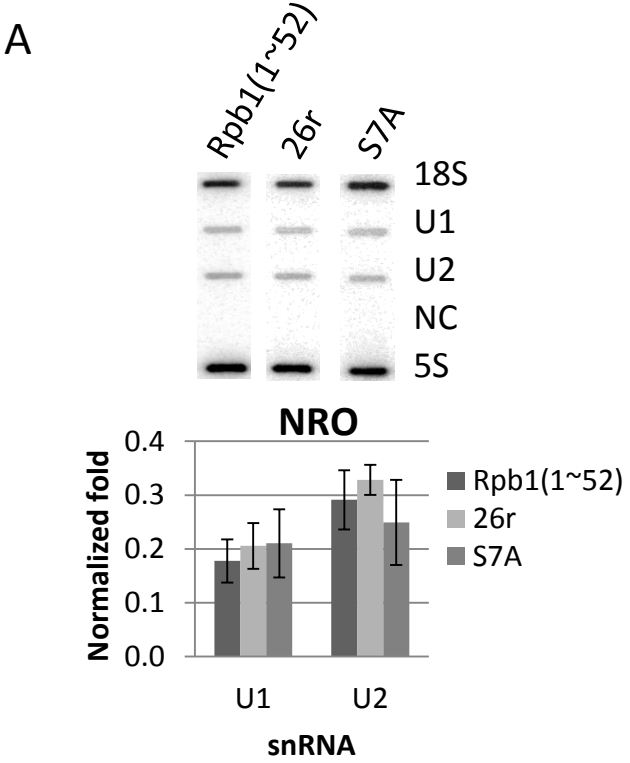
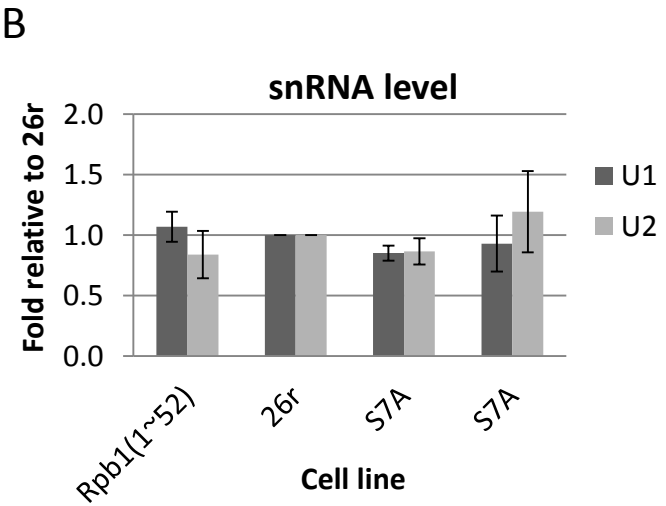
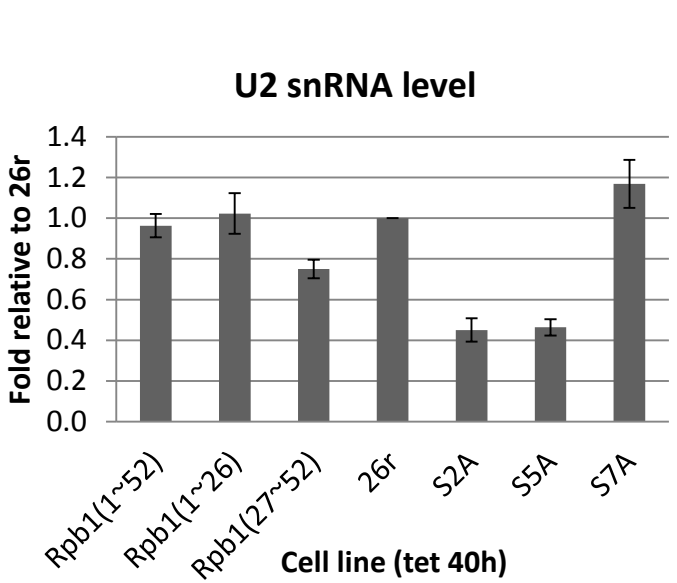

D
Rpb1 ChIp (tet 24h)
Fold occupancy
350
300
250
200
150
100
50
0
26r
S2A
S5A
H2A TSS
H2A 3' end
U1
U2
Amplicon

Figure 4.


A
Rpb1(1~52)
26r
S7A
18S
U1
U2
NC
5S
NRO
Normalized fold
0.4
0.3
0.2
0.1
0.0
Rpb1(1~52)
26r
S7A
U1
U2
snRNA


B
snRNA level
Fold relative to 26r
2.0
1.5
1.0
0.5
0.0
U1
U2
Rpb1(1~52)
26r
S7A
S7A
Cell line


C
U2 snRNA level
Fold relative to 26r
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
Rpb1(1~52)
Rpb1(1~26)
Rpb1(27~52)
26r
S2A
S5A
S7A
Cell line (tet 40h)

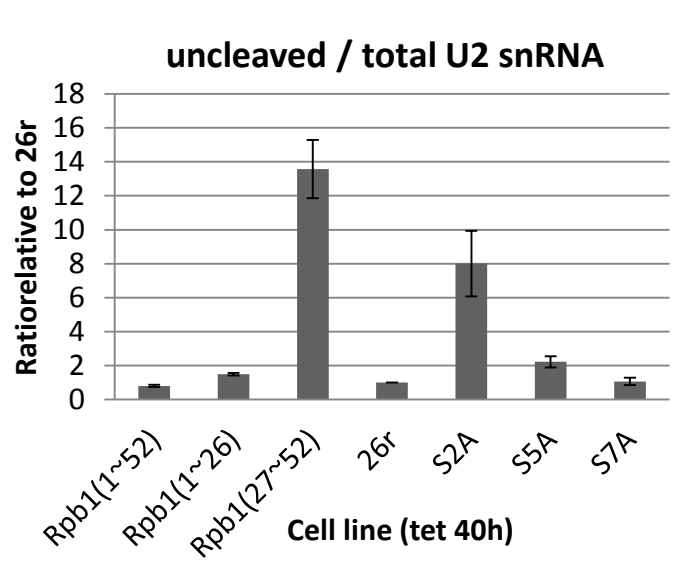

D
uncleaved / total U2 snRNA
Ratiorelative to 26r
18
16
14
12
10
8
6
4
2
0
Rpb1(1~52)
Rpb1(1~26)
Rpb1(27~52)
26r
S2A
S5A
S7A
Cell line (tet 40h)

Figure 5.

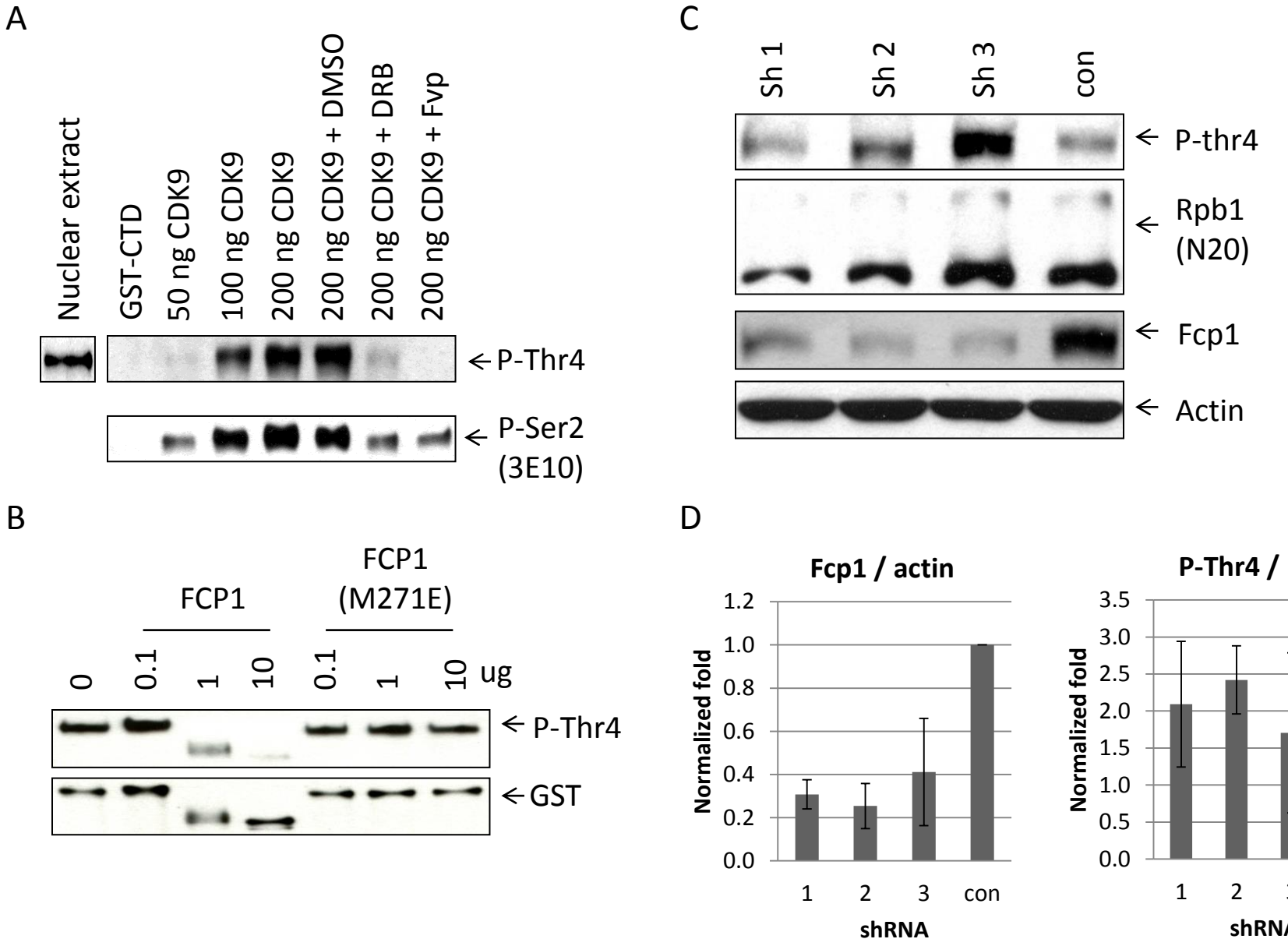

A
Nuclear extract
GST-CTD
50 ng CDK9
100 ng CDK9
200 ng CDK9
200 ng CDK9 + DMSO
200 ng CDK9 + DRB
200 ng CDK9 + Fvp
P-Thr4
P-Ser2 (3E10)
B
FCP1
FCP1 (M271E)
0
0.1
1
10
0.1
1
10
ug
P-Thr4
GST
C
Sh 1
Sh 2
Sh 3
con
P-thr4
Rpb1 (N20)
Fcp1
Actin
D
Fcp1 / actin
Normalized fold
1.2
1.0
0.8
0.6
0.4
0.2
0.0
1
2
3
con
shRNA
P-Thr4 / Rpb1
Normalized fold
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
1
2
3
con
shRNA

Figure S1.

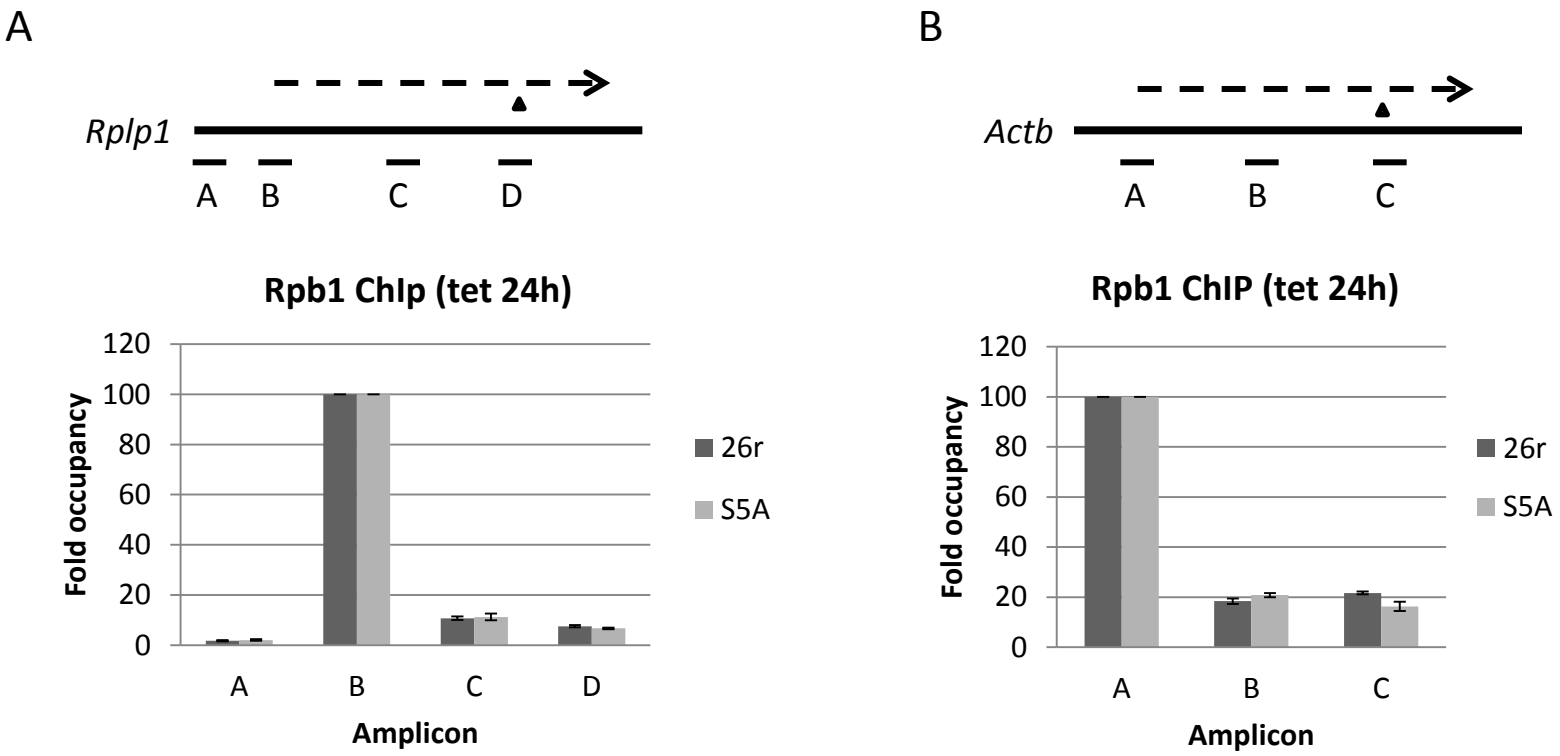

A
Rplp1
A
B
C
D
Rpb1 ChIp (tet 24h)
Fold occupancy
120
100
80
60
40
20
0
A
B
C
D
Amplicon
26r
S5A
B
Actb
A
B
C
Rpb1 ChIP (tet 24h)
Fold occupancy
120
100
80
60
40
20
0
A
B
C
Amplicon
26r
S5A

Figure S2.

