# Regulation of RNA Synthesis and Decay via the C-terminal Domain of Pol II

By

Sandra C. Tseng

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

(Biochemistry)

at the

UNIVERSITY OF WISCONSIN-MADISON

2019

Date of final oral examination: 12/19/2018

The dissertation is approved by the following members of the Final Oral Committee:

Aseem Z. Ansari, Professor, Biochemistry
David A. Brow, Professor, Biomolecular Chemistry
Elizabeth A. Craig, Professor, Biochemistry
Aaron A. Hoskins, Associate Professor, Biochemistry
Rupa Sridharan, Assistant Professor, Cell and Regenerative Biology


*For being my friends.*

## Table of Contents

# List of Tables and Figures

**Chapter 1: Introduction**



**Chapter 2: A Rational Approach to Targeted Inactivation of the TFIIH-Kinase Clarifies Its Regulatory Roles in Transcription**

## Acknowledgments

First, I would like to acknowledge the following funding sources for supporting my research during various periods of my tenure in graduate school: NSF pre-doctoral fellowship (DGE-1256259), NIH National Research Service Award T32 (GM007215), Biochemistry WARF Research Assistant Award, and the University of Wisconsin-Madison Integrated Program in Biochemistry. Next, I would like to thank my Thesis Committee for each instance they have graciously taken time out of their day to review my work (including this document) and provide their invaluable scientific feedback. Finally, I would like to thank the members of the Ansari Lab, past and present, for all their contributions, scientific and otherwise, to my graduate school experience. Specifically, I thank Aseem Ansari for teaching me how research is conducted at the highest level and his intellectual input on my projects; he stands alone in always appreciating and understanding the minutiae, as well as the broader context, of my research.

## Abstract

RNA expression can be thought of as a sum of the products of RNA synthesis and decay: on the one hand, cells are transcribing the genes required for a given situation, while on the other, RNA degradation machinery is constantly acting on normal as well as aberrantly synthesized RNA. The kinases which phosphorylate the C-terminal domain (CTD) of RNA Polymerase II (Pol II) are canonically associated with being on the additive scale of the balancing act between RNA synthesis and decay. One such kinase, and the focus of my thesis, Kin28/CDK7 activates and promotes transcription by phosphorylating serine residues on the CTD. I begin my thesis with an introduction into our current understanding of how transcription is regulated via post-translational modification of the CTD, and also present the open questions, which this thesis seeks to address, on the role of Kin28/CDK7 in mediating gene expression (Chapter 1).

In Chapter 2, I describe the development of a chemical genetic approach to covalently inhibit Kin28 in budding yeast. The strategy and lessons learned therein can be generally applied to covalently inhibit kinases in other model organisms. Combined with transcriptomic sequencing, I show that Kin28 is required for synthesis of nearly all annotated yeast mRNAs, and demonstrate that previously conflicting views on the matter can be explained by the phenomenon known as “RNA buffering,” which is the stabilization of the steady-state level of RNAs during insults to transcription or decay. Also discussed in this chapter is our novel discovery that Kin28 plays a role in advancing Pol II from the initiation into the elongation phase of transcription.

In Chapter 3, I investigate determinants of RNA buffering in a Kin28-inhibited regime, and reveal that the relative binding of the RNA binding proteins Nab2 and Ski2 can predict mRNA stability. Furthermore, I show inhibition of Kin28 rapidly induces the formation of P bodies and perturbs translation. From my analyses, I present a compendium of the mRNAs that are unstable, buffered, translated, and or poorly

translated in situations of transcriptional crisis, a state I show is distinct from that of the canonical stress response.

In Chapter 4, I study RNA buffering from the "other side" of the gene expression equation by exploring the combined effect of deleting a component of the RNA decay machinery (*rai1::URA3*) and inhibition of Kin28. Despite being on opposite sides of the scale of gene expression, both defects in decay (*rai1::URA3*) and synthesis (inhibition of Kin28) result in similar outcomes, such as the formation of P bodies and defects to translation, suggesting translation as a key regulatory node for the crosstalk between RNA synthesis and decay.

Finally, in Chapter 5, I comment on the potential of Kin28/CDK7-based therapies for the treatment of disease, and how studies in yeast can inform on the mechanism of action of such therapies. Appendix I describes my efforts into characterizing the role of CTD methylation on gene expression. In sum, this work has significantly contributed to our understanding of the myriad ways Kin28/CDK7 functions in regulating gene expression beyond its canonical roles in promoting RNA synthesis.

# Chapter 1: Introduction

## 1.1 Introduction

Gene expression, *i.e.* the manifestation of the information encoded in the genome, and its precise regulation through development and in response to environmental stimuli fundamentally dictate all aspects of an organism's biology. The control of a gene's expression is complex, beginning with *cis* regulatory elements encoded in the very nucleic acid bases of the gene, and involving multiple layers of regulation thereafter that modulate the levels of a gene's products, RNA and protein. This chapter reviews our current understanding of how eukaryotic gene expression is regulated via the C-terminal domain (CTD) of RNA Polymerase II, with an emphasis on the role of the essential CTD kinase Kin28/CDK7. Kin28 and CDK7 are budding yeast and human homologs, respectively; this delineation in yeast/human protein names will be used hereafter.

## 1.2 The CTD Code of RNA Polymerase II

The core catalytic subunit of DNA-dependent RNA polymerases is extremely conserved from bacteria to man; while prokaryotes utilize one polymerase to perform all transcription, eukaryotes contain multiple RNA polymerase genes with non-overlapping functions [1]. Rpb1/POLR2A, the largest subunit of RNA polymerase II (Pol II), specializes in transcription of mRNA. Pol II, whilst still in possession of a nearly identical core catalytic domain, differs substantially from the other eukaryotic Pol I and III by an unstructured C-terminal domain (CTD). The CTD contains a highly repetitive amino acid sequence motif '$Y_1S_2P_3T_4S_5P_6S_7$' in tandem 26 times in budding yeast and 52 times in mammals (**Figure 1.1**). Each residue of the heptad will be referred to hereafter by its three-letter abbreviation and position in the heptad, *e.g.* the fifth serine as Ser5.

Since its discovery [2, 3], the unique sequence of the CTD was recognized as a potential node of transcriptional regulation because five of the seven residues in the

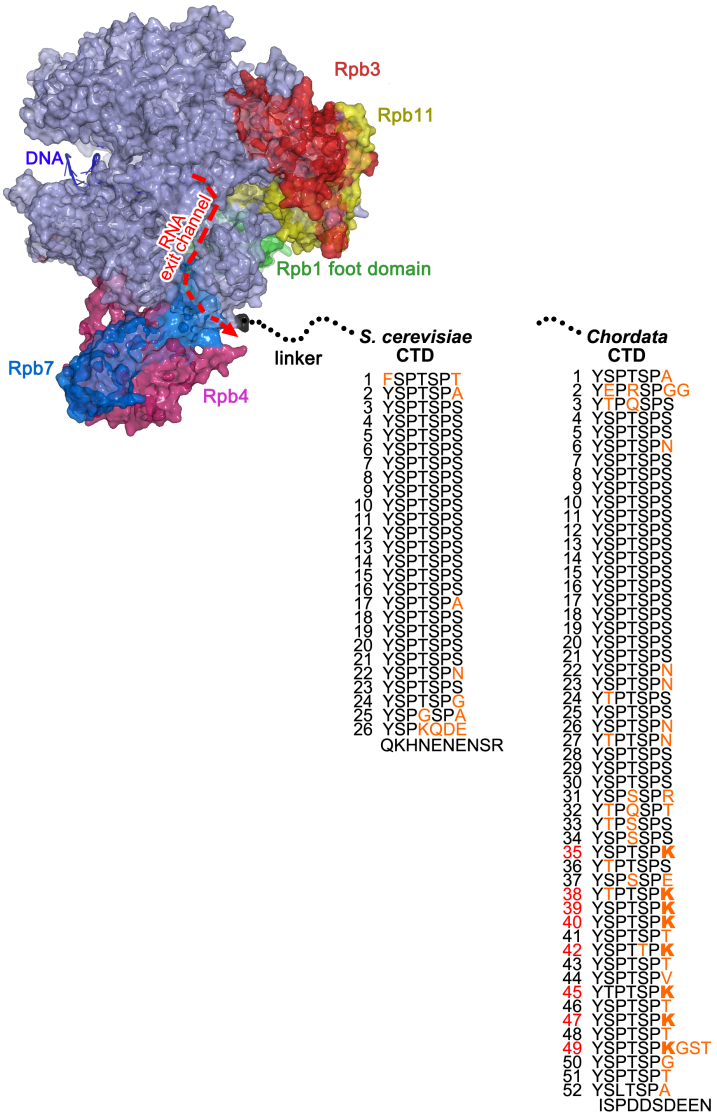



**Figure 1.1 The CTD of Pol II Has Expanded and Diversified from the 'YSPTSPS' Motif Conserved Among Eukaryotes**

Deviations from consensus are highlighted in orange, and chordate heptads which contain the S7K polymorphism are highlighted in red. The CTD of Rpb1 extends from an unstructured linker, which together with the CTD can theoretically—though unlikely—extend 10X the diameter of Rpb1. Like the CTD, Pol II subunits are known to interact with RNA processing and other transcription-coupled proteins. The exit tunnel from which nascent RNA emerges is also indicated [4]. Pol II graphic was generated from PDB# 1Y1W [5].

repeating heptads can be phosphorylated and, being in close proximity to the catalytic center of Pol II, is optimally placed for action [6]. Indeed, the CTD is extensively, reversibly phosphorylated and further, the phosphorylation state of the CTD is critical for activation of transcription in vivo through interaction with histone chaperones and methyltransferases, and numerous other complexes [7-9]. Based on the difference in mobility of hypo- and hyperphosphorylated Pol II in polyacrylamide gels, it is apparent that Pol II exists predominately at these extremes in vivo and less so at an indiscrete number of intermediate phosphorylated states [10, 11]. The exact distribution and patterns of phosphorylation along the CTD have been difficult to ascertain because current biochemical methods (*e.g.* mass spectrometry or antibody-based methods) typically used to characterize post-translational modifications are unable to differentiate one heptad from another. Using cells expressing CTD with sequences modified to allow for differentiation by mass spectrometry, two groups independently demonstrate that phosphorylation of Ser2 and Ser5 are the most abundant modifications to the CTD [12, 13].

The combinations of post-translational modifications of the CTD are often referred to as the “CTD code” [14] because each modification or set of modifications is known to correspond to various stages of RNA synthesis and processing (**Figure 1.2**). Through regulation of CTD modifying enzymes, such as CTD kinases and phosphatases, or “writers” and “erasers” of the code, gene expression is controlled. “Readers” of the code can be thought of as the proteins that bind to the CTD, and they do so selectively, their binding affinity enhanced or reduced by the combination of modifications on the one or multiple CTD repeats they bind. Once associated with the CTD, reader proteins can interact with chromatin, chromatin-bound factors, and, in the case of elongating Pol II, with nascent RNA. Further expanding the code are the prolines

within the CTD that can isomerize, which has been demonstrated to affect the binding affinity of CTD interacting proteins such as the Ser5 and 7 phosphatase Ssu72 [15] and the short RNA termination factor Nrd1 [16]. In addition to phosphorylation, the serine and threonine residues may undergo O-linked glycosylation [17], occurring predominantly before active transcription occurs, and mutually exclusive to Ser2 and 5 phosphorylation.

Historically, phosphorylation of Ser2 and 5 are the most studied CTD marks as the first CTD antibodies were raised against epitopes containing phosphorylated Ser2 and 5 [18], and the first CTD modifying enzymes discovered were the Ser2 and 5 kinases, Ctk1 [19] and Kin28/CDK7 [20-22]. Soon after, the primary Ser2 kinase Bur1/CDK9 was also characterized [23]. However, the CTD code continues to be updated with the discovery of new writers, erasers, readers, modifications, and new functions for previous discoveries [24]. Contrasting the clear picture presented in **Figure 1.2**, recent developments reveal that multiple enzymes can act on the same CTD substrate in a gene- or environment-specific manner. For example, instead of CDK7, ERK1/2 has been shown to specifically phosphorylate Ser5 at the promoter of developmental regulator genes in embryonic stem cells [25].

Furthermore, a vertebrate-specific CTD code is becoming apparent. Thr4 is dispensable in yeast, but essential in chicken DT40 cells and human and mouse cells [26, 27]. (Though mutation of Thr4 in the yeast CTD is tolerated, Thr4 mutants display a severe growth defect, and the phosphorylation of Thr4 enhances recruitment of transcription termination machinery in yeast [28].) The phosphorylation of Ser7 recruits Integrator, a small non-coding RNA (snRNA) processing complex found in mammals but without an ortholog in yeast [29, 30]. In addition to differences in usage of the canonical CTD, mammals contain additional repeats of non-consensus CTD, which differ predominately at the 7th position of the heptad (**Figure 1.1**). These non-consensus residues are also modified and functionally relevant to transcriptional processes.

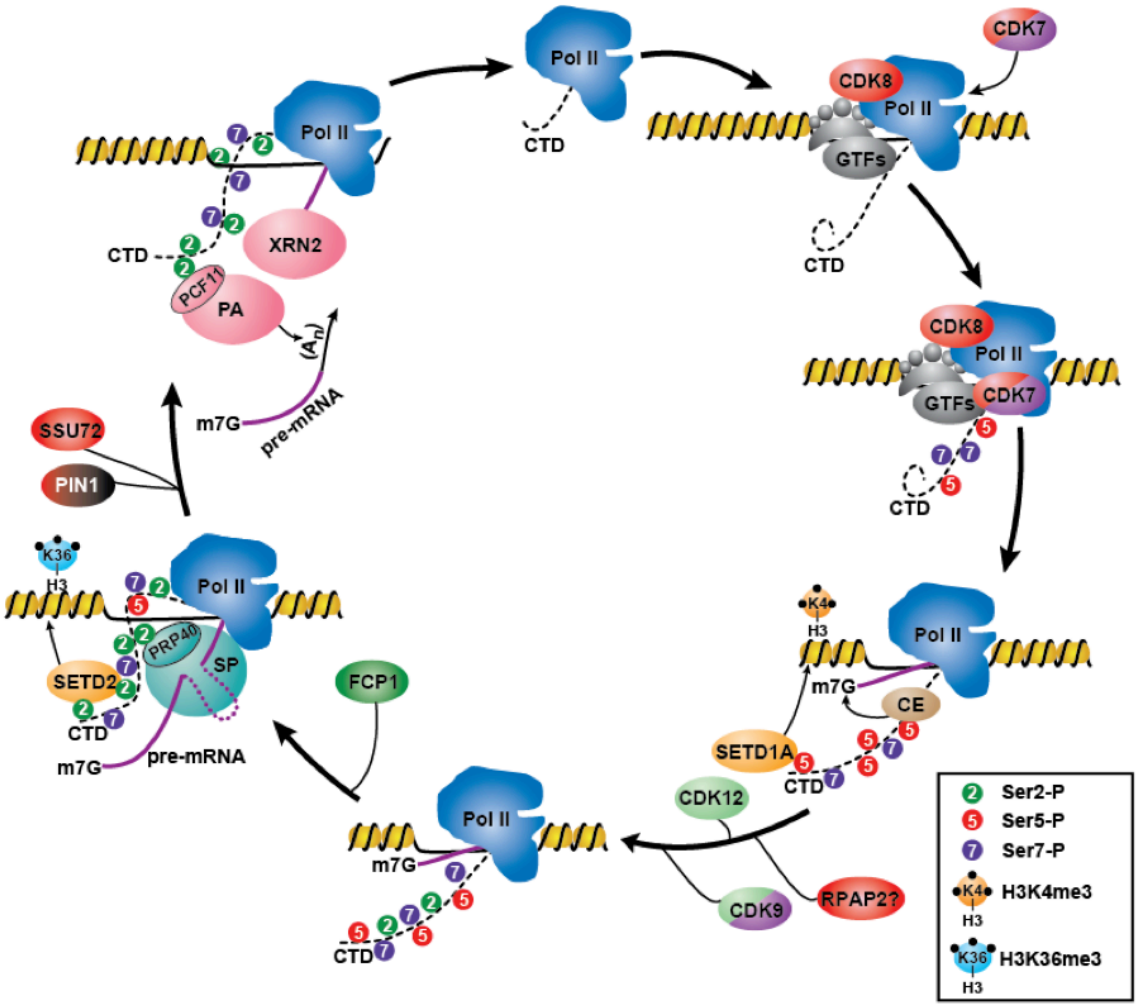



**Figure 1.2 The Phosphorylation of the CTD Coordinates RNA Synthesis and Processing**

At the beginning of the transcription cycle, hypo-phosphorylated Pol II is recruited to promoter elements by general transcription factors (GTFs). CDK7 and 8 phosphorylate Ser5 and 7 [31]; Ser5 phosphorylation is read by mRNA capping enzymes (CE) and histone H3K4 methyltransferase SetD1A [32] and in turn, H3K4me3 recruits CDK9 and 12, initiating promoter release and active elongation with phosphorylation of Ser2 [33]. With increasing Ser2p, Ser5p and Ser7p, phosphatase RPAP2 begins removal of these marks [34]. Ser2p interacts with splicing proteins (SP) and termination factors such as PRP40 and PCF11 of the polyadenylation complex (PA) [35]. SETD2, also recruited by Ser2p, methylates H3K36, recruiting histone deacetylases (not shown), which “close” chromatin in the wake of transcribing Pol II [36]. Exonuclease XRN2 promotes termination of transcription and release of Pol II from chromatin [37, 38]. FCP1 erases Ser2p [39], and SSU72 removes remaining Ser5p and Ser7p; prolylisomerase PIN1 catalyzes dephosphorylation of Ser5p by converting Sp-Pro bonds to the cis conformation, which is a higher affinity substrate for SSU72 [40]. Figure adapted from [9].

Notably, a CTD arginine (R1810) is methylated [41], and the lysines are methylated [42], acetylated [43], and ubiquitinated [44].

**Table 1.1** summarizes the CTD code deciphered to date, and **Figure 1.3** presents profiles for CTD modifications over a typical transcribed gene, noting differences between budding yeast and human cells.

### 1.3 Regulation of Transcription by Kin28/CDK7

The predominant Ser5 and 7 kinase, Kin28 in budding yeast and CDK7 in humans, is required for viability in both organisms because it serves a range of essential biological functions which, for the most part, are conserved. The fission yeast (*S. pombe*) homolog Mcs6 is also required for viability and exhibits similar functions as Kin28/CDK7 [45]. Deviations between species will be noted when appropriate in the following discussion of the kinase's manifold functions in regulating transcription.

Kin28/CDK7 was first identified as part of the ten-subunit TFIIH general transcription factor complex [31, 46, 47], which as a whole has been implicated in DNA damage repair [48, 49], rRNA transcription [50], and cell cycle control [51, 52] in addition to its roles in regulating Pol II. Interestingly, three of the ten subunits (Kin28/CDK7, Tfb3/MAT1, Ccl1/CCNH) can reversibly disassociate from the rest of the TFIIH complex [53, 54]. This trimeric complex is called the cyclin-dependent kinase (CDK) Activating Kinase (CAK) complex because in metazoans and *S. pombe*, it does indeed activate CDKs during specific stages of the cell cycle [55]. CDK7 and Mcs6 regulate the activity of CDKs through phosphorylation of a threonine in a conserved CDK domain known as the T-loop, which induces a conformational switch of the CDK into an active state [56]. However, a different kinase Cak1, rather than Kin28, serves this function in budding yeast [57, 58].

**Table 1.1 Summary of the Known Functions of CTD Post-translational Modifications**

| Modification | Known Function | References |
|---|---|---|
| **Ser2p** | promotes transcription elongation, histone remodeling/modification, proper termination, cleavage and polyadenylation | [59-64] |
| **Ser5p** | promotes transcription initiation and entry into elongation, recruits capping enzymes/chromatin remodelers and modifiers, splicing, termination | [7, 65-74] |
| **Ser7p** | snRNA expression, recruits Integrator complex* | [29, 30] |
| **Tyr1p** | stabilizes Rpb1, prevents antisense transcription and prevents pre-mature termination | [75-77] |
| **Thr4p** | transcription termination, chromatin remodeling, splicing*, histone mRNA processing* | [26-28, 78, 79] |
| **Arg1810me** | sn(o)RNA expression* | [41] |
| **Lys7ac** | transcription initiation of growth factor genes* | [43] |
| **Lys7me** | transcription initiation*, maintenance of gene expression levels* | [42, 80] |
| **Lys7ub** | degrades Rpb1* | [44, 81] |
| **Ser *O*-GlcNAc** | promotes transcription initiation* | [17, 44, 82] |

* in human cells only

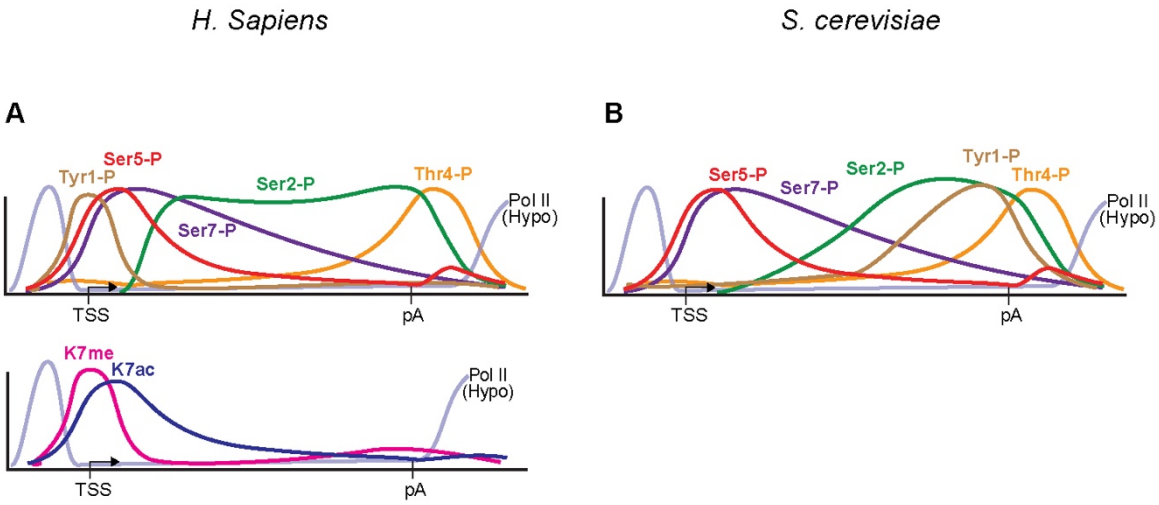



**Figure 1.3 Average Profiles of CTD Modifications across a Gene**

Pol II begins transcription at the transcription start site (TSS), unphosphorylated (Hypo). **(A)** The human CTD (*H. sapiens*), unlike the **(B)** budding yeast CTD (*S. cerevisiae*), contains lysines (K7) which are methylated (me) and acetylated (ac) early in the transcription cycle. The serine (Ser), tyrosine (Tyr), and threonine (Thr) residues of the CTD heptad are phosphorylated (-P) at distinct stages of transcription. Past the polyadenylation site (pA), the CTD returns to its hypo-phosphorylated state by the action of phosphatases throughout the transcription cycle.

Shortly after its discovery, Kin28/CDK7 was shown to phosphorylate the CTD [21, 31], and was surmised to regulate transcription by this activity. Recent structural studies have provided finely resolved insight into how Kin28/CDK7 is situated as part of the transcriptional pre-initiation complex (PIC), which has added to our understanding of its functions at gene promoters and agrees with previous biochemical studies [83, 84]. As shown in **Figure 1.4**, TFIIH makes contacts with both Pol II, the Mediator complex, and promoter DNA. The Ssl2/XPB translocase subunit of TFIIH has been shown to "reel in" DNA, causing melting of the DNA duplex, and thus opening of the promoter for transcription by Pol II [85, 86]. Phosphorylation of the CTD by Kin28/CDK7 further promotes transcription initiation [73, 74] by disrupting high-affinity interactions between Pol II and the Mediator that prevent Pol II escape from the promoter [72, 87-89].

In metazoans and *S. pombe*, but absent in budding yeast, Pol II pauses after promoter escape in a phenomenon known as promoter-proximal pausing [90]. Inhibition of CDK7 was recently shown to further increase Pol II pausing [91]. Remarkably we show Pol II accumulation near the +2 nucleosome, reminiscent of promoter-proximal pausing, in budding yeast when Kin28 is inhibited [67] (discussed further in Chapter 3 of this thesis), which highlights the crucial role of Kin28/CDK7 in promoting Pol II clearance from the promoter into active elongation. Promoter-proximal pausing is thought to be partially regulated by Pol II interactions with the DSIF (DRB sensitivity inducing factor) complex, homologous to budding yeast Spt4 and 5; phosphorylation of DSIF by CDK7 in metazoans releases Pol II from the promoter-proximal pause [92]. Spt5 phosphorylation in yeast by Kin28 is known only to occur *in vitro*, and is thought rather to be a substrate for the Ser2 kinase Bur1 [93]. Promoter-proximal pausing is also thought to be established by the NELF (negative elongation factor) complex, absent in budding yeast, which is disassociated from Pol II by phosphorylation by the Ser2 kinase CDK9 (Bur1 in budding yeast) [94]. However, because Kin28/CDK7 stimulates Bur1/CDK9 activity, it is

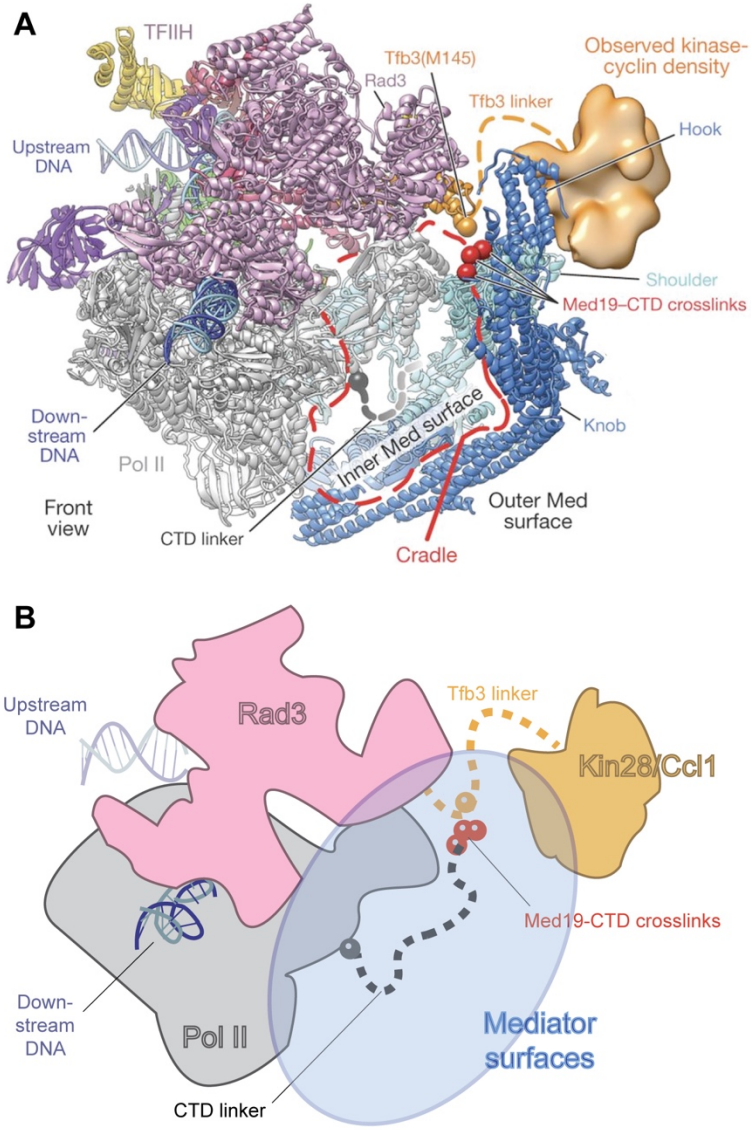



**Figure 1.4 Schematic of Mediator Contacts with Components of the PIC**

Figure **(A)** is adapted from [84]; it presents a view of the cryo-electron microscopy structure of Mediator in complex with the PIC containing TFIIH. Regions of the TFIIH subunits Rad3, Tfb3, and Kin28/Ccl1 are shown. **(B)** A simplified diagram of (A) illustrates Kin28 and the associated cyclin Ccl1 are positioned to disrupt contacts between Mediator and Pol II through phosphorylation of the CTD, which would accumulate repulsive negative charge in the 'Cradle' region.

difficult to disentangle Kin28/CDK7's from Bur1/CDK9's role in promoting some aspects of transcription [93].

Apart from its functions in the early stages of transcription, Kin28/CDK7-mediated phosphorylation is also required for efficient RNA processing and transcription termination. Capping enzymes, which catalyze the addition of 7-methylguanosine caps on the 5' ends of nascent mRNA, are recruited to nascent mRNA by binding to CTD with phosphorylated Ser5 and 7 residues [66, 95, 96]. In metazoan cells, Ser5 phosphorylation has been shown to affect alternative splicing [70], which may be caused by altered Pol II elongation rates, which are linked to splicing efficiency [97, 98]. Transcription termination is thought to be regulated by Kin28/CDK7 because the affinity of termination machinery for Pol II is highly dependent on the phosphorylation state of the CTD [28, 78]. For example, the Nrd1-Nab3-Sen1 complex in yeast is required for efficient termination of short RNAs, including a class of noncoding RNAs known as small nuclear and small nucleolar RNA (snRNA and snoRNA) [99, 100]. This complex binds CTD phosphorylated on Ser5 residues [65], and increased phospho-Ser5 levels in attenuated Ser5 phosphatase mutant yeast cause readthrough defects in RNA, *i.e.* 3' ends extend past canonical termination sites [101-103]. Additionally, Pta1, a factor of the cleavage and polyadenylation factor complex, which processes the 3' ends of mRNAs, also interacts with Kin28-phosphorylated CTD [104].

In sum, the TFIIH kinase Kin28/CDK7 regulates transcription from its start, middle, and end. **Figure 1.5** summarizes these modes of regulation.

### 1.4 Merging Conflicting Views from Kin28/CDK7 Loss-of-function Studies

Given the large body of work which demonstrates Kin28/CDK7's crucial roles in regulating transcription, it may come as a surprise that conflicting views exist on whether Kin28/CDK7 is dispensable for RNA synthesis. Early studies of an *in vitro* reconstituted

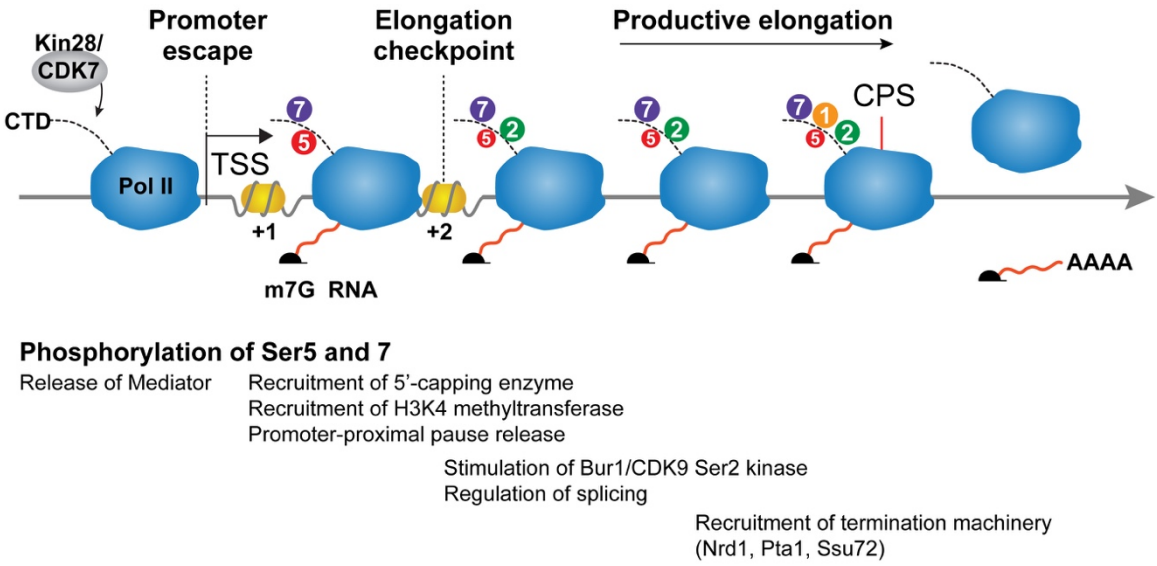



**Figure 1.5 Mechanisms of Transcriptional Regulation Mediated by Kin28/CDK7**

Adapted from [67]; Kin28/CDK7 phosphorylates Ser5 and 7 of the CTD, triggering a series of events which regulate multiple stages of transcription. Below the diagram, these events are listed horizontally offset by their relative occurrence in the transcription cycle. TSS denotes the transcription start site, and CPS indicates the cleavage and polyadenylation site.

transcription reaction showed that purified extracts of Pol II, without the CTD, in the absence of Kin28/CDK7, could transcribe naked DNA templates as efficiently as intact Pol II [105]; Kin28/CDK7 activity was shown to be nonessential for transcription *in vitro*, but was shown to combat the effects of an unidentified negative elongation factor in less pure extracts [106]. These findings conflicted with subsequent *in vitro* studies, which demonstrated a dependence on CTD and Kin28/CDK7 activity for tested genes [107].

Later *in vivo* work provided two examples (copper-inducible *CUP1* and *HSP82* genes) of genes that did not require the CTD, Kin28, and Mediator subunits [108]. In this study, yeast expressing Pol II lacking the CTD could transcribe those genes normally and without binding of Kin28 and Mediator components at the gene loci. The CTD-less Pol II was the only copy of Pol II expressed, while endogenous Pol II was subject to degradation by introduction of a temperature-sensitive allele (*ts)*. Though Pol II was able to transcribe *CUP1* at a high level without Kin28, the *CUP1* mRNAs that were produced utilized alternative polyA sites and were more rapidly degraded. Thus, it was concluded that Kin28 was non-essential, *not* a global transcriptional activator, and that its main function lay in recruitment of 5'-capping enzymes to protect nascent mRNAs from degradation.

However, a subsequent transcriptome-wide study that also utilized temperature-sensitive mutant yeast strains—including *kin28ts*—revealed that nearly all mRNAs were reduced in cells devoid of Kin28 at the restrictive temperature, akin to cells devoid of the Rpb1 enzyme itself [109]. Thus, contrary to several aforementioned reports, it was postulated that Kin28/CDK7 *is* globally required for transcription. Still, a subsequent report emerged arguing against the point; in this study, a kinase-attenuated strain of yeast carrying a point mutation of *kin28* showed reduced levels of Ser5 phosphorylation and capping enzyme (Ceg1) occupancy at a few genes, but had unaffected levels of Pol II occupancy and levels of mRNAs at the same tested genes [66].

A previous member in our lab hypothesized that these contrasting views might arise from differences in the loss-of-function strategy employed to study Kin28. Her studies revealed that the *ts* mutants of Kin28 impacted processes separate from the kinase activity of Kin28 [110]. Specifically, at the restrictive temperature used to inactivate Kin28, other subunits of the TFIIH complex were disassociated. The devastating consequences observed transcriptome-wide in *kin28ts* at restrictive temperatures were likely caused by inactivation of the other members of the TFIIH complex, and perhaps the PIC as a whole, rather than solely as a result of the loss of Kin28 activity. Indeed, by utilizing a chemical-genetic approach to specifically inhibit Kin28 with a non-hydrolysable competitive ATP analog (1-NA-PP1), which did not disrupt its interactions with other proteins, it was shown that the gene expression of only a small fraction of the transcriptome was affected by loss-of-function of Kin28 [110]. **Figure 1.6** highlights the key features of the chemical-genetic approach.

Doubt still persisted as to the dispensability of Kin28/CDK7 for transcription, despite the findings from this study. For example, another study utilizing the same approach introduced in [110] came to an opposing conclusion with regards to the effect of inhibiting Kin28 on the transcriptome, namely most mRNAs were reduced in expression [111]; differences in differential gene expression normalization methods were cited as the reason for the discrepancy. Furthermore, communications with other researchers in the community (unpublished) indicated concern for the potential for "leaky" activity in the chemical-genetic approach utilized in [110, 111]. It was noted that inhibition of Kin28 with 1-NA-PP1 was reversible, and furthermore, slowed but did not completely kill cells, as might be expected for inactivation of an essential kinase.

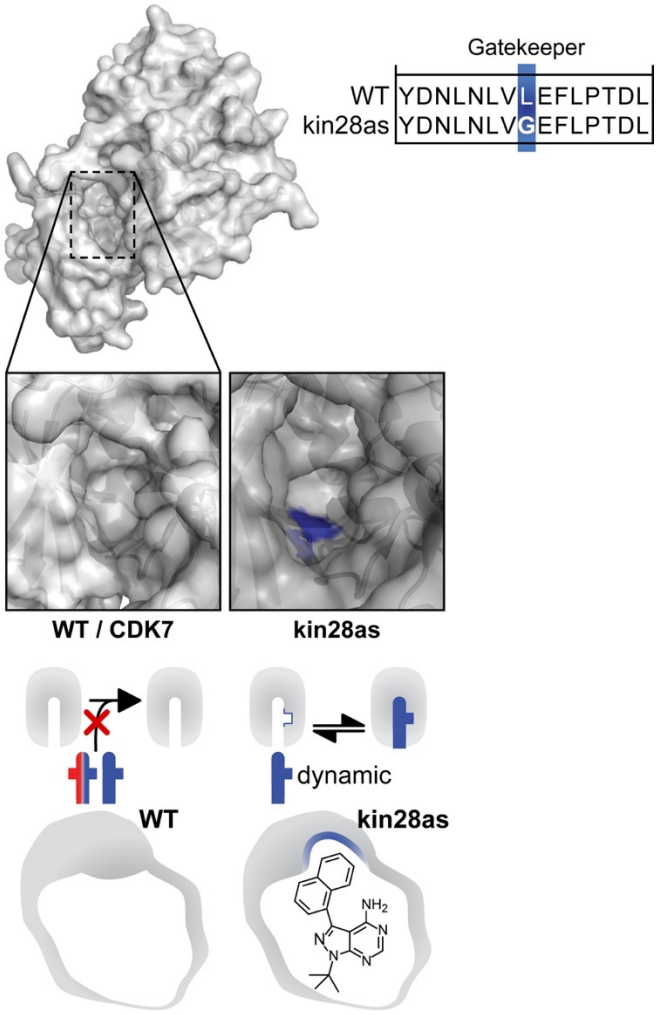



**Figure 1.6 Chemical-genetic Approach to Inhibit Kin28/CDK7**

The structure of CDK7 (PDB# 1UA2 [112]) is shown alongside a zoomed-in view of the ATP-binding pocket of wild-type (WT) and mutant Kin28 (kin28as). WT and kin28as differ only at one position: the "Gatekeeper" residue (L83, blue) is substituted for a glycine in kin28as, as shown at the right in the sequence alignment of WT and kin28as. The smaller Gatekeeper residue allows for binding of, and thus inhibition by, ATP analogs containing bulky substituents. Below, a schematic of one such inhibitor (1-NA-PP1) is docked into the mutated ATP-binding pocket of kin28as.

### 1.5 Thesis Overview

This thesis clarifies the role of Kin28/CDK7 kinase activity in transcription initiation and gene expression. In Chapter 2, I describe improvements to the chemical-genetic strategy of inhibiting Kin28 and more accurate methods of quantifying differential gene expression, which allowed us to obtain a sharper picture of Kin28 functions. Though I hesitate to say I have dispelled all controversy on the subject, I show Kin28 activity is globally required for RNA synthesis, and that discrepancies between studies may have arisen from not only differences in methods of attenuating Kin28 function, but also in differences of differential gene expression analysis. Specifically, I show Kin28's role in RNA synthesis has been overlooked in studies which only capture the steady-state level of RNA, which is prone to stabilize in situations of defective transcription and decay in a phenomenon known as RNA buffering.

Though the feedback between RNA synthesis and decay has been independently observed in several instances, the mechanism by which these two processes communicate is less established. Translation is a potential cellular meeting point at which signals imprinted on mRNA during RNA synthesis (*i.e.* transcription) can converge with RNA decay machinery. In Chapters 3 and 4, I investigate the crosstalk between RNA synthesis, decay, and translation via inhibition of the essential CTD kinase Kin28/CDK7.

This work reveals fundamental biology of a highly conserved, central kinase. Importantly, I have determined mechanisms by which Kin28 regulates transcription genome-wide, and the vast datasets collected can be applied to cancer cell biology,

where the fine-tuning of transcription is of high interest. This potential will be described in Chapter 5.

## 1.6 References


1. Chapman, R.D., et al., *Molecular evolution of the RNA polymerase II CTD.* Trends Genet, 2008. **24**(6): p. 289-96.

2. Allison, L.A., et al., *Extensive homology among the largest subunits of eukaryotic and prokaryotic RNA polymerases.* Cell, 1985. **42**(2): p. 599-610.

3. Corden, J.L., et al., *A unique structure at the carboxyl terminus of the largest subunit of eukaryotic RNA polymerase II.* Proc Natl Acad Sci U S A, 1985. **82**(23): p. 7934-8.

4. Chen, C.Y., et al., *Mapping RNA exit channel on transcribing RNA polymerase II by FRET analysis.* Proc Natl Acad Sci U S A, 2009. **106**(1): p. 127-32.

5. Kettenberger, H., K.J. Armache, and P. Cramer, *Complete RNA polymerase II elongation complex structure and its interactions with NTP and TFIIS.* Mol Cell, 2004. **16**(6): p. 955-65.

6. Meinhart, A., et al., *A structural perspective of CTD function.* Genes Dev, 2005. **19**(12): p. 1401-15.

7. Ng, H.H., et al., *Targeted recruitment of Set1 histone methylase by elongating Pol II provides a localized mark and memory of recent transcriptional activity.* Mol Cell, 2003. **11**(3): p. 709-19.

8. Phatnani, H.P. and A.L. Greenleaf, *Phosphorylation and functions of the RNA polymerase II CTD.* Genes Dev, 2006. **20**(21): p. 2922-36.

9. Zhang, D.W., et al., *Emerging Views on the CTD Code.* Genet Res Int, 2012. **2012**: p. 347214.

10. Dahmus, M.E., *Reversible phosphorylation of the C-terminal domain of RNA polymerase II.* J Biol Chem, 1996. **271**(32): p. 19009-12.

11. Morris, D.P., et al., *Assaying CTD kinases in vitro and phosphorylation-modulated properties of RNA polymerase II in vivo.* Methods, 1997. **12**(3): p. 264-75.

12. Schuller, R., et al., *Heptad-Specific Phosphorylation of RNA Polymerase II CTD.* Mol Cell, 2016. **61**(2): p. 305-14.

13. Suh, H., et al., *Direct Analysis of Phosphorylation Sites on the Rpb1 C-Terminal Domain of RNA Polymerase II.* Mol Cell, 2016. **61**(2): p. 297-304.

14. Buratowski, S., *The CTD code.* Nat Struct Biol, 2003. **10**(9): p. 679-80.

15. Werner-Allen, J.W., et al., *cis-Proline-mediated Ser(P)5 dephosphorylation by the RNA polymerase II C-terminal domain phosphatase Ssu72.* J Biol Chem, 2011. **286**(7): p. 5717-26.

16. Kubicek, K., et al., *Serine phosphorylation and proline isomerization in RNAP II CTD control recruitment of Nrd1.* Genes Dev, 2012. **26**(17): p. 1891-6.

17. Ranuncolo, S.M., et al., *Evidence of the involvement of O-GlcNAc-modified human RNA polymerase II CTD in transcription in vitro and in vivo.* J Biol Chem, 2012. **287**(28): p. 23549-61.

18. Du, L. and S.L. Warren, *A functional interaction between the carboxy-terminal domain of RNA polymerase II and pre-mRNA splicing.* J Cell Biol, 1997. **136**(1): p. 5-18.

19. Lee, J.M. and A.L. Greenleaf, *A protein kinase that phosphorylates the C-terminal repeat domain of the largest subunit of RNA polymerase II.* Proc Natl Acad Sci U S A, 1989. **86**(10): p. 3624-8.

20. Feaver, W.J., O. Gileadi, and R.D. Kornberg, *Purification and characterization of yeast RNA polymerase II transcription factor b.* J Biol Chem, 1991. **266**(28): p. 19000-5.

21. Lu, H., et al., *Human general transcription factor IIH phosphorylates the C-terminal domain of RNA polymerase II.* Nature, 1992. **358**(6388): p. 641-5.

22. Serizawa, H., R.C. Conaway, and J.W. Conaway, *A carboxyl-terminal-domain kinase associated with RNA polymerase II transcription factor delta from rat liver.* Proc Natl Acad Sci U S A, 1992. **89**(16): p. 7476-80.

23. Marshall, N.F., et al., *Control of RNA polymerase II elongation potential by a novel carboxyl-terminal domain kinase.* J Biol Chem, 1996. **271**(43): p. 27176-83.

24. Harlen, K.M. and L.S. Churchman, *The code and beyond: transcription regulation by the RNA polymerase II carboxy-terminal domain.* Nat Rev Mol Cell Biol, 2017. **18**(4): p. 263-273.

25. Tee, W.W., et al., *Erk1/2 activity promotes chromatin features and RNAPII phosphorylation at developmental promoters in mouse ESCs.* Cell, 2014. **156**(4): p. 678-90.

26. Hintermair, C., et al., *Threonine-4 of mammalian RNA polymerase II CTD is targeted by Polo-like kinase 3 and required for transcriptional elongation.* EMBO J, 2012. **31**(12): p. 2784-97.

27. Hsin, J.P., A. Sheth, and J.L. Manley, *RNAP II CTD phosphorylated on threonine-4 is required for histone mRNA 3' end processing.* Science, 2011. **334**(6056): p. 683-6.

28. Nemec, C.M., et al., *Different phosphoisoforms of RNA polymerase II engage the Rtt103 termination factor in a structurally analogous manner.* Proc Natl Acad Sci U S A, 2017. **114**(20): p. E3944-E3953.

29. Chapman, R.D., et al., *Transcribing RNA polymerase II is phosphorylated at CTD residue serine-7.* Science, 2007. **318**(5857): p. 1780-2.

30. Egloff, S., et al., *Serine-7 of the RNA polymerase II CTD is specifically required for snRNA gene expression.* Science, 2007. **318**(5857): p. 1777-9.

31. Feaver, W.J., et al., *Relationship of CDK-activating kinase and RNA polymerase II CTD kinase TFIIH/TFIIK.* Cell, 1994. **79**(6): p. 1103-9.

32. Xiao, T., et al., *Phosphorylation of RNA polymerase II CTD regulates H3 methylation in yeast.* Genes Dev, 2003. **17**(5): p. 654-63.

33. Jang, M.K., et al., *The bromodomain protein Brd4 is a positive regulatory component of P-TEFb and stimulates RNA polymerase II-dependent transcription.* Mol Cell, 2005. **19**(4): p. 523-34.

34. Xiang, K., J.L. Manley, and L. Tong, *The yeast regulator of transcription protein Rtr1 lacks an active site and phosphatase activity.* Nat Commun, 2012. **3**: p. 946.

35. Zhang, Z., J. Fu, and D.S. Gilmour, *CTD-dependent dismantling of the RNA polymerase II elongation complex by the pre-mRNA 3'-end processing factor, Pcf11.* Genes Dev, 2005. **19**(13): p. 1572-80.

36. Govind, C.K., et al., *Phosphorylated Pol II CTD recruits multiple HDACs, including Rpd3C(S), for methylation-dependent deacetylation of ORF nucleosomes.* Mol Cell, 2010. **39**(2): p. 234-46.

37. Lunde, B.M., et al., *Cooperative interaction of transcription termination factors with the RNA polymerase II C-terminal domain.* Nat Struct Mol Biol, 2010. **17**(10): p. 1195-201.

38. Sun, M., et al., *A tandem SH2 domain in transcription elongation factor Spt6 binds the phosphorylated RNA polymerase II C-terminal repeat domain (CTD).* J Biol Chem, 2010. **285**(53): p. 41597-603.

39. Cho, E.J., et al., *Opposing effects of Ctk1 kinase and Fcp1 phosphatase at Ser 2 of the RNA polymerase II C-terminal domain.* Genes Dev, 2001. **15**(24): p. 3319-29.

40. Zhang, Y., M. Zhang, and Y. Zhang, *Crystal structure of Ssu72, an essential eukaryotic phosphatase specific for the C-terminal domain of RNA polymerase II, in complex with a transition state analogue.* Biochem J, 2011. **434**(3): p. 435-44.

41. Sims, R.J., 3rd, et al., *The C-terminal domain of RNA polymerase II is modified by site-specific methylation.* Science, 2011. **332**(6025): p. 99-103.

42. Voss, K., et al., *Site-specific methylation and acetylation of lysine residues in the C-terminal domain (CTD) of RNA polymerase II.* Transcription, 2015. **6**(5): p. 91-101.

43. Schroder, S., et al., *Acetylation of RNA polymerase II regulates growth-factor-induced gene transcription in mammalian cells.* Mol Cell, 2013. **52**(3): p. 314-24.

44. Li, H., et al., *Wwp2-mediated ubiquitination of the RNA polymerase II large subunit in mouse embryonic pluripotent stem cells.* Mol Cell Biol, 2007. **27**(15): p. 5296-305.

45. Saiz, J.E. and R.P. Fisher, *A CDK-activating kinase network is required in cell cycle control and transcription in fission yeast.* Curr Biol, 2002. **12**(13): p. 1100-5.

46. Conaway, J.W. and R.C. Conaway, *A multisubunit transcription factor essential for accurate initiation by RNA polymerase II.* J Biol Chem, 1989. **264**(4): p. 2357-62.

47. Conaway, R.C. and J.W. Conaway, *An RNA polymerase II transcription factor has an associated DNA-dependent ATPase (dATPase) activity strongly stimulated by the TATA region of promoters.* Proc Natl Acad Sci U S A, 1989. **86**(19): p. 7356-60.

48. Svejstrup, J.Q., et al., *Different forms of TFIIH for transcription and DNA repair: holo-TFIIH and a nucleotide excision repairosome.* Cell, 1995. **80**(1): p. 21-8.

49. Winkler, G.S., et al., *TFIIH with inactive XPD helicase functions in transcription initiation but is defective in DNA repair.* J Biol Chem, 2000. **275**(6): p. 4258-66.

50. Iben, S., et al., *TFIIH plays an essential role in RNA polymerase I transcription.* Cell, 2002. **109**(3): p. 297-306.

51. Larochelle, S., et al., *Requirements for Cdk7 in the assembly of Cdk1/cyclin B and activation of Cdk2 revealed by chemical genetics in human cells.* Molecular Cell, 2007. **25**(6): p. 839-850.

52. Rossi, D.J., et al., *Inability to enter S phase and defective RNA polymerase IICTD phosphorylation in mice lacking Mat1.* Embo Journal, 2001. **20**(11): p. 2844-2856.

53. Serizawa, H., et al., *Association of Cdk-Activating Kinase Subunits with Transcription Factor Tfiih.* Nature, 1995. **374**(6519): p. 280-282.

54. Shiekhattar, R., et al., *Cdk-Activating Kinase Complex Is a Component of Human Transcription Factor Tfiih.* Nature, 1995. **374**(6519): p. 283-287.

55. Roy, R., et al., *The Mo15 Cell-Cycle Kinase Is Associated with the Tfiih Transcription DNA-Repair Factor.* Cell, 1994. **79**(6): p. 1093-1101.

56. Morgan, D.O., *Principles of Cdk Regulation.* Nature, 1995. **374**(6518): p. 131-134.

57. Espinoza, F.H., et al., *A cyclin-dependent kinase-activating kinase (CAK) in budding yeast unrelated to vertebrate CAK.* Science, 1996. **273**(5282): p. 1714-1717.

58. Kaldis, P., A. Sutton, and M.J. Solomon, *The Cdk-activating kinase (CAK) from budding yeast.* Cell, 1996. **86**(4): p. 553-564.

59. Yoh, S.M., et al., *The Spt6 SH2 domain binds Ser2-P RNAPII to direct Iws1-dependent mRNA splicing and export.* Genes Dev, 2007. **21**(2): p. 160-74.

60. Kizer, K.O., et al., *A novel domain in Set2 mediates RNA polymerase II interaction and couples histone H3 K36 methylation with transcript elongation.* Mol Cell Biol, 2005. **25**(8): p. 3305-16.

61. Dermody, J.L., et al., *Unphosphorylated SR-like protein Npl3 stimulates RNA polymerase II elongation.* PLoS One, 2008. **3**(9): p. e3273.

62. Hollingworth, D., et al., *RNA polymerase II CTD phosphopeptides compete with RNA for the interaction with Pcf11.* RNA, 2006. **12**(4): p. 555-60.

63. Sims, R.J., 3rd, R. Belotserkovskaya, and D. Reinberg, *Elongation by RNA polymerase II: the short and long of it.* Genes Dev, 2004. **18**(20): p. 2437-68.

64. Peterlin, B.M. and D.H. Price, *Controlling the elongation phase of transcription with P-TEFb.* Mol Cell, 2006. **23**(3): p. 297-305.

65. Vasiljeva, L., et al., *The Nrd1-Nab3-Sen1 termination complex interacts with the Ser5-phosphorylated RNA polymerase II C-terminal domain.* Nat Struct Mol Biol, 2008. **15**(8): p. 795-804.

66. Komarnitsky, P., E.J. Cho, and S. Buratowski, *Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription.* Genes Dev, 2000. **14**(19): p. 2452-60.

67. Rodriguez-Molina, J.B., et al., *Engineered Covalent Inactivation of TFIIH-Kinase Reveals an Elongation Checkpoint and Results in Widespread mRNA Stabilization.* Mol Cell, 2016. **63**(3): p. 433-44.

68. Krogan, N.J., et al., *Methylation of histone H3 by Set2 in Saccharomyces cerevisiae is linked to transcriptional elongation by RNA polymerase II.* Mol Cell Biol, 2003. **23**(12): p. 4207-18.

69. Morris, D.P. and A.L. Greenleaf, *The splicing factor, Prp40, binds the phosphorylated carboxyl-terminal domain of RNA polymerase II.* J Biol Chem, 2000. **275**(51): p. 39935-43.

70. Nojima, T., et al., *Mammalian NET-Seq Reveals Genome-wide Nascent Transcription Coupled to RNA Processing.* Cell, 2015. **161**(3): p. 526-540.

71. Schwer, B. and S. Shuman, *Deciphering the RNA polymerase II CTD code in fission yeast.* Mol Cell, 2011. **43**(2): p. 311-8.

72. Wong, K.H., Y. Jin, and K. Struhl, *TFIIH phosphorylation of the Pol II CTD stimulates mediator dissociation from the preinitiation complex and promoter escape.* Mol Cell, 2014. **54**(4): p. 601-12.

73. Akhtar, M.S., et al., *TFIIH Kinase Places Bivalent Marks on the Carboxy-Terminal Domain of RNA Polymerase II.* Molecular Cell, 2009. **34**(3): p. 387-393.

74. 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**(20): p. 5455-5464.

75. Descostes, N., et al., *Tyrosine phosphorylation of RNA polymerase II CTD is associated with antisense promoter transcription and active enhancers in mammalian cells.* Elife, 2014. **3**: p. e02105.

76. Hsin, J.P., et al., *RNAP II CTD tyrosine 1 performs diverse functions in vertebrate cells.* Elife, 2014. **3**: p. e02112.

77. Mayer, A., et al., *CTD tyrosine phosphorylation impairs termination factor recruitment to RNA polymerase II.* Science, 2012. **336**(6089): p. 1723-5.

78. Harlen, K.M., et al., *Comprehensive RNA Polymerase II Interactomes Reveal Distinct and Varied Roles for Each Phospho-CTD Residue.* Cell Rep, 2016. **15**(10): p. 2147-2158.

79. Rosonina, E., et al., *Threonine-4 of the budding yeast RNAP II CTD couples transcription with Htz1-mediated chromatin remodeling.* Proc Natl Acad Sci U S A, 2014. **111**(33): p. 11924-31.

80. Dias, J.D., et al., *Methylation of RNA polymerase II non-consensus Lysine residues marks early transcription in mammalian cells.* Elife, 2015. **4**.

81. Daulny, A., et al., *Modulation of RNA polymerase II subunit composition by ubiquitylation.* Proc Natl Acad Sci U S A, 2008. **105**(50): p. 19649-54.

82. Lu, L., et al., *Distributive O-GlcNAcylation on the Highly Repetitive C-Terminal Domain of RNA Polymerase II.* Biochemistry, 2016. **55**(7): p. 1149-58.

83. Greber, B.J., et al., *The cryo-electron microscopy structure of human transcription factor IIH.* Nature, 2017. **549**(7672): p. 414-+.

84. Schilbach, S., et al., *Structures of transcription pre-initiation complex with TFIIH and Mediator.* Nature, 2017. **551**(7679): p. 204-+.

85. Fishburn, J., et al., *Double-stranded DNA translocase activity of transcription factor TFIIH and the mechanism of RNA polymerase II open complex formation.* Proceedings of the National Academy of Sciences of the United States of America, 2015. **112**(13): p. 3961-3966.

86. Grunberg, S., L. Warfield, and S. Hahn, *Architecture of the RNA polymerase II preinitiation complex and mechanism of ATP-dependent promoter opening.* Nature Structural & Molecular Biology, 2012. **19**(8): p. 788-796.

87. Jeronimo, C. and F. Robert, *Kin28 regulates the transient association of Mediator with core promoters.* Nature Structural & Molecular Biology, 2014. **21**(5): p. 449-455.

88. Robinson, P.J., et al., *Structure of a Complete Mediator-RNA Polymerase II Pre-Initiation Complex.* Cell, 2016. **166**(6): p. 1411-+.

89. Sogaard, T.M.M. and J.Q. Svejstrup, *Hyperphosphorylation of the C-terminal repeat domain of RNA polymerase II facilitates dissociation of its complex with mediator.* Journal of Biological Chemistry, 2007. **282**(19): p. 14113-14120.

90. Adelman, K. and J.T. Lis, *Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans.* Nature Reviews Genetics, 2012. **13**(10): p. 720-731.

91. Ebmeier, C.C., et al., *Human TFIIH Kinase CDK7 Regulates Transcription-Associated Chromatin Modifications.* Cell Reports, 2017. **20**(5): p. 1173-1186.

92. Larochelle, S., et al., *Dichotomous but stringent substrate selection by the dual-function Cdk7 complex revealed by chemical genetics.* Nature Structural & Molecular Biology, 2006. **13**(1): p. 55-62.

93. Qiu, H.F., et al., *Pol II CTD kinases Bur1 and Kin28 promote Spt5 CTR-independent recruitment of Paf1 complex.* Embo Journal, 2012. **31**(16): p. 3494-3505.

94. Fujinaga, K., et al., *Dynamics of human immunodeficiency virus transcription: P-TEFb phosphorylates RD and dissociates negative effectors from the transactivation response element.* Molecular and Cellular Biology, 2004. **24**(2): p. 787-795.

95. Bharati, A.P., et al., *The mRNA capping enzyme of Saccharomyces cerevisiae has dual specificity to interact with CTD of RNA Polymerase II.* Scientific Reports, 2016. **6**.

96. Schroeder, S.C., et al., *Dynamic association of capping enzymes with transcribing RNA polymerase II.* Genes & Development, 2000. **14**(19): p. 2435-2440.

97. de la Mata, M., et al., *A slow RNA polymerase II affects alternative splicing in vivo.* Molecular Cell, 2003. **12**(2): p. 525-532.

98. Fong, N., et al., *Pre-mRNA splicing is facilitated by an optimal RNA polymerase II elongation rate.* Genes & Development, 2014. **28**(23): p. 2663-2676.

99. Kim, M., et al., *Distinct pathways for snoRNA and mRNA termination.* Molecular Cell, 2006. **24**(5): p. 723-734.

100. Steinmetz, E.J., et al., *RNA-binding protein Nrd1 directs poly(A)-independent 3 '-end formation of RNA polymerase II transcripts.* Nature, 2001. **413**(6853): p. 327-331.

101. Gudipati, R.K., et al., *Phosphorylation of the RNA polymerase II C-terminal domain dictates transcription termination choice.* Nature Structural & Molecular Biology, 2008. **15**(8): p. 786-794.

102. Mosley, A.L., et al., *Rtr1 Is a CTD Phosphatase that Regulates RNA Polymerase II during the Transition from Serine 5 to Serine 2 Phosphorylation.* Molecular Cell, 2009. **34**(2): p. 168-178.

103. Steinmetz, E.J. and D.A. Brow, *Ssu72 protein mediates both poly(A)-coupled and poly(A)-independent termination of RNA polymerase II transcription.* Molecular and Cellular Biology, 2003. **23**(18): p. 6339-6349.

104. Rodriguez, C.R., et al., *Kin28, the TFIIH-associated carboxy-terminal domain kinase, facilitates the recruitment of mRNA processing machinery to RNA polymerase II.* Molecular and Cellular Biology, 2000. **20**(1): p. 104-112.

