Title
Chemical-genomic dissection of the CTD code

Journal
Nature Structural & Molecular Biology, 17(9)

ISSN
1545-9993

Authors
Tietjen, Joshua R
Zhang, David W
Rodríguez-Molina, Juan B
et al.

Publication Date
2010-09-01

DOI
10.1038/nsmb.1900

Peer reviewed

HHS Public Access

Author manuscript

Nat Struct Mol Biol. Author manuscript; available in PMC 2014 May 27.

Published in final edited form as:

Nat Struct Mol Biol. 2010 September ; 17(9): 1154–1161. doi:10.1038/nsmb.1900.

Chemical-genomic dissection of the CTD code

Joshua R. Tietjen[1], **David W. Zhang**[1], **Juan B. Rodríguez-Molina**[1], **Brent E. White**[1], **Md. Sohail Akhtar**[1], **Martin Heidemann**[2], **Xin Li**[1,3], **Rob D. Chapman**[2], **Kevan Shokat**[4], **Sündüz Keles**[3], **Dirk Eick**[2], and **Aseem Z. Ansari**[1,*]

[1] Department of Biochemistry and The Genome Center, University of Wisconsin, Madison, WI 53706

[2] Institute of Clinical Molecular Biology and Tumour Genetics, Helmholtz Center Munich, Center of Integrated Protein Science (CIPSM), Munich, Germany

[3] Department of Biostatistics & Medical Informatics, University of Wisconsin, Madison, WI

[4] Cellular and Molecular Pharmacology, University of California, San Francisco, CA 94158

Summary

Sequential modifications of the RNA polymerase II (Pol II) carboxyl-terminal domain (CTD) coordinate the stage-specific association and release of cellular machines during transcription. Here we examine the genome-wide distributions of the “early” (phospho-serine 5), “mid” (phospho-serine 7) and “late” (phospho-serine 2) CTD marks. We identify gene-class specific patterns and find widespread co-occurrence of the CTD marks. Contrary to its role in 3’ processing of non-coding RNA, the Ser7-P marks are placed early and retained until transcription termination at all Pol II-dependent genes. Chemical-genomic analysis reveals that the promoter-distal Ser7-P marks are not remnants of early phosphorylation, but are placed anew by the CTD kinase Bur1. Consistent with the ability of Bur1 to facilitate transcription elongation and suppress cryptic transcription, high levels of Ser7-P are observed at highly transcribed genes. We propose that Ser7-P could facilitate elongation and suppress cryptic transcription.

Introduction

The carboxyl-terminal domain (CTD) of the largest subunit of RNA polymerase II (Pol II) orchestrates dynamic interactions with proteins that are required for various stages of transcription [1]. The structural plasticity of the CTD and its proximity to the RNA exit tunnel of the polymerase enables it to interact with multiple protein complexes, including those that

* Address correspondence to: ***Aseem Z. Ansari*** Department of Biochemistry & The Genome Center of Wisconsin University of Wisconsin-Madison 433 Babcock Drive Madison, WI 53706 Phone: (608) 265 4690 Fax: (608) 262 3453 ansari@biochem.wisc.edu.

Contributions

JRT, DWZ, and BEW performed the ChIP-chip experiments in the JTY1 and Kin28as+Srb10as strains. JBRM performed the ChIP-chip experiments with the BY4743 and Bur1as strains. JRT analyzed the ChIP-chip data for all the experiments. MSA and DWZ performed the kinase assays. XL and SK assisted with data analysis. MH performed the ELISA assays. AZA, RC, and DE planned and analyzed the Ser7-P experiments. KS provided the unpublished small molecule inhibitors and assisted in the planning of the kinase inhibition experiments. JRT, DWZ, JBRM, and AZA wrote the manuscript and all authors commented on the manuscript.

process the nascent transcript (Supplementary Fig. 1). The CTD is composed of 26 repeating heptapeptides (**$Y_1S_2P_3T_4S_5P_6S_7$**) in budding yeast. Five of the seven residues (Y_1, S_2, T_4, S_5, and S7) can be phosphorylated or glycosylated and the proline residues (P_3 and P_6) can exist in two stereoisomeric states (cis/trans). The stage-specific association and exchange of protein partners is mediated by sequential post-translational modifications of different residues of the heptapeptide repeats [1-4]. During transcription initiation, Ser5 residues of the CTD are phosphorylated by the Cdk7/Kin28 subunit of TFIIH and by the Cdk8/Srb10 subunit of the Mediator complex [5-10]. This “early” modification releases Pol II from the promoter bound preinitiation complex [8,11] and facilitates the association of the capping enzyme complex and the Set1 histone methyltransferase [12-16]. Shortly after promoter release, Rtr1, an atypical phosphatase, erases the phospho-Ser5 (**Ser5-P**) marks on the elongating Pol II [17]. Next, the Cdk9 kinase of the P-TEFb complex phosphorylates the Ser2 residues of the CTD [7,18]. This “late” post-translational mark facilitates transcription elongation, as well as the association of splicing factors and the Set2 histone methyltransferase that places repressive marks to prevent cryptic transcription within coding regions [1,4,19,20]. In *Saccharomyces cerevisiae*, the role of Cdk9 is split between two homologous kinases Ctk1 and Bur1 [21]. Ctk1 is the primary Ser2 kinase whereas Bur1 is thought to stimulate elongation and suppress cryptic transcription by acting on non-CTD substrates [21-25]. Toward the end of the coding region, the phospho-Ser2 marks (**Ser2-P**) are recognized in conjunction with the polyadenylation sequence by proteins involved in polyadenylation of the nascent transcript [26-28]. During transcription termination, Fcp1 and Ssu72, two CTD phosphatases, de-phosphorylate the CTD, making it available for the next cycle of transcription [29-31]. In mammalian cells, phospho-Ser7 (**Ser7-P**) marks are important for recruitment of the Integrator, a complex that plays an important role in 3’ processing of non-coding RNA [32,33]. While the substitution of this evolutionarily conserved residue with an alanine (S7A) has modest affects on growth, the substitution with a charged glutamic acid (S7E) is lethal [34,35]. Thus, the placement and more importantly the removal of this phosphate moiety plays a critical role in the transcription cycle.

The CTD code hypothesis posits that sequential CTD marks specify a stage-specific recognition code in a manner akin to the histone code [1-3,7,36,37]. Whether or not the patterns that underlie the CTD code are retained across different classes of genes in the genome remains a fundamentally important question. The observations that Ser2-P marks are not required for the expression of the p53-regulated p21 gene, histone genes or snRNA genes raises the spectre that this mark is not universally needed for RNA synthesis [38,39]. The early Ser5-P mark, thought to be essential for promoter release, is largely dispensable for mRNA synthesis [8,40-42]. Despite its role in 3’ processing of non-coding genes [32,34], the Ser7-P mark is observed at promoters of all Pol II transcribed genes [43-45]. It is unclear what role Ser7-P marks play at protein-coding genes. Furthermore, the crosstalk between different modifications is an important, but poorly understood phenomenon. Early experiments suggested that Ser5-P and Ser2-P modifications were placed independently [29]. In fact, inhibition of one mark was shown to increase the abundance of the other, suggesting an antagonistic relationship [46]. However, recently Ser5-P was shown to prime the CTD for subsequent phosphorylation of Ser2 by Bur1/Cdk9 [22,47,48]. Importantly, the interdependence

and co-occurrence of Ser7-P is only now beginning to be explored. Thus, many questions about the universality of the sequential patterning of CTD marks remain to be resolved.

Our genome-wide analysis reveals an unexpected degree of co-occurrence of all three CTD marks, suggesting a bivalent or even multivalent mode of recognition by docking partners. A particularly surprising observation is that Ser7-P is placed early in transcription and retained at significant levels until transcription termination. Chemical inhibition studies demonstrate that the promoter-distal Ser7-P marks are not remnants of early phosphorylation events, but are replenished by Bur1. This defines a new role for Bur1 and suggests that Ser7-P marks are important for efficient transcription elongation. Finally, while we highlight the differential patterns of CTD marks at coding and non-coding genes, our genome-wide atlas of CTD marks serves as a general resource to identify new regulatory mechanisms that underlie the CTD code.