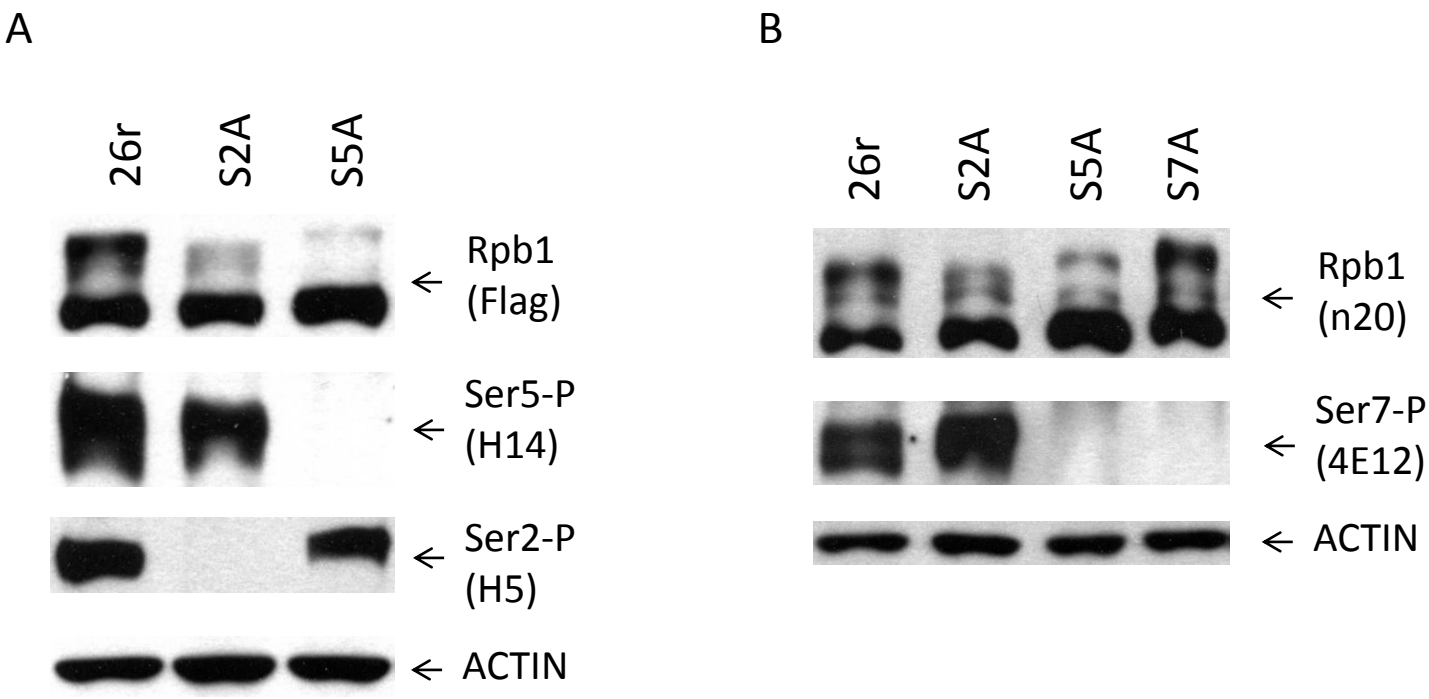

A
26r
S2A
S5A
Rpb1
(Flag)
Ser5-P
(H14)
Ser2-P
(H5)
ACTIN
B
26r
S2A
S5A
S7A
Rpb1
(n20)
Ser7-P
(4E12)
ACTIN

Figure S3.

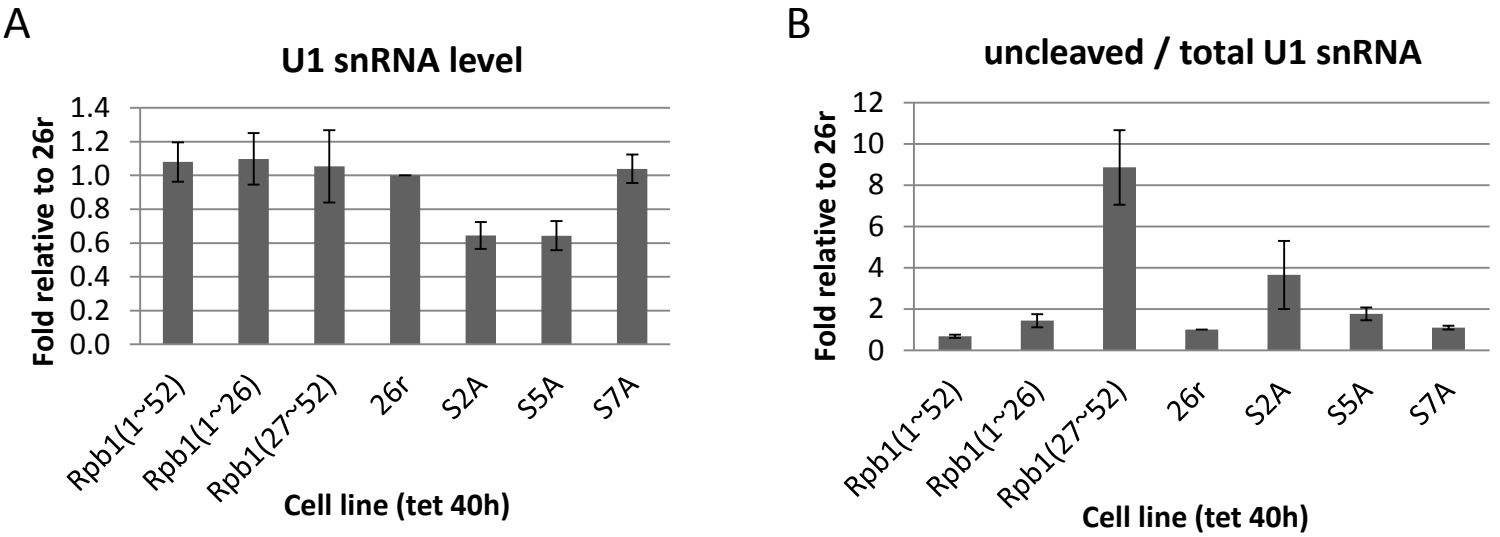

A
U1 snRNA level
Fold relative to 26r
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
Rpb1(1~52)
Rpb1(1~26)
Rpb1(27~52)
26r
S2A
S5A
S7A
Cell line (tet 40h)
B
uncleaved / total U1 snRNA
Fold relative to 26r
12
10
8
6
4
2
0
Rpb1(1~52)
Rpb1(1~26)
Rpb1(27~52)
26r
S2A
S5A
S7A
Cell line (tet 40h)

Figure S4.

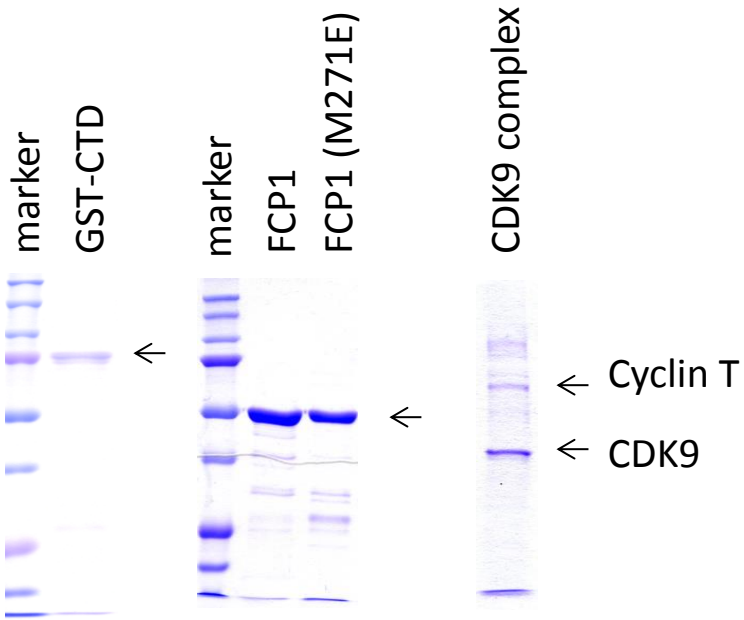
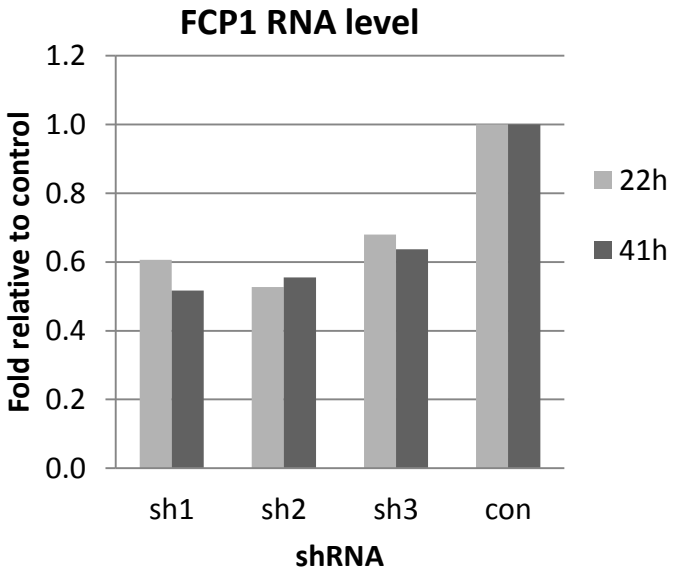

marker
GST-CTD
marker
FCP1
FCP1 (M271E)
CDK9 complex
Cyclin T
CDK9

Figure S5.


FCP1 RNA level
Fold relative to control
1.2
1.0
0.8
0.6
0.4
0.2
0.0
sh1
sh2
sh3
con
shRNA
22h
41h

Chapter 4

RNAP II CTD tyrosine 1 performs multiple functions in vertebrate cells

Jing-Ping Hsin[1], Wencheng Li[2], Bin Tian[2], James L. Manley[1]

[1]Department of Biological Sciences, Columbia University, New York, NY 10027, USA

[2]Department of Biochemistry and Molecular Biology, University of Medicine and Dentistry of New Jersey (UMDNJ)- New Jersey Medical School, Newark, New Jersey, USA

Abstract

The RNA polymerase II largest subunit (Rpb1) contains a unique C-terminal domain (CTD) that is known to coordinate transcription with RNA processing in a phosphorylation-dependent manner. The CTD is composed of tandem $Y^1S^2P^3T^4S^5P^6S^7$ repeats, in which Ser, Thr and Tyr residues can be phosphorylated. We have shown previously that Thr 4 is phosphorylated throughout eukaryotes, and provided evidence that this modification functions in histone mRNA 3' end formation, using genetically tractable chicken DT40 cells. Recently, Tyr 1 was shown to be phosphorylated in *S. cerevisiae*, perhaps functioning to prevent premature transcription termination. Here, we report analysis of CTD Tyr 1 function using DT40 cells. We found that cells expressing an Rpb1 derivative with all Tyr residues mutated to Phe (Rpb1-Y1F) were inviable. Western blotting revealed that Rpb1-Y1F was unstable, degraded to a CTD-less form. Remarkably, Rpb1-Y1F stability, but not cell viability, was fully rescued by restoration of a single Tyr in the last heptad repeat. Indeed, cytoplasmic Rpb1 was phosphorylated exclusively on Tyr 1, and phosphorylation of Tyr 1, but not of other residues, prevented CTD degradation by the 20S proteasome *in vitro*. Further analysis revealed a second function of Tyr 1, in transcription initiation. Consistent with this, Tyr 1 phosphorylation was found by ChIP to peak at promoters, while RNAP II levels were reduced in Y1F mutant cells. Lastly, global RNA sequencing coupled with ChIP analysis revealed a posttranscriptional accumulation of upstream antisense RNAs in cells with Y1F, but not other CTD, mutations, indicative of a role for Tyr 1 in turnover of these RNAs. Together, our data indicate that CTD Tyr 1 and its phosphorylation plays multiple roles in RNAP II function.

Introduction

RNA polymerase II (RNAP II) is a large multisubunit enzyme responsible for transcription of all mRNAs and many non-coding RNAs. Rpb1, the largest RNAP II subunit, contains a unique C-terminal domain, composed of tandem repeats with a consensus sequence of Tyr^{1}-Ser^{2}-Pro^{3}-Thr^{4}-Ser^{5}-Pro^{6}-Ser^{7} (YSPTSPS), in which Tyr, Thr and Ser are all potential sites for posttranslational modifications including acetylation, glycosylation, and phosphorylation. The complexity of organisms is to some degree reflected in the number of repeats in the CTD, e.g. 26 repeats in *S. cerevisiae* up to 52 in vertebrate CTDs. Thus, a great number of modification patterns can be achieved, and through such modifications, the CTD functions dynamically during the transcription cycle, for example to link transcription with RNA processing events such as 5' capping, splicing and 3' end formation (for reviews, see Egloff et al. 2012; Hsin and Manley 2012; Heidemann et al. 2013).

The CTD is extensively phosphorylated during transcription. At promoters, RNAP II is initially bound with an unphosphorylated CTD, but concomitant with initiation becomes phosphorylated on Ser 5 and Ser 7 residues by a cyclin-dependent kinase, CDK7 (Hengartner et al. 1998; Komarnitsky et al. 2000; Akhtar et al. 2009; Glover-Cutter et al. 2009). As RNAP II elongates towards the 3' ends of genes, Ser 2 (Marshall et al. 1996; Cho et al. 2001), and likely Thr 4 (Hsin et al. 2011, 2013; Hintermair et al. 2012) are phosphorylated by the cyclin-dependent kinase CDK9 (P-TEFb), while Ser 5 and Ser 7 are gradually dephosphorylated (Krishnamurthy et al. 2004; Bataille et al. 2012; Zhang et al. 2012). The phosphorylation pattern of the CTD along genes provides a mean to couple different RNA processing reactions with transcription. For example, Ser 5 phosphorylation has been shown to promote recruitment of capping enzyme (Fabrega et al. 2003; Schwer and Shuman 2011), and Ser 2 phosphorylation facilitates recruitment of factors involved in 3' processing (Kim et al. 2004; Meinhart and Cramer 2004; Lunde et al. 2010). In addition, Ser 7 phosphorylation was shown to function in

expression and 3' processing of U1 and U2 snRNAs (Egloff et al. 2007; Egloff et al. 2010), and Thr 4 phosphorylation is important for histone mRNA 3' end formation (Hsin et al. 2011). Ser 2 and Ser 5 are also known to function in other processes, including chromatin modification and pre-mRNA splicing, facilitating recruitment of appropriate factors required for these reactions (Munoz et al. 2010; Spain and Govind 2011; Hsin and Manley 2012).

The function of Tyr 1 phosphorylation (Tyr 1-P) is less well understood. Recent studies in yeast have provided evidence that Tyr 1 is indeed phosphorylated (Mayer et al. 2012), despite the scarcity of Tyr-specific kinases. Genome-wide chromatin immunoprecipitation (ChIP) studies provided evidence that Tyr 1-P occurred in coding regions of genes, and in vitro assays found that Tyr phosphorylation prevented interaction between the CTD and 3' processing/termination factors, e.g. Pcf11 and Rtt103 (Mayer et al. 2012). This has the potential to prevent premature recruitment of termination factors in coding regions of genes, although this remains to be investigated. Consistent with an important function, mutation of Tyr residues to Phe (Y1F) in a truncated CTD in *S. cerevisiae* was lethal (West and Corden 1995), although a similar Y1F mutation in *S. pombe* was viable, displaying only a cold sensitive phenotype (Schwer and Shuman 2011). Whether the requirement of CTD Tyr residues for viability differs between these two yeast species, or can be attributed to experimental differences, these results highlight the importance of examining the functions of Tyr 1-P in metazoan organisms. Tyr 1 was shown to be phosphorylated in mammals by the c-Abl tyrosine kinase (Baskaran et al. 1993), which is absent in yeast (Colicelli 2010). c-Abl binds to a motif consisting of the last ten amino acids (ISPDDSDEEN in vertebrates) of the CTD via a domain at its C-terminal end (Baskaran et al. 1999), and phosphorylates the CTD in a processive manner (Duyster et al. 1995). Intriguingly, this ten residue motif has been shown to be important for Rpb1 stability, as removal of the motif resulted in Rpb1 degradation to a CTD-less form (Chapman et al. 2004, 2005; de la Mata and Kornblihtt 2006). c-Abl, and Tyr 1-P, might therefore be involved in affecting Rpb1 stability, although there is currently no evidence to suggest this.

Rbp1 turnover has been shown in several instances to be mediated by the proteasome. For example, the proteasome has been shown to clear stalled RNAP II in response to DNA damage, in an ubiquitin (Ub)-dependent manner (for review, see Wilson et al. 2013), to be coimmunoprecipitated with RNAP II, and to associate with transcriptionally active genes (Gillette et al. 2004; Geng and Tansey 2012). Studies in *S. cerevisiae* showed that Rsp5 (a ubiquitin E3 ligase) bound to Ser 2 phosphorylated-CTD and, together with a Ub-conjugating enzyme, mono-ubiquinated Rpb1 stalled by DNA lesions, and this led to Rpb1 degradation by the proteasome pathway (Beaudenon et al. 1999; Somesh et al. 2005; Somesh et al. 2007). Similarly, the mouse E3 ligase Wwp2 was found to interact with the CTD, and compromised Rpb1 uniquitination and increased Rpb1 protein levels were detected after Wwp2 knock-down (Li et al. 2007). Wwp2 was shown to be able to ubiquitinate the CTD on lysine residues, indicating that modification of the CTD may play a role in regulating Rpb1 levels in a manner mediated by the proteasome.

Differential CTD phosphorylation plays a role in synthesis and processing of all types of RNAP II transcripts characterized to date. However, one relatively new class of transcripts are upstream antisense (ua) RNAs, which are products of bidirectional transcription that occurs from a large fraction of promoters throughout eukaryotes (Core et al. 2008; He et al. 2008; Seila et al. 2008; Neil et al. 2009; Xu et al. 2009). Unlike downstream, sense RNAs, the upstream transcripts are usually short with a median size of about 250 nucleotides, and are unstable and targeted for degradation by the nuclear exosome (for review, see Seila et al. 2009; Wei et al. 2011). Like their counterpart downstream mRNAs, most upstream transcripts are 5’ capped and cleaved and polyadenylated at their 3’ ends, in mammals likely by the machinery that generates mRNA 3’ ends (Flynn et al. 2011; Preker et al. 2011). How uaRNAs are specifically targeted for rapid degradation is unknown, but one possibility is that differential CTD phosphorylation is involved (Ntini et al. 2013). Although Ser 2 and Ser 5 phosphorylation seems similar during sense and anti-sense transcription (Flynn et al. 2011; Preker et al. 2011), phosphorylation of other residue could contribute to distinguishing theses transcripts.