105. Serizawa, H., J.W. Conaway, and R.C. Conaway, *Phosphorylation of C-Terminal Domain of RNA Polymerase-II Is Not Required in Basal Transcription.* Nature, 1993. **363**(6427): p. 371-374.

106. Li, Y. and R.D. Kornberg, *Interplay of Positive and Negative Effectors in Function of the C-Terminal Repeat Domain of Rna-Polymerase-Ii.* Proceedings of the National Academy of Sciences of the United States of America, 1994. **91**(6): p. 2362-2366.

107. Akoulitchev, S., et al., *Requirement for Tfiih Kinase-Activity in Transcription by Rna-Polymerase-Ii.* Nature, 1995. **377**(6549): p. 557-560.

108. McNeil, J.B., H. Agah, and D. Bentley, *Activated transcription independent of the RNA polymerase II holoenzyme in budding yeast.* Genes & Development, 1998. **12**(16): p. 2510-2521.

109. Holstege, F.C.P., et al., *Dissecting the regulatory circuitry of a eukaryotic genome.* Cell, 1998. **95**(5): p. 717-728.

110. Kanin, E.I., et al., *Chemical inhibition of the TFIIH-associated kinase Cdk7/Kin28 does not impair global mRNA synthesis.* Proceedings of the National Academy of Sciences of the United States of America, 2007. **104**(14): p. 5812-5817.

111. Hong, S.W., et al., *Phosphorylation of the RNA polymerase II C-terminal domain by TFIIH kinase is not essential for transcription of Saccharomyces cerevisiae genome.* Proceedings of the National Academy of Sciences of the United States of America, 2009. **106**(34): p. 14276-14280.

112. Lolli, G., et al., *The crystal structure of human CDK7 and its protein recognition properties.* Structure, 2004. **12**(11): p. 2067-2079.

# Chapter 2: A Rational Approach to Targeted Inactivation of the TFIIH-Kinase Clarifies Its Regulatory Roles in Transcription

This chapter is adapted from Rodríguez-Molina JB*, Tseng SC*, Simonett SP, Taunton J, Ansari AZ (2016) Engineered Covalent Inactivation of TFIIH-Kinase Reveals an Elongation Checkpoint and Results in Widespread mRNA Stabilization. *Mol Cell* 63(3):433-44. *Co-authors; JBR-M designed the initial *kin28is* alleles and performed inhibitor wash-out experiments, performed and analyzed data from ChIP-chip, and designed the Pol II genetic interaction assays conducted by SPS; all other experiments presented in this chapter were performed by SCT.

## 2.1 Introduction

The transition of Pol II from a promoter-bound pre-initiation complex (PIC) to active transcription initiation is gated by TFIIH [1]. The kinase subunit (Kin28 in *S. cerevisiae* or CDK7 in humans) marks the transcriptionally active polymerase by hyperphosphorylating the C-terminal domain (CTD) of its largest subunit [2-4]. The CTD is comprised of a repeating heptapeptide (Y1-S2-P3-T4-S5-P6-S7) that is phosphorylated at positions 5 and 7 by Kin28/CDK7 [5-7]. Hyperphosphorylation of the CTD by Kin28/CDK7 is thought to play an essential role in releasing Pol II from the promoter-bound PIC. Kin28/CDK7-mediated phosphorylation of the CTD also creates a scaffold to recruit factors that act on nascent transcripts, the underlying chromatin, and the CTD itself. Promoter-proximal marks on Ser5 and Ser7 are also thought to "prime" the CTD for subsequent modifications that coordinate sequential association of different cellular machines that facilitate transcription [8-13].

Facilitating promoter release was thought to be the primary function of Kin28/CDK7, because certain temperature-sensitive (*ts*) alleles of Kin28 severely disable transcription initiation at non-permissive temperatures [14, 15]. Genetic studies further highlighted the essential nature of this kinase, as deletion of Kin28 or substitution with a catalytically inactive form resulted in loss of cellular viability [16]. More recently, a small molecule inhibitor that conjugates to a cysteine residue outside the kinase domain of human CDK7 was shown to inhibit kinase activity and result in cytotoxicity in cancer cell lines [17]. These observations suggest that blocking Kin28/CDK7 from hyper-phosphorylating the CTD effectively blocks promoter release and disables transcription initiation by Pol II.

In contrast, a catalytically impaired *kin28* (T17D) mutant showed defective placement of the m7G cap on nascent transcripts but only had a minor impact on mRNA synthesis [18]. Non-covalent inhibition of an analog-sensitive allele of Kin28 (*kin28as*)

with purine analogs demonstrated that loss of Kin28-dependent CTD hyperphosphorylation did not efficiently block promoter release in vivo [19, 20]. Using this chemical-genetic approach, subsequent studies confirmed the observation that Kin28/CDK7 kinase activity was not essential for promoter release or Pol-II-dependent RNA synthesis [5, 6, 21-27]. Depleting Kin28 from the nucleus using the in vivo "anchor-away" approach in *S. cerevisiae* also did not abrogate transcription initiation [26]. Similarly, attenuating CDK7 function in mammalian cells did not block transcription initiation [6, 28-30]. Furthermore, chemical inhibition of CDK7 with THZ1 did not dramatically impact transcription initiation and promoter release [31]. In sum, these observations suggest that Kin28/CDK7 functions to prime the CTD for subsequent binding of factors that coordinate the transcription cycle.

Despite significant inhibition of kinase activity with non-covalent inhibitors, the lack of cellular lethality led to the possibility that activity of other promoter-bound kinases, such as Cdc28 [32] or Srb10 [27], or residual activity due to reversible inhibition of Kin28as permitted low levels of CTD phosphorylation to persist. Minimal CTD phosphorylation may suffice for promoter release and productive RNA synthesis.

To resolve these discrepancies and definitively address the critical role of Kin28 activity in RNA synthesis, we developed a targeted covalent inhibition approach that robustly inactivates Kin28 in vivo. We recapitulate the lethal phenotype of a catalytically inactive *kin28* allele. Using our "irreversibly sensitized" Kin28 allele (*kin28is*), we observe that covalent inhibition impacts promoter escape to differing degrees and impairs transition to productive elongation, revealing a rate-limiting "checkpoint" coincident with the +2 nucleosome. Promoter-proximal pausing of Pol II, a hallmark of the metazoan transcription cycle, is not observed by chromatin immunoprecipitation (ChIP) in budding yeast [33]. However, upon irreversible inhibition of Kin28is, a clear enrichment of Pol II is observed within a transition window where promoter-proximal factors are exchanged for

elongation and 3′ end processing factors [34-37]. Consistent with a potential defect in early transcription elongation, a catalytically “slow” version of Pol II is particularly sensitive to any degree of Kin28 inhibition, whereas a “fast” transcribing Pol II is more resistant [38]. Unexpectedly, we observe a global decrease in nascent transcription along with a stabilization of extant mRNA levels. Taken together, our results reconcile long-standing controversies and reveal a role for Kin28 in facilitating promoter release, enhancing the transition to productive elongation, and priming the CTD for downstream stages of the transcription cycle. Moreover, given the highly conserved ATP-binding site across kinases, our strategy for structure-guided covalent inhibition of Kin28 readily lends itself to targeted irreversible inhibition of other kinases *in vivo*.

### 2.2 Design of an Irreversibly Sensitive Kinase Allele

To definitively dissect the role of Kin28, we used structure-guided protein design to generate an “irreversibly sensitized” version of Kin28 (*kin28is*) that can be selectively and covalently inhibited by an ATP analog bearing a thiol-reactive chloromethyl-ketone moiety (CMK). The approach was inspired by the use of CMK to inhibit CDC5, which has a naturally occurring solvent-exposed cysteine within its ATP-binding pocket. CMK docks into an enlarged ATP-binding pocket of CDC5 (*cdc5as*) and becomes covalently conjugated to the reactive cysteine, thereby irreversibly inhibiting kinase function [39, 40]. We reasoned that rational engineering of Kin28 to possess an enlarged ATP-binding pocket and a reactive Cys in the structurally identical position to CDC5 would render it sensitive to CMK. We created a mutant allele of Kin28 that contains two substitutions within its ATP-binding pocket: (1) a smaller gatekeeper residue (L83G) to accommodate the para-methyl-phenyl group of CMK and (2) a solvent-exposed Cys positioned to react with CMK in place of a conserved Val (V21C) (**Figure 2.1A**). Our design predicts that CMK would meet the two selectivity filters by docking into the enlarged active site and

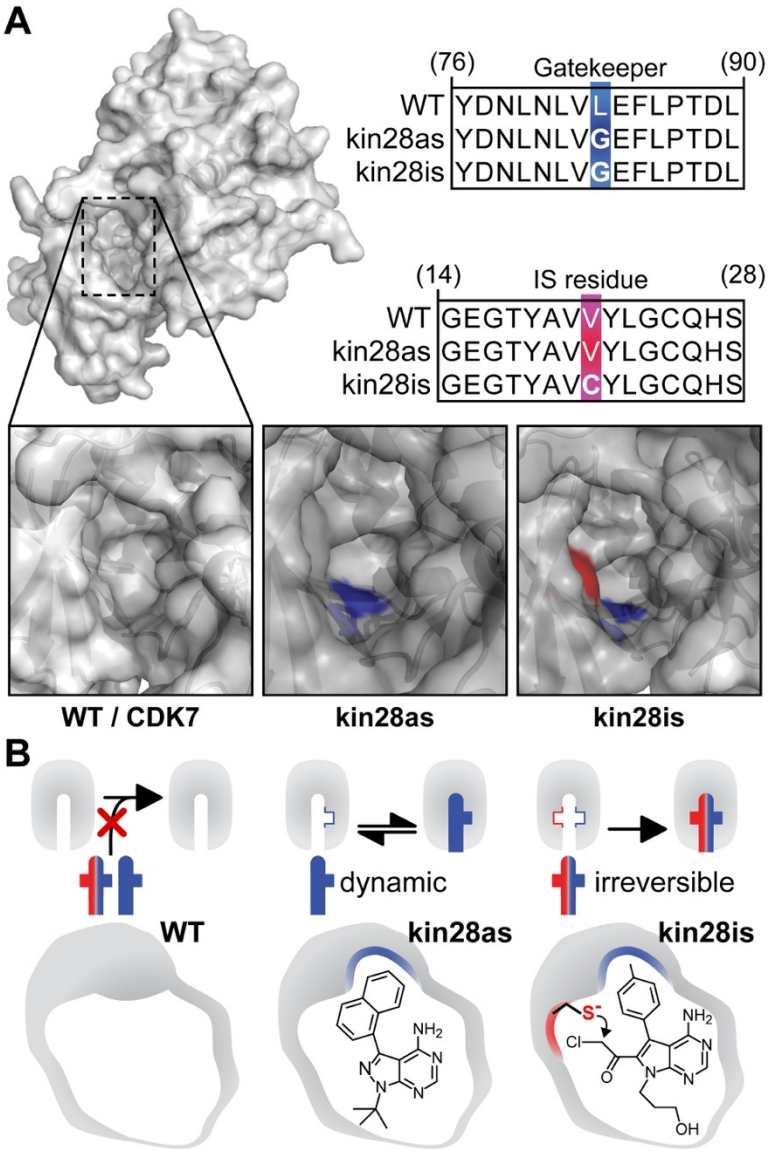



**Figure 2.1 Rational Design of Irreversibly Sensitive Kin28 Allele**

**(A)** Left: structure of human CDK7 (PDB: 1UA2) highlighting the kinase active site. Right: sequence alignment of WT, Kin28as, or Kin28is with indicated residue positions. To generate the *kin28is* allele, the gatekeeper residue (L83, blue) was substituted for a glycine, and a conserved valine (V21, red) was substituted for a solvent-exposed cysteine. The homologous gatekeeper and conserved valine substitutions were modeled into the CDK7 structure using PyMOL. **(B)** Schematic representation of a WT kinase (left), an analog-sensitive kinase with a gatekeeper substitution (center), and the irreversibly sensitive kinase containing both the smaller gatekeeper and the reactive cysteine (right).

irreversibly conjugating to Kin28is, thereby preventing engagement of cellular ATP. In contrast, non-covalent inhibitors equilibrate with the kinase as a function of on/off rates, inhibitor concentrations, and cellular ATP levels (**Figure 2.1B**). In other words, CMK would irreversibly silence the catalytic activity of the kinase.

### 2.3 Covalent Inhibition of Kin28is Irreversibly Arrests Cell Growth

*kin28is* cells grown in the presence of CMK display striking sensitivity (growth rate, as measured by doubling time, is halved at 40 nM CMK) and growth arrest on solid media as well as in liquid cultures (**Figures 2.2** and **2.3**). In the presence of 1-NA-PP1 (a non-covalent reversible inhibitor), *kin28as* and *kin28is* cells display partial decrease in growth rates and, over time, attain growth densities comparable to uninhibited cells. Importantly, cells with Kin28 bearing only the gatekeeper mutation (L83G) or only the V21C mutation (not shown) are not sensitive to CMK (**Figure 2.2**). Given the strict requirement for two selectivity filters, the ATP-binding sites of endogenous yeast kinases would not permit covalent conjugation by CMK. Consistent with this expectation, CMK does not perturb the transcriptome in wild-type cells or in cells bearing the *kin28as* allele (detailed below).

To determine whether the CMK-rendered growth arrest was irreversible, we treated cells with 1-NA-PP1 or CMK for 24 hours, followed by cell harvest, extensive washes with fresh media, and further growth for 24 hours in media lacking either inhibitor. Wild-type (WT) and *kin28as* cells were able to recover from inhibition after drug washout, whereas *kin28is* cells treated with CMK were unable to recover (**Figure 2.4**). Together, these data suggest our covalent inhibition strategy is specific, irreversible, and able to recapitulate the growth-arrest phenotype manifested by a catalytically inactive *kin28* allele. Although we have been cognizant of partial kinase inhibition by reversible inhibitors, others may not realize this caveat [21].

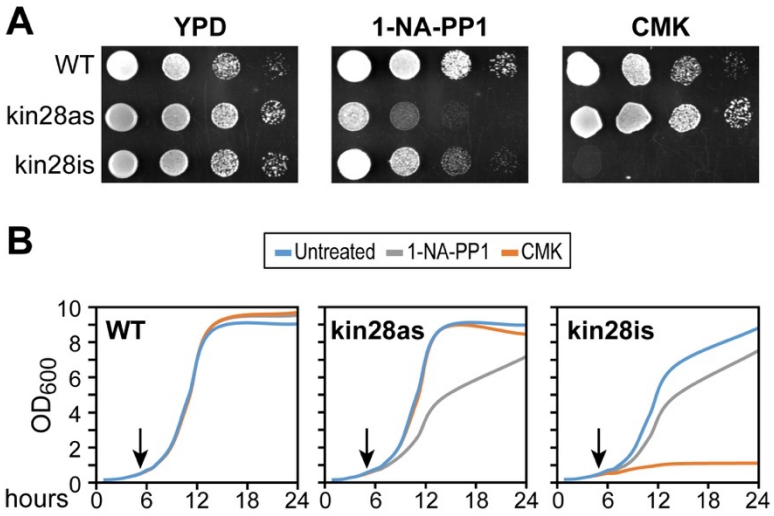



**Figure 2.2 Covalent Inhibition of Kin28is Irreversibly Arrests Cell Growth**

**(A)** Drop tests of serially diluted cultures and **(B)** growth curves of WT, *kin28as*, and *kin28is* cells grown in 5 μM 1-NA-PP1, 5 μM CMK, or DMSO solvent control. Arrows in the growth curves denote addition of the respective inhibitor.

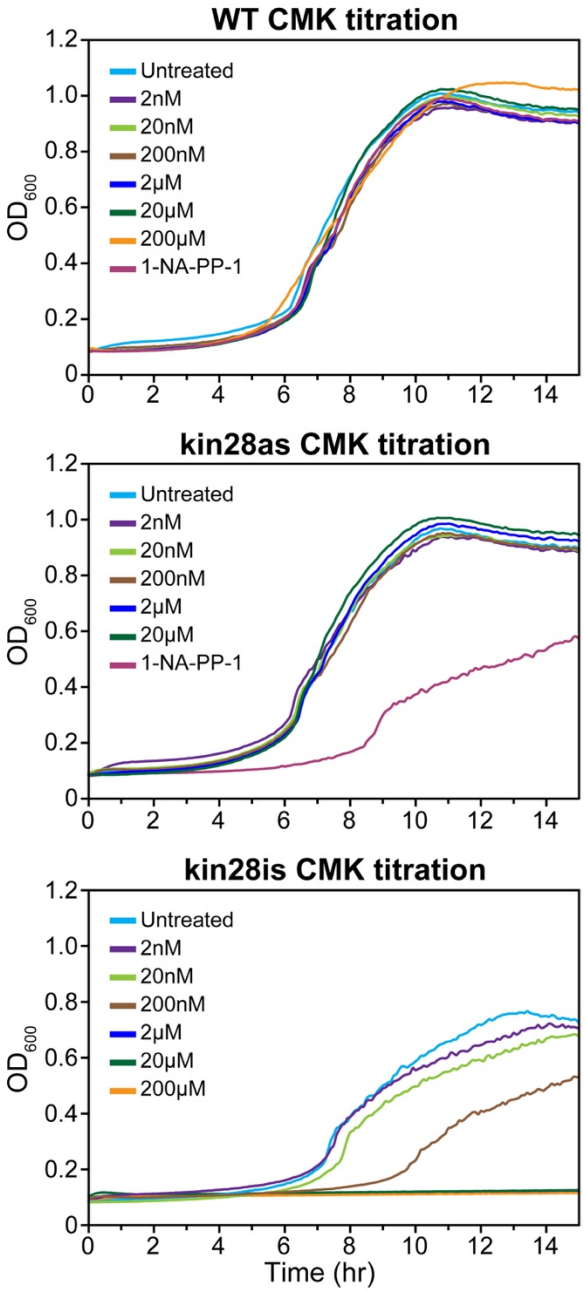



**Figure 2.3 Dose-dependent Decrease in *kin28is* Growth Rate upon Covalent Inhibition**

WT (top), *kin28as* (middle), and *kin28is* (bottom) cells were grown in increasing concentrations of CMK. Growth rate was monitored for 16 hours. As a control, 2 μM 1-NA-PP1 was also added to WT and *kin28as* growth curves.

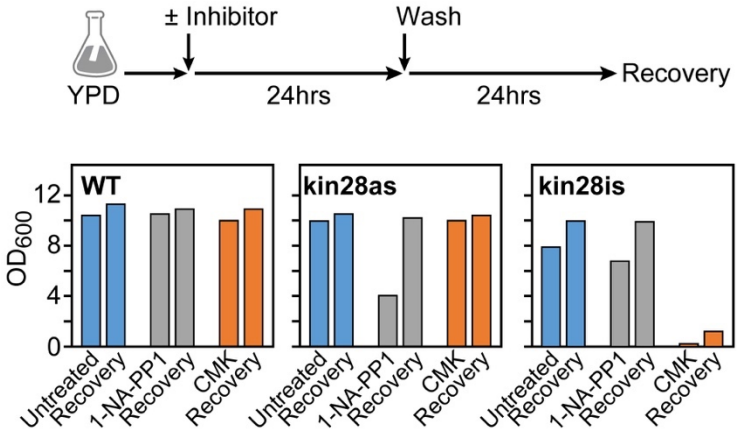



**Figure 2.4 Inhibitor Washout Assay**

To test whether CMK could be washed out, overnight cultures of WT, *kin28as*, and *kin28is* cells were subcultured with 5 μM of 1-NA-PP1, CMK, or volume equivalent of DMSO control for 24 hours. Equal number of cells were harvested and washed with fresh media lacking the inhibitor. Washed cells were used to inoculate fresh media for 24 hours. OD600 measurements were used to monitor cell growth at each end point.

## 2.4 Irreversible Inhibition of Kin28 Drastically Reduces Cellular Ser5-P Levels

Promoter-proximal phospho-Ser5 (Ser5-P) and phospho-Ser7 (Ser7-P) marks are placed primarily by Kin28 [5]. Inhibition of Kin28 with reversible ligands, however, does not result in bulk decrease in Ser5-P or Ser7-P marks [5, 20, 32]. To assess the effect of covalent inhibition of Kin28is on bulk levels of phospho-CTD marks, we prepared whole-cell extracts from exponentially growing cells treated with the DMSO solvent control ("Untreated"), the reversible inhibitor 1-NA-PP1, or the irreversible inhibitor CMK. Western blots show a striking decrease in Ser5-P marks in cells treated with CMK ($p < 0.05$, Student's t test) (**Figure 2.5** and **2.6**). Partial loss of bulk Ser7-P marks is consistent with the role of Bur1 (CDK9 in humans) in placing Ser7-P marks [27]. Ser2-P, phosphorylated tyrosine-1 (Tyr1-P), and phosphorylated threonine-4 (Thr4-P) levels do not appear to change appreciably, suggesting that these are not major targets of Kin28 in cells. Importantly, treatment of WT cells does not affect phospho-CTD levels in line with the requirement of an "*is*"-sensitized kinase for CMK to inhibit kinase activity.

Interestingly, using an antibody that detects hyper-Ser5-P, our blotting shows residual Ser5-P after inhibition of Kin28, which suggests alternative minor Ser5 kinases are still functional (**Figure 2.6B**). However, after 15 min of Kin28 inhibition, bulk Ser5-P levels are visibly diminished in *kin28is* cells (**Figure 2.6C** and **2.6D**) with no difference in overall Kin28 protein abundance after 1 hour, the time span of our inhibition experiments (**Figure 2.6E**). Together, these data further highlight the selectivity and effectiveness of our covalent inhibition approach.

## 2.5 Genome-wide Ablation of Promoter-Proximal Ser5-P and Ser7-P Marks

To determine the impact of covalent inhibition of Kin28 on Pol II distribution, we performed ChIP with antibodies against the tandem affinity purification (TAP) tag (Rpb3-

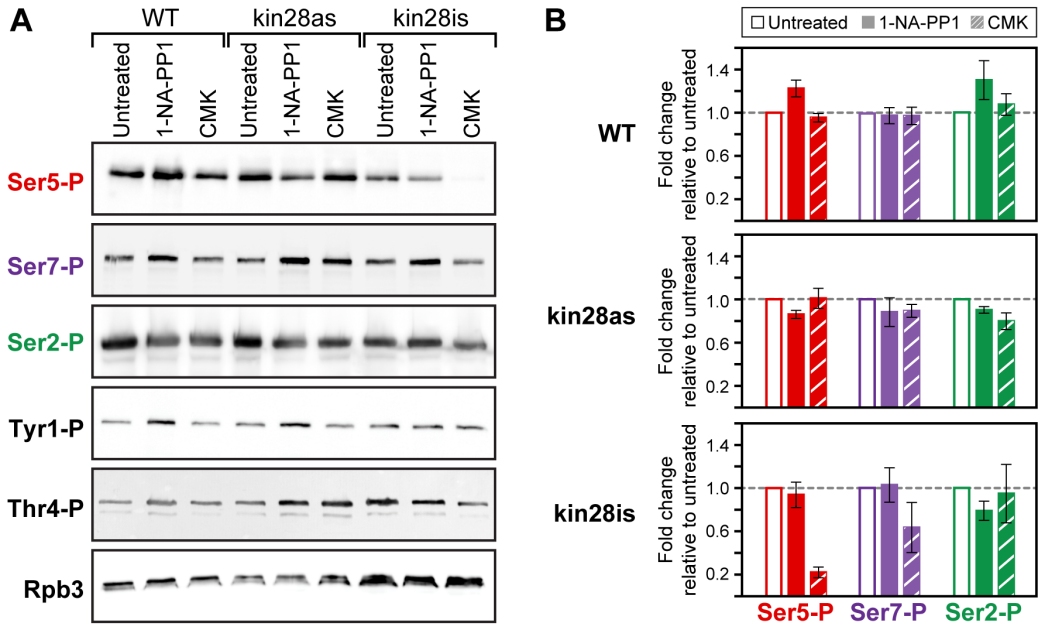



**Figure 2.5 Covalent Inhibition of *kin28is* Leads to Global Decrease in Ser5-P CTD Marks**

**(A)** Immunoblots of co-immunoprecipitations from cells bearing WT, *kin28as*, or *kin28is* alleles treated with DMSO or 5 μM of inhibitors. Blots were probed with phospho-CTD specific antibodies against Ser5-P (H14), Ser7-P (4D12), Ser2-P (S2, Bethyl Laboratories), Tyr1-P (3D12), Thr4-P (1G7), and against the TAP tag in Rpb3-TAP as loading control. **(B)** Signal from each phospho-CTD mark was normalized to its corresponding abundance of Rpb3-TAP, and the change in signal relative to DMSO was calculated for individual replicates. The average fold changes are plotted. Error bars represent the standard error of the mean, n = 3-5.

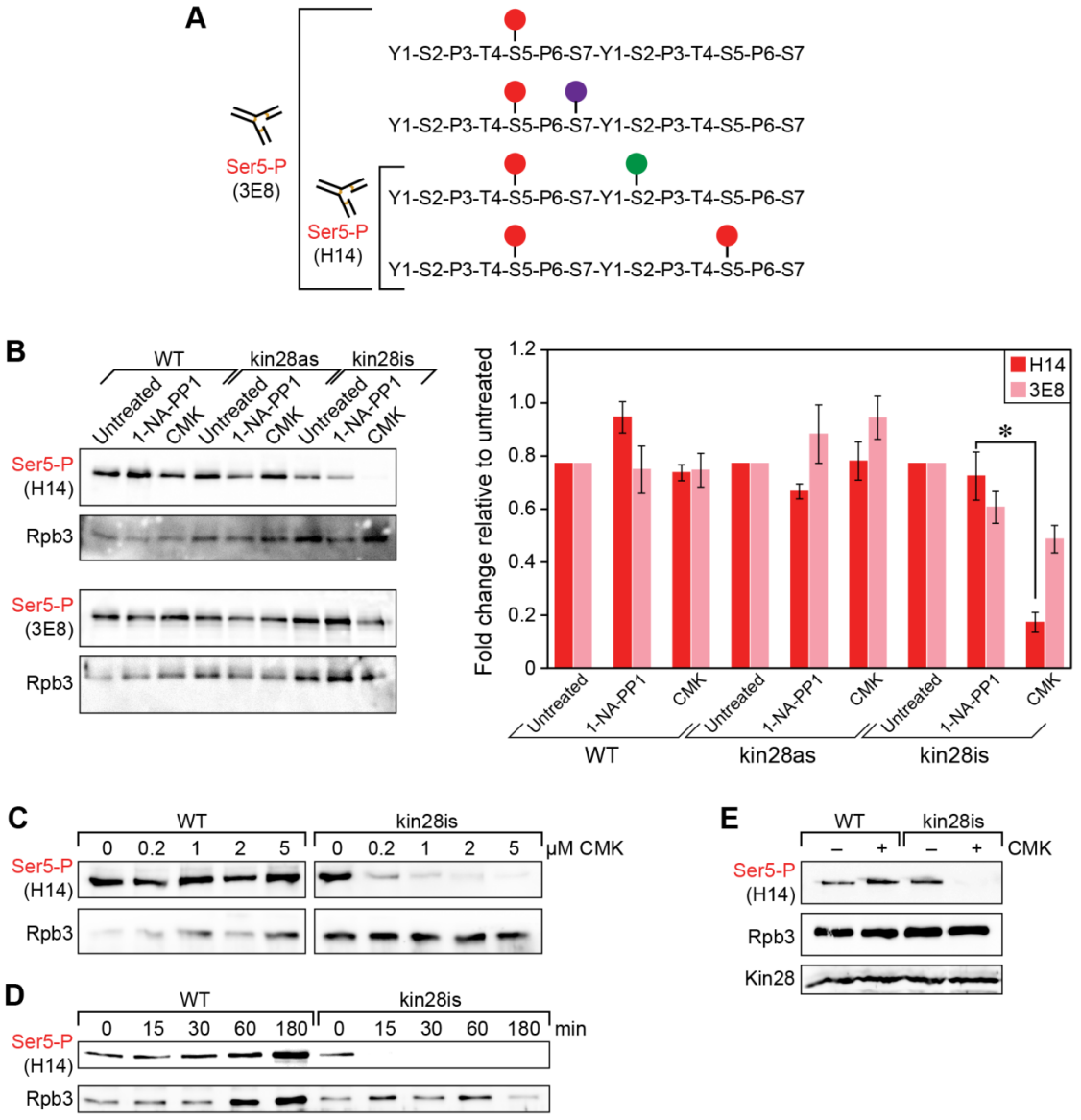



**Figure 2.6 Covalent Inhibition of *kin28is* Abolishes Hyper-Ser5-P CTD Marks**

**(A)** Differential recognition of single and multiple Ser5-P marks by H14 and 3E8 antibodies. 3E8 is able to detect a single Ser5-P mark whereas H14 detects multiple Ser5-P marks [64, 65]. **(B)** Immunoblots of whole cell extracts from cells bearing WT, kin28as or kin28is alleles treated with DMSO control or 5 μM of inhibitors. Blots against Rpb3-TAP were used as a loading control. Signal from each phospho-CTD mark was normalized to the abundance of Rpb3-TAP in its corresponding blot, and the change in signal relative to DMSO was calculated for individual replicates. Error bars represent the standard error of the mean, n = 3-5. $P<0.05$, two-sided Student's t-test. **(C)** CMK titration leads to dose-dependent depletion of bulk Ser5-P marks. **(D)** Ser5-P marks are ablated within 15 minutes of treatment with 5 μM CMK, and remain undetectable after 180 minutes of treatment. **(E)** Covalent inhibition of Kin28is does not affect bulk protein levels of Kin28is.

TAP) or phospho-CTD marks. Covalent inhibition of Kin28is caused a global decrease in Ser5-P marks and a precipitous decrease in promoter-proximal Ser7-P marks (**Figure 2.7**). The loss of Ser5-P and Ser7-P marks were more pronounced when cells were treated with CMK compared to 1-NA-PP1 at the same concentration (**Figure 2.7B**). We validated these decreases in Ser5-P and Ser7-P marks near active gene promoters by ChIP-qPCR at *ADH1* (**Figure 2.8**). Residual Ser5-P signal within genes, as well as remaining Ser7-P signal near promoters and within genes, is consistent with the activity of Bur1 on CTD at these regions [27].

Phosphorylation of the Pol II CTD by Kin28 is thought to release the CTD from its interaction with the Mediator Head module [41]. Covalent inhibition of Kin28is, followed by ChIP-chip analysis of the Mediator Head module subunit, Med8, reveals a shift in Med8 occupancy to −42 bp, rather than its typical location at −116 bp with respect to the transcription start site (TSS) (**Figure 2.8**), consistent with previous studies [26, 42]. While we do not observe a significant difference in Pol II occupancy at genes regulated by TFIID or SAGA complexes (**Figure 2.9**), nascent transcription is slightly more affected at TFIID-dependent rather than SAGA-dependent genes. Kin28 inhibition is also reported to alter start site selection at TFIID-targeted promoters in vitro [43]. However, under covalent inhibition conditions, no such shift was apparent in vivo (**Figure 2.9**). The sum of these results highlights the specific and potent effects of covalent inhibition of Kin28is with CMK previously not attained by reversible inhibition strategies.

### 2.6 Depleted Pol II Occupancy at 3' Ends of Transcribed Regions

Covalent inhibition of Kin28is revealed a reproducible Pol II decrease toward the 3′ ends of protein-coding genes (pc-genes) (**Figure 2.7**). The graded decrease in Pol II levels is inversely correlated with gene length (**Figure 2.10A**, Pearson correlation −0.509, $p = 8.6 \times 10^{-42}$ versus mean $p = 0.501$ from $10^4$ randomized iterations of the

data), suggesting a length-dependent mechanism by which Pol II is depleted from chromatin in the absence of Kin28 activity. To parse out different patterns of Pol II perturbation, we performed k-means clustering of changes in Pol II occupancy at pc-genes with robust levels of Pol II occupancy (top 10%). Visual inspection of the clusters revealed six distinct patterns of Pol II change. For 71% of these genes, Pol II shows depletion at the cleavage and polyadenylation site (CPS) (elongation coefficient), which does not necessarily correlate with Pol II changes at the TSS (promoter escape coefficient) (**Figure 2.10B**). However, the ratio of Pol II at the CPS versus the TSS (Pol II travel index) invariably shows Pol II has a reduced travel index within most of the clusters identified (**Figure 2.10C**). These data suggest Kin28 impacts Pol II promoter escape to different degrees at different promoters and unexpectedly plays an important role in Pol II elongation.

### 2.7 Global Decrease in Nascent RNA Synthesis and Concomitant Stabilization of mRNA

To determine the role of Kin28 activity on RNA synthesis, we performed comparative dynamic transcriptome analysis [44]. The newly synthesized transcripts were pulse labeled with 4-thiouracil, which is incorporated into growing RNA chains. The thiolated transcripts were biotinylated and purified using streptavidin-coated magnetic beads (Nascent, **Table 2.1** and **2.3**). In parallel, we purified Total RNA (Total, **Table 2.2**) as a measure of steady-state transcript levels. To account for any global defects in transcription upon inhibition of Kin28 activity, we spiked in *S. pombe* cells for which RNA was simultaneously prepared and sequenced (6:1 S. *cerevisiae*:*S. pombe* cellular ratio). We fit local linear regressions of gene expression across replicates from S. pombe, which were used to normalize *S. cerevisiae* Nascent and Total RNA sequencing (RNA-Seq) data across replicates and treatments. Thus, we circumvent potentially misleading

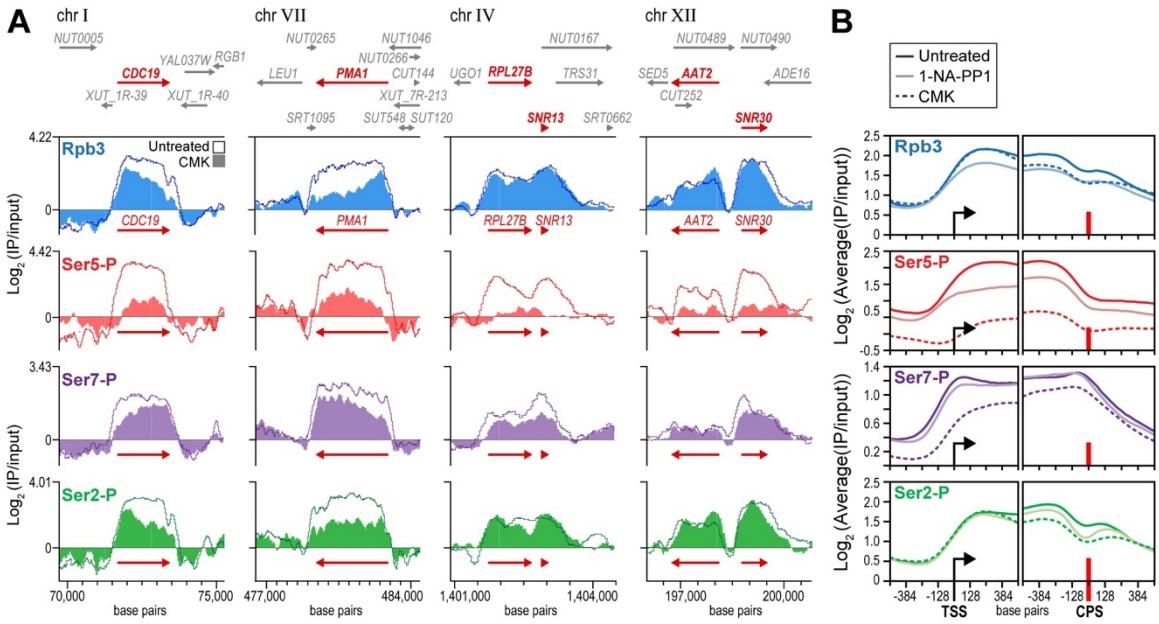



**Figure 2.7 Genome-wide Ablation of Promoter-Proximal Ser5-P and Ser7-P Marks**

**(A)** Single gene traces of Rpb3 (blue), Ser5-P (red), Ser7-P (purple), and Ser2-P (green) from *kin28is* cells treated with DMSO or 5 μM CMK. **(B)** Average occupancy for Rpb3, Ser2-P, Ser5-P, and Ser7-P from *kin28is* cells at the TSS (black arrow) and CPS (red bar) of pc-genes with the top 10% of Pol II occupancy (average Log2 [IP/input], n = 615 genes).

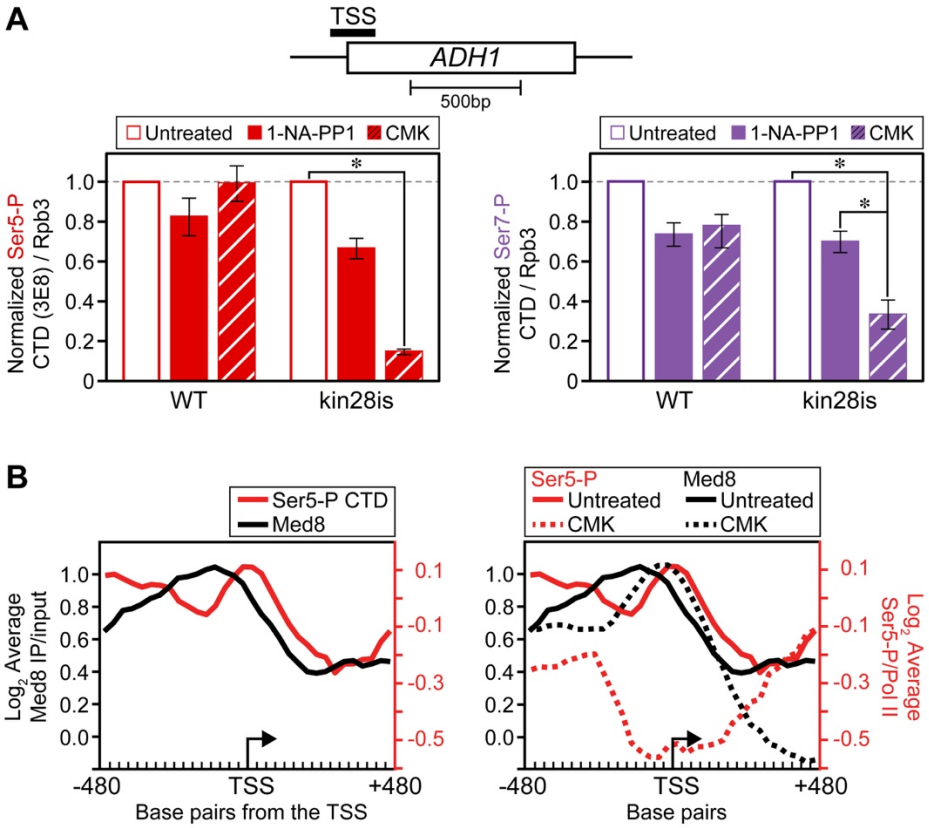



**Figure 2.8 Observed Changes to Ser5-P, Ser7-P, and Med8 Match Previous Observations**

**(A)** ChIP from WT or *kin28is* cells treated with DMSO (solvent control), 2 μM 1-NA- PP1 or 2 μM CMK for 1 hour. Chromatin was immunoprecipitated using specific antibodies against Ser5-P (10 μl, 3E8), Ser7-P (50 μl, 4E12), or the TAP tag (5 μl). DNA was quantified by qPCR using primers that target the transcription start site of the *ADH1* gene (ADH1 -196 Fwd and ADH1 +13 Rev). Ser5-P and Ser7-P values were normalized to Rpb3 levels at that same site (error bars depict standard error of the mean) (*, $P < 0.05$, one-sided paired Student's t-test, n=4). **(B)** (Left) Average Med8 (black trace) and Pol II-normalized Ser5-P marks (red trace) at the TSS (black arrow) of 60 pc-genes isolated from neighboring genetic elements. (Right) Inhibition of Kin28is leads to ablation of Ser5-P marks (dotted red trace) and a corollary downstream shift in Med8 occupancy (dotted black trace).

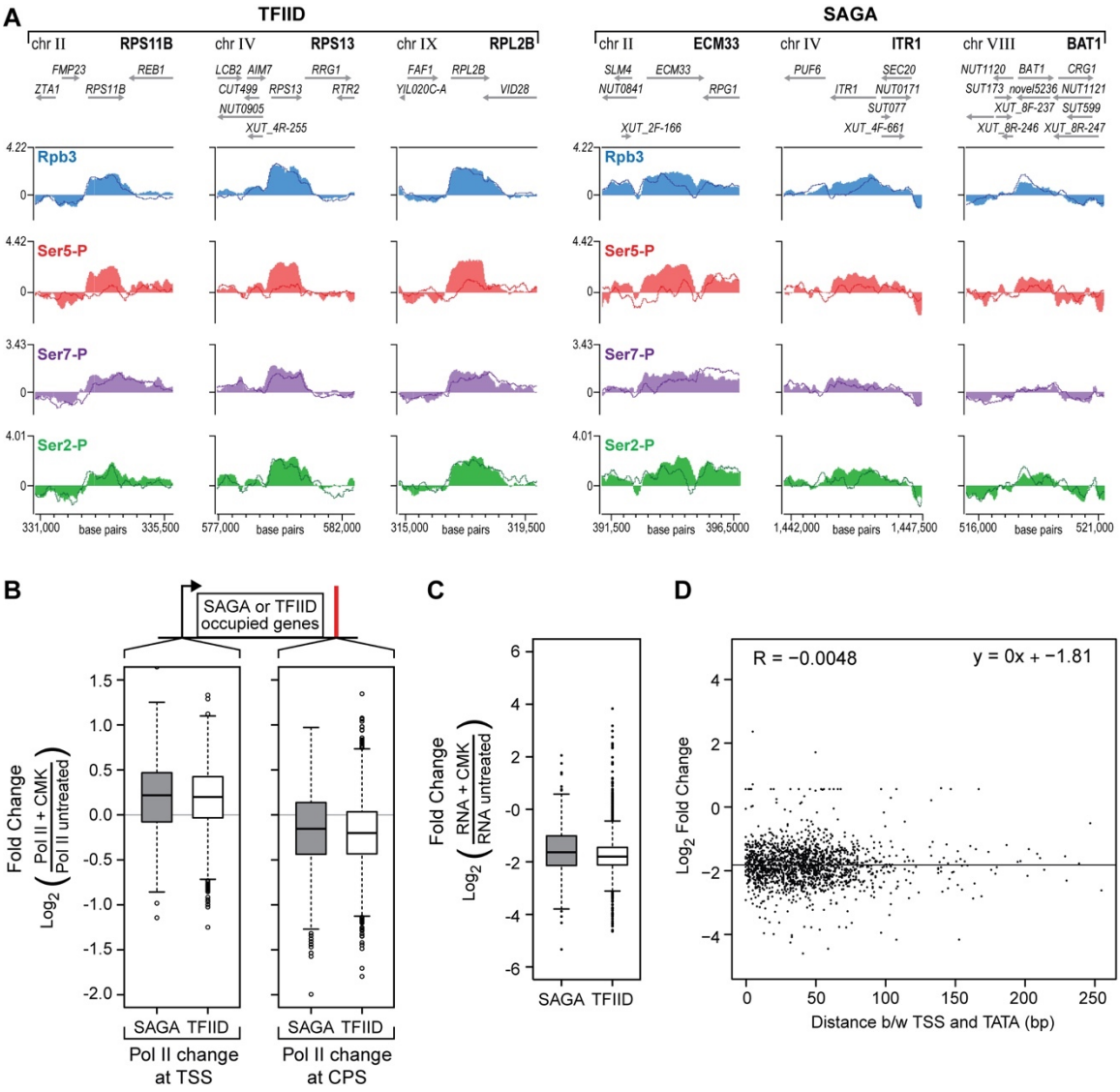



**Figure 2.9 Comparison of Changes in Pol II Occupancy and RNA-Seq Coverage at TFIID versus SAGA Gene Classes**

**(A)** Single gene traces of Rpb3 (blue), Ser5-P (red), Ser7-P (purple), and Ser2-P (green) marks from untreated *kin28is* cells (filled bars) or treated with 5 μM CMK (lines) at TFIID and SAGA genes. **(B)** Pol II (Rpb3-TAP) occupancy changes at the TSS or the CPS of TFIID or SAGA genes [66] are not significantly different. **(C)** However, the distribution of gene expression changes (RNA-Seq of Nascent fraction) is significantly different between TFIID and SAGA genes (Mann-Whitney-Wilcoxon Test, $P < 0.01$). **(D)** The expression change after inhibition of Kin28is for each TFIID gene was plotted against its respective TATA-TSS distance [43]. Outliers in TATA-TSS distance were removed (TATA-TSS distances >50,000 bp). The equation of the best-fit line and Pearson correlation coefficient are shown.

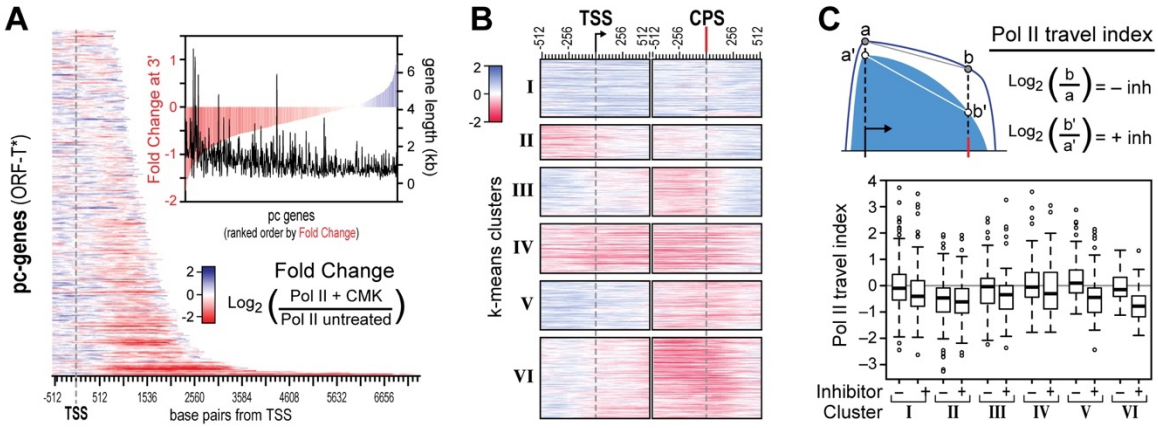



**Figure 2.10 Kin28 Protects the Elongation Complex from Premature Termination**

**(A)** Change in Pol II (Rpb3-TAP) after Kin28is inhibition with 5 μM CMK arranged by gene length (same genes as averaged in Figure 2.7B). **(B)** k-means clustering of changes in Pol II occupancy at the TSS and the CPS of genes in (A). **(C)** Pol II travel index describes Pol II occupancy at the CPS relative to the TSS. A negative Pol II travel index indicates there is less Pol II at the CPS than there was loaded onto the TSS.

normalization approaches that rely on a handful of model genes that do not reflect the full range of transcriptome perturbations in cells [22]. Thus, if Kin28is inhibition were to cause global reduction in transcription, it would be reliably detected by our analyses. We have consistently used spike in controls to account for global changes in the yeast transcriptome [19], contrary to assertions which claim otherwise [22].

Unexpectedly, the data revealed that Total RNA levels remain generally unchanged, whereas Nascent RNA levels show nearly 4-fold decrease upon covalent inhibition of Kin28is (**Figure 2.11**). To determine the number of genes that were significantly downregulated upon covalent inhibition of Kin28is, we employed edgeR and DESeq, two widely used statistical packages. Although both generally agreed on the effect of Kin28is inhibition on Total RNA (edgeR: 283 downregulated genes, DESeq: 643 downregulated genes, with ≥2-fold decrease, 5% FDR), we found edgeR and DESeq differed in their analysis of the Nascent RNA data. Analysis of the raw Nascent RNA data clearly shows a global decrease in abundance upon covalent inhibition of Kin28is. However, DESeq, but not edgeR, preserved this initial observation (edgeR: 166 versus DESeq: 4,321, with ≥2-fold decrease, 5% FDR, out of 5,390 pc-genes) (**Figure 2.12**). Indeed, in simulation studies, edgeR and other commonly used analysis methods performed poorly when there was global reduction in transcript abundance [45]. We conclude that covalent inhibition of Kin28is leads to downregulation of the majority of the Nascent RNA fraction.

For both Total and Nascent RNA fractions, irreversible inhibition of Kin28is with CMK causes the largest overall change in RNA abundance (**Figure 2.12 and 2.13**). Treatment of WT cells with either 1-NA-PP1 or CMK did not result in any significant changes in Total or Nascent RNA levels (**Figure 2.13**). Likewise, non-covalent inhibition of Kin28is or Kin28as with 1-NA-PP1 did not result in significant changes in RNA levels, consistent with previous reports (**Figure 2.13A**) [19]. These results reconcile previous

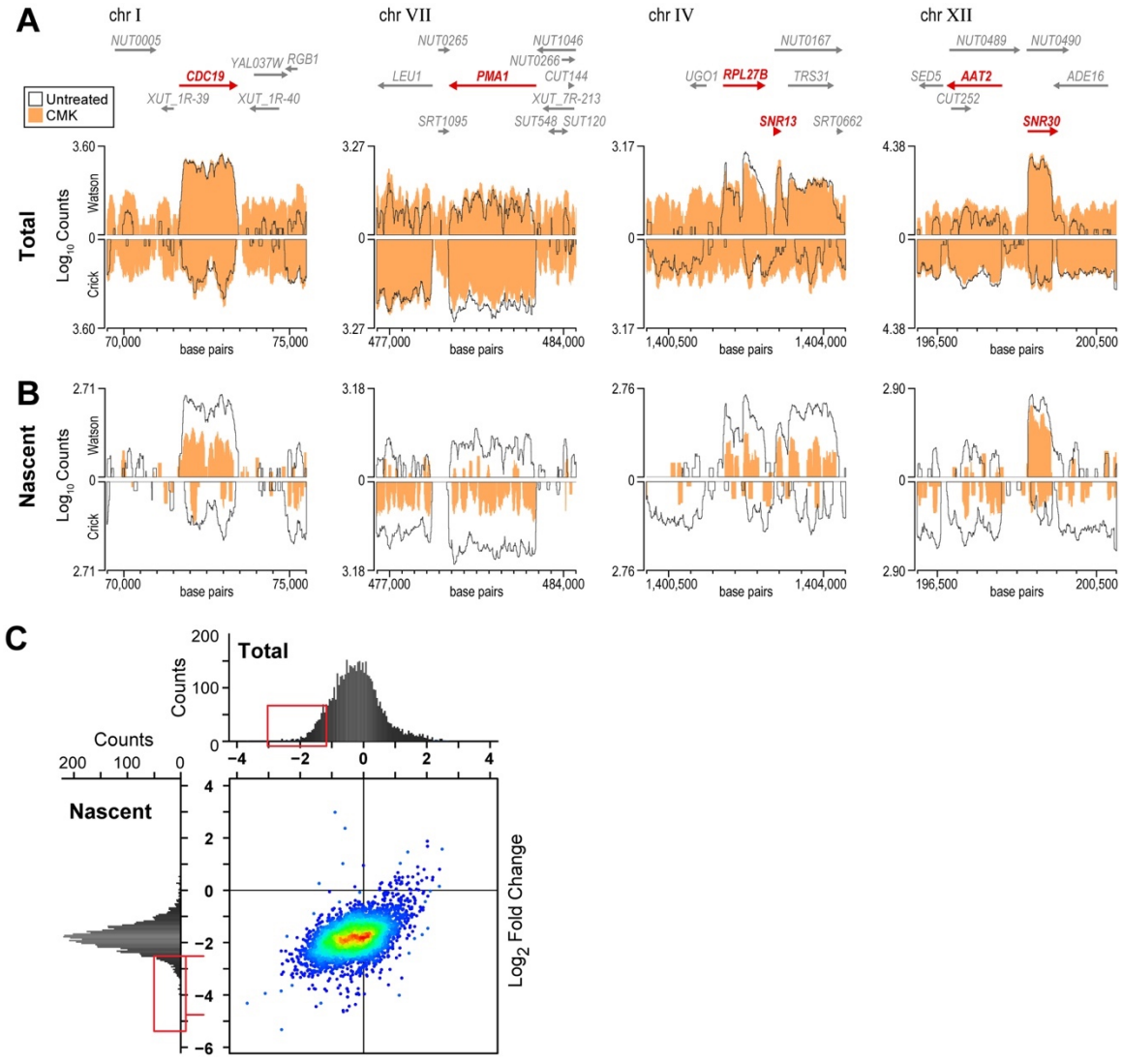



**Figure 2.11 Inhibition of Kin28is Leads to Stabilization of mRNA Abundance and Reduction in Nascent Transcripts**

Coverage of RNA-Seq reads at the indicated genes for the **(A)** Total and **(B)** Nascent fractions of RNA from untreated (black line) and CMK-treated (orange fill) *kin28is* cells. **(C)** Scatterplot of expression changes in the Total (x axis) and Nascent (y axis) fraction plotted for each pc-gene. The distributions of expression changes are plotted as histograms, adjacent. Differentially expressed genes with greater than 2-fold decrease, as called by edgeR, are boxed in red.

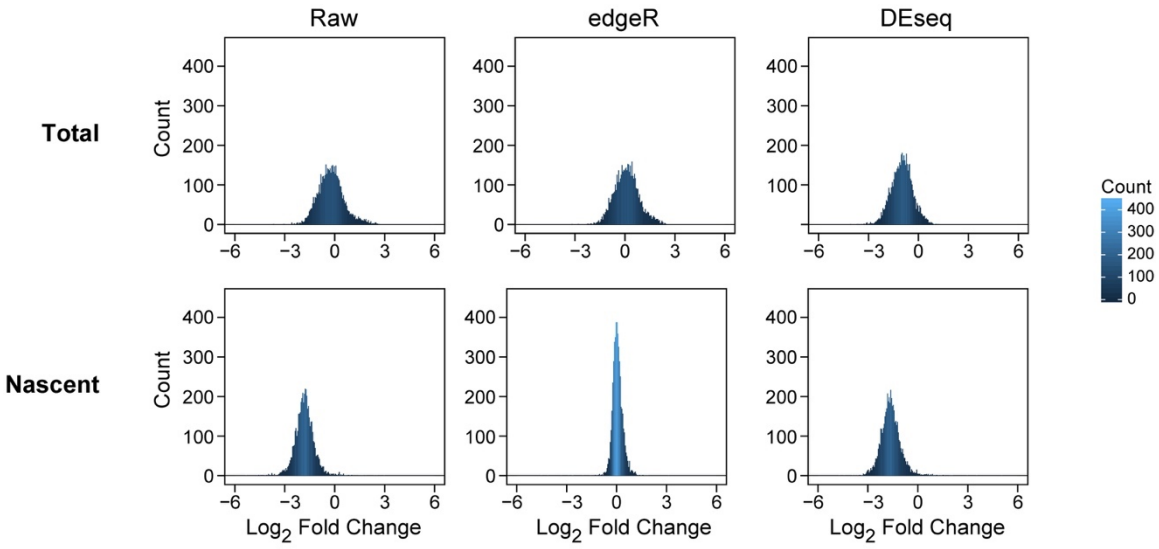



**Figure 2.12 Results from Differential Gene Expression Analysis Vary by Software Used**

Comparison of read count fold-change values in CMK-treated versus untreated cells using commonly used R packages for calculating differential gene expression analysis.

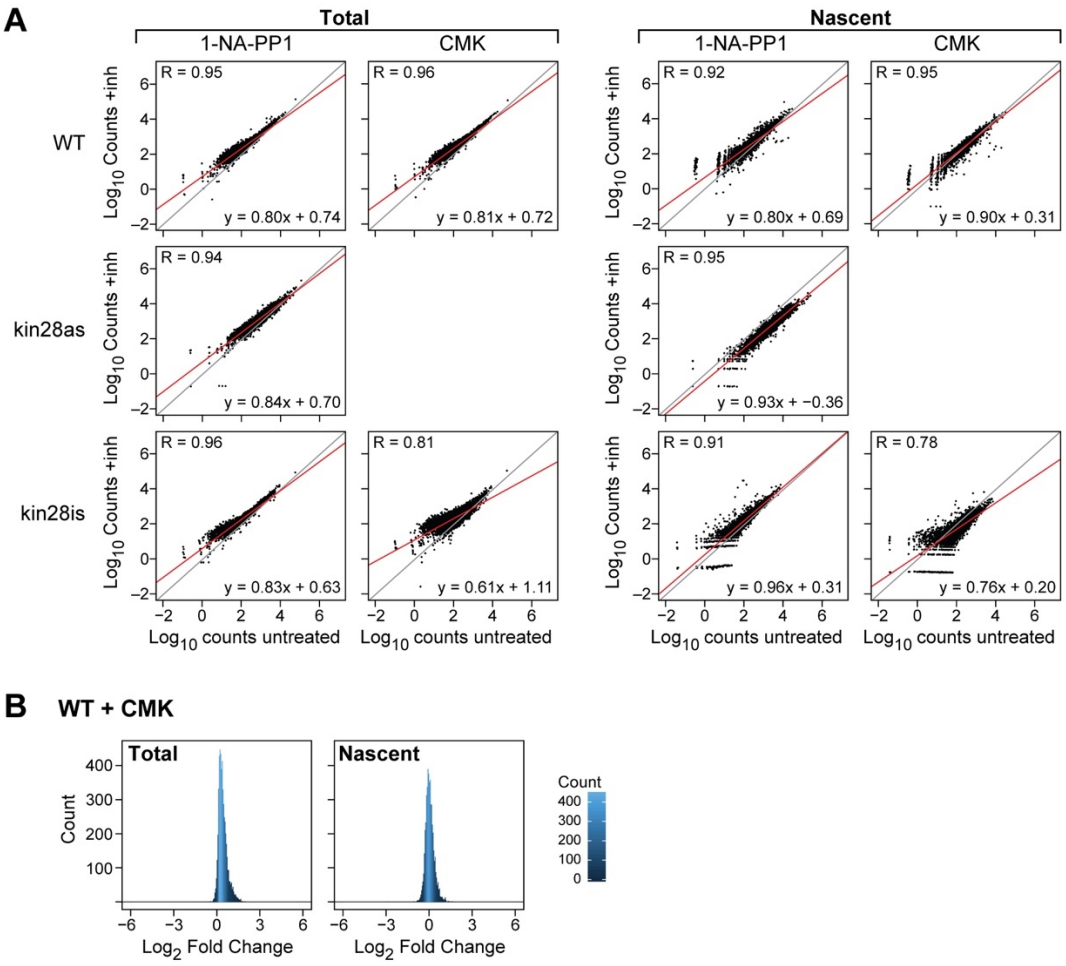



**Figure 2.13 Irreversible Inhibition of Kin28is Impacts mRNA Expression Globally**

**(A)** Plot of read count (Log10) per pc-gene in untreated versus inhibited cells. Points along the gray line of slope = 1 represent genes that do not change in expression between treatment conditions (absolute fold change relative to DMSO ≤ 2, 5% FDR). The equation of the regression line (red) and Pearson correlation coefficient are displayed. **(B)** Distribution of expression changes for WT cells treated with CMK does not significantly affect RNA levels in the Total or the Nascent fractions.

discrepancies and indicate that different transcriptional processes depend on different degrees of CTD phosphorylation, with capping requiring maximal levels of phosphorylation, whereas promoter escape and transcription elongation can occur at low levels of Kin28-dependent CTD phosphorylation.

### 2.8 Highly Repressed Genes Reveal a Checkpoint Prior to Productive Elongation

Among the genes with the most pronounced decrease in Pol II travel index, we observed genes with Pol II accumulation 100–200 nt downstream of the TSS (**Figure 2.14**). Likewise, Pol II profiles at the genes most severely affected in their Total and Nascent pools of RNA revealed accumulation of Pol II, which coincides with the +2 nucleosome and the transition from initiation to elongation (**Figure 2.15**) [23, 34-37, 46]. This Pol II accumulation is consistent with a checkpoint at the onset of productive elongation after successful transcription initiation and promoter escape. The Pol II buildup also coincides with the region of maximal Spt5 recruitment, which occurs upstream of maximal Bur1 and Paf1 occupancy [47, 48]. Phosphorylation of Spt5 and Pol II by Bur1 contributes to efficient recruitment of the Paf1 complex, which is a critical step for efficient transcription elongation [49-51]. To further define the nature of the defect in the transition to elongation, we examined the levels of Spt5 and phosphorylated Spt5 (Spt5-P) at two representative genes (**Figure 2.16**). The data show that inhibiting Kin28 significantly reduces the level of Spt5-P, indicative of compromised Bur1 function.