Results

Genome-wide analyses reveal unexpected profiles

To explore the universality of the CTD code, we examined the occupancy profiles of Pol II and the three major CTD marks across the *S. cerevisiae* genome. We performed chromatin immunoprecipitation (ChIP) experiments and identified enriched DNA fragments via high-resolution tiled genomic microarrays (ChIP-chip). Pol II was immunoprecipitated using a monoclonal antibody against Rpb3, an integral subunit of the polymerase that is not influenced by CTD phosphorylation. CTD phosphorylations were examined using epitope-specific antibodies (see methods for details). The high-resolution profiles revealed novel patterns of Pol II association across some genes while confirming known binding patterns at other genes (Fig. 1, traces in blue). For example, high occupancy of Pol II across the ribosomal protein gene RPL16B and rapid depletion of Pol II across the NRD1 gene have been well documented (Fig. 1a) [49]. On the other hand, the enrichment of Pol II at the 3' end of the MRPL4 gene or the enrichment at the 5' and 3' ends of the RIM1 gene are new findings (Fig. 1a). Although cryptic unstable transcripts (CUTs), stable unannotated transcripts (SUTs) [50,51], and neighboring convergent genes may contribute to some of these profiles, there are some genes at which there is no neighboring or overlapping transcription to account for the 3' enrichment of Pol II (Supplementary Fig. 2).

We then examined the genome-wide patterns of CTD phosphorylation. To ensure that the profiles of CTD modifications were not skewed by unusual Pol II profiles, we focused on genes bearing uniformly high levels of Pol II across the transcription unit (Fig. 1b). Particular examples include the protein-coding (**pc**) genes: PDC1, COX18 and PSA1, as well as the non-coding (**nc**), polycistronic cluster: SNR78-72. The Ser7-P profile across the SNR cluster is enriched at the 5' end of the gene with dissipation of the signal towards the 3' end of the transcription unit. The pc-genes show different Ser7-P profiles. Unlike previous reports in human cells [32,34], we fail to detect Ser7-P enrichment solely in the middle of the coding region across the yeast genome. On the other hand, the reciprocal enrichment of Ser5-P at promoters and Ser2-P at the 3' ends is frequently observed and defines the current paradigm. However, the levels of Ser5-P across COX18 and Ser2-P on the SNR78-72 cluster do not conform to the paradigm and are lower than expected. The unexpected

patterns of CTD marks at different genes are not due to experimental variability (Supplementary Fig. 3) but may arise from CUTs and SUTs.

Clustering genome-wide profiles of Pol II and the three major CTD marks

To examine commonalities between patterns of Pol II and its modifications across the genome, we used an average transcription unit analysis (Fig. 2a). In this analysis, the entire transcription unit, from the transcription start site (TSS, arrow at 5' end) to the cleavage and polyadenylation site (CPS, vertical red bar at the 3' end), was represented by 10 evenly scaled bins across all annotated genes in the genome (see methods for details). The Pol II and CTD modification profiles are sorted by unrestrained k-means clustering. The profiles coalesce into four general clusters (Fig. 2b): uniform enrichment across the transcription unit (denoted by **U**), 5' enrichment (denoted by **5**), 3' enrichment (denoted by **3**), or 5' and 3' enrichments (denoted by **5+3**). The four panels show the clustered profiles of Pol II and its modifications across all genes in the genome (Fig. 2b).

The compact nature of the yeast genome complicated analysis due to signal bleed-over from neighboring genes and their regulatory elements. Therefore, we parsed the genes into three sets: the first included all 6147 annotated genes in the genome, the second focused on 615 genes that have the highest levels (top 10%) of Pol II occupancy, and the third set is comprised of 60 genes that display robust Pol II occupancy but are isolated from other known or predicted genes by at least 400 base pairs. Importantly, genes in the third set are also devoid of overlapping or neighboring CUTs and SUTs. The profiles of Pol II and its modifications across the three sets of genes are summarized as pie charts (Fig. 2c).

The summary profiles of the 60 well-isolated pc-genes (Set III) show that Ser7-P and Ser5-P marks are enriched at 5' end of genes (combining **5**, **5+3**, **U**) whereas the Ser2-P mark is enriched at the 3' end of genes. A quantitative fit of the profiles show that Ser5-P and Ser7-P maxima occur within 50 bp of the start site whereas the Ser2-P marks reach their first maxima by 517±226 bp (Supplementary Fig. 4a-b). Unexpectedly, at ~36% of the genes, the Ser5-P "early" mark is also enriched at the 3' end (**5+3** profile). The non-canonical enrichment of Ser5-P at the 3' ends and the occurrence of "late" Ser2-P mark at the 5' ends of genes are surprising. Moreover, only 15% of the Set III genes display the canonical Ser5-P enrichment only at the 5' end followed by the reciprocal Ser2-P enrichment at the 3' end of transcription units. Thus, the accepted paradigm that Ser5-P is enriched only at promoters, Ser7-P in the middle of the coding regions and Ser2-P toward the end does not adequately describe the patterns of CTD marks observed across the genome and therefore should be revised.

To assess the functional importance of Ser7-P profiles, we examined gene ontology (GO) enrichment within each cluster. The clusters obtained using k-means were refined using silhouette analysis in which scores were assigned to the genes within each cluster. A silhouette score gives a measure of how well a gene profile fits in a given cluster. Iterative optimization allowed us to obtain the highest-confidence genes for each cluster (Supplementary Fig. 5). We then examined the clusters to find groups of ontologically related genes unique to each cluster (Fig. 2d, Supplementary Fig. 6, and Supplementary Table 1). The strong enrichment of ribosomal biogenesis genes with the uniform (**U**) Ser7-P

profile was especially striking ($P < 10^{-14}$). To further validate this result, we performed a hypergeometric comparison of genes within the **U** cluster with known genome-wide binding sites of all yeast transcription factors. Consistent with the GO enrichment, transcription factors involved in ribosome biogenesis (FHL1, RAP1, and SFP1) were significantly enriched in the **U** cluster (Supplementary Fig. 6). To examine the generality of our initial observations, we compared the known rates of RNA synthesis for genes within each cluster [52]. The **U** cluster was enriched for highly transcribed genes. The reciprocal analysis of the genes transcribed at the highest frequency revealed that 86% had the **U** Ser7-P profile. (Fig. 2e and Supplementary Fig. 6c). Further comparison with genome-wide maps of components of the transcriptional machinery, chromatin remodeling machines, and transcription factors revealed that genes within the **U** cluster were highly enriched for factors that stimulate high levels of transcription (Fig. 2f and Supplementary Fig. 7) [53]. In essence, multiple analyses of datasets from independent sources strongly tie highly expressed genes to uniform patterns of Ser7-P marks across the transcribed regions. Thus, Ser7-P mark may contribute to efficient transcription elongation.

Distinct patterns at non-coding versus protein-coding genes

Ser7-P is only thought to be important for 3' processing of snRNA; however, it is also placed robustly at pc-genes. It is unknown what role this CTD mark may play at pc-genes. We therefore examined if the patterns of CTD marks differ between nc- and pc-genes (Fig. 3a). The Ser7-P and Ser5-P profiles on SNR7 and SNR14 are strikingly similar but they differ from the profiles at pc-genes (YDJ1 and FBA1 in Fig. 3a). At pc-genes, Ser7-P and Ser5-P profiles coincide at the promoter but rapidly deviate within coding regions. While Ser5-P levels abate, Ser7-P levels are retained at high levels until the 3' end. Moreover, dramatically lower levels of Ser2-P are apparent across nc-genes when compared to pc-genes. None of the profiles display any correlation with the well-positioned nucleosomes at promoters of genes (Fig. 3a).

To extend the comparison beyond a few representative genes, we calculated the average levels of Ser7-P and Ser2-P marks for the 82 nc-genes (snRNA and snoRNA) transcribed by Pol II and the 60 pc-genes of Set III (from Fig. 2b). Other than distance from neighboring genes, the Set III pc-genes are ontologically unrelated to each other and they provide unambiguous occupancy profiles. The genome-scale data clearly show that the extent of Ser7-P modification on both classes of genes is equivalent. However, the levels of Ser2-P are 400% lower on nc-genes ($P = 4.4\times10^{-21}$) (Fig. 3b). Even when correcting for the shorter size of nc-genes, we find that the levels of Ser2-P are 275% lower than the signals observed within the first 300bp of pc-genes ($P = 1.9\times10^{-10}$) (Supplementary Fig. 4c). The lower levels of Ser2-P marks and the abundance of Ser7-P at nc-genes likely serve as a signal for enzymes that process nc-transcripts.