Here, we describe experiments examining the functions of Tyr 1 of the CTD in vertebrate cells. Using a chicken DT40 cell line we created previously (Hsin et al. 2011), we found that cells expressing exclusively an Rpb1 derivative in which all CTD Tyr 1 residues were mutated to Phe (Rpb1-Y1F) were inviable. Western blots revealed that Rpb1-Y1F was degraded to a size of a CTD-less Rpb1, indicating the involvement of Tyr 1 in Rpb1 stability. However, when a single Phe, in the last heptad repeat of Rpb1-Y1F, was replaced with Tyr, to create Rpb1-25F+Y, we found that degradation was blocked, and that this one Tyr was phosphorylated, suggesting that Tyr phosphorylation stabilizes the CTD. Consistent with this, we show that the purified 20S proteasome degraded unphosphorylated as well as CDK 7-phosphorylated GST-CTD, but not GST-CTD phosphorylated on Tyr 1 by c-Abl. However, despite the stability of Rpb1-25F+Y, these cells were, like Y1F cells, inviable. ChIP analysis revealed that Tyr 1 was phosphorylated on Rpb1 at promoters, and 25F+Y cells were defective in RNAP II transcription initiation. In line with this, genome-wide RNA-Seq data showed an unusual accumulation of upstream antisense (ua) transcripts in these cells, but not in cells harboring other CTD mutant derivatives. As RNAP II levels were reduced in the antisense as well as sense directions, this indicates a posttrancriptional defect caused by the Y1F mutation stabilizes uaRNAs. Together, our data provide evidence that CTD Tyr 1 phosphorylation functions in stabilizing Rpb1, facilitates transcription initiation, and plays a role in ensuring turnover of upstream antisense RNAs.

Results

Tyr 1 is required for viability

We previously utilized a genetic system based on chicken DT40 cells to study the properties of the Rpb1 CTD. We showed for example that an Rpb1 derivative containing a CTD with 26 YSPTSPS repeats (Rpb1-26r) plus the ten C-terminal non-consensus residues necessary for stability (see above) confers cell viability, as did a version in which all Ser 7 residues were mutated to Ala (S7A), while comparable

S2A, T4V and S5A derivatives were inviable (Hsin et al. 2011, 2013). To investigate the functions of Tyr 1, we constructed plasmid encoding an N-terminal Flag-tagged Rpb1 derivative, Rpb1-Y1F, identical to Rpb1-26r but with all Tyr 1 residues replaced by Phe, and expressed this in Rpb1 conditional knock-out cells (DT40-Rpb1, in which Rpb1 expression is driven by a tetracycline (tet)-sensitive promoter; (Hsin et al. 2011). As mentioned above, Tyr 1 was shown to be essential for viability in *S. cerevisiae* (West and Corden 1995), but not in *S. pombe* (Schwer and Shuman 2011), and it is unclear whether it is required for growth in higher eukaryotic cells. We first performed a complementation assay to determine if Rpb1-Y1F could substitute for wild-type Rpb1. In this assay, DT40-Rpb1 cells were transfected with the Rpb1-Y1F vector, and tet was added to the culture to turn off wild-type Rpb1 expression. We found that cells expressing Rpb1-Y1F as the only source of Rpb1 were inviable, whereas Rpb1-26r, as shown previously, fully restored viability.

We next established cell lines stably expressing Rpb1-Y1F to analyze how the Y1F mutation affects cell growth and gene expression. In agreement with the complementation assay, cells expressing Rpb1-Y1F (Y1F) stopped growing around 24 h in medium containing tet. In contrast, cells expressing Rpb1-26r (26r) were as expected fully viable (Figure 1A). To examine whether the inviability of Y1F cells might result from different expression levels of the Rpb1 derivatives, we analyzed several independent Y1F cell lines by Western blotting with anti-FLAG antibodies. As shown in Figure 1B, levels of Rpb1-Y1F were reduced by at least half compared to Rpb1-26r. Additionally, accumulation of a truncated, lower molecular weight form (indicated by *) was observed in all Y1F cell lines. This corresponds to a derivative likely precisely lacking the CTD, as it was detected with FLAG Abs and also migrated slightly more rapidly than an Rpb1 derivative containing six heptads (Figure 1C). As Rpb1-26r and -Y1F transcript levels were equivalent (results not shown), these results indicate that a function of CTD Tyr1 residues is to ensure Rpb1 protein stability. It is conceivable that the inviabilty of Rpb1-Y1F cells is due to the significantly reduced levels of full-length Rpb1.

To begin to investigate the basis for the instability of Rpb1-Y1F, we determined how many Tyr1 residues were necessary to restore stability. We first constructed and expressed a derivative of Rpb1-Y1F in which the F residues in the C-terminal six heptads were reverted to Y, and found this Rpb1 derivative was completely stable (results not shown). Next we constructed an Rpb1-Y1F derivative in which only a single F, in the C terminal-most heptad, was changed back to Y, and isolated cell lines stably expressing this derivative (Rpb1-25F+Y). Strikingly, this single Tyr residue was sufficient to prevent Rpb1 degradation, as the truncated isoform, which we denote Rpb1-b, was absent, and Rpb1-25F+Y levels were comparable to Rpb1-26r in several independent 25F+Y cell lines (Figure 1D). However, despite the fact that Rpb1 stability was restored, 25F+Y cells, similar to Y1F cells, ceased growing and died around 24 h of growth in the presence of tet (Figure 1A and 1E).

We next set out to investigate how a single Tyr residue in the C-terminal heptad repeat stabilizes Rpb1. A first question was whether Rpb1 is indeed Tyr1 phosphorylated in DT40 cells. To address this, we utilized an anti-phospho-Tyr1 Ab (Mayer et al. 2012) to examine the CTD Tyr 1 phosphorylation status of Rpb1-25F+Y and Rpb1-26r by Western blot. The results (Figure 2A) indicate that both proteins were indeed phosphorylated on Tyr 1, supporting the idea that Tyr 1 phosphorylation (Tyr 1-P) is important for Rpb1 stability. We next investigated where in the cell the Rpb1-b isoform accumulates. We prepared cytoplasmic, nuclear, and chromatin-bound fractions from 26r and Y1F cells and analyzed them by Western blot with an N-terminal Rpb1 Ab (N20). Rpb1-b (indicated by *) was detected in all three fractions from the Y1F cells, but barely or not at all in the 26r cytoplasmic and nuclear fractions (Figure 2B). Absolute levels were highest in the nucleoplasm, while Rpb1-b was essentially the only form on Y1F chromatin. As anticipated, Rpb1-b was not detected in 25F+Y cell fractions (Figure S1A). We next determined whether Tyr 1-P could also be detected on Rpb1 in all three subcellular fractions, in this case using extracts from wild-type DT40 (Figure 2C) and human HEK293 (Figure S1B) cells. Indeed in both cell types, robust Tyr 1-P was detected in all three fractions. It is

notable that in the cytoplasm and nucleoplasm, Tyr 1-P was observed only on hypophosphorylated Rpb1 (the lower band), while it was found primarily on the hyperphosphorylated isoform on chromatin. This suggests both that CTD phosphorylation is limited to only a few residues in the cytoplasm and nucleoplasm and that Tyr1-P is present on hyperphosphorylated RNAP II found on active genes (see Discussion). We also examined phosphorylation on CTD Ser 2, 5 and 7 and Thr 4 (Figure 2C and S1B). Strikingly, all these modifications were nearly undetectable in cytoplasmic and nuclear fractions, present almost exclusively on chromatin-associated, hyperphosphorylated Rpb1. Together, our data show that Tyr 1, and only Tyr 1, is phosphorylated before RNAP II engages in transcription, and support the idea that Tyr 1-P functions in stabilizing the CTD when RNAP II is not transcribing, and perhaps also plays a role during transcription.

Tyr 1 phosphorylation prevents Rpb1 degradation by the 20S proteasome

We next wished to determine how the CTD is degraded, and whether Tyr 1-P indeed plays a role. Since one Tyr residue in the final heptad repeat confers stability to the CTD, it is unlikely that endoproteolytic cleavage occurred between the CTD and the body of Rpb1. Consistent with this, a CTD fragment resulting from an internal cleavage was not detected using CTD Abs by us (results not shown) and others (Chapman et al. 2005). One possibility is that the CTD is degraded by the proteasome. The proteasome was shown to co-IP with RNAP II, and to associate with active genes (Gillette et al. 2004; Geng and Tansey 2012). Additionally, certain naturally unstructured proteins can be degraded by the proteasome in a Ub-independent manner (Sheaff et al. 2000; Tofaris et al. 2001). Since the CTD is considered a structure-less domain (Meinhart et al. 2005), we hypothesized that the CTD is a natural proteasome substrate, and that Tyr 1-P might prevent it from being digested by the proteasome. To test this directly, we performed *in vitro* proteasome assays using a GST-CTD substrate and purified 20S proteasome. Following incubation, products were resolved by SDS-PAGE and detected by Western blot.

Using an anti-GST Ab, the amount of full-length GST-CTD was greatly diminished and a ladder-like pattern of low molecular weight bands was detected, indicating that the 20S proteasome degraded the GST-CTD protein from the C-terminus (Figure 3A, lane 1 and 2). Consistent with this, the low molecular weight products were undetectable by western blotting using an antibody (8WG16) against the CTD (Figure 3B). Notably, the 20S proteasome used in this assay was in a latent status with a closed gate and minimal enzymatic activity (Groll et al. 2000; Liu et al. 2003). Thus, GST-CTD, like for example the unstructured protein p21 (Liu et al. 2003), was capable of activating the 20S proteasome. Low concentrations of SDS have been used to render the 20S proteasome gate disordered, leading to proteasome activation (Forster and Hill 2003). Indeed, addition of 0.01% SDS to reaction mixtures increased CTD degradation (Figure 3A, lane 3 and 4). In contrast, the proteasome inhibitor MG132 inhibited CTD degradation (Figure 3A, lane 5 and 6).

We next investigated whether phosphorylation of GST-CTD affects its stability in the proteasome assay. For this, we used a recombinat c-Abl derivative to phosphorylate GST-CTD. This resulted in time-dependent conversion of a fraction of the GST-CTD to a low-mobility, Tyr 1-P isoform, although the majority remained unphosphorylated (Figure 3C), which is consistent with the processive phosphorylation observed previously (Duyster et al. 1995). We then performed the proteasome assay describe above using the c-Abl-modified GST-CTD (Figure 3D). Strikingly, the Tyr 1 hyperphosphorylated GST-CTD (top panel, upper band, and lower panel) was resistant to degradation by the proteasome (lanes 1-2), while the remaining unphosphorylated GST-CTD (top panel, bottom band) was degraded as above. Addition of 0.01% SDS again promoted degradation of unphosphorylated GST-CTD, but Tyr 1 phosphorylated isoform was resistant (lane 2 and 3). Significantly, GST-CTD phosphorylated by CDK7, which converted essentially all of the susbstrate to the hyperphosphorylated form, was not protected from degradation (lane 4, 5 and 6), indicating a specific role of Tyr 1 phosphorylation in preventing degradation by the proteasome. These data provide substantial evidence that Tyr 1-P prevents CTD

degradation by the proteasome, and are entirely consistent with the effects of the Y1F mutation on Rpb1 stability observed *in vivo*.

Transcription defects in 25F+Y cells

As shown above, 25F+Y cells are inviable, even though the Rpb1 protein is stable and expressed at comparable levels to 26r cells. To investigate whether this reflects a defect in transcription, we first used ChIP assays to measure Rpb1 levels on select genes in 25F+Y cells. The results indicated that Rpb1 levels were greatly reduced (~70%) compared to 26r cells on several genes, including a ribosomal protein gene, *Rplp1* (Figure 4A), β-actin (Figure 4B), a histone H2A gene (Figure S2), and U1 and U2 snRNA genes (Figure S2). Decreased Rpb1 levels were observed at transcription start sites (TSS), coding regions, and 3' untranslated regions (3' UTR) in the examined genes. To extend these results, we examined transcription of an inducible gene, *Egr1,* which we showed previously can be induced ~300-fold in 26r cells by a 20-min treatment with ionomycin and PMA (Hsin et al. 2011). Consistent with the ChIP results, *Egr1* mRNA levels following induction were reduced ~five folds in two independent 25F+Y cell lines (Figure S3A). Furthermore, we also detected a two-fold increase in unspliced *Egr1* pre-mRNA and a six-fold increase in 3' uncleaved RNAs in 25F+Y cells, suggesting defects in splicing and 3' end processing (Figure S3B and S3C). Together, we conclude that transcription as well as RNA processing were affected by the Y1F mutation.

We next examined the Tyr 1-P status of RNAP II along the length of actively transcribed genes. For this, we performed ChIP assays with 26r cells using an anti-Tyr 1-P Ab. We found that Tyr 1-P levels were highest at the TSS on the *Rplp1* (Figure 4C) and *β-actin* (Figure 4D) genes, but detectable downstream as well. Reanalyzing the ChIP results by normalizing the phosphotyrosine signal to Rpb1 showed that Tyr 1-P increased modestly in the coding regions, and its levels were maintained at and beyond polyadenylation sites of these transcripts (Figure S4). These results, which are somewhat different than

those observed with *S. cerevisiae* (Mayer et al. 2012; see Discussion) indicate that Tyr 1-P is present on active genes, and are consistent with the results shown above indicating that Tyr 1-P was detected exclusively on hyperphosphorylated Rpb1 in the chromatin fraction (Figure 2C). To extend these results, we next examined if Tyr 1-P could be found globally in Rpb1 phosphorylated on Ser 5 and Ser 2. Indeed, Tyr phosphorylation was detected on Rpb1 IPed by Abs recognizing Ser 5-P (3E8) and Ser 2-P (3E10) (Figure S5A). Taken together, these results provide evidence that Tyr 1-P plays roles throughout the transcription cycle.

The above data supports the view that Tyr 1-P plays a direct role in transcription initiation by RNAP II. To investigate this further, we performed ChIP assays to examine occupancy of general transcription factors, specifically the TATA box binding protein (TBP) and TFIIB, at gene promoters in 25F+Y cells after 24 hrs growth in tet. Compared to 26r cells, TBP levels at the TSS of both the *Rplp1* and *Actb* genes were reduced by about 40% (Figure 5A and 5B), and TFIIB levels by ~60% (Figure 5C and 5D). The reduced TBP and TFIIB occupancy was not a result from decreased levels of these proteins in 25F+Y cells (Figure S5B). In fact, TFIIB protein levels were increased ~ two fold in both Y1F and 25F+Y cells compared to 26r cells. Together, these data suggest that Tyr 1 is important for efficient preinitiation complex assembly at RNAP II promoters.

Increased levels of upstream antisense RNAs in 25F+Y cells

We next sought to investigate the genome-wide phenotypes of 25F+Y cells by RNA deep sequencing. RNA from 25r and 25F+Y, as well as for comparison S2A and S5A, cells treated with tet for 24 h was processed for sequencing, and analyzed as described (Hoque et al. 2013). This method is designed to identify and quantitate levels of all polyadenylated RNAs in a cell. We generated a total of ~8 million reads for each sample, which were mapped onto the chicken genome. We were able to define more than 46% of the total reads as authentic in each sample (Figure S6). Next, we analyzed the

transcripts in about 3,500 chicken reference protein coding genes. The reads were then classified into three types according to poly(A) site location: in the most 3' exon (D), in the body of the gene (G), and in the upstream region of the gene (U). Significantly up-regulated and down-regulated transcripts, compared with 26r cells, that passed Fisher's Exact Test with a *p* value <0.05 were documented, and categorized according to type. Unexpectedly, we found significantly increased abundance of U-type transcripts in about one hundred genes in 25F+Y cells (Figure 6A). This strong tendency was not observed in Rpb1-S2A and Rpb1-S5A cells, suggesting it is specific for Tyr 1 mutation. Notably, the number of annotated genes with U-type transcripts may have been underestimated because the employed RNA-Seq assay, for technical reasons, selectively sequenced RNAs with poly(A) tails of at least 45 adenine bases, and this type of transcript may have atypically short tails (see Discussion). Further analysis showed that the direction of these U-type reads was antisense to the downstream neighboring genes, and no reads in the sense direction were detected. Using RT-qPCR, we validated several of the U-type transcripts. As shown in Figure 6B, about two-fold more upstream transcripts within one kilo-base of the *ARGLU1*, *METTL14*, *SH3PB5* and *WEE1* genes were found in two independent 25F+Y cell lines, consistent with results from RNA-seq. While a significant increase of U transcripts upstream of the *YWHAH* gene was observed by RNA-Seq, only a slight increase was detected by RT-qPCR. Importantly, UA transcripts were undetectable in *RPLP1* and *CCNB2* genes by both methods.