### 2.9 Impact of Pol II Alleles with Different Elongation Rates

To test whether a functional link between Kin28 activity and the rate of Pol II elongation exists, we examined the impact of Kin28 inhibition in strains bearing Pol II alleles with either fast or slow elongation rates. We first replaced WT Kin28 with the *kin28is* allele in the anchor-away system [52]. In this system, rapamycin-induced

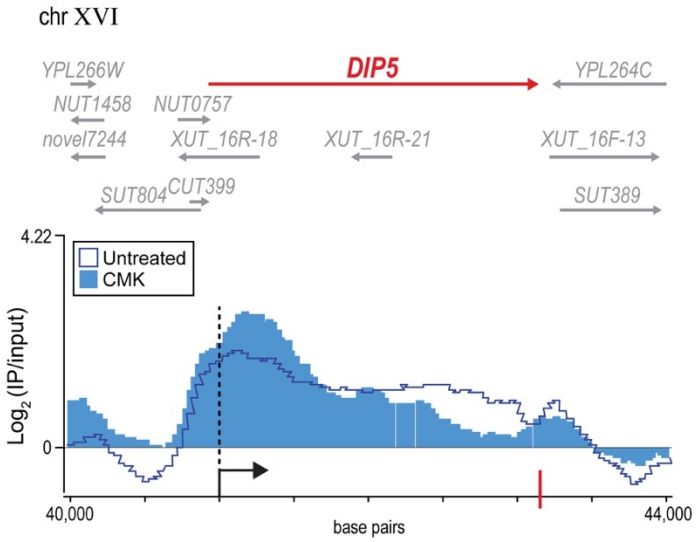



**Figure 2.14 Pol II Traces at *DIP5***

Pol II traces at *DIP5* before (line) and after (filled) inhibition of Kin28is.

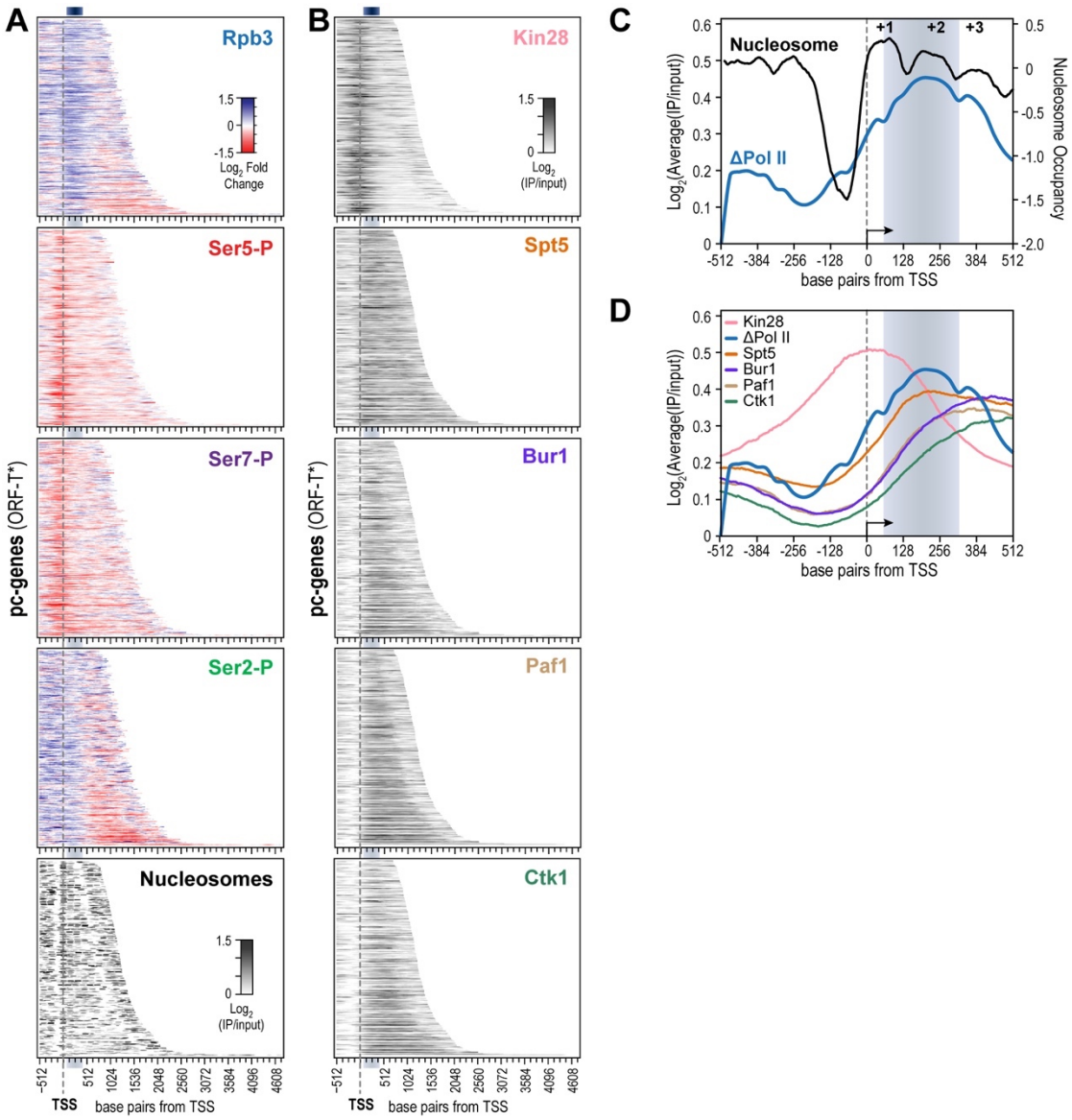



**Figure 2.15 Kin28 Plays a Crucial Role in Priming Pol II for Elongation**

**(A)** Difference in Rpb3, Ser5-P, Ser7-P, and Ser2-P occupancy after inhibition of Kin28is with 2 μM CMK at genes most affected by loss of Kin28 function (*i.e.* genes boxed in Figure 2.11C) and sorted by gene length. Following a decrease in Ser5-P and Ser7-P marks, Rpb3 ChIP signal accumulates on average 177 bp downstream of the TSS (top blue bar) and coincides with the +2 nucleosome (bottom panel). **(B)** Occupancy of factors recruited during initiation and elongation [47] **(C)** Average occupancy traces for Pol II and nucleosome data shown in (A). **(D)** Average occupancy traces for Pol II and data shown in (B). Vertical dotted line denotes TSS. Blue highlighted regions denote the site of Pol II stalling.

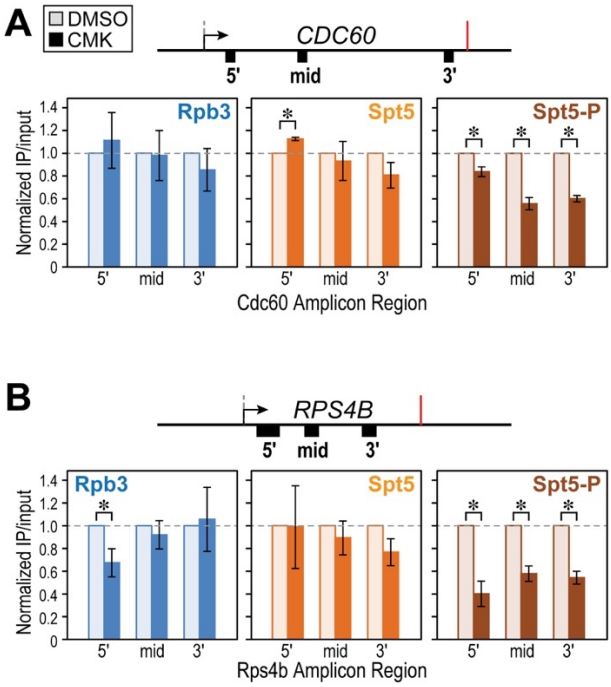



**Figure 2.16 Levels of Spt5 and Phosphorylated Spt5 (Spt5-P) at Two Representative Genes**

ChIP-qPCR of Rpb3 (blue), Spt5 (orange), and Spt5-P (maroon) at the indicated regions (5′, mid, and 3′) of **(A)** CDC60 and **(B)** RPS4B. Average fold changes (n = 3−6) relative to DMSO control are plotted. ∗P < 0.05, one sample two-sided Student's t test.

heterodimerization of an Rpb1-FRB fusion with RPL13A-FKBP12 actively exports Rpb1 from the nucleus due to the normal nucleo-cytoplasmic trafficking of RPL13A (**Figure 2.17A**). Expression of Rpb1 variants in this system permits the replacement of endogenous WT Pol II with Pol II variants with reduced/slow (*rpb1-N488D*) or enhanced/fast (*rpb1-E1103G*) elongation rates [38]. The slow Pol II variant is thought to hinder initiation or early elongation and should therefore exacerbate the effects of irreversible inactivation of Kin28 and show a negative genetic interaction with sub-optimal levels of Kin28 activity. Consistent with a shared function in elongation, *kin28is* cells grown in otherwise sub-lethal concentrations of CMK (0.2 μM) displayed a synthetic lethal phenotype when combined with the slow Pol II variant (**Figure 2.17B**). Reciprocally, we predicted the fast Pol II variant would be less sensitive to Kin28is inhibition. In further agreement, the fast variant of Pol II did not display the synthetic lethal phenotype when Kin28is activity was dialed down with sub-lethal concentrations of CMK. Remarkably, the fast Pol II partially rescued the temperature-sensitive phenotype of *kin28is* when grown at 37 °C. In contrast, WT and slow Pol II strains were unable to grow under these conditions. Thus, our data reveal a role for Kin28 activity in placing marks that facilitate transition from transcription initiation to efficient transcription elongation.

## 2.10 Discussion

To resolve the long-standing discrepancies on the function of Kin28/CDK7, we rationally designed a sensitized variant of Kin28 (*kin28is*) that can be inhibited specifically and irreversibly in vivo by CMK. To achieve targeted inhibition, CMK requires the simultaneous presence of two selectivity filters, a small gatekeeper residue and a precisely positioned cysteine in the ATP-binding site. Conveniently, none of the

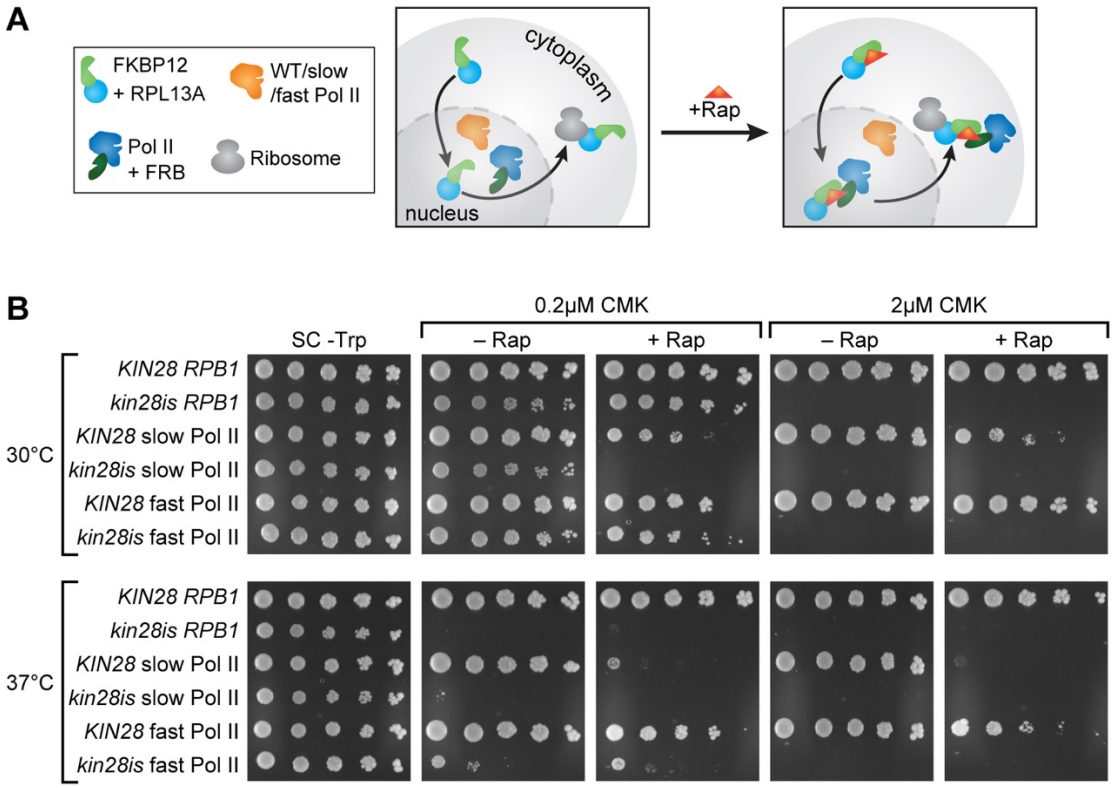



**Figure 2.17 Functional Interaction between Kin28 Activity and Pol II Elongation Rates**

**(A)** Scheme of anchor-away strategy used to exchange endogenous Pol II fused to FKBP12-rapamycin-binding protein (Pol II + FRB) with a slow (*rpb1-N488D*), fast (*rpb1-E1103G*), or WT Pol II fused to human FKBP12. Rapamycin (Rap) was added to anchor-away endogenous Rpb1 while mutant Pol II remains in the nucleus. **(B)** Top: a sub-lethal dose of CMK (0.2 μM) was combined with WT, slow Pol II, or fast Pol II in a *KIN28* or *kin28is* background. Inhibited *kin28is* shows a synthetic lethal interaction with a Pol II mutant with reduced elongation rate. Bottom: at 37 °C, 0.2 μM CMK is lethal for *kin28is* cells bearing WT or slow Pol II. Exchange of endogenous Pol II for a fast Pol II partially restores viability to *kin28is* (compare *kin28is* RPB1 to *kin28is* fast Pol II). However, presence of a fast Pol II is insufficient to rescue *kin28is* viability at higher concentrations of CMK (2 μM).

endogenous kinases in budding yeast have both sequence elements, enabling selective inhibition of the doubly mutated *kin28is* allele.

We find that Kin28 plays a crucial role at multiple stages of Pol II transcription and that when the kinase is chemically inhibited in vivo, the impact on transcription is masked by increased stability of mature mRNA. Furthermore, we find Kin28 enhances transcription elongation by priming the CTD and enabling the engagement of complexes that help overcome an “elongation checkpoint” well after promoter release and transcription initiation (**Figure 2.18**). Thus, by irreversibly inhibiting Kin28 with a covalent ligand, we have effectively dialed down Kin28 activity past a physiologically relevant threshold previously unattained with non-covalent inhibitors. By achieving more complete inactivation, our study has revealed the essential role of Kin28 catalytic activity at different stages of the transcription cycle.

We conclude that differential levels of kinase inhibition reveal distinct transcriptional processes that display differing degrees of dependence on CTD phosphorylation states. Promoter-proximal recruitment of Ceg1 and 5′ capping of nascent pre-mRNAs is highly sensitive to partial Kin28 inhibition and partial reduction of CTD phosphorylation. The heightened sensitivity of Ceg1 for any perturbation in Kin28 activity may arise in part from the kinetic competition between transcribing Pol II and 5′ surveillance factors. This competition may restrict the essential function of the capping enzyme to a narrow window of opportunity immediately following initiation. Thus, capping defects are readily detectable when using non-covalent inhibition strategies against Kin28 [19]. On the other hand, further reduction in Kin28-dependent Ser5-P and Ser7-P levels diminishes the formation of elongation-competent complexes. Upon irreversible inhibition of Kin28, the position of Pol II buildup coincides with proposed “checkpoints” that are thought to coordinate 5′ capping of nascent transcripts with productive elongation and splicing [34-37]. This checkpoint was not detected by previous

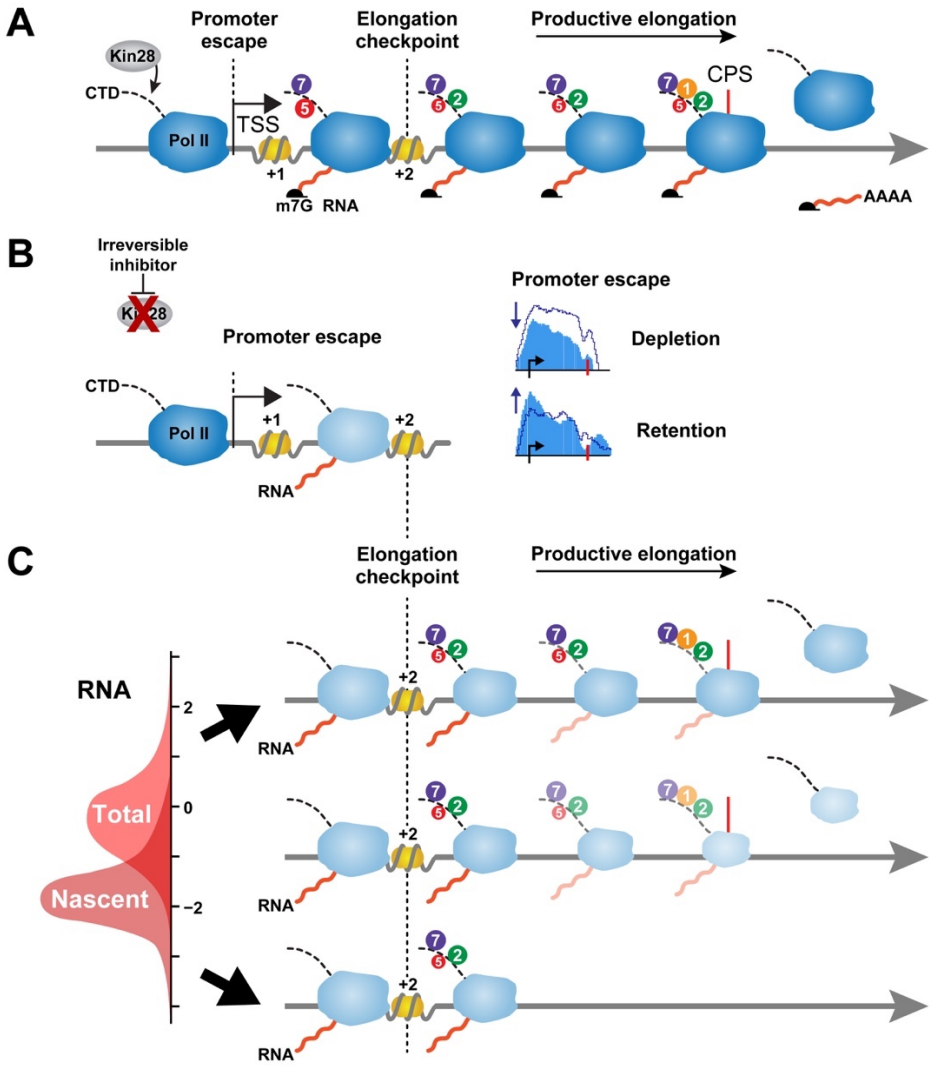



**Figure 2.18 Model Depicting the Various Roles of Kin28 during Transcription, which Are Revealed by Covalent Inhibition of Kin28is**

**(A)** Normal Pol II progression through the transcription cycle when Kin28 is active. **(B)** Upon covalent inactivation of Kin28is, Pol II is depleted to different degrees at the TSS of the majority (82%) of highly transcribed genes and retained at a smaller fraction of genes (11%). **(C)** Left: although Total RNA is generally unaffected by covalent inhibition of Kin28is, Nascent RNA is globally decreased, consistent with stabilization of the Total RNA pool upon a global decrease in RNA synthesis. Right: genes that show decrease in Total and Nascent RNA also show delayed transition through the elongation checkpoint, as evidenced by a buildup of Pol II ~177 bp from the TSS. These genes also display attrition of Pol II toward their 3′ end.

genome-wide chromatin immunoprecipitation studies of Pol II but is readily evident upon covalent Kin28 inhibition.

Consistent with a role for Kin28 in productive elongation, we find that upon Kin28 inactivation, Pol II abundance at the 3′ end of genes correlates negatively with gene length. Interestingly, a slow-transcribing Pol II (*rpb1-N488D*) shows a synthetic lethal interaction with Kin28, suggesting a functional link between Pol II elongation rates and Kin28 activity. In line with a role for Kin28 in early transcription elongation, recruitment of the cap binding complex is required for efficient recruitment of Bur1 and Ctk1, proper levels of Ser2-P marks, and efficient elongation [25, 34-36, 50]. Fast Pol II, on the other hand, is proposed to display transcription rates that may bypass elongation checkpoints and compromise co-transcriptional processing [53, 54]. Our observations suggest a critical role for Kin28 in priming Pol II for elongation competency by facilitating timely recruitment of elongation and pre-mRNA processing factors and protecting the elongation complex from capping surveillance factors that would remove Pol II molecules carrying uncapped pre-mRNAs [55]. Our findings are also consistent with a model wherein Kin28 activity primes Pol II and Spt5 for downstream phosphorylation by Bur1. These modifications recruit elongation complexes required to overcome a transition checkpoint imposed by the +2 nucleosome. This “kinetic” checkpoint contrasts with the stable promoter-proximal Pol II pause in mammalian systems that occurs within 100 bp of the TSS and is mediated by the Negative Elongation Factor (NELF) and DSIF (Spt4/5). While NELF does not exist in yeast and the kinetic checkpoint is detected downstream of the mammalian pause site, both mechanisms present a rate-limiting step prior to transition into productive transcription elongation.

It is important to note that Nascent RNA synthesis decreases upon covalent inhibition of Kin28is but is not dramatically impaired upon partial inhibition with a reversible inhibitor. Unexpectedly, steady-state RNA levels remain mostly unaffected for

an extended period after covalent inhibition of Kin28is, reflecting increased stabilization of the cellular mRNA pool in response to defective transcription by Pol II. This dichotomous response to Kin28 inhibition supports a built-in safeguard against defective Pol II transcription [44, 56, 57]. Our findings help reconcile initial observations that suggested an essential role for Kin28 in transcription and more recent findings showing that mRNA levels remain unaffected when Kin28 activity is compromised by certain mutations [16, 58] or blocked with reversible ligands [5, 19-23].

Pol II occupancy is not affected to the same degree as Nascent RNA upon covalent inhibition of Kin28. This is reminiscent of reports wherein inhibition of general transcription using the *rpb1-1* allele shut down RNA synthesis within 5–15 min after switching to non-permissive temperatures [59], yet Pol II remained robustly associated with chromatin for an additional hour [60, 61]. Similarly, under certain nutrient conditions, Pol II association with transcribed regions increases subtly even though the rate of Nascent RNA synthesis is dramatically reduced [56]. Therefore, although RNA synthesis is reduced, remaining Pol II occupancy may reflect a combination of slowly transcribing, backtracking, or inactively engaged Pol II. Taken together, our results converge on multiple roles for Kin28/CDK7 in regulating post-initiation stages of the transcriptional cycle and reconcile long-standing controversies.

Furthermore, unlike kinase-specific inhibitors that require exhaustive and cost-prohibitive small molecule screens and may have off-target effects due to non-specific interaction with related kinases [17], our approach has the potential to be broadly applied to other kinases. Our results clearly demonstrate that the active site of a kinase can be rationally sensitized to irreversible inhibition and that CMK does not perturb cellular growth rates, CTD phosphorylation patterns, or the transcriptome in wild-type cells. Importantly, our strategy faithfully recapitulates the phenotype of a catalytically inactive *kin28* allele and achieves a greater degree of inhibition than reversible ligands. These

outcomes are likely to extend to other kinases as our structurally and bioinformatically guided engineering of Kin28 can be readily applied to confer “irreversible sensitization” to thiol-reactive ATP analogs. Thus, the strategy that we describe here may serve as a resource for the broader scientific community studying kinases, as well as ATP-binding proteins and enzymes that are amenable to rationally designed chemical-genetic retrofitting.

## 2.11 Materials and Methods

*Kin28is Allele Design and Isolation*

The gatekeeper (L83) and conserved valine (V21) residues of Kin28 were identified by aligning the primary sequence of Kin28 to that of Cdc5 which has a naturally occurring solvent exposed Cys in place of the conserved Val using ClustalW2 (http://www.ebi.ac.uk/Tools/msa/clustalw2/). Protein sequences were gathered from SGD. After identification of sites for mutagenesis, we used previously reported plasmids encoding kin28-L83G (*pSH573*) [20] to create the *kin28is* allele. Primers were designed to direct the V21C substitution in *KIN28* following QuickChange II manufacturing instructions (formerly Stratagene, now Agilent Technologies). PCR and DpnI digestion of template DNA was also carried out following manufacturer instructions. Purified plasmids were sequenced to confirm mutagenesis at the University of Wisconsin-Madison Biotechnology Center sequencing facility. The endogenous *KIN28* gene was substituted with its *kin28is* counterpart by two-step allele replacement into a BY4705 strain.

To increase recombination efficiency for subsequent strains we amplified the *kin28is* genomic locus from the *kin28is* strain (above) with increased up- and downstream sequence using the KpnI-Kin28 Fwd and BamHI-Kin28 Rev primers. We subsequently cloned the resulting amplicon into the KpnI and BamHI sites of *pRS306* to create *pSCT001*. *pSCT001* was digested upstream of the *KIN28* coding region

with *BmgBI* and transformed into the Rpb3-TAP (Open Biosystems, now Thermo Fisher) or HHY170 background [52]. Cells were selected for *URA3*+ transformants. After recombination, loss of *URA3* was screened on SC+5FOA and potential *kin28is* colonies were screened by growing them on rich media (YPD) or media containing CMK (YPD+CMK). Colonies that displayed a growth defect on YPD+CMK, but not on YPD, were selected and confirmed by sequencing to carry the modified gatekeeper and the reactive Cys.

*Growth Curves*

Cells were grown overnight in respective dropout media and sub-cultured to an OD600 of 0.1 two to three hours before the assay. When the OD600 was between 0.2 and 0.4, cells were plated onto a 96-well clear, flat-bottom polystyrene plate in a volume of 100 µL. Drugs were added directly to each well, taking care that the volume of the drug did not exceed 0.5% of the total volume in the well. Cells were incubated with constant shaking at 30 °C in an Infinite M1000 Pro (Tecan) plate reader. OD600 measurements were recorded every 5 min over the duration of the assay (12-16 hours) and plotted as growth curves. $IC_{50}$ was calculated by plotting *kin28is* percentage growth (i.e. $OD_{max}/OD_{initial}$ normalized to growth in DMSO) as a function of drug concentration. The drug concentration at which 50% growth was achieved was calculated from a logarithmic trend line that was fitted to the data.

*Extracts and Antibodies for Western Blotting*

Yeast strains were grown overnight (14-16 hours) in YPD and subcultured in 50 mL of YPD at a starting OD600 of 0.1 and allowed to grow to exponential phase (OD600 0.5-1). Unless otherwise noted, cells were treated with 5 µM 1-NA-PP1 (Toronto Research Chemicals), 5 µM CMK or an equivalent volume of DMSO for 1 hour. For CMK

titrations, cells were treated with the indicated concentrations of CMK for 1 hour as well. Cells were harvested by centrifugation, and lysis was conducted by bead-beating in lysis buffer (100 mM HEPES pH 7.4, 100 mM potassium acetate, 0.5% Triton X-100, 0.1% Tween-20), Protease Inhibitor Cocktail (Roche), 0.2 mM PMSF, and 5 mM of phosphatase inhibitors $Na_3VO_4$, NaF, and $NaN_3$. Lysates were treated with 25 U of DNase I (Sigma) and incubated on ice for 20 min before addition of SDS loading buffer. In the case of H14, extracts were first incubated with anti-Rpb3 (Neoclone, 1:400) overnight at 4 °C and immunoprecipitated with Protein G Dynabeads, washed in lysis buffer and eluted with SDS loading buffer. After Rpb3 enrichment, supernatant was used to probe for bulk Kin28 levels. Samples were resolved by 10% SDS-polyacrylamide gel electrophoresis. Anti-phosphorylated CTD antibodies were generous gifts from Dr. Dirk Eick (1:500), except anti-Ser5p (Covance H14, 1:000) anti-Ser2-P (S2 antibody Bethyl Laboratories, 1:2500). Rpb3-TAP levels were probed with anti-TAP (Pierce CAB1001, 1:2000) and Kin28 levels were probed with anti-Kin28 (BioLegend 924901, 1:250).

*PCR-mediated FLAG-tag Fusion of Med8*

A *FLAGx5::KanMX6* cassette was fused to the C-terminal end of Med8 using the *pFA6a-FLAGx5* plasmid (a gift from Dr. John Pringle) as template following a published protocol [62]. A PCR amplicon bearing the in frame FLAGx5 tag and the G418 resistance cassette (*KANMX6*) was amplified using *pFA6a*-MED8-Ctermtag-F2 and *pFA6a*-Med8-Ctermtag-R1. The ~2kb fragment was transformed into *kin28is* cells following the standard LiAc protocol [63]. After 6 hours of outgrowth, cells were plated on YPD + 500 µg/mL G418 and after 3-4 days at 30 °C, G418 resistant colonies were screened by colony PCR (Ling et al., 1995) for the correct insertion. The *pFA6a* specific primer, pFA6a insertion check, and the Med8 specific primer, Med8 +301 Fwd were used to screen G418 resistant colonies for the appropriate insertion. Med8-FLAGx5 was

confirmed by Sanger sequencing through the UW Biotechnology Center sequencing facility.

*Chromatin Immunoprecipitation and Microarray Hybridization*

ChIP experiments were carried out essentially as described previously (Chinchilla et al., 2012; Tietjen et al., 2010) with minor modifications. Briefly, cells in mid-log phase were treated with 2 μM or 5 μM 1-NA-PP1 or CMK (as indicated) for 1 hour followed by crosslinking with 1% formaldehyde for 5min. Pulldowns were carried out with the following amounts of antibody: Rpb3-TAP, with 3 μl of anti-TAP antibody (Thermo), Ser5-P marks, with 3 μl of 3E8 (a generous gift from D. Eick), Ser7-P marks with 50 μl of 4D12 (Chapman et al., 2007), Ser2-P marks, with S2 antibody (Bethyl Laboratories), Med8-FLAG with 3 μl of anti-FLAG (M2, Sigma), Spt5 with 3 μl of anti-Spt5 (a generous gift from Dr. Grant Hartzog) and phosphorylated Spt5 with 3 μl of anti-Spt5-P (a generous gift from Dr. Steven Hahn) overnight. For every immunoprecipitation, 50ul of protein G conjugated Dynabeads (Life Technologies), pre- washed with lysis buffer (50 mM HEPES-KOH [pH 7.5], 140 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% Na-Deoxycholate, 1 mM, Aprotinin, 1 mM Leupeptin, 1 mM PMSF, 1 mM Benzamidine, 1.45 mM Pepstatin, 1mM $NaN_3$, 1 mM $Na_3VO_4$, 1 mM NaF), were mixed with the whole cell extract-antibody mixture to pulldown Pol II complexes. Washes and elution were performed as previously described (Tietjen et al., 2010). ChIP DNA was end-repaired, linkers were ligated and ligation-mediated PCR was performed as previously described (Chinchilla et al., 2012).

*ChIP-chip Data Analysis*

Single gene traces were gathered using Integrated Genome Viewer (IGV) (Thorvaldsdottir et al., 2013). Changes in Pol II, Ser2-P, Ser5-P and Ser7-P were

calculated for genes with the top 10% of Pol II occupancy (**Figure 2.9A**) or for genes which showed a significant decrease in Total or Nascent RNA (**Figure 2.13**) and arranged by length using R. k-means clustering was performed as previously described (Tietjen et al., 2010). Briefly the changes in Pol II occupancy at the TSS and CPS for genes with the top 10% of Pol II occupancy were initially clustered in 10 k- means clusters in Multiple Experiment Viewer using an uncentered Pearson correlation and visually inspected. General Pol II change patterns coalesced to 6 distinct patterns. Pol II 3' end decrease (**Figure 2.9**) was calculated by taking the Log2 of the average change in Pol II occupancy within the last 200bp of pc genes with the top 10% highest Pol II occupancy (Tietjen et al., 2010). Pearson correlation statistic between Pol II 3' end decrease and gene length were calculated in R. To further test if the correlation between the change in 3' Pol II occupancy and gene length was non-random, we performed the Pearson correlation statistic with $10^4$ randomized iterations of the data and calculated the average p-value. Pol II travel index was calculated by taking the ratio of Pol II occupancy at the cleavage and polyadenylation site (CPS) (b) vs. Pol II occupancy at the transcription start site (TSS) (a) for DMSO (b/a) and CMK (b'/a') conditions. Likewise, Promoter Escape and Elongation coefficients were calculated by taking the ratio of Pol II at the TSS (a'/a) and CPS (b'/b) upon inhibition of Kin28 respectively. Position of Pol II stalling with respect to the TSS was calculated by finding the mode of the distribution of maximum Pol II change at genes whose Total or Nascent RNA levels decreased significantly.

*RNA Preparation and Next Generation Sequencing*

Total and nascent (*i.e.* labeled) RNA was prepared as described in (Miller et al., 2011) but was used in an RNA-Seq rather than microarray format. For cDTA, cells were treated essentially as described in (Sun et al., 2012) with a labeling time of 6 min with 4-

thiouracil (Sigma). *S. pombe* was harvested and labeled identically to the experimental samples, and was added in a 1:6 ratio to *S. cerevisiae* prior to lysis. The strand-specific RNA-Seq libraries were constructed with the ScriptSeq Complete Yeast Kit (Epicentre). Sequencing (1 x 100 bp) was performed on an Illumina HiSeq2000 sequencer following standard Illumina clustering and quantification protocols. Reads were de- multiplexed and quick statistics generated with FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc). Reads were simultaneously clipped of 3' adapters and aligned to a concatenated version of the most recent *S. cerevisiae* (SacCer3, SGD, http://www.yeastgenome.org/) and *S. pombe* genome (ASM294v.2.23, PomBase, http://www.pombase.org/) with STAR 2.3.0 (Dobin et al., 2013). Reads mapping to more than one locus were discarded to prevent ambiguous alignment of sequences shared between species. Alignments were then subset by species with SAMtools and visualized with IGV (Li et al., 2009).

*RNA-Seq Data Normalization and Differential Gene Expression Analysis*

Because we cannot assume total RNA levels are comparable between cells with and without Kin28 function, we could not normalize by library size as is common practice. Instead, we utilized an *S. pombe* spike-in to normalize and compare RNA expression between samples. The number of reads, which mapped to each annotated gene, was counted using HTSeq 0.6.0 (Anders et al., 2015). Annotation files were obtained from the same databases from where the genome sequences used in initial read alignment were downloaded. *S. cerevisiae* read counts were normalized with loess regression using *S. pombe* read counts to fit the loess as described by (Loven et al., 2012). To scale coverage plots, the slope of the best fit line between *S. pombe* genes for each total and nascent sample was scaled to the largest slope for total and nascent fractions, respectively. Subsequently, these scaling factors were used to scale *S.*

*cerevisiae* genome-wide coverage plots. Differential expression of annotated regions was calculated with edgeR (Robinson et al., 2010), DESeq (Anders and Huber, 2010), or manually with Excel and the R qvalue package (Bass JDSwcfAJ et al., 2015).

*Anchor Away Assays*

HHY170 and HHY170 *kin28is* were transformed with plasmids bearing WT (*pL-RPB1*), slow (*pL-rpb1-N488D*) or fast (*pL-rpb1-E1103G*) variants of RNA Polymerase II. Cells were grown in SC -Trp overnight (typically 16-20 hours) at 30 °C. Cultures were serially diluted by a factor of 0.5 and plated on SC -Trp media alone or containing rapamycin (1 µg/mL), 0.2 µM or 2 µM CMK or respective CMK +rapamycin combinations. Cells were incubated at 30 °C or 37 °C for 3-4 days and photographed to score for growth and viability. At least three independent biological replicates were performed.

**Table 2.1 Read Counts in Nascent RNA Fractions**

Replicate 1

| Strain | Treatment | All Reads | Aligned | *Sc* Reads | *Sp* Reads |
|---|---|---|---|---|---|
| **WT** | DMSO | 7,716,701 | 6,766,911 | 6,387,702 | 379,209 |
| | 1-NAPP-1 | 19,231,330 | 16,664,541 | 15,553,964 | 1,110,577 |
| | CMK | 6,557,336 | 5,149,371 | 4,819,252 | 330,119 |
| **kin28as** | DMSO | 8,554,860 | 7,010,016 | 6,512,344 | 497,672 |
| | 1-NAPP-1 | 3,863,661 | 3,461,856 | 2,795,611 | 666,245 |
| **kin28is** | DMSO | 12,784,391 | 10,305,178 | 5,766,954 | 4,538,224 |
| | 1-NAPP-1 | 10,696,662 | 705,501 | 473,197 | 232,304 |
| | CMK | 8,051,363 | 1,425,733 | 624,130 | 801,603 |

Replicate 2

| Strain | Treatment | All Reads | Aligned | *Sc* Reads | *Sp* Reads |
|---|---|---|---|---|---|
| **WT** | DMSO | 12,364,023 | 1,038,016 | 809,998 | 228,018 |
| | 1-NAPP-1 | 13,475,974 | 7,439,215 | 5,976,996 | 1,462,219 |
| | CMK | 11,022,435 | 8,631,108 | 7,004,567 | 1,626,541 |
| **kin28is** | DMSO | 5,602,058 | 4,773,654 | 3,134,732 | 1,638,922 |
| | 1-NAPP-1 | 5,574,112 | 4,366,742 | 2,584,020 | 1,782,722 |
| | CMK | 3,764,657 | 1,551,544 | 667,602 | 883,942 |

Replicate 3

| Strain | Treatment | All Reads | Aligned | *Sc* Reads | *Sp* Reads |
|---|---|---|---|---|---|
| **kin28is** | DMSO | 7,531,965 | 6,421,525 | 4,221,258 | 2,200,267 |
| | 1-NAPP-1 | 7,720,014 | 6,051,090 | 3,582,496 | 2,468,594 |
| | CMK | 6,048,661 | 2,497,509 | 1,074,166 | 1,423,343 |

Replicate 4

| Strain | Treatment | All Reads | Aligned | *Sc* Reads | *Sp* Reads |
|---|---|---|---|---|---|
| **kin28is** | CMK | 5,048,144 | 3,750,264 | 960,318 | 2,789,946 |

Replicate 5

| Strain | Treatment | All Reads | Aligned | *Sc* Reads | *Sp* Reads |
|---|---|---|---|---|---|
| **kin28is** | CMK | 7,458,865 | 5,543,416 | 1,421,922 | 4,121,494 |

Aligned refers to uniquely aligned reads to either the *S. cerevisiae* (*Sc*) or *S. pombe* (*Sp*) genome.

**Table 2.2 Read Counts in Total RNA Fractions**

Replicate 1

| Strain | Treatment | All Reads | Aligned | *Sc* Reads | *Sp* Reads |
|---|---|---|---|---|---|
| **WT** | DMSO | 6,717,512 | 5,211,778 | 4,466,352 | 745,426 |
| | 1-NAPP-1 | 6,372,173 | 4,532,813 | 3,793,781 | 739,032 |
| | CMK | 4,910,988 | 3,687,483 | 3,089,200 | 598,283 |
| **kin28as** | DMSO | 11,066,959 | 8,614,187 | 7,230,601 | 1,383,586 |
| | 1-NAPP-1 | 7,499,402 | 4,250,509 | 3,426,600 | 823,909 |
| **kin28is** | DMSO | 8,961,864 | 6,375,489 | 5,432,326 | 943,163 |
| | 1-NAPP-1 | 9,625,429 | 7,259,703 | 6,167,925 | 1,091,778 |
| | CMK | 19,303,394 | 14,962,674 | 12,706,714 | 2,255,960 |

Replicate 2

| Strain | Treatment | All Reads | Aligned | *Sc* Reads | *Sp* Reads |
|---|---|---|---|---|---|
| **WT** | DMSO | 9,980,891 | 8,758,874 | 5,841,791 | 2,917,083 |
| | 1-NAPP-1 | 16,728,332 | 13,136,172 | 9,602,654 | 3,533,518 |
| | CMK | 13,808,827 | 11,912,806 | 8,882,815 | 3,029,991 |
| **kin28is** | DMSO | 18,515,063 | 14,459,419 | 8,397,047 | 6,062,372 |
| | 1-NAPP-1 | 11,534,266 | 8,386,767 | 5,826,290 | 2,560,477 |
| | CMK | 13,079,359 | 8,978,793 | 6,079,555 | 2,899,238 |

Aligned refers to uniquely aligned reads to either yeast genome.

**Table 2.3 *S. cerevisiae* to *S. pombe* Counts in Mock Nascent RNA Pulldowns**

| Strain | Treatment | Total | Aligned* | *Sc* Reads | *Sp* Reads |
|---|---|---|---|---|---|
| **WT** | DMSO (-4tU, Total) | 13,840,989 | 10,812,142 | 8,308,803 | 2,503,339 |
| **WT** | DMSO (-4tU, Nascent) | 10,502,224 | 4,111,410 | **848,378** | **3,263,032** |

Nascent RNA enrichment from cells without 4tU show a substantially lower proportion of *S. cerevisiae* to spike-in reads.

**Table 2.4 Yeast Strains Used in This Chapter**

| Strain | Relevant Genotype | Reference |
|---|---|---|
| WT | *MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 RPB3-TAP::HIS3* | [67] |
| BY4705 | *MATα, ade2::hisG, his3Δ200, leu2Δ0, lys2Δ0, met15Δ, trp1Δ63, ura3Δ0* | [68] |
| kin28is Rpb3-TAP | *MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 RPB3-TAP::HIS3 kin28 L83G, V21C* | This study |
| kin28is | *MATα, ade::hisG, his3Δ200, leu2Δ0, lys2Δ0, met15Δ, trp1Δ63, ura3Δ0, kin28 L83G, V21C* | This study |
| kin28is Med8-FLAG | *MATα, ade::hisG, his3Δ200, leu2Δ0, lys2Δ0, met15Δ, trp1Δ63, ura3Δ0, kin28 L83G, V21C MED8-FLAGx5::kanMX6* | This study |
| HHY170 | *MATα tor1-1 fpr::NAT RPL13A-2xFKB12::TRP1 RPO21-FRB::kanMX6* | [52] |
| HHY170 kin28is | *MATα tor1-1 fpr::NAT RPL13A-2xFKB12::TRP1 RPO21-FRB::kanMX6 kin28 L83G, V21C* | This study |
| KIN28 RPB1 | *MATα tor1-1 fpr::NAT RPL13A-2xFKB12::TRP1 RPO21-FRB::kanMX6 pL-RPB1* | This study |
| kin28is RPB1 | *MATα tor1-1 fpr::NAT RPL13A-2xFKB12::TRP1 RPO21-FRB::kanMX6 kin28 L83G, V21C pL-RPB1* | This study |
| KIN28 slow Pol II | *MATα tor1-1 fpr::NAT RPL13A-2xFKB12::TRP1 RPO21-FRB::kanMX6 pL-rpb1-N488D* | This study |
| kin28is slow Pol II | *MATα tor1-1 fpr::NAT RPL13A-2xFKB12::TRP1 RPO21-FRB::kanMX6 kin28 L83G, V21C pL-rpb1-N488D* | This study |
| KIN28 fast Pol II | *MATα tor1-1 fpr::NAT RPL13A-2xFKB12::TRP1 RPO21-FRB::kanMX6 pL-rpb1-E1103G* | This study |
| kin28is fast Pol II | *MATα tor1-1 fpr::NAT RPL13A-2xFKB12::TRP1 RPO21-FRB::kanMX6 kin28 L83G, V21C pL-rpb1-E1103G* | This study |

**Table 2.5 Plasmids Used in This Chapter**

| Plasmid | Backbone | Reference |
|---|---|---|
| *pSH573* | *pRS306* | (Liu et al., 2004) |
| *pSCT001 (pRS306 kin28is)* | *pRS306* | This study |
| *pL-RPB1* | *pRS415* | (Malagon et al., 2006) |
| *pL-rpb1-N488D* | *pRS415* | (Malagon et al., 2006) |
| *pL-rpb1-E1103G* | *pRS415* | (Malagon et al., 2006) |
| *pFA6a-FLAGx5* | | (Longtine et al., 1998) |

**Table 2.6 Oligonucleotides Used in This Chapter**

| Name | Sequence |
|---|---|
| ADH1 -196 Fwd | CGGTATACGGCCTTCCTTCCAG |
| ADH1 +13 Rev | CTGGGATAGACATTGTATATGAGATAGTTGATTGTATG |
| CDC60-5'F | GTTCGGCTTCCCTCTTTAGTT |
| CDC60-5'R | GAGCTTTGTTTCCATTGGGTTT |
| CDC60-midF | TGCCGTAGTATTCAGGTTTGTG |
| CDC60-midR | GTCATTGCTACCAAGGGTACTG |
| CDC60-3'F | GCTTGTGCTGCCTTCTTAAC |
| CDC60-3'R | AAGCCAGAGGATGTGTTTGA |
| RPS4B-5'F | CCCGCAATGCTTATCTGTTACT |
| RPS4B-5'R | CCCTCTCATGATTGGTACAAGTG |
| RPS4B-midF | GAGGCTTCTTCATCGGTGATAC |
| RPS4B-midR | GTTTCATGGACGTCATCACTCT |
| RPS4B-3'F | AGACGAAGAGAGGGCAGATAA |
| RPS4B-3'R | GCCAAAGGGTAAGGGTATCA |
| pFA6a-MED8-C-term tag-F2 | AAGCCCATTATTACGGGATCTACTTCAACATCATCTAGTAATCGGATCCCCGGGTTAATT |
| pFA6a Med8 C-term tag-R1 | CAATAAAATGTTTTAAAATATCAGTAGCGTGAAGTAAAATGAATTCGAGCTCGTTTAAAC |
| pFA6a insertion check | CCTGAGAAAGCAACCTGACC |
| Med8 +301 Fwd | TTCCCAACCACCTCTCATGAAAGTTTGG |
| Kin28 V21C SDM Fwd | GGTGAGGGTACTTATGCGGTTTGTTACTTGGGTTGTCAACACTC |
| Kin28 V21C SDM Rev | GAGTGTTGACAACCCAAGTAACAAACCGCATAAGTACCCTCACC |
| KpnI-Kin28 Fwd | TAAGCAGGTACCGTTGTGCCTGTAGGT |
| BamHI-Kin28 Rev | TGCTTAGGATCCTCAGTTACGTATTTT |

## 2.12 References


1. Egly, J.M. and F. Coin, *A history of TFIIH: Two decades of molecular biology on a pivotal transcription/repair factor.* DNA Repair, 2011. **10**(7): p. 714-721.

2. Lee, T.I. and R.A. Young, *Transcription of eukaryotic protein-coding genes.* Annual Review of Genetics, 2000. **34**: p. 77-137.

3. Liu, X., D.A. Bushnell, and R.D. Kornberg, *RNA polymerase II transcription: Structure and mechanism.* Biochimica Et Biophysica Acta-Gene Regulatory Mechanisms, 2013. **1829**(1): p. 2-8.

4. Roeder, R.G., *The role of general initiation factors in transcription by RNA polymerase II.* Trends in Biochemical Sciences, 1996. **21**(9): p. 327-335.

5. Akhtar, M.S., et al., *TFIIH Kinase Places Bivalent Marks on the Carboxy-Terminal Domain of RNA Polymerase II.* Molecular Cell, 2009. **34**(3): p. 387-393.

6. 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**(20): p. 5455-5464.

7. Kim, M., et al., *Phosphorylation of the Yeast Rpb1 C-terminal Domain at Serines 2, 5, and 7.* Journal of Biological Chemistry, 2009. **284**(39): p. 26421-26426.

8. Buratowski, S., *Progression through the RNA Polymerase II CTD Cycle.* Molecular Cell, 2009. **36**(4): p. 541-546.

9. Corden, J.L., *RNA Polymerase II C-Terminal Domain: Tethering Transcription to Transcript and Template.* Chemical Reviews, 2013. **113**(11): p. 8423-8455.

10. Eick, D. and M. Geyer, *The RNA Polymerase II Carboxy-Terminal Domain (CTD) Code.* Chemical Reviews, 2013. **113**(11): p. 8456-8490.

11. Perales, R. and D. Bentley, *"Cotranscriptionality": The Transcription Elongation Complex as a Nexus for Nuclear Transactions.* Molecular Cell, 2009. **36**(2): p. 178-191.

12. Phatnani, H.P. and A.L. Greenleaf, *Phosphorylation and functions of the RNA polymerase II CTD.* Genes Dev, 2006. **20**(21): p. 2922-36.

13. Zhang, D.W., et al., *Emerging Views on the CTD Code.* Genet Res Int, 2012. **2012**: p. 347214.

14. Cismowski, M.J., et al., *Kin28 Encodes a C-Terminal Domain Kinase That Controls Messenger-Rna Transcription in Saccharomyces-Cerevisiae but Lacks Cyclin-Dependent Kinase-Activating Kinase (Cak) Activity.* Molecular and Cellular Biology, 1995. **15**(6): p. 2983-2992.

15. Valay, J.G., et al., *The Kin28 Gene Is Required Both for Rna-Polymerase-Ii Mediated Transcription and Phosphorylation of the Rpb1p Ctd.* Journal of Molecular Biology, 1995. **249**(3): p. 535-544.

16. Keogh, M.C., et al., *Kin28 is found within TFIIH and a Kin28-Ccl1-Tfb3 trimer complex with differential sensitivities to T-loop phosphorylation.* Molecular and Cellular Biology, 2002. **22**(5): p. 1288-1297.

17. Kwiatkowski, N., et al., *Targeting transcription regulation in cancer with a covalent CDK7 inhibitor.* Nature, 2014. **511**(7511): p. 616-620.

18. Rodriguez, C.R., et al., *Kin28, the TFIIH-associated carboxy-terminal domain kinase, facilitates the recruitment of mRNA processing machinery to RNA polymerase II.* Molecular and Cellular Biology, 2000. **20**(1): p. 104-112.

19. Kanin, E.I., et al., *Chemical inhibition of the TFIIH-associated kinase Cdk7/Kin28 does not impair global mRNA synthesis.* Proceedings of the National Academy of Sciences of the United States of America, 2007. **104**(14): p. 5812-5817.

20. 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**(4): p. 1721-1735.

21. 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.* Molecular Cell, 2012. **45**(2): p. 158-170.

22. Hong, S.W., et al., *Phosphorylation of the RNA polymerase II C-terminal domain by TFIIH kinase is not essential for transcription of Saccharomyces cerevisiae genome.* Proceedings of the National Academy of Sciences of the United States of America, 2009. **106**(34): p. 14276-14280.

23. Kim, H., et al., *Gene-specific RNA polymerase II phosphorylation and the CTD code.* Nature Structural & Molecular Biology, 2010. **17**(10): p. 1279-+.

24. Lee, K.M., et al., *Impairment of the TFIIH-associated CDK-activating kinase selectively affects cell cycle-regulated gene expression I in fission yeast.* Molecular Biology of the Cell, 2005. **16**(6): p. 2734-2745.

25. Viladevall, L., et al., *TFIIH and P-TEFb Coordinate Transcription with Capping Enzyme Recruitment at Specific Genes in Fission Yeast.* Molecular Cell, 2009. **33**(6): p. 738-751.

26. Wong, K.H., Y. Jin, and K. Struhl, *TFIIH phosphorylation of the Pol II CTD stimulates mediator dissociation from the preinitiation complex and promoter escape.* Mol Cell, 2014. **54**(4): p. 601-12.

27. Tietjen, J.R., et al., *Chemical-genomic dissection of the CTD code.* Nature Structural & Molecular Biology, 2010. **17**(9): p. 1154-U16.

28. Akoulitchev, S., et al., *Requirement for Tfiih Kinase-Activity in Transcription by Rna-Polymerase-Ii.* Nature, 1995. **377**(6549): p. 557-560.

29. Helenius, K., et al., *Requirement of TFIIH kinase subunit Mat1 for RNA Pol II C-terminal domain Ser5 phosphorylation, transcription and mRNA turnover.* Nucleic Acids Research, 2011. **39**(12): p. 5025-5035.

30. Larochelle, S., et al., *Cyclin-dependent kinase control of the initiation-to-elongation switch of RNA polymerase II.* Nature Structural & Molecular Biology, 2012. **19**(11): p. 1108-+.

31. Nilson, K.A., et al., *THZ1 Reveals Roles for Cdk7 in Co-transcriptional Capping and Pausing.* Molecular Cell, 2015. **59**(4): p. 576-587.

32. Chymkowitch, P., et al., *Cdc28 kinase activity regulates the basal transcription machinery at a subset of genes.* Proceedings of the National Academy of Sciences of the United States of America, 2012. **109**(26): p. 10450-10455.

33. Adelman, K. and J.T. Lis, *Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans.* Nature Reviews Genetics, 2012. **13**(10): p. 720-731.

34. Gornemann, J., et al., *Cotranscriptional spliceosome assembly occurs in a stepwise fashion and requires the cap binding complex.* Molecular Cell, 2005. **19**(1): p. 53-63.

35. Hossain, M.A., et al., *The Yeast Cap Binding Complex Modulates Transcription Factor Recruitment and Establishes Proper Histone H3K36 Trimethylation during Active Transcription.* Molecular and Cellular Biology, 2013. **33**(4): p. 785-799.

36. Lidschreiber, M., K. Leike, and P. Cramer, *Cap Completion and C-Terminal Repeat Domain Kinase Recruitment Underlie the Initiation-Elongation Transition of RNA Polymerase II.* Molecular and Cellular Biology, 2013. **33**(19): p. 3805-3816.

37. Pokholok, D.K., N.M. Hannett, and R.A. Young, *Exchange of RNA polymerase II initiation and elongation factors during gene expression in vivo.* Molecular Cell, 2002. **9**(4): p. 799-809.

38. Malagon, F., et al., *Mutations in the Saccharomyces cerevisiae RPB1 gene conferring hypersensitivity to 6-azauracil.* Genetics, 2006. **172**(4): p. 2201-2209.

39. Cohen, M.S., et al., *Structural bioinformatics-based design of selective, irreversible kinase inhibitors.* Science, 2005. **308**(5726): p. 1318-1321.

40. Snead, J.L., et al., *A coupled chemical-genetic and bioinformatic approach to polo-like kinase pathway exploration.* Chemistry & Biology, 2007. **14**(11): p. 1261-1272.

41. Robinson, P.J.J., et al., *Structure of the Mediator Head module bound to the carboxy-terminal domain of RNA polymerase II.* Proceedings of the National

Academy of Sciences of the United States of America, 2012. **109**(44): p. 17931-17935.

42. Jeronimo, C. and F. Robert, *Kin28 regulates the transient association of Mediator with core promoters.* Nature Structural & Molecular Biology, 2014. **21**(5): p. 449-455.

43. Murakami, K., et al., *Uncoupling Promoter Opening from Start-Site Scanning.* Molecular Cell, 2015. **59**(1): p. 133-138.

44. Sun, M., et al., *Comparative dynamic transcriptome analysis (cDTA) reveals mutual feedback between mRNA synthesis and degradation.* Genome Research, 2012. **22**(7): p. 1350-1359.

45. Soneson, C. and M. Delorenzi, *A comparison of methods for differential expression analysis of RNA-seq data.* Bmc Bioinformatics, 2013. **14**.

46. Albert, I., et al., *Translational and rotational settings of H2A.Z nucleosomes across the Saccharomyces cerevisiae genome.* Nature, 2007. **446**(7135): p. 572-576.

47. Mayer, A., et al., *Uniform transitions of the general RNA polymerase II transcription complex.* Nature Structural & Molecular Biology, 2010. **17**(10): p. 1272-+.

48. Segal, E., et al., *A genomic code for nucleosome positioning.* Nature, 2006. **442**(7104): p. 772-778.

49. 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**(17): p. 4852-4863.

50. Qiu, H.F., et al., *Pol II CTD kinases Bur1 and Kin28 promote Spt5 CTR-independent recruitment of Paf1 complex.* Embo Journal, 2012. **31**(16): p. 3494-3505.

51. Yu, M., et al., *RNA polymerase II-associated factor 1 regulates the release and phosphorylation of paused RNA polymerase II.* Science, 2015. **350**(6266): p. 1383-1386.

52. Haruki, H., J. Nishikawa, and U.K. Laemmli, *The anchor-away technique: Rapid, conditional establishment of yeast mutant phenotypes.* Molecular Cell, 2008. **31**(6): p. 925-932.

53. Hazelbaker, D.Z., et al., *Kinetic Competition between RNA Polymerase II and Sen1-Dependent Transcription Termination.* Molecular Cell, 2013. **49**(1): p. 55-66.

54. Jimeno-Gonzalez, S., et al., *Rat1p maintains RNA polymerase II CTD phosphorylation balance.* Rna-a Publication of the Rna Society, 2014. **20**(4): p. 551-558.

55. Chang, J.H., et al., *Dxo1 is a new type of eukaryotic enzyme with both decapping and 5 '-3 ' exoribonuclease activity.* Nature Structural & Molecular Biology, 2012. **19**(10): p. 1011-U62.

56. Pelechano, V., et al., *Regulon-Specific Control of Transcription Elongation across the Yeast Genome.* Plos Genetics, 2009. **5**(8).

57. Haimovich, G., et al., *Gene Expression Is Circular: Factors for mRNA Degradation Also Foster mRNA Synthesis.* Cell, 2013. **153**(5): p. 1000-1011.

58. Komarnitsky, P., E.J. Cho, and S. Buratowski, *Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription.* Genes Dev, 2000. **14**(19): p. 2452-60.

59. Nonet, M., et al., *Eukaryotic Rna-Polymerase Conditional Mutant That Rapidly Ceases Messenger-Rna Synthesis.* Molecular and Cellular Biology, 1987. **7**(5): p. 1602-1611.

60. Kim, T.S., et al., *RNA polymerase mapping during stress responses reveals widespread nonproductive transcription in yeast.* Genome Biology, 2010. **11**(7).

61. Zanton, S.J. and B.F. Pugh, *Full and partial genome-wide assembly and disassembly of the yeast transcription machinery in response to heat shock.* Genes & Development, 2006. **20**(16): p. 2250-2265.

62. Longtine, M.S., et al., *Additional modules for versatile and economical PCR-based gene deletion and modification in Saccharomyces cerevisiae.* Yeast, 1998. **14**(10): p. 953-961.

63. Gietz, R.D. and R.H. Schiestl, *Quick and easy yeast transformation using the LiAc/SS carrier DNA/PEG method.* Nature Protocols, 2007. **2**(1): p. 35-37.

64. Chapman, R.D., et al., *Transcribing RNA polymerase II is phosphorylated at CTD residue serine-7.* Science, 2007. **318**(5857): p. 1780-2.

65. Cho, E.J., et al., *Opposing effects of Ctk1 kinase and Fcp1 phosphatase at Ser 2 of the RNA polymerase II C-terminal domain.* Genes Dev, 2001. **15**(24): p. 3319-29.

66. Huisinga, K.L. and B.F. Pugh, *A genome-wide housekeeping role for TFIID and a highly regulated stress-related role for SAGA in Saccharomyces cerevisiae.* Molecular Cell, 2004. **13**(4): p. 573-585.

67. Ghaemmaghami, S., et al., *Global analysis of protein expression in yeast.* Nature, 2003. **425**(6959): p. 737-741.

68. Brachmann, C.B., et al., *Designer deletion strains derived from Saccharomyces cerevisiae S288C: a useful set of strains and plasmids for PCR-mediated gene disruption and other applications.* Yeast, 1998. **14**(2): p. 115-132.

# Chapter 3: Defining Post-transcriptional Consequences of Kin28 Inhibition

Parts of this chapter are adapted from Tseng SC*, McClean MN, Steinmetz LM, Ansari AZ, Inhibition of a Transcriptional Kinase Triggers Coordinated Inhibition of Protein Translation and mRNA Decay, *intended for publication*.

### 3.1 Introduction

The TFIIH kinase, Kin28 in budding yeast and CDK7 in mammals, has critical roles in regulating transcription and pre-mRNA processing events. Notably, Kin28/CDK7 phosphorylates the fifth and seventh serine residues (Ser5/7) of the C-terminal domain (CTD) of RNA polymerase II (Pol II) [1-3], transitioning Pol II from a promoter-bound pre-initiation complex to an elongation-competent form [4-6]. CDK7 has received renewed attention as a chemotherapeutic target because of the success of the inhibitor, THZ1, in selectively killing tumor cells through covalent conjugation with a cysteine residue outside of the ATP-binding pocket of CDK7 [7]. However, it is somewhat of a conundrum that inhibition of CDK7, which has essential functions in all cells, is particularly harmful only to cancer cells. The current theory for the propensity of cancer cells to be “addicted” to CDK7 hinges on the notion that cancer cells uniquely require continued, high-level expression of oncogenes to maintain their cancerous state [8-11].

We have previously shown that inhibition of Kin28 dramatically attenuates nascent RNA synthesis, but because of concurrent decrease in decay of pre-existing mRNAs, the steady-state mRNA levels actually remained somewhat constant [12]. The phenomenon of maintenance of mRNA steady-state levels in response to insults to transcription or decay has been termed “RNA buffering” and has been shown to be mediated by the cytoplasmic 5’-to-3’ exonuclease Xrn1 [13, 14]. Interestingly, a small set of mRNAs were not buffered, and their stability was highly dependent on Kin28 activity.

In this current study, we sought to determine why certain mRNAs are not buffered in cells that encounter transcriptional stress. Our analyses reveal unstable mRNAs are often highly expressed, have short open reading frames (ORFs), and encode proteins destined for the endoplasmic reticulum (ER) or cell periphery. While buffering is mediated by inactivation of Xrn1, we find that non-buffered mRNAs are generally unstable in other situations of defective transcription, and surprisingly,

translation. Of particular note, is our finding that the buffering status of an mRNA can be predicted from the ratio of Nab2-to-Ski2 association.