Chemical-genomic inhibition reveals crosstalk between CTD marks

To examine the crosstalk between the CTD marks, we used a chemical-genetic approach to inhibit the two major kinases that phosphorylate Ser5 during transcription initiation: TFIIH-associated Cdk7/Kin28, and to a lesser extent the Mediator-associated Cdk8/Srb10 (Supplementary Fig. 8a) [8,40]. To inhibit these two kinases selectively, their ATP-binding

pockets were genetically altered to accept cell-permeable inhibitor analogs (Supplementary Fig. 8a) [8,54]. Consistent with our previous studies, targeted chemical inhibition of Kin28/Cdk7 in combination with Srb10/Cdk8 (Kin28as+Srb10as) blocks a significant portion of the kinase activity and leads to a rapid decrease in cell growth (Supplementary Fig. 8a) [8,40]. Chemical inhibition of Kin28as+Srb10as *in vivo* leads to a precipitous decrease in Ser5-P marks at the 5' end of several genes (SED1 in Fig. 4a). Unexpectedly, under similar treatment Ser7-P marks are only attenuated at the promoter. The promoter-proximal loss of both marks is particularly evident upon normalization to Pol II (Supplementary Fig. 8b).

To further clarify the remodeling of occupancy patterns, we examined the Set II genes that show uniformly high levels of Pol II occupancy throughout the transcription unit (Fig. 4b). The profiles from the untreated and the chemically inhibited double mutant strain were binned as described above (see Fig. 2). The mean of 415 genes is displayed (Fig. 4b). Upon inhibition, the Pol II profile shows clear decrease in the middle of the coding region resulting in a bimodal distribution with peaks at the 5' and 3' ends of the transcription unit (profile $\mathbf{I^{5'+3'}}$). Ser5-P and Ser7-P marks are dramatically reduced at 5' ends but they are unperturbed within coding regions. In agreement with recent reports [47,48], this decrease in Ser5-P leads to a subtle but detectable decrease in Ser2-P. However, it is important to note that in contrast to Ser5-P and Ser7-P, the overall profile of the Ser2-P mark does not change.

We then examined the consequences of inhibiting Kin28as+Srb10as across the genome (Fig. 4c and Supplementary Fig. 10). The profile clusters for the inhibited strain resemble those seen previously (**U**, **5**, **3**, **5+3**), but they were labeled with an "**I**" (inhibitor treated) to differentiate them from the clusters from the untreated wild type strain (Fig. 2c and Fig. 4c). Using this nomenclature, we compared the profiles from wild type strains (central pie charts: **U**, **5**, **3**, **5+3**) with the profiles for the same genes in the chemically inhibited Kin28as +Srb10as strain (surrounding pie charts: $\mathbf{I^U}$, $\mathbf{I^5}$, $\mathbf{I^3}$, $\mathbf{I^{5+3}}$; Fig. 4c and Supplementary Fig. 10). The inhibition data for Pol II is striking, as a significant fraction of genes from the uniform and 5'-enriched clusters (**U**, **5**) are altered to the 5'+3' enriched profile ($\mathbf{I^{5+3}}$ in dashed circles). In addition, the Ser7-P profiles are remodeled with a loss of 5'-enriched patterns and an increase of 3' enrichment ($\mathbf{I^3}$) within the coding regions. Ser5-P profiles are also remodeled, as the 5'-enrichment is lost. The 3'-enriched fraction of Ser2-P profiles is not dramatically altered. Taken together, the data indicate that promoter-proximal Ser5-P and Ser7-P marks are strongly coupled, whereas promoter-distal Ser2-P and Ser7-P marks are not strictly dependent on early CTD marks. The coupling of Ser5-P and Ser7-P at promoters is consistent with the role of Kin28/Cdk7 in phosphorylating these residues [43-45]. However, the occurrence of Ser7-P within coding regions is unexpected.

An internal Ser7 kinase acts within coding regions of pc-genes

To determine if Ser7-P marks detected within coding regions are not remnants of marks placed at the promoter we examined the distribution of this mark at the inducible GAL1 gene. We first inhibited Kin28as+Srb10as for 60 minutes in non-inducing conditions. Under these conditions, we detect no Pol II at the GAL1 promoter and have previously shown that both Kin28as and Srb10as remain chemically inhibited [8,40]. Next, we induce the GAL genes for 30 minutes, and use ChIP-qPCR to examine the distribution of Pol II and Ser7-P marks

at the 5' and 3' ends of the GAL1 gene (Fig. 5a). The data reveal an unambiguous diminution of Ser7-P marks at the 5' end and a robust signal for this mark at the 3' end. In other words, the internal Ser7-P marks are placed anew by a different kinase(s), one that is not sensitive to inhibitors of Kin28as and Srb10as.

To explore if internal Ser7-P marks are also placed anew in nc-genes, we compared the changes in the Ser7-P profiles at pc- and nc-genes that are of similar length and have similar polymerase occupancy. The SNR78-72 polycistronic cluster and GLN1 gene display comparable levels of Ser7-P and Ser5-P marks at the promoter and are equally responsive to the inhibitor at the 5' end (Fig. 5b). However, in contrast to GLN1, Ser7-P marks are not replenished at the 3' end of the SNR78-72 cluster. We systematically examined the levels of Ser7-P marks at the 3' end of 82 nc-genes and the 60 pc-genes from Set III (Fig. 5c). The data unambiguously demonstrate that inhibiting promoter-proximal kinases leads to a dramatic loss of Ser7-P at nc-genes but not at pc-genes. Thus, chemical-genomic analysis reveals the existence of a Ser7 kinase that acts on elongating polymerases.

To identify the internal Ser7 kinase, we purified the four kinases that phosphorylate the CTD during different stages of transcription: Kin28/Cdk7, Srb10(Srb11)/Cdk8, Bur1/Cdk9, and Ctk1/Cdk9 (Supplementary Fig. 1). Consistent with our previous work, Kin28 phosphorylated Ser5 and Ser7 residues on a GST fusion protein bearing four or more heptapeptide repeats (Fig. 6a) [43]. While Bur1 does not phosphorylate substrates bearing three heptapeptide repeats[54], it phosphorylates Ser7 on substrates bearing four or fourteen repeats (Fig. 6a and Supplementary Fig. 11a). Our results are consistent with recent reports of Bur1 preferentially phosphorylating longer CTD substrates [37]. The kinase activity was quantified using an ELISA (Fig. 6b). Importantly, neither Srb10 nor the control Cka1 kinase phosphorylated Ser7 in parallel assays (Supplementary Fig. 11a).

To validate these observations, we used a strain where Bur1 was engineered to bind a bulky purine analog, 3-MB-PP1 [22]. Unlike the wild type enzyme, the enlarged ATP-binding site of the engineered Bur1 (Bur1as) is inhibited by 3-MB-PP1 and can utilize the bulky ATP analog N^6-benzyl-ATP to phosphorylate Ser7 *in vitro* (Fig. 6c). The inhibitor has only a modest affect on the growth rate of Bur1as cells, suggesting a partial *in vivo* inhibition of this kinase whose function is vital for cell viability (Supplementary fig. 11b). However, this degree of inhibition is sufficient to reduce the ability of Bur1 to modify Spt5 or CTD *in vivo* [22,48] In agreement with its *in vitro* activity, chemical inhibition of the Bur1as strain led to a loss of Ser7-P marks *in vivo* at the SNR cluster and GLN1 (Supplementary Fig. 11d). This loss of Ser7-P within the coding region is not observed in chemically treated isogenic wild type strain or the Kin28as+Srb10as strain (Supplementary Fig. 11c). In contrast, the loss of Ctk1 nearly abolishes Ser2-P marks but has no affect on Ser7-P at the SNR cluster and the internal site of GLN1. The decrease in Ser7-P levels at GLN1 in ctk1Δ strain may result from latent Ser7-P activity of this kinase *in vivo*, or the inability of Pol II to release basal transcription factors in the absence of Ctk1 [55].