We next wished to gain insight into the mechanism underling the enhanced accumulation of U-type transcripts in 25F+Y cells. As mentioned above, upstream antisense (ua) RNAs have been found across organisms (Seila et al. 2009; Wei et al. 2011). These transcripts are low-abundant,usually rapidly degraded by the nuclear exosome. Increased levels of ua transcripts, also known as PROMPTs, were reported after knock-down of several subunits of the exosome (Preker et al. 2008; Ntini et al. 2013). While mRNA levels of several exosome subunits were somewhat reduced in 25F+Y cells, based on the RNA seq analysis, protein levels were unaffected. Western blotting revealed that levels of two subunits,

Exosc9 and 10, were comparable in 26r and 25F+Y cells (Figure S5C), suggesting that the increase in ua transcript levels in 25F-Y cells was likely not due to decreased levels of the exosome. Another possibility was that transcription of these transcripts was increased. To address this, ChIP assays were performed to determine if Rpb1 levels were elevated in regions encoding several of these transcripts in 25F+Y cells. Interestingly, we found that Rpb1 levels were in fact reduced, not increased, upstream of the *ARGLU1*, *METTL14*, *SH3PB5*, and *WEE1* genes (Figure 6C). This indicates that the observed accumulation of ua RNAs results from a posttranscriptional defect. Thus our results point to a role for CTD Tyr 1 in regulating accumulation of uaRNAs by contributing to their rapid turnover. The implications of this finding, as well as other Tyr 1 functions we have described, will be discussed below.

Discussion

In this study, we investigated the functions of the RNAP II CTD residue Tyr 1 using chicken DT40 cells. We have used DT40 cells previously to study a number of factors involved in gene expression, and found them to serve as an excellent model for elucidating functions of evolutionarily conserved proteins. Most recently, we analyzed the functions and requirement of other residues in the CTD heptad repeat. We found for example that cells with mutations altering all Ser 2, Thr 4 or Ser 5, but not Ser 7, were inviable (Hsin et al. 2011, 2013), and we can now add Tyr 1 to the list of essential CTD residues. However, unlike the other mutant Rpb1 proteins, we found that Rpb1-Y1F was unstable. A truncated isoform, denoted Rpb1-b, was found to accumulate in these cells, reflecting precise, or nearly so, loss of the CTD. The involvement of Tyr 1 phosphorylation in stabilizing the CTD was suggested by the restored stability of Rpb1-Y1F brought about by reintroduction of a single Tyr 1 in the last heptad. We further showed that the CTD is degraded by the 20S proteasome *in vitro*, and that Tyr 1 phosphorylation prevents this degradation. Rpb1 protein stability conferred by Tyr phosphorylation was not the only function of Tyr 1, however, as 25F+Y cells, expressing a stable Rpb1, displayed defects in transcription initiation and RNA

processing. Using RNA deep sequencing, we discovered increased accumulation of upstream antisense RNAs in 25F+Y cells, and attributed this to yet another function of Tyr 1, which is to facilitate turnover of these transcripts. The importance of these findings will be discussed below.

The instability of Rpb1-Y1F was unexpected. One explanation was that this reflects the inviability of the Y1F cells when grown in tet, and the labile, unstructured CTD was simply a substrate for proteolysis in the dying cells. For example it has been known almost since the discovery of RNAP II that the CTD can be readily proteolyzed during purification, generating an isoform known as RNAP IIB (Kim and Dahmus 1986). However, this seems very unlikely for a number of reasons. First, the Rpb1-b isoform is present even when cells are grown in the absence of tet, and are perfectly healthy (unpublished data). Second, none of the several other inviable Rpb1 derivatives reconstructed have shown any indication of instability. Third, the subcellular fractionation data revealed that Rpb1-b was first detected in the cytoplasm, but the ratio to full-length was relatively low. The ratio increased in the nucleoplasm, and indeed virtually all of the chromatin-bound Rpb1-Y1F was CTD-less form. This is consistent with the view that the CTD was gradually proteolyzed beginning soon after translation and entirely removed by the time the CTD-less RNAP II reached chromatin.

Our finding that Tyr 1 phosphorylation, catalyzed by c-Abl, plays an important role in stabilizing Rpb1 provides a parsimonious explanation to a number of previous findings. A possible role for Tyr 1-P in stabilizing Rpb1 was first suggested by our unexpected finding that both cytoplasmic and nucleoplasmic Rpb1 contains Tyr 1, and only Tyr 1, phosphorylation. This is consistent with our detection of Rpb1-b in both cytoplasm and nucleus of Y1F cells, and raised the possibility that Tyr 1-P normally stabilizes Rpb1 in these compartments. These results, coupled with previous findings mentioned in the Introduction, that the C-terminal ten residues of the CTD are required for Rpb1/CTD stability (Chapman et al. 2004; Chapman et al. 2005; de la Mata and Kornblihtt 2006), and that they also

constitute a c-Abl binding site necessary for processive CTD phosphorylation (Duyster et al. 1995; Baskaran et al. 1999), are nicely reconciled by our finding that c-Abl- mediated Tyr 1-P is sufficient to stabilize the CTD from proteasomal degradation *in vitro*. Our finding that a single Tyr 1 residue in the C-terminal heptad repeat is sufficient to stabilize Rpb1 is also consistent with these observations. Although we detected somewhat reduced levels of Rpb1 in DT40 cells upon treatment with c-Abl kinase inhibitors, e.g. imatinib and nilotinib (Rix et al. 2007), no Rpb1-b was detected as it was in Y1F cells (unpublished data). One plausible explanation for this is that secondary effects of these drugs interfered in some way with Rpb1 metabolism. Another explanation may be the existence of unknown tyrosine kinases capable of phosphorylating Tyr 1 that are not inhibited by these drugs.

It is intriguing that that the 20S proteasome degrades the CTD lacking Tyr 1-P in a Ub-independent manner. The proteolytic activities of the proteasome lie within the 20S proteasome, which is composed of four stacked rings to form a barrel shape structure with proteolytic sites inside (Groll et al. 1997). To prevent inappropriate digestion, the entry of substrates into the proteolytic chamber is blocked by a gate formed by the N-terminal tails of the outer-ring subunits (Groll et al. 2000). Our *in vitro* proteasome assay revealed that the latent 20S proteasome was able to digest the CTD without activation, i.e., by the addition of SDS, suggesting the CTD can interact with amino acid residues surrounding the gate to render it disordered, allowing access into the catalytic chamber. Several proteins have been reported to display this type of interaction with the proteasome, e.g. $p21^{Cip1}$ and α-synuclein (Touitou et al. 2001; Liu et al. 2003). This behavior of the CTD supports the idea that the naturally unfolded CTD is a substrate for the 20S proteasome, and that Tyr 1-P functions to prevent such degradation in cells. A plausible mechanism to explain how Tyr 1-P prevents CTD degradation by the proteasome is that phosphorylation prevents an interaction between the CTD and the proteasome, thereby preventing passage through the gate. Another possibility, not mutually exclusive, is that Tyr 1-P provides steric hindrance that prevents the CTD passing through the gate. Whatever the mechanism, it will be interesting in the future to

determine how, and if, regulation of CTD Tyr 1-P contributes to control of gene expression by modulating Rpb1 levels in cells.

The inviability of Y1F cells extends beyond CTD instability, and is consistent with our finding that Tyr 1, and likely Tyr 1-P, is important for several steps in the gene expression pathway. In *S. cerevisiae*, Tyr 1-P was found by genome-wide ChIP to increase during transcription elongation, peaking before the 3' cleavage sites, and it was suggested that this pattern functions to prevent premature recruitment of termination factors, e.g. Pcf11, and Rtt103 (Mayer et al. 2012). In contrast to these findings, we detected high levels of Tyr 1-P around the TSS, increasing only mildly thereafter. This difference suggests that this modification might function differently in vertebrate cells than in yeast cells. Indeed, we provided evidence showing that Tyr 1-P functions in transcription initiation, and appears in fact, based on ChIP analyses, to be required for efficient PIC formation *in vivo*. This is consistent with our finding that Tyr 1, and only Tyr 1, is phosphorylated prior to recruitment to promoters. The involvement of the CTD in PIC assembly was reported previously, as PIC assembly was impaired in extracts from yeast cells expressing an Rpb1 carrying a CTD with 9 heptads, instead of full-length 26 repeats (Ranish et al. 1999). However, partly because of the flexibility of the CTD, insightful structural studies regarding the interaction between the CTD and PIC are lacking (Meinhart et al. 2005). A recent cryo-EM study showed that the CTD is in a close proximity to CDK7/cyclin H (He et al. 2013), which is intuitive since CDK7 phosphorylates Ser 5 and perhaps Ser 7 during initiation. How Tyr 1-P might function in PIC assembly remains to be determined, but its involvement emphasizes the importance of the CTD in influencing the PIC assembly. It will be of interest in the future to understand the differences in Tyr 1 function in *S. cerevisiae* and vertebrates, as well as the now surprising finding that Tyr 1 is dispensable for viability of *s. pombe* (Schwer and Shuman 2011).

Our results have also provided evidence that Tyr 1 plays an unanticipated role in ensuring rapid turnover of upstream antisense RNAs. Previous reports have provided evidence that uaRNAs are present at the majority of active genes in humans (Almada et al. 2013), but only ~200 genes (of ~3500 reference genes) revealed associated uaRNAs in our RNA-Seq analysis. This difference likely reflects for the most part the limited sequencing depth of our analysis. However, it might also be due our strategy for purifying polyadenylated RNA (Hoque et al. 2013). As mentioned above, we sequenced polyadenylated RNAs with at least 45 As at their 3' ends, which is almost double the length of the minimal poly(A) tail of the RNAs analyzed by (Almada et al. 2013). This selection, which was employed to avoid priming at internal oligo(A) tracts, would exclude transcripts with very short poly(A) tails. However, the importance of our findings are not diminished by relatively low number of uaRNAs identified, as we detected upregulated uaRNA accumulation in fully half of the 200 reference genes with the associated uaRNAs analyzed in 25Y+F cells treated with tet for only 24 h. Coupled with the fact that this was not detected in S2A or S5A cells, this highlights the importance of Tyr 1 in uaRNA metabolism.

How might Tyr 1 contribute to ensuring low accumulation of uaRNA? Our RNAP II ChIP data indicated that transcription of these transcripts is in fact decreased in 25F+Y cells. Therefore, a significant change in the turnover rate of ua RNAs must be responsible for the elevated levels detected in these cells. As mentioned above, the abundance of uaRNAs is usually very low due to their sensitivity to the nuclear exosome. Since these transcripts, which contain typical mRNA polyadenylation signals and appear to be cleaved and polyadenylated at by the machinery that generates 3' ends of their downstream counterpart mRNAs (Ntini et al. 2013), what is the mechanism that distinguishes upstream and downstream RNAs for degradation by the nuclear exosome? One possibility is that the UA transcripts contain some cis-elements that can interact with and recruit the exosome. However, another possibility, not mutually exclusive, is that the CTD, known to function in recruitment of factors involved in 3' end formation of all other classes of RNAP II transcripts (Hsin and Manley 2012), specifically recruits

the exosome or an associated factor to facilitate rapid degradation of uaRNAs, and that this recruitment is dependent on Tyr 1-P.

In summary, our studies have provided evidence that Tyr 1 phosphorylation of the RNAP II CTD plays multiple important roles in RNAP II structure and function. We can thus add this modification to the repertoire of CTD modifications that play important roles in regulating and orchestrating gene expression in eukaryotic cells.

Materials and Methods

Cell culture and cloning

DT40 cells and HEK293 cells were cultured at 37^{o}C with 5% CO_2 in RPMI1640 medium containing 10% FBS and 1% chicken serum, and in DMEM medium containing 10% FBS, respectively.

Rpb1 CTD derivatives were cloned as previously described (Hsin et al. 2011). Briefly, a fragment of beta-actin promoter and FLAG tag was inserted into pBlueScript containing Neomycin resistance gene. The human *Rpb1* body without the CTD was inserted immediately after the FLAG tag, and various CTD fragments were inserted directly 3' to the *Rpb1* body.

Complementation assay and construction of stable cell lines

Procedures for complementation assays and for constructing stable cell clines were followed as previously described (Hsin et al. 2011). Briefly, 10^7 cells were transfected with linearized DNA, and selected in the presence of appropriate antibiotics. Surviving cell clones were isolated, and the identity of these cells was further confirmed using western blotting.

Western blotting

Cells lysates were resolved in SDS-PAGE with indicated percentage of acrylamide. Western blotting was performed using standard protocols. For quantification, western blots were analyzed by ImageJ. Antibodies used in this paper as follows: Flag tag (M2; Sigma), actin (Sigma), phospho CTD Tyr 1 (3D12; Active Motif), U2AF65 (Sigma), histone H3 protein (abcam), phospho CTD Ser 2 (3E10; Millipore), phospho CTD Ser 5 (3E8; Millipore), phospho CTD Ser 7 (4E12, Millipore), Rpb1 CTD (8WG16; abcam), GST tag (Invitrogen), Rpb1 (N20; Santa Cruz), TBP (N-12; Santa Cruz), TFIIB (biorbyt), Exosc10 (Rrp6) (Novus), and Exosc9 (Rrp45) (Novus).

Subcellular fractionation

Subcellular fractionation was performed using a modified protocol as described (Mapendano et al. 2010). Briefly, Cells ($1\sim2x10^7$) were harvested, washed in PBS, and resuspended in 0.5 ml of RSB100 (50 mM Tris-HCl PH 7.4, 100 mM NaCl) containing 40 ug/ml digitonin. Cell lysates were incubated on ice for 5 min. The cytoplasmic fraction was separated from nuclear fraction by centrifugation (2000g, 5 min). The pellets were resuspended in 0.5 ml of RSB100 containing 0.5% Triton X-100, and the reactions were incubated on ice for 5 min. Separation of soluble nuclear proteins from insoluble chromatin bound proteins was carried by centrifugation (2000g, 5 min). The pellets containing chromatin bound proteins were resuspended in 0.5 ml of RSB100 (0.5% Triton-X100), and sonicated briefly.

***In vitro* proteasome assay**

In vitro proteasome assay was performed as described (Asher et al. 2005) with the following modifications. Briefly, GST-CTD or GST-CTD phosphorylated by abl tyrosine kinase was incubated with 2~10 nM bovine 20S proteasome (UBPBio) in a buffer (50 mM Tris-HCl PH 7.4, 100 mM NaCl, 0.5 mM EDTA) at 37 ^{o}C for 1 hour. The reactions were stopped by add equal amount of 2X SDS PAGE sampling buffer.

***In vitro* phosphorylation of GST-CTD**

CDK7 complexes were expressed in insect cells, and purified using Ni-NTA agarose (QIAGEN) as described (Larochelle et al. 2006). GST-CTD was expressed in *E. coli* and purified using glutathione Sepharose 4B (GE Healthcare). Phosphorylation of GST-CTD by CDK7 complexes was carried out at 30^{o}C for 1 hour in a kinase buffer (25 mM Hepes PH 7.5, 10 mM $MgCl_2$, 150 mM NaCl, 1 mM ATP). GST-CTD phosphorylation by recombinant active abl kinase (NEB) was performed as described in manual. Briefly, 500 nM GST-CTD was incubated with 25 U abl kinase at 30^{o}C for 2 h. The phosphorylated GST-CTD was purified using glutathione Sepharose 4B (GE Healthcare).

RT-qPCR

RNA was extracted using Trizol (Invitrogen), and further treated with DNase I. Reverse transcription and qPCR analysis were performed as previously described (Hsin et al. 2011). Primer sequences will be provided upon request.

Immunoprecipitation

About 2x10^{7} cells were collected, and washed with PBS. Then, 1 ml cold RIPA (150 mM NaCl, 1 mM EDTA, 50 mM Tris-HCl pH 7.4, 0.5% NP-40, 0.25% sodium deoxycholate) buffer containing 1X PhosSTOP (Roche), and 1X protease inhibitors (1.4 ug/mL Pepstatin A, 0.35 ug/mL Leupeptin, and 1.7 ug/mL Aprotinin). After brief sonication, debris was centrifuged at 12,000g, 4^{o}C, for 10 min, and the supernatant was removed to a new tube. 50 ul of the lysate were kept for input control, and the rest of the extract was incubated with 20 ul of pre-washed protein G Sepharose and 1-4 ug of antibody. Samples were rotated at 4^{o}C for 1~2 h, and beads were washed with cold RIPA buffer for 3 min three times, and then were resuspended in 100 ul of 1X SDS sample buffer for western blotting.