The stabilization of extant RNA pools upon experiencing stress is well documented [15]. However, it was not clear whether irreversible inhibition of Kin28 elicits known stress signaling pathways to buffer extant mRNA pools from degradation. Importantly, whether the buffered mRNAs are capped and translated in inhibited Kin28 cells was, prior to this study, unclear. Contrary to expectations, we discovered that inhibition of Kin28 reduces translation globally, but not from lack of m7G or canonical stress-related pathways. Rather, we show mRNAs are rapidly sequestered into P-bodies and identify classes of mRNAs that escape translation repression. Moreover, we find that selective preference for Ski2 over Nab2 correlates with stabilization of mRNA levels. Our results underscore the close coupling of transcription initiation in the nucleus with translation and RNA decay in the cytoplasm by a quintessential general transcription factor. Understanding the various mechanisms by which Kin28 regulates gene expression in addition to its canonical role in transcription initiation may benefit potential CDK7-based therapies, as well as expand our understanding of the web of regulatory decisions that couple nuclear and cytoplasmic processes and stabilize extant RNA pools under situations of cellular stress.

### 3.2 The Transcriptomic Landscape of Irreversibly Inhibited Kin28 Cells Resembles That of Defective Translation

To study the transcriptomic and post-transcriptomic consequences of inhibiting Kin28, we utilized our chemical genetic strategy to covalently and potently inhibit this kinase in vivo. Two point mutations in KIN28 irreversibly sensitize the kinase (*kin28is*) to an ATP analog that covalently reacts via a chloromethylketone (CMK) moiety without discernable off-target effects [12]. Transient transcriptome sequencing (TT-Seq), an

RNA-Seq method that combines 4-thiouracil labeling and biotinylation with fragmentation was utilized because it reduces the 5' bias in reads during capture of nascent transcripts [16]. We also improved the efficiency of recovering labeled transcripts by utilizing a more reactive biotinylation reagent [17]. Differential gene expression analysis of our new TT-Seq dataset closely correlated with our previous RNA-Seq datasets (**Figure 3.1**). Therefore, we utilized the data from all RNA-Seq libraries in subsequent bioinformatic analyses.

In agreement with our previous findings [12], synthesis of transcripts after inhibition of Kin28 decreased significantly (median fold-change = 0.12), while the steady-state level of mRNA was much less affected (median fold-change = 0.80) (**Figure 3.2A**). This is indicative of RNA buffering [13, 14]. Transcripts that are most reduced in quantity, *i.e.*, unstable, by inhibition of Kin28 tend to be highly-expressed (**Figure 3.2B**) and possess short ORFs (**Figure 3.2C**). This is in contrast to the inherent tendency for longer mRNAs to have faster degradation rates [18, 19]. The few transcripts that are induced or unaffected when Kin28 is inhibited have longer 5' and 3' UTR lengths compared to the transcripts that are buffered or decrease in expression (**Figure 3.2D**). Interestingly, there is little correlation between the change in transcript levels upon inhibition of Kin28 with previously reported mRNA half-lives [18, 20, 21], synthesis rates [20], and 5'/3'-UTR lengths [22] (**Table 3.1**).

Next, we wondered if the buffered and unstable mRNAs in inhibited Kin28 cells were the same in other situations of attenuated transcription. We performed correlation analysis between the expression changes in inhibited *kin28is* and the expression changes of cells in 52 other instances where buffering was examined [14], which included the transcriptome of decay factor deletion strains, cells treated with a general transcription inhibitor (1,10-phenanthroline), and cells treated with a low dose of translation elongation inhibitor cycloheximide (CHX). Inhibition of Kin28 had similar

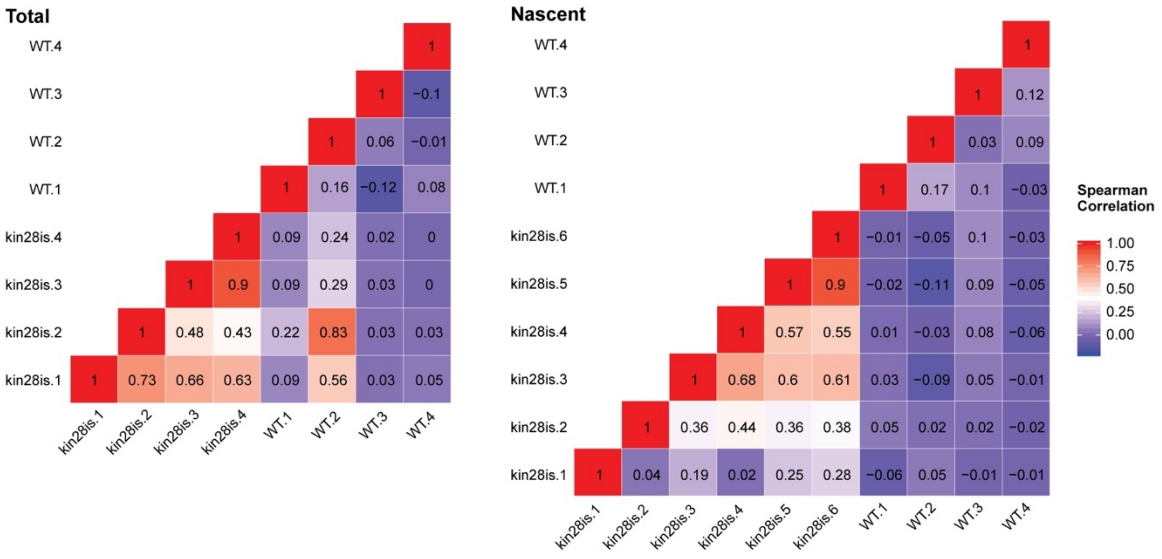



**Figure 3.1 Correlation Analysis of Expression Changes between RNA-Seq Replicates**

Correlation analysis of expression changes in wild-type (WT) and *kin28is* cells treated with inhibitor CMK presented as a heatmap of pairwise Spearman correlation coefficients.

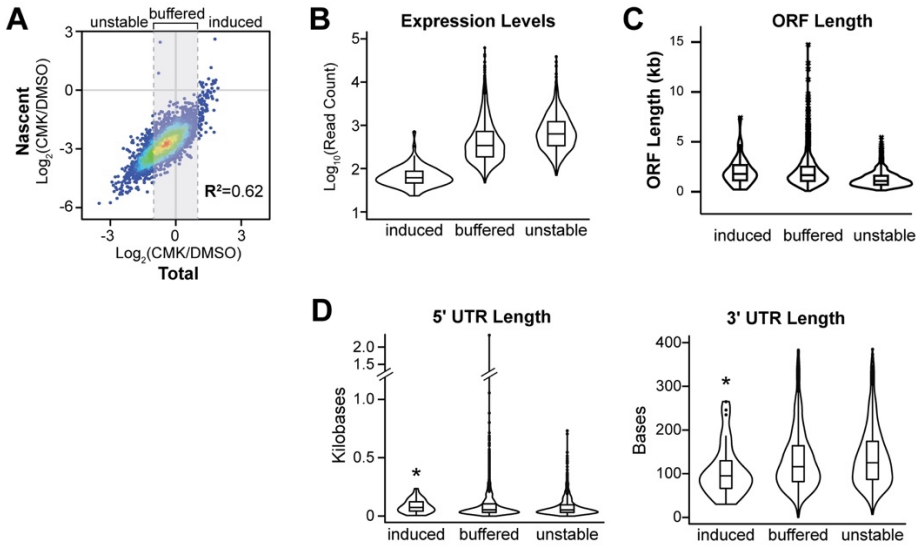



**Figure 3.2 Bioinformatic Analysis of Buffered versus Unstable Transcripts**

**(A)** Scatter plot of expression changes of newly synthesized, y axis, and steady-state mRNA, x axis, after inhibition of Kin28 with CMK. The square of the Pearson correlation coefficient ($R^2$) between the two variables is indicated. Buffered transcripts have expression changes between -1 and 1, shaded. **(B)** Violin plot showing the distribution of expression level in uninhibited cells of mRNA whose steady-state levels are increased (Log2(CMK/DMSO)>1; n=92), buffered (-1<Log2(CMK/DMSO)<1; n=3557), or decreased (Log2(CMK/DMSO)<-1; n=1050) after inhibition of Kin28. One-way ANOVA indicates the difference in mean expression level between the three groups of mRNA is statistically significant, p<0.0001. **(C)** Violin plot showing the distribution of ORF length. **(D)** Violin plots showing the distribution of UTR lengths. *One-way ANOVA indicates the difference in expression level, 5' and 3' UTR, and ORF length between the three groups of mRNA is statistically significant, p<0.0001.

**Table 3.1 Correlation between Published Datasets of Half-life (H.L.), Synthesis Rate (S.R.), 5' and 3' UTR and ORF Length (L.)**

| | H.L. [20] | S.R. [20] | H.L. [21] | H.L. [18] | 5'UTR-L [22] | 3' UTR-L [22] | ORF-L SGD |
|---|---|---|---|---|---|---|---|
| Total | 0.133 | 0.081 | 0.007 | 0.003 | 0.004 | 0.000 | 0.069 |
| Nascent | 0.041 | 0.020 | 0.012 | 0.008 | 0.003 | 0.001 | 0.000 |

effects to cells treated with low-dose CHX for 10 and 60 minutes (Spearman R=0.44 and 0.38, respectively) (**Figure 3.3**). Additionally, transcriptomic changes in inhibited *kin28is* also correlated with those observed from deletion of *XRN1* (R=0.41), *UPF2* and *UPF3* (R=0.35 and 0.36, respectively), and *THP2* (R=0.31). Deletion of the *UPF* factors [23] and *XRN1* [14], as well as inhibition of translation [24], are all known to generally increase the steady-state levels of mRNA. To our surprise, the effect of inhibition of Kin28 on the transcriptome more closely resembles that of defective translation rather than defective transcription by treatment with 1,10-phenanthroline (R=0.11). In sum, we find that inhibition of Kin28 leads to a unique transcriptomic landscape, wherein synthesis of nearly all nascent transcripts is severely diminished, but the steady-state levels of transcripts parallel the differential gene expression changes observed in conditions of translational attenuation and *XRN1*-deletion.

### 3.3 Inhibition of Kin28 Reduces Translation Globally

To probe the state of translation in inhibited Kin28 cells, we examined the polysome profiles from *kin28is* cell lysates (**Figure 3.4A**). The profiles of inhibited *kin28is* showed significantly less polysome formation and a slight accumulation of monosomes (**Figure 3.4B**). Polysome profiles may not reflect active translation, therefore we assayed translation in inhibited cells by metabolic labeling of nascent proteins in vivo. Methionine analog, L-homopropargylglycine (HPG), was added to the media and its incorporation into nascent polypeptides was measured using copper-catalyzed “click chemistry.” Aided by the catalyst, nascent proteins bearing a HPG side chain form a triazole adduct with the fluorescent Cy3-azide (**Figure 3.5A**) [25]. The results clearly demonstrate that nascent protein synthesis is drastically reduced when Kin28 is inhibited, but some synthesis continues at a low level (**Figure 3.5B** and **3.5C**).

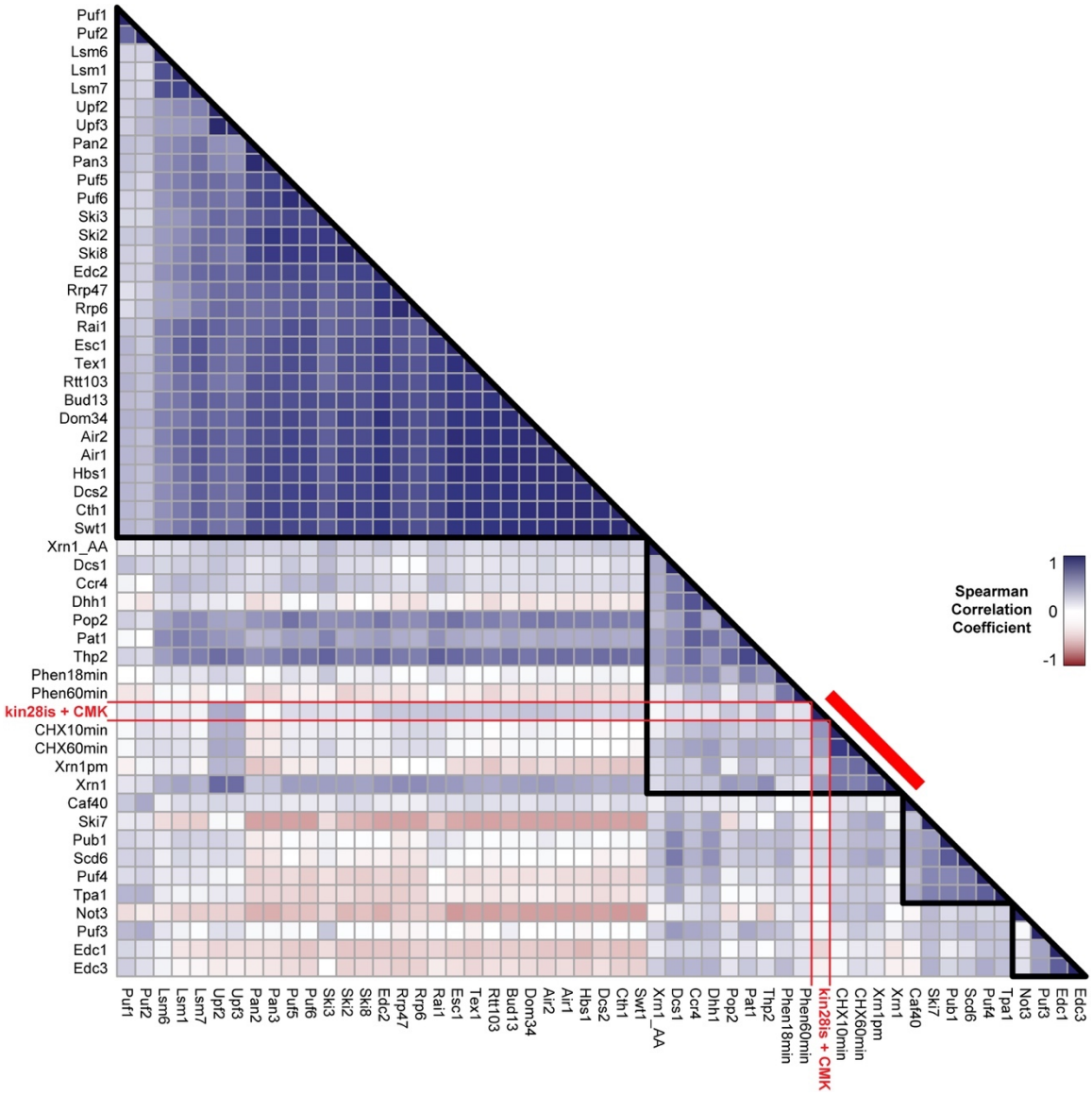



**Figure 3.3 The Transcriptomic Landscape of Irreversibly Inhibited Kin28 Cells Resembles That of Defective Translation**

Spearman correlation coefficients between the change in steady-state RNA levels (i.e., 'Total' Log2[treatment or mutant versus vehicle]) of Sun *et al.* decay factor knockout strains/treatments and inhibited *kin28is*. The pairwise correlations were clustered hierarchically with black outlines to emphasize the four most related groups. A strong positive correlation is indicated by blue, while a strong negative correlation is indicated by red..

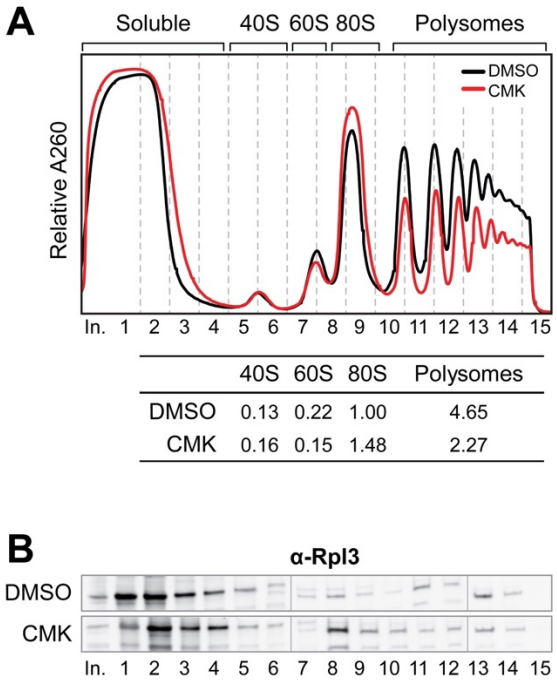



**Figure 3.4 Polysome Profiling of Inhibited *kin28is***

**(A)** Polysome profiling of *kin28is* cells treated with vehicle control (DMSO) or inhibitor (CMK) for 1 hr. The area under each peak, relative to the area under the DMSO 80S peak, is reported below. **(B)** Western blotting of Rpl3, a component of the large ribosomal subunit.

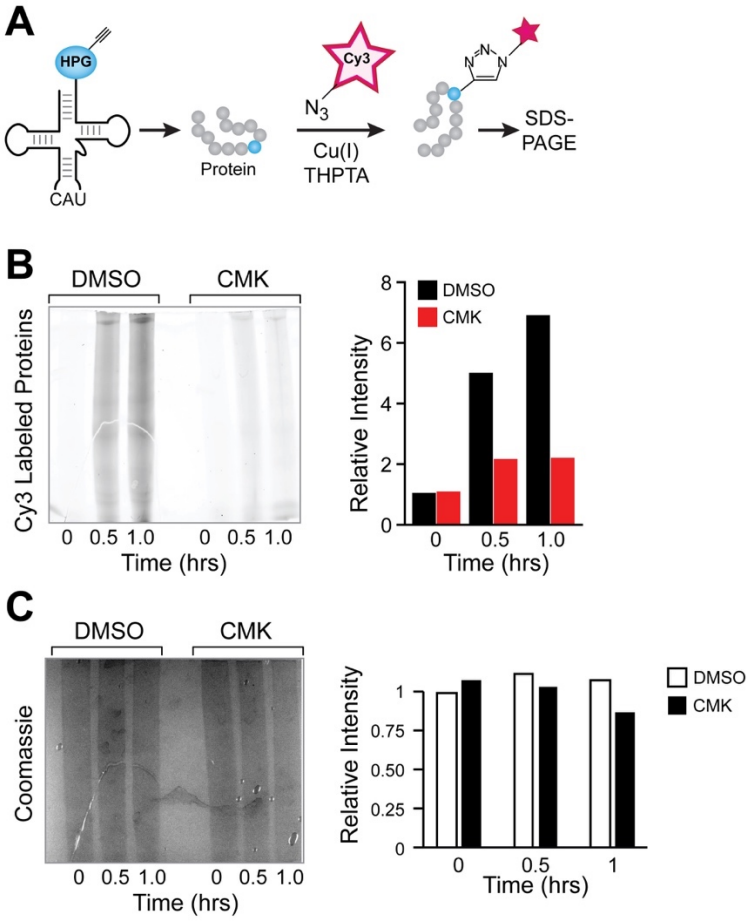



**Figure 3.5 Labeling of Nascent Proteins**

**(A)** Schematic of nascent protein labeling *in vivo* and the click chemistry performed after *in vitro*. L-homopropargylglycine (HPG) is incorporated into proteins in place of naturally occurring methionine. Cell lysate is reacted with Cy3-azide in the presence of Cu(I) and tris-hydroxypropyltriazolylmethylamine (THPTA), forming the protein-traizole-Cy3 adduct that can be detected fluorescently. **(B)** Cy3 fluorescence and **(C)** Coomassie staining of SDS-PAGE of whole cell extract from *kin28is* cells treated with vehicle control (DMSO) or inhibitor (CMK) for 1 hour. After 10 minutes of incubation with DMSO or CMK, proteins were labeled for the indicated times with methionine analog L-HPG to which Cy3-azide was conjugated through Cu(I)-catalyzed “click chemistry.”

Next, we examined common causes for translational attenuation to determine if they were mediating the effect in inhibited Kin28 cells. It is well understood that the m7G cap at the 5' end of mRNAs greatly enhances translation through binding with translation initiation machinery [26, 27]. Furthermore, the phosphorylation of the Ser5 and 7 residues of the Pol II CTD, which decreases upon inhibition of Kin28 [12, 28], recruits mRNA capping enzymes [29], so it is possible that inhibition of Kin28 would reduce m7G levels. Recently, the methylation of adenosines (m6A) was shown to increase on mRNAs that are derived in conditions of defective or slow transcription, and that increased levels of m6A impede translation [30]. We performed RNA immunoprecipitation with m7G- and m6A-specific antibodies and included an exogenous S. pombe spike-in to control for immunoprecipitation efficiency and to enable accurate quantification of m7G enrichment between samples (**Figure 3.6A**). The level of both buffered and unstable mRNAs in Input and immunoprecipated samples were quantified by RT-qPCR. Controlling for mRNA levels in the Input as well as efficiency of the immunoprecipitation, we found no significant changes in the cap state of steady-state mRNAs from either the buffered or non-buffered groups (**Figure 3.6B**) and a slight enrichment of m6A in some of the inhibited *kin28is* mRNA tested (**Figure 3.6C**).

The expected mechanism of attenuating translation through depletion of m7G cap was not the primary attenuator of protein translation in the context of Kin28 inhibition, so we next examined phosphorylation of the α subunit of the translation initiation factor eIF2 (eIF2α). Stress-responsive kinases phosphorylate eIF2α preventing the formation of translation initiation complexes, and this event, the integrated stress response (ISR), is a general mechanism to rapidly arrest translation when cells encounter unfavorable growth conditions [31]. Immunoblotting of *kin28is* cells treated with CMK showed no significant increase in eIF2α phosphorylation. This result demonstrates that inhibition of Kin28 does not trigger the ISR and the observed

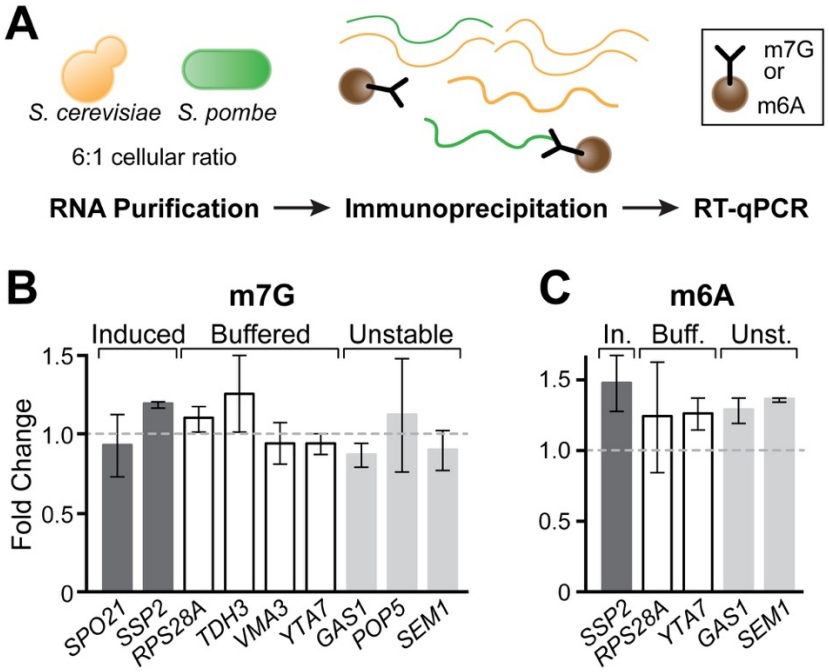



**Figure 3.6 Immunoprecipitation of Modified RNAs**

**(A)** Schematic of quantification of m7G-capped and m6A-modified mRNA. *S. pombe* cells are mixed with *S. cerevisiae* cells in a 6:1 cellular ratio. After RNA purification, an antibody that binds to the m7G cap or the m6A modification is used to purify capped mRNA from both yeast species. This is followed by RT-qPCR of specific *S. cerevisiae* and *S. pombe* transcripts, with the *S. pombe* mRNA serving as an external control. After *kin28is* was inhibited with CMK, RNA was immunoprecipitated with an antibody with affinity for (**B**) m7G or **(C)** m6A, which was quantified by RT-qPCR. cDNA levels were normalized to an *S. pombe* exogenous control before calculation of % Input, then fold change; mean +/- SEM, n=3.

reduction in translation is not mediated by phosphorylation of eIF2α (**Figure 3.7**). To ensure that this stress-responsive pathway is active in Kin28 inhibited cells, phosphorylation was induced by treatment with rapamycin, a known simulator of nutrient deprivation and inducer of eIF2α phosphorylation [32]. The phosphorylation of eIF2α is readily evident in *kin28is* cells, irrespective of their treatment with the covalent inhibitor. Furthermore, eIF2α phosphorylation was absent from yeast lacking *GCN2*, which is a highly conserved kinase that phosphorylates eIF2α in response to a range of cellular stresses [33].

Upon various perturbations, non-translating mRNAs are known to accumulate in cytoplasmic foci called P-bodies and stress granules [34]. To examine if extant mRNA were sequestered in P-bodies we examined the distribution of fluorescently tagged Edc3 and Pab1. Immunofluorescence of P-body (Edc3-GFP) and stress granule markers (Pab1-GFP) demonstrate that P-bodies, but not stress granules, formed within minutes of inhibiting Kin28 (**Figure 3.8**). The absence of eIF2α phosphorylation in inhibited *kin28is*, the lack of significantly increased m6A modification [35], the lack of similarity between the ISR transcriptome (**Figure 3.9**), and the lack of stress granule accumulation, reinforces the conclusion that inhibition of Kin28 does not trigger the ISR during the time scale of our studies but it stabilizes extant mRNA by sequestering them in P-bodies.

### 3.4 Inhibition of Kin28 Reduces Co-translational Decay

Ribosome dynamics can be captured by extracting 5'P-Seq signal that displays a 3-nt periodicity in open reading frames (ORFs) (**Figure 3.10**) [36]. 5'P-Seq selectively captures RNA with a 5'-monophosphate terminus. From a set of 3765 transcripts with robust signal, we found a positive correlation between changes in 5'P and changes in RNA-Seq when Kin28 was inhibited (R2=0.28) (**Figure 3.11A**). Similar to the expression

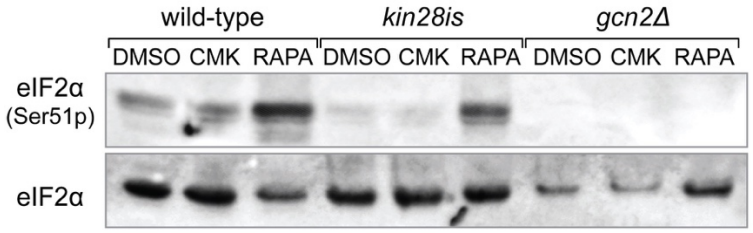



**Figure 3.7 Immunoblotting for Phosphorylated or Total eIF2α**

The wild-type, *kin28is*, or *gcn2Δ* cells were treated with vehicle control (DMSO), Kin28is inhibitor (CMK), or 400 ng/mL rapamycin (RAPA).

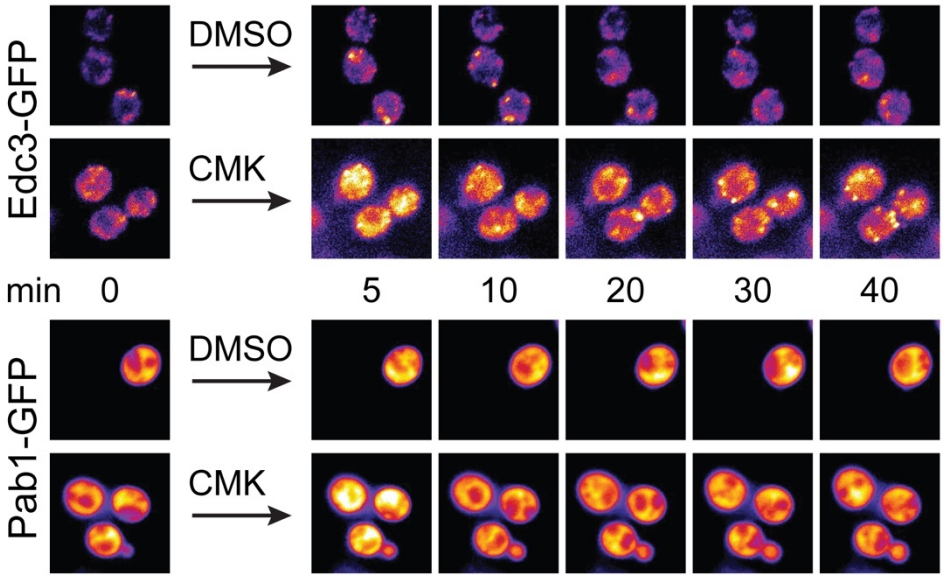
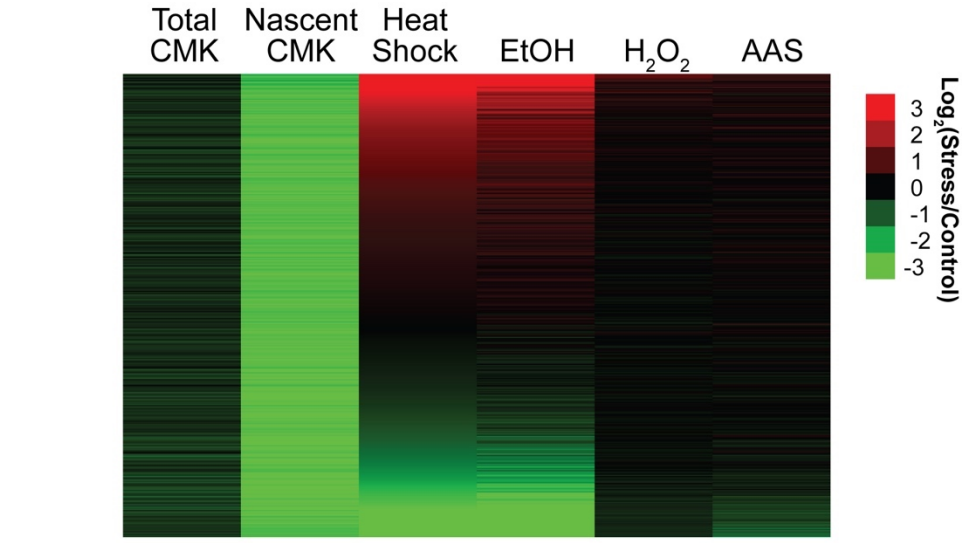



**Figure 3.8 Immunofluorescence of P Body and Stress Granule Markers**

Representative images of *kin28is* cells expressing P body marker Edc3-GFP or stress granule marker Pab1-GFP before and after the addition of vehicle control (DMSO) or inhibitor (CMK) for the indicated time.




| | | | | | | |
|---|---|---|---|---|---|---|
| **Total** | 1.000 | 0.614 | 0.009 | 0.025 | 0.000 | 0.032 |
| **Nascent** | 0.614 | 1.000 | 0.015 | 0.032 | 0.007 | 0.017 |

**Figure 3.9 Expression Changes Compared to the Induced Stress Response**

Heatmap of the expression changes from this study (inhibition of *kin28is*) plotted adjacent to the expression changes of yeast treated with heat shock, ethanol (EtoH), hydrogen peroxide ($H_2O_2$), and amino acid starvation (AAS) [37]. The Pearson correlation between our dataset and the respective stress dataset is listed below.

changes quantified by RNA-Seq, mRNA with decreased 5'P in inhibited Kin28 cells tend to be higher expressed (**Figure 3.11B**).

By comparing the pattern of 5'P across a transcript, we could infer its translation state. The buffered mRNA RPS28A has increased 5'P reads across its length in inhibited *kin28is* (**Figure 3.12A**), which suggests *RPS28A* is still translated in inhibited *kin28is*. In contrast, the unstable mRNA *SEM1* has decreased 5'P reads, which is most apparent at the peak -20 nt from the stop codon (**Figure 3.12B**). This 5'P peak at -20 nt from the stop codon corresponds to the pause of the last ribosome terminating translation. As expected, the 5'P reads from the isogenic strain of yeast bearing wild-type KIN28 was unaffected by treatment with the inhibitor (**Figure 3.13**). In general, the change in 5'P at the start of the ORF and at the 5'-UTR scales with the change in steady-state RNA levels determined by RNA-Seq (**Figure 3.13A**). Conversely, a dramatic loss of 5'P signal at the end of the ORF is apparent in most transcripts from both the buffered and unstable groups of mRNAs (**Figure 3.13B**). The general loss in 5'P is uniform ~100 bp, or three to four ribosome lengths, from the stop codon, suggesting a lack of ribosome processivity and/or Xrn1 activity (discussed in the next section) in inhibited Kin28 cells.

Plotting the fraction of 5'P reads in the codon-protected frame across gene length shows the proportion of reads is reduced but still greater than 0.33, the theoretical fraction indicative of no co-translational decay (**Figure 3.14A**). Importantly, the fraction of codon-protected reads is not increased before the start codon relative to the downstream nucleotides when Kin28 is inhibited, which would indicate a stalling of ribosomes at the start codon. A severe elongation defect would be reflected by an accumulation of codon-protected reads relatively closer to the start codon because of the increased residence time of ribosomes [36].

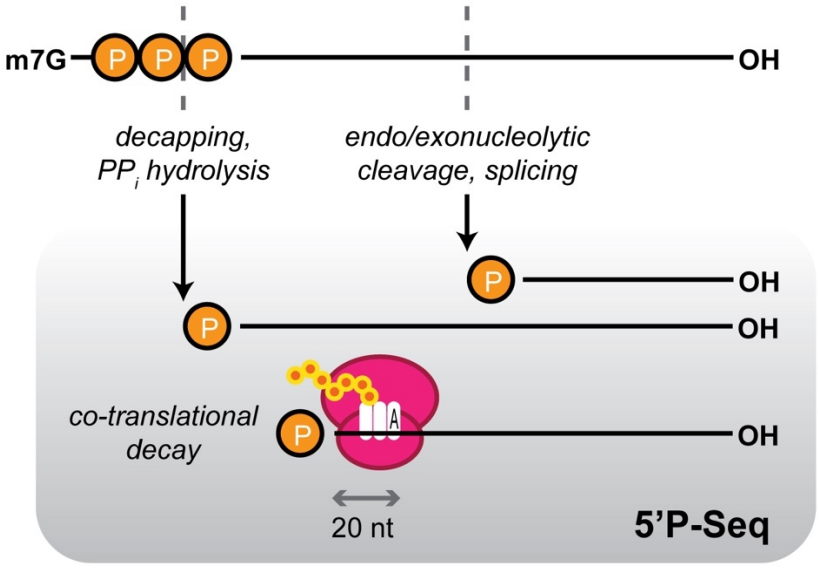



**Figure 3.10 Schematic of RNA Captured via 5'P-Seq**

A diagram of processes which generate the 5'P degradation intermediates captured by 5'P-Seq is shown. The ribosome occupies 20 nucleotides of RNA from the 5'P to the end of the 'A' site.

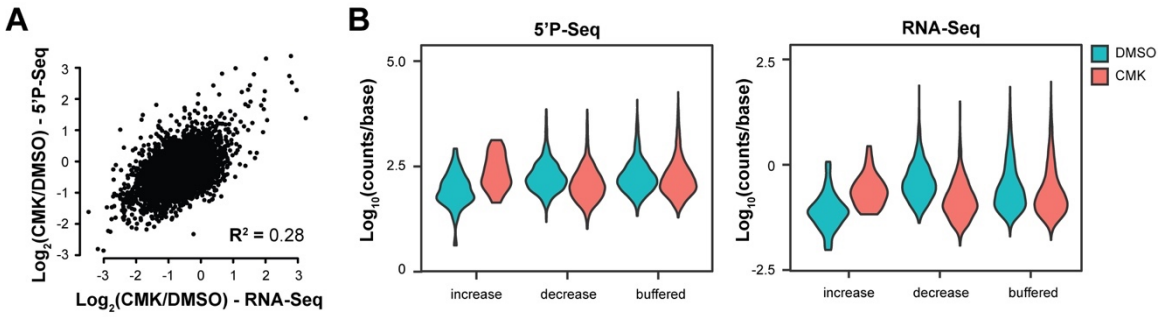



**Figure 3.11 A Comparison of RNA-Seq and 5'P-Seq Read Counts in Inhibited *kin28is***

**(A)** Scatter plot of expression changes (RNA-Seq), x axis, and changes in 5'P signal, y axis, in *kin28is* cells treated with inhibitor CMK. The square of the Pearson correlation ($R^2$) is displayed. **(B)** Violin plots showing the distribution of 5'P-Seq, left, and RNA-Seq, right, read counts of mRNA whose steady-state levels are increased, buffered, or decreased after inhibition of Kin28.

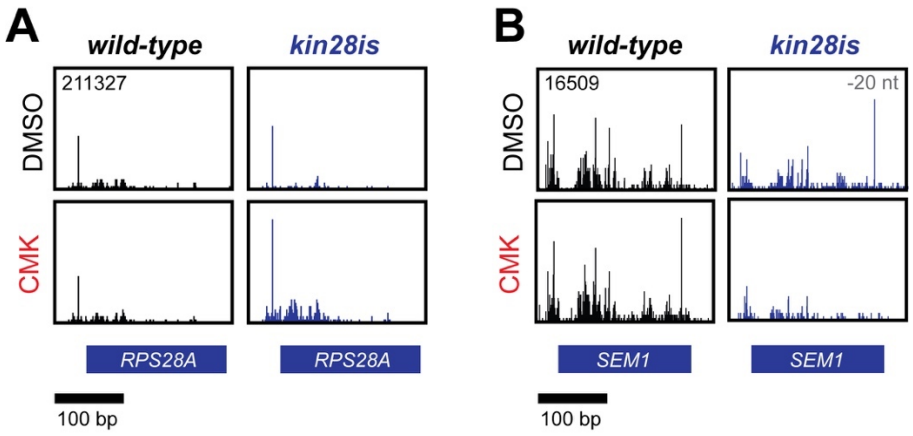



**Figure 3.12 Example 5'P-Seq Coverage over Genes**

5'P-Seq reads mapped to a **(A)** buffered mRNA, *RPS28A*, and a **(B)** unstable mRNA, *SEM1*, in wild-type and kin28is cells treated with the vehicle control (DMSO) or inhibitor (CMK). The peak at -20 nt corresponding to the last terminating ribosome is indicated.

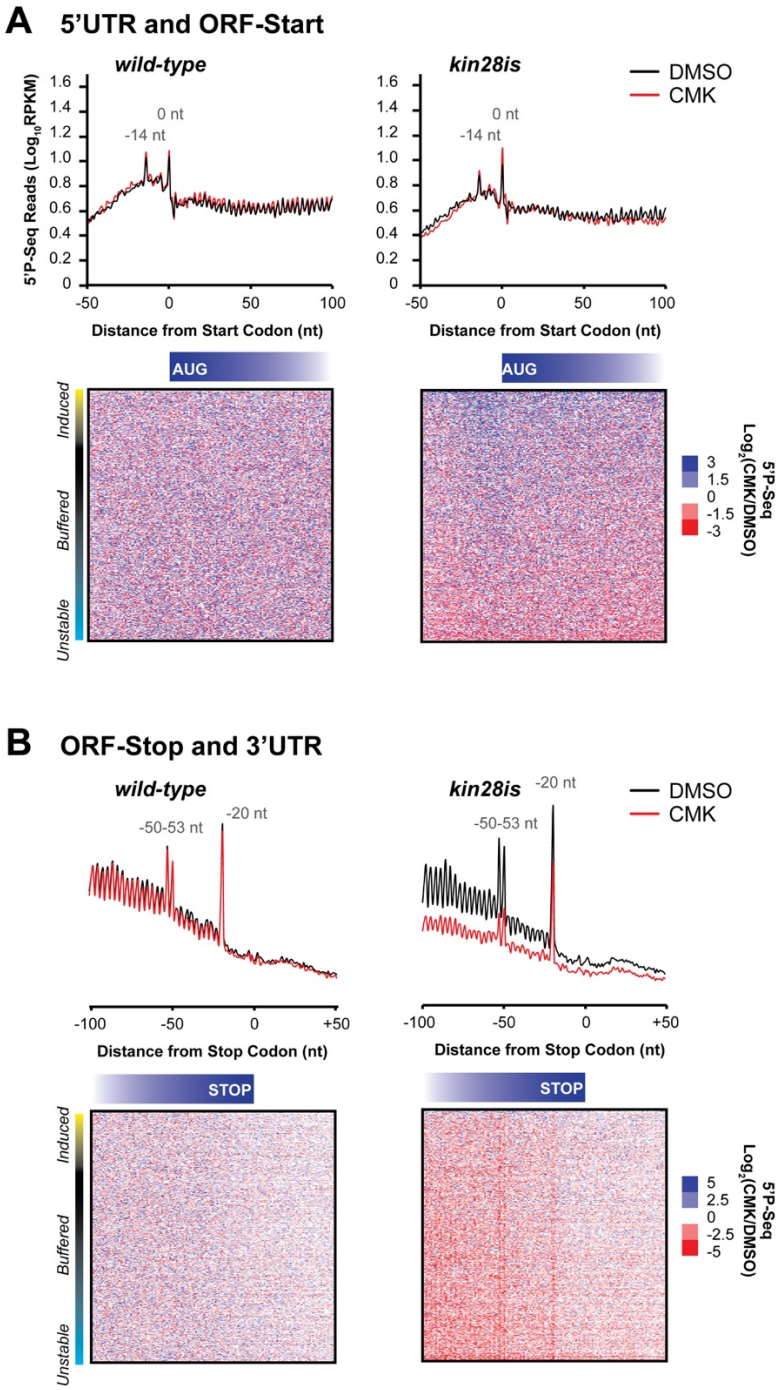



**Figure 3.13 Changes in 5'P at the ORF**

The average 5'P-Seq reads relative to **(A)** the first nucleotide of the start codon or to **(B)** the last nucleotide of the stop codon of the ORF is plotted. The peaks at -14 and 0 nt correspond to a ribosome parked at the start and 6th codon, respectively, in the 'P' site. The peaks at -20 and -50-53 nt correspond to one or two ribosomes, respectively, parked at the stop codon in the 'A' site. Heatmap of the change in 5'P of treated cells per gene ordered by steady-state expression change in *kin28is*, below. Only genes with over 100 RNA-Seq reads and over 32 5'P-Seq reads/base are included, n=3765.

The 3-nt periodicity in the 5'P signal, which is associated with Xrn1 chasing the last translating ribosome, decreases at both the start and end of the ORF (**Figure 3.14B**), indicating that the defect in co-translational decay occurs early in translation. The decrease in 3-nt periodicity is apparent in both buffered and non-buffered transcripts, suggesting these changes to co-translational decay do not dictate mRNA stability in inhibited *kin28is* (**Figure 3.14C**). Interestingly the few mRNAs that increase in expression in inhibited *kin28is* continue to be translated, as indicated by sustained 3-nt periodicity and intensity of 5'P-Seq signal.

On closer inspection of the change in 5'P signal at the start codon for the mRNAs that increase in inhibited Kin28 cells, a slight accumulation of 5'P reads relative to the end of transcript is visible. This increase in 5'P is diffuse over the start of the ORF and does not present as a distinct peak (*e.g.* not -14 nt from the start codon that is associated with the 48S pre-initiation pause [36]), suggesting the increase in 5'P does not stem from pausing of the pre-initiation complex. Rather, the increase in 5'P signal upstream of the annotated start codon in inhibited Kin28 cells might represent an increase in alternate TSS and start codon usage from long undecoded transcript isoforms (LUTIs), which are known to be utilized by cells during certain phases of meiosis [38]. In support of this idea, we found several examples, in both our 5'P-Seq and RNA-Seq data, of increased expression of LUTIs in inhibited *kin28is* (**Figure 3.15**).

Overall, our 5'P-Seq analysis supports the idea that the global reduction in translation of inhibited Kin28 cells observed in polysome profiling is the result of reduced association of mRNA with ribosomes, rather than a defect in a particular stage of translation (*i.e.* initiation or termination). Many mRNAs are still translated, albeit with a defect in elongation.

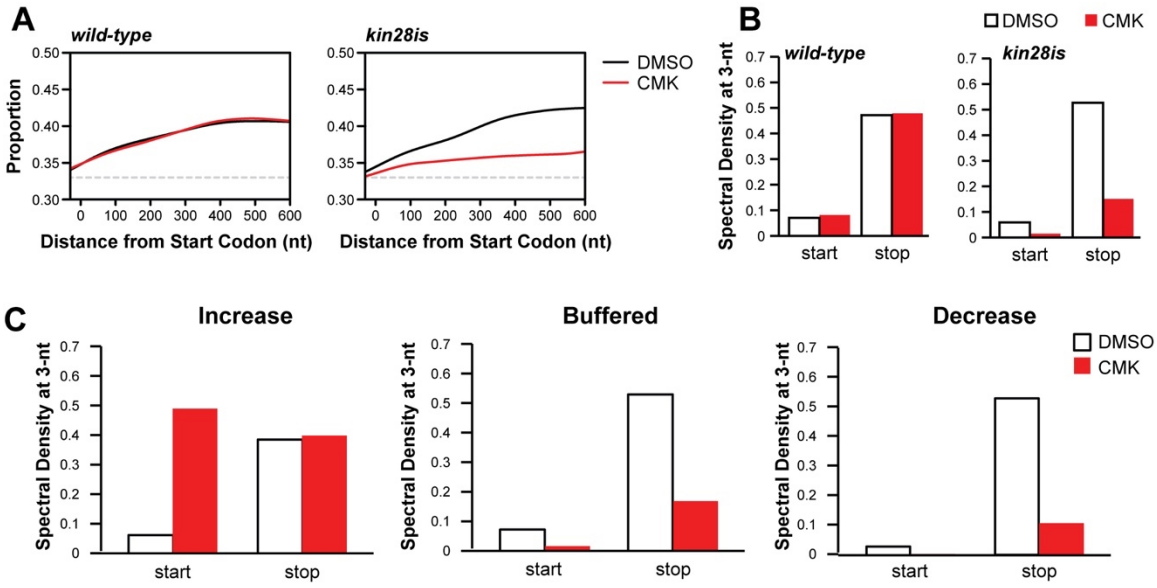



**Figure 3.14 Analysis of 5'P-Seq Signal in the Translated Frame**

**(A)** The proportion of 5'P reads in the ribosome-protected frame plotted across the first 600 nt of a meta-ORF for untreated or inhibited wild-type and *kin28is* cells. The smoothed average is displayed for genes without introns and longer than 600 nt. The gray dashed line indicates 0.33, which is the expected proportion of 5'P in one frame if not biased by ribosome protection. **(B)** The power of the 3-nt periodicity in the 5'P reads is plotted as bars for wild-type and *kin28is* cells treated with the vehicle control (DMSO) or inhibitor (CMK) in a 45-nt window near the start and stop codon. **(C)** Same as (B) but grouped by mRNA whose steady-state levels are increased, buffered, or decreased after inhibition with CMK.

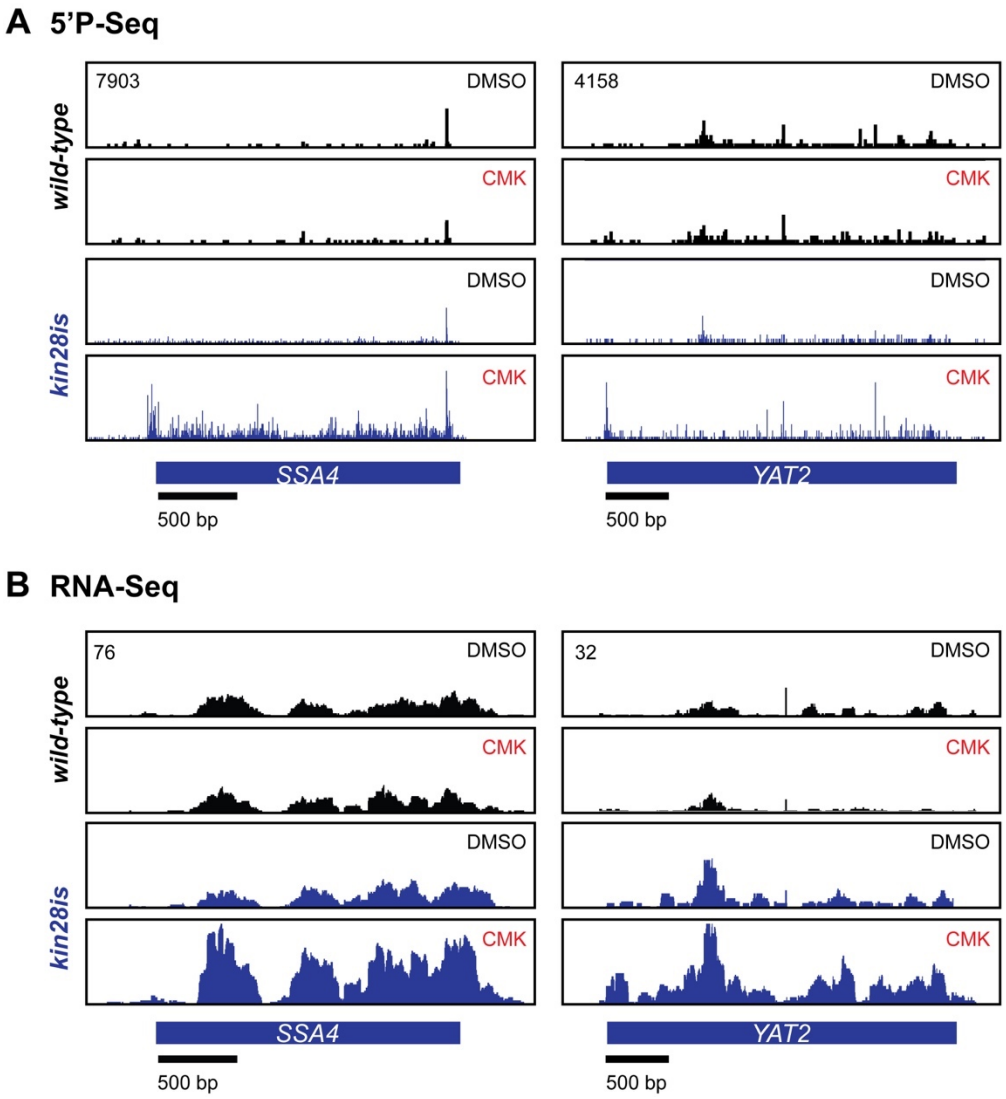



**Figure 3.15 Examples of LUTIs in Inhibited *kin28is***

*SSA4* and *YAT2* express long undecoded transcript isoforms (LUTIs) in inhibited *kin28is* cells, but not wild-type yeast upon treatment with CMK. LUTIs are characterized by alternate TSS usage, which is distinguishable in both **(A)** 5'P- and **(B)** RNA-Seq data.

### 3.5 Xrn1-mediated Co-translational Decay Is Attenuated During Defective Transcription

Because 5'P signal intensity is an indicator of decay as well as the state of translation, we wondered if the reduction in co-translational decay observed in 5'P-Seq could also be explained by attenuated Xrn1 activity. To test if exonucleolytic activity was reduced in inhibited *kin28is*, we depleted Pol II from the nucleus using the anchor-away technique [39], and monitored mRNA levels with RT-qPCR (**Figure 3.16A**). Removing Pol II from the nucleus effectively ablates synthesis of additional mRNAs, such that the decline in mRNA levels over time only reflects inherent or exonuclease-catalyzed degradation. Buffered mRNA, *e.g.*, *RPS28A* and *YTA7*, were indeed stabilized after transcriptional arrest, while mRNA with decreased steady-state expression, *e.g.*, *SEM1* and *GAS1*, were unstable (**Figure 3.16B**). Thus, differences in steady-state level are not simply a function of synthesis rates, but are actually a result of differential RNA decay. Surprisingly, transcripts that were induced upon inhibition of Kin28, such as *SSP2*, increased in quantity despite removal of Pol II, suggesting they may actually be Pol I or III transcripts. This is supported by ChIP, which reveals low levels of Pol II at *SSP2* (**Figure 3.16C**). Interestingly, the differential mRNA stability in inhibited *kin28is* cells seems to mirror that of anchor-away of Pol II. In other words, transcripts that are buffered when Kin28 is inhibited are also buffered when synthesis of new transcripts is stopped in general.

Next, we sought to determine which decay factors are inactivated when Kin28 is inhibited. Like most mRNA, the mRNA of decay factors was dramatically reduced in synthesis while the steady-state level was moderately reduced (**Figure 3.17**). This modest reduction of decay factors may be sufficient to initiate buffering. Indeed, deletion of the major cytoplasmic 5'-to-3' exonuclease Xrn1 was able to partially restore steady-state levels of transcripts that were normally not buffered in inhibited *kin28is* cells

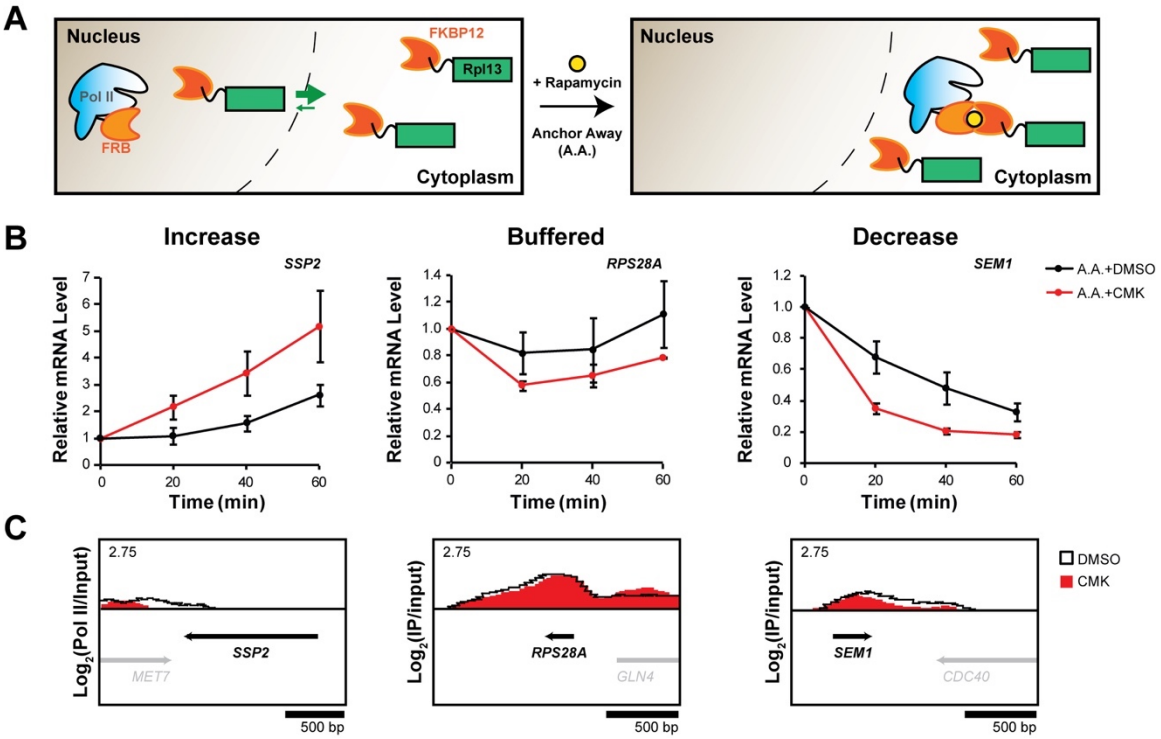



**Figure 3.16 RNA Buffering Occurs in Other Situations of Transcriptional Crisis**

**(A)** Schematic of anchor-away of Rpb1. Pol II tagged with FRB is effectively removed from the nucleus via rapamycin-induced dimerization of FRB to FKBP12, which itself is attached to Rpl13, an abundant cytoplasmic, ribosomal protein that shuttles in and out of the nucleus. **(B)** Relative steady-state levels of example transcripts from the indicated groups after anchor-away of Rpb1 (A. A.). *kin28is* cells were treated with vehicle control or CMK for 10 minutes before addition of rapamycin at time point 0 minutes to induce A. A. Expression is normalized to an *S. pombe* exogenous control; mean +/- SEM, n = 3 biological replicates. All values are relative to the respective value at time = 0 minutes. **(C)** Rpb3 ChIP profiles of genes in (B) from *kin28is* cells treated with vehicle control (DMSO) or inhibitor (CMK).

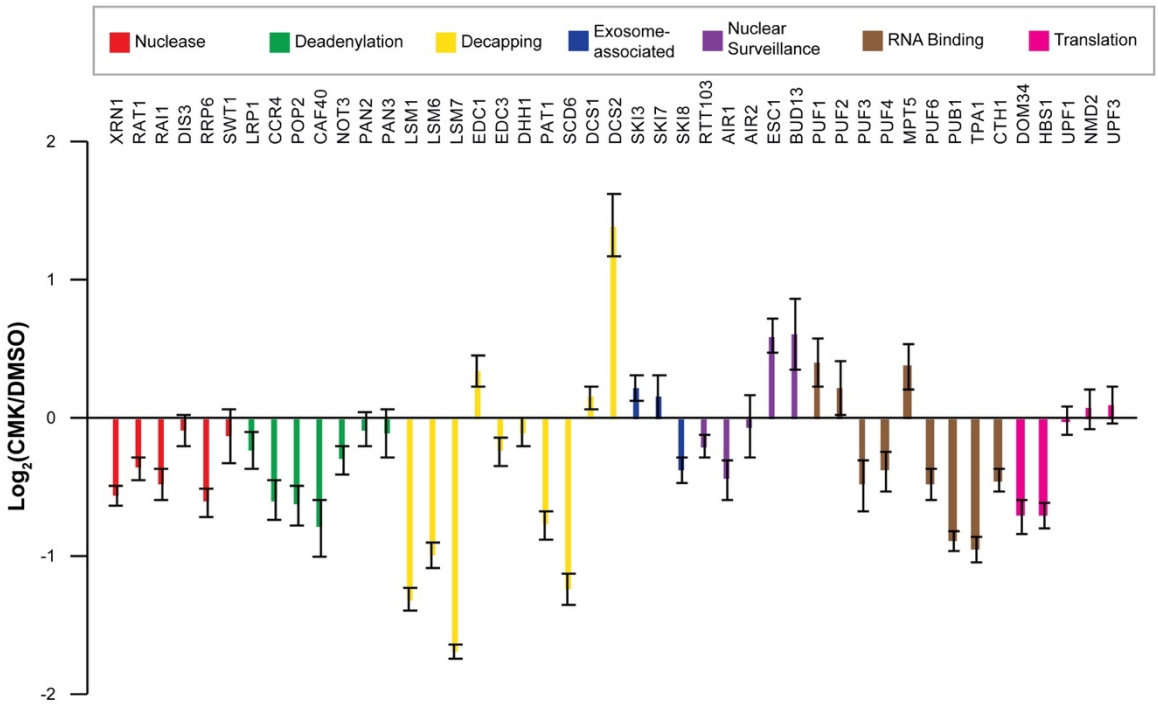



**Figure 3.17 Change in Expression of mRNA Decay Factors in Inhibited *kin28is***

The mean change in expression of factors involved in mRNA decay when Kin28 is inhibited is presented as bars colored by the process with which they are associated. Data from RNA-Seq; mean +/- SEM, n = 4 biological replicates.

(**Figure 3.18**). This suggests that the spectrum of differential gene expression observed from Kin28 inhibition is caused in some part by differential Xrn1 activity. Buffered mRNAs are protected from Xrn1-mediated decay during transcriptional arrest, and, unstable mRNAs are either particularly sensitive to Xrn1, or are decayed by other factors. Though deletion of XRN1 was not able to reverse the growth defect of inhibited *kin28is* (**Figure 3.19A**), it allowed *kin28is* colonies to eventually form on YPD+CMK plates (**Figure 3.19B**).

### 3.6 mRNA Decay in Inhibited *kin28is* Is Characterized by Subcellular Localization

To untangle the identity of transcripts with reduced translation from those with decreased steady-state expression, we performed k means clustering of the mRNAs by changes in their 5'P-Seq signal (**Figure 3.20A**). More specifically, steady-state expression changes are reflected by the change in 5'P signal at the ORF start and the UTRs (**Figure 3.20B**), whereas changes to translation are apparent at the ORF stop. Thus, clustering changes to 5'P signal at a window centered around the start as well as the stop codon allowed us to deconvolute the effect of inhibiting Kin28 on transcription and translation, respectively. Overall, we find that the translation state of an mRNA did not necessarily match with its buffering state in inhibited *kin28is*. That is, a group of buffered as well as unstable mRNAs (Clusters 2 and 4, respectively) showed relatively unchanged 5'P signal at the end of their ORFs. Similarly, a group of buffered as well as unstable mRNAs (Clusters 3 and 5, respectively) exhibited a severe defect in translation.