To examine the genome-wide contribution of Bur1 to Ser7-P profiles, we performed ChIP-chip analysis on the Bur1as strain in the presence or absence of the inhibitor. Unlike the promoter-proximal decrease in Ser7-P upon inhibition of Kin28as+Srb10as, inhibition of

Bur1 reduces Ser7-P within the coding region (Fig. 6d). The summarized data on Set II genes with uniformly Pol II profiles show that the decrease is most apparent within the coding region, with the greatest effect toward the 3' end (Fig. 6e). Moreover, a direct comparison of changes in Ser7-P levels upon chemical inhibition of Kin28as+Srb10as or Bur1as further highlights the role of Kin28 at the 5' end and Bur1 at the 3' end of genes (Fig. 6f). The results are consistent with the ability of Bur1 to act on elongating polymerases from promoter-proximal regions until the 3' end of genes.

Discussion

This study provides the first genome-wide atlas of CTD modifications that orchestrate stage-specific processes in RNA biogenesis. The data reveal: 1) ubiquitous co-occurrence of CTD marks across the transcribed units, 2) existence of gene-class specific patterns of Pol II and its modifications, 3) weak coupling between promoter-proximal and promoter-distal CTD marks, 4) Ser7-P marks within coding regions are placed anew and are not remnants of marks placed by Kin28 at the promoter, and 5) a novel role of Bur1 as a Ser7 kinase that acts on elongating Pol II. When considered in its entirety, our data suggest that accepted paradigms underlying the "CTD code" should be revised to accommodate combinatorial patterns of CTD modifications.

The widespread co-occurrence of CTD marks does not mean that they are placed at equivalent levels at all genes across the genome. We observe distinct patterns of the three marks between nc- and pc-genes. Ser7-P closely mirrors the patterns of Ser5-P across the entire nc-gene, whereas at pc-genes the Ser7-P mark persists well beyond the point where Ser5-P levels drop significantly. Moreover, Ser2-P marks are underrepresented at nc-genes, even when length of the transcript is taken into consideration. The distinct patterns of CTD marks at these two gene classes reflect the different mechanisms of transcription termination and 3' end processing that act on these two classes of RNA [49,56-59].

The co-occurrence of CTD marks is consistent with our observation that two kinases known to phosphorylate the CTD also place bivalent marks. We and others previously identified the ability of Kin28/Cdk7 to phosphorylate Ser5 and Ser7 [43-45]. Because, Kin28/Cdk7 associates with Pol II at promoters, the co-occurrence of Ser5 and Ser7 marks within 50 bp of the transcription start site can be ascribed to the action of Kin28/Cdk7. We also show that Srb10, another Ser5 kinase, does not contribute to Ser7-P marks at promoters or within coding regions (Fig. 4). Inhibition of Kin28as+Srb10as revealed, for the first time, the existence of other kinases that act on Ser7 residues of the elongating polymerase. A series of kinase assays on CTD substrates bearing more than three heptapeptide repeats revealed that Bur1 phosphorylates Ser7. The requirement for longer CTD substrates was previously noted in studies that examined the ability of Bur1 to phosphorylate Ser2 [48]. The inhibition of Bur1as with cell permeable inhibitors further confirmed its role as a Ser7-P kinase *in vivo* (Fig. 6d-e, Supplementary Fig. 11c-d). This is a new role for Bur1, one that couples the placement of Ser7-P with Ser2-P on elongating polymerases.

In performing these experiments, we were surprised to find that promoter-distal Ser7-P marks were not eliminated by inhibition of the promoter-proximal Ser7 kinase. This

observation is inconsistent with the notion that promoter-distal Ser7-P marks were remnants of the marks placed by Kin28 at the promoter. Similarly, promoter-distal Ser2-P and Ser5-P marks were not eliminated upon simultaneous inhibition of promoter-proximal Ser5 kinases (Srb10 and Kin28). Moreover, it was recently suggested that promoter-proximal Ser5-P marks were critical for sequential placement of Ser2-P marks on the elongating Pol II. Thus, contrary to the expectation that early marks prime the CTD for subsequent modifications, our data indicate that promoter-distal placement of CTD modifications is not acutely dependent on early marks.

Previous mutational analyses highlight the importance of dynamic placement of Ser7-P marks. The substitution of Ser7 with a glutamic acid (S7E) is lethal whereas the inability to phosphorylate this position (S7A in T4A containing CTDs) only results in a decreased growth rate at suboptimal temperatures [34,35]. The removal and replenishment of Ser7-P marks within coding regions may serve as a critical signal for the exchange of factors during the transition from promoter-release to productive elongation. A role for Ser7-P in efficient elongation is consistent with the observation that highly transcribed genes show high levels of this CTD mark. Moreover, the ability of Bur1 to place this mark is consistent with its ability to associate with transcribing Pol II and stimulate transcriptional elongation [24,25]. In addition to improving elongation via phosphorylation of Spt5 [22,23], Bur1-mediated placement of Ser7-P marks could help recruit cellular machinery that facilitates elongation and suppresses cryptic transcripts. This could well explain why Ser7-P marks are observed across all protein-coding genes even though their function is only defined at non-coding genes.

Methods

Genome-wide location and transcriptional analyses

Chromatin immunoprecipitation (ChIP), with antibodies against Pol II (αRPB3) and the Ser7-P (4E12), Ser5-P (H14), and Ser2-P forms of Pol II, was performed as described previously with several modifications[1] (see Supp. Methods). The ChIP samples were amplified using ligation-mediated PCR and hybridized to high density tiling microarrays from NimbleGen (Roche NimbleGen, Inc.; Madison, WI). Total RNA samples were isolated, prior to formaldehyde crosslinking, during the ChIP experiments, and were processed by Roche NimbleGen hybridization to high density tiling microarrays having the same design as those used for the ChIP-chip experiments.

Data Analysis

ChIP-chip Immunoprecipitated (IP) data was mean-scaled against its respective “Input” sample data and then the ratio of scaled IP to the Input was Log_2 transformed. For total RNA data, the probe intensities were divided by the peak intensity from the raw data histogram and then Log_2 transformed. Both data types were subjected to computational repeat sequence masking based on probe sequence repetitiveness relative to the sequence composition of the probes on the microarray (see Supp. Methods). A moving average was used to smooth the microarray data, and baseline corrections were applied through comparison between polymerase and transcript data (see Supp. Methods).

Average transcription unit analysis was applied to the data to obtain occupancy profiles normalized for gene length for every gene in gene sets I, II, and III. This allowed for alignment of the occupancy patterns for all the genes in the genome (see Supp. Methods). K-means clustering was performed to partition the genes into categories based on their occupancy profiles. These clusters were then manually collapsed into four representative groups. Clusters were refined through iterative silhouette score calculations, cluster reassignments, and gene silhouette score filtering[2] (see Supp. Methods). Transcription factor enrichments were determined via hypergeometric probabilities calculated from published binding data[3].

Kinase Assays

The kinase assay was done as described previously[4]. Briefly, 200ng of GST-CTD_4 or $GSTCTD_{14}$ (four or fourteen repeats of YSPTSPS fused to GST) or its alanine substitutions (2A, 5A and 7A) were phosphorylated by four different kinases or whole cell extract at 25°C for one hour in a buffer containing 20mM HEPES [pH 7.5], 2.5mM EGTA, 15mM magnesium acetate, 0.8mM ATP, 10% glycerol, protease inhibitors, and phosphatase inhibitors [1mM NaN3, 1mM NaF, 0.4mM Na3VO3]. Four micro-liters from the reaction mix were directly spotted onto nitrocellulose membrane (GE Healthcare) and was processed further as a standard dot blotting protocol. For the assay, a 1:50(αSer2-P), 1:500(αSer5-P), or 1:200(αSer7-P) dilution of primary rat IgG and 1:10000 dilution of secondary HRP anti-rat IgG (Southern Biotech) antibodies were used.

For the Bur1as inhibition study, 200ng of GST-CTD14 was used. TAP-tagged WT Bur1 and FLAGx3-tagged Bur1as were incubated with 5uM 3MB-PP1 for 5 minutes in kinase assay buffer sans ATP at room temperature. Following the inhibition, 0.8mM ATP or 0.6mM N6-Benzyl-ATP was added and the reaction was allowed to run for 30 minutes at room temperature. Five micro-liters from the reaction mix were directly spotted onto nitrocellulose membrane and processed as a standard dot blotting protocol. For the assay, a 1:30(αSer2-P), 1:1000(αSer5-P), or 1:100(αSer7-P) dilution of primary rat IgG and 1:12500 dilution of secondary HRP anti-rat IgG (Southern Biotech) antibodies were used.