Chromatin immunoprecipitation (ChIP)

Cells were grown to 70% confluence (~$2x10^6$/ml), cross-linked with 1% formaldehyde for 10 minutes, and processed for ChIP as previously described (Hsin et al. 2011). ChIP was performed using antibody against Flag tag (M2; Sigma), phospho CTD Tyr 1 (3D12; Active Motif), TBP (N-12; Santa Cruz), and TFIIB (biorbyt). Primers sequences will be provided upon request.

RNA-Seq

Total RNA was extracted from cells treated with tetracycline for 24 hours. RNA was further processed, and analyzed by deep sequencing as described (Hoque et al. 2013). Briefly, to avoid the issue of internal priming at A-rich sequences, RNAs with a poly (A) tail of 45 nucleotides was purified and analyzed by deep sequencing. We generated four million reads for each cell lines, and repeated the experiment on a second independent RNA sample. Therefore, for each cell line the total number of reads generated is about 8 millions. The reads then were mapped to the chicken genome, and those with at least two nongenomic As at 3' end were considered as poly(A) site-supporting (PASS) reads.

Acknowledgements

We thank our lab members for insightful discussions. We are thankful for CDK7 baculoviruses, a gift from Dr. Stephane Larochelle and Dr. Robert Fisher. This work was funded by grants from the NIH.

References

Akhtar MS, Heidemann M, Tietjen JR, Zhang DW, Chapman RD, Eick D, Ansari AZ. 2009. TFIIH kinase places bivalent marks on the carboxy-terminal domain of RNA polymerase II. *Mol Cell* **34**: 387-393.

Almada AE, Wu X, Kriz AJ, Burge CB, Sharp PA. 2013. Promoter directionality is controlled by U1 snRNP and polyadenylation signals. *Nature* **499**: 360-363.

Asher G, Tsvetkov P, Kahana C, Shaul Y. 2005. A mechanism of ubiquitin-independent proteasomal degradation of the tumor suppressors p53 and p73. *Genes Dev* **19**: 316-321.

Baskaran R, Dahmus ME, Wang JYJ. 1993. Tyrosine phosphorylation of mammalian RNA polymerase II carboxyl-terminal domain. *Proc Natl Acad Sci U S A* **90**: 11167-11171.

Baskaran R, Escobar SR, Wang JYJ. 1999. 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 & Differentiation* **10**: 387-396.

Bataille AR, Jeronimo C, Jacques PE, Laramee L, Fortin ME, Forest A, Bergeron M, Hanes SD, Robert F. 2012. A universal RNA polymerase II CTD cycle is orchestrated by complex interplays between kinase, phosphatase, and isomerase enzymes along genes. *Mol Cell* **45**: 158-170.

Beaudenon SL, Huacani MR, Wang G, McDonnell DP, Huibregtse JM. 1999. Rsp5 ubiquitin-protein ligase mediates DNA damage-induced degradation of the large subunit of RNA polymerase II in Saccharomyces cerevisiae. *Mol Cell Biol* **19**: 6972-6979.

Chapman RD, Conrad M, Eick D. 2005. Role of the mammalian RNA polymerase II C-terminal domain (CTD) nonconsensus repeats in CTD stability and cell proliferation. *Mol Cell Biol* **25**: 7665-7674.

Chapman RD, Palancade B, Lang A, Bensaude O, Eick D. 2004. The last CTD repeat of the mammalian RNA polymerase II large subunit is important for its stability. *Nucleic Acids Res* **32**: 35-44.

Cho EJ, Kobor MS, Kim M, Greenblatt J, Buratowski S. 2001. Opposing effects of Ctk1 kinase and Fcp1 phosphatase at Ser 2 of the RNA polymerase II C-terminal domain. *Genes Dev* **15**: 3319-3329.

Colicelli J. 2010. ABL tyrosine kinases: evolution of function, regulation, and specificity. *Science signaling* **3**: re6.

Core LJ, Waterfall JJ, Lis JT. 2008. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science* **322**: 1845-1848.

de la Mata M, Kornblihtt AR. 2006. RNA polymerase II C-terminal domain mediates regulation of alternative splicing by SRp20. *Nat Struct Mol Biol* **13**: 973-980.

Duyster J, Baskaran R, Wang JY. 1995. Src homology 2 domain as a specificity determinant in the c-Abl-mediated tyrosine phosphorylation of the RNA polymerase II carboxyl-terminal repeated domain. *Proc Natl Acad Sci U S A* **92**: 1555-1559.

Egloff S, Dienstbier M, Murphy S. 2012. Updating the RNA polymerase CTD code: adding gene-specific layers. *Trends Genet* **28**: 333-341.

Egloff S, O'Reilly D, Chapman RD, Taylor A, Tanzhaus K, Pitts L, Eick D, Murphy S. 2007. Serine-7 of the RNA polymerase II CTD is specifically required for snRNA gene expression. *Science* **318**: 1777-1779.

Egloff S, Szczepaniak SA, Dienstbier M, Taylor A, Knight S, Murphy S. 2010. The integrator complex recognizes a new double mark on the RNA polymerase II carboxyl-terminal domain. *J Biol Chem* **285**: 20564-20569.

Fabrega C, Shen V, Shuman S, Lima CD. 2003. Structure of an mRNA Capping Enzyme Bound to the Phosphorylated Carboxy-Terminal Domain of RNA Polymerase II. *Mol Cell* **11**: 1549-1561.

Flynn RA, Almada AE, Zamudio JR, Sharp PA. 2011. Antisense RNA polymerase II divergent transcripts are P-TEFb dependent and substrates for the RNA exosome. *Proc Natl Acad Sci U S A* **108**: 10460-10465.

Forster A, Hill CP. 2003. Proteasome degradation: enter the substrate. *Trends Cell Biol* **13**: 550-553.

Geng F, Tansey WP. 2012. Similar temporal and spatial recruitment of native 19S and 20S proteasome subunits to transcriptionally active chromatin. *Proc Natl Acad Sci U S A* **109**: 6060-6065.

Gillette TG, Gonzalez F, Delahodde A, Johnston SA, Kodadek T. 2004. Physical and functional association of RNA polymerase II and the proteasome. *Proc Natl Acad Sci U S A* **101**: 5904-5909.

Glover-Cutter K, Larochelle S, Erickson B, Zhang C, Shokat K, Fisher RP, Bentley DL. 2009. 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**: 5455-5464.

Groll M, Bajorek M, Kohler A, Moroder L, Rubin DM, Huber R, Glickman MH, Finley D. 2000. A gated channel into the proteasome core particle. *Nat Struct Biol* **7**: 1062-1067.

Groll M, Ditzel L, Lowe J, Stock D, Bochtler M, Bartunik HD, Huber R. 1997. Structure of 20S proteasome from yeast at 2.4 A resolution. *Nature* **386**: 463-471.

He Y, Fang J, Taatjes DJ, Nogales E. 2013. Structural visualization of key steps in human transcription initiation. *Nature* **495**: 481-486.

He Y, Vogelstein B, Velculescu VE, Papadopoulos N, Kinzler KW. 2008. The antisense transcriptomes of human cells. *Science* **322**: 1855-1857.

Heidemann M, Hintermair C, Voss K, Eick D. 2013. Dynamic phosphorylation patterns of RNA polymerase II CTD during transcription. *Biochim Biophys Acta* **1829**: 55-62.

Hengartner CJ, Myer VE, Liao SM, Wilson CJ, Koh SS, Young RA. 1998. Temporal regulation of RNA polymerase II by Srb10 and Kin28 cyclin-dependent kinases. *Mol Cell* **2**: 43-53.

Hintermair C, Heidemann M, Koch F, Descostes N, Gut M, Gut I, Fenouil R, Ferrier P, Flatley A, Kremmer E et al. 2012. Threonine-4 of mammalian RNA polymerase II CTD is targeted by Polo-like kinase 3 and required for transcriptional elongation. *The Embo Journal*.

Hoque M, Ji Z, Zheng D, Luo W, Li W, You B, Park JY, Yehia G, Tian B. 2013. Analysis of alternative cleavage and polyadenylation by 3' region extraction and deep sequencing. *Nature methods* **10**: 133-139.

Hsin JP, Sheth A, Xiang K, Manley JL. 2013. Function and control of RNA polymerase II CTD phosphorylaiton in veterbrate transcription and RNA processing. (Submitted)

Hsin JP, Manley JL. 2012. The RNA polymerase II CTD coordinates transcription and RNA processing. *Genes Dev* **26**: 2119-2137.

Hsin JP, Sheth A, Manley JL. 2011. RNAP II CTD phosphorylated on threonine-4 is required for histone mRNA 3' end processing. *Science* **334**: 683-686.

Kim M, Krogen NJ, Vasiljeva L, Rando OJ, Nedea E, Greenblatt J, Buratowski S. 2004. The yeast Rat1 exonuclease promotes transcription termination by RNA polymerase II. *Nature* **432**: 517-522.

Kim WY, Dahmus ME. 1986. Immunochemical analysis of mammalian RNA polymerase II subspecies. Stability and relative in vivo concentration. *J Biol Chem* **261**: 14219-14225.

Komarnitsky P, Cho EJ, Buratowski S. 2000. Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. *Genes Dev* **14**: 2452-2460.

Krishnamurthy S, He X, Reyes-Reyes M, Moore C, Hampsey M. 2004. Ssu72 Is an RNA polymerase II CTD phosphatase. *Mol Cell* **14**: 387-394.

Larochelle S, Batliner J, Gamble MJ, Barboza NM, Kraybill BC, Blethrow JD, Shokat KM, Fisher RP. 2006. Dichotomous but stringent substrate selection by the dual-function Cdk7 complex revealed by chemical genetics. *Nat Struct Mol Biol* **13**: 55-62.

Li H, Zhang Z, Wang B, Zhang J, Zhao Y, Jin Y. 2007. Wwp2-mediated ubiquitination of the RNA polymerase II large subunit in mouse embryonic pluripotent stem cells. *Mol Cell Biol* **27**: 5296-5305.

Liu CW, Corboy MJ, DeMartino GN, Thomas PJ. 2003. Endoproteolytic activity of the proteasome. *Science* **299**: 408-411.

Lunde BM, Reichow SL, Kim M, Suh H, Leeper TC, Yang F, Mutschler H, Buratowski S, Meinhart A, Varani G. 2010. Cooperative interaction of transcription termination factors with the RNA polymerase II C-terminal domain. *Nat Struct Mol Biol* **17**: 1195-1201.

Mapendano CK, Lykke-Andersen S, Kjems J, Bertrand E, Jensen TH. 2010. Crosstalk between mRNA 3' end processing and transcription initiation. *Mol Cell* **40**: 410-422.

Marshall NF, Peng J, Xie Z, Price DH. 1996. Control of RNA polymerase II elongation potential by a novel carboxyl-terminal domain kinase. *J Biol Chem* **271**: 27176-27183.

Mayer A, Heidemann M, Lidschreiber M, Schreieck A, Sun M, Hintermair C, Kremmer E, Eick D, Cramer P. 2012. CTD tyrosine phosphorylation impairs termination factor recruitment to RNA polymerase II. *Science* **336**: 1723-1725.

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

Meinhart A, Kamenski T, Hoeppner S, Baumli S, Cramer P. 2005. A structural perspective of CTD function. *Genes Dev* **19**: 1401-1415.

Munoz MJ, de la Mata M, Kornblihtt AR. 2010. The carboxy terminal domain of RNA polymerase II and alternative splicing. *Trends in Biochemical Sciences* **35**: 497-504.

Neil H, Malabat C, d'Aubenton-Carafa Y, Xu Z, Steinmetz LM, Jacquier A. 2009. Widespread bidirectional promoters are the major source of cryptic transcripts in yeast. *Nature* **457**: 1038-1042.

Ntini E, Jarvelin AI, Bornholdt J, Chen Y, Boyd M, Jorgensen M, Andersson R, Hoof I, Schein A, Andersen PR et al. 2013. Polyadenylation site-induced decay of upstream transcripts enforces promoter directionality. *Nat Struct Mol Biol*.

Preker P, Almvig K, Christensen MS, Valen E, Mapendano CK, Sandelin A, Jensen TH. 2011. PROMoter uPstream Transcripts share characteristics with mRNAs and are produced upstream of all three major types of mammalian promoters. *Nucleic Acids Res* **39**: 7179-7193.

Preker P, Nielsen J, Kammler S, Lykke-Andersen S, Christensen MS, Mapendano CK, Schierup MH, Jensen TH. 2008. RNA exosome depletion reveals transcription upstream of active human promoters. *Science* **322**: 1851-1854.

Ranish JA, Yudkovsky N, Hahn S. 1999. Intermediates in formation and activity of the RNA polymerase II preinitiation complex: holoenzyme recruitment and a postrecruitment role for the TATA box and TFIIB. *Genes Dev* **13**: 49-63.

Rix U, Hantschel O, Durnberger G, Remsing Rix LL, Planyavsky M, Fernbach NV, Kaupe I, Bennett KL, Valent P, Colinge J et al. 2007. Chemical proteomic profiles of the BCR-ABL inhibitors imatinib, nilotinib, and dasatinib reveal novel kinase and nonkinase targets. *Blood* **110**: 4055-4063.

Schwer B, Shuman S. 2011. Deciphering the RNA polymerase II CTD code in fission yeast. *Mol Cell* **43**: 311-318.

Seila AC, Calabrese JM, Levine SS, Yeo GW, Rahl PB, Flynn RA, Young RA, Sharp PA. 2008. Divergent transcription from active promoters. *Science* **322**: 1849-1851.

Seila AC, Core LJ, Lis JT, Sharp PA. 2009. Divergent transcription: a new feature of active promoters. *Cell Cycle* **8**: 2557-2564.

Sheaff RJ, Singer JD, Swanger J, Smitherman M, Roberts JM, Clurman BE. 2000. Proteasomal turnover of p21Cip1 does not require p21Cip1 ubiquitination. *Mol Cell* **5**: 403-410.

Somesh BP, Reid J, Liu WF, Sogaard TM, Erdjument-Bromage H, Tempst P, Svejstrup JQ. 2005. Multiple mechanisms confining RNA polymerase II ubiquitylation to polymerases undergoing transcriptional arrest. *Cell* **121**: 913-923.

Somesh BP, Sigurdsson S, Saeki H, Erdjument-Bromage H, Tempst P, Svejstrup JQ. 2007. Communication between distant sites in RNA polymerase II through ubiquitylation factors and the polymerase CTD. *Cell* **129**: 57-68.

Spain MM, Govind CK. 2011. A role for phosphorylated Pol II CTD in modulating transcription coupled histone dynamics. *Transcription* **2**: 78-81.

Tofaris GK, Layfield R, Spillantini MG. 2001. alpha-synuclein metabolism and aggregation is linked to ubiquitin-independent degradation by the proteasome. *FEBS letters* **509**: 22-26.

Touitou R, Richardson J, Bose S, Nakanishi M, Rivett J, Allday MJ. 2001. A degradation signal located in the C-terminus of p21WAF1/CIP1 is a binding site for the C8 alpha-subunit of the 20S proteasome. *Embo J* **20**: 2367-2375.

Wei W, Pelechano V, Jarvelin AI, Steinmetz LM. 2011. Functional consequences of bidirectional promoters. *Trends Genet* **27**: 267-276.

West ML, Corden JL. 1995. Construction and analysis of yeast RNA polymerase II CTD deletion and substitution mutations. *Genetics* **140**: 1223-1233.

Wilson MD, Harreman M, Svejstrup JQ. 2013. Ubiquitylation and degradation of elongating RNA polymerase II: the last resort. *Biochim Biophys Acta* **1829**: 151-157.

Xu Z, Wei W, Gagneur J, Perocchi F, Clauder-Munster S, Camblong J, Guffanti E, Stutz F, Huber W, Steinmetz LM. 2009. Bidirectional promoters generate pervasive transcription in yeast. *Nature* **457**: 1033-1037.

Zhang DW, Mosley AL, Ramisetty SR, Rodriguez-Molina JB, Washburn MP, Ansari AZ. 2012. 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**: 8541-8551.