Gene ontology (GO) analysis of the mRNAs with the most reduced translation in inhibited Kin28 cells (Clusters 3 and 5) revealed these mRNA are associated with nuclear and cell cycle processes such as 'transcription,' 'nucleolus,' 'ribosome biogenesis,' and 'nucleosome assembly' (**Table 3.2**). Clusters 2 and 4, which contain mRNA that continue to be translated, share GO terms related to 'membrane' and

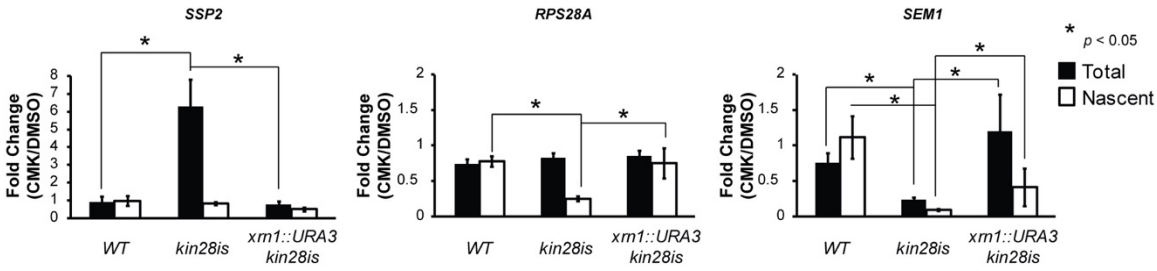



**Figure 3.18 Differential Xrn1 Activity Mediates Buffering and Co-translational Decay in Inhibited Kin28 Cells**

Change in steady-state (Total) and newly synthesized (Nascent) mRNA levels of example transcripts from the groups indicated in Figure 3.16 after 1 hour inhibition of wild-type (WT), *kin28is*, or *XRN1*-deleted *kin28is* (*xrn1::URA3 kin28is*) cells. Expression was normalized to an *S. pombe* exogenous control before calculation of fold change; mean +/- SEM, n = 3-7 biological replicates, one-tailed *p < 0.05.

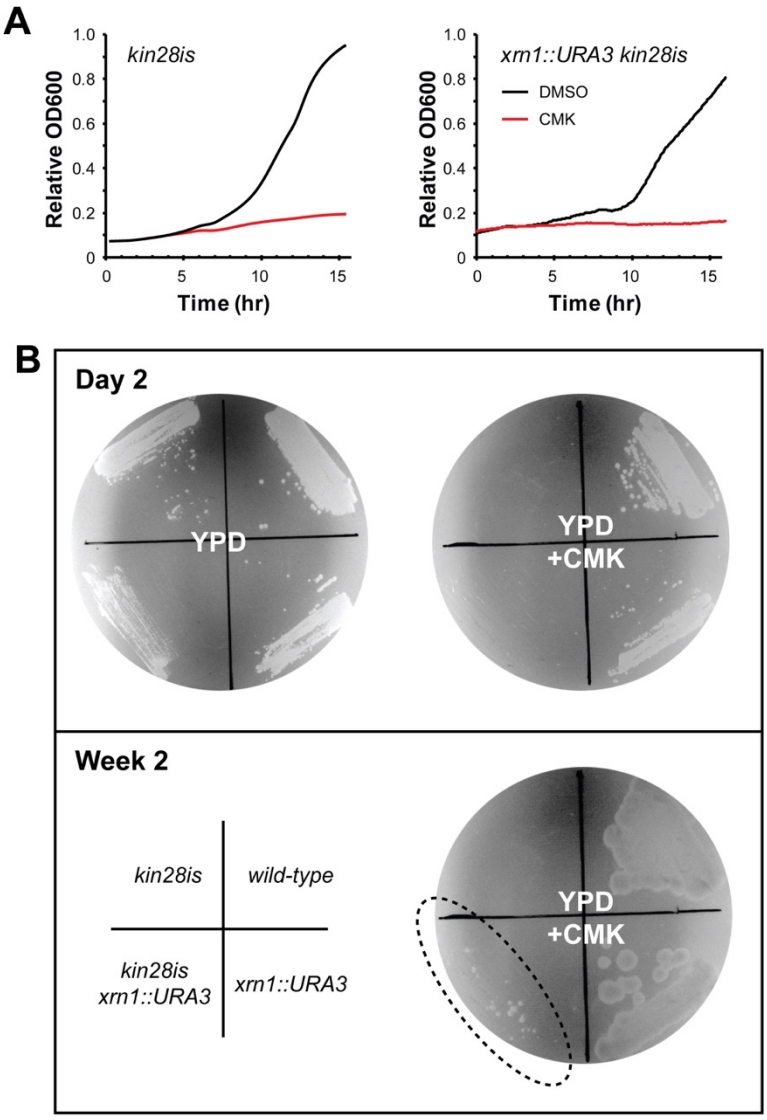



**Figure 3.19 Growth of *xrn1::URA3 kin28is* In Inhibitor-treated Media**

**(A)** Growth curves for *kin28is* and *xrn1::URA3 kin28is* cells treated with vehicle control (DMSO) or 5 μM of inhibitor (CMK). **(B)** Viability assay of the yeast strains indicated in the bottom-left on normal media (YPD) or media with 10 μM inhibitor (YPD+CMK). The time elapsed after plating the yeast is shown.

'transport.' Cluster 1 mRNAs, which are stable and translated in inhibited Kin28 cells, are associated with protein folding and the stress response. Taken together, most mRNAs experience a translation defect when Kin28 is inhibited, but many mRNAs are still translated, with associations to cytoplasmic "house-keeping" functions and response to stress. In contrast, the mRNAs with the most severe translation defects are related to nuclear processes.

### 3.7 Buffered mRNA Display Differential Preference for Ski2 over Nab2

The associations between subcellular localization, stability, and translation state of mRNAs in the inhibited Kin28 transcriptome prompted us to examine the relative binding of RNA-binding proteins (RBPs). We hypothesized mRNA ribostasis in inhibited *kin28is* might be correlated with differential binding of cytoplasmic versus nuclear RBPs. To test this idea, we compared the relative binding of the RBPs from the "transcriptome-wide atlas" of crosslinking and analysis of cDNA (CRAC) [40] to the inhibited *kin28is* transcriptome. Interestingly, the relative binding of Nab2 and Ski2 was strikingly different between buffered and unstable transcripts (**Figure 3.20C**). Predominantly Nab2-bound mRNAs are unstable in inhibited *kin28is*, while Ski2-bound mRNAs are typically buffered. Nab2 is a nuclear polyA binding protein [41], and Ski2 is a cytoplasmic helicase associated with the exosome [42]. Though less apparent than the relative binding of Nab2/Ski2, other nuclear RBPs (Cbc1, Hrp1) are bound in greater proportion to unstable transcripts (**Figure 3.21A**). Xrn1 binding did not seem to correlate with changes in mRNA stability. A recent study was published that revealed defects in transport and loss of Nab2 resulted in instability of mRNAs in the nucleus [43], a result consistent with our observation that cytoplasmic Ski2-associated mRNAs are more protected from degradation during transcriptional stress. Thus, the absolute amount of

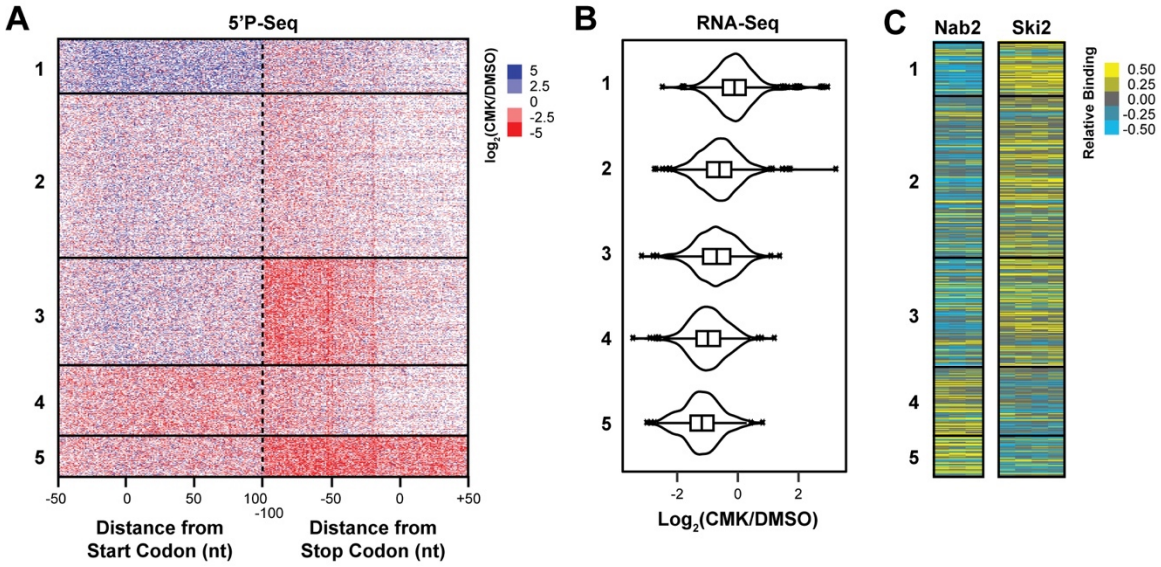



**Figure 3.20 Properties of Unstable mRNAs in Inhibited *kin28is***

**(A)** k-means clustering of the change in 5'P at the start and stop codon. See Figure 3.13. **(B)** Violin plot showing the distribution of steady-state mRNA expression changes in inhibited *kin28is* cells by clusters identified in (A). **(C)** Relative binding (normalized across each gene) of Nab2 and Ski2 from Tuck and Tollervey [40] CRAC datasets plotted in the same order as in (A).

**Table 3.2. Gene Ontology Analysis of Clusters Identified in Figure 3.20**

Significant ($p < 0.01$, Bonferroni-adjusted) gene ontology (GO) terms shared among genes in each cluster grouped by 5'P-Seq profile. The number of genes (n) in each cluster, median open-reading frame (ORF) length in nucleotides (nt), and median fold-change in steady-state RNA-Seq is listed for each cluster. The number of genes in the cluster (k) is reported along with the number of genes in each GO term category (f).

| Cluster | n | ORF Length (nt) | RNA-Seq Log2FC | GO Terms | *p*-value | k | f |
|---|---|---|---|---|---|---|---|
| **1** | 478 | 1437 | -0.11 | catalytic activity | 8.2E-10 | 69 | 455 |
| | | | | response to stress | 2.6E-07 | 30 | 152 |
| | | | | protein refolding | 3.4E-07 | 9 | 16 |
| | | | | ATPase activity | 8.8E-07 | 24 | 112 |
| **2** | 1409 | 1494 | -0.61 | nucleotide binding | 2.2E-10 | 235 | 778 |
| | | | | endoplasmic reticulum membrane | 5.6E-09 | 111 | 318 |
| | | | | ATP binding | 6.2E-09 | 190 | 622 |
| | | | | endoplasmic reticulum | 1.3E-08 | 136 | 416 |
| | | | | membrane | 1.3E-08 | 437 | 1671 |
| | | | | transport | 4.1E-07 | 229 | 815 |
| | | | | transferase activity | 8.0E-07 | 178 | 611 |
| **3** | 902 | 1701 | -0.70 | transcription, DNA-dependent | 2.9E-12 | 131 | 540 |
| | | | | nucleolus | 1.1E-11 | 75 | 253 |
| | | | | regulation of transcription, DNA-dependent | 1.6E-10 | 120 | 507 |
| | | | | rRNA processing | 2.8E-10 | 60 | 195 |
| | | | | ribosome biogenesis | 1.9E-09 | 53 | 170 |
| | | | | DNA binding | 1.7E-07 | 100 | 449 |
| | | | | ATP binding | 2.0E-07 | 129 | 622 |
| | | | | nucleotide binding | 2.6E-07 | 154 | 778 |
| **4** | 620 | 1123.5 | -0.99 | transport | 9.4E-08 | 119 | 815 |
| | | | | membrane | 1.7E-08 | 214 | 1671 |
| | | | | integral to membrane | 3.1E-07 | 170 | 1303 |
| **5** | 356 | 1149 | -1.19 | P-P-bond-hydrolysis-driven protein transmembrane transporter activity | 4.1E-06 | 6 | 10 |
| | | | | nucleosome | 1.7E-10 | 9 | 11 |
| | | | | chromatin assembly or disassembly | 7.2E-09 | 11 | 23 |
| | | | | nuclear nucleosome | 9.5E-09 | 8 | 11 |
| | | | | nucleosome assembly | 1.1E-07 | 9 | 18 |
| | | | | nucleolus | 5.6E-07 | 34 | 253 |

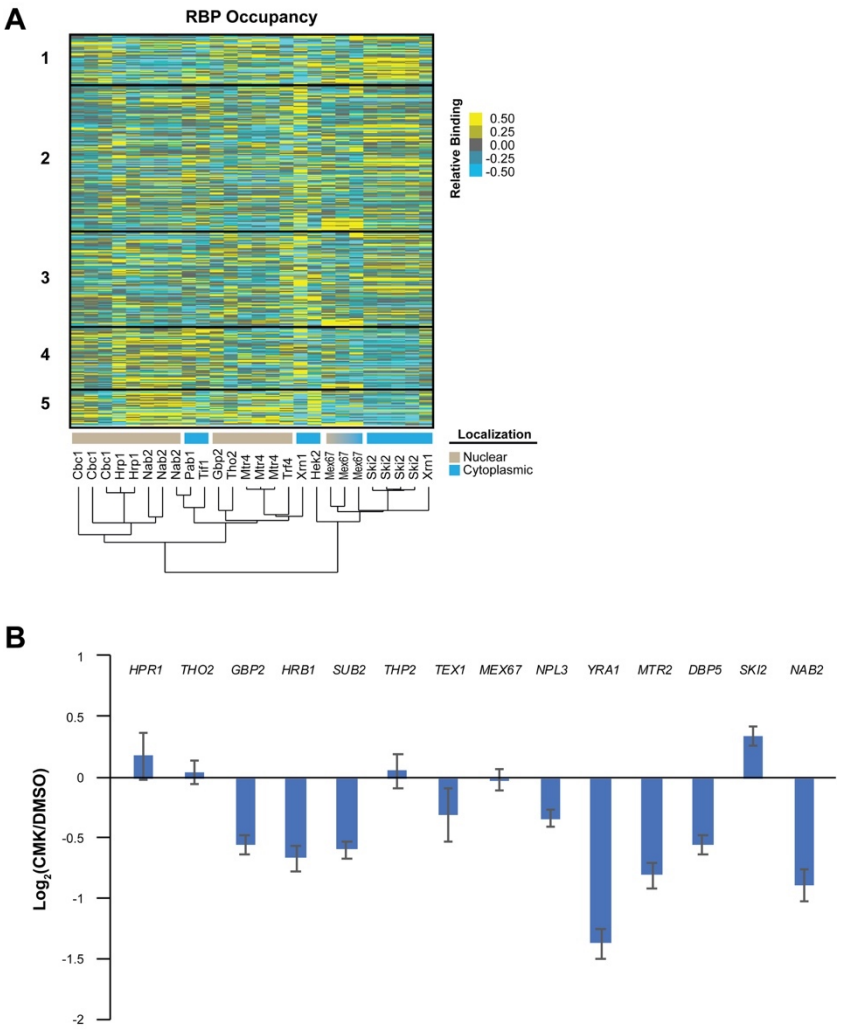



**Figure 3.21 Nuclear Association of Unstable mRNA in Inhibited *kin28is***

**(A)** Relative binding (normalized across each gene) of RNA binding proteins from CRAC datasets [40] plotted in the same order as in Figure 3.20. **(B)** Change in expression of factors involved in mRNA export when Kin28 is inhibited is presented as bars. Data from RNA-Seq; mean +/- SEM, n = 4 biological replicates.

Nab2 binding is less relevant than the relative binding of Nab2-to-Ski2 in predicting mRNA stability in inhibited *kin28is*. Notably, the mRNA of export factors are fairly unstable in inhibited *kin28is* (**Figure 3.21B**).

We further explored the relationship between mRNA instability and the cell periphery by analyzing the subcellular localization of the proteins that are encoded by transcripts that increase, decrease, or are buffered when Kin28 is inhibited. The proportion of genes that encode cell periphery and ER-associated proteins is significantly greater in the group of genes with decreased expression when Kin28 is inhibited (**Figure 3.22**). Conversely, the proportion of genes that encode cytosolic proteins is greater in the group of genes with buffered expression. Analysis of another dataset of the subcellular localization of the yeast proteome also revealed ER proteins tend to be encoded by non-buffered mRNAs (**Figure 3.23**).

### 3.8 Discussion

Kin28 regulates gene expression in a myriad of ways beginning with its role in transcription initiation and recruitment of capping enzymes via phosphorylation of Ser5 and Ser7 residues of the CTD heptad repeat. In this study, we show inhibition of Kin28 has far-reaching roles beginning in, but extending well beyond the nucleus (**Figure 3.24**). The defects in new RNA synthesis are severe and affect nearly all protein-coding genes with a few exceptions, discussed below. Surprisingly, we find that the Kin28-inhibited transcriptome does not mirror the canonical yeast ISR (**Figure 3.9**). In the ISR many genes are acutely up- and down-regulated [37], while in inhibited Kin28 cells, very few genes are up-regulated, and most genes show little change in steady-state level. The ~2% of transcripts that consistently increase over two-fold when Kin28 is inhibited share many associations with meiosis (**Table 3.3**). This group of induced meiosis-related mRNA is lowly expressed under normal conditions in part due to RNA interference by

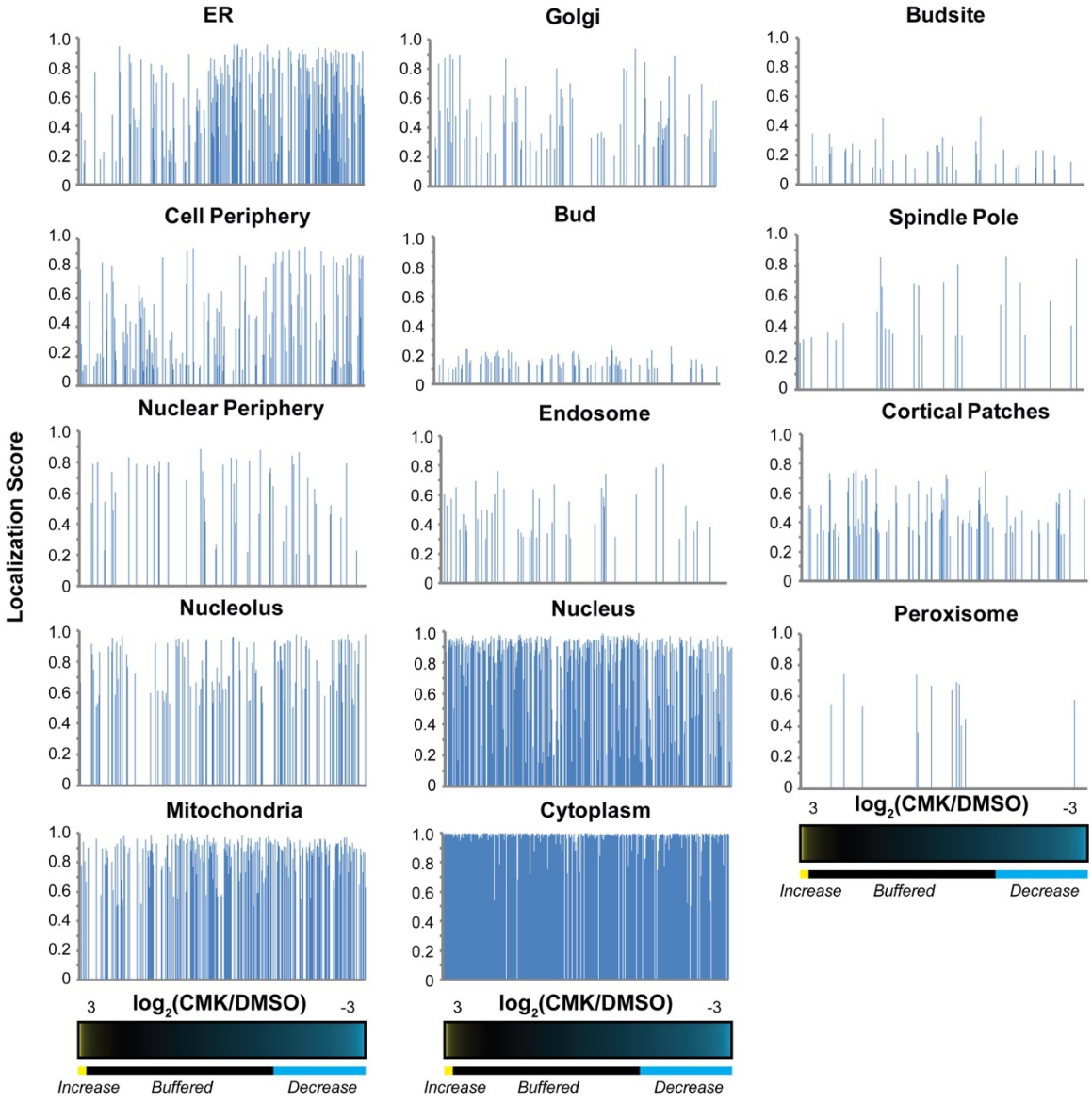



**Figure 3.22 Subcellular Localization of Proteins Ranked by mRNA Stability in Inhibited *kin28is***

Proportion of subcellular localization in listed compartments is plotted in order of decreasing expression change, indicated below the graphs. Subcellular localization data is from the CYCLoPS database [44].

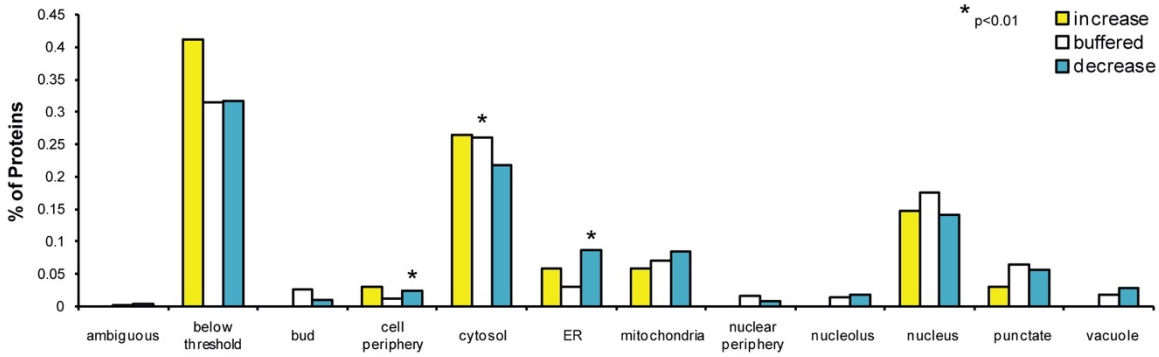



**Figure 3.23 Enrichment of Subcellular Localization of Proteins**

The normal cellular localization of proteins encoded by mRNA that are increased, buffered, or decreased when Kin28 is inhibited is plotted as a percent of the respective group. Localization data is from LoQAtE [45], which is derived from microscopy of GFP-tagged proteins. The proportion of genes that encode cell periphery and ER-associated proteins is significantly greater in the group of genes with decreased expression when Kin28 is inhibited; conversely, the proportion of genes that encode cytosolic proteins is greater in the group of genes with buffered expression, z-test two-tailed *$p < 0.01$.

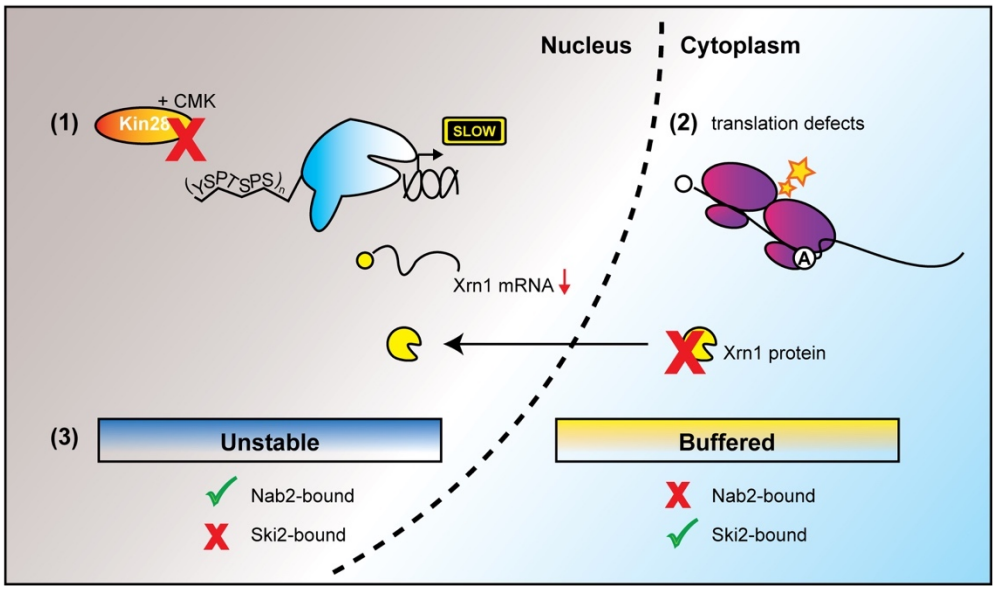



**Figure 3.24 Model Depicting the Post-transcriptional Events after Inhibition of Kin28**

(1) Inhibiting the nuclear kinase Kin28 causes a massive decrease in nascent mRNA synthesis. (2) In turn, activity of the cytoplasmic exonuclease Xrn1 is reduced, through a decrease in Xrn1 mRNA and/or translocation of Xrn1 protein into the nucleus. Conversely, the loss of transcription de-represses m6A methyltransferase Ime4 expression, which potentially increases deposition of m6A. Translation is reduced, at least in part, because the pool of translatable mRNAs is decreased and the obstruction from m6A. (3) While the levels of many cytoplasmic Ski2-bound mRNAs are buffered because of loss of cytoplasmic Xrn1 activity, the levels of nuclear Nab2-bound mRNAs are severely reduced through nuclear surveillance mechanisms.

**Table 3.3 Gene Ontology Analysis of Unstable and Induced mRNAs**

Gene ontology (GO) terms associated with genes that are unstable or induced upon Kin28 inhibition. P-values displayed are Bonferroni-corrected. The number of genes in the indicated group (k) —unstable or induced—is indicated along with the total number of mRNA in the GO category (f).

**Repressed/Unstable ($Log_2$CMK/DMSO < -1.5, RNA-Seq; n = 335)**

| | **p-value** | **k** | **f** |
|---|---|---|---|
| nucleosomal protein complex | 5.8E-09 | 7 | 8 |
| vacuole | 1.7E-08 | 33 | 224 |
| purine anabolism | 6.7E-08 | 11 | 29 |
| cell periphery | 8.7E-08 | 31 | 216 |
| membrane (ER, Golgi, nuclear envelope) | < E-06 | - | - |
| 20S proteasome | 3.5E-06 | 7 | 15 |
| cell wall | 5.8E-05 | 25 | 213 |
| MFS proteins | 6.6E-05 | 19 | 140 |

**Induced ($Log_2$CMK/DMSO > 1, RNA-Seq; n = 92)**

| | p-value | k | f |
|---|---|---|---|
| PP1-associated | 1.0E-05 | 4 | 11 |
| meiosis | 2.8E-05 | 10 | 146 |

upstream transcripts [38]. Interestingly, in the *XRN1* deletion strain, two of the sporulation genes tested (*SPO21* and *SSP2*) were over three-fold that of the wild-type strain (**Figure 3.25**), which emphasizes the similarity in the Kin28-inhibited transcriptome to that of *XRN1*-deleted yeast. Early studies of Xrn1 highlighted its essential function in successful meiosis [46], which has largely been overshadowed by its role as the master regulator of a cell's steady-state mRNA levels [13, 14]. Potentially, deletion of *XRN1* and inhibition of Kin28, which effectively results in decreased Xrn1 activity, operate in the same pathway that releases the transcriptional block or hyper-degradation of sporulation-associated genes.

How exactly Xrn1 exerts differential activity during transcriptional crises as in inhibited *kin28is* was, prior to this study, an open question. We show here that Ski2-binding, relative to Nab2-binding, may define mRNAs that are buffered when Kin28 is inhibited. *SKI2* mutations are synthetic lethal with *XRN1* deletion [47], which hints at a role for Ski2 in Xrn1-related pathways. Ski2 is a RNA helicase associated with the cytoplasmic exosome (notably subunit Rrp44) in 3'-to-5' RNA decay [42, 48] and in the decay and prevention of translation of mRNAs lacking a polyA tail [49]. Nab2, on the other hand, has been implicated in co-transcriptional mRNA end processing, polyA-binding, and with protecting nascent mRNAs from the nuclear exosome (notably subunits Rrp6 and Dis3) [43, 50, 51]. The relative binding of these factors may reflect a difference in localization of the bound mRNAs.

As such, localization of a given mRNA to the nucleus (or lack of export to the cytoplasm) could contribute to its decay during inhibition of Kin28 by nuclear decay factors whose activity remains intact. Furthermore, Xrn1 has been shown to mediate the transit of decay factors Ccr4, Dcp1/2, Edc3, and Lsm1 from the cytoplasm to the nucleus, where they accumulate in response to heat shock and starvation [13]. In the

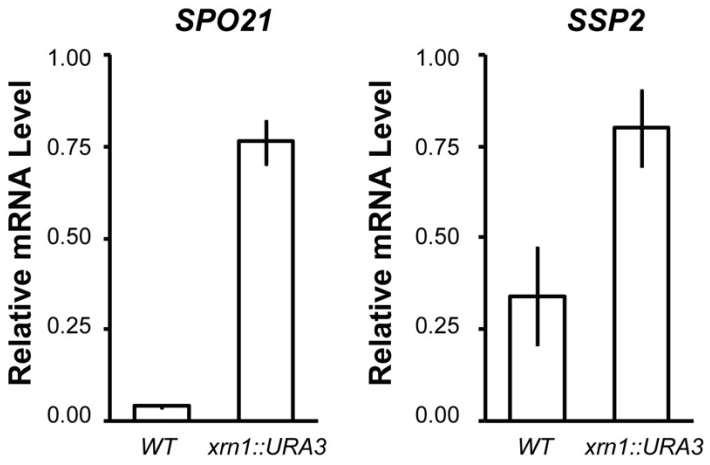



**Figure 3.25 Expression of Meiosis-Related mRNA in *XRN1*-deleted Yeast**

The expression of induced transcripts *SPO21* and *SSP2* in the wild-type (WT) and *xrn1::URA3* strain was determined by RT-qPCR and plotted as bars; mean +/- SEM, n=3-5 biological replicates.

same study, Xrn1 itself was also shown to reside in the nucleus where it could modulate transcription under stress, a feature that required its intact RNA binding and degradation capacity. It will be of interest to determine if Xrn1 and these other decay factors are enriched in the nucleus of inhibited Kin28 cells, thus accelerating the decay of nuclear, Nab2-bound mRNAs.

Potentially, the significance of the increased Nab2-to-Ski2 binding on unstable mRNAs goes beyond differential export and localization of mRNA. One possibility is that differential Nab2/Ski2 binding may instead reflect differences in polyA length [52, 53], which are known enhancers of translation and mRNA turnover [51]. The mammalian homologs of Nab2 and Ski2, PABPN1 and SKIV2L, respectively, share similar functions as their yeast counterparts [54], and may also be linked to mRNA stability in CDK7 inhibited regimes.

### 3.9 Materials and Methods

*Yeast Strains and Growth Conditions*

Unless specified otherwise, cells were treated for 1 hour with 5 μM CMK and harvested during early log-phase (OD600 between 0.3-0.7) in histidine-dropout media with 2% dextrose or standard yeast peptone dextrose complete media with 2% dextrose. Analog-sensitive kinase alleles were introduced using the two-step allele replacement strategy described in [12]. Strains used in this study are listed in Table 3.4.

*RNA Purification*

RNA was extracted from cells by incubating with occasional vortexing in TE Lysis Buffer (10 mM TrisHCl pH 7.5, 10 mM EDTA, 0.5% SDS) and acidic phenol at 65 °C, followed by ethanol precipitation. All RNA preparations were treated with DNase I (Ambion) prior to downstream assays.

*TT-Seq Library Preparation*

Nascent RNA was metabolically labeled with 4-thiouracil (4tU, Sigma) as previously described [12]. Briefly, 5 mM 4tU was added to cell media 5 minutes before harvest. An aliquot of cells was removed for counting, and cell pellets were flash frozen in liquid nitrogen and stored at -80 °C until lysis and RNA purification. 4tU-labeled *S. pombe* cells were added to each pellet at a ratio of 1:6 *S. pombe* to *S. cerevisiae* before RNA purification as detailed above. 100 µg of RNA was diluted to 240 ng/µL in Biotinylation Buffer (10 mM HEPES pH 7.5, 1 mM EDTA) and fragmented on a BioRuptor at high power for 24 minutes total time (12 cycles of 30 seconds/30 seconds ON/OFF). Biotin was conjugated to nascent RNA by adding 5 µg of MTSEA biotin-XX (Biotium) and nutating at room temperature for 1 hr covered in aluminum foil. Before pulldown with 100 µL of Dynabeads Streptavidin C1 (Thermo Fisher Scientific), RNA was purified using 25:24:1 phenol:chloroform:isoamyl alcohol followed by isopropanol precipitation. Nascent biotinylated RNA was eluted from the beads with 5% BME. A final ethanol precipitation was performed with 1 uL of GlycoBlue (Thermo Fisher Scientific). ScriptSeq Complete Gold Kit for Yeast (Epicentre) was used for preparing stranded rRNA-depleted sequencing libraries. Single end, 100 bp reads were collected on an Illumina HiSeq 2500.

*RT-qPCR*

cDNA was prepared using the SuperScript IV First-Strand Synthesis System (Thermo Fisher Scientific) from 1 µg of RNA using random hexamers. Before qPCR, cDNA was treated with RNaseH (Thermo Fisher Scientific). qPCR was performed on an ABI 7500 machine with PowerUp SYBR Green Master Mix (Thermo Fisher Scientific) under standard cycling conditions. Primers used for qPCR are listed in Table 3.5.

*Polysome Profiling*

Sucrose was dissolved in Polysome Profile Buffer (20 mM TrisHCl pH 7.5, 10 mM $MgCl_2$, 100 mM KCl) to 7, 17, 27, 37, and 47% weight by volume. Sucrose gradients were formed by layering sucrose solutions into an ultracentrifuge tube, freezing in liquid nitrogen between layers. Sucrose gradients were thawed overnight at 4 °C before use or stored at -20 °C. Cycloheximide (CHX, 100 µg/mL final concentration, DOT Scientific Inc.) was added to cells immediately before harvest. Cells were lysed in Lysis Buffer (2 mM DTT, 1% Triton X-100, 1 µL/mL RNase Inhibitor (Promega), 100 µg/mL CHX, 1X EDTA-free cOmplete Protease Inhibitor (Roche), in Polysome Profile Buffer) by bead-beating at 4 °C, five cycles of 2 minutes with 2 minutes rest. Lysates (15 A260 units) were layered onto the top of sucrose gradients and separated by spinning at 45,000 rpm (SW50.1 Ti rotor) for 1 hour and 20 minutes at 4 °C with 3 minutes acceleration and deceleration. Profiles and fractions were collected on an ISCO Foxy Jr with PeakTrak software.

*Fluorescence Microscopy*

Images were collected on a Nikon Eclipse Ti system. Cells grown in YPD without riboflavin and folic acid were plated on Nunc 96-well coverslip bottom plates that were pre-treated with concanavalin A (2 mg/mL concanavalin A, 5 mM $CaCl_2$, 5 mM $MnCl_2$). Approximately 15 minutes after plating, 1 µL of DMSO or CMK (equivalent to 1% of the total volume) was added. Images were processed using ImageJ software with the plugin StackReg [55] to align stacks.

*Nascent Protein Labeling*

Methionine-prototroph *w303a kin28is* cells were used for these experiments, though labeling was comparable in BY4741 yeast. Cells were scraped into 15 mL of

methionine-dropout media to an OD600 of ~0.3 and were treated with either DMSO or CMK for 10 minutes before addition of 100 μM L-HPG (Click Chemistry Tools). The 'time 0' samples were collected before addition of L-HPG. Cells were harvested by pelleting 5 mL samples, rinsing in water, and flash freezing in liquid nitrogen until lysis. Pellets were lysed by bead-beating at 4 °C for 10 minutes in 80 μL of HPG Lysis Buffer (50 mM TrisHCl pH 8, 1% SDS 1X EDTA-free cOmplete Protease Inhibitor (Roche), 1 mM PMSF, 1 μg/mL pepstatin A, 1 μg/mL aprotinin). To 50 μL of lysate, 90 μL of 1X PBS and 20 μL of 2.5 mM Cy3-azide (Click Chemistry Tools) was mixed, followed by 10 μL each of 100 mM THPTA (Click Chemistry Tools), 20 mM $CuSO_4$, and 300 mM sodium ascorbate, vortexing gently after each addition. Click-chemistry reactions were incubated at room temperature, protected from light for 30 minutes. Proteins were purified with chloroform-methanol extraction before resuspending in 1X SDS Loading Buffer (Bio-Rad) and using for SDS-PAGE, described below. After running, gels were scanned using the Typhoon FLA 9500 for Cy3, followed by staining with Coomassie Blue.

*SDS-PAGE and Immunoblotting*

Lysates were prepared for SDS-PAGE by bead-beating cells and adding SDS Loading Buffer. Samples were heated for 5 minutes at 95 °C, and insoluble material was pelleted prior to loading onto tris-glycine gels. SDS-PAGE and transfer to nitrocellulose membranes were conducted with the Bio-Rad Mini-PROTEAN system following standard protocols. Blots were blocked for 1 hour in 5% milk in TBST and incubated with primary antibodies overnight at 4 °C at the following dilutions: anti-eIF2α (1:2000), anti-eIF2α (pS52) (1:1000, #44-728G, Thermo Fisher Scientific), anti-Mpp10 (1:10,000), anti-RpL3 (1:5000). The yeast eIF2α antibody was a generous gift of Dr. Thomas Dever; Mpp10 of Dr. Susan Baserga. Secondary antibodies were used at a 1:5000 dilution.

Blots were incubated with Clarity Western ECL substrate and developed on the ImageQuant LAS 4000.

*m7G and m6A RNA Immunoprecipitation*

For m7G, 10 µg of antibody (#RN016M, Medical and Biological Laboratories) was used per reaction. For m6A, 5 µg of antibody (#MABE1006, EMD Millipore) was used per reaction. Antibody was added to 20 uL of Protein A/G beads (Thermo Fisher Scientific) and incubated with nutating at 4 °C for 1 hour. Antibody-conjugated beads were washed twice with 500 µL of RIP Buffer (150 mM NaCl, 25 mM Tris pH 7.5, 5 mM EDTA, 0.5 mM DTT, 0.5% NP-40) and resuspended in 200 µL of RIP Buffer to which 1 µg RNA and 1 µL RNase Inhibitor was added. The reaction was incubated with nutating for 1 hour at room temperature. Beads were washed three times with 500 µL of RIP Buffer before eluting immunoprecipitated RNA with phenol-chloroform extraction followed by ethanol precipitation.

*5'P-Seq Library Preparation*

RNA was purified as described above, and the 5'P-Seq libraries were prepared with rRNA reduction as detailed in [56].

*Anchor-Away of Pol II*

Cells were treated with DMSO or CMK for 10 minutes before inducing dimerization of FRB and FKBP12, and thus anchor-away of Pol II in HHY170, by addition of 1 µg/mL rapamycin (LC Laboratories). The 'time 0' sample was collected before adding rapamycin. An aliquot of cells was removed for counting at each time point, and cell pellets were flash frozen in liquid nitrogen and stored at -80 °C until lysis

and RNA purification. *S. pombe* cells were added to each pellet at a ratio of 1:6 S. pombe to *S. cerevisiae* before RNA purification, described above.

*TT-Seq Data Analysis*

Reads were demultiplexed and quick statistics generated with FastQC. Reads were clipped of adapter sequences and aligned to a concatenated sacCer3 and *S. pombe* genome using STAR (v2.4.2a). Reads mapping to more than one position were filtered to prevent ambiguous alignment of sequences shared between species. Aligned reads were sorted and indexed with SAMtools and counted with HTSeq (0.6.0). S. cerevisiae read counts were normalized with loess regression using S. pombe read counts to fit the loess as described previously [12]. Differential gene expression analysis was performed with DESeq. Alignments were scaled and converted to bigwig files with deepTools 2 and visualized in IGV. The heatmap of correlations between inhibited *kin28is* differential gene expression and the matrix in Figure 1 was plotted with R 'cormat' and 'corrplot' packages.

*5'P-Seq Data Analysis*

Reads were demultiplexed and quick statistics generated with FastQC. UMI-tools was used to move the unique molecular identifier (UMI) into the read header before alignment to sacCer3 with STAR; clipped 5' reads were filtered with –alignEndsType Extend5pOfRead1. Aligned reads with a quality score over 30 were sorted and indexed with SAMtools. UMI-tools was used to de-duplicate reads. The 5' nucleotide was plotted with --Offset 1 and scaled with --normalizeUsingRPKM during conversion to the bigwig format with deepTools 2. Alignments were visualized in IGV. Signal was extracted across ORFs with deepTools 2 with single nucleotide bin sizes. Analysis was continued only on transcripts that had over 100 RNA-Seq reads and over 32 5'P-Seq reads per

base in at least one condition to reduce the effect of noise in interpretation of data. Heatmaps of 5'P signal and relative Nab2/Ski2 binding intensity were generated with Java TreeView 3.0 and clustered with Cluster 3.0. To calculate the spectral density at 3-nucleotide, the R package 'spectrum' was used on 5'P signal over the window [2:47] with base 1 being the first nucleotide of the start codon or -100 from the last nucleotide of the stop codon. The proportion of 5'P reads in the ribosome-protected frame was calculated for each gene across 600 nucleotides (200 frames) by dividing the read count in frame 1 by the sum of read counts in frame 1, 2, and 3. The smoothed average proportion for genes without introns and longer than 600 nucleotides was plotted as a spline with R 'smooth.spline' function. Gene ontology analysis was performed with FunSpec.

*RNA Binding Protein Bioinformatic Analysis*

The CRAC dataset from Tuck and Toll was downloaded from GEO, and UCSC's bigWigAverageOverBed software was used to obtain a count/base for each transcript and each RNA-binding protein. The resultant matrix was quantile-normalized across RNA-binding protein arrays with the R 'preprocessCore' package and 'normalize.quantiles' function. Next, Cluster 3.0 was used to determine relative binding across genes (mean-centered).

*Quantification and Statistical Analyses*

Quantitative data are presented as the mean and standard error from biological replicates. To determine statistical significance between groups, comparisons were made using Student's two-tailed t tests and ANOVA, using VassarStats (http://vassarstats.net/). The number of biological replicates (n), test used, and p values are indicated in the legend of each figure. For immunoblots, the reported images are representative of at least three independent experiments.

**Table 3.4 Yeast Strains Used in This Chapter**

| Relevant Genotype | Reference |
|---|---|
| *S. cerevisiae*: wild-type: *MAT***a** *his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 RPB3-TAP::HIS3* | [57] |
| *S. cerevisiae*: kin28is: *MAT***a** *his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 RPB3-TAP::HIS3 kin28 L83G, V21C* | [12] |
| *S. cerevisiae*: HHY170 kin28is: MAT**α** *tor1-1 fpr::NAT RPL13A-2xFKBP12::TRP1 RPO21-FRB::kanMX6 kin28 L83G, V21C* | [12] |
| *S. cerevisiae*: gcn2Δ: gcn2Δ MAT**α** ura3-52 trp1-Δ63 leu2-3 leu2-112 GAL2+ | Dr. Thomas E. Dever |
| *S. cerevisiae*: xrn1::URA3: *MAT***a** *his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 RPB3-TAP::HIS3 xrn1::URA3* | This Study |
| *S. cerevisiae*: xrn1::URA3 kin28is: *MAT***a** *his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 RPB3-TAP::HIS3 xrn1::URA3 kin28 L83G, V21C* | This Study |
| *S. cerevisiae*: w303a: *MAT***a** *leu2-3,112 trp1-1 can1-100 ura3-1 ade2-1 his3-11,15* | [58] |
| *S. cerevisiae*: w303a kin28is: *MAT***a** *leu2-3,112 trp1-1 can1-100 ura3-1 ade2-1 his3-11,15 kin28 L83G, V21C* | This Study |
| *S. cerevisiae*: Edc3-GFP: *MAT***a** *his4-539 trp1 leu2-3,112 ura3-52 EDC3-GFP-NEO cup1::LEU2/PGK1pG/MFA2pG* | Dr. Roy Parker |
| *S. cerevisiae*: Pab1-GFP: *MAT***a** *his4-539 trp1 leu2-3,112 ura3-52 PAB1-GFP-NEO cup1::LEU2/PGK1pG/MFA2pG* | Dr. Roy Parker |
| *S. cerevisiae*: Edc3-GFP kin28is: *MAT***a** *his4-539 trp1 leu2-3,112 ura3-52 EDC3-GFP-NEO cup1::LEU2/PGK1pG/MFA2pG kin28 L83G, V21C* | This Study |
| *S. cerevisiae*: Pab1-GFP kin28is: *MAT***a** *his4-539 trp1 leu2-3,112 ura3-52 PAB1-GFP-NEO cup1::LEU2/PGK1pG/MFA2pG kin28 L83G, V21C* | This Study |

**Table 3.5 Oligonucleotides Used in This Chapter**

| Name | Sequence (5'-to-3') |
|---|---|
| GAS1-Forward | TCAGCGACAAGAACTACAGAAA |
| GAS1-Reverse | CATCATCACCACAAGCGAAATAA |
| POP5-Forward | CCTGCGGATGTGTCCATAAA |
| POP5-Reverse | TTCAACTGCAAGAGAGAGTTACA |
| RPS28A-Forward | GTCACTTTAGCCAAGGTCATCAA |
| RPS28A-Reverse | GCCCTTCACGTTTCTGACAATA |
| SEM1-Forward | AGCAATGCTGTAACGCAAAC |
| SEM1-Reverse | GGCCTTCAATTCATTGGTGAAA |
| SPO21-Forward | TGCTCTTCTTCTTCATCATCATCT |
| SPO21-Reverse | CTCATAATCGATGCGGTGTTAATG |
| SSP2-Forward | GCACAGGCATCTCTAGTGTATT |
| SSP2-Reverse | AAAGAACGTTCCGGAGTATCG |
| TDH3-Forward | GTTGCTTTGAACGACCCATTC |
| TDH3-Reverse | GGAAACTTCACCAGCGTATCT |
| VMA3-Forward | GGTTGTGCCTCTGCAATTATC |
| VMA3-Reverse | GGTCTGGTCTCAACACACAA |
| YTA7-Forward | CAGAACGAACGAAGAGCTTACTA |
| YTA7-Reverse | CTGCTTCTTCCTCCTCTTCATC |

## 3.10 References


1. Akhtar, M.S., et al., *TFIIH Kinase Places Bivalent Marks on the Carboxy-Terminal Domain of RNA Polymerase II.* Molecular Cell, 2009. **34**(3): p. 387-393.

2. 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**(20): p. 5455-5464.

3. Kim, M., et al., *Phosphorylation of the Yeast Rpb1 C-terminal Domain at Serines 2, 5, and 7.* Journal of Biological Chemistry, 2009. **284**(39): p. 26421-26426.

4. Rodriguez-Molina, J.B., et al., *Engineered Covalent Inactivation of TFIIH-Kinase Reveals an Elongation Checkpoint and Results in Widespread mRNA Stabilization.* Mol Cell, 2016. **63**(3): p. 433-44.

5. Wong, K.H., Y. Jin, and K. Struhl, *TFIIH phosphorylation of the Pol II CTD stimulates mediator dissociation from the preinitiation complex and promoter escape.* Mol Cell, 2014. **54**(4): p. 601-12.

6. Bentley, D.L., *Coupling mRNA processing with transcription in time and space.* Nature Reviews Genetics, 2014. **15**(3): p. 163-175.

7. Kwiatkowski, N., et al., *Targeting transcription regulation in cancer with a covalent CDK7 inhibitor.* Nature, 2014. **511**(7511): p. 616-620.

8. Bradner, J.E., D. Hnisz, and R.A. Young, *Transcriptional Addiction in Cancer.* Cell, 2017. **168**(4): p. 629-643.

9. Cayrol, F., et al., *THZ1 targeting CDK7 suppresses STAT transcriptional activity and sensitizes T-cell lymphomas to BCL2 inhibitors.* Nature Communications, 2017. **8**: p. 14290.

10. Chipumuro, E., et al., *CDK7 Inhibition Suppresses Super-Enhancer-Linked Oncogenic Transcription in MYCN-Driven Cancer.* Cell, 2014. **159**(5): p. 1126-1139.

11. Franco, Hector L. and W.L. Kraus, *No Driver behind the Wheel? Targeting Transcription in Cancer.* Cell, 2015. **163**(1): p. 28-30.

12. Rodríguez-Molina, Juan B., et al., *Engineered Covalent Inactivation of TFIIH-Kinase Reveals an Elongation Checkpoint and Results in Widespread mRNA Stabilization.* Molecular Cell, 2016. **63**(3): p. 433-444.

13. Haimovich, G., et al., *Gene Expression Is Circular: Factors for mRNA Degradation Also Foster mRNA Synthesis.* Cell, 2013. **153**(5): p. 1000-1011.

14. Sun, M., et al., *Global Analysis of Eukaryotic mRNA Degradation Reveals Xrn1-Dependent Buffering of Transcript Levels.* Molecular Cell, 2013. **52**(1): p. 52-62.

15. Timmers, H.T.M. and L. Tora, *Transcript Buffering: A Balancing Act between mRNA Synthesis and mRNA Degradation.* Molecular Cell, 2018. **72**(1): p. 10-17.

16. Schwalb, B., et al., *TT-seq maps the human transient transcriptome.* Science, 2016. **352**(6290): p. 1225-1228.

17. Duffy, Erin E., et al., *Tracking Distinct RNA Populations Using Efficient and Reversible Covalent Chemistry.* Molecular Cell, 2015. **59**(5): p. 858-866.

18. Geisberg, Joseph V., et al., *Global Analysis of mRNA Isoform Half-Lives Reveals Stabilizing and Destabilizing Elements in Yeast.* Cell, 2014. **156**(4): p. 812-824.

19. Neymotin, B., V. Ettore, and D. Gresham, *Multiple Transcript Properties Related to Translation Affect mRNA Degradation Rates in Saccharomyces cerevisiae.* G3: Genes|Genomes|Genetics, 2016.

20. Miller, C., et al., *Dynamic transcriptome analysis measures rates of mRNA synthesis and decay in yeast.* Molecular Systems Biology, 2014. **7**(1): p. 458-458.

21. Presnyak, V., et al., *Codon Optimality Is a Major Determinant of mRNA Stability.* Cell, 2015. **160**(6): p. 1111-1124.

22. Park, D., et al., *Simultaneous mapping of transcript ends at single-nucleotide resolution and identification of widespread promoter-associated non-coding RNA governed by TATA elements.* Nucleic Acids Research, 2014. **42**(6): p. 3736-3749.

23. Lelivelt, M.J. and M.R. Culbertson, *Yeast Upf Proteins Required for RNA Surveillance Affect Global Expression of the Yeast Transcriptome.* Molecular and Cellular Biology, 1999. **19**(10): p. 6710-6719.

24. Hu, W., et al., *Co-translational mRNA decay in Saccharomyces cerevisiae.* Nature, 2009. **461**(7261): p. 225-229.

25. Wiltschi, B., et al., *Expanding the genetic code ofSaccharomyces cerevisiaewith methionine analogues.* Yeast, 2008. **25**(11): p. 775-786.

26. Gingras, A.-C., B. Raught, and N. Sonenberg, *eIF4 Initiation Factors: Effectors of mRNA Recruitment to Ribosomes and Regulators of Translation.* Annual Review of Biochemistry, 1999. **68**(1): p. 913-963.

27. Paterson, B.M. and M. Rosenberg, *Efficient translation of prokaryotic mRNAs in a eukaryotic cell-free system requires addition of a cap structure.* Nature, 1979. **279**(5715): p. 692-696.

28. Kanin, E.I., et al., *Chemical inhibition of the TFIIH-associated kinase Cdk7/Kin28 does not impair global mRNA synthesis.* Proceedings of the National Academy of Sciences, 2007. **104**(14): p. 5812-5817.

29. Schroeder, S.C., *Dynamic association of capping enzymes with transcribing RNA polymerase II.* Genes & Development, 2000. **14**(19): p. 2435-2440.

30. Slobodin, B., et al., *Transcription Impacts the Efficiency of mRNA Translation via Co-transcriptional N6-adenosine Methylation.* Cell, 2017. **169**(2): p. 326-337.e12.

31. Sonenberg, N. and A.G. Hinnebusch, *Regulation of Translation Initiation in Eukaryotes: Mechanisms and Biological Targets.* Cell, 2009. **136**(4): p. 731-745.

32. Zhan, K., J. Narasimhan, and R.C. Wek, *Differential Activation of eIF2 Kinases in Response to Cellular Stresses in Schizosaccharomyces pombe.* Genetics, 2004. **168**(4): p. 1867-1875.

33. Dever, T.E., et al., *Phosphorylation of initiation factor 2α by protein kinase GCN2 mediates gene-specific translational control of GCN4 in yeast.* Cell, 1992. **68**(3): p. 585-596.

34. Parker, R. and U. Sheth, *P Bodies and the Control of mRNA Translation and Degradation.* Molecular Cell, 2007. **25**(5): p. 635-646.

35. Zhou, J., et al., *N 6 -Methyladenosine Guides mRNA Alternative Translation during Integrated Stress Response.* Molecular Cell, 2018. **69**(4): p. 636-647.e7.

36. Pelechano, V., W. Wei, and Lars M. Steinmetz, *Widespread Co-translational RNA Decay Reveals Ribosome Dynamics.* Cell, 2015. **161**(6): p. 1400-1412.

37. Gasch, A.P., et al., *Genomic Expression Programs in the Response of Yeast Cells to Environmental Changes.* Molecular Biology of the Cell, 2000. **11**(12): p. 4241-4257.

38. Cheng, Z., et al., *Pervasive, Coordinated Protein-Level Changes Driven by Transcript Isoform Switching during Meiosis.* Cell, 2018. **172**(5): p. 910-923.e16.

39. Haruki, H., J. Nishikawa, and U.K. Laemmli, *The Anchor-Away Technique: Rapid, Conditional Establishment of Yeast Mutant Phenotypes.* Molecular Cell, 2008. **31**(6): p. 925-932.

40. Tuck, Alex C. and D. Tollervey, *A Transcriptome-wide Atlas of RNP Composition Reveals Diverse Classes of mRNAs and lncRNAs.* Cell, 2013. **154**(5): p. 996-1009.

41. Anderson, J.T., et al., *NAB2: a yeast nuclear polyadenylated RNA-binding protein essential for cell viability.* Molecular and Cellular Biology, 1993. **13**(5): p. 2730-2741.

42. Anderson, J.S.J., *The 3′ to 5′ degradation of yeast mRNAs is a general mechanism for mRNA turnover that requires the SKI2 DEVH box protein and 3′ to 5′ exonucleases of the exosome complex.* The EMBO Journal, 1998. **17**(5): p. 1497-1506.

43. Tudek, A., et al., *A Nuclear Export Block Triggers the Decay of Newly Synthesized Polyadenylated RNA.* Cell Reports, 2018. **24**(9): p. 2457-2467.e7.

44. Koh, J.L., et al., *CYCLoPs: A Comprehensive Database Constructed from Automated Analysis of Protein Abundance and Subcellular Localization Patterns in Saccharomyces cerevisiae.* G3 (Bethesda), 2015. **5**(6): p. 1223-32.

45. Breker, M., et al., *LoQAtE--Localization and Quantitation ATlas of the yeast proteomE. A new tool for multiparametric dissection of single-protein behavior in response to biological perturbations in yeast.* Nucleic Acids Res, 2014. **42**(Database issue): p. D726-30.

46. Solinger, J.A., D. Pascolini, and W.-D. Heyer, *Active-Site Mutations in the Xrn1p Exoribonuclease ofSaccharomyces cerevisiaeReveal a Specific Role in Meiosis.* Molecular and Cellular Biology, 1999. **19**(9): p. 5930-5942.

47. Johnson, A.W. and R.D. Kolodner, *Synthetic lethality of sep1 (xrn1) ski2 and sep1 (xrn1) ski3 mutants of Saccharomyces cerevisiae is independent of killer virus and suggests a general role for these genes in translation control.* Molecular and Cellular Biology, 1995. **15**(5): p. 2719-2727.

48. Araki, Y., *Ski7p G protein interacts with the exosome and the Ski complex for 3'-to-5' mRNA decay in yeast.* The EMBO Journal, 2001. **20**(17): p. 4684-4693.

49. Searfoss, A.M. and R.B. Wickner, *3' poly(A) is dispensable for translation.* Proceedings of the National Academy of Sciences, 2000. **97**(16): p. 9133-9137.

50. Klass, D.M., et al., *Quantitative proteomic analysis reveals concurrent RNA-protein interactions and identifies new RNA-binding proteins in Saccharomyces cerevisiae.* Genome Research, 2013. **23**(6): p. 1028-1038.

51. Schmid, M., et al., *The Nuclear PolyA-Binding Protein Nab2p Is Essential for mRNA Production.* Cell Reports, 2015. **12**(1): p. 128-139.

52. Beilharz, T.H. and T. Preiss, *Widespread use of poly(A) tail length control to accentuate expression of the yeast transcriptome.* RNA, 2007. **13**(7): p. 982-997.

53. Subtelny, A.O., et al., *Poly(A)-tail profiling reveals an embryonic switch in translational control.* Nature, 2014. **508**(7494): p. 66-71.

54. Kühn, U. and E. Wahle, *Structure and function of poly(A) binding proteins.* Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression, 2004. **1678**(2-3): p. 67-84.

55. Thevenaz, P., U.E. Ruttimann, and M. Unser, *A pyramid approach to subpixel registration based on intensity.* IEEE Transactions on Image Processing, 1998. **7**(1): p. 27-41.

56. Pelechano, V., W. Wei, and L.M. Steinmetz, *Genome-wide quantification of 5'-phosphorylated mRNA degradation intermediates for analysis of ribosome dynamics.* Nature Protocols, 2016. **11**(2): p. 359-376.

57. Ghaemmaghami, S., et al., *Global analysis of protein expression in yeast.* Nature, 2003. **425**(6959): p. 737-41.

58. Rothstein, R.J., *One-step gene disruption in yeast.* Methods Enzymol, 1983. **101**: p. 202-11.

# Chapter 4: Exploring the Dark Side of RNA Expression: Exonucleases and Decay

This chapter is composed of unpublished, preliminary data.

## 4.1 Introduction

RNA is constantly being degraded and synthesized in cells, with the median half-life of budding yeast mRNA estimated as 2-30 minutes under normal circumstances [1]. Interestingly, in many instances where insults are made to the synthesis of new RNA in a cell, cellular mechanisms which stabilize pre-existing RNA are initiated [2]. Relatedly, when RNA decay is defective, RNA synthesis decreases to prevent accumulation of abnormally high levels of steady-state RNA. This phenomenon of maintenance of steady-state RNA levels is known as 'RNA buffering.' Xrn1, a 5'-to-3' processive RNA exonuclease [3, 4], has been implicated in regulating RNA buffering, but the mechanism of action remains unclear. Haimovich *et al*. report Xrn1 translocates to the nucleus and generally promotes transcription at gene promoters when RNA levels in the cytoplasm are depleted [5]. However, another report published contemporaneously to [5] was unable to detect Xrn1-chromatin interactions, and instead attributed regulation of buffering to the transcription factor Nrg1 [2]. How Nrg1 and other decay factors, such as Xrn1, can 'sense' defective decay or synthesis remains to be determined.

In the previous Chapters, I have shown inhibition of Kin28 kinase leads to decreased mRNA synthesis and induction of RNA buffering. However, a subset of mRNAs are not stabilized when RNA synthesis is perturbed, and I identified a link between these unstable RNAs and nuclear decay. The connection between nuclear localization and sensitivity to Kin28 inhibition led us to examine whether deletion of nuclear decay factors might rescue expression of the Kin28-sensitive transcripts. We focused on Rai1, a nuclear exonuclease with RNA pyrophosphohydrolase and phosphodiesterase activity toward 5' triphosphates and unmethylated mRNA caps, respectively [6-10]. Both activities result in an mRNA with a 5' monophosphate end, which is then vulnerable to decay by the processive 5'-3' exonucleases Rat1/XRN2 (nuclear) and Xrn1 (cytoplasmic).

In this Chapter, I present my preliminary investigations into the role of both cytoplasmic Xrn1 and nuclear Rai1 in RNA buffering in the context of *kin28is* inhibition. I show deletion of *XRN1* (*xrn1::URA3*), but not *RAI1* (*rai1::URA3*), results in a non-buffered state consistent with previous reports identifying Xrn1 as a crucial factor in maintaining RNA buffering. I show Xrn1 is detectable in both the nucleus and the cytoplasm, and localizes into distinct subcellular puncta during inhibition of Kin28. Interestingly, cells lacking Rai1 show increased levels of P body formation and defects in translation. Specifically, we show translation initiation stalls after formation of the 48S translation pre-initiation (PIC) complex in *rai1::URA3* cells. Although deletion of *RAI1* did not rescue the expression of Kin28-sensitive transcripts during inhibition of *kin28is*, it inadvertently allowed us to elucidate translation pre-initiation dynamics of inhibited Kin28 cells. Clarifying the functions of these decay factors will add to our understanding of a fundamental question in the field, namely, “How are signals that initiate RNA buffering (*i.e.* attenuation of decay or synthesis) transmitted to and from the nucleus, the site of transcription?”