For ELISA, the recombinant kinase (100ng) was incubated with excess CTD-peptide (YSPTSPSYSPTSPSYSPTSPSYSPTSPSC; Peptide Specialty Laboratories GmbH, Heidelberg) linked to 96-well plates in a 25μl kinase buffer containing 20mM Tris-Cl [pH 7.4], 20mM NaCl, 10mM MgCl2, 1μM DTT and 2μM ATP at 28°C for 60 minutes. Phosphorylation of CTD was quantitated via ELISA experiments after incubation with αSer2-P, αSer5-P and αSer7-P specific antibodies.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank Steve Hahn (Fred Hutchinson Cancer Research Center) for sharing strains before publication and for helpful discussions, John Nau and Andrew Nett for optimizing strains, and Elenita Kanin and Alison Leaf for exploratory experiments at UW-Madison. We thank Dick Burgess and Nancy Thompson (UW-Madison) for the

gift of Pol II antibody (RPB3). We also gratefully acknowledge the support of the NSF (MCB 07147), W. M. Keck, Shaw Scholar and Vilas Associate awards (to AZA). JRT and JRM were supported by an NHGRI training grant to the Genomic Sciences Training Program (5T32HG002760). DE was supported by Deutsche Forschungsgemeinschaft SFB/TR5 and SFB684 and SK was supported by HG03747 (NIH). MH was funded by Boehringer Ingelheim Fonds travel grant

References

1. Phatnani HP, Greenleaf AL. Phosphorylation and functions of the RNA polymerase II CTD. Genes & Development. 2006; 20:2922–2936. [PubMed: 17079683]
2. Buratowski S. The CTD code. Nature Structural Biology. 2003; 10:679–680. [PubMed: 12942140]
3. Corden JL. Transcription: Seven Ups the Code. Science. 2007; 318:1735–1736. [PubMed: 18079391]
4. Perales R, Bentley D. "Cotranscriptionality": The Transcription Elongation Complex as a Nexus for Nuclear Transactions. Molecular Cell. 2009; 36:178–191. [PubMed: 19854129]
5. Lee T, Young R. Transcription of Eukaryotic Protein-Coding Genes. Annual Review of Genetics. 2000; 34:77–137.
6. Myers LC, Kornberg RD. Mediator of transcriptional regulation. Annual Review of Biochemistry. 2000; 69:729–749.
7. Sims RJ III, Belotserkovskaya R, Reinberg D. Elongation by RNA polymerase II: the short and long of it. Genes & Development. 2004; 18:2437–2468. [PubMed: 15489290]
8. Liu Y, et al. Two cyclin-dependent kinases promote RNA polymerase II transcription and formation of the scaffold complex. Molecular and Cellular Biology. 2004; 24:1721–1735. [PubMed: 14749387]
9. Ansari A, Ogirala A, Ptashne M. Transcriptional activating regions target attached substrates to a cyclin-dependent kinase. Proceedings of the National Academy of Sciences of the United States of America. 2005; 102:2346–2349. [PubMed: 15687503]
10. Riedl T, Egly JM. Phosphorylation in transcription: the CTD and more. Gene Expr. 2000; 9:3–13. [PubMed: 11097421]
11. Max T, Sogaard M, Svejstrup J. Hyperphosphorylation of the C-terminal repeat domain of RNA polymerase II facilitates dissociation of its complex with mediator. Journal of Biological Chemistry. 2007; 282:14113–14120. [PubMed: 17376774]
12. Schroeder SC, Schwer B, Shuman S, Bentley D. Dynamic association of capping enzymes with transcribing RNA polymerase II. Genes & Development. 2000; 14:2435–2440. [PubMed: 11018011]
13. Komarnitsky P, Cho EJ, Buratowski S. Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. Genes & Development. 2000; 14:2452–2460. [PubMed: 11018013]
14. Ng HH, Robert F, Young RA, Struhl K. Targeted recruitment of Set1 histone methylase by elongating Pol II provides a localized mark and memory of recent transcriptional activity. Molecular Cell. 2003; 11:709–719. [PubMed: 12667453]
15. Krogan NJ, et al. Methylation of Histone H3 by Set2 in Saccharomyces cerevisiae Is Linked to Transcriptional Elongation by RNA Polymerase II. Molecular and Cellular Biology. 2003; 23:4207–4218. [PubMed: 12773564]
16. Carrozza M, et al. Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. Cell. 2005; 123:581–592. [PubMed: 16286007]
17. Mosley A, et al. Rtr1 Is a CTD Phosphatase that Regulates RNA Polymerase II during the Transition from Serine 5 to Serine 2 Phosphorylation. Molecular Cell. 2009; 34:168–178. [PubMed: 19394294]
18. Brès V, Yoh S, Jones K. The multi-tasking P-TEFb complex. Current opinion in Cell Biology. 2008; 20:334–340. [PubMed: 18513937]
19. Buratowski S. Progression through the RNA polymerase II CTD cycle. Molecular Cell. 2009; 36:541–546. [PubMed: 19941815]

20. Kornblihtt A, De La Mata M, Fededa J, Munoz M, Nogues G. Multiple links between transcription and splicing. RNA. 2004; 10:1489–1498. [PubMed: 15383674]
21. Wood A, Shilatifard A. Bur1/Bur2 and the Ctk Complex in Yeast: The Split Personality of Mammalian P-TEFb. Cell Cycle. 2006; 5:1066–1068. [PubMed: 16721054]
22. Liu Y, et al. Phosphorylation of the transcription elongation factor Spt5 by yeast Bur1 kinase stimulates recruitment of the PAF complex. Molecular and Cellular Biology. 2009; 29:4852–4863. [PubMed: 19581288]
23. Zhou K, Kuo W, Fillingham J, Greenblatt J. Control of transcriptional elongation and cotranscriptional histone modification by the yeast BUR kinase substrate Spt5. Proceedings of the National Academy of Sciences USA. 2009; 106:6956–6961.
24. Chu Y, Simic R, Warner M, Arndt K, Prelich G. Regulation of histone modification and cryptic transcription by the Bur1 and Paf1 complexes. The EMBO Journal. 2007; 26:4646–4656. [PubMed: 17948059]
25. Keogh MC, Podolny V, Buratowski S. Bur1 Kinase Is Required for Efficient Transcription Elongation by RNA Polymerase II. Molecular and Cellular Biology. 2003; 23:7005–7018. [PubMed: 12972617]
26. Proudfoot NJ, Furger A, Dye MJ. Integrating mRNA Processing with Transcription. Cell. 2002; 108:501–512. [PubMed: 11909521]
27. Licatalosi DD, et al. Functional Interaction of Yeast Pre-mRNA 3' End Processing Factors with RNA Polymerase II. Molecular Cell. 2002; 9:1101–1111. [PubMed: 12049745]
28. Ahn SH, Kim M, Buratowski S. Phosphorylation of Serine 2 within the RNA Polymerase II C-Terminal Domain Couples Transcription and 3′ End Processing. Molecular Cell. 2004; 13:67–76. [PubMed: 14731395]
29. Cho EJ, Kobor MS, 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 & Development. 2001; 15:3319–3329. [PubMed: 11751637]
30. Krishnamurthy S, He X, Reyes-Reyes M, Moore C, Hampsey M. Ssu72 Is an RNA Polymerase II CTD Phosphatase. Molecular Cell. 2004; 14:387–394. [PubMed: 15125841]
31. Steinmetz E, Conrad N, Brow D, Corden J. RNA-binding protein Nrd1 directs poly (A)-independent 3 -end formation of RNA polymerase II transcripts. Nature. 2001; 413:327–331. [PubMed: 11565036]
32. Egloff S, et al. Serine-7 of the RNA Polymerase II CTD Is Specifically Required for snRNA Gene Expression. Science. 2007; 318:1777–1779. [PubMed: 18079403]
33. Baillat D, et al. Integrator, a multiprotein mediator of small nuclear RNA processing, associates with the C-terminal repeat of RNA polymerase II. Cell. 2005; 123:265–276. [PubMed: 16239144]
34. Chapman RD, et al. Transcribing RNA Polymerase II Is Phosphorylated at CTD Residue Serine-7. Science. 2007; 318:1780–1782. [PubMed: 18079404]
35. Stiller J, Mcconaughy B, Hall B. Evolutionary complementation for polymerase II CTD function. Yeast. 2000; 16:57–64. [PubMed: 10620775]
36. Egloff S, Murphy S. Cracking the RNA polymerase II CTD code. Trends in Genetics. 2008; 24:280–288. [PubMed: 18457900]
37. Jenuwein T, Allis C. Translating the histone code. Science. 2001; 293:1074–1080. [PubMed: 11498575]
38. Gomes N, et al. Gene-specific requirement for P-TEFb activity and RNA polymerase II phosphorylation within the p53 transcriptional program. Genes & Development. 2006; 20:601–612. [PubMed: 16510875]
39. Medlin J, et al. P-TEFb is not an essential elongation factor for the intronless human U2 snRNA and histone H2b genes. The EMBO Journal. 2005; 24:4154–4165. [PubMed: 16308568]
40. Kanin EI, et al. Chemical inhibition of the TFIIH-associated kinase Cdk7/Kin28 does not impair global mRNA synthesis. Proceedings of the National Academy of Sciences USA. 2007; 104:5812–5817.
41. Lee K, et al. Impairment of the TFIIH-associated CDK-activating kinase selectively affects cell cycle-regulated gene expression in fission yeast. Molecular Biology of the Cell. 2005; 16:2734–2745. [PubMed: 15829570]