Figure legends

Figure 1. Growth properties analysis in Rpb1 cell lines. (A) Cells were cultured in medium containing 1 ug/ml tetracycline (tet). The control cells, 26r, were split on day 2. Average cell counts from two independent experiments were plotted. (B) Western blot analysis was performed in lysates from cells treated with tet for 24 h. Flag-tagged Rpb1 proteins were detected using Flag antibody. Samples from three independent Y1F cell lines were shown. Asterisk indicates the degraded Rpb1 protein. (C) Cells were treated with tet for 24 h. Whole cell lysates from 26r, Y1F cells, and cells (6r) expressing an Rpb1

with 6 YSPTSPS repeats were analyzed by western blotting. The degraded Rpb1 is indicated by asterisk (*). (D) Growth curves of two independent 25F+Y cell lines and 26r cells were plotted as in (A). (E) Cell lysates from 4 independent 25F+Y and 3 independent Y1F cell lines were analyzed as in (B). Asterisk indicates the degraded Rpb1 protein.

Figure 2. Western blotting analysis of Y1F and 25F+Y cells. (A) Flag-tagged Rpb1 proteins were immunoprecipitated from cells treated with tet for 24 h, and analyzed using western blotting. Phosphorylation on Tyr 1 (P-Tyr 1) was detected by the 3D12 antibody. (B) 26r and Y1F cells were treated with tet for 24 h, subcellular fractionation was performed, and cytoplasmic, nuclear, and chromatin-bound fractions were analyzed by western blotting. U2AF65 (a nuclear protein), and chromatin bound histone H3 protein were served as controls for subcellular fractionation assay. Asterisk (*) indicates the degraded Rpb1 protein. (C) Wild-type DT40 cells were subjected to subcellular fractionation. The localization of Rpb1 phosphorylated on Tyr 1, Ser 2, 5, and 7, was determined using antibodies as described in Methods.

Figure 3. *In vitro* 20S proteasome assay. (A) 200 nM GST-CTD was incubated with 5 nM bovine 20S proteasome, and the reaction was carried with or without 0.01% SDS. MG132, selective inhibitor of 20S proteasome, was used to inhibit 20S proteasome, and reaction with 2.5% ethanol (ETOH) was served as control. Western blotting was performed using antibody against GST, and the CTD (B). Molecular weight markers at 25 and 75 kD are indicated. (C) GST-CTD protein was phosphorylated by Abl tyrosine kinase for indicated time *in vitro*. Reactions were analyzed by western blotting with indicated antibodies. (D) About 40 nM GST-CTD phosphorylated by Abl tyrosine kinase or by CDK7 was incubated with 2 nM bovine 20S proteasome with or without 0.01% SDS for 1 h, and reactions were analyzed by western blotting.

Figure 4. Rpb1 ChIP analysis. (A) Cells were treated with tet for 24 h. ChIP was performed as described. The recruitment of Flag-tagged Rpb1 on the *Rplp1* gene was determined using qPCR. Diagrams depict genes analyzed. Thick line, dashed line, and triangle denote gene, transcript, and 3' cleavage site, respectively. On *Rplp1*, amplicon B is at transcription start site (TSS), and amplicon D covers 3' cleavage site. Average % input from three independent experiments (*N=3*) was plotted. (B) ChIP analysis was performed as in (A). Rpb1 recruitment on the β-actin (*Actb*) gene was determined. Amplicon A is at TSS, and amplicon B covers 3' cleavage site. (C) ChIP analysis of phospho CTD Tyr 1 in the *Rplp1* gene, and in the *Actb* gene (D). *N=3*. Error bars display standard deviation.

Figure 5. ChIP analysis of the recruitment of TBP and TFIIB basal transcription factors. ChIP was performed, and analyzed as in Figure 4. (A) The levels of TBP on *Rplp1* were plotted. (B) TBP ChIP analysis in the *Actb* gene. (C) ChIP analysis of TFIIB distribution in the *Rplp1* gene, and in the *Actb* gene (D). *N=3*. Error bars denote standard deviation.

Figure 6. Analyses of upstream antisense transcription. (A) RNA from cells treated with tet for 24 h was processed for RNA-Seq analysis as described. The PASS reads, defined as described in text, were categorized into three types according to the poly(A) site location, in most 3' exon (D), in the body of gene (G), and in the upstream region of the gene (U). Changes in the usage of poly (A) sites were measured, and compared to 26r cells, those with a significance (*p* value <0.05, tested by Fisher Exact Test) were counted. The number of genes with significant up- or down-regulated use of poly(A) sites relative to 26r cells was shown. (B) RT-qPCR was performed to measure the upstream transcripts in genes with significant up-regulated use of U-type poly(A) site, uncovered in RNA-Seq analysis. Fold relative to 26r cells was plotted. *N=3*. (C) ChIP was performed as described in Figure 4. Rpb1 levels on genes were determined using primers as indicated. The number in the parenthesis next to the examined

gene indicates the distance between the amplicon the TSS (minus sign denotes upstream). *N=2*. Error bars display standard deviation.

Supplemental figure legends

Figure S1. Subcellular localization of Rpb1 protein in cells. (A) Cells were treated with tet for 24 h, and then subjective to subcellular fractionation. Rpb1 localization was determined by western blotting. Nuclear protein U2AF65, and chromatin-bound histone H3 were served as controls for the fractionation. Asterisk (*) indicates the degraded Rpb1 fragment. (B) Subcellular fractionation assay was performed in HEK293 cells. The localization of Rpb1 phosphorylated on Tyr 1, Ser 2, 5, 7, and Thr 4 was determined using antibodies as described in Methods.

Figure S2. ChIP analysis of Rpb1 recruitment on histone H2A, U1, and U2 snRNA genes. ChIP was performed and analyzed as in Figure 4. Average % input from three independent experiments was plotted. Error bars display standard deviation.

Figure S3. The influences of Tyr 1 mutation in *Egr1* expression, splicing and 3' end processing. (A) Cells were cultured in the presence of tet for 24 h, and the synthesis of *Egr1* mRNA was induced with 0.5 uM ionomycin and 10 ng/ml phorbo 12-myristate 13- acetate (PMA) for 20 min. The levels of *Egr1* mRNA were measured using RT-qPCR, and plotted relatively to 26r cells. *N=3*. (B) Cells were treated as in (A). The ratios of unspliced over spliced *Egr1* mRNA were determined by RT-qPCR, and plotted relatively to that of 26r cells. Diagram depicts the two-exon *Egr1* gene. The arrows denote primers (bottom) for detecting spliced *Egr1* mRNA, and the top primer set detects unspliced products. *N=3*. (C) The ratios of uncleaved to total Egr1 mRNA were measured using RT-qPCR, and plotted relatively to that detected in 26r cells. The left primers were used for measuring total *Egr1* mRNA, and the right primer set to detect uncleaved RNA. *N=3*. Error bars present standard deviation.

Figure S4. DATA from Figure 4 were reanalyzed by normalizing the average signal of phospho CTD Tyr 1 to that of Rpb1. The ratios of Y1P to Rpb1 on *Rplp1* gene (A) and *Actb* gene (B) were plotted.

Figure S5. Western blotting analysis. (A) Rpb1 proteins from DT40 cell lysates were immunoprecipitated using antibodies recognizing phosphoserine 5 (3E8) or phosphoserine 2 (3E10). The association of tyrosine phosphorylation with phosphoserine 5 or phosphoserine 2 was determined by western blotting using the 3D12 antibody. (B) The levels of TBP and TFIIB proteins were determined using western blotting. Cells were treated with tet for 24 h. (C) Cells were treated with tet for 24 h. The levels of exosome subunits, Exosc9 (Rrp45) and Exoec10 (Rrp 6), were determined by western blotting.

Figure S6. RNA-Seq analysis in the Rpb1 cell lines. Cells were treated with tet for 24 h, RNA was extracted, and processed for RNA-Seq analysis as described. 4 million reads were generated for each cell line, and with two independent samples, a total 8 million reads were obtained. The resulting reads were aligned to the chicken genome, the poly(A) site-supporting (PASS) reads were defined, and the percentage of PASS reads over total reads was calculated and presented.

Figure 1.

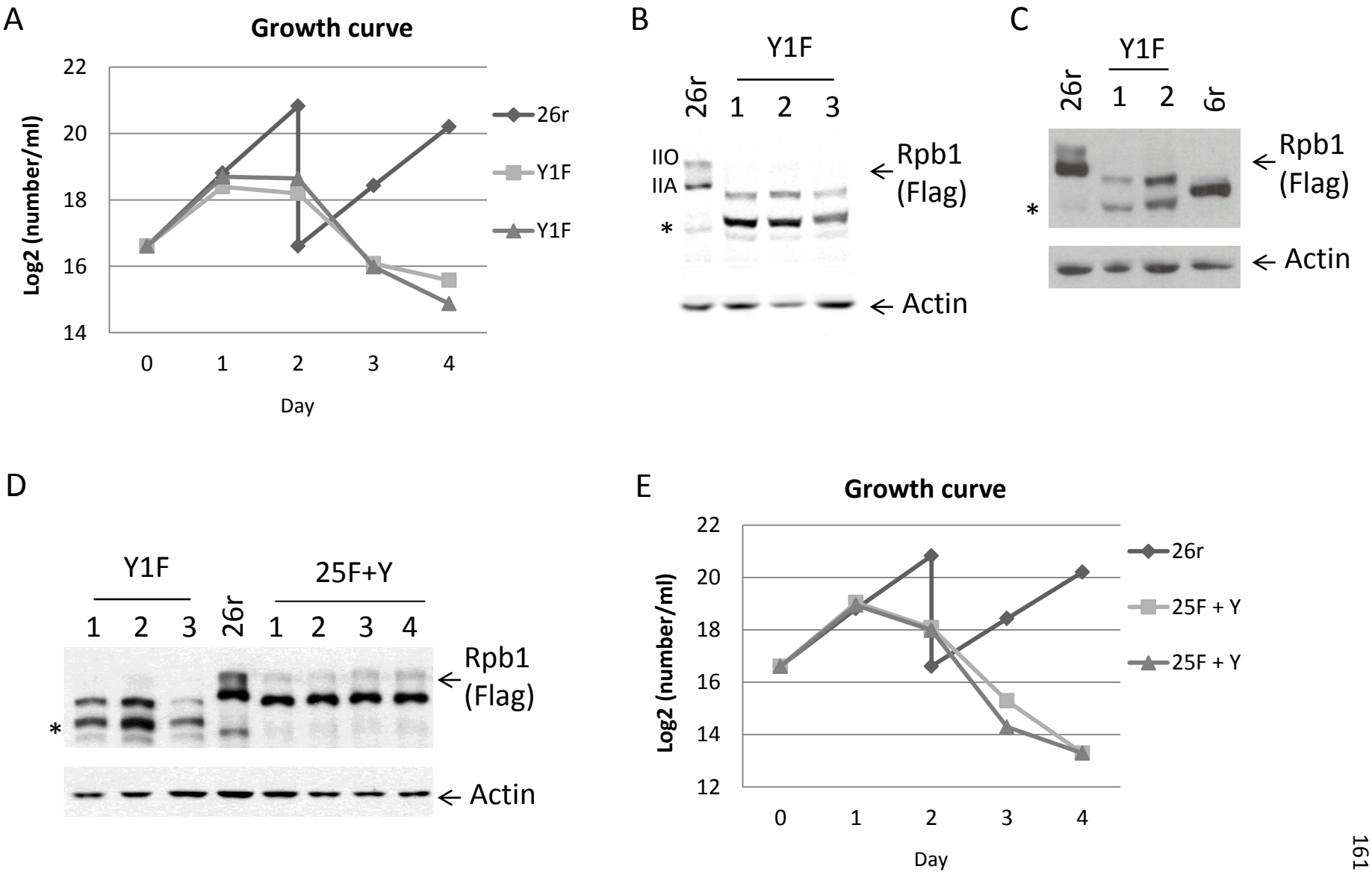

A
Growth curve
Log2 (number/ml)
22
20
18
16
14
0
1
2
3
4
Day
26r
Y1F
Y1F
B
Y1F
26r
1
2
3
IIO
IIA
*
Rpb1
(Flag)
Actin
C
Y1F
26r
1
2
6r
*
Rpb1
(Flag)
Actin
D
Y1F
25F+Y
1
2
3
26r
1
2
3
4
*
Rpb1
(Flag)
Actin
E
Growth curve
Log2 (number/ml)
22
20
18
16
14
12
0
1
2
3
4
Day
26r
25F + Y
25F + Y

Figure 2.

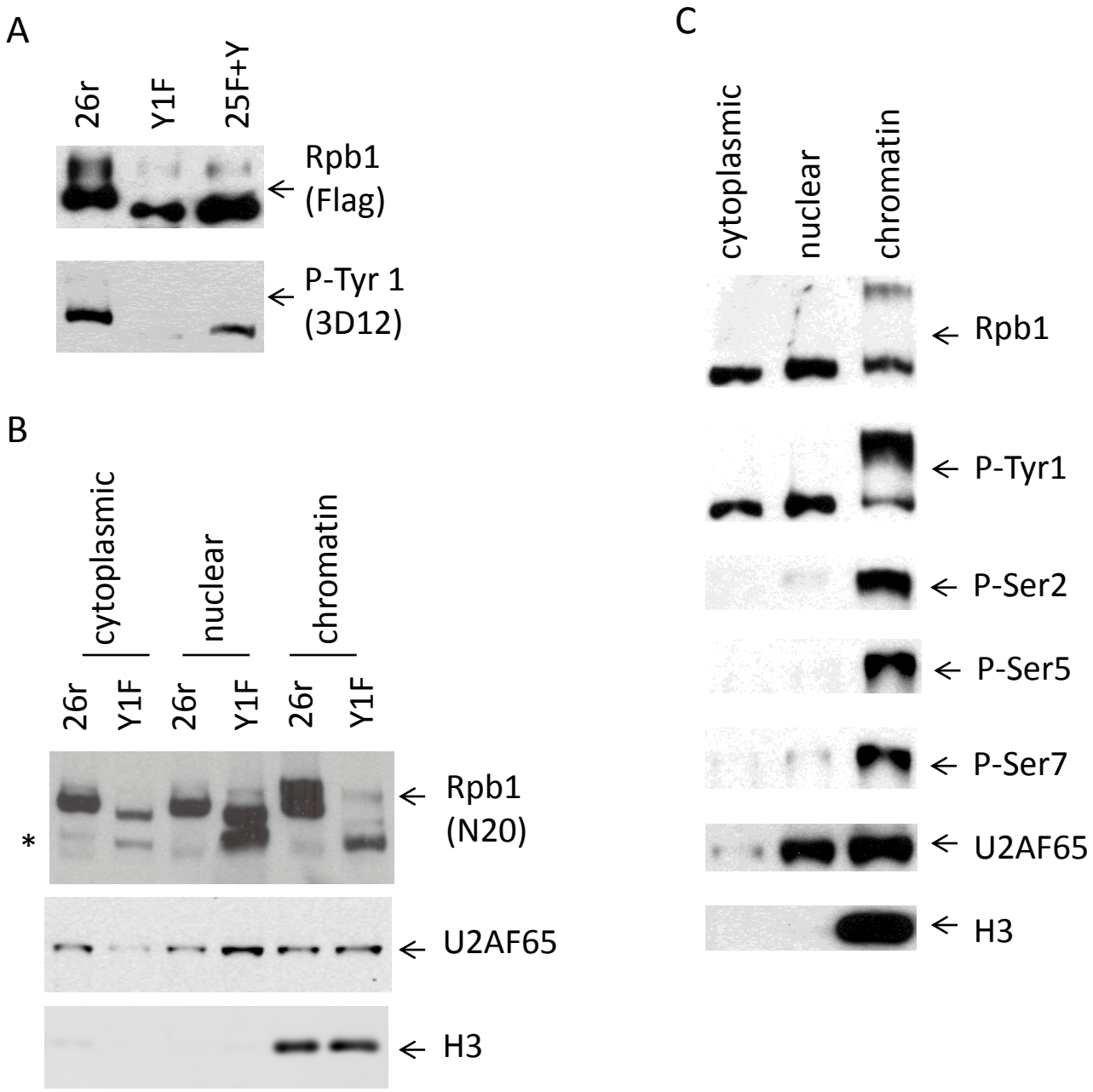

A
26r
Y1F
25F+Y
Rpb1
(Flag)
P-Tyr 1
(3D12)
B
cytoplasmic
nuclear
chromatin
26r
Y1F
26r
Y1F
26r
Y1F
*
Rpb1
(N20)
U2AF65
H3
C
cytoplasmic
nuclear
chromatin
Rpb1
P-Tyr1
P-Ser2
P-Ser5
P-Ser7
U2AF65
H3

Figure 3.