### 4.2 Xrn1 Sub-cellular Localization Changes in Inhibited *kin28is*

To determine if the normally cytoplasmic exonuclease Xrn1 changes its cellular localization in inhibited *kin28is*, we tracked Xrn1-GFP with fluorescent microscopy in inhibited Kin28 cells (**Figure 4.1**). As previously observed [11, 12], Xrn1-GFP is abundant throughout the cell, but presents as predominantly cytoplasmic. When Kin28 is inhibited, the Xrn1-GFP fluorescence appears as distinct puncta. As discussed in Chapter 3, P bodies form when Kin28 is inhibited (Figure 3.8). Xrn1 is a known P body component [13], which when deleted in yeast, also results in increased P body formation. Thus, Xrn1 is likely localizing to P bodies when Kin28 is inhibited. To test for nuclear co-localization, we counterstained cells with DAPI. We did not observe

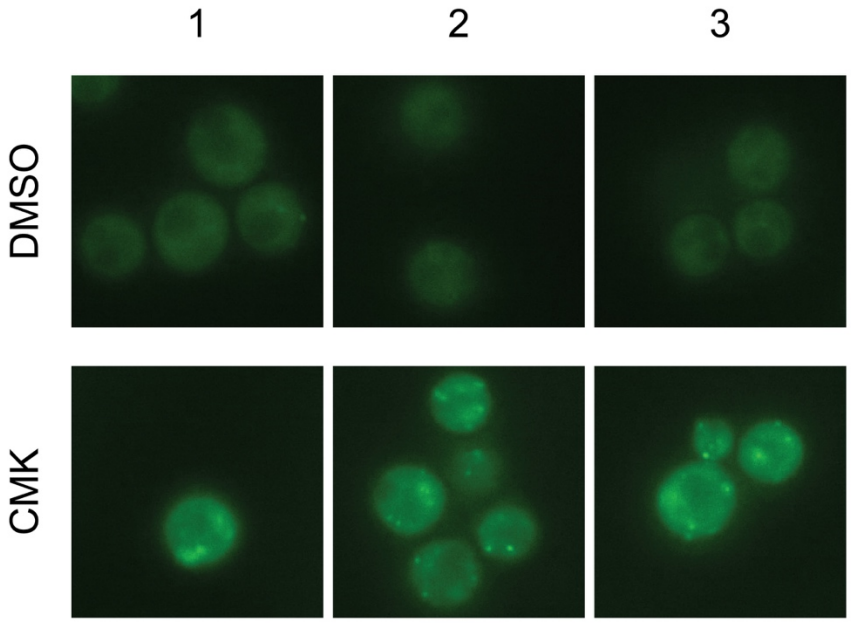



**Figure 4.1 Fluorescence Microscopy of Xrn1-GFP in *kin28is***

*kin28is* was inhibited with 5 μM of CMK or vehicle control (DMSO) for 1 hour. Three representative images from each treatment are shown.

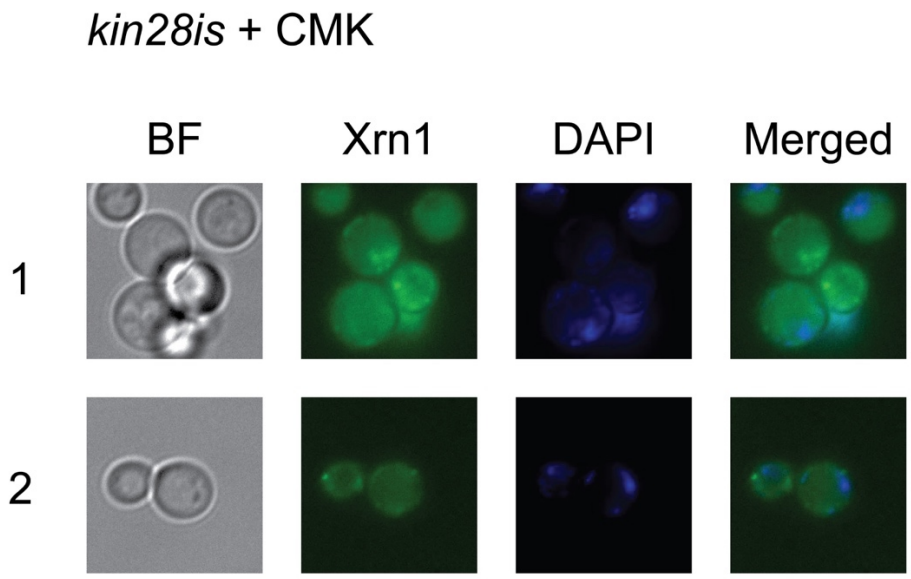



**Figure 4.2 Fluorescence Microscopy of DAPI-stained Xrn1-GFP in Inhibited *kin28is***

*kin28is* was inhibited with 5 μM of CMK for 1 hour and stained with nuclear dye DAPI. Two representative images of the bright-field (BF), Xrn1-GFP (Xrn1), and DAPI channel are shown. Merged images of the Xrn1-GFP and DAPI fluorescence are also presented.

significant overlap in the Xrn1-GFP puncta and the DAPI-stained regions (**Figure 4.2**). However, it is possible that a smaller portion of Xrn1, not readily detectable via our fluorescence microscopy method, is translocating to the nucleus.

Thus, we performed immunoblotting of Xrn1 with nuclear and cytoplasmic lysates of *kin28is* (**Figure 4.3**). As a negative control for the Xrn1 antibody, we also performed blotting in *xrn1::URA3*, which validates the specificity of the antibody for binding of Xrn1. Immunoblotting of nuclear histone protein H3 confirmed that our preparations of nuclear and cytoplasmic lysates were not cross-contaminated. Interestingly, a smaller sized form of histone H3 is abundant in *XRN1*-deleted yeast. Both H3 forms are reduced in protein level in inhibited Kin28 cells, as was predicted in Chapter 3 via 5'P-Seq analysis of translation. Xrn1 itself does not appreciably change in protein level during the time span of our Kin28 inhibition studies, and a subtle increase in nuclear levels of Xrn1 is apparent. Our immunoblotting of Xrn1 shows it localizes to both nuclear and cytoplasmic compartments, and potentially the shuttling between compartments occurs during RNA buffering regimes.

We also performed immunoblotting of Rpb4, a detachable subunit of Pol II, which is known to translocate from the nucleus to the cytoplasm; Rpb4 has been linked to enhanced cytoplasmic decay for certain mRNAs [14, 15]. Interestingly, similar to H3, we also found an increased abundance of a smaller sized form of Rpb4 in *xrn1::URA3* cells. However, we did not find shuttling of the Rpb4 protein to be dependent on Kin28 function. Taken together with our microscopy studies of Xrn1-GFP, we cannot exclude the possibility that a small amount of Xrn1 enters the nucleus when Kin28 is inhibited, but a majority of Xrn1 likely localizes to P bodies during inhibition of Kin28.

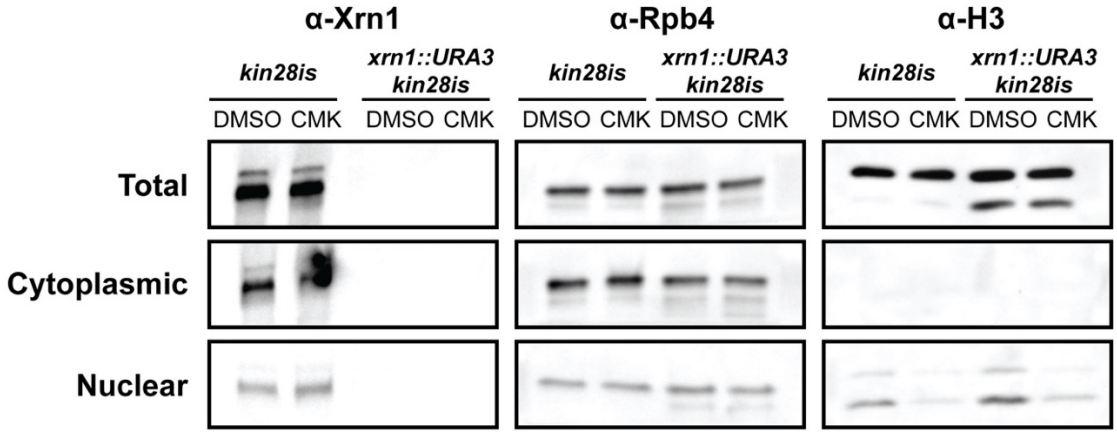



**Figure 4.3 Immunoblotting of Nuclear and Cytoplasmic Lysates of *kin28is***

*kin28is* and the *XRN1*-deleted strain (*xrn1::URA3 kin28is*) were inhibited with 5 μM of CMK or vehicle control (DMSO) for 1 hour. Total cellular lysates were divided into cytoplasmic and nuclear fractions and used in blotting of the indicated proteins.

### 4.3 Deletion of *XRN1* Short-circuits RNA Buffering

To elucidate their function in RNA buffering of inhibited *kin28is*, We deleted *XRN1* and *RAI1* and used RT-qPCR to quantify steady-state levels of mRNA (**Figure 4.4**). In line with its central role in RNA decay and buffering, deletion of *XRN1* resulted in a significant increase in the steady-state level of mRNAs tested. In contrast, *rai1::URA3* cells did not exhibit elevated steady-state levels of mRNAs, which is consistent with buffering observed in [2]. Rather, the steady-state level of the mRNAs tested were slightly decreased in *rai1::URA3*. As discussed in Chapter 3, deletion of *XRN1* was able to partially restore stability to mRNAs that were unstable in inhibited *kin28is*. Deletion of *RAI1* only marginally restores stability to an unstable mRNA in inhibited *kin28is* (*e.g. SEM1*) compared to *xrn1::URA3* deletion (**Figure 4.5**). Although a transcriptome-wide survey of expression changes in these exonuclease mutants would be more conclusive, this result suggests differential activity of Xrn1, rather than Rai1, is the main decay factor causing instability of mRNAs in inhibited *kin28is*.

### 4.4 *rai1::URA3* Cells Exhibit Defects in Translation

We also performed polysome profiling of the exonuclease mutant *kin28is* strains to test if translation—the process we hypothesized is involved with mediating RNA stability—was affected in the absence of these decay factors. Interestingly, 'halfmer peaks,' which are attributed to stalled 40S pre-initiation complexes, are visible as shoulders on the monosome (80S) and polysome peaks in profiles from exonuclease-deleted cells (**Figure 4.6**). A previous study [9] suggested translation pre-initiation stalling in *RAI1*-deleted yeast was caused by an insufficiency of mature 60S subunits, which could be reversed by overexpressing the Rai1 cofactor and rRNA processing exonuclease, Rat1. We also observe a greater ratio of 40S:60S ribosomal subunits in exonuclease-deleted yeast (**Figure 4.7A**, white bars). Interestingly, the halfmer peak

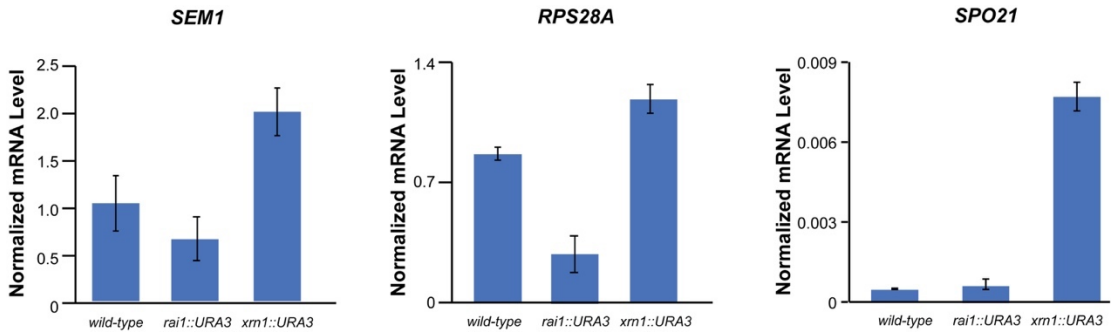



**Figure 4.4 RNA Levels in Exonuclease Deletion Strains**

RNA levels of *SEM1*, *RPS28A*, and *SPO21* were quantified after 1 hour inhibition of wild-type, *RAI1*-deleted (*rai1::URA3*), or *XRN1*-deleted (*xrn1::URA3*) cells. Expression was normalized to an *S. pombe* exogenous control before calculation of fold change; mean +/- SEM, n = 3-4 biological replicates.

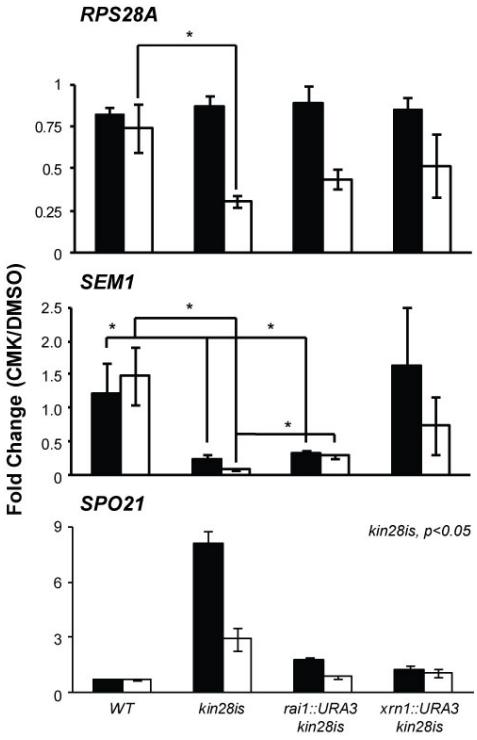



**Figure 4.5 Differential Xrn1 Activity Mediates Buffering**

Change in steady-state (Total, black bars) and newly synthesized (Nascent, white bars) mRNA levels of example transcripts after 1 hour inhibition of wild-type (WT), *kin28is*, *RAI1*-deleted (*rai1::URA3 kin28is*) or *XRN1*-deleted *kin28is* (*xrn1::URA3 kin28is*) cells. Expression was normalized to an *S. pombe* exogenous control before calculation of fold change; mean +/- SEM, n = 3-7 biological replicates.

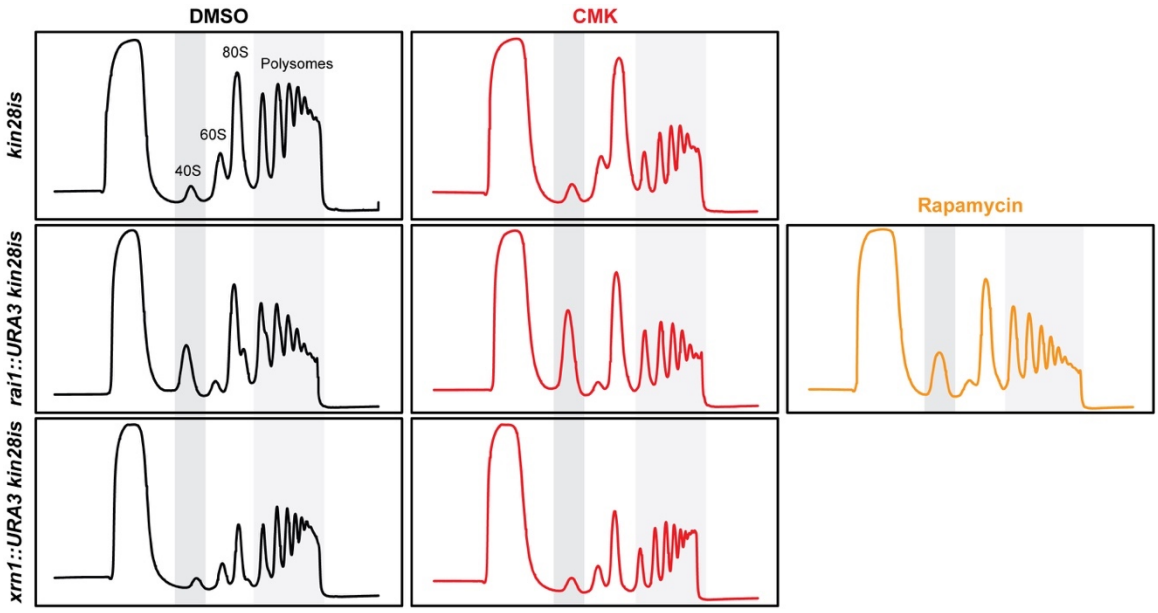



**Figure 4.6 Polysome Profiling of Exonuclease-Deleted *kin28is***

Polysome profiling of exonuclease-deleted *kin28is* cells treated with vehicle control (DMSO), *kin28is* inhibitor (CMK), or rapamycin (200 ng/mL) for 1 hour. The 40S, 60S, 80S, and polysome peaks are indicated in the *kin28is* DMSO panel. The 40S and Polysome peaks are highlighted in grey for ease of comparison. See also Figure 4.7.

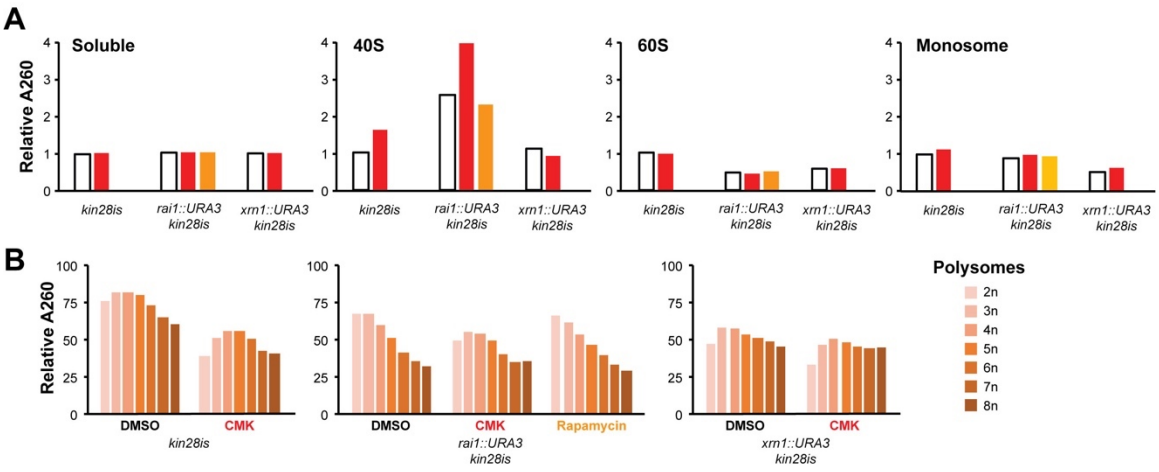



**Figure 4.7 Inhibition of Kin28 Disassembles Paused Pre-initiation Complexes**

**(A)** Bar graphs showing quantification of the relative A260 values of the profiles in Figure 4.6 for the soluble, 40S, 60S, and monosome peaks. All values are shown relative to the profile for *kin28is* treated with DMSO. **(B)** Bar graphs showing quantification of the A260 values of the polysome peaks for the profiles in Figure 4.6.

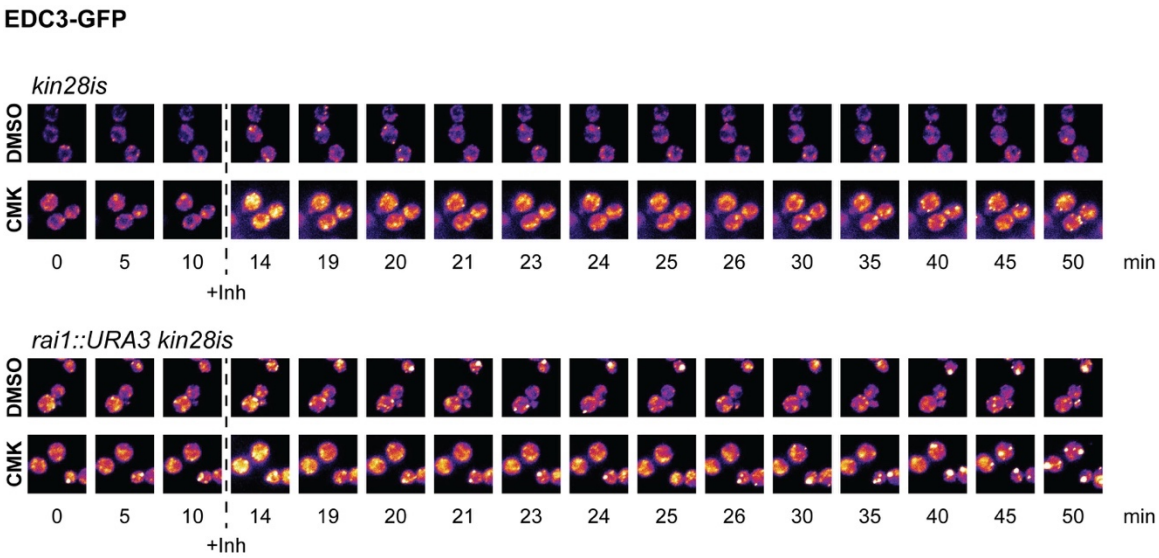



**Figure 4.8 Immunofluorescence of P Body Markers in *rai1::URA3 kin28is***

Representative images of *kin28is* cells of the wild-type or *RAI1*-deleted background (*rai1::URA3 kin28is*) expressing P body marker Edc3-GFP before and after the addition of vehicle control (DMSO) or inhibitor (CMK) for the indicated time.

disappears with inhibition of Kin28 without restoration of 60S levels (**Figure 4.7A**, red bars). Rather, the 40S peak increases in height, suggesting inhibition of Kin28 disrupts stalled translation pre-initiation complexes and returns 40S subunits to the pool of mRNA-unassociated subunits.

General translation repression via inhibition of the Tor1/2 kinase with rapamycin also causes the halfmer peak to disappear, but does not increase the level of free 40S (**Figure 4.7A**, orange bars) as was observed with inhibition of Kin28. Rapamycin represses Pol I and III transcription [16], prevents translation initiation via activation of Gcn2 and phosphorylation of eIF2α [17], and leads to the rapid degradation of pre-existing ribosomes (40-60%) [18]. In Chapter 3, I report inhibition of Kin28 leads to a defect, but not a global shutdown, of translation caused by sequestration of mRNAs to P bodies and a decrease in total mRNA levels. This is reflected in the decrease in polysome peaks in inhibited *kin28is* cells that is somewhat Xrn1-dependent (**Figure 4.7B**). Unlike the global shutdown to translation caused by rapamycin treatment, inhibition of Kin28 affects only a subset of mRNAs. Also unlike rapamycin treatment [19], inhibition of Kin28 causes formation of P bodies. Kin28-mediated effects on translation are distinct from the general translation stress induced by rapamycin, and appear protective of ribosomes, rather than destructive.

We hypothesized *rai1::URA3* cells might also show increased formation of P bodies given its defects in translation. To test this we utilized fluorescence microscopy of Edc3-GFP, a marker for P bodies. Indeed, *rai1::URA3* cells exhibit an increase in P body number compared to wild-type yeast (**Figure 4.8**). When Kin28 is inhibited in *rai1::URA3*, these P bodies grow larger and appear to coalesce. In sum, we discover *RAI1*-deleted yeast exhibit increased P body formation and translation defects.

### 4.5 Deciphering Rai1 and Kin28 Function in Translation via 5'P-Seq

To explore the translation defects in *rai1::URA3*, as well as determine if *RAI1* deletion affected stability of any mRNAs when Kin28 is inhibited, we utilized 5'P-Seq, which quantitatively captures the position of nucleotides bearing 5'-monophosphates (5'P) in RNA [20]. **Figure 4.9** shows a heatmap of 5'P in ascending order of mRNA expression levels in wild-type yeast. As observed before, 5'P levels are generally proportional to RNA levels; intuitively, 5'P ends are more abundant if the mRNA itself is more abundant. Interestingly, deletion of *RAI1* results in an increase in 5'P at the transcription start site (TSS) and the cleavage and polyadenylation site (CPS), which is unlikely to be a consequence of increased mRNA levels in *rai1::URA3*, as our RT-qPCR experiments show mRNA levels in *rai1::URA3* are buffered (**Figure 4.4**). Rather, the increase in 5'P of the *RAI1* mutant is consistent with Rai1's role in promoting Rat1-mediated decay. After both de-capping and cleavage of mRNA at the TSS and PAS, respectively, 5'P are generated, but are rapidly, processively cleared by Rat1. Inefficient Rat1 activity, as expected in *rai1::URA3*, would prolong the abundance of 5'P species at the TSS and CPS site of mRNA.

Next, we analyzed the 5'P-Seq data centered around the open reading frame (ORF) start (**Figure 4.10**) and stop (**Figure 4.11**) codon to elucidate the defects in translation of *rai1::URA3*. At -14 nt from the start codon, we find a dramatic increase in 5'P of *rai1::URA3*. This suggests that the halfmer peak in polysome profiles of *rai1::URA3* (**Figure 4.6**) are representative of the translation pre-initiation complex stalling at the start codon (48S). Because the increase in 5'P is distinctly positioned and not diffuse over the 5'UTR, we conclude that the defect in translation of cells lacking Rai1 is specifically a result of 48S pausing versus a defect in start codon scanning or regulation via eIF2. **Figure 4.12** diagrams the formation of the translation pre-initiation complex for reference.

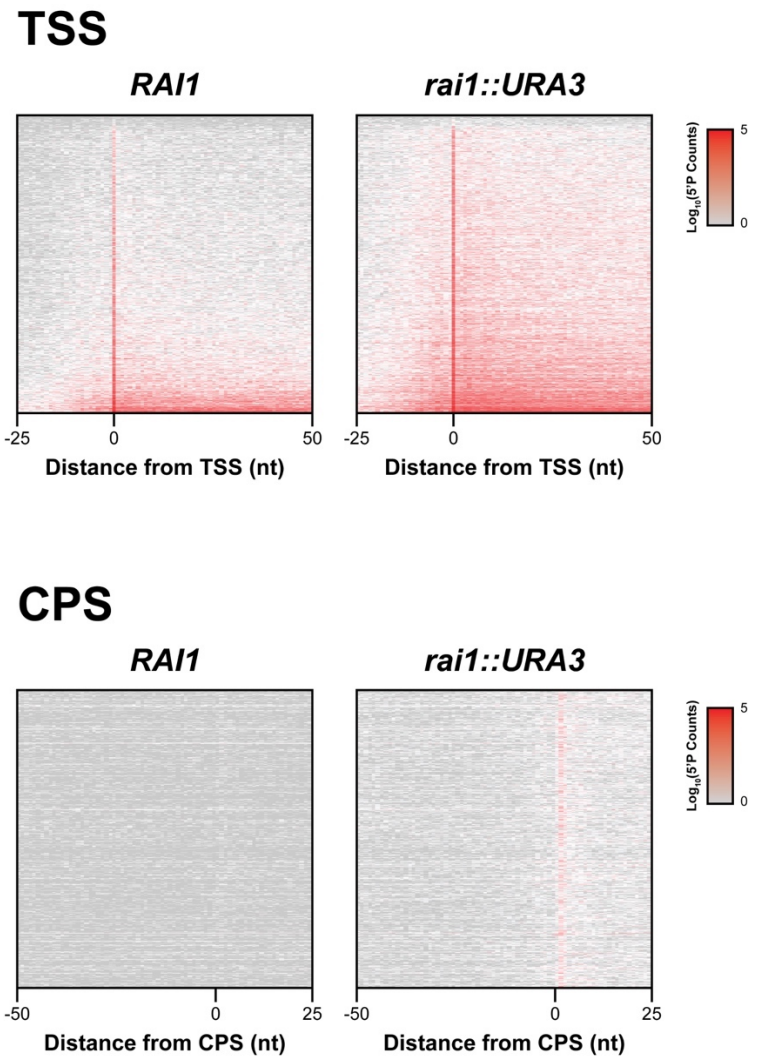



**Figure 4.9 Deletion of *RAI1* Increases Capture of 5'P at the TSS and CPS**

Heatmap of 5'P-Seq reads relative to the transcription start site (TSS) or the cleavage and polyadenylation site (CPS) is plotted for wild-type (*RAI1*) and *RAI1*-deleted (*rai1::URA3*) cells. mRNAs are ordered by steady-state expression level of wild-type cells, as determined by RNA-Seq. Only genes with over 100 RNA-Seq reads and over 32 5'P-Seq reads/base are included, n=3765.

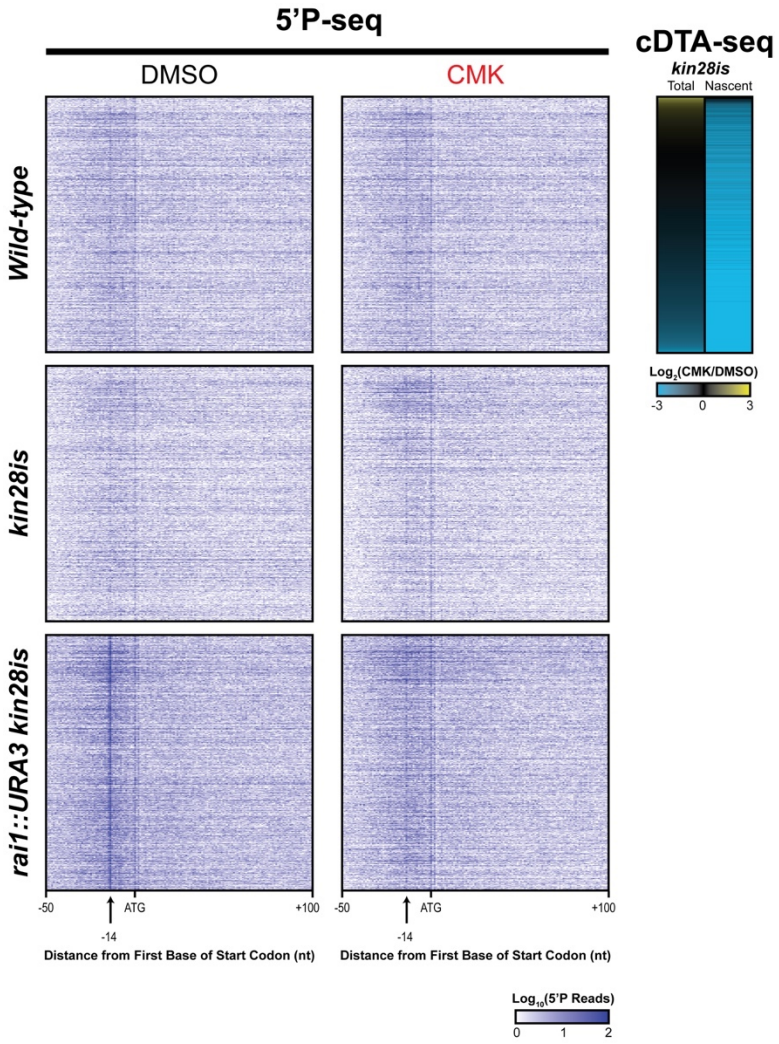



**Figure 4.10 Changes in 5'P at the Start Codon**

Heatmaps of 5'P-Seq reads relative to the first nucleotide of the start are plotted for the indicated strains. Plots are in descending order of change in expression level upon inhibition of *kin28is*; the cDTA-Seq data (described in Chapter 3) is adjacent. The peak at -14 nt corresponds to a ribosome parked at the start codon in the 'P' site. Only genes with over 100 RNA-Seq reads and over 32 5'P-Seq reads/base are included, n=3765.

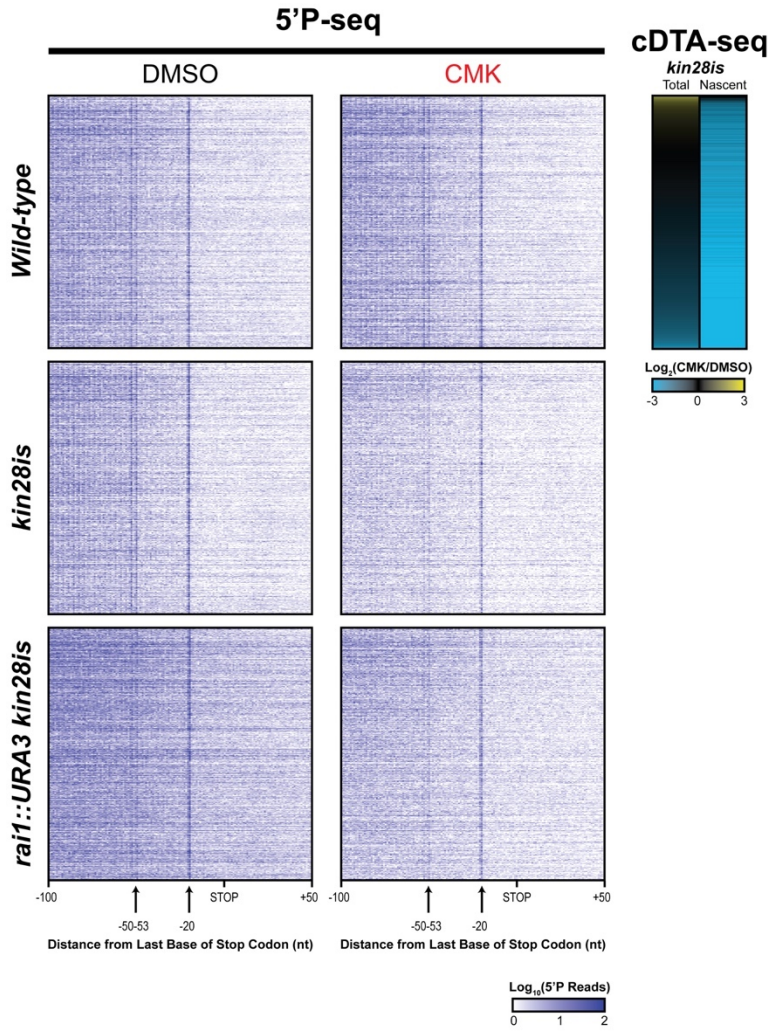



**Figure 4.11 Changes in 5'P at the Stop Codon**

Heatmaps of 5'P-Seq reads relative to the last nucleotide of the stop codon of the ORF are plotted. Plots are in descending order of change in expression level upon inhibition of *kin28is*; the cDTA-Seq data (described in Chapter 3) is adjacent. The peaks at -20 and -50-53 nt correspond to one or two ribosomes, respectively, parked at the stop codon in the 'A' site. Only genes with over 100 RNA-Seq reads and over 32 5'P-Seq reads/base are included, n=3765.

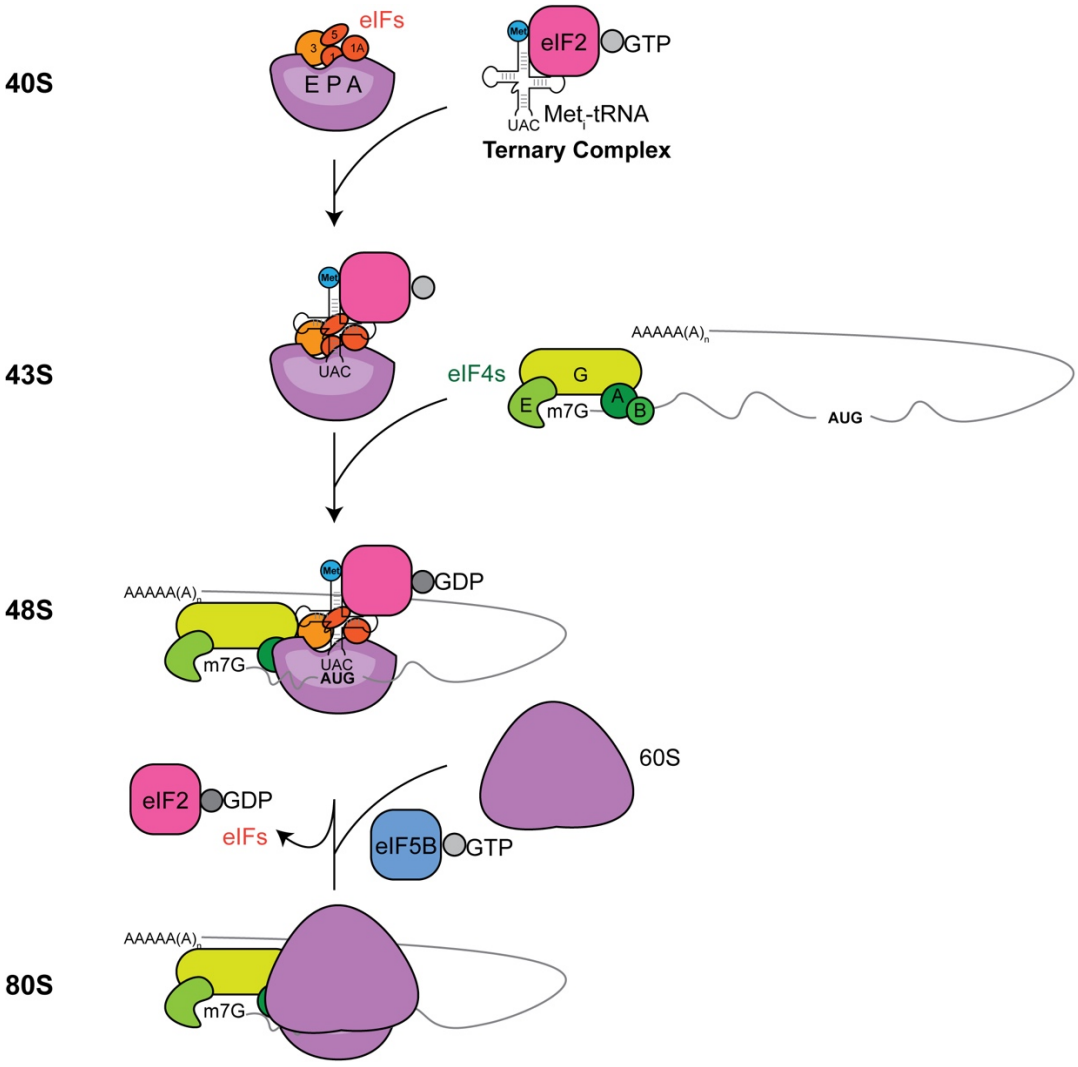



**Figure 4.12 Schematic of Translation Pre-initiation Events**

In contrast to changes in 5'P at the start codon, the peaks every 3-nt (*i.e.* 5'P generated from co-translational decay) around the stop codon of *rai1::URA3* are not noticeably changed compared to wild-type yeast (**Figure 4.11**). Rather, the 5'P signal around the ORF stop codon is more random in *rai1::URA3* cells. This observation is quantified by the decrease in 3-nt periodicity of the intensity of 5'P in *rai1::URA3* (**Figure 4.13**). The abundance of 5'P unrelated to translation may be indicative of Rai1's function as Rat1's cofactor, discussed above. Further, in conjunction with Kin28 inhibition a decrease in co-translational decay is also observed in the *rai1::URA3* background, indicating Rai1 is not responsible for this change—instead, most likely altered Xrn1 co-translational decay is the cause, as discussed previously.

Interestingly, the 48S 5'P peak observed in *rai1::URA3* cells disappears when Kin28 is inhibited. Fortuitously, the change in the 48S 5'P peak served as a proxy for translation defects in inhibited Kin28 cells; mRNAs with reduced 48S in inhibited *kin28is* are likely not translated. Furthermore, the disappearance of the 48S 5'P signal upon inhibition of Kin28 suggests translation defects in inhibited *kin28is* occur early, before ribosome scanning to identify the start codon. We plotted the change in 5'P signal of mRNAs in *rai1::URA3 kin28is* upon inhibition of Kin28 in the same order as Figure 3.20, which clustered the mRNAs by what we believed at the time to be their translatability in inhibited *kin28is* (**Figure 4.14**). Notably, the mRNAs in Cluster 1 exhibit less change in 48S pausing.

We then wondered if 48S pausing was unchanged for Cluster 1 mRNAs in inhibited *kin28is*, or if 48S pausing did not occur at all for these mRNAs. To test for these possibilities, we broke-down the change in 5'P signal into its component parts (*i.e.* we plotted heatmaps of 5'P in DMSO and CMK treated cells separately) (**Figure 4.15** and **4.16**). In this representation of 5'P, we observe co-translational decay at the stop codon is more pronounced under normal conditions for mRNA in Clusters 1, 3, and 5. Thus,

changes to co-translational decay would be more prominent for mRNAs in these clusters, as is the case in **Figure 4.14**. Potentially, inhibition of Kin28 affects translation of all genes, but is only observable as a change in 5'P for mRNAs subject to co-translational decay (mRNA in Clusters 1, 3, and 5). In answer to our original question about 48S pausing in Cluster 1 mRNAs, we find the 48S pause is also prominent for these mRNAs in *rai1::URA3 kin28is* cells treated with either DMSO or CMK (**Figure 4.16**), suggesting they are still translated when Kin28 is inhibited. Changes in 5'P from 48S pausing in *rai1::URA3 kin28is* cells may be a better indicator of translation state than 5'P at the stop codon.

Therefore, we re-clustered mRNAs by their changes in 48S pausing in *rai1::URA3 kin28is* to clarify which mRNAs are still translated, *i.e.* exhibit 48S pausing, in inhibited Kin28 cells (**Figure 4.17**). We identify three such groups of mRNAs, Clusters i, iii, and vi; many of the mRNAs in Cluster iii overlap with those previously identified in **Figure 4.14**. Gene ontology (GO) analysis reveals these mRNAs are associated with membrane, cytoplasm, mitochondrion, redox activity, the stress response, amino acid biosynthesis, the ribosome, RNA binding, and RNA export. In contrast, the mRNAs with reduced 48S pausing (Clusters ii, iv, v; untranslated) are enriched for GO terms related to rRNA processing, ribosome biogenesis, the proteasome, nucleosome, endoplasmic reticulum, pre-ribosome, nucleus, and nucleolus. The full list of GO terms associated with each cluster is presented in **Table 4.1**.

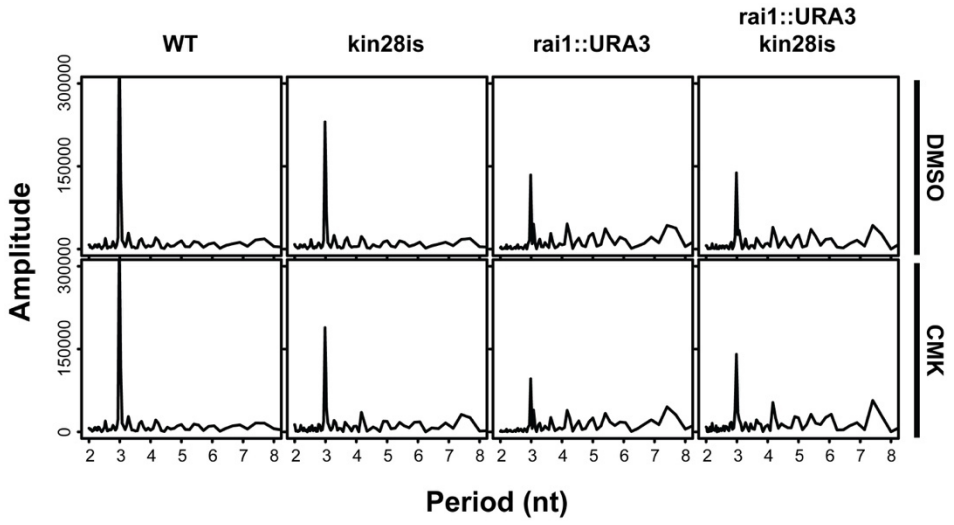



**Figure 4.13 Analysis of Periodicity in 5'P Signal**

5'P signal from the last 200 nt of the ORF was analyzed for periodicity in intensity. Those genes for which there was no signal at any of the 200 nt were excluded from analysis.

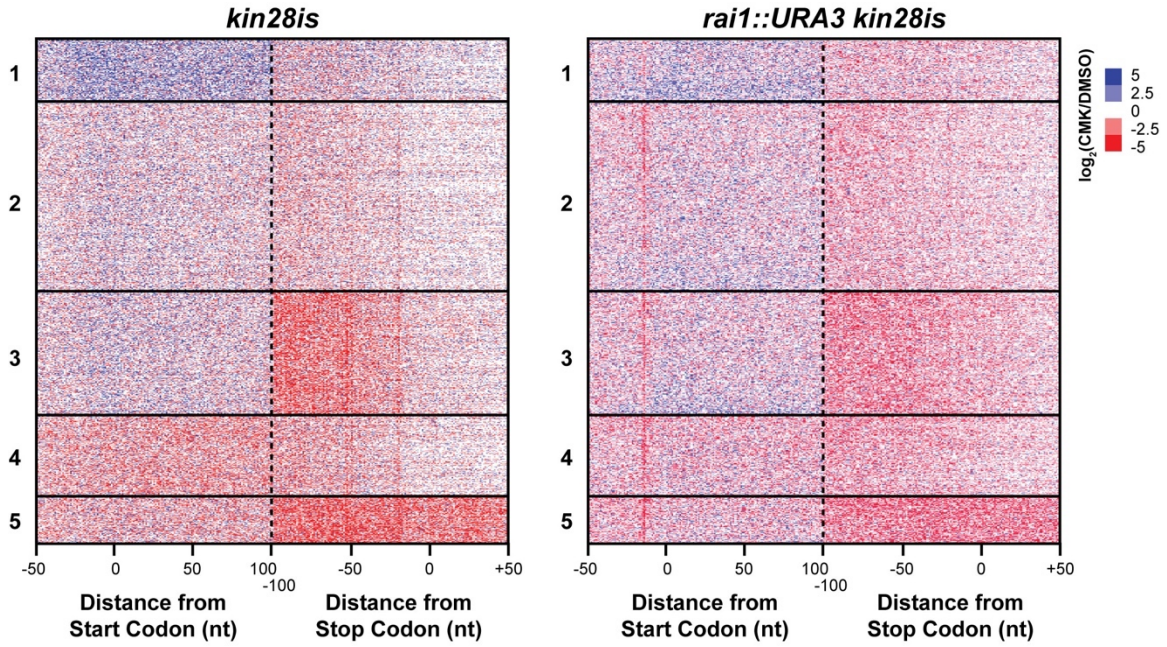



**Figure 4.14 Effect of *RAI1*-deletion on 5'P-Seq Changes in *kin28is***

See Figure 3.20; k-means clustering of the change in 5'P at the start and stop codon of *kin28is*. The 5'P data from *rai1::URA3 kin28s* is plotted adjacent in the same order.

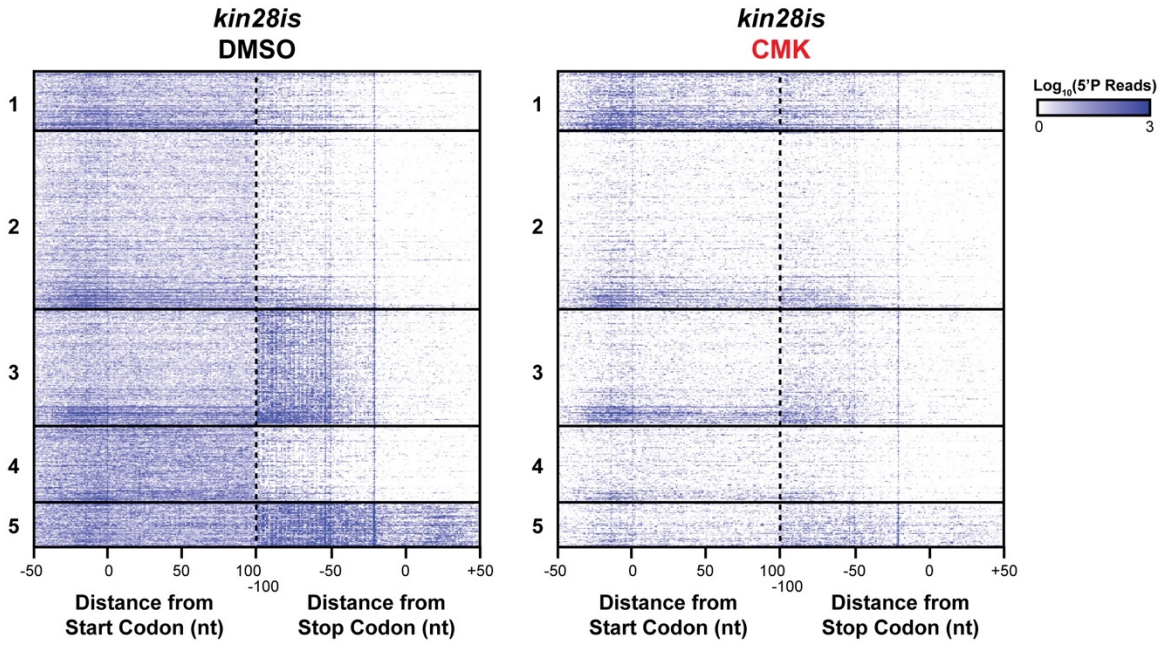



**Figure 4.15 5'P-Seq Changes in *kin28is***

See Figure 4.14; 5'P-Seq changes were de-convoluted into signal from inhibited (CMK) and control (DMSO) cells.

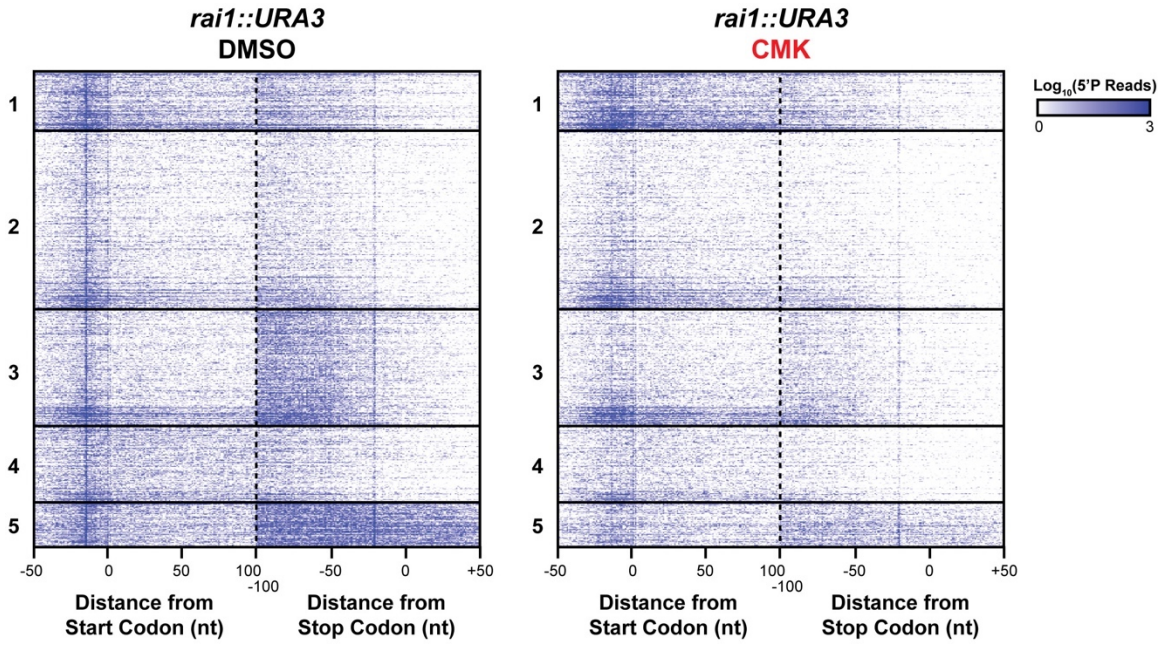



**Figure 4.16 5'P-Seq Changes in *rai1::URA3 kin28is***

See Figure 4.14; 5'P-Seq changes were de-convoluted into signal from inhibited (CMK) and control (DMSO) cells.

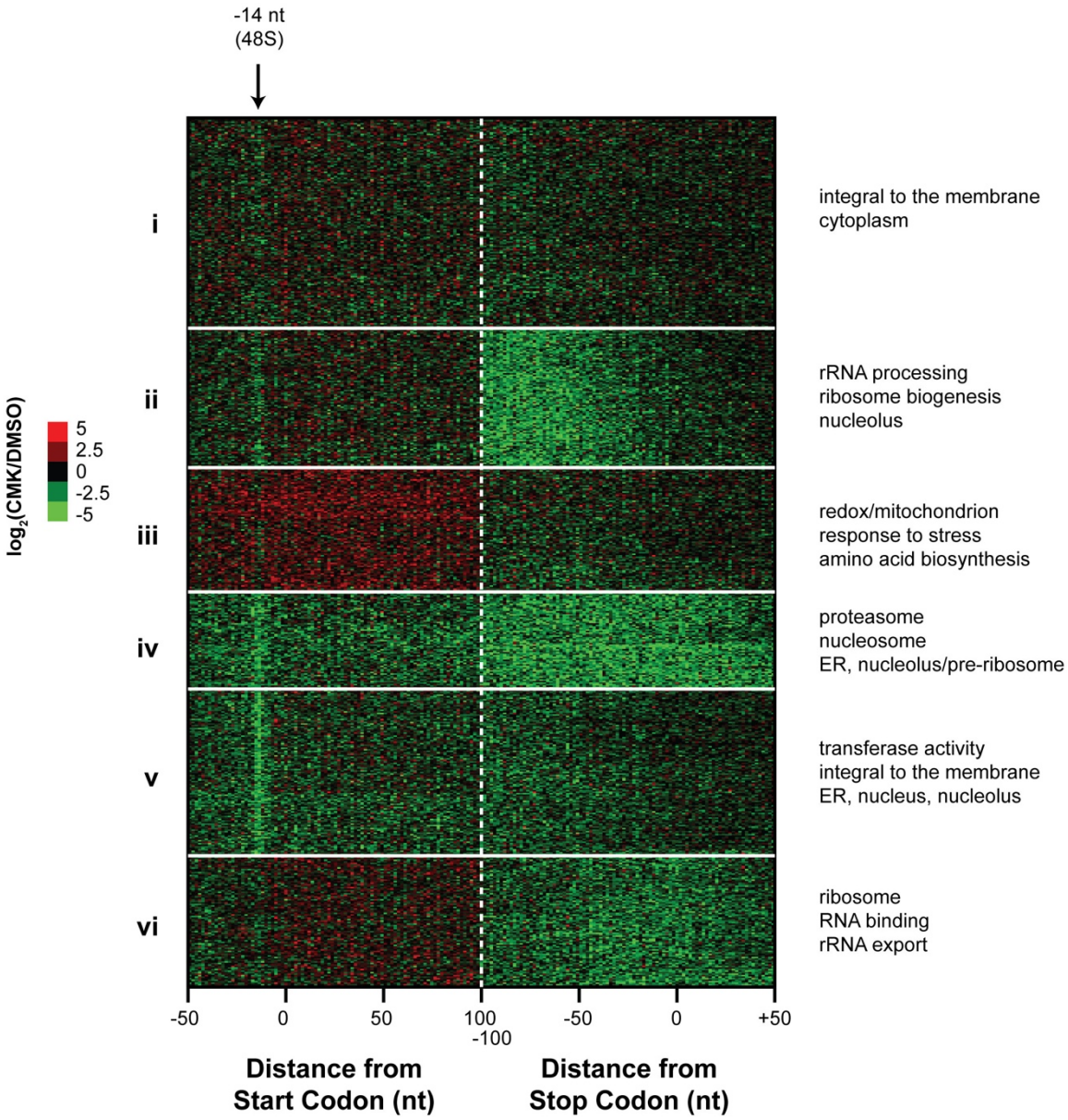



**Figure 4.17 Re-clustering of 5'P Using Data from *rai1::URA3 kin28is***

k-means clustering of the change in 5 at the start and stop codon of *rai1::URA3 kin28s*. The position of the -14 nt peak, which corresponds to a 48S ribosomal subunit parked at the start codon, is indicated. Relevant Gene Ontology terms from each cluster are listed adjacent. See Table 4.1 for a more comprehensive list.

**Table 4.1 Gene Ontology Analysis of Clusters Identified in Figure 4.17**

Significant ($p < 0.01$, Bonferroni-adjusted) Gene Ontology (GO) terms shared among genes in each cluster grouped by 5'P-Seq profile. The number of genes (n) in each cluster and the number of genes in the cluster with the GO term (k) are reported along with the number of genes in each GO term category (f).

| Cluster | n | Category | p-value | k | f |
|---|---|---|---|---|---|
| i | 908 | membrane [GO:0016020] | 3.98E-10 | 307 | 1671 |
| | | cytoplasm [GO:0005737] | 2.48E-07 | 345 | 2026 |
| | | integral to membrane [GO:0016021] | 1.14E-05 | 228 | 1303 |
| ii | 600 | ribosome biogenesis [GO:0042254] | 6.72E-11 | 44 | 170 |
| | | nucleolus [GO:0005730] | 1.29E-10 | 56 | 253 |
| | | cytosolic large ribosomal subunit [GO:0022625] | 2.83E-10 | 29 | 88 |
| | | rRNA processing [GO:0006364] | 2.32E-09 | 45 | 195 |
| | | endoplasmic reticulum [GO:0005783] | 2.40E-09 | 75 | 416 |
| | | ribonucleoprotein complex [GO:0030529] | 4.99E-09 | 60 | 307 |
| | | translation [GO:0006412] | 7.79E-09 | 61 | 318 |
| | | structural constituent of ribosome [GO:0003735] | 3.08E-08 | 46 | 218 |
| | | ribosome [GO:0005840] | 6.67E-07 | 55 | 310 |
| | | cytoplasm [GO:0005737] | 1.42E-06 | 236 | 2026 |
| | | small-subunit processome [GO:0032040] | 1.83E-06 | 16 | 47 |
| | | ribosomal large subunit assembly [GO:0000027] | 2.77E-06 | 14 | 38 |
| | | endoplasmic reticulum membrane [GO:0005789] | 3.44E-06 | 54 | 318 |
| | | nucleus [GO:0005634] | 3.46E-06 | 228 | 1965 |
| | | endonucleolytic cleavage to generate mature 5'-end of SSU-rRNA from (SSU-rRNA, 5.8S rRNA, LSU-rRNA) [GO:0000472] | 3.47E-06 | 12 | 29 |
| | | transferase activity [GO:0016740] | 8.03E-06 | 87 | 611 |
| | | endonucleolytic cleavage in 5'-ETS of tricistronic rRNA transcript (SSU-rRNA, 5.8S rRNA, LSU-rRNA) [GO:0000480] | 1.07E-05 | 11 | 27 |
| | | intracellular [GO:0005622] | 1.17E-05 | 60 | 381 |
| | | alpha-1,6-mannosyltransferase activity [GO:0000009] | 1.31E-05 | 6 | 8 |
| | | 90S preribosome [GO:0030686] | 2.20E-05 | 19 | 74 |
| | | nucleotide binding [GO:0000166] | 2.79E-05 | 103 | 778 |
| | | alpha-1,6-mannosyltransferase complex [GO:0000136] | 3.39E-05 | 5 | 6 |
| iii | 538 | catalytic activity [GO:0003824] | 8.88E-14 | 85 | 455 |
| | | cytoplasm [GO:0005737] | 2.95E-13 | 265 | 2026 |
| | | oxidation-reduction process [GO:0055114] | 3.40E-11 | 56 | 272 |
| | | oxidoreductase activity [GO:0016491] | 1.07E-10 | 55 | 272 |
| | | binding [GO:0005488] | 4.85E-09 | 55 | 300 |
| | | metabolic process [GO:0008152] | 1.01E-08 | 69 | 425 |

| | | | | | |
|---|---|---|---|---|---|
| | | oxidoreductase activity, acting on the aldehyde or oxo group of donors, NAD or NADP as acceptor [GO:0016620] | 2.30E-08 | 7 | 7 |
| | | cellular amino acid biosynthetic process [GO:0008652] | 3.85E-08 | 26 | 98 |
| | | ATP binding [GO:0005524] | 1.18E-06 | 84 | 622 |
| | | ligase activity [GO:0016874] | 4.20E-06 | 27 | 130 |
| | | nucleotide binding [GO:0000166] | 6.11E-06 | 97 | 778 |
| | | mating projection tip [GO:0043332] | 8.60E-06 | 23 | 105 |
| | | response to stress [GO:0006950] | 1.09E-05 | 29 | 152 |
| | | 2 iron, 2 sulfur cluster binding [GO:0051537] | 2.55E-05 | 7 | 13 |
| | | actin cortical patch [GO:0030479] | 3.36E-05 | 15 | 57 |
| | | mitochondrial nucleoid [GO:0042645] | 4.28E-05 | 9 | 23 |
| | | mitochondrion [GO:0005739] | 4.92E-05 | 121 | 1072 |
| | | cytosol [GO:0005829] | 6.62E-05 | 32 | 192 |
| iv | 415 | threonine-type endopeptidase activity [GO:0004298] | 4.89E-07 | 8 | 14 |
| | | proteasomal ubiquitin-independent protein catabolic process [GO:0010499] | 4.89E-07 | 8 | 14 |
| | | nucleosome [GO:0000786] | 9.76E-07 | 7 | 11 |
| | | proteasome core complex [GO:0005839] | 9.91E-07 | 8 | 15 |
| | | proteasome storage granule [GO:0034515] | 1.84E-06 | 10 | 26 |
| | | proteasomal ubiquitin-dependent protein catabolic process [GO:0043161] | 2.05E-06 | 11 | 32 |
| | | proteasome complex [GO:0000502] | 2.99E-06 | 13 | 46 |
| | | nucleoside metabolic process [GO:0009116] | 4.54E-06 | 7 | 13 |
| | | proteolysis involved in cellular protein catabolic process [GO:0051603] | 5.70E-06 | 8 | 18 |
| | | endopeptidase activity [GO:0004175] | 1.47E-05 | 8 | 20 |
| | | endoplasmic reticulum membrane [GO:0005789] | 1.54E-05 | 40 | 318 |
| | | glucosidase activity [GO:0015926] | 1.81E-05 | 5 | 7 |
| | | proteasome core complex, alpha-subunit complex [GO:0019773] | 1.81E-05 | 5 | 7 |
| | | nuclear nucleosome [GO:0000788] | 2.10E-05 | 6 | 11 |
| | | nucleolus [GO:0005730] | 4.26E-05 | 33 | 253 |
| | | nucleus [GO:0005634] | 4.61E-05 | 160 | 1965 |
| | | endoplasmic reticulum [GO:0005783] | 4.69E-05 | 47 | 416 |
| | | preribosome, large subunit precursor [GO:0030687] | 6.13E-05 | 11 | 44 |
| v | 724 | nucleus [GO:0005634] | 4.56E-11 | 293 | 1965 |
| | | endoplasmic reticulum [GO:0005783] | 1.88E-08 | 83 | 416 |
| | | nucleolus [GO:0005730] | 1.32E-07 | 56 | 253 |
| | | endoplasmic reticulum membrane [GO:0005789] | 1.43E-06 | 63 | 318 |
| | | integral to membrane [GO:0016021] | 3.76E-06 | 190 | 1303 |
| | | transferase activity [GO:0016740] | 7.77E-06 | 101 | 611 |
| | | membrane [GO:0016020] | 1.36E-05 | 231 | 1671 |

| | | | | | |
|---|---|---|---|---|---|
| vi | 569 | structural constituent of ribosome [GO:0003735] | <1e-14 | 72 | 218 |
| | | translation [GO:0006412] | <1e-14 | 92 | 318 |
| | | cytosolic small ribosomal subunit [GO:0022627] | <1e-14 | 28 | 62 |
| | | cytosolic large ribosomal subunit [GO:0022625] | <1e-14 | 34 | 88 |
| | | ribosome [GO:0005840] | <1e-14 | 86 | 310 |
| | | ribonucleoprotein complex [GO:0030529] | <1e-14 | 76 | 307 |
| | | cytoplasm [GO:0005737] | <1e-14 | 272 | 2026 |
| | | intracellular [GO:0005622] | 2.43E-13 | 77 | 381 |
| | | rRNA export from nucleus [GO:0006407] | 8.17E-08 | 13 | 27 |
| | | RNA binding [GO:0003723] | 1.89E-05 | 52 | 337 |
| | | cellular bud tip [GO:0005934] | 4.71E-05 | 18 | 75 |
| | | small ribosomal subunit [GO:0015935] | 6.82E-05 | 7 | 14 |

## 4.6 Discussion

The connections between RNA decay, translation, and P body formation are manifold but a unifying picture of the three processes is only just becoming apparent. Our study suggests that the signal that initiates P body formation may be reduced mRNA decay. Both *RAI1*-deletion and Kin28 inhibition—both of which result in reduced mRNA decay—result in the increased formation of P bodies. Potentially, RNA binding proteins are titrated to sites of accumulating mRNA, causing aggregation into compartments such as P bodies. However, a subset of mRNAs continues to be translated, and is presumably *not* associated with P bodies.