42. Serizawa H, Conaway J, Conaway R. Phosphorylation of C-terminal domain of RNA polymerase II is not required in basal transcription. Nature. 1993; 363:371–374. [PubMed: 8497323]
43. Akhtar M, et al. TFIIH Kinase Places Bivalent Marks on the Carboxy-Terminal Domain of RNA Polymerase II. Molecular Cell. 2009; 34:387–393. [PubMed: 19450536]
44. Kim M, Suh H, Cho E, Buratowski S. Phosphorylation of the Yeast Rpb1 C-terminal Domain at Serines 2, 5, and 7. Journal of Biological Chemistry. 2009; 284:26421–26426. [PubMed: 19679665]
45. 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. Molecular and Cellular Biology. 2009; 29:5455–5464. [PubMed: 19667075]
46. Patturajan M, Conrad N, Bregman D, Corden J. Yeast carboxyl-terminal domain kinase I positively and negatively regulates RNA polymerase II carboxyl-terminal domain phosphorylation. Journal of Biological Chemistry. 1999; 274:27823–27828. [PubMed: 10488128]
47. Viladevall L, et al. TFIIH and P-TEFb Coordinate Transcription with Capping Enzyme Recruitment at Specific Genes in Fission Yeast. Molecular Cell. 2009; 33:738–751. [PubMed: 19328067]
48. Qiu H, Hu C, Hinnebusch A. Phosphorylation of the Pol II CTD by KIN28 Enhances BUR1/BUR2 Recruitment and Ser2 CTD Phosphorylation Near Promoters. Molecular Cell. 2009; 33:752–762. [PubMed: 19328068]
49. Steinmetz EJ, et al. Genome-Wide Distribution of Yeast RNA Polymerase II and Its Control by Sen1 Helicase. Molecular Cell. 2006; 24:735–746. [PubMed: 17157256]
50. Neil H, et al. Widespread bidirectional promoters are the major source of cryptic transcripts in yeast. Nature. 2009; 457:1038–1042. [PubMed: 19169244]
51. Xu Z, et al. Bidirectional promoters generate pervasive transcription in yeast. Nature. 2009; 457:1033–1037. [PubMed: 19169243]
52. Holstege F, Young R. Dissecting the regulatory circuitry of a eukaryotic genome. Cell. 1998; 95:717–728. [PubMed: 9845373]
53. Venters B, Pugh B. A canonical promoter organization of the transcription machinery and its regulators in the Saccharomyces genome. Genome Research. 2009; 19:360–371. [PubMed: 19124666]
54. Knight ZA, Shokat KM. Features of Selective Kinase Inhibitors. Chemistry & Biology. 2005; 12:621–637. [PubMed: 15975507]
55. Ahn S, Keogh M, Buratowski S. Ctk1 promotes dissociation of basal transcription factors from elongating RNA polymerase II EMBO Open. The EMBO Journal. 2009; 28:205–212. [PubMed: 19131970]
56. Arigo J, Eyler D, Carroll K, Corden J. Termination of cryptic unstable transcripts is directed by yeast RNA-binding proteins Nrd1 and Nab3. Molecular Cell. 2006; 23:841–851. [PubMed: 16973436]
57. Thiebaut M, Kisseleva-Romanova E, Rougemaille M, Boulay J, Libri D. Transcription termination and nuclear degradation of cryptic unstable transcripts: a role for the nrd1-nab3 pathway in genome surveillance. Molecular Cell. 2006; 23:853–864. [PubMed: 16973437]
58. Kim M, et al. Distinct pathways for snoRNA and mRNA termination. Molecular Cell. 2006; 24:723–734. [PubMed: 17157255]
59. Sheldon K, Mauger D, Arndt K. A requirement for the Saccharomyces cerevisiae Paf1 complex in snoRNA 3 end formation. Molecular Cell. 2005; 20:225–236. [PubMed: 16246725]
60. Lee W, et al. A high-resolution atlas of nucleosome occupancy in yeast. Nature Genetics. 2007; 39:1235–1244. [PubMed: 17873876]
1. Kanin EI, et al. Chemical inhibition of the TFIIH-associated kinase Cdk7/Kin28 does not impair global mRNA synthesis. Proceedings of the National Academy of Sciences USA. 2007; 104:5812–5817.
2. Rousseeuw P. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. Journal of Computational and Applied Mathematics. 1987; 20:53–65.
3. Harbison C, et al. Transcriptional regulatory code of a eukaryotic genome. Nature. 2004; 431:99–104. [PubMed: 15343339]

4. Ansari A, Ogirala A, Ptashne M. Transcriptional activating regions target attached substrates to a cyclin-dependent kinase. Proceedings of the National Academy of Sciences of the United States of America. 2005; 102:2346–2349. [PubMed: 15687503]

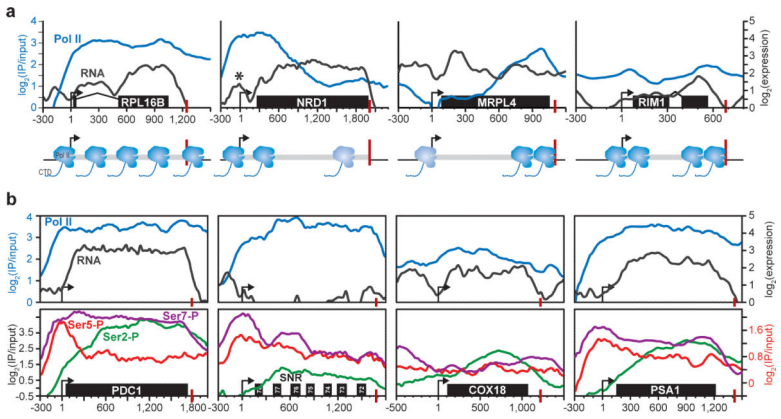

Figure 1. Pol II and CTD Phosphorylation Profiles

(a) ChIP-chip profiles for representative genes chosen to display the diversity of Pol II profiles across the genome. Pol II is shown in blue and total RNA is shown in black. Translation boundaries are indicated by the black boxes, and transcription boundaries shown as an arrow (Transcription Start Site, TSS) and a red bar (Cleavage and Polyadenylation Site, CPS). Introns are marked with a ^ symbol. Scale on x-axis corresponds to distance in base pairs from the TSS. Scale on y-axis represents fold enrichment of IP over input on a Log_2 scale for ChIP-chip data and fold expression over background for total RNA data. The cartoon beneath the plots illustrates the potential Pol II distributions that would yield the observed ChIP-chip occupancy profiles.

(b) ChIP-chip profiles of genes for which Pol II occupancy profile (blue) is constant across the ORF. Phosphorylation profiles are shown in purple (Ser7-P), red (Ser5-P), and green (Ser2-P) and total RNA is shown in black.