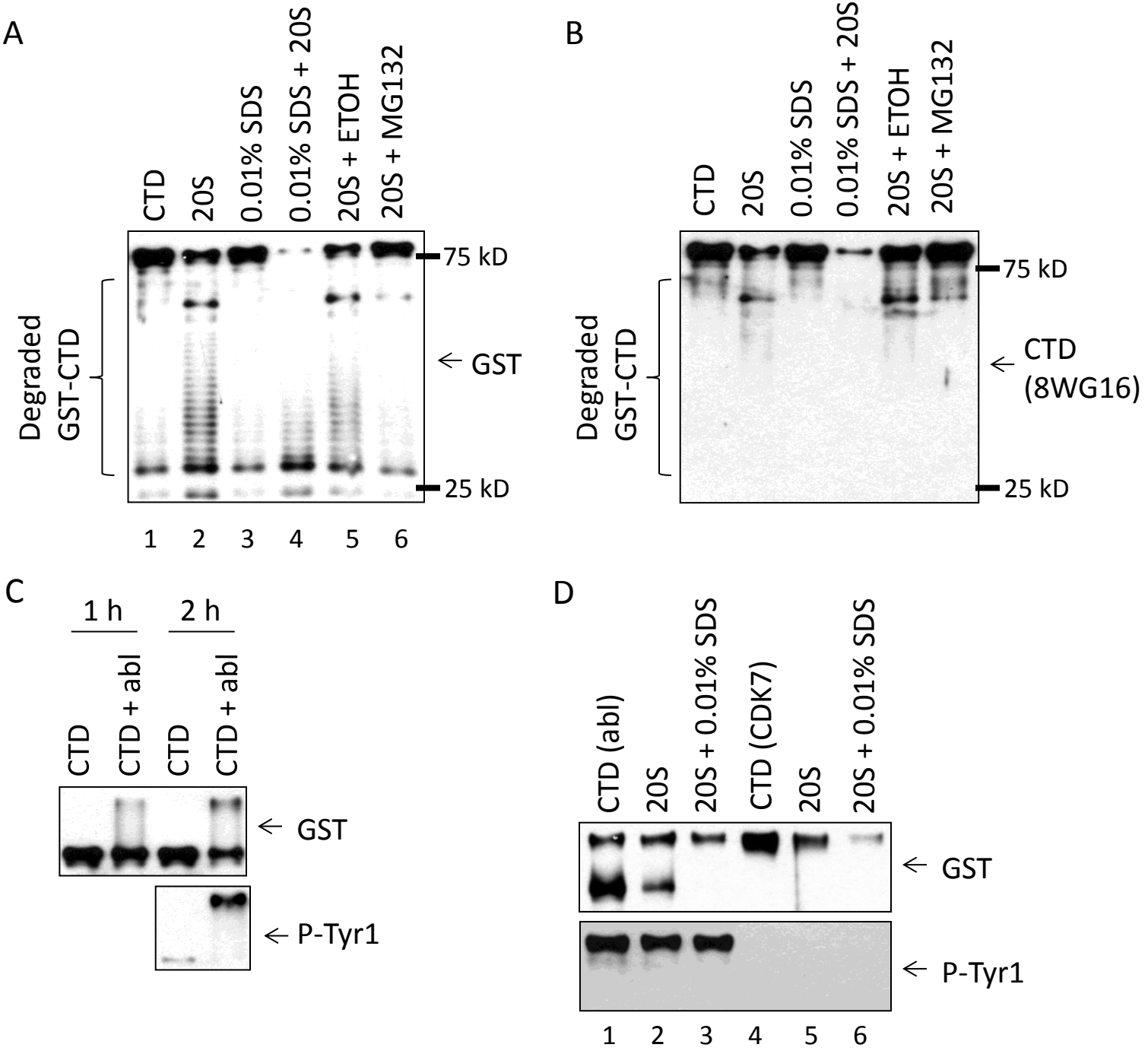

A
CTD
20S
0.01% SDS
0.01% SDS + 20S
20S + ETOH
20S + MG132
75 kD
GST
25 kD
Degraded GST-CTD
1 2 3 4 5 6
B
CTD
20S
0.01% SDS
0.01% SDS + 20S
20S + ETOH
20S + MG132
75 kD
CTD (8WG16)
25 kD
Degraded GST-CTD
C
1 h
2 h
CTD
CTD + abl
CTD
CTD + abl
GST
P-Tyr1
D
CTD (abl)
20S
20S + 0.01% SDS
CTD (CDK7)
20S
20S + 0.01% SDS
GST
P-Tyr1
1 2 3 4 5 6

Figure 4.

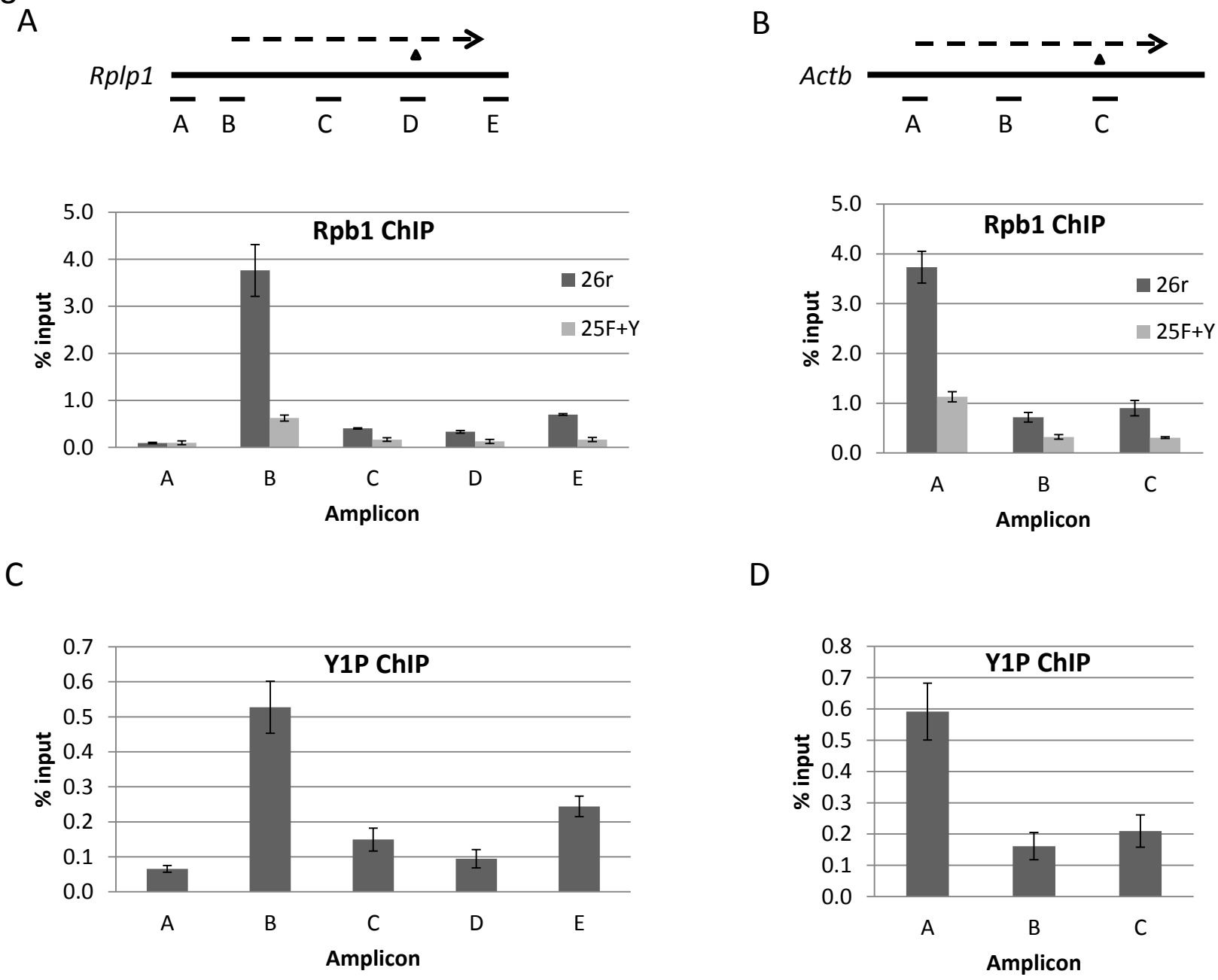

A
Rplp1
A B C D E
Rpb1 ChIP
% input
26r
25F+Y
Amplicon
B
Actb
A B C
Rpb1 ChIP
% input
26r
25F+Y
Amplicon
C
Y1P ChIP
% input
Amplicon
D
Y1P ChIP
% input
Amplicon

Figure 5.

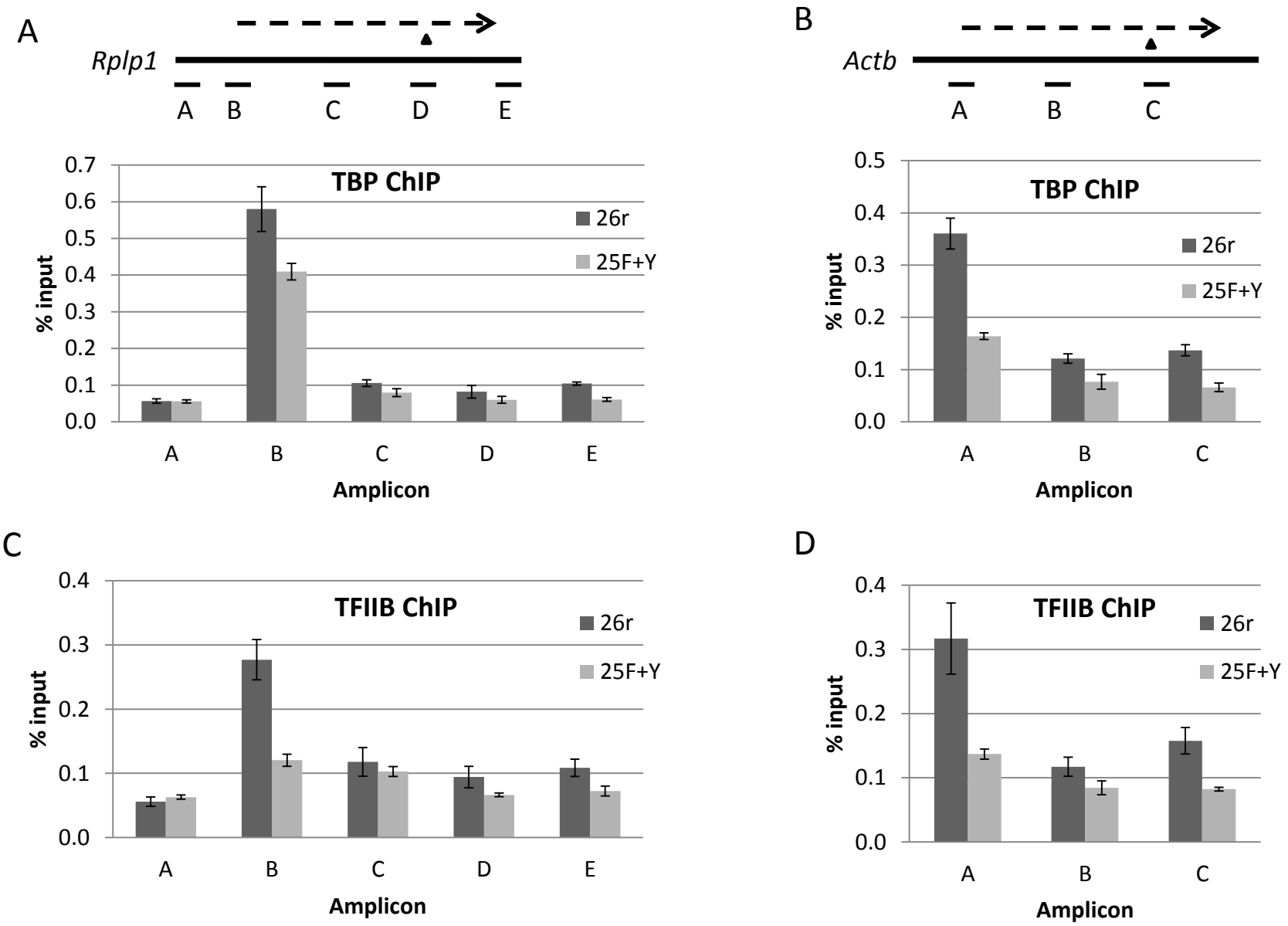
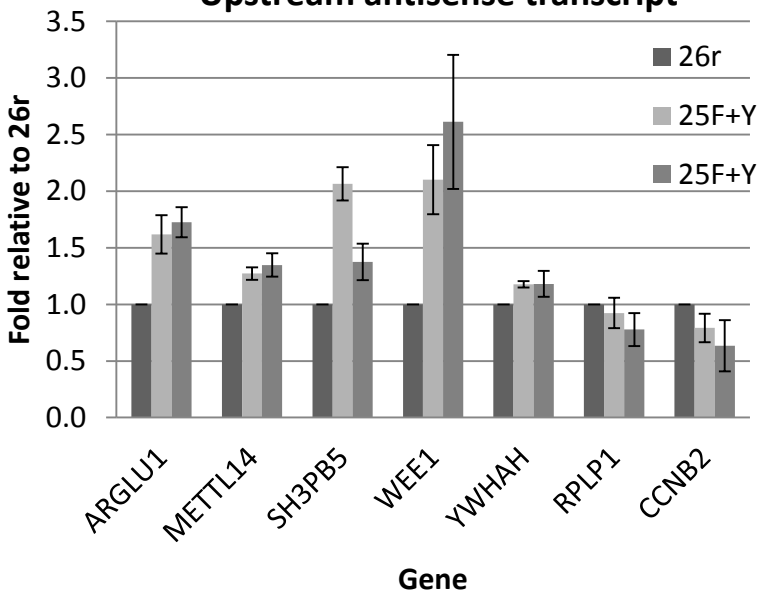

A
Rplp1
A
B
C
D
E
TBP ChIP
26r
25F+Y
% input
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
A
B
C
D
E
Amplicon
B
Actb
A
B
C
TBP ChIP
26r
25F+Y
% input
0.5
0.4
0.3
0.2
0.1
0.0
A
B
C
Amplicon
C
TFIIB ChIP
26r
25F+Y
% input
0.4
0.3
0.2
0.1
0.0
A
B
C
D
E
Amplicon
D
TFIIB ChIP
26r
25F+Y
% input
0.4
0.3
0.2
0.1
0.0
A
B
C
Amplicon

Figure 6.

A

	25F+Y #1		25F+Y #2		S2A #1		S2A #2		S5A #1		S5A #2	
	Up	Down	Up	Down	Up	Down	Up	Down	Up	Down	Up	Down
D	494	546	551	586	581	716	702	739	541	612	561	732
G	84	34	96	49	178	60	242	66	125	39	173	43
U	69	9	81	14	55	28	55	21	42	16	55	22

B



C

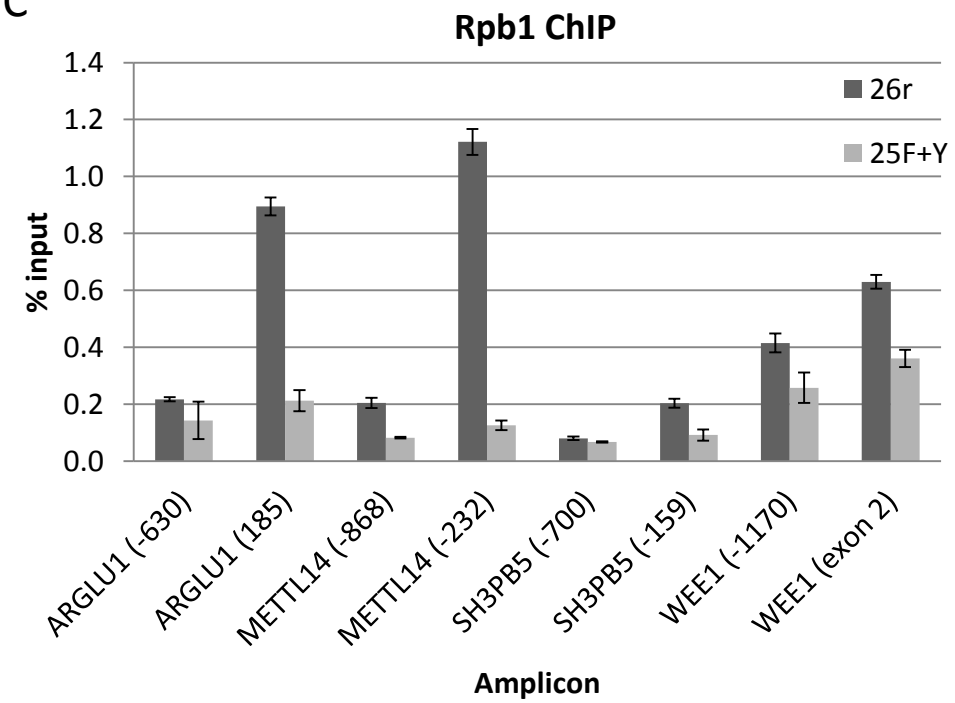


Figure S1.