In this study, we were able to refine our understanding of translation in inhibited *kin28is* through polysome profiling and analysis of 5'P-Seq of *rai1::URA3*. In our 5'P-Seq analysis, we find mRNAs involved with nuclear processes and proteolysis are poorly translated in inhibited *kin28is* cells. We speculate cells have evolved to prioritize translation of mRNAs required for basic cellular function and survival during transcriptional stress. The effectors of this differential translation in inhibited *kin28is* remain elusive. Potentially, ribosome biogenesis, which involves the action of decay factors and Pol I and III transcripts, is a nexus for cellular sensing of RNA metabolism and translation. Whether P bodies are sites of decay or storage in inhibited *kin28is* is also an open question. The yeast strains and methods described here can be utilized in future studies to answer these questions.

## 4.7 Materials and Methods

*Yeast Strains and Growth Conditions*

**Table 4.2** lists all strains used in this study. Exonuclease-deletion mutant strains were created by replacing the coding regions of the *XRN1* and *RAI1* gene loci with a *URA3* gene cassette amplified from *pRS306*. Clones were selected on uracil-dropout

**Table 4.2 Yeast Strains Used in This Chapter**

| Relevant Genotype | Reference |
|---|---|
| *S. cerevisiae*: wild-type: *MAT***a** *his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 RPB3-TAP::HIS3* | [21] |
| *S. cerevisiae*: kin28is: *MAT***a** *his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 RPB3-TAP::HIS3 kin28 L83G, V21C* | [22] |
| *S. cerevisiae*: xrn1::URA3: *MAT***a** *his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 RPB3-TAP::HIS3 xrn1::URA3* | This Study |
| *S. cerevisiae*: xrn1::URA3 kin28is: *MAT***a** *his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 RPB3-TAP::HIS3 xrn1::URA3 kin28 L83G, V21C* | This Study |
| *S. cerevisiae*: rai1::URA3: *MAT***a** *his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 RPB3-TAP::HIS3 rai1::URA3* | This Study |
| *S. cerevisiae*: rai1::URA3 kin28is: *MAT***a** *his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 RPB3-TAP::HIS3 rai1::URA3 kin28 L83G, V21C* | This Study |
| *S. cerevisiae*: Edc3-GFP: *MAT***a** *his4-539 trp1 leu2-3,112 ura3-52 EDC3-GFP-NEO cup1::LEU2/PGK1pG/MFA2pG* | Dr. Roy Parker |
| *S. cerevisiae*: Edc3-GFP rai1::URA3: *MAT***a** *his4-539 trp1 leu2-3,112 ura3-52 EDC3-GFP-NEO cup1::LEU2/PGK1pG/MFA2pG rai1::URA3* | This Study |
| *S. cerevisiae*: Edc3-GFP kin28is: *MAT***a** *his4-539 trp1 leu2-3,112 ura3-52 EDC3-GFP-NEO cup1::LEU2/PGK1pG/MFA2pG kin28 L83G, V21C* | This Study |
| *S. cerevisiae*: Edc3-GFP rai1::URA3 kin28is: *MAT***a** *his4-539 trp1 leu2-3,112 ura3-52 EDC3-GFP-NEO cup1::LEU2/PGK1pG/MFA2pG rai1::URA3 kin28 L83G, V21C* | This Study |
| *S. cerevisiae*: Xrn1-GFP: *MAT***a** *his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 XRN1::GFP::HIS3MX6* | [23] |
| *S. cerevisiae*: Xrn1-GFP kin28is: *MAT***a** *his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 XRN1::GFP::HIS3MX6* | This Study |

plates, and deletion was confirmed via Sanger sequencing. Unless otherwise specified, cells were treated as described in Chapters 2 and 3.

*Fluorescence Microscopy*

Images were collected on a Nikon Eclipse Ti system. Cells grown in YPD without riboflavin and folic acid were plated on Nunc 96-well coverslip bottom plates that were pre-treated with concanavalin A (2 mg/mL concanavalin A, 5 mM $CaCl_2$, 5 mM $MnCl_2$). Images were processed using ImageJ.

*Cytoplasmic/Nuclear Lysate Preparation and Immunoblotting*

Adapted from [24]. All buffers used contained 1X protease inhibitors (1X EDTA-free cOmplete Protease Inhibitor, 1 mM PMSF, 1 µg/mL pepstatin A, 1 µg/mL aprotinin). Briefly, 50 mL of cells were pelleted and resuspended in SB (1 M Sorbitol, 20 mM TrisHCl, pH 7.4) and treated with 125 µL of Zymolyase 20T (10 mg/mL) and nutated for at least 30 minutes until cell walls were degraded, confirmed by > 80% decrease in OD600. Spheroplasts were rinsed twice in 1 mL of SB and gently resuspended with 500 µL of EBX (20 mM TrisHCl, pH 7.4, 100 mM NaCl, 0.25% Triton X-100, 15 mM beta-mercaptoethanol), with additional Triton X-100 added to 0.5% final concentration. Samples were gently mixed and incubated on ice for 10 minutes. An aliquot of lysate was removed (“Total” fraction), and the rest was layered on 1 mL of NIB (20 mM TrisHCl, pH 7.4, 100 mM NaCl, 1.2 mM sucrose, 15 mM beta-mercaptoethanol). The samples were centrifuged at 10,000*g* for 15 min at 4 °C. Avoiding the lipid interphase, the top layer was removed and saved as “Cytoplasmic” fraction. The nuclei was visible as a glassy white pellet, which was then resuspended in 500 µL of EBX. Accounting for the pellet volume, Triton X-100 was added to 1% final concentration. Samples were

lysed on ice for 10 min with occasional gentle mixing. The resultant lysate was saved as the “Nuclear” fraction.

SDS Loading Buffer was added to each preparation and were heated for 5 minutes at 95 °C prior to loading onto tris-glycine gels. SDS-PAGE and transfer to nitrocellulose membranes were conducted with the Bio-Rad Mini-PROTEAN system following standard protocols. Blots were blocked for 1 hour in 5% milk in TBST and incubated with primary antibodies overnight at 4 °C at the following dilutions: anti-Xrn1 (1:5000), anti-H3 (1:5000), and anti-Rpb4 (1:1000). The yeast Xrn1 antibody was a generous gift of Dr. Arlen Johnson. Secondary antibodies were used at a 1:5000 dilution. Blots were incubated with Clarity Western ECL substrate and developed on the ImageQuant LAS 4000.

*RNA Purification and RT-qPCR*

As described in Chapter 3; RNA was extracted from cells by incubating with occasional vortexing in TE Lysis Buffer (10 mM TrisHCl pH 7.5, 10 mM EDTA, 0.5% SDS) and acidic phenol at 65 °C, followed by ethanol precipitation. All RNA preparations were treated with DNase I prior to downstream assays.

*Polysome Profiling*

Sucrose was dissolved in Polysome Profile Buffer (20 mM TrisHCl pH 7.5, 10 mM $MgCl_2$, 100 mM KCl) to 7, 17, 27, 37, and 47% weight by volume. Sucrose gradients were formed by layering sucrose solutions into an ultracentrifuge tube, freezing in liquid nitrogen between layers. Sucrose gradients were thawed overnight at 4 °C before use or stored at -20 °C. Cycloheximide (100 µg/mL final concentration) was added to cells immediately before harvest. Cells were lysed in Lysis Buffer (2 mM DTT, 1% Triton X-100, 1 µL/mL RNase Inhibitor, 100 µg/mL CHX, 1X EDTA-free cOmplete Protease

Inhibitor, in Polysome Profile Buffer) by bead-beating at 4 °C, five cycles of 2 minutes with 2 minutes rest. Lysates (15 A260 units) were layered onto the top of sucrose gradients and separated by spinning at 45,000 rpm (SW50.1 Ti rotor) for 1 hour and 20 minutes at 4 °C with 3 minutes acceleration and deceleration. Profiles and fractions were collected on an ISCO Foxy Jr instrument with PeakTrak software.

*5'P-Seq Data Analysis*

Reads were demultiplexed and quick statistics generated with FastQC. UMI-tools was used to move the unique molecular identifier (UMI) into the read header before alignment to sacCer3 with STAR; clipped 5' reads were filtered with –alignEndsType Extend5pOfRead1. Aligned reads with a quality score over 30 were sorted and indexed with SAMtools. UMI-tools was used to de-duplicate reads. The 5' nucleotide was plotted with --Offset 1 and scaled with --normalizeUsingRPKM during conversion to the bigwig format with deepTools 2. Alignments were visualized in IGV. Signal was extracted across ORFs with deepTools 2 with single nucleotide bin sizes. Analysis was continued only on transcripts which had over 100 RNA-Seq reads and over 32 5'P-Seq reads per base in at least one condition to reduce the effect of noise in interpretation of data. Heatmaps of 5'P signal were generated with Java TreeView 3.0 and clustered with Cluster 3.0. The R package '*GeneCycle*' was used to examine periodicity of 5'P. Gene ontology analysis was performed with FunSpec.

## 4.8 References


1. Baudrimont, A., et al., *Multiplexed gene control reveals rapid mRNA turnover.* Science Advances, 2017. **3**(7).

2. Sun, M., et al., *Global Analysis of Eukaryotic mRNA Degradation Reveals Xrn1-Dependent Buffering of Transcript Levels.* Molecular Cell, 2013. **52**(1): p. 52-62.

3. Houseley, J. and D. Tollervey, *The Many Pathways of RNA Degradation.* Cell, 2009. **136**(4): p. 763-776.

4. Parker, R. and U. Sheth, *P bodies and the control of mRNA translation and degradation.* Molecular Cell, 2007. **25**(5): p. 635-646.

5. Haimovich, G., et al., *Gene Expression Is Circular: Factors for mRNA Degradation Also Foster mRNA Synthesis.* Cell, 2013. **153**(5): p. 1000-1011.

6. Jiao, X.F., et al., *A Mammalian Pre-mRNA 5 ' End Capping Quality Control Mechanism and an Unexpected Link of Capping to Pre-mRNA Processing.* Molecular Cell, 2013. **50**(1): p. 104-115.

7. Jiao, X.F., et al., *Identification of a quality-control mechanism for mRNA 5 '-end capping.* Nature, 2010. **467**(7315): p. 608-U137.

8. Xiang, S., et al., *Structure and function of the 5 '-> 3 ' exoribonuclease Rat1 and its activating partner Rai1.* Nature, 2009. **458**(7239): p. 784-U130.

9. Xue, Y., et al., *Saccharomyces cerevisiae RAI1 (YGL246c) is homologous to human DOM3Z and encodes a protein that binds the nuclear exoribonuclease Rat1p.* Molecular and Cellular Biology, 2000. **20**(11): p. 4006-4015.

10. Stevens, A. and T.L. Poole, *5'-Exonuclease-2 of Saccharomyces-Cerevisiae - Purification and Features of Ribonuclease Activity with Comparison to 5'-Exonuclease-1.* Journal of Biological Chemistry, 1995. **270**(27): p. 16063-16069.

11. Breker, M., et al., *LoQAtE--Localization and Quantitation ATlas of the yeast proteomE. A new tool for multiparametric dissection of single-protein behavior in response to biological perturbations in yeast.* Nucleic Acids Res, 2014. **42**(Database issue): p. D726-30.

12. Koh, J.L., et al., *CYCLoPs: A Comprehensive Database Constructed from Automated Analysis of Protein Abundance and Subcellular Localization Patterns in Saccharomyces cerevisiae.* G3 (Bethesda), 2015. **5**(6): p. 1223-32.

13. Decker, C.J. and R. Parker, *P-Bodies and Stress Granules: Possible Roles in the Control of Translation and mRNA Degradation.* Cold Spring Harbor Perspectives in Biology, 2012. **4**(9).

14. Goler-Baron, V., et al., *Transcription in the nucleus and mRNA decay in the cytoplasm are coupled processes.* Genes & Development, 2008. **22**(15): p. 2022-2027.

15. Selitrennik, M., et al., *Nucleocytoplasmic shuttling of the Rpb4p and Rpb7p Subunits of Saccharomyces cerevisiae RNA polymerase II by two pathways.* Eukaryotic Cell, 2006. **5**(12): p. 2092-2103.

16. Zaragoza, D., et al., *Rapamycin induces the G(0) program of transcriptional repression in yeast by interfering with the TOR signaling pathway.* Molecular and Cellular Biology, 1998. **18**(8): p. 4463-4470.

17. Cherkasova, V.A. and A.G. Hinnebusch, *Translational control by TOR and TAP42 through dephosphorylation of eIF2 alpha kinase GCN2.* Genes & Development, 2003. **17**(7): p. 859-872.

18. Pestov, D.G. and N. Shcherbik, *Rapid Cytoplasmic Turnover of Yeast Ribosomes in Response to Rapamycin Inhibition of TOR.* Molecular and Cellular Biology, 2012. **32**(11): p. 2135-2144.

19. Ramachandran, V. and P.K. Herman, *Antagonistic Interactions Between the cAMP-Dependent Protein Kinase and Tor Signaling Pathways Modulate Cell Growth in Saccharomyces cerevisiae.* Genetics, 2011. **187**(2): p. 441-U131.

20. Pelechano, V., W. Wei, and Lars M. Steinmetz, *Widespread Co-translational RNA Decay Reveals Ribosome Dynamics.* Cell, 2015. **161**(6): p. 1400-1412.

21. Ghaemmaghami, S., et al., *Global analysis of protein expression in yeast.* Nature, 2003. **425**(6959): p. 737-41.

22. Rodríguez-Molina, Juan B., et al., *Engineered Covalent Inactivation of TFIIH-Kinase Reveals an Elongation Checkpoint and Results in Widespread mRNA Stabilization.* Molecular Cell, 2016. **63**(3): p. 433-444.

23. Huh, W.K., et al., *Global analysis of protein localization in budding yeast.* Nature, 2003. **425**(6959): p. 686-691.

24. Keogh, M.C., et al., *The Saccharomyces cerevisiae histone H2A variant Htz1 is acetylated by NuA4.* Genes Dev, 2006. **20**(6): p. 660-5.

# Chapter 5: Perspective on Kin28/CDK7 Inhibitor-based Therapies

## 5.1 Introduction

Protein kinases are involved in modulating virtually every known signaling and regulatory pathway, and in many cases their function is co-opted to promote specific disease states. Despite their obvious relevance as druggable targets, there are only 28 FDA-approved kinase inhibitors as of 2015 [1]. Kinases have retained nearly identical active site structure throughout evolutionary time, which contrasts with their varied functional roles in cells [2]. Because of high structural conservation, it has been exceedingly difficult to develop inhibitors that selectively target a single kinase's catalytic ATP-binding domain. Over the past two decades, however, several druggable residues, unique to a handful of human kinases have been exploited with potent inhibitors that show 'lock-and-key' complementarity with the target kinase [3]. These druggable residues rarely occur naturally in model organisms, such as budding yeast, which allows for the design and construction of orthogonal systems of inhibition that may prove useful for elucidating kinase function.

Of the more than 500 human kinases, cyclin-dependent kinases (CDKs) have received renewed attention as a chemotherapeutic target because of the success of an anti-cancer therapeutic, THZ1, which selectively kills tumor cells through a novel covalent binding mechanism outside of the CDK7 ATP-binding pocket [4]. CDK7 along with CDK8, CDK9, and CDK12/13 have key roles in regulating RNA transcription [5] (**Figure 5.1A**). Phosphorylation of the carboxyl-terminal domain (CTD) of the largest subunit of RNA polymerase II (Pol II) by CDK7, CDK9, and CDK12/13 regulates the spatiotemporal recruitment of transcriptional machinery, RNA processing factors, and chromatin modifying factors. Because all of these CDKs are required for normal cellular function and show high levels of sequence and structural similarities, it is not readily apparent how THZ1 can selectively inhibit cancerous cell growth without harming healthy cells.

In this Chapter, I discuss how targeted covalent inhibition of Kin28/CDK7 in yeast can help explain the anti-cancer properties of the recently characterized CDK7 inhibitor THZ1 [4].

### 5.2 Pause: Licensing for Elongation

To address the challenges of non-specific kinase inhibition, we recently applied a structure-guided approach to target a specific kinase for covalent inhibition in vivo by exploiting the high conservation of kinase active sites [6, 7]. Specifically, the ATP-binding pocket of Kin28/CDK7 was rationally repurposed by introducing two obligate mutations which do not occur in the yeast kinome in order to accept a covalent inhibitor. Our rationally engineered inhibition strategy provided confidence of the specificity of the inhibitor for its target, thus allowing us to directly link its effect to the function of Kin28/CDK7 (**Figure 5.1B**).

Covalent inhibition of Kin28 caused a Pol II buildup at the +2 nucleosome and a progressive depletion of Pol II occupancy downstream (**Figure 5.1C**). This Pol II redistribution is reminiscent of the promoter-proximal Pol II pausing seen in metazoans in which NELF restricts Pol II transition to elongation until phosphorylated by P-TEFb (Bur1/CDK9). Importantly, Ser2 phosphorylation of RNA Pol II CTD, which is required for elongation and termination of transcription, persists even after Kin28 is covalently inhibited, albeit at generally lower levels (**Figure 5.2B**). Our findings underscore the role of Kin28/CDK7 in directly promoting the initiation–elongation transition by overcoming a potential nucleosome roadblock and in the absence of NELF. These observations may also help clarify why the canonical Pol II pause is normally absent in yeast and only emerges upon complete inhibition of Kin28/CDK7.

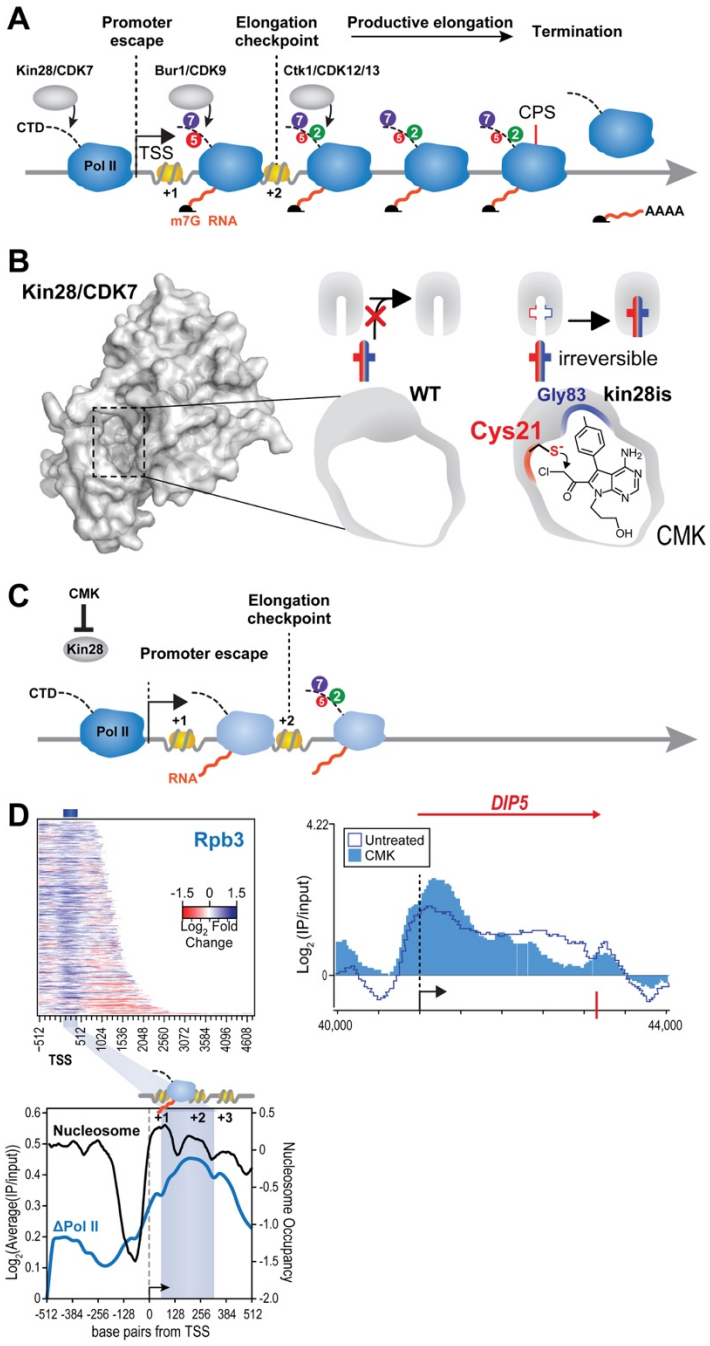



**Figure 5.1 Findings from Kin28 Inhibition Studies in Yeast**

**(A)** Diagram of CTD phosphorylation events and related kinases during the transcription cycle. **(B)** Schematic of chemical-genetic inhibition of Kin28 with CMK. **(C)** Inhibition of Kin28 is not essential for promoter escape, but defects in RNA synthesis and **(D)** entry into the elongation phase of transcription occur.

### 5.3 Effects of Kin28/CDK7 Inhibition Are Gradated and Cell-type Specific

Interestingly, inhibiting Kin28 did not have an equal impact on transcription genome-wide. Rather, though Pol II occupancy was dramatically reduced after the post-initiation pause at most genes, it was relatively unchanged at others (**Figure 5.1C**). Newly synthesized RNA levels were depleted genome-wide after inhibition of Kin28, although steady-state levels varied by transcript [7]. This spectrum of effect is reminiscent of the recent observation that transcription of oncogenes seems particularly sensitive to CDK7 inhibition by THZ1 [8].

THZ1, which has recently entered pre-clinical trials [9], can selectively dampen transcription of oncogenes that promote small cell lung cancer [10] and triple-negative breast cancer [11], both of which are notorious for their aggressive progression and lack of effective treatments. Unlike flavopiridol ,the first clinically approved broad spectrum CDK inhibitor  with varying affinity to CDKs 1, 2, 4, 6, 7, 8, 9, 12, and 13 [12, 13], THZ1 has remarkable affinity for only CDK7 (3 nM) and modest affinity for CDK12/13 [4]. Akin to our finding that inhibition of Kin28/CDK7 impacts Pol II to varying degrees, THZ1 stops the growth of only those cancerous cells that are “addicted to” or highly reliant on high levels of transcription of certain oncogenes such as MYC/MYCN [8]. The transcription factor MYC/MYCN globally amplifies transcription and is itself highly expressed in several tumor types [14]. Although some low level of Pol II may properly complete transcription without functional Kin28/CDK7, oncogenes may be especially dependent on efficient transcription or high levels of CTD kinase activity for maintenance of a cancerous cell state.

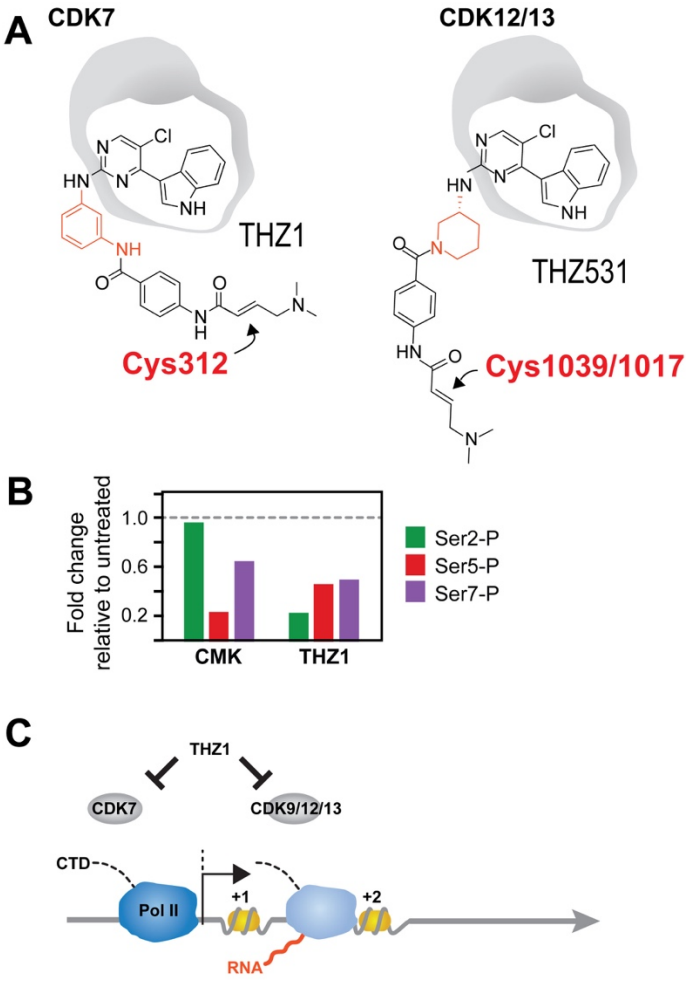



**Figure 5.2 THZ1 May Exert Its Effects by Inhibiting Kinases Other Than CDK7**

**(A)** Schematic of THZ1 and THZ531 molecules. **(B)** Summary of CTD phosphorylation changes, determined by western blotting in CMK- versus THZ1-inhibited cells. **(C)** Diagram illustrating that THZ1 may target multiple CDKs.

### 5.4 THZ1 May Exert Its Effects by Inhibiting Kinases Other Than CDK7

Though THZ1 treatment also elicited the anti-proliferative effects observed in our Kin28/CDK7 inhibition studies, we noted some key differences in phosphorylation and genomic occupancy of Pol II. THZ1, as well as its close analog THZ2 [11], are thought to work by covalently binding to a C-terminal Cys residue (Cys312) that is external the ATP-binding pocket of CDK7 (**Figure 5.2A**). Though the structure of the C-terminal region, including Cys312, of CDK7 has not yet been solved, docking models predict that the Cys312-THZ1 moiety occludes the ATP-binding pocket. CDKs 8, 12, 13, and 19 have a C-terminal Cys within four residues of where C312 would occur in peptide sequence alignment. THZ1 binds to CDK7 with 100-fold greater affinity than to the other CDKs in vitro (3 nM compared to 300 nM), but can nonetheless affinity enrich other CDKs to varying degrees. Intriguingly, mass spectrometry of co-bound THZ1-biotin proteins reveals enrichment for several other kinases, though of the CDKs, only CDK7 is highly ranked [4]. Further, binding studies of THZ1 with a panel of kinases shows 10 other kinases (TBK1, MAP3K11, MERTK, TYK2, CDK5, CDK2, INSR, CDK1, IFG1R, and MAPK8) with nanomolar binding affinities to THZ1, though none showed indication of covalent binding.

Furthermore, unlike the Pol II pause we observe after inhibiting Kin28/CDK7 inhibitor, THZ1 treatment caused a reduction in Pol II occupancy throughout all stages of transcription. Likewise, THZ1 caused a marked reduction in Ser2, Ser5, and Ser7 phosphorylation levels (inhibition of Kin28 affected mainly Ser5 phosphorylation), which suggests that THZ1 may also impact the activity of other CTD kinases (**Figure 5.2B**). In an elegant experiment, Kwiatkowski et al. substituted the reactive cysteine in CDK7 (Cys312Ser), and observed rescue of Ser5 and 7 phosphorylation, but only partial rescue of Ser2 phosphorylation when treated with THZ1. These data indicate that THZ1 may derive its anti-cancer effects by acting combinatorically on CDK7 as well as other

CTD kinases, particularly those that promote other stages of transcription such as the Ser2 kinases CDK9, CDK12, and CDK13 (**Figure 5.2C**). In fact, the THZ-scaffold was recently adapted to inhibit CDK12 and CDK13 [15] (**Figure 5.2A**, right).

### 5.5 Conclusion

THZ1 and other drugs that aim to target the general transcriptional machinery are an exciting and promising line of therapeutics. The mode of action by which these drugs can selectively kill certain tumor cells and leave healthy cells relatively unharmed may be better understood by testing the hypothesized mechanisms of action in budding yeast with chemical genetic inhibition of rationally designed kinase mutants. The cysteine targeted by the THZ-scaffold lies in a C-terminal region that is not found in Kin28. It would be interesting to introduce this region and test whether it sensitizes Kin28 to THZ1 in yeast. Also, as we have proposed here, combination therapy which targets multiple CTD kinases may provide the synergistic effects needed for anti-cancer activity and should not be overlooked in the design and development of new and effective therapeutics.

## 5.6 References


1. Wu, P., T.E. Nielsen, and M.H. Clausen, *FDA-approved small-molecule kinase inhibitors.* Trends in Pharmacological Sciences, 2015. **36**(7): p. 422-439.

2. Buzko, O. and K.M. Shokat, *A kinase sequence database: sequence alignments and family assignment.* Bioinformatics, 2002. **18**(9): p. 1274-1275.

3. Miller, R.M. and J. Taunton, *Targeting Protein Kinases with Selective and Semipromiscuous Covalent Inhibitors.* Protein Kinase Inhibitors in Research and Medicine, 2014. **548**: p. 93-116.

4. Kwiatkowski, N., et al., *Targeting transcription regulation in cancer with a covalent CDK7 inhibitor.* Nature, 2014. **511**(7511): p. 616-+.

5. Lapenna, S. and A. Giordano, *Cell cycle kinases as therapeutic targets for cancer.* Nature Reviews Drug Discovery, 2009. **8**(7): p. 547-566.

6. Cohen, M.S., et al., *Structural bioinformatics-based design of selective, irreversible kinase inhibitors.* Science, 2005. **308**(5726): p. 1318-1321.

7. Rodriguez-Molina, J.B., et al., *Engineered Covalent Inactivation of TFIIH-Kinase Reveals an Elongation Checkpoint and Results in Widespread mRNA Stabilization.* Molecular Cell, 2016. **63**(3): p. 433-444.

8. Chipumuro, E., et al., *CDK7 Inhibition Suppresses Super-Enhancer-Linked Oncogenic Transcription in MYCN-Driven Cancer.* Cell, 2014. **159**(5): p. 1126-1139.

9. Bunn, P.A., et al., *Small Cell Lung Cancer: Can Recent Advances in Biology and Molecular Biology Be Translated into Improved Outcomes?* Journal of Thoracic Oncology, 2016. **11**(4): p. 453-474.

10. Christensen, C.L., et al., *Targeting Transcriptional Addictions in Small Cell Lung Cancer with a Covalent CDK7 Inhibitor.* Cancer Cell, 2014. **26**(6): p. 909-922.

11. Wang, Y.B., et al., *CDK7-Dependent Transcriptional Addiction in Triple-Negative Breast Cancer.* Cell, 2015. **163**(1): p. 174-186.

12. Carlson, B.A., et al., *Flavopiridol induces G(1) arrest with inhibition of cyclin-dependent kinase (CDK)2 and CDK4 in human breast carcinoma cells.* Cancer Research, 1996. **56**(13): p. 2973-2978.

13. Chao, S.H. and D.H. Price, *Flavopiridol inactivates P-TEFb and blocks most RNA polymerase II transcription in vivo.* Journal of Biological Chemistry, 2001. **276**(34): p. 31793-31799.

14. Loven, J., et al., *Revisiting Global Gene Expression Analysis.* Cell, 2012. **151**(3): p. 476-482.

15. Zhang, T.H., et al., *Covalent targeting of remote cysteine residues to develop CDK12 and CDK13 inhibitors.* Nature Chemical Biology, 2016. **12**(10): p. 876-+.

## Appendix I: Function of RNA Polymerase II CTD Methylation

Asuka Eguchi cultured cells and performed immunohistochemistry. John D. Leonard performed dot blotting. All other experiments presented in this Appendix were performed by SCT.

## I.1 Introduction

The C-terminal domain of the catalytic subunit of RNA Polymerase II (Pol II CTD) is highly repetitive, consisting of the amino acids 'Y1-S2-P3-T4-S5-P6-S7' repeated in tandem 26 times in yeast and, with slight variation in distal repeats, 52 times in human (**Figure I.1A**). The CTD is extensively, reversibly phosphorylated, and the phosphorylation state of the CTD is critical for gene transcription through interaction of phosphorylated S2 and S5 residues (S2p and S5p) with histone chaperones, methyltransferases, and a vast array of co-transcriptional RNA processing complexes [1-3]. Recent developments reveal that multiple enzymes can act on the CTD substrate in a gene- or environment-specific manner [4]. Furthermore, a vertebrate-specific CTD code is becoming apparent. It has been hypothesized that the expansion of the CTD co-evolved with gene duplication and evolution of chordate-specific proteins [5-8]. For example, T4 is dispensable in yeast, but essential in chicken DT-40 cells and human and mouse cells [9]. Phosphorylation of Y1 (Y1p) prevents antisense transcription at promoters and is essential for the function of certain enhancers, which is also a phenomenon specific to mammalian cells and not observed in yeast [10, 11]. The differences in the function of CTD modification between mammalian species and yeast may result from the set of transcriptional machinery present in higher-order eukaryotes but absent in yeast. For example, the phosphorylation of S7 recruits Integrator, a small non-coding RNA processing complex, which is expressed in mammals but without an ortholog in yeast [12, 13].

In addition to differences in usage of the canonical CTD, mammals contain additional repeats of non-consensus CTD, which differ predominately at the 7th position of the heptad. Evidence that these non-consensus CTD residues and their modifications are functionally significant is beginning to accumulate. The asymmetric dimethylation of a non-consensus CTD arginine (S7R) by CARM1/PRMT4 in HeLa cells and mouse

embryonic fibroblasts was found to regulate the expression of a subset of snRNA and snoRNA expression [14]. The most frequent variation of the consensus sequence of the CTD is S7K, with eight repeats punctuated by K7 rather than S7. The ubiquitinylation of K7 by Wwp2 was found to promote the degradation of Pol II [15]. And more recently, Voss *et al.* detected the acetylation, and mono-, di-, and tri-methylation of the CTD from Raji cells by mass spectrometry [16]. In an independent study, acetylation of K7 (K7ac) by p300 and ChIP-Seq of K7ac was also reported, and was found to pause Pol II at growth-factor response genes in mouse ESCs [17]. Interestingly, mass spectrometry experiments indicate modification of CTD lysines is always detected on phosphorylated peptides with phosphorylation of other CTD residues in the same or adjacent heptads [16], but phosphorylation of adjacent S2 and S5 residues was found to inhibit the process of acetylation of K7 (and vice versa—pre-acetylated CTD substrates are phosphorylated less efficiently at S2 and S5) [17].

Here we utilized the monoclonal antibody '1F5', generated and characterized in [16], to elucidate the function of CTD methylation in the regulation of gene expression in human embryonic stem cells (hESCs). We found that chromatin-associated Pol II was constitutively methylated, and, consistent with previous data [16], that detection of K7me diminished as Pol II departed from the transcription start site (TSS) of highly transcribed genes as a result of epitope masking by hyper-phosphorylation with S5p. In contrast, we found extensive co-enrichment of Y1p with K7me. A comparison of K7ac and K7me ChIP peaks suggest a sequential modification of K7 residues that may serve to prime the CTD for subsequent phosphorylation. Finally, through comprehensive ChIP-Seq correlation analysis, we suggest proteins that may potentially interact with the methylated CTD of Pol II.

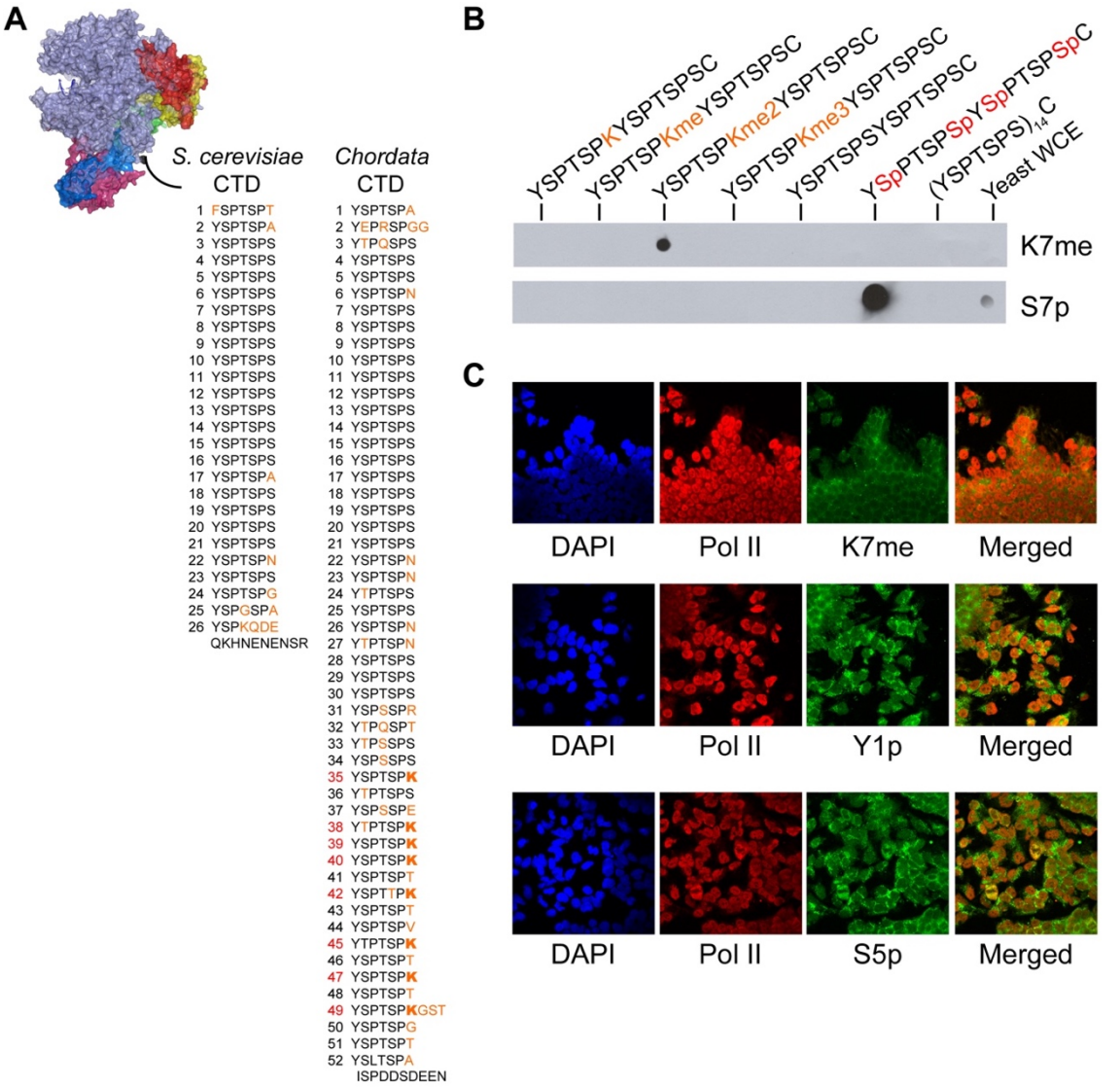



**Figure I.1 Antibody 1F5 Recognizes Mono- and Di-Methylated Lysines on the CTD**

**(A)** Crystal structure of Pol II (PDB# 1Y1W) showing the sequence of yeast and chordate CTD, adjacent. Deviations from consensus are highlighted in orange, and chordate heptads which contain the S7K polymorphism are highlighted in red. **(B)** Dot blot of synthetic CTD peptides and yeast whole cell extract (WCE) showing the 1F5 antibody binds strongly to di-methylated K7 (K7me2), but not mono- or tri-methylated K7 (K7me, K7me3). An antibody, which detects S7p, serves as a positive control for yeast Rpb1, which does not stain positive for K7me because yeast lacks non-consensus heptads containing K7 residues. **(C)** Immunocytochemistry of hESCs with antibodies, which recognize Pol II irrespective of phosphorylation state (Pol II), methylation of K7 (K7me), phosphorylated Y1 CTD (Y1p), and phosphorylated S5 CTD (S5p). Merged images with DAPI nuclear stain showing methylated CTD detected predominantly in the cytoplasm, which contrasts with the more nuclear Y1p and S5p modifications.

### I.2 The CTD of Pol II Is Methylated in Human Embryonic Stem Cells

We chose to explore the methylation of Pol II in hESCs (H1 cell line) because the non-consensus CTD repeats, particularly the K7 residues, function in regulating the expression of growth and development genes [17-19]. Dot blot assays confirmed the 1F5 antibody primarily recognizes CTD peptides containing di-methylated lysines (**Figure I.1B**). It did not detect yeast whole cell extract, which does not contain Pol II with lysine in its CTD, or mono- and tri-methylated CTD. Voss *et al.* demonstrate 1F5 can also bind to mono-methylated CTD, but with less affinity (~1/10) than to di-methylated CTD, which is consistent with the results of our immunoblotting experiments [16]. Also consistent with previous observations [16], K7me was predominantly detected in the cytoplasm by immunofluorescence staining, though methylated Pol II decorated both nuclear and cytoplasmic Pol II in hESCs (**Figure I.1C**).

### I.3 K7me Is Detected at the 5' End of Genes before Transcription Elongation

Next, we performed ChIP-Seq with the K7me-specific antibody to identify binding sites of methylated Pol II genome-wide in hESCs. Again, we chose to study K7me in hESCs because the functional significance of non-consensus CTD repeats has been associated with early development and growth, and has not yet been characterized in human pluripotent cells [14, 17, 18]. Further, H1 cells are classified by ENCODE Consortium as a Tier 1 cell line, which is a group designation reserved for widely used cell types that have an abundance of pre-existing datasets [20]. Previous ChIP-Seq experiments of K7me were conducted in Raji cells [16], for which substantially less data has been generated as compared to H1 cells. In Raji cells, K7me was mostly associated with the 5' end of genes because subsequent S5p past the TSS masked the ability for the antibody to detect K7me. Y1p of the CTD was shown to also peak at the 5' end of genes in Raji and chicken DT-40 cells [10, 11]. As such, we also performed ChIP-Seq of

**Table I.1 Summary of ChIP-Seq Alignment Statistics**

| Sample | Antibody | Replicate | Total Reads ($10^6$) | Overall Alignment (%) |
|---|---|---|---|---|
| Pol II | N-20 | 1 | 40.8 | 98.98 |
| | | 2 | 48.6 | 99.06 |
| K7me | 1F5 | 1 | 46.2 | 97.71 |
| | | 2 | 49.8 | 97.86 |
| Y1p | 3D12 | 1 | 30.5 | 96.47 |
| | | 2 | 34.5 | 98.29 |
| S5p | 3E8 | 1 | 44.8 | 95.56 |
| | | 2 | 52.5 | 95.81 |
| Input | N/A | 1 | 55.2 | 98.47 |
| | | 2 | 64.1 | 98.23 |

Pol II irrespective of phosphorylation state (antibody N-20) and Pol II phosphorylated at Y1 and S5 (antibodies 3D12 and 3E8, respectively). This allowed us to compare co-occurrence of methylation with phosphorylation of CTD residues in hESCs. **Table I.1** summarizes the ChIP-Seq experiments performed in this study and genome alignment statistics.

Overall, the enrichment of ChIP-Seq reads across the genome correlated strongly (Pearson correlation coefficient R > 0.7) between Pol II, K7me, Y1p, and S5p ChIPs (**Figure I.2**). As expected, K7me and Y1p occurred near TSSs, whereas S5p persisted into the gene body with transcribing Pol II (**Figure I.3** and **I.4**). Y1p and K7meChIP-Seq profiles were essentially the same, with Y1p and K7me ChIP signal derived from the same biological replicate correlating more strongly than replicate ChIPs from different replicates (R = 0.93 versus 0.89 for Y1p ChIP, **Figure I.2**). This suggests that Y1p does not interfere with the 1F5 antibody's ability to detect K7me, and that CTD containing phosphorylated Y1 may in fact be simultaneously methylated. Indeed, the co-occurrence of Y1p on the same and adjacent CTD repeats as methylated K7 residues is also observed by mass spectrometry [16].

K-means clustering of input-normalized K7me ChIP signal centered over the TSSs of protein-coding and non-coding genes (n=55,671, *hg19*) resulted in distinct profiles of K7me enrichment, ranging from high (*Cluster 1-3*) to low (*Cluster 4-6*) abundance of detectable CTD methylation, which was indicative of gene expression levels (further discussion of this point below) (**Figure I.4A**). Pol II ChIP profiles overlapped K7me profiles at the 5' end of genes, which suggests Pol II is methylated at K7 early in transcription *or* that Pol II is methylated before binding to chromatin. The low-signal *Clusters 4*, *5*, and *6* appear depleted of Pol II, and thus also K7me, which is detected at all Pol II-bound genes (**Figure I.4B** and **Figure I.5B**). Genes in the high-signal *Clusters 1*, *2*, and *3* did not differ significantly in length between clusters (**Figure**

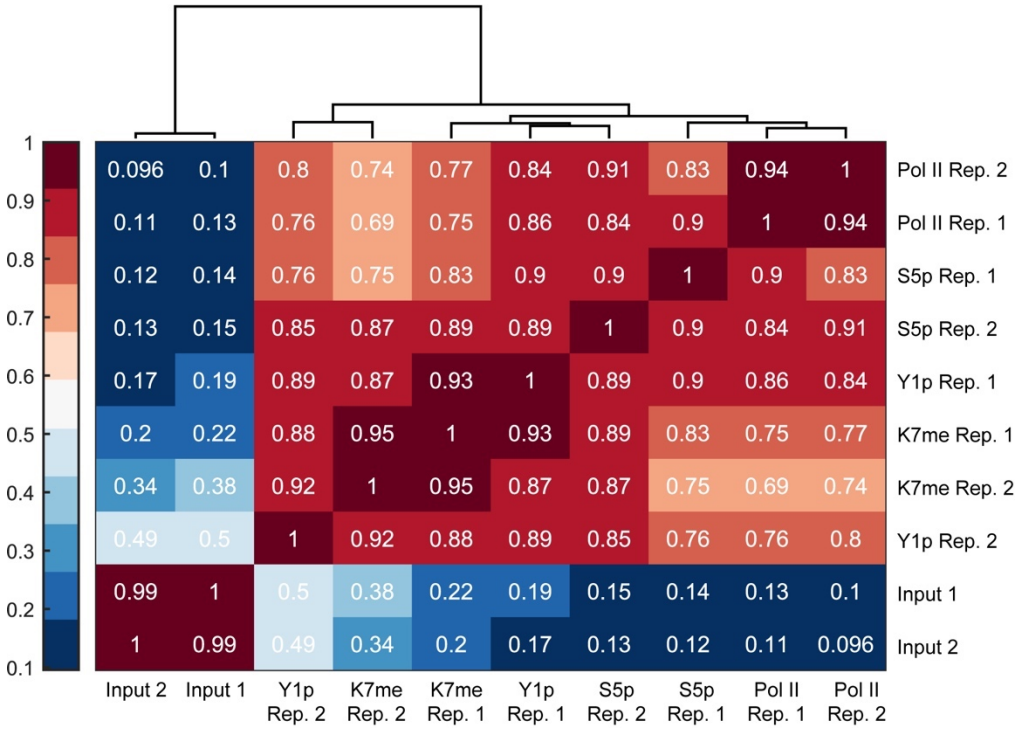



**Figure I.2 ChIP-Seq Results Are Comparable between Biological Replicates**

Heat map showing Pearson correlation coefficients between ChIP-Seq reads across the genome. Biological replicates (Rep. 1 and 2) cluster together by hierarchical clustering.

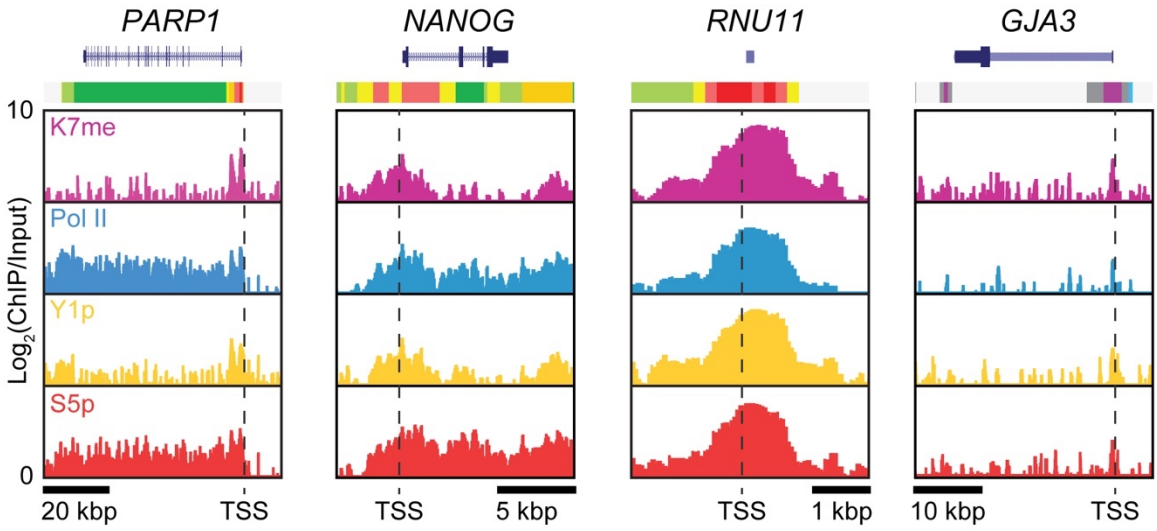



**Figure I.3 K7me ChIP Profiles at Example Genes**

Example input-normalized ChIP-Seq reads for Pol II and Pol II modifications at indicated gene in navy. ChromHMM state is color-coded as in Figure I.6B. The transcription start site (TSS) is indicated by the black dashed line, and the scale is annotated below by the black bar.

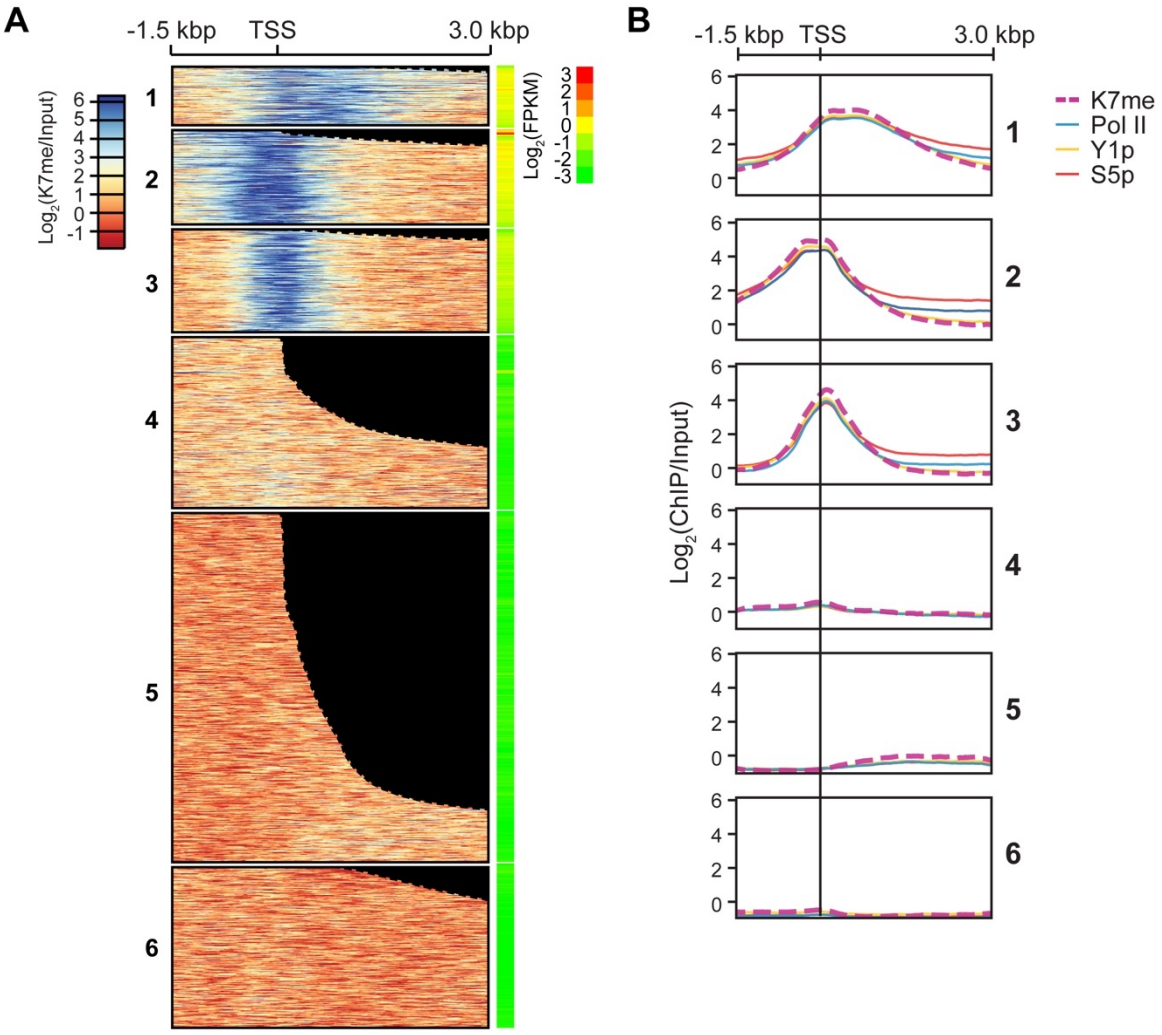



**Figure I.4 Genome-wide Profile of K7me, Pol II, Y1p, and S5p in hESCs**

**(A)** Heat map of K7me ChIP-Seq signal generated by k-means clustering of K7me ChIP-Seq signal centered around the transcription start site (TSS), flanked by 1.5 and 3.0 kbp before and after the TSS, respectively; heat map of RNA expression, adjacent. **(C)** Metagene profiles of K7me, Pol II, Y1p, and S5p ChIP-Seq signal in indicated clusters.

**I.5A**). However, the width of the K7me peaks differed in broadness over the TSS between genes in each cluster, despite having similar peak intensities. Though it has been demonstrated that K7me epitopes are masked by subsequent hyper-phosphorylation of the CTD (*i.e.* S5p) [16], methylated CTD remained detectable over ~1.5 kbp into the gene body for *Cluster 1* genes. Decreasing K7me signal may also be explained by de-methylation after transcription initiation, followed by acetylation or further methylation. Potentially, K7me may also be encapsulated by CTD binding proteins, which would also result in diminished ChIP signal.

Despite the differences in K7me peak width between clusters, *Cluster 1*, *2*, and *3* were composed of similar types of genes (**Figure I.6A**) with similar chromatin states (**Figure I.6B**). Interestingly, we did not observe enrichment of K7me or Pol II at Polycomb-repressed regions (**Figure I.7**) but could readily detect K7me at bivalent or poised chromatin states. The enrichment of K7me at enhancer and insulator elements is consistent with the previously described binding of Pol II at these important cis-regulatory regions.

### I.4 Methylation of Pol II Does Not Correlate Linearly with Gene Expression Levels

Next, we asked whether intensity of K7me, which more or less mimicked Pol II levels, correlated with gene expression levels. Pol II may be bound to genes but may not be actively transcribing, as is the case with promoter-proximal paused and poised Pol II [21, 22]. To answer this question, we compared K7me ChIP-Seq enrichment to RNA-Seq reads collected from hESCs [23]. Plotting the best-fit line between maximum K7me signal (in an interval centered around the gene flanked by 3000 bp) and FPKM per gene revealed only a weak positive correlation ($R^2$=0.146). Genes bound by greater levels of methylated Pol II—though on average—were not necessarily more expressed (**Figure I.6C**).

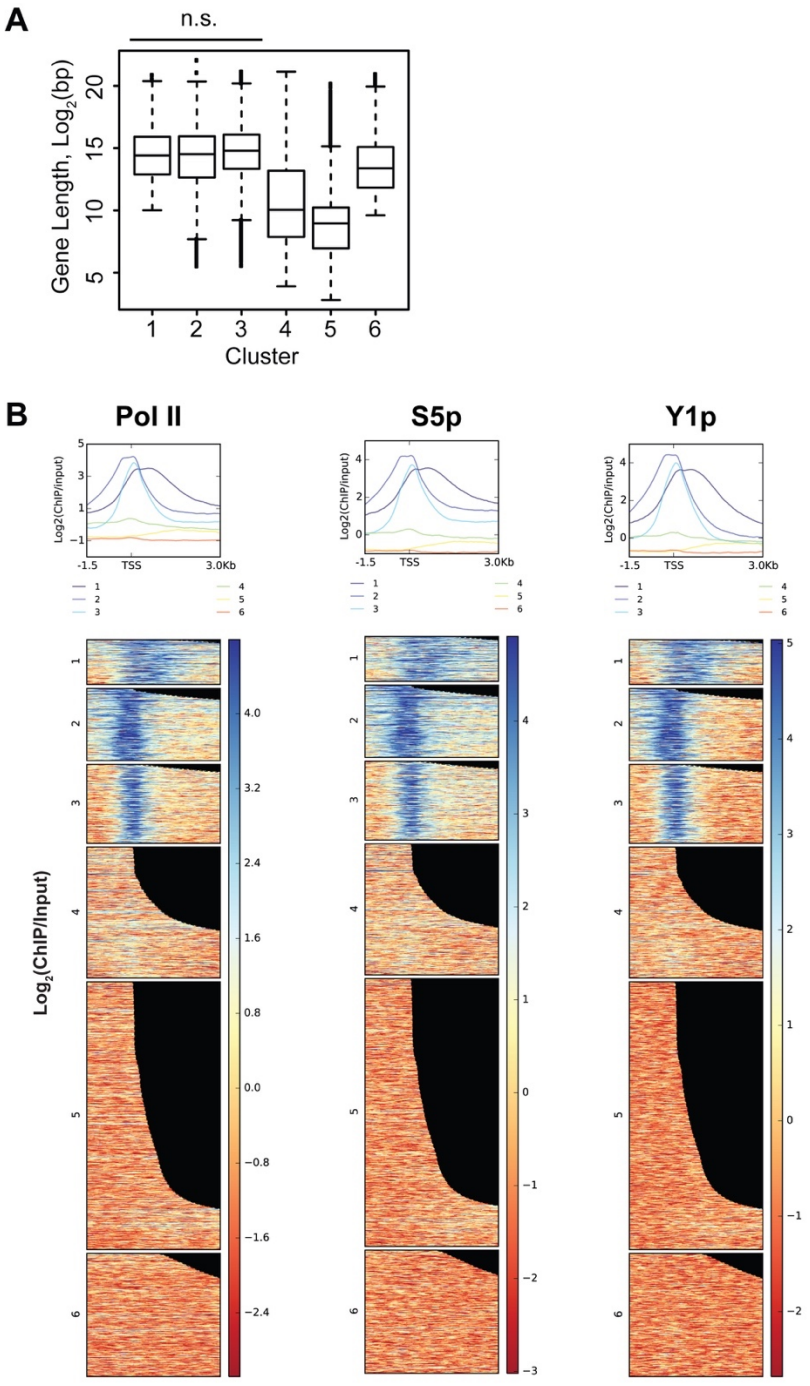



**Figure I.5 K7me Enriched Clusters Are Also Enriched for Pol II, Y1p, and S5p**

**(A)** Boxplot showing length of genes in each cluster is not significantly (n.s.) in clusters with K7me signal. **(B)** Heat maps and metagene plots for Pol II, S5p, and Y1p ChIP-Seq in clusters defined by K7me in Figure 2.