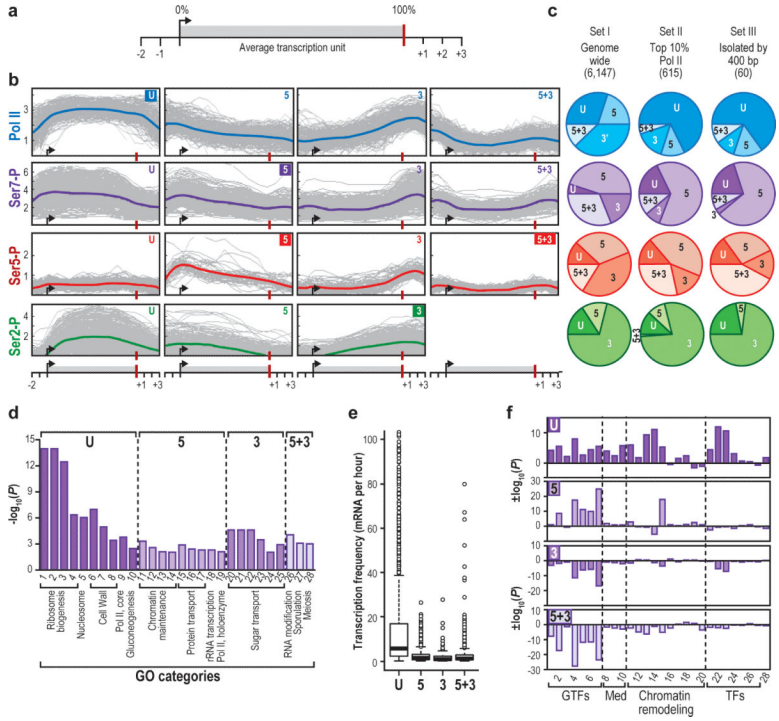

Figure 2. Genome-wide Pol II and CTD Phosphorylation Profiles

(a) The diagram summarizes the scale used in the average transcription unit analysis. Bins within transcribed region are equivalent to 10% of the transcribed region length and bins flanking this region (−1, +1) are a constant 157bp.

(b) Representative clusters from the K-means clustering analysis for Pol II, Ser7-P, Ser5-P, and Ser2-P occupancy profiles across the genome using average transcription unit analysis (see Methods). **U**: uniform enrichment across the transcription unit; **5**: 5’ enrichment; **3**: 3’ enrichment ; **5+3**: or 5’ and 3’ enrichments. The dominant profile(s) for each is highlighted.

(c) Pie chart diagrams displaying the distribution of genes within each gene set for Pol II, Ser7-P, Ser5-P, and Ser2-P having each of the profiles shown in (b). Set I –Annotated protein-coding genes (6147 genes); Set II – pc-genes with an average Pol II enrichment in the top 10% (615 genes); Set III –pc-genes isolated by 400bp and having an average Pol II enrichment greater than or equal to 1 (60 genes).

(d) Histogram of significantly enriched gene ontology (GO) categories within individual Ser7-P clusters (**U, 5, 3, and 5+3**). Details for each GO category are provided in Supplementary Table 1.

(e) Box plots of transcription frequency for genes within each Ser7-P cluster [52]. The uniform (**U**) Ser7-P cluster is strongly enriched for genes with higher rates of transcription frequency compared to the genome average (*P*-value = 3.35×10^{-30}).

(f) Bar graph of transcription machinery at the upstream activating sequence (UAS) within individual Ser7-P clusters. The individual proteins are listed in Supplementary Fig. 6. Proteins that are enriched within the cluster are represented by positive values, while proteins that are depleted are represented by negative values [53]. Dashed lines segregate the components of each complex.

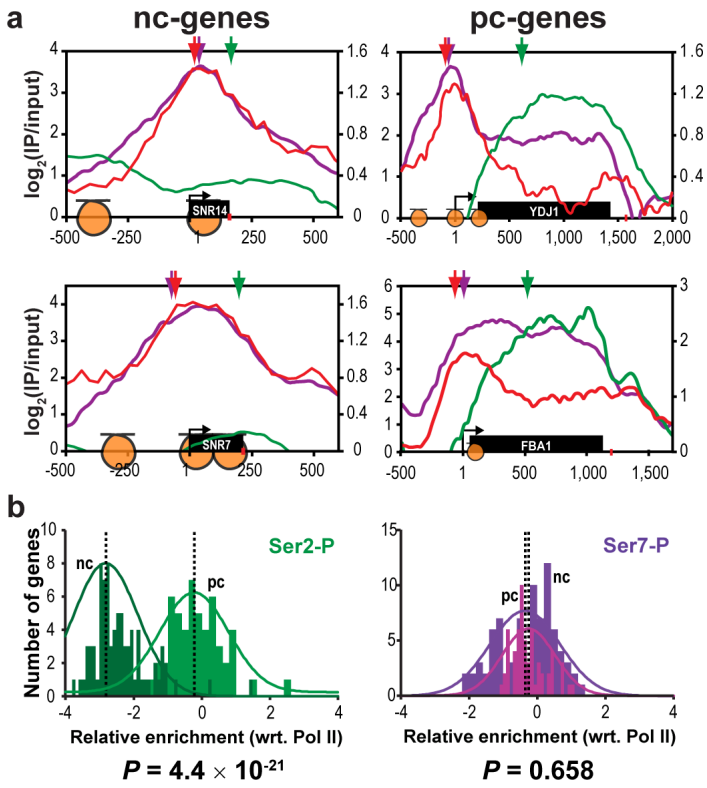

Figure 3. CTD Phosphorylation Profile for Protein-coding and Non-coding Genes

(a) Phosphorylation profiles across several representative nc-genes (left column) and pc-genes (right column). Color-coded arrows indicate the respective phosphorylation peak. The nc-genes display overlapping Ser5-P (red) and Ser7-P (purple) profiles, while varying levels of 3' Ser7-P enrichment is observed in pc-genes. Ser2-P (green) appears to be less abundant in nc-genes. The nucleosome positions near the TSS [60] are displayed as orange circles (drawn to scale).

(b) Histograms displaying the relative enrichment (with respect to Pol II levels) across nc- and pc-genes for Ser7-P and Ser2-P. The Ser2-P levels are significantly lower in nc-genes than pc-genes (*P*-value = 4.4×10^{-21}). No significant difference in Ser7-P was detected between nc-genes and pc-genes (*P*-value = 0.658).

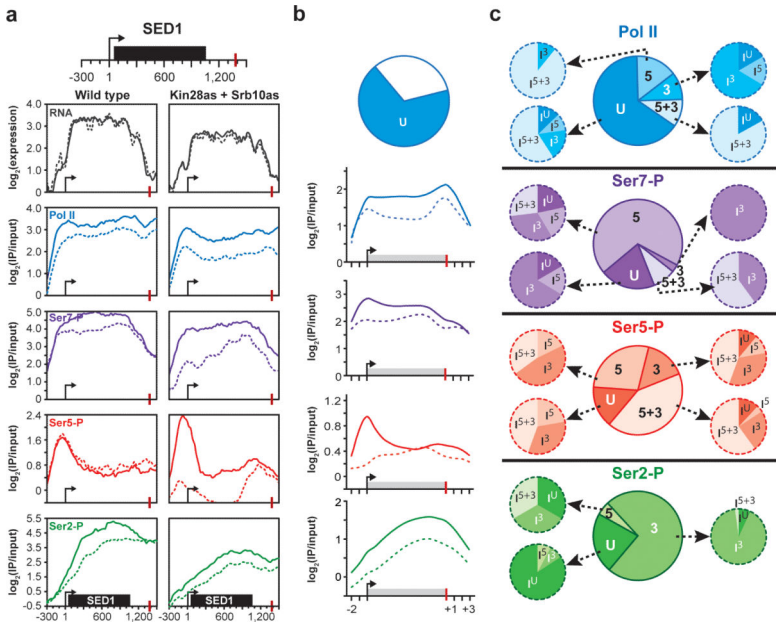

Figure 4. Small Molecule Inhibition of CTD Phosphorylation

(a) Inhibition effects on total RNA (black), Pol II (blue), Ser7-P (purple), Ser5-P (red), and Ser2-P (green) occupancy profiles for SED1. Uninhibited profiles are shown as a solid line and inhibited profiles shown with a dashed line.

(b) Summary of average Kin28as+Srb10as responses to inhibition for Pol II (blue), Ser7-P (purple), Ser5-P (red), and Ser2-P (green). Data displayed for 415 genes with "uniform" Pol II profiles from Set II in the Wild Type strain (see Supplementary Fig. 11, central Pol II Pie). The solid and dashed lines indicate the uninhibited and inhibited Kin28as+Srb10as average profiles, respectively. Scale on the x-axis is shown as distance from the TSS using average transcription unit analysis (described in Fig. 2a). Scale on y-axis represents fold enrichment of IP over input on a Log_2 scale.