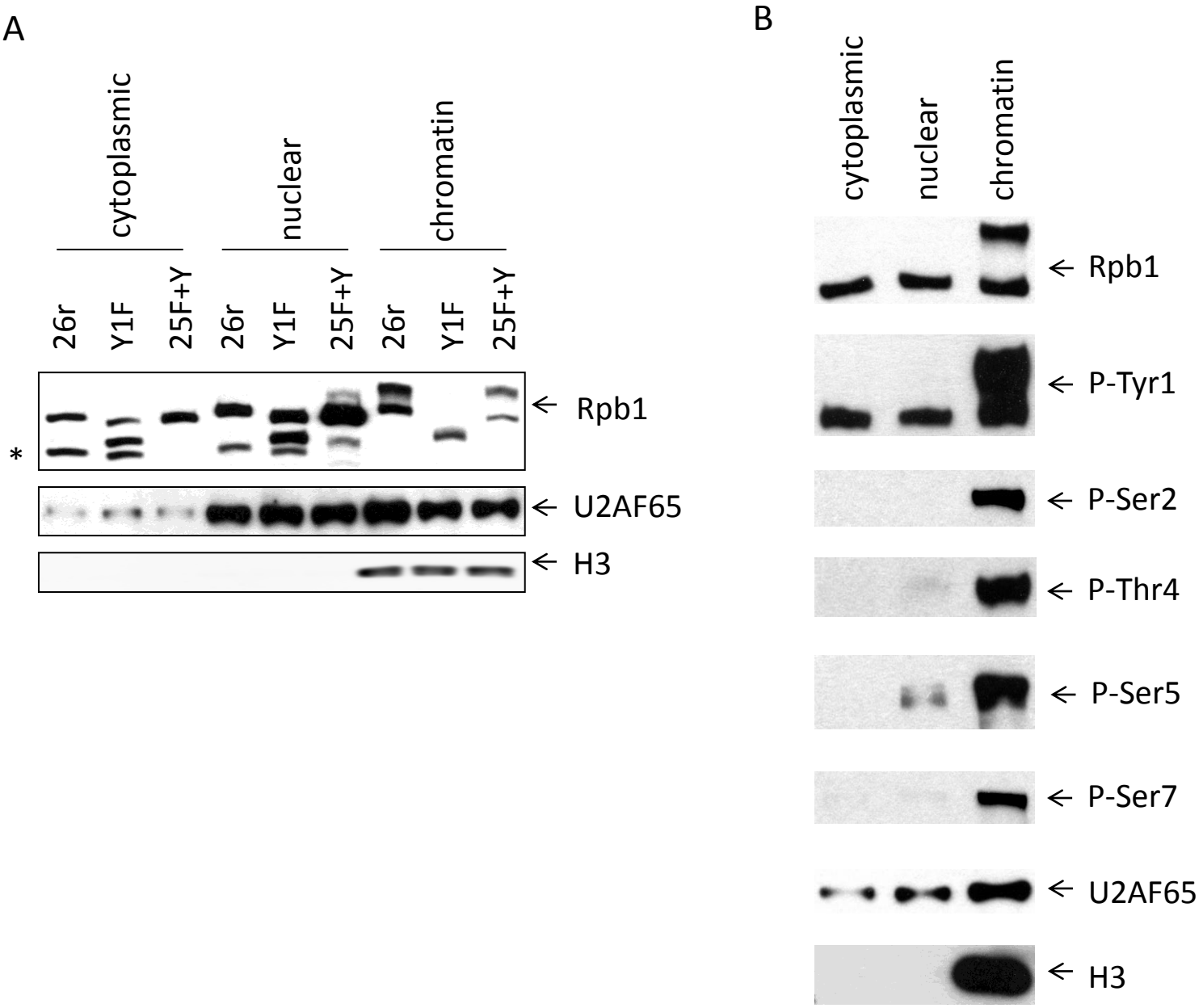

A
cytoplasmic
nuclear
chromatin
26r
Y1F
25F+Y
26r
Y1F
25F+Y
26r
Y1F
25F+Y
*
Rpb1
U2AF65
H3
B
cytoplasmic
nuclear
chromatin
Rpb1
P-Tyr1
P-Ser2
P-Thr4
P-Ser5
P-Ser7
U2AF65
H3

Figure S2.

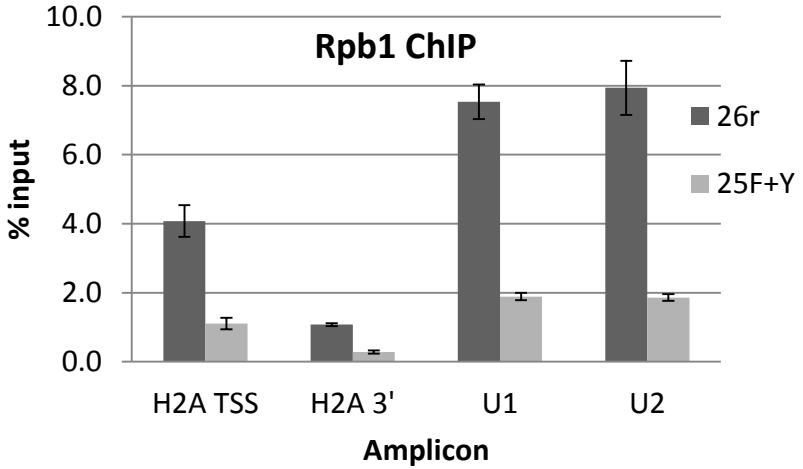
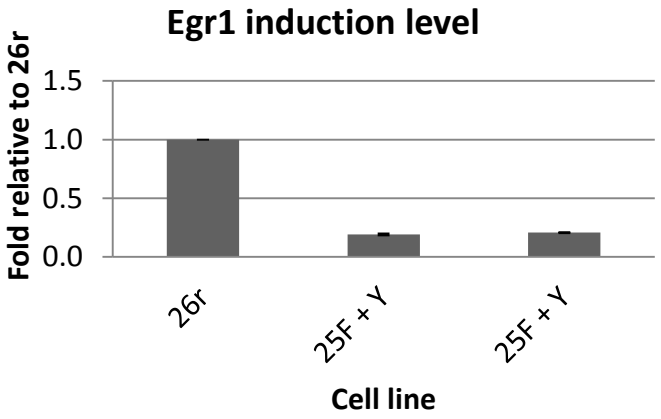
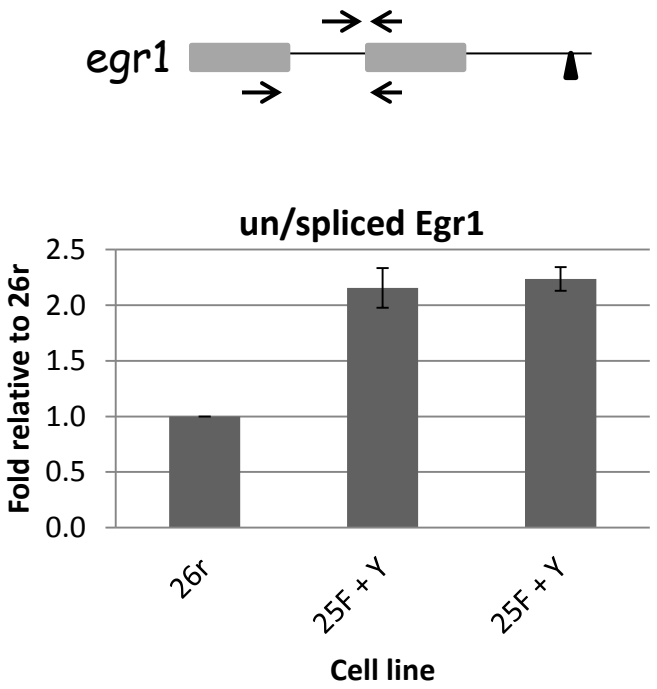

Rpb1 ChIP
10.0
8.0
6.0
4.0
2.0
0.0
% input
26r
25F+Y
H2A TSS
H2A 3'
U1
U2
Amplicon

Figure S3.

A



B



C

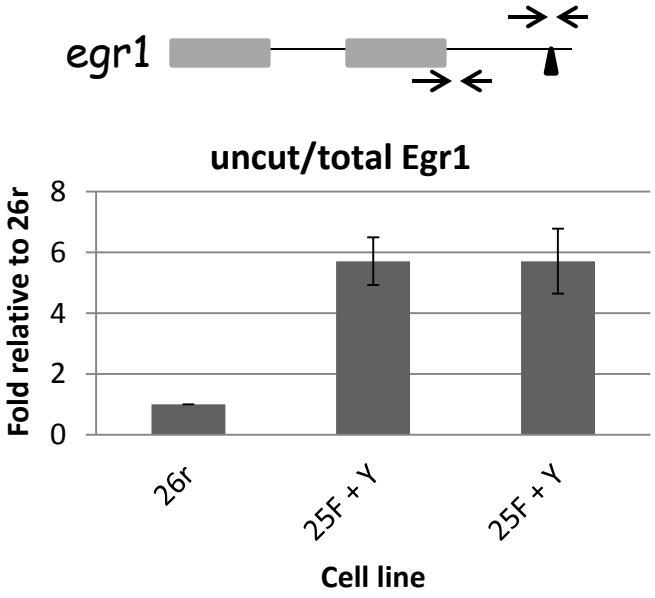


Figure S4.

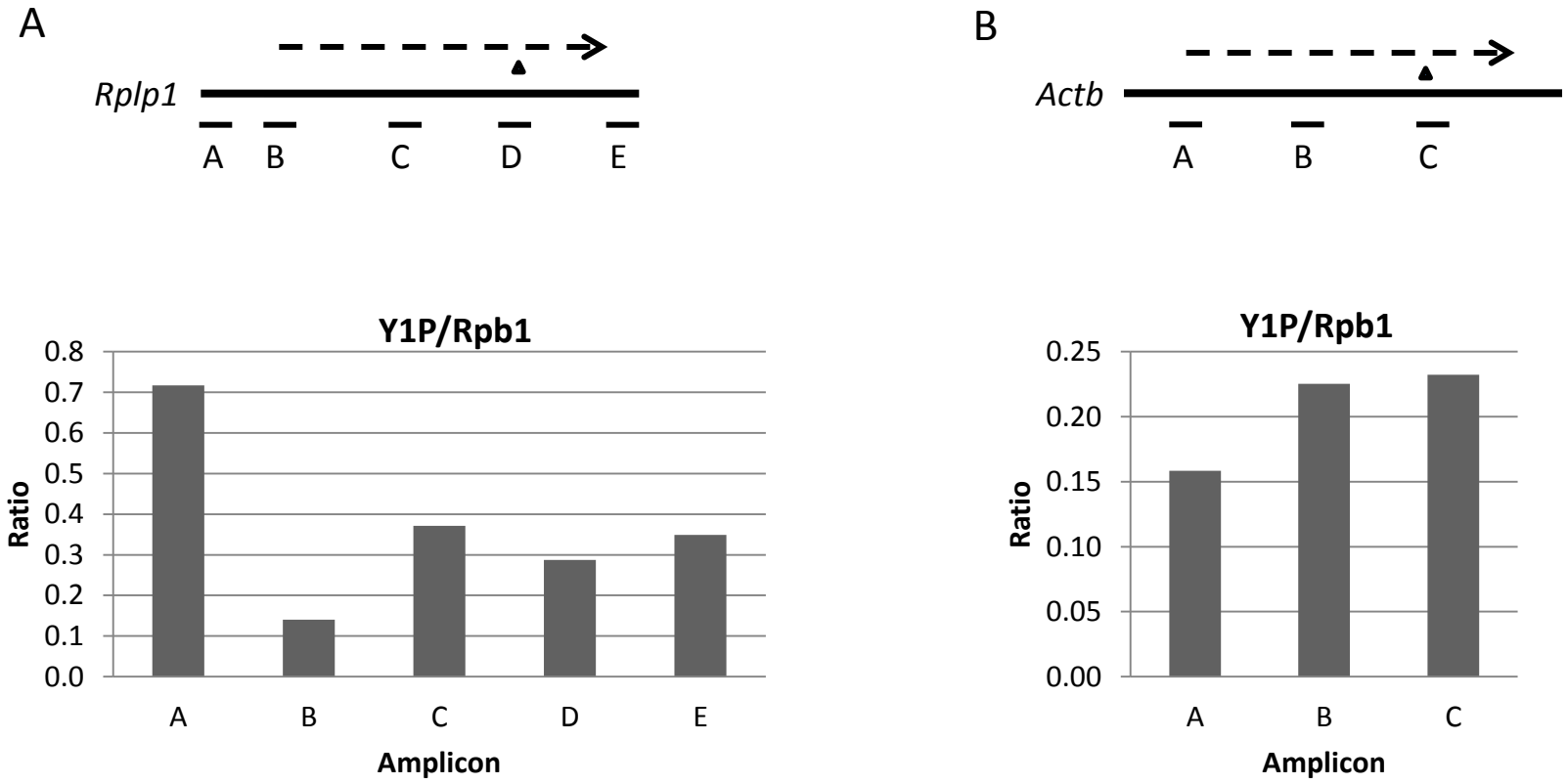

A
Rplp1
A
B
C
D
E
Y1P/Rpb1
Ratio
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
A
B
C
D
E
Amplicon
B
Actb
A
B
C
Y1P/Rpb1
Ratio
0.25
0.20
0.15
0.10
0.05
0.00
A
B
C
Amplicon

Figure S5.

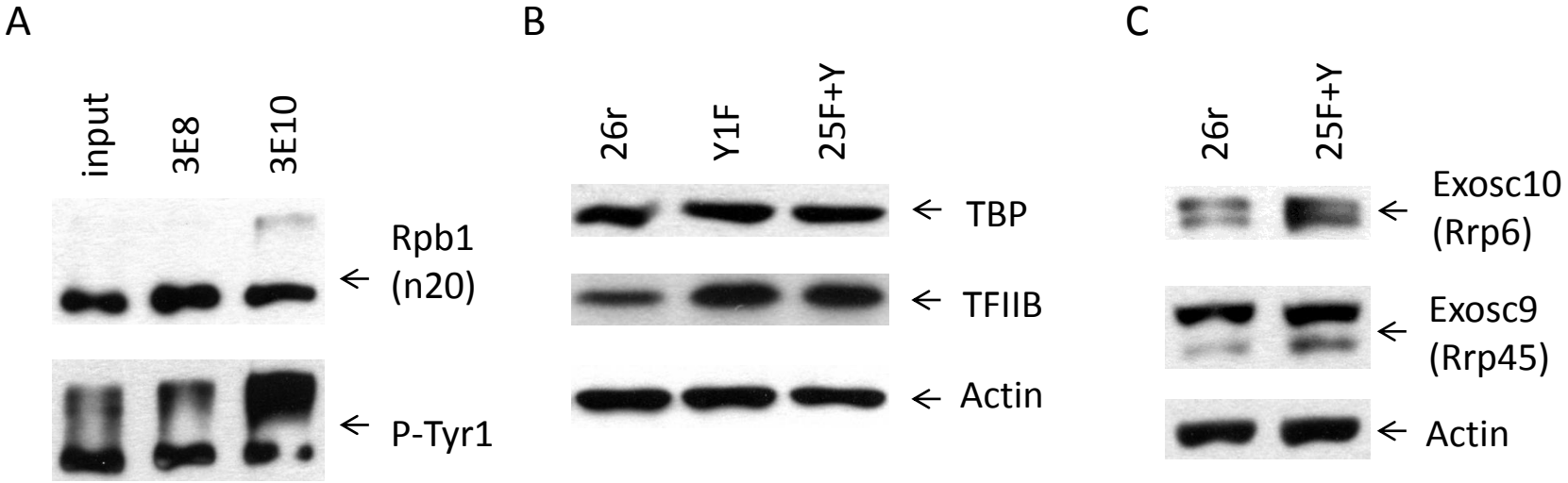

A
input
3E8
3E10
Rpb1
(n20)
P-Tyr1
B
26r
Y1F
25F+Y
TBP
TFIIB
Actin
C
26r
25F+Y
Exosc10
(Rrp6)
Exosc9
(Rrp45)
Actin

Figure S6.

Sample	Read	Pass read	% PASS
26r #1	4,508,084	2,257,806	50.08
26r #2	6,388,457	3,064,875	47.98
25F+Y #1	3,902,921	1,818,503	46.59
25F+Y#2	5,017,448	2,436,559	48.56
S2A #1	5,978,939	3,122,647	52.22
S2A #2	4,911,291	2,440,795	49.70
S5A #1	3,966594	1,865975	47.04
S5A #2	6,209,492	3,032,265	48.83