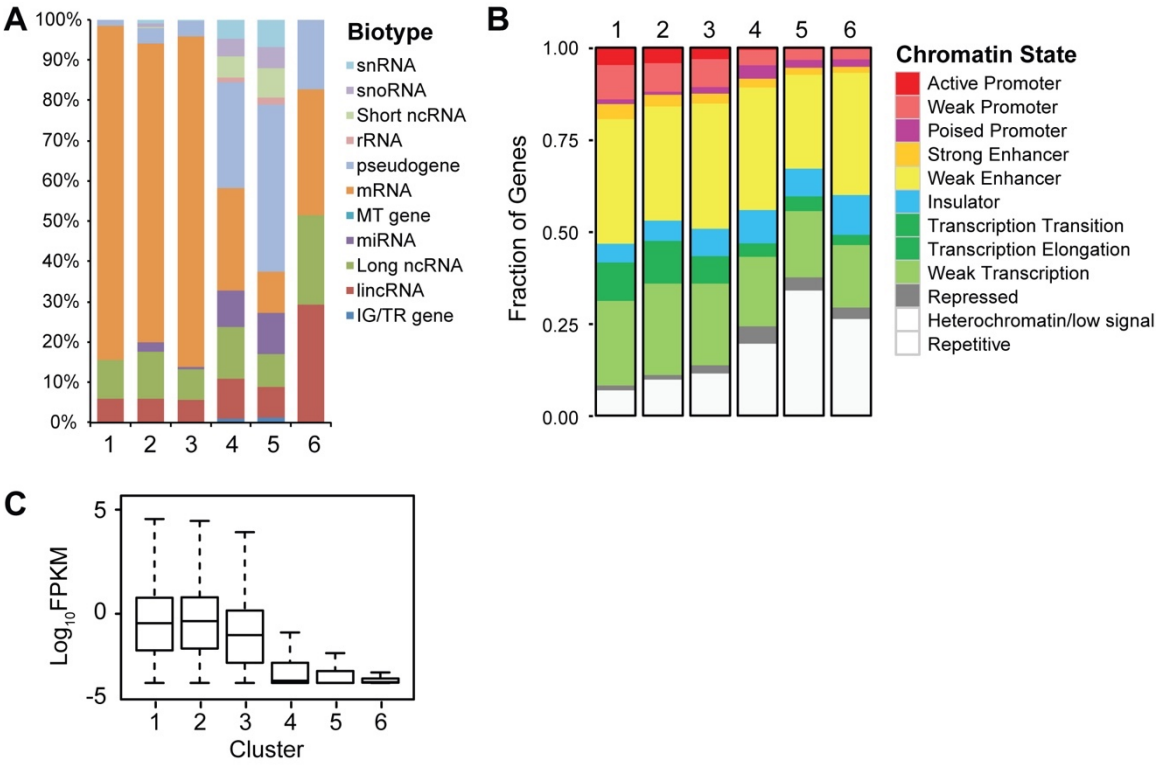



**Figure I.6 Analysis of Gene Expression and Chromatin State of K7me-bound Regions**

**(A)** Bar graph showing distribution of gene biotypes for each cluster as defined by ENSEMBL. **(B)** Bar graph showing distribution of chromatin states for each cluster as in ChromHMM defined by the Broad Institute. **(C)** Boxplot showing distribution of RNA expression (Log2FPKM) of genes in each cluster, outlier points not plotted.

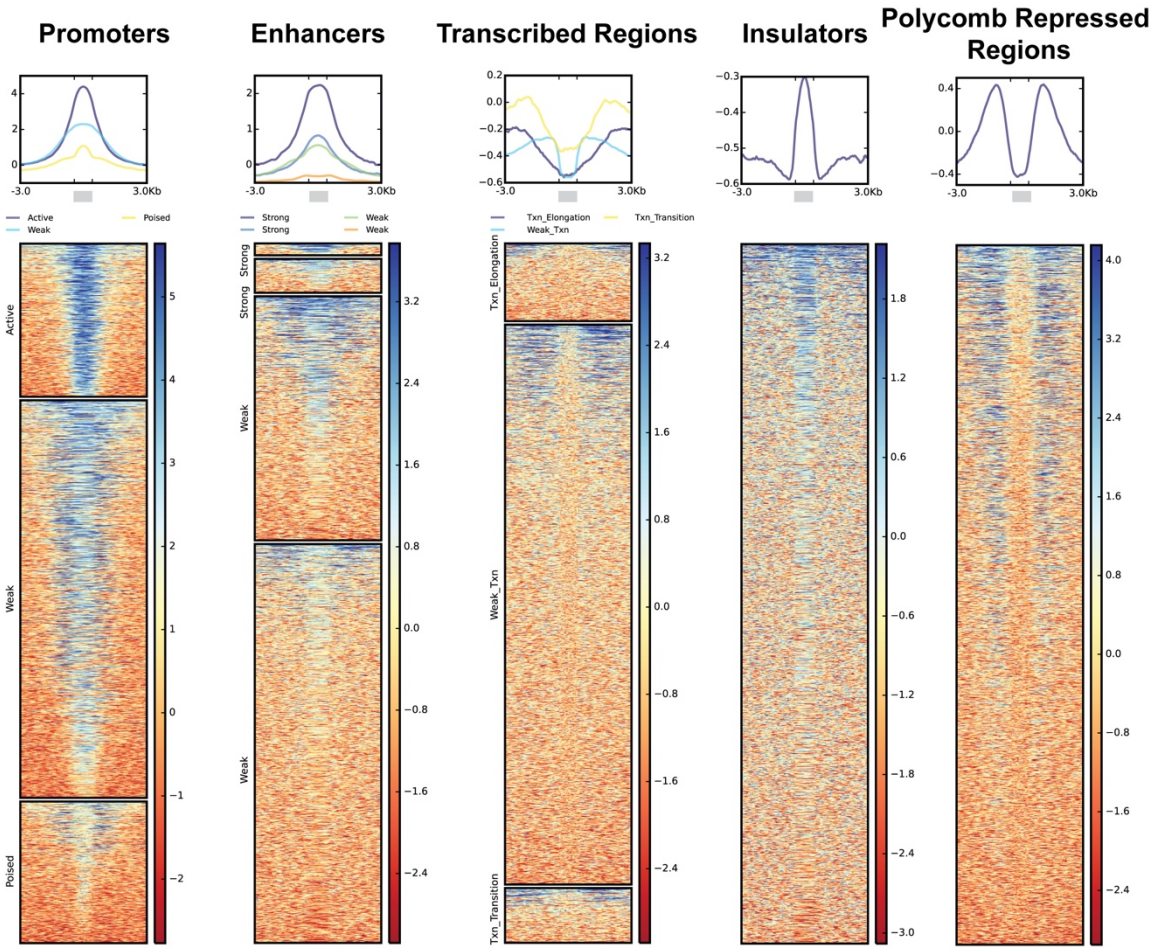



**Figure I.7 K7me ChIP-Seq Signal Centered Over ChromHMM Features**

Heat maps and metagene plots for K7me ChIP-Seq signal centered over the different ChromHMM features defined in Figure I.6, shown here as a gray rectangle compressed into 100 bins depicted in total as 1 kbp. The signal in the 3 kbp flanking regions is absolute, *i.e.* not binned.

To determine what dictated gene expression levels despite being bound by similar levels of Pol II, we compared the levels of Pol II, K7me, Y1p, and S5p in highly expressed (above median FPKM) and lowly expressed (below median FPKM) genes from K7me-enriched clusters, *i.e. Clusters 1*, *2*, and *3*. The enrichment of K7me and Y1p was slightly less between highly and lowly expressed genes over the TSS (**Figure I.8**). However, although Pol II and S5p signal also appeared slightly less at the TSS, Pol II and S5p were considerably more enriched in the gene body of highly expressed genes compared to lowly expressed genes. Thus, when the ChIP signals for Pol II modifications were normalized to the amount of Pol II signal, we observed that genes with high expression rapidly lost K7me ChIP signal after the TSS (**Figure I.9A**). This trend was also observed for Y1p, and was opposite for S5p (*i.e.* increased relative S5p levels corresponded to highly expressed genes). When compared to relative Pol II levels, it was apparent that K7me peaks over the TSS, which is relatively depleted of S5p. This phenomenon is consistent with the idea that K7me epitopes are masked by hyper-phosphorylation of the CTD with S5p. Phosphorylation of S5 occurred later in the gene body for *Cluster 1*, which is consistent with detection of K7me and Y1p further past the TSS for genes in *Cluster 1*. It has been shown previously that the antibody for Y1p (3D12) also does not efficiently detect Y1p that is adjacent to phosphorylated T4, S5, and S7 residues [24]. Taken together, Pol II may be constitutively methylated, but detection of K7me ChIP diminishes because of epitope masking by other CTD modifications. However, this does not exclude the possibility that K7me is also removed, bound, further methylated, or exchanged for another modification after Pol II departs from the TSS.

To clarify if Y1p and K7me are actually confined to non-elongating Pol II, we examined the transcription state of the promoters in the high and low expression groups of genes as defined by the methylation state of histone lysines at their promoters,

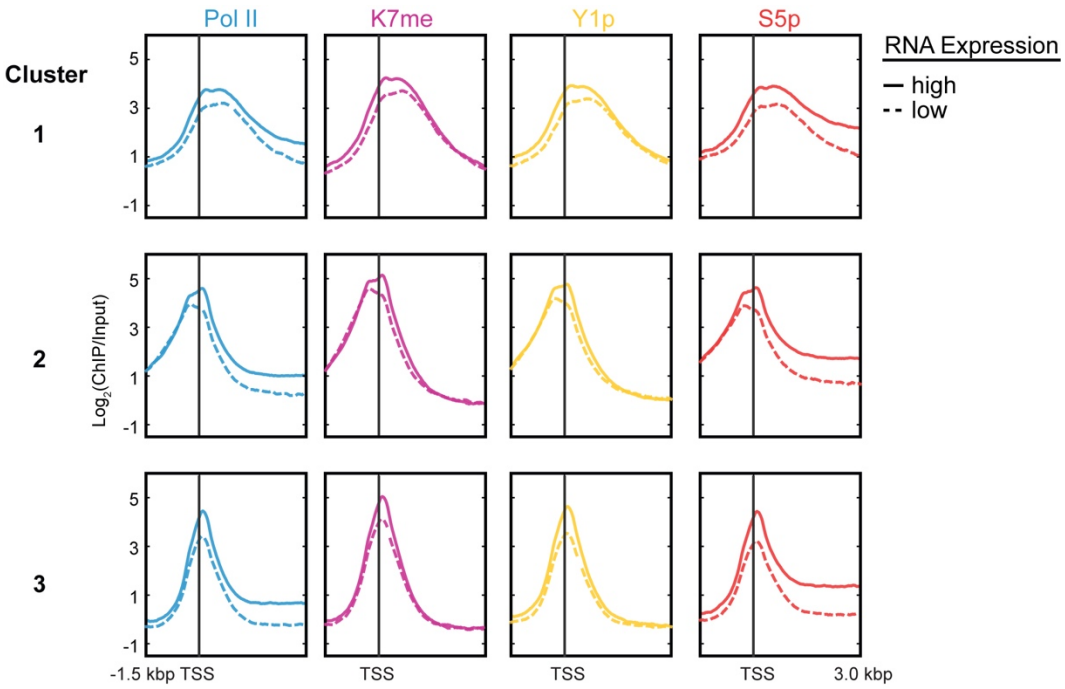



**Figure I.8 ChIP-Seq Average Profiles at the TSS by Expression Level**

Metagene profiles of input-normalized ChIP-Seq reads centered around the transcripton start site (TSS )of high (greater than median FPKM, solid line) or low (less than median FPKM, dashed line) expression genes from Clusters 1-3, median = 0.5405 FPKM.

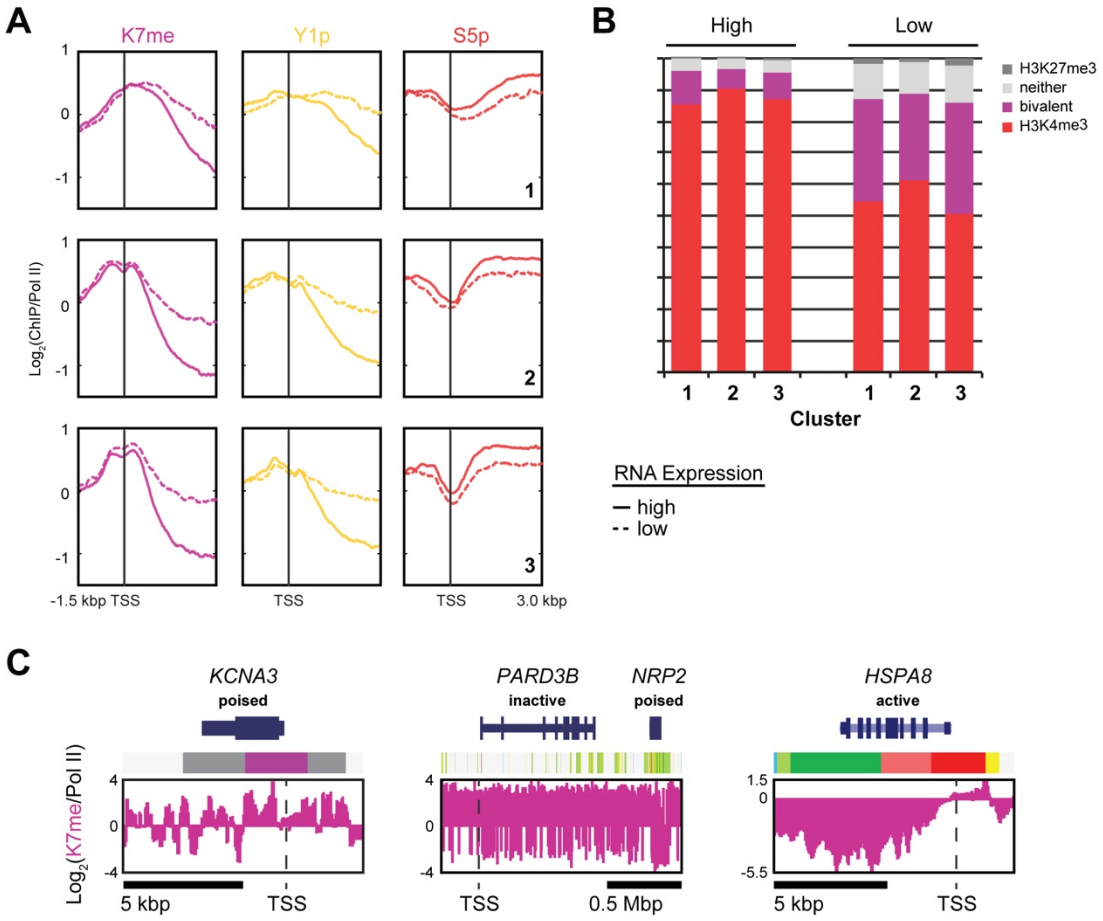



**Figure I.9 Methylation of the CTD Is Enriched at the TSS before Hyper-phosphorylation of S5p**

**(A)** Metagene profiles of input-normalized or Pol II-normalized ChIP-Seq reads centered around the transcription start site (TSS) of high (greater than median FPKM, solid line) or low (less than median FPKM, dashed line) expression genes from Clusters 1-3, median = 0.5405 FPKM. **(B)** Bar graph showing fraction of high and low expression genes in transcription states defined by histone H3 lysine methylation at the promoter. 'Neither' designation has neither H3K4me3 nor H3K27me3 enriched at the gene promoter. 'Bivalent' designation has both marks. **(C)** Examples of Pol II-normalized K7me ChIP-Seq reads at active, poised, and inactive genes indicated in navy. ChromHMM states are color-coded as in Figure 3B. The TSS is indicated by the black dashed line, and the scale is annotated below by the black bar.

previously annotated in H1 cells [25] (**Figure I.9B**). Genes from the low expression group, *i.e.* genes with broad K7me and Y1p peak past the TSS, were marked by comparably more poised (H3K4me3 and H3K27me3) and inactive or repressed (H3K27me3) promoters than genes from the high expression group. Genes from the high expression group, *i.e.* genes with K7me and Y1p peaks confined to the TSS, were associated with more active chromatin (H3K4me3). K7me is relatively enriched on genes containing inactive and poised Pol II, but depleted after the TSS on actively transcribed genes (**Figure I.9C**). These results suggest K7me and Y1p may mark non-elongating Pol II. Whether this is an inadvertent consequence of epitope masking, an actual representation of removal or replacement of Y1p and K7me modifications, or a combination of both, remains to be tested.

### I.5 Methylation and Acetylation of the Same K7 Residue May Occur Sequentially

Recently, it was shown by mass spectrometry that methylated repeats were also acetylated on the same K7 residue, which suggests de-methylation/acetylation and methylation/de-acetylation events occur *in vivo [16]*. Additionally, di- and tri-heptads were simultaneously methylated and acetylated on adjacent repeats. To elucidate the interplay between K7me to K7ac, we compared the K7me data described here from hESCs to previously published K7ac ChIP-Seq data from mouse ESCs [17]. We compiled a list of homologous mouse genes by identifying mouse genes with over 10% sequence homology to human genes using UCSC's liftOver program [26]; 11,319 human genes were matched to 14,946 mouse genes. Like K7me, K7ac is enriched at the 5' end of genes. Interestingly, K7ac levels at the TSS did not correlate linearly with K7me levels (**Figure I.10A**); acetylation of Pol II at genes with low K7me signal was also observed (**Figure I.11**). Furthermore, the differences in peak broadness observed for K7me between *Clusters 1-3* was not as apparent for K7ac. Whereas K7me peaked over the

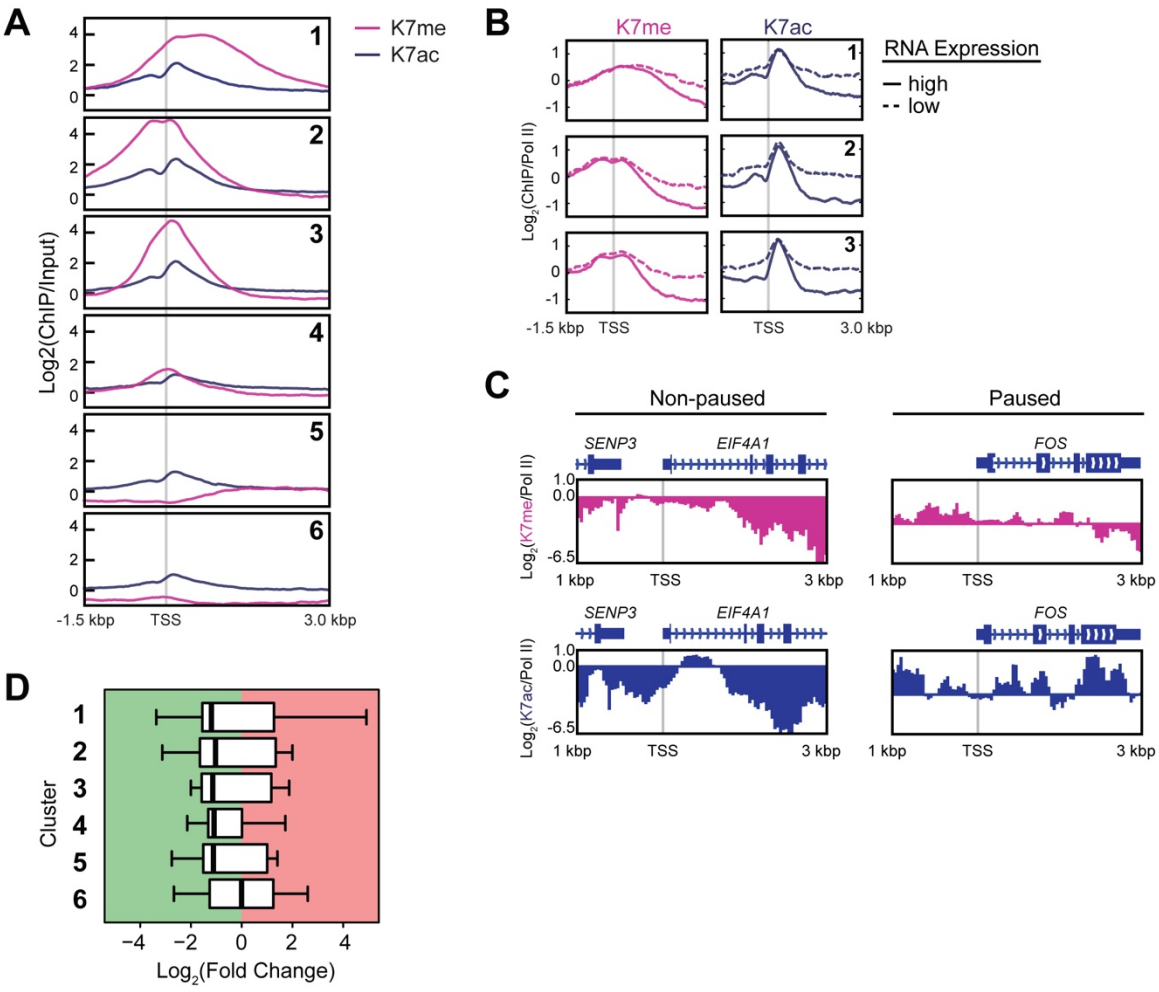



**Figure I.10 Methylation and Acetylation of Pol II Lysines May Be Sequential**

**(A)** Metagene profiles showing K7ac versus K7me after and over the transcription starte site (TSS), respectively. Clusters are as defined previously in Figure I.4. **(B)** Metagene profiles of Pol II-normalized ChIP-Seq reads centered around the TSS of high (greater than median FPKM, solid line) or low (less than median FPKM, dashed line) expression genes from Clusters 1-3, median = 0.5405 FPKM. **(C)** K7me and K7ac ChIP profiles centered over the TSS, flanked by 1.5 and 3.0 kbp, of non-paused (EIF4E, and adjacent SENP3) and paused (FOS) genes, as indicated. The human and mouse gene diagrams are shown above the profiles for K7me and K7ac ChIP, respectively. **(D)** Boxplot showing the distribution of fold-changes in differentially expressed genes in K7R mouse embryonic fibroblasts (MEFs) by the clusters defined by K7me. Green highlights decreased gene expression in K7R MEFs relative to wild-type, and red highlights increased gene expression in K7R MEFs relative to wild-type.

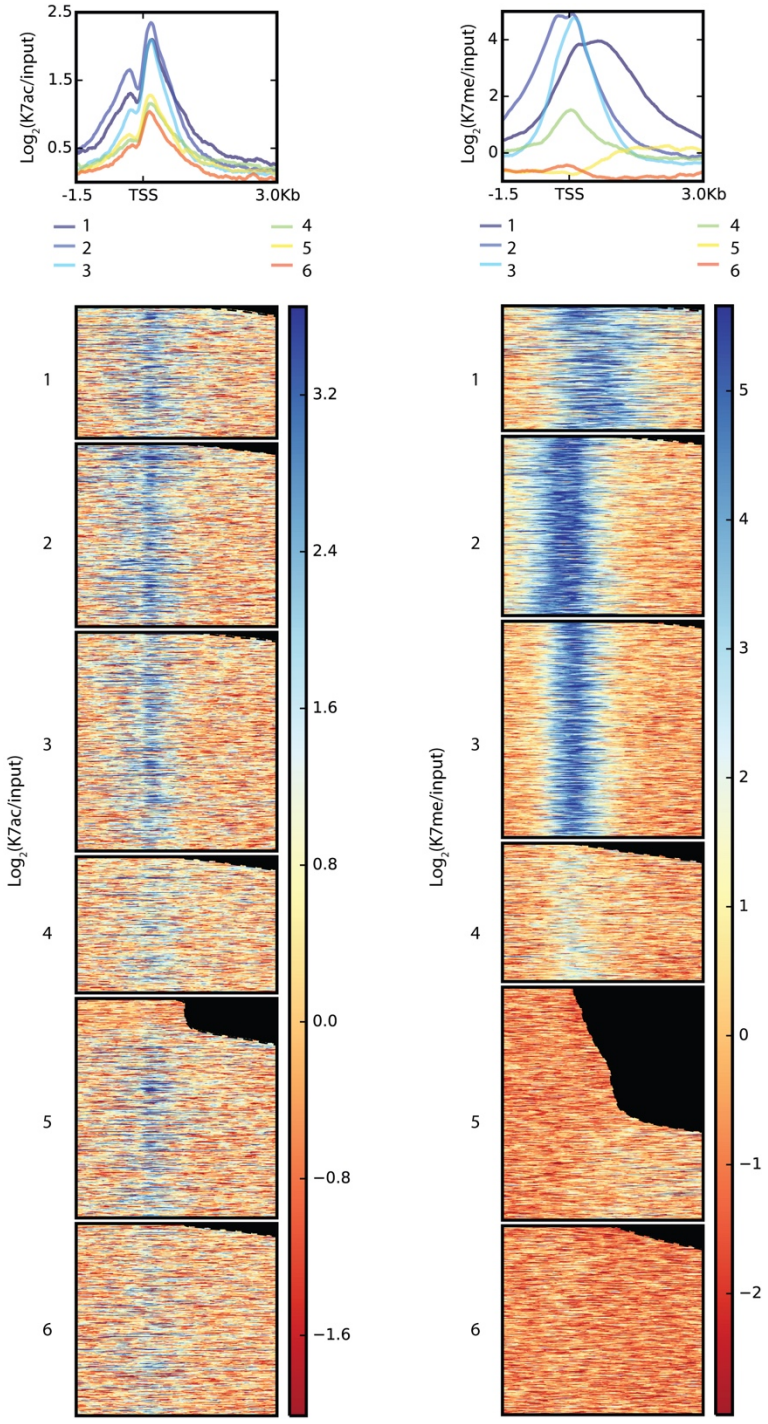



**Figure I.11 A Comparison of K7ac and K7me**

Heat maps and metagene plots for K7ac and K7me in genes conserved in mouse and human embryonic stem cells showing ChIP-Seq signal separated by the clusters defined in Figure I.4.

TSS, K7ac peaked immediately *before* and even more *after* the TSS, *i.e.* K7ac was slightly depleted over the TSS. This is consistent with the replacement of methylation with acetylation after Pol II begins transcription.

The sequential enrichment for K7me and then K7ac was more apparent when ChIP signal was normalized to Pol II ChIP (**Figure I.10B**). K7ac diminishes slightly further into the gene body for genes in *Cluster 1*, and especially for lowly expressed genes in all three K7me-associated clusters as was observed in **Figure I.9** for K7me and Y1p. Previously, K7ac was shown to be required for promoter-proximal pausing at growth-factor-induced genes (*e.g. FOS*), and replacement of K7 residues with arginine (K7R) resulted in severely decreased induction of gene expression of these genes but not housekeeping genes (*e.g. EIF4A1*) [17]. The authors noted K7ac peaks were narrowly restricted to the pause region around the TSS for paused genes, whereas K7ac peaks were shifted into the gene body for non-paused genes, and ascribed the phenomenon to increased dwell time of promoter-proximal Pol II at paused genes. In our analysis, we also observed K7ac peaks further into the gene body for highly expressed genes such as *EIF4A1*, and K7ac peaks abut the TSS for paused genes such as *FOS* (**Figure I.10C**). However, we find this is a general phenomenon for highly expressed versus lowly expressed genes, and may not be specific to growth-factor-induced genes (**Figure I.10B**).

Next, we compared our data to the previously published RNA-array data from mouse embryonic fibroblasts containing mutated K7R-CTD in place of wild-type CTD [19]. In this study, the authors found that the 1882 genes that were differentially expressed (false discovery rate of 5%) were enriched for gene ontology terms “cell adhesion”, “vasculature development”, and “blood vessel development.” Of these differentially expressed genes, 1091 were annotated in the homologous mouse gene list described here. Interestingly, clusters enriched for K7me (*i.e. Clusters 1-3*) included a

greater proportion of these dis-regulated genes than clusters not enriched for K7me (*i.e. Clusters 4-6*) ($p < 0.0001$; **Table I.2**). This suggests methylation of K7 may be especially important for the proper expression of growth and development genes. However, plotting the distribution of fold-change in expression of the most affected genes in wild-type versus K7R cells (multiple-test adjusted $p < 0.05$ and $|\text{fold-change}| > 2$, $n=250$) by K7me clusters, showed only modest differences in distribution between clusters (**Figure I.10D**). *Cluster 1* genes contained the most greatly impacted genes by fold-change by the loss of K7. In summary, the change in transcriptional program incurred by the mutation in mouse fibroblasts, particularly in the clusters associated with higher K7me modification, suggests modification of K7 is important for regulating transcription and that this form of regulation is conserved in mouse and human cells.

### I.6 K7me Co-localizes with Known Chromatin Modifiers and Binding Proteins

Finally, to identify proteins that can potentially interact with methylated CTD, we calculated the Pearson correlation coefficient between our ChIP datasets and several chromatin marks and modifiers (**Figure I.12**) and transcription factors (**Figure I.13**) deposited in the ENCODE Project database for hESCs [20]. Though co-localization is not definitive evidence of association, it may narrow the list of candidate proteins to investigate. As expected, K7me was enriched on chromatin that was similarly enriched by H3K4me2/3 ($R = 0.54/0.59$), which are chromatin marks associated with poised and active genes [27, 28]. Similarly, ChIP of RBBP5, a core component of the H3K4 methyltransferase complex was also correlated with K7me ($R = 0.47$). Notably, the histone methylation peaks are bifurcated, unlike those for K7me, which confirms the specificity of the 1F5 antibody for methylated CTD and not histones (**Figure I.14**).

Interestingly, ChIP of PHF8, a lysine demethylase known to physically associate with the CTD [29] also showed moderate correlation with K7me ChIP ($R = 0.48$). PHF8

**Table I.2 Chi-square Goodness of Fit Test**

Contigency table and results of chi-square “goodness of fit” test showing an unequal number of genes between clusters defined by K7me ChIP signal are dis-regulated in mouse embryonic fibroblasts containing 8KR-CTD

| Cluster | Genes Dis-regulated in 8KR | Expected Proportion | Percentage Deviation | Standardized Residuals |
|---|---|---|---|---|
| 1 | 209 | 2206/14946 | +29.79% | +3.78 |
| 2 | 257 | 2902/14946 | +21.32% | +3.10 |
| 3 | 279 | 3440/14946 | +11.11% | +1.76 |
| 4 | 103 | 2029/14946 | -30.46% | -3.71 |
| 5 | 104 | 2192/14946 | -35.00% | -4.43 |
| 6 | 139 | 2177/14946 | -12.53% | -1.58 |

Observed frequency = 1091
Expected proportion = 1.0

Chi-square = 62.86
Degrees of freedom = 5
P = < 0.0001

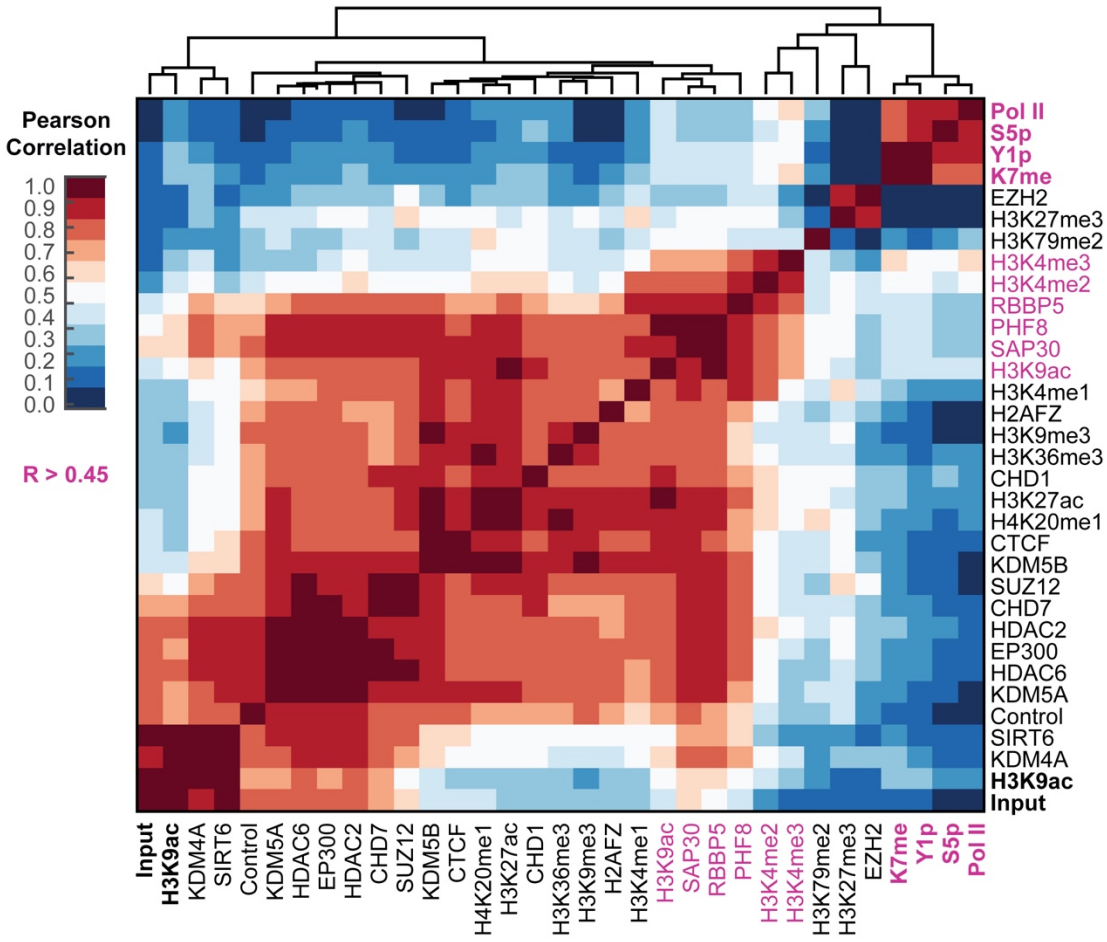



**Figure I.12 Correlation of Methylated Pol II and Chromatin Modification**

Methylated Pol II co-localizes with known chromatin modifiers and binding proteins. Heat map showing the Pearson correlation coefficients of ChIP-Seq performed in this study (bold) versus the ChIP-Seq from the Bernstein Lab. Proteins with associated correlation coefficients between K7me ChIP-Seq greater than 0.4 are highlighted in pink.

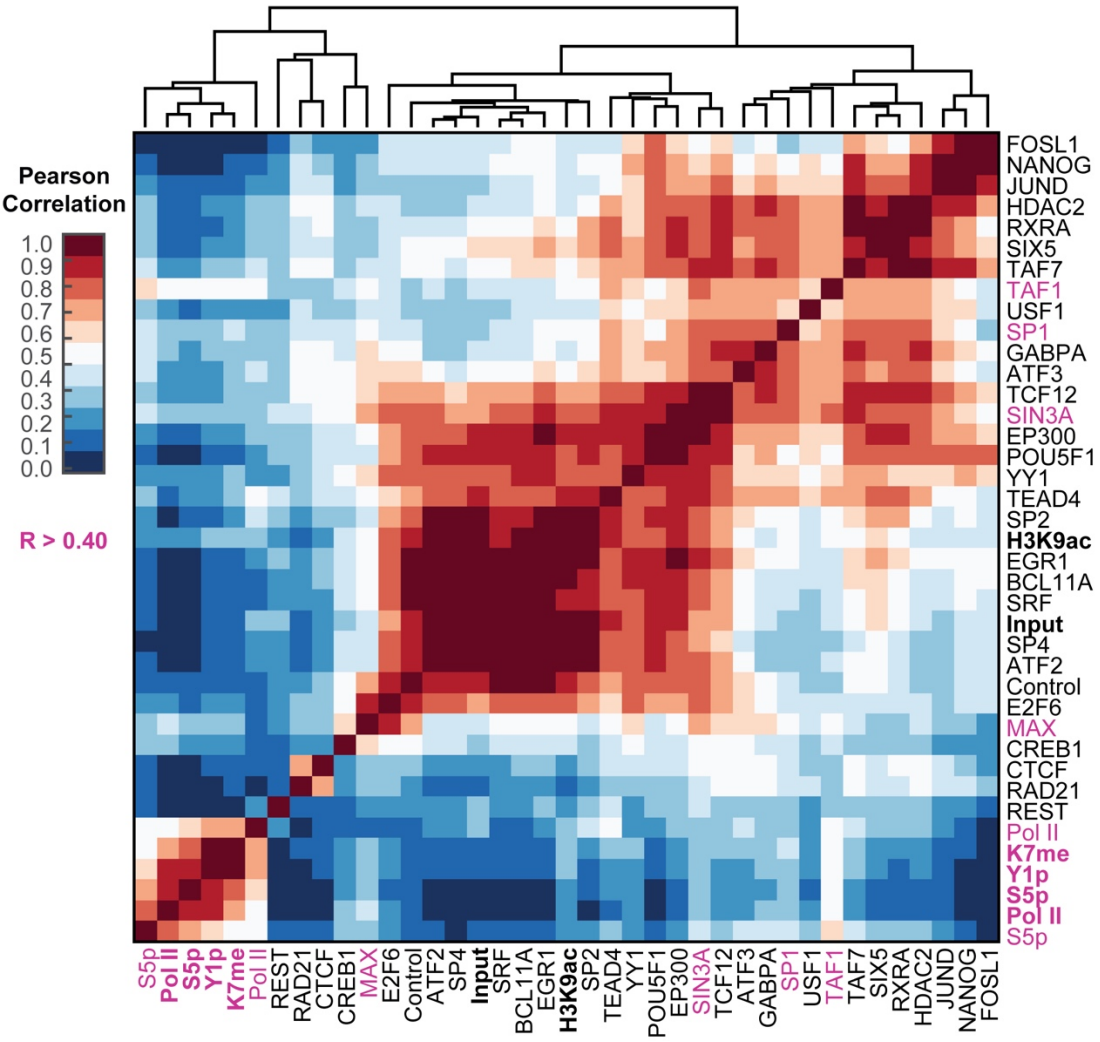



**Figure I.13 Correlation of Methylated Pol II and Transcription Factors**

Methylated Pol II co-localizes with known chromatin modifiers and binding proteins. Heat map showing the Pearson correlation coefficients of ChIP-Seq performed in this study (bold) versus the ChIP-Seq from the Meyers Lab. Proteins with associated correlation coefficients between K7me ChIP-Seq greater than 0.4 are highlighted in pink.

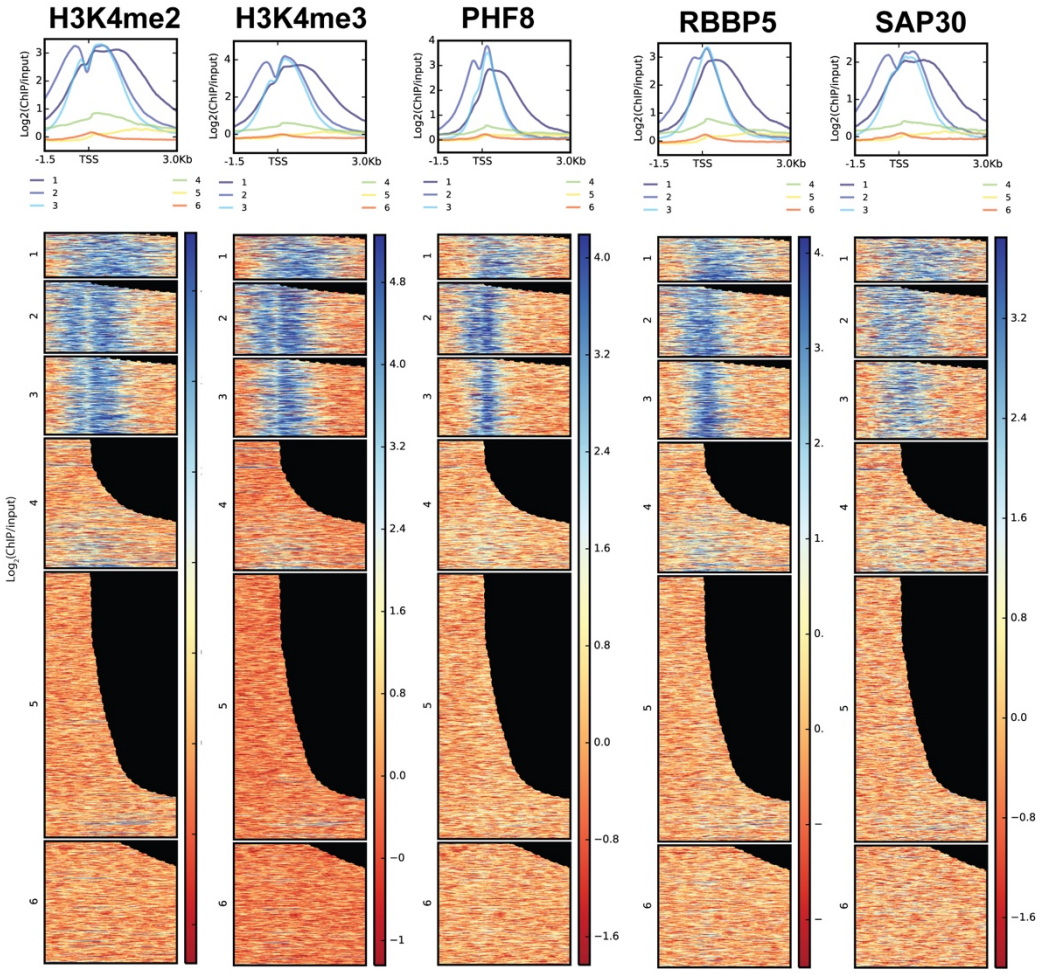



**Figure I.14 ChIP-Seq of Correlated Factors**

Heat maps and metagene plots for ChIP-Seq signal of proteins identified as highly correlated in enrichment (Pearson correlation coefficient R > 0.45) to K7me, separated by the clusters defined in Figure I.4.

co-immunoprecipitates with the distal region of the CTD, and the phosphorylation state of the CTD does not affect the association. Mutations in PHF8 link the protein to X linked mental retardation and cleft lip/cleft palate [30]and aberrant expression of specific genes involved in neuronal differentiation [31]. ChIP of SAP30, SIN3A and MAX also correlate moderately well with K7me ChIP ($R > 0.40$). These are all constituents of the Rpd3 histone de-acetylase complex, which is involved with transcriptional repression [32]. It remains to be tested whether PHF8 and the Rpd3 complex, actually "read" K7me on the CTD and co-activate or repress transcription in developmentally critical genes.

### I.7 Discussion

The non-consensus CTD can be functionally replaced with consensus repeats without apparent consequence to cellular growth and propagation [33]. Paradoxically, the non-consensus CTD residues are completely conserved among all members of the vertebrate subphylum, which strongly suggests they serve an important function unique to the biology that has evolved with backbone-possessing animals. Rather than being essential for cellular viability during the senescent stages of cellular division, the non-consensus CTD may regulate transcription of genes key to differentiation of pluripotent cells and early organismal development [19, 33, 34]. Here, we comprehensively characterized the genomic regions enriched for binding of the non-consensus CTD modification K7me in hESCs. We found the CTD of chromatin-bound Pol II to be constitutively methylated in hESCs at genes and gene regulatory regions, such as enhancers and insulators. The pervasive methylation of Pol II at K7 may prevent subsequent ubiquitinylation, which has been shown to result in the degradation of Pol II [15].

Likely, methylated K7 residues interact with transcription-regulatory proteins in conjunction with modification of adjacent or the same K7 residues. Though detection of

K7me diminishes in conjunction with hyper-phosphorylation with S5p, we found Y1p to co-occur with K7me. The mutation of K7 to arginine in mouse embryonic fibroblasts results in reduced response to epidermal growth factor mediated gene expression of developmentally important genes, which were shown to harbor paused Pol II acetylated at K7 immediately after the promoter region [17]. Intuitively, K7ac and K7me are mutually exclusive CTD modifications on the same K7 residue. As such, K7 deacetylases and methyltransferases presumably act early during transcription to pattern Pol II K7 for active transcription. In this study, we show K7me is detectable at the TSS, peaking over the region relatively depleted of K7ac, in both active and poised genes. Methylation of K7 may prevent premature K7ac and stalling of Pol II at actively transcribed genes.

Finally, the identity of the methyltransferases, demethylases, and methyl-lysine readers of K7me remain elusive. Transcriptional activator p300/CBP, which was found to acetylate K7 [17], may potentially interact with K7me through its PHD finger domain, which has yet to be assigned a function. Through our ChIP correlation analysis, we identified PHF8, a known CTD-binding protein and demethylase, as a potential K7me interactor. Additionally, we found binding of components of the Rpd3 repressive, de-acetylase complex, to correlate with K7me. p300/CBP and Rpd3 complexes acetylate and de-acetylate histones, respectively, to finely regulate transcription throughout embryonic development [35]. The potential for K7-modifying enzymes to mirror histone-modifying enzymes is an attractive model that is suggested by our analyses but remains to be tested.

## I.8 Materials and Methods

*Cell Culture Conditions*

Human embryonic stem cells (line H1) were obtained from WiCell and cultured in E8 medium. E8 media was made with DMEM/F-12, HEPES (ThermoFisher), 64 µg/mL L-ascorbic acid 2-phosphate (Sigma), 14 ng/mL sodium selenite (Sigma), transferrin (Sigma), 20 µg/mL insulin (ThermoFisher), 100 ng/mL FGF2 (Peprotech), and 2 ng/mL TGF-β1 (Peprotech). Cells were maintained on Matrigel (BD Biosciences) and passaged with 0.5 mM EDTA into E8 media containing Y-27632 (Selleck).

*Immunocytochemistry and Dot Blotting*

All the rat monoclonal antibodies used in this study, except Pol II N-20 (Santa Cruz Biotechnology), were graciously provided by Dirk Eick. For immunocytochemistry experiments, cells were fixed with 4% formaldehyde for 10 min, permeabilized with 0.25% Triton X-100 for 10 min, and blocked with 3% BSA in PBS for 30 min, all at room temperature. Primary antibodies were added to cells with the following dilutions in 1X PBS: Pol II 8WG16 (1:50), K7me 1F5 (1:5), Y1p 8G5 (1:20), and S5p 3E8 (1:200). Cells were incubated overnight at 4 °C with the primary antibody. Secondary AlexaFluor488-conjugated antibody (Santa Cruz Biotechnology) was added at a dilution of 1:250 in 1X PBS for 1 hr at room temperature. Between all steps, cells were rinsed twice in 1X PBS. Before imaging on a Nikon ECLIPSE TE2000-U confocal microscope, 0.3 µM DAPI solution and a drop of VECTASHIELD mounting media (Vector Laboratories) was added to the cells. For immunoblotting experiments, we directly spotted CTD peptides (X company) onto nitrocellulose membranes (GE Healthcare) and processed it as described previously [36]. A 1:100 dilution of primary rat antibody and 1:12,500 dilution of secondary HRP-conjugated secondary (Santa Cruz Biotechnology) was used.

*ChIP-Seq of Pol II and CTD Modifications*

H1 cells were fixed in 1.5% formaldehyde with gentle agitation for 15 min at room temperature. To quench the reaction, cells were incubated in 0.125 M glycine for 5 min. Cells were washed with cold PBS and pelleted via centrifugation. The pelleted cells were flash frozen and stored at -80 °C before lysis. Frozen cells were resuspended in ChIP Buffer 1 (10 mM HEPES, 10 mM EDTA, 0.5 mM EGTA, 0.25% Triton X-100, pH 6.5) and 1X Protease Inhibitor Cocktail (Roche) at a volume of 125 µL to $10^7$ cells, spun at 4000 rpm for 5 min at 4 ºC. Pelleted cells were resuspended in 125 µL of ChIP Buffer 2 (200 mM NaCl, 10 mM HEPES, 1 mM EDTA, 0.5 mM EGTA, pH 6.5). Cells were pelleted again and lysed in 175 µL of Lysis Buffer (50 mM Tris-HCl, 10 mM EDTA, 0.5% Empigen BB, 1% SDS) per $10^7$ cells. Cells were sonicated with a Misonix sonicator at 60% power for 32 min total pulse time, 10 sec 'ON' and 10 sec 'OFF'. Spinning samples at 13,000 rpm for 10 min retrieved sheared complexes in the supernatant. A small portion was collected for Input and the ChIP protocol was performed in parallel sans antibodies. For Pol II ChIP with N-20 antibody (Santa Cruz Biotechnology), 25 µL was used for $10^8$ cells in 1750 µL volume. For Pol II modification ChIPs (*i.e.*K7me, Y1p, S5p), 50 µL of antibody sera was used for $10^8$ cells. Samples were incubated overnight with rotation at 4 ºC. Each sample received 200 µL of pre-cleared magnetic Protein G Dynabeads (Life Technologies) and was incubated for 1 hour at 4 ºC. The beads and antibody-bound chromatin complexes were separated from the supernatant by using a magnetic tube rack. Beads were washed sequentially with Wash Buffer 1 (2 mM EDTA, 20 mM Tris-HCl, 0.1% SDS, 1% Triton X-100, 150 mM NaCl), Wash Buffer 2 (Wash Buffer 1, except 500 mM NaCl), and Wash Buffer 3 (1 mM EDTA, 10 mM Tris-HCl, 250 mM LiCl, 1% deoxycholate, 1% NP-40), and four times with Wash Buffer 4 (10 mM Tris-HCl, 1 mM EDTA). Washed beads were resuspended in 300 µL of Elution Buffer (1% SDS, 0.1 M $NaHCO_3$), and samples were incubated in a 65 ºC water bath for six hours with 12 µL of

5 M NaCl added to reverse-crosslink DNA. DNA was purified with EconoSpin DNA purification columns (Epoch Life Science) and stored at 4 ºC until barcoding and sequencing on an IlluminaHighSeq 2000.

*Bioinformatics Analysis of ChIP-Seq and RNA-Seq Data*

Reads were mapped to the human genome (version *hg19*) with Bowtie 2 [37] using default settings after adapter and barcode sequences were trimmed with FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Alignment files were converted to binary format with SAMtools [38], and normalized to Input or Pol II ChIP and converted to the bigwig format with deepTools using the *bamCompare* function [39]. Pearson correlation coefficients, heat maps, and metagene profiles were all calculated with deepTools' *bamCorrelate*, *computeMatrix* (TSS reference point, 10 bp bins, flanked by 1.5 and 3.0 kbp), *heatmapper*, and *profiler* functions, respectively.
Annotated gene lists for human were as described in [23], and a list of homologous mouse genes was generated by finding mouse genes with over 10% sequence homology to human genes using UCSC's liftOver program [26] and mouse genome version *mm9*. Chromatin states in H1 were categorized based on the ChromHMM states described in [40]. K-means clustering was performed with Cluster 3.0 [41] and visualized with Java TreeView[42]. Correlation coefficients from Figure I.2, I.12, and I.13 were clustered and plotted with MATLAB's (MathWorks, Inc.) *clustergram* function. ChIP-Seq signal at specific genes were visualized on the UCSC Genome Browser [43] and Broad Institute's Integrative Genomics Viewer [44].

See **Table I.3** for information regarding datasets analyzed from previous publications.

**Table I.3 Datasets from ENCODE Project [20] or Other Publications Analyzed in this Study**

| Assay | Laboratory | GEO/SRA Accession | Sample |
|---|---|---|---|
| ChIP-Seq | Bernstein [20] | GSE29611 | H1 hESC chromatin marks and modifiers |
| ChIP-Seq | Meyers [20] | GSE32465 | H1 hESC transcription factors |
| ChIP-Seq | Ott [17] | SRX338012 | mESC K7ac |
| RNA-seq | Belmonte [23] | GSE54969 | H1 hESC RNA expression |
| RNA array | Capra [19] | GSE66088 | NIH/3T3 expression profiling |

## I.9 References


1. Zhang, D.W., et al., *Emerging Views on the CTD Code.* Genet Res Int, 2012. **2012**: p. 347214.

2. Sun, M., et al., *A tandem SH2 domain in transcription elongation factor Spt6 binds the phosphorylated RNA polymerase II C-terminal repeat domain (CTD).* J Biol Chem, 2010. **285**(53): p. 41597-603.

3. Corden, J.L., *RNA polymerase II C-terminal domain: Tethering transcription to transcript and template.* Chem Rev, 2013. **113**(11): p. 8423-55.

4. Tee, W.W., et al., *Erk1/2 activity promotes chromatin features and RNAPII phosphorylation at developmental promoters in mouse ESCs.* Cell, 2014. **156**(4): p. 678-90.

5. Burton, Z.F., *The Old and New Testaments of gene regulation: Evolution of multi-subunit RNA polymerases and co-evolution of eukaryote complexity with the RNAP II CTD.* Transcription, 2014. **5**.

6. Chapman, R.D., et al., *Molecular evolution of the RNA polymerase II CTD.* Trends Genet, 2008. **24**(6): p. 289-96.

7. Toll-Riera, M.C., Jose; Albà, M. Mar *Accelerated Evolution of Genes of Recent Origin*, in *Evolutionary Biology from Concept to Application*, P. Pontarotti, Editor. 2008, Springer Berlin Heidelberg. p. 45-59.

8. Yang, C. and J.W. Stiller, *Evolutionary diversity and taxon-specific modifications of the RNA polymerase II C-terminal domain.* Proc Natl Acad Sci U S A, 2014. **111**(16): p. 5920-5.

9. Hintermair, C., et al., *Threonine-4 of mammalian RNA polymerase II CTD is targeted by Polo-like kinase 3 and required for transcriptional elongation.* EMBO J, 2012. **31**(12): p. 2784-97.

10. Descostes, N., et al., *Tyrosine phosphorylation of RNA polymerase II CTD is associated with antisense promoter transcription and active enhancers in mammalian cells.* Elife, 2014. **3**: p. e02105.

11. Hsin, J.P., et al., *RNAP II CTD tyrosine 1 performs diverse functions in vertebrate cells.* Elife, 2014. **3**: p. e02112.

12. Egloff, S., et al., *Serine-7 of the RNA polymerase II CTD is specifically required for snRNA gene expression.* Science, 2007. **318**(5857): p. 1777-9.

13. Egloff, S., et al., *The integrator complex recognizes a new double mark on the RNA polymerase II carboxyl-terminal domain.* J Biol Chem, 2010. **285**(27): p. 20564-9.

14. Sims, R.J., 3rd, et al., *The C-terminal domain of RNA polymerase II is modified by site-specific methylation.* Science, 2011. **332**(6025): p. 99-103.

15. Li, H., et al., *Wwp2-mediated ubiquitination of the RNA polymerase II large subunit in mouse embryonic pluripotent stem cells.* Mol Cell Biol, 2007. **27**(15): p. 5296-305.

16. Voss, K., et al., *Site-specific methylation and acetylation of lysine residues in the C-terminal domain (CTD) of RNA polymerase II.* Transcription, 2015: p. 0.

17. Schroder, S., et al., *Acetylation of RNA polymerase II regulates growth-factor-induced gene transcription in mammalian cells.* Mol Cell, 2013. **52**(3): p. 314-24.

18. Litingtung, Y., et al., *Growth retardation and neonatal lethality in mice with a homozygous deletion in the C-terminal domain of RNA polymerase II.* Mol Gen Genet, 1999. **261**(1): p. 100-5.

19. Simonti, C.N., et al., *Evolution of lysine acetylation in the RNA polymerase II C-terminal domain.* BMC Evol Biol, 2015. **15**: p. 35.

20. ENCODE, *An integrated encyclopedia of DNA elements in the human genome.* Nature, 2012. **489**(7414): p. 57-74.

21. Adelman, K. and J.T. Lis, *Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans.* Nat Rev Genet, 2012. **13**(10): p. 720-31.

22. Brookes, E. and A. Pombo, *Modifications of RNA polymerase II are pivotal in regulating gene expression states.* EMBO Rep, 2009. **10**(11): p. 1213-9.

23. Kurian, L., et al., *Identification of novel long noncoding RNAs underlying vertebrate cardiovascular development.* Circulation, 2015. **131**(14): p. 1278-90.

24. Mayer, A., et al., *CTD tyrosine phosphorylation impairs termination factor recruitment to RNA polymerase II.* Science, 2012. **336**(6089): p. 1723-5.

25. Pan, G., et al., *Whole-genome analysis of histone H3 lysine 4 and lysine 27 methylation in human embryonic stem cells.* Cell Stem Cell, 2007. **1**(3): p. 299-312.

26. Karolchik, D., et al., *The UCSC Table Browser data retrieval tool.* Nucleic Acids Res, 2004. **32**(Database issue): p. D493-6.

27. Stock, J.K., et al., *Ring1-mediated ubiquitination of H2A restrains poised RNA polymerase II at bivalent genes in mouse ES cells.* Nat Cell Biol, 2007. **9**(12): p. 1428-35.

28. Guenther, M.G., et al., *A chromatin landmark and transcription initiation at most promoters in human cells.* Cell, 2007. **130**(1): p. 77-88.

29. Fortschegger, K., et al., *PHF8 targets histone methylation and RNA polymerase II to activate transcription.* Mol Cell Biol, 2010. **30**(13): p. 3286-98.

30. Laumonnier, F., et al., *Mutations in PHF8 are associated with X linked mental retardation and cleft lip/cleft palate.* J Med Genet, 2005. **42**(10): p. 780-6.

31. Qiu, J., et al., *The X-linked mental retardation gene PHF8 is a histone demethylase involved in neuronal differentiation.* Cell Res, 2010. **20**(8): p. 908-18.

32. Kadosh, D. and K. Struhl, *Targeted recruitment of the Sin3-Rpd3 histone deacetylase complex generates a highly localized domain of repressed chromatin in vivo.* Mol Cell Biol, 1998. **18**(9): p. 5121-7.

33. Chapman, R.D., M. Conrad, and D. Eick, *Role of the mammalian RNA polymerase II C-terminal domain (CTD) nonconsensus repeats in CTD stability and cell proliferation.* Mol Cell Biol, 2005. **25**(17): p. 7665-74.

34. Corden, J.L., et al., *A unique structure at the carboxyl terminus of the largest subunit of eukaryotic RNA polymerase II.* Proc Natl Acad Sci U S A, 1985. **82**(23): p. 7934-8.

35. McDonel, P., et al., *Sin3a is essential for the genome integrity and viability of pluripotent cells.* Dev Biol, 2012. **363**(1): p. 62-73.

36. Tietjen, J.R., et al., *Chemical-genomic dissection of the CTD code.* Nat Struct Mol Biol, 2010. **17**(9): p. 1154-61.

37. Langmead, B. and S.L. Salzberg, *Fast gapped-read alignment with Bowtie 2.* Nat Methods, 2012. **9**(4): p. 357-9.

38. Li, H., et al., *The Sequence Alignment/Map format and SAMtools.* Bioinformatics, 2009. **25**(16): p. 2078-9.

39. Ramirez, F., et al., *deepTools: a flexible platform for exploring deep-sequencing data.* Nucleic Acids Res, 2014. **42**(Web Server issue): p. W187-91.

40. Ernst, J. and M. Kellis, *Discovery and characterization of chromatin states for systematic annotation of the human genome.* Nat Biotechnol, 2010. **28**(8): p. 817-25.

41. de Hoon, M.J., et al., *Open source clustering software.* Bioinformatics, 2004. **20**(9): p. 1453-4.

42. Saldanha, A.J., *Java Treeview--extensible visualization of microarray data.* Bioinformatics, 2004. **20**(17): p. 3246-8.

43. Kent, W.J., et al., *The human genome browser at UCSC.* Genome Res, 2002. **12**(6): p. 996-1006.

44. Robinson, J.T., et al., *Integrative genomics viewer.* Nat Biotechnol, 2011. **29**(1): p. 24-6.