(c) Comprehensive diagrams displaying the comparison between the uninhibited wild type strain (large central pies) and the inhibited double mutant strain (smaller pies) for all the genes in Set III (described in Fig. 2c). Summaries shown for Pol II (blue), Ser7-P (purple), Ser5-P (red), and Ser2-P (green). The smaller pie charts correspond to a slice of the wild type chart and show the partitioning of the genes from the double mutant data within that slice after the inhibitor was added. The profiles **U**, **5**, **3**, and **5+3** are as described in Figure 2 and "**I**" is used to indicate the profiles seen after inhibition ($\mathbf{I^U}$, $\mathbf{I^5}$, $\mathbf{I^3}$, and $\mathbf{I^{5+3}}$).

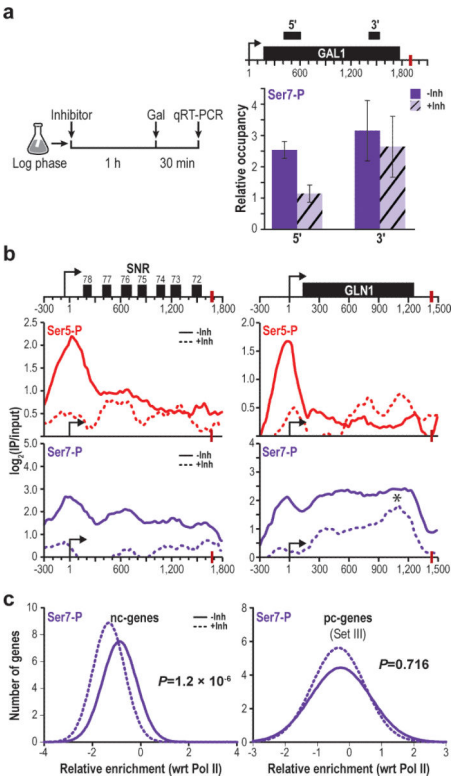

Figure 5. Promoter-Distal Ser7-P Marks are Not Remnants of Promoter-Proximal Kin28 Phosphorylation

(a) Ser7-P marks within the coding region are not remnants of phosphorylation at the promoter. The Kin28as+Srb10as strain was treated for 60 min with DMSO or the inhibitors 1-NA-PP1/1-NM-PP1 under non-inducing conditions. The GAL genes were induced for 30 minutes, and ChIP-qPCR was performed against the 5' and 3' ends of the *GAL1* genes (black bars indicate primer location). Upon inhibition, Ser7-P was significantly reduced at the 5' end but not the 3' end of *GAL1*.

(b) Representative ChIP-chip profiles for Ser5-P (red) and Ser7-P (purple) at a nc-gene polycistronic cluster (SNR78-72) and a pc-gene (GLN1) demonstrating the effects of Kin28 and Srb10 inhibition. The inhibition of Ser5-P is observed predominantly at the 5’ ends of both nc- and pc-genes. At nc-genes, the total level of Ser7-P decreases uniformly across the transcript, while at pc-genes the primary effect of the inhibition is seen at the 5’ end of the transcript. This suggests the presence of a secondary Ser7 kinase acting near the 3’ end of pc-genes (indicated by a *).

(c) Histogram demonstrating the effects of inhibition on the relative Ser7-P enrichment (with respect to Pol II levels) at the 3’ end of the ORF at of nc-Genes (left panel) and pc-Genes (right panel). (nc-genes: $P = 1.2\times10^{-6}$; pc-genes: $P = 0.716$)

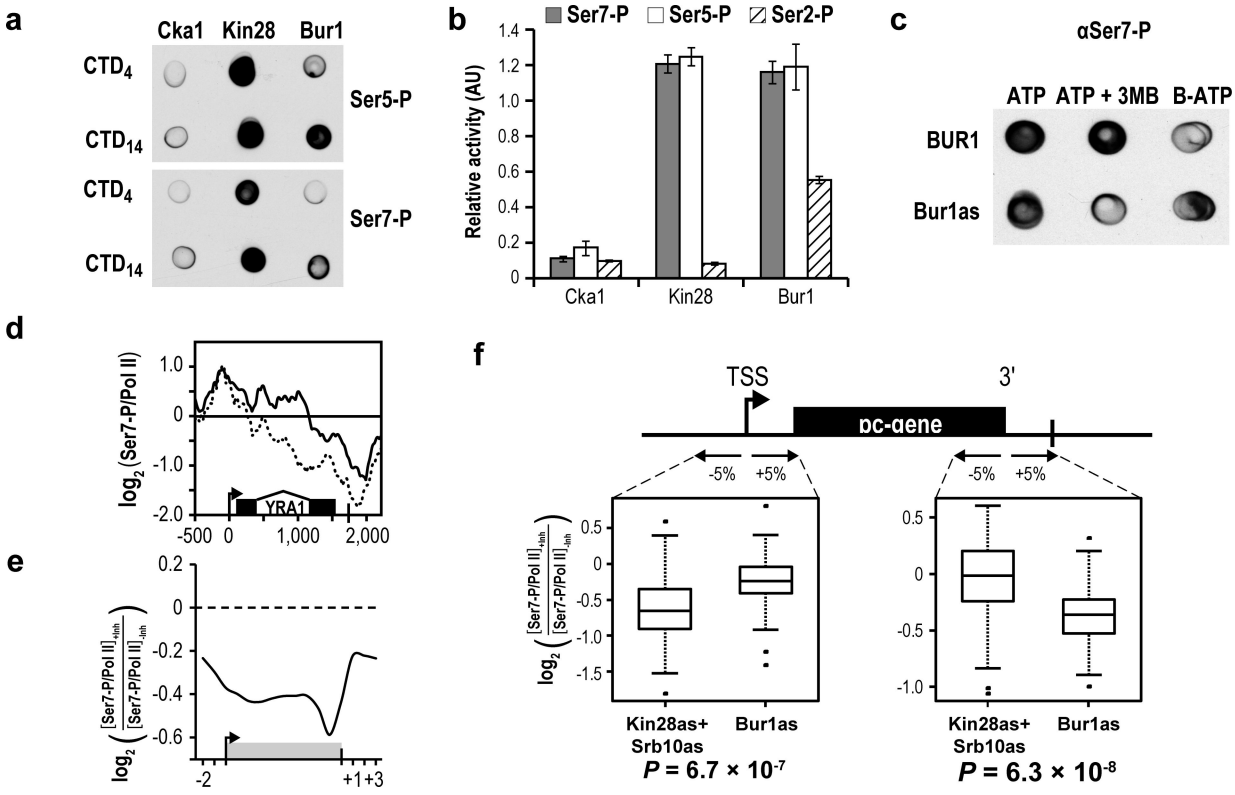

Figure 6. Bur1 Kinase Directly Phosphorylates Ser7

(a) Kinase assay using a GST-CTD_4 and GST-CTD_{14} substrates and purified Cka1, Kin28, and Bur1 kinases. The blots were probed with antibodies against Ser5-P (top) and Ser7-P (bottom).

(b) ELISA of GST-CTD_4 phosphorylated by purified yeast kinases and probed with antibodies against Ser7-P (purple), Ser5-P (red), and Ser2-P (green).

(c) Dot blot of GST-CTD_{14} phosphorylated by Bur1 and the analog-sensitive mutant Bur1as (Bur1-L149G) probed with anti-Ser7-P antibodies.

(d) Ser7-P enrichment profile (normalized to Pol II) across the YRA1 gene showing the effects of inhibition of either Kin28as+Srb10as or Bur1as.

(e) Average Pol II (blue) and Ser7-P (purple) from uninhibited (solid line) and inhibited (dashed lines) Bur1as cells at 415 genes with uniform and robust Pol II occupancy from Set II (top panels). Pol II normalized fold change in Ser7-P across the same subset of genes (bottom panel) is shown to highlight the role of Bur1 in Ser7 phosphorylation across the transcribed unit.

(f) Distribution of average Pol II normalized fold changes in Ser7-P within a 10% window centered around the TSS and the 3' ORF boundary of Set III genes. Distributions from Kin28as+Srb10as and Bur1as cells were compared using a two-sided student's T-